

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE

ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

DEPARTEMENT HYDRAULIQUE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

PROJET FIN D'ETUDE
POUR L'OBTENTION DU DIPLOME
D'INGENIEUR D'ETAT EN HYDRAULIQUE

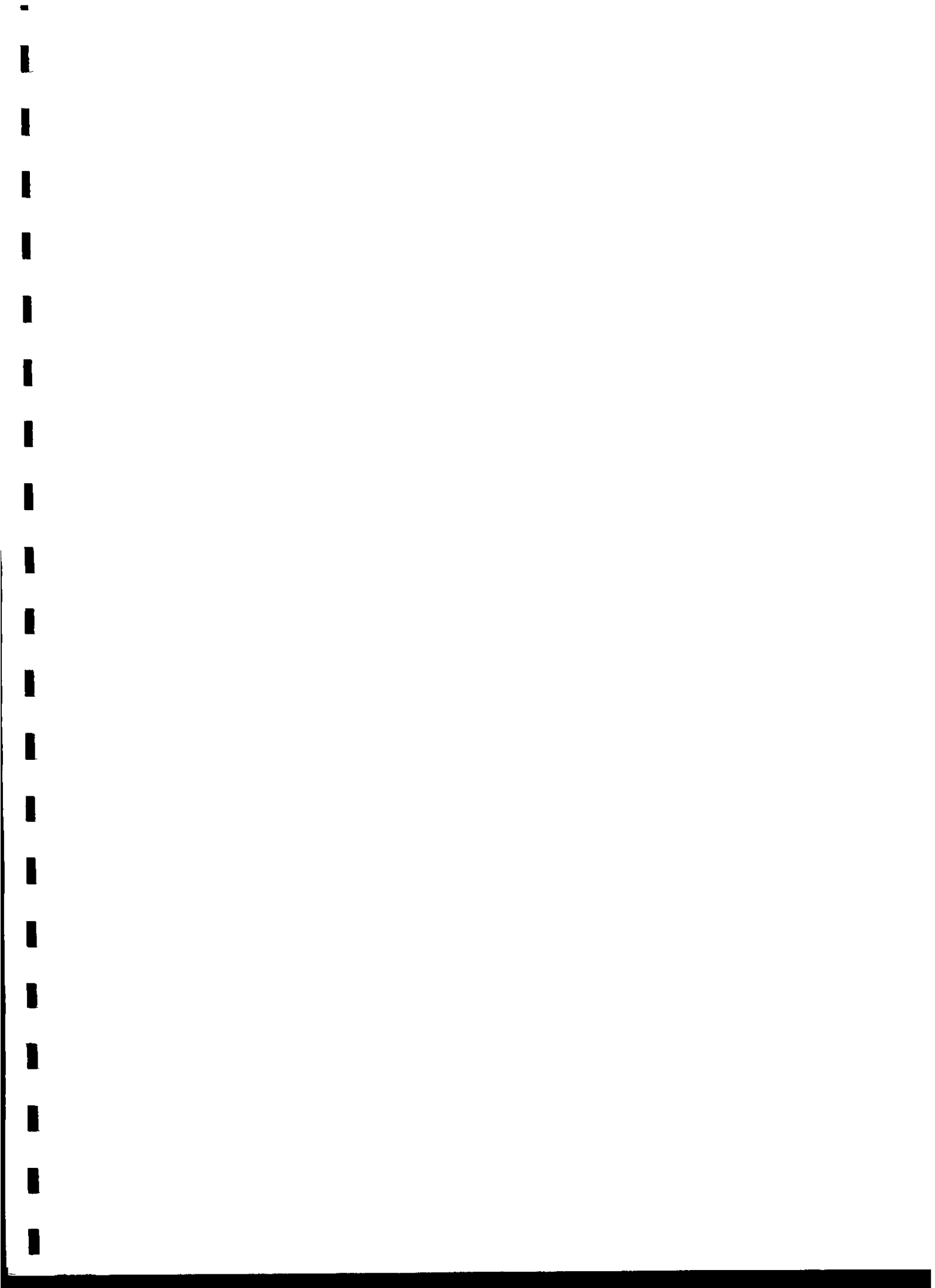
THEME

Prévision de la pluviométrie
par combinaison de l'ACP
et des chaînes de MARKOV

Réaliser par :
M^{lle} : Amel RAHAI

Promoteur :
M^r A. BERMED

PROMOTION 2002



Résumé :

Afin de mettre en évidence cette tendance à la baisse de la pluviométrie, on a effectué un traitement statistique sur les données de la station d'Alger pour comprendre globalement le phénomène étudié et contribuer mieux gérer les ressources hydriques existantes en modélisent cet aléa par des techniques mathématiques telles que l'analyse des composantes principales et les chaînes de Markov.

Mots clés :

Analyse en composante principale, chaîne de Markov, ressources hydriques.....

Abstract :

In order to highlight this downward trend of pluviometry, one carried out a statistical processing on the data of the station of Algiers to include/understand the phenomenon overall studies and to better contribute to manage the existing hydrous resources model this risk by mathematical techniques of it such that the analysis of the principal components and the chains of Markov.

Key words:

Analyze in principal component, chain of Markov, hydrous resources.....

الملخص:

قمنا بمعالجة إحصائية على المعطيات المخصصة بالجزار من أجل فهم الظاهرة المدروسة و>| من أجل تحسين الموارد المائية المتوفرة بتطبيق المثال الخاص بالمتغيرة بتقنيات رياضية تحليل المكونات الرئيسية و سلسلة مالكوف .

الكلمات المفتاحية: التحليل بالمكونات الرئيسية ؛ سلسلة مالكوف ؛ الموارد المائية.

Sommaire

Introduction	
Chapitre I : Recherche bibliographique	3
Chapitre II : Collecte et critique des données	
♦ II.1 Introduction	9
♦ II.2 Inventaire des données	9
♦ II.3 Critique préliminaire des données	10
- II.3.1 Introduction	10
- II.3.2 Méthodes de comblement de lacune	11
- II.3.3 Détection des valeurs singulières	13
♦ II.4 Traitement statistique des données	14
- II.4.1 Test d'Homogénéité	14
- II.4.2 Test d'indépendance	15
- II.4.3 Analyse des résultats	16
Chapitre III : Théories de l'ACP et des chaînes de Markov	
♦ III.1 Généralité sur les modèles	18
♦ III.2 Analyse en composantes principales	19
- III.2.1 Théorie	19
- III.2.2 Application	27
♦ III.3 Les chaînes de Markov	28
- III.3.1 Introduction	28
- III.3.2 Ordre des chaînes de Markov	29
- III.3.3 Les modèles multiclassés	31
- III.3.4 Les modèles multiclasse saisonnier	32
Chapitre IV : Analyse du régime pluviométrique	
♦ IV.1 Introduction	34
♦ IV.2 Evolution de la pluviométrie	34
- IV.2.1 Pas de temps annuel	34
- IV.2.2 Pas de temps mensuel	38
- IV.2.3 Pas de temps journalier	43
♦ IV.3 Caractérisation des années sèches	46
- IV.3.1 Définition de les années sèche	46
- IV.3.2 Variation de la séquence sèche	49
- IV.3.3 Détermination des années sèches	49
♦ IV.4 Etude de la pluviométrie par les chaînes de Markov	54
♦ IV.5 Conclusion	

Chapitre V : le modèle de prévision

♦ V.1 Introduction	58
♦ V.2 Le modèle de prévision	58
- V.2.1 Calage du modèle	59
- V.2.2 Test du modèle	59
♦ V.3 Critère de validation	59
♦ V.4 Elaboration du modèle de prévision	60
- V.4.1 1 ^{ère} approche	63
- V.4.2 1 ^{ème} approche	76
♦ V.5 Conclusion	80
Conclusion générale.	81

Introduction

Les phénomènes météorologiques et climatiques constituent un danger potentiel pour la vie et l'environnement. Ils le sont particulièrement lorsqu'ils revêtent une intensité extrême, inhabituelle ou un caractère persistant (orages, inondations ou sécheresses).

L'étude de l'information météorologique devient donc une nécessité qui permet aux services concernés de mieux contribuer à assurer la sécurité, des biens et des personnes et de mieux maîtriser les apports qui engendrent une meilleure planification du développement.

La prévision des facteurs climatiques prépondérants revêt une extrême importance, en particulier pour la gestion des ressources en eau, ainsi que pour la maîtrise des activités agricoles.

L'Algérie, à l'instar des pays de l'Afrique du nord est confrontée ces dernières années, à un déficit pluviométrique qui affecte négativement la ressource hydrique sur qui la demande croît sans cesse.

✕ Afin de mettre en évidence cette tendance à la baisse de la pluviométrie, on a effectué un traitement statistique sur les données de la station d'Alger pour comprendre globalement le phénomène étudié et contribuer à mieux gérer les ressources hydriques existantes en modélisant cet aléa par des techniques mathématiques telles que l'analyse des composantes principales et les chaînes de Markov.

Pour cela, on a adopté la démarche suivante :

- Après une introduction, le chapitre un est consacré à la recherche bibliographique sur les travaux réalisés dans le domaine de l'analyse et de la prévision de la pluviométrie.
- La collecte et la critique des données sont étudiées dans le chapitre deux.
- Dans le chapitre trois, on a présenté les théories de la de modélisation basées sur l'analyse en composantes principales (ACP) et les chaînes de Markov.
- L'étude du régime pluviométrique et l'analyse des précipitations aux différents pas du temps ont fait l'objet du quatrième chapitre.
- Finalement, dans le dernier chapitre, on a procédé au développement d'un modèle de prévision basé sur l'ACP et le modèle Markovien multiclasse, appliqué sur des variables brutes et normées, au pas du temps mensuel.

Recherche bibliographique

Aujourd'hui, la sécheresse exerce sur les pays non préparés à cet aléa climatique un impact économique et sanitaire qui risque de bouleverser la vie dans certaines régions du monde, et devient l'une des catastrophes climatiques majeures qui affecte aussi l'environnement.

Le rapport de la banque mondiale montre que [1] :

- 40 % de la population mondiale vit dans un état de pénurie d'eau chronique.
- Plus de 80 pays manqueraient d'eau, et plus d'un milliard d'hommes n'aurait pas accès à l'eau potable.
- Au rythme actuel de la croissance démographique, la demande en eau doublerait tous les 20ans.
- La situation de la Méditerranée serait grave : la moitié des 20 pays de cette zone aurait une consommation dépassant leurs ressources réelles au cours du siècle prochain.

Le concept de la sécheresse diffère selon le domaine considéré, on distingue plusieurs types [2] :

- En météorologie : la sécheresse est caractérisée par un déficit, une absence ou une mauvaise distribution des pluies par rapport à la normale (la moyenne).
- La sécheresse dite pédologique : est caractérisée par la disparition de l'eau utile pour le sol.
- La sécheresse hydrologique (phréatique) : est liée à la diminution ou même la disparition de l'ensemble des nappes phréatiques et la baisse des niveaux des cours d'eau, ce qui entraîne la surexploitation du potentiel existant, la déplétion des nappes, leurs assèchements progressifs et leurs pollutions (il n'y a pas de renouvellement des nappes).
- La sécheresse agricole (biovégétale) : est un déficit pluviométrique qui entraîne la dégradation du couvert végétal ou même la disparition des végétaux, ainsi que la réduction agricole (récolte inexistante ou insuffisante).
- La sécheresse socio – économique : se traduit par la pénurie d'eau pour les populations avec toutes les conséquences connues sur le plan social et sanitaire en matière d'alimentation en eau potable et de mode de vie, cette sécheresse entraîne aussi le ralentissement ou même l'arrêt d'unités de production industrielles, ce qui influe négativement sur le développement économique et social.

Les grandes souffrances qu'a entraîné ce phénomène dès son déclenchement en Afrique sahélienne en 1968 et encore plus en s'étendant par la suite à d'autres parties du monde (Afrique du sud, Afrique de l'Est, Brésil et Australie) ont mené à de vastes programmes de surveillance et de recherches sur le climat.

L'évolution pluviométrique, par rapport à l'indice de normalité dans quelques régions méditerranéennes a fait l'objet de nombreuses études.

L'analyse des données des précipitations de ces régions a montré que [3] :

- Une occurrence de périodes de déficit en précipitation dans une très vaste zone durant les années 1944, 1945, 1947, 1970, 1973, 1980, 1981, 1982, 1989 et 1990 où la majorité des régions était touchée.
- Certaines zones semblent plus affectées que d'autres et notamment :
 - L'Algérie et l'Espagne dans la partie occidentale.
 - La Grèce et le Moyen Orient dans la partie orientale.
- Une quasi- généralisation de la sécheresse pour la majorité de pays du pourtour de la Méditerranée semble apparaître depuis le début de la décennie 80.
- La sécheresse est chronique dans certains zones (les périodes de déficit chroniques se sont produites presque aux même époques).

L'étude des précipitations et des pressions de surface dans le bassin méditerranéen au pas de temps saisonniers de 1915- 1916 à 1987 - 1988, basée sur une ACP avec rotation (ACPR) a montré que les régions pluviométriques étudiées présentent pour toutes les saisons des évolutions tendancielle non significatives à l'exception de la Grèce où les pluies de la saison pluvieuse diminuent de façon significative [4].

Le résultat de l'étude des précipitations en utilisant 74 stations également réparties sur le Nord algérien a donné lieu aux constatations suivantes [5] :

Une occurrence de sécheresse sévère :

1. Durant des années très isolées (1937- 1961- 1970) avec particulièrement une saison agricole 1960- 1961 très déficitaire où la station de Guelma a enregistré 51% de déficit par rapport à la normale.

2. Durant deux longues périodes :

- De 1943 à 1948, il y a eu des répercussions importantes sur les récoltes et le cheptel (durant la saison agricole 1944/ 1945) les pluies mesurées à Oran et Mascara ont été à peine de l'ordre de 45% par rapport à la normale et les rendements de céréales en cette année ont été inférieurs à 25% des rendements moyens.

-A partir de 1980, elle a en fait commencé à certains endroits en 1973 et n'a pas cessé depuis de se manifester à des intervalles n'excédant pas deux années consécutives bien qu'elle était entrecoupée d'année de bonne pluviométrie (1979, 80, 85, 86).

Entre ces deux périodes de déficit : s'étale une période à pluviométrie annuelle moyenne souvent supérieur à la normale avec quelques perturbations limitées dans l'espace.

La contribution à l'étude de l'évolution de la pluviosité dans quelques stations steppiques algériennes (Sud-Oranais et Sud-Algérois) basée sur des tests de tendances non paramétriques (test de Mann Withey, Wilcoxon et Spearman) a montré une tendance à l'assèchement qui diffère d'une région à l'autre [6].

En outre, l'analyse du déficit pluviométrique qui touche actuellement l'Algérie a montré que les années de sécheresses sont principalement occasionnées par une diminution des pluies d'hiver, et une baisse de la fréquence des fortes précipitations, il a été remarqué que certaines années de sécheresse sont influencées par le phénomène EL NINO, les résultats ont montré que lors d'une épisode EL NINO, les pluies d'hiver présentent une forte probabilité d'être inférieures à la moyenne [7].

Dans la recherche des causes de la sécheresse, plusieurs études sont menées afin d'établir la liaison entre les indicateurs météorologiques et ce phénomène, pour pouvoir fournir des prévisions acceptables [3] :

- Les anomalies de la circulation générale :

L'Oscillation Nord Atlantique (ONA) :

L'équipe du Dr. PJ Lamb a établie la première liaison de la sécheresse au Maroc entre 1979 et 1984 avec les anomalies de la circulation générale , se basant sur les travaux précédents de Walker (1924) et Walker et Bliss (1932) et ceux de Rogers et Vanloon, cette recherche identifie des corrélations acceptables entre un indice des précipitations au Maroc et un indice de l'Oscillation Nord Atlantique (ONA) qui est définie comme étant la résultante de la valeur de pression inférieure à la normale à Ponta Delgadas (Açores) et de la valeur de la pression supérieure à la normale à Akureyri (Islande).

Cette étude a montré que :

- Les pluies sont inversement liées à l'état de l'ONA.
- Les précipitations sont liées non seulement à la position et la puissance de l'anticyclone (négativement) mais aussi aux mêmes caractéristiques de la dépression d'Islande (positivement).
- Les coefficients de corrélations indiquent clairement l'influence d'autres mécanismes sur les pluies au Maroc.

Les corrélations entre l'indice de l'ONA durant les mois précédents les pluies hivernales ne permettent pas d'établir des prévisions, mais il a été remarqué que les anomalies des températures de surface de l'océan (SST) dans les deux extrêmes de l'ONA d'hiver existent déjà durant l'été précédent et persistent durant l'automne.

L'oscillation méditerranéenne :

La tendance à la diminution des pluies en Italie et les signes de désertification ont conduit à un second concept : l'Oscillation Méditerranéenne.

L'analyse de cet indice (différence de géopotentiel à la surface 500HPA entre Alger (Méditerranée occidentale) et le Caire (Méditerranée orientale) indiquerait une plus fréquente incursion de l'anticyclone subtropical dans la région. D'autres indicateurs ont été mis en évidence tels que la position de certaines caractéristiques saisonnières de la circulation atmosphérique.

▪ La température de surface de l'océan (SST)

Les derniers développements de la prévision saisonnière montrent le caractère persistant de la température de surface de l'océan et sa forte corrélation avec les précipitations.

Les indicateurs météorologiques de la sécheresse cités ci dessus constituent la justification scientifique d'un projet Algéro- Marroco- Tuniso- Français qui était lancé avec un appui financier de la Communauté Economique Européenne dans le cadre du programme «AVICENNA » [8].

Bien que les causes de la sécheresse ne sont pas totalement expliquées, notamment les événements de l'ENSO (El Nino Southern Oscillation), l'assèchement de l'humidité du sol et le comportement non linéaire des perturbations du système climatique, elles continuent d'être analysées et font l'objet de ce projet de prévision mensuelle et saisonnière du régime des précipitations sur le bassin Méditerranéen.

Le programme de travail comprend deux parties essentielles :

- Procéder à des études de corrélations entre l'ensemble des indicateurs ci dessus indiqués et les précipitations.
- Contribuer à l'amélioration des modèles globaux de prévision numérique existants (ARPEGE) par la validation des prévisions historiques en temps réel qui se feront par des centres mondiaux.

Les services météorologiques commencent à mettre en expérimentation des méthodologies sur la prévision mensuelle et saisonnière des précipitations qui devient un outil important pour le développement [9].

Une quantification de la relation entre les précipitations sur le nord d'Algérie et l'anomalie mensuelle de la SST a été concrétisée par l'élaboration de plusieurs modèles statistiques.

A cet effet, un modèle «océan – Atmosphère » basé sur la régression linéaire multiple a été élaborée, celui-ci consiste à établir une relation statistique entre les données de la température de la surface de l'océan Atlantique (SST) observées au cours de la période qui s'étale de 1964 à 1994 et l'indice régional de précipitations pour les régions du Centre et de l'Est de l'Algérie. Les décalages de temps considérés sont un, deux et trois mois.

Le champ de la SST a été réduit en procédant à une ACP classé par ordre d'importance du taux de corrélation par rapport à l'indice de pluie à l'aide de la méthode de sélection progressive ascendante. Les meilleures valeurs des coefficients de corrélation multiples ont été obtenues pour un décalage de deux et trois mois en utilisant dix composantes principales [10].

La recherche de l'existence d'une relation entre les oscillations méditerranéennes et la distribution géographique des précipitations au nord de l'Algérie a fait l'objet d'une étude basée sur les corrélations et l'analyse canonique [11].

Pour procéder à la reconstitution des données climatiques d'autres travaux ont été élaboré sur l'application d'une méthode d'interprétation automatique dite «méthode des réseaux neurones », les variables explicatives utilisées sont la longitude, latitude et l'altitude.

La comparaison des valeurs pluviométriques estimées par rapport aux données fournies par les stations de référence avoisinant les sites a montré une certaine cohérence pour les secteurs Centre et Ouest [12].

En outre, la régionalisation des paramètres climatiques est l'objectif d'autres chercheurs. La méthode utilisée est dite des vecteurs propres, cette analyse permet de délimiter les régions homogènes de point de vue origine des précipitations [13].

En conclusion, on peut dire que plusieurs approches ont été préconisé, la prévision de qualité reste le seul moyen qui permet d'amélioré les outils d'aide à une meilleure planification des opérations agricoles, par une connaissance préalable des tendances pluviométriques saisonnières et constituerait un outil d'assistance à une gestion rationnelle des ressources en eau en fournissant des prévisions objectives sur les réserves.

Chapitre II

Collecte et critique des données

Collecte et critique des données

II.1 Introduction :

Les données hydrologiques constituent la plupart du temps le seul outil dont dispose l'hydrologue pour étudier et comprendre le comportement d'un phénomène naturel, il est donc important de connaître à fond les techniques appropriées afin d'être en mesure d'obtenir le maximum d'information sur l'échantillon et de garantir la qualité des données, de ce fait, il est indispensable d'effectuer une analyse des données et un traitement statistique des séries chronologiques des pluies recueillies aux niveaux des stations de mesure

Il s'agit d'abord de compléter et de combler le manque des données, de détecter les observations anomalies, et puis de vérifier l'indépendance, l'homogénéité et la stationnarité des données disponibles.

II.2 Inventaire des données :

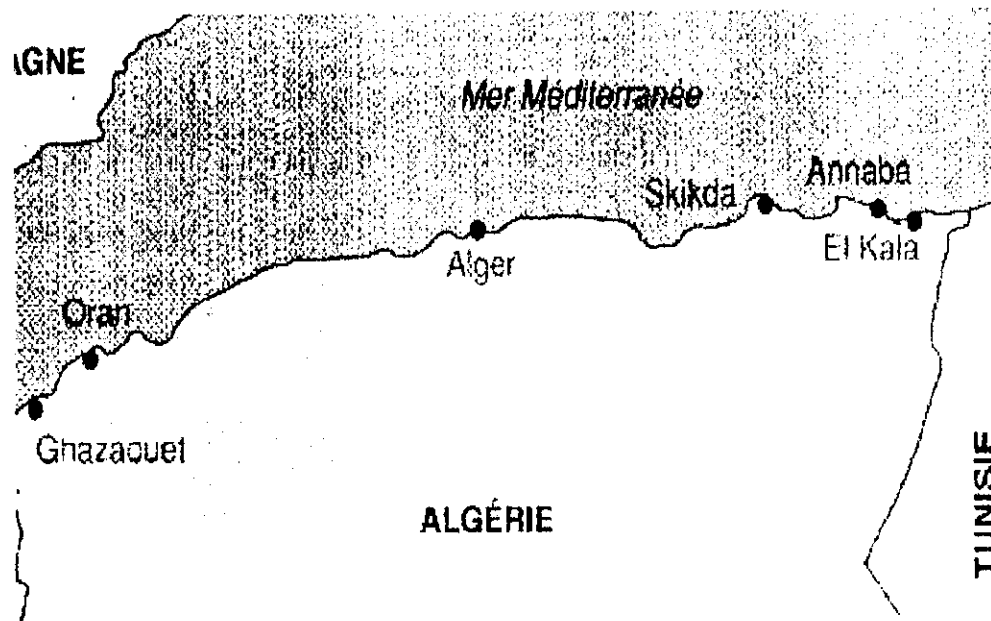
Les données utilisées nous ont été aimablement fournies par l'Office Nationale de la Météorologie (O.N.M). Un fichier des précipitations de la station d'Alger (Dar Baida), s'étalant sur 65 ans(1936-2000), au pas du temps journalier et sur 66 ans (1936-2001) au pas du temps mensuel ainsi qu'un fichier de températures de la même station au pas de temps mensuel mais s'étalant de 1952-2001.

Les caractéristiques de la station étudiée sont présentées dans le tableau II.1.

Tableau II.1 : caractéristiques de la station étudiée.

Station	Numéro de la station	Code	Latitude (Nord)	Longitude (E/W)	Altitude (mètre)
Alger	6	390	36°43	03°15 E	24

Figure II.1 Représentation de la station d'Alger



II.3 Critique préliminaire des données :

II.3.1 Introduction :

La mesure des précipitations peut se faire par :

- Le pluviomètre : c'est l'instrument le plus utilisé, il nécessite une intervention humaine pour obtenir une lecture directe.
- Le pluviographe : c'est un instrument qui permet non seulement l'enregistrement automatique de la hauteur des précipitations mais aussi sa répartition dans le temps (son intensité) par des diagrammes pluviographiques.

Les types de problèmes le plus souvent rencontrés lors du traitement des données en hydrologie sont [14]:

- Le manque des données : station non existante ou fichier incomplet.
- Non- continuité des données : manque d'enregistrement.
- La mauvaise qualité des données, qui peut dépendre de deux causes :
 - Données inconsistantes : lorsqu'elles proviennent d'une erreur systématique due à un instrument de mauvaise qualité ou non calibré ou à un mauvais réglage de l'appareil de mesure.
 - Données non homogènes : lorsqu'il se produit à un moment déterminé, un changement de l'emplacement de l'appareil qui affecte la mesure, ces erreurs dite accidentelles surviennent inopinément, et ne peuvent être prévues ni évitées, elles peuvent être dues à une erreur de manipulation ou une défectuosité imprévue de l'instrumentation (déplacement de la station de mesure ou changement d'appareil).

Afin de garantir la qualité des données, on doit suivre les étapes suivantes :

- Comblent les données manquantes dans les séries d'observations.
- Détecter et corriger les erreurs.
- Vérifier l'homogénéité et l'indépendance des données dans le temps et dans l'espace.

II.3.2 Méthodes de comblement de lacune [15] :

Dans un enregistrement continu de la pluviométrie d'une station, certaines données sont manquantes pour différentes périodes, afin de ne pas perdre la continuité de l'information, il est nécessaire d'estimer les valeurs manquantes.

Dans ce paragraphe, on va décrire les méthodes de comblement des lacunes, malgré que les deux dernières méthodes ne peuvent pas être utilisées dans ce cas car elles nécessitent plusieurs stations.

- Calcul des moyennes.
- Régression linéaire sur les variables mesurées.
- Simulation.

a) Calcul des moyennes :

Lorsqu'un fichier d'observations comporte moins de trois valeurs manquantes, le comblement de ces lacunes se fait par la méthode de la moyenne.

- On remplace une valeur annuelle manquante par la moyenne interannuelle de la série considérée.
- S'il s'agit d'une valeur mensuelle, on la remplace par la moyenne mensuelle des mois disponibles ou par la moyenne mensuelle interannuelle du mois considéré.

Exemple :

Soit la matrice d'un fichier mensuelle de n lignes (les années) et p colonnes (les mois).

$$[X] = \begin{pmatrix} X(1,1) \text{ inconnu} & \dots & X(1,j) & \dots & X(1,p) \\ X(2,1) & X(2,2) & \text{inconnu} & X(2,j) & \dots & X(2,p) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X(n,1) & \text{inconnu} & \dots & X(n,j) & \dots & X(n,p) \end{pmatrix}$$

Cette matrice comporte des vides correspondant aux lacunes d'observations (inconnu), le comblement de ces valeurs se fait par la moyenne mensuelle interannuelle telle que :

$$M_j = 1/q_j \sum_{i=1}^n X(i, j)$$

avec :

q_j : la différence entre la taille du vecteur colonne et le nombre d'inconnues de la colonne j

- Une donnée journalière manquante peut être remplacée soit par la moyenne entre la valeur des jours précédent et suivant, soit par la moyenne du mois ou bien par la moyenne journalière interannuelle du jour considéré.

b) La régression linéaire :

Soit X et Y deux échantillons, avec n observations (x_i, y_i) ; $i = 1, n$, correspondant chacune à des mesures simultanées de précipitations sur deux stations assez rapprochées, les observations sont donc similaires, et la connaissance d'une des variables permet, alors dans une certaine mesure, la prédiction quantitative ou qualitative de l'autre.

Théoriquement, on appelle modèle de régression linéaire sur les données expérimentales, la méthode des moindres carrés reprend sur l'échantillon la propriété que $E(y/x) = \alpha + \beta x$

La méthode des moindres carrés permet de déterminer l'équation de régression $\hat{y} = ax + b$ de telle sorte que $\sum (y_i - \hat{y}_i)^2$ soit minimal.

$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 = f(a, b)$ ce minimum est atteint pour $\partial F / \partial a = \partial F / \partial b$

$$\text{Ce qui donne : } -2 \sum (y_i - a - bx_i) = 0 \quad \longrightarrow \quad y = a + bx$$

$$-2x_i \sum (y_i - a - bx_i) = 0$$

la solution est donc :

$$b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = r (S_y / S_x)$$

$$\text{D'où : } \hat{y} = \bar{y} + r \frac{S_y}{S_x} (x - \bar{x})$$

c) La simulation :

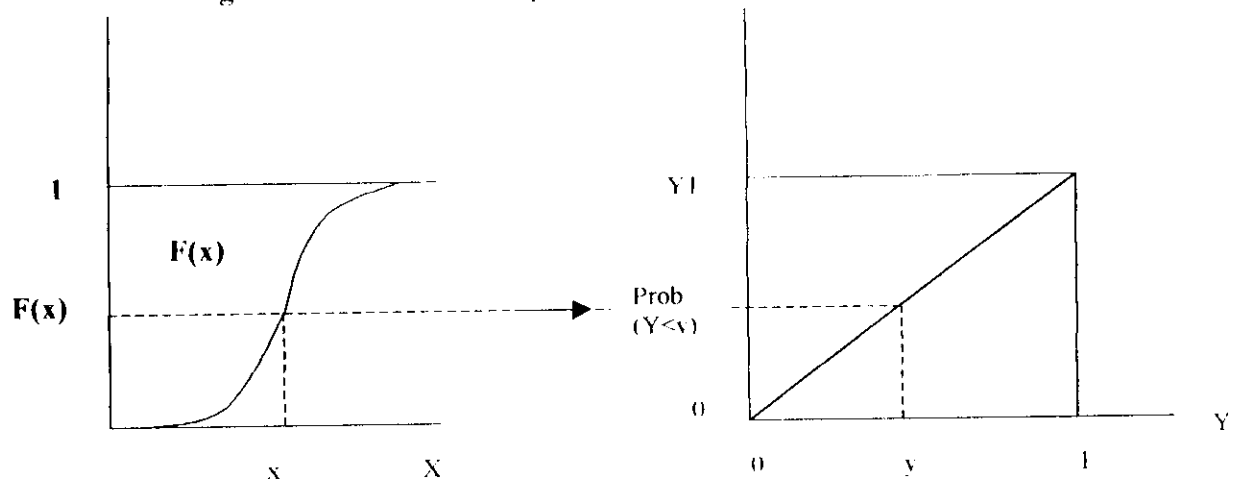
La notion de la simulation est basée essentiellement sur le concept du hasard.

C'est une méthode de tirage d'un échantillon artificiel de (n) valeurs d'une variable aléatoire continue. Elle s'applique lorsque F^{-1} a une forme analytique simple. Soit $F(x)$ la fonction de répartition de x, la variable $Y = f(x)$ est distribuée sur $[0, 1]$ en effet :

$$\text{Prob}(Y < y) = \text{prob}[F(x) < y] = y$$

Si on tire (n) nombres uniformément répartis sur $[0, 1]$: $r_1, r_2, r_3, \dots, r_n$ l'échantillon cherché $(x_1, x_2, x_3, \dots, x_n)$ sera déterminé par $x_i = F^{-1}(y_i)$, cette méthode est dite «de l'anamorphose» (fig. II.2). On dit qu'on a simulé la variable x.

Figure II.2 : simulation à partir de la fonction de répartition



L'analyse des fichiers fournis par l'ONM ne présentent pas de valeurs manquantes.

II.3.3 Détection des valeurs singulières :

Les observations qui s'écartent fortement du centre de répartition sont les résultats d'une erreur systématique, d'une lecture incorrecte des indications de l'instrument de mesure ou à cause des valeurs intégrées après le comblement des lacunes.

Il est important avant tout traitement de détecter ces erreurs qui se glissent dans les données de base, afin que l'on puisse effectuer une analyse plus fine pour décider si cette donnée douteuse est aberrante et doit donc être éliminée, ou si c'est une donnée exceptionnelle, qui par conséquent doit être au contraire précieusement conservée.

Test de Grubbs et Beck (valeurs singulières) : [15]

Ce test est très utilisé à cause de sa simplicité. Il peut détecter les points singuliers de la série en calculant les statistiques suivantes :

$$X_n = \text{Exp} (m + K_n * S)$$

$$X_1 = \text{Exp} (m - K_n * S)$$

Où :

m et **S** : sont respectivement, la moyenne et l'écart type des logarithmes naturels des éléments constituant l'échantillon.

K_n : est une valeur statistique du test de Grubbs et Beck, tabulée pour différentes tailles d'échantillons de niveaux de signification.

Au niveau de risque de 10%, elle est donnée par l'expression suivante :

$$K_n = 3.62201 + 6.28446 * n^{1/4} - 2.49835 * n^{1/2} + 0.491436 * n^{3/4} - 0.037911 * n$$

Avec n : Taille de l'échantillon.

En appliquant ce test, tous les points de l'échantillon supérieurs à X_h ou inférieurs à X_l sont considérés comme points singuliers.

Analyse des résultats :

Le tableau II.2 récapitule les résultats obtenus après l'application du test de Grubbs et Beck.

Tableau II.2 résultats du test Grubbs et Beck

	n	m (mm)	Σ (mm)	K_n	X_h (mm)	X_l (mm)
Echelle mensuelle	792	1.26	0.93	2.75	6717.7	0.273
Echelle annuelle	66	6.44	0.29	2.87	1439.85	272.51

L'application de ce test a détecté au pas du temps mensuel les valeurs nulles comme singulières, mais on ne peut rejeter ces points car ils correspondent aux mois secs.

II.4 Traitement statistique des données :

Après la critique des données pluviométriques, il est indispensable de vérifier l'indépendance et l'homogénéité de la série.

II.4.1 Test d'homogénéité : [15]

Les échantillons sont dits **homogènes**, s'ils sont caractérisés par une même loi de probabilité et par les mêmes paramètres, les données recueillies sont alors considérées comme extraites d'une même population.

Il existe plusieurs tests pour vérifier l'homogénéité des séries de données, on a utilisé dans ce cas celui de **WILCOXON** :

C'est un test non paramétrique qui utilise la série des rangs au lieu de la série des observations.

Soit deux échantillons X et Y : N_1 et N_2 sont les tailles de ces échantillons.

Convenons que le premier échantillon est celui dont la taille est la plus petite (ou dont la taille n'est pas supérieure à celle du second) de sorte que dans ce qui suit nous aurons :

$$N_1 \leq N_2.$$

Le test de WILCOXON se base sur le principe suivant

Si l'échantillon X est issu de la même population Y , l'échantillon $X \cup Y$ en est également issu.

On construit à l'aide des échantillons réunis, la série de variations qui sera composée de $N_1 + N_2$ valeurs, disposées dans l'ordre de croissance.

Par la suite, on s'intéresse au rang de chacun des éléments des deux échantillons dans cette série.

Si une même valeur se répète plusieurs fois, on lui associe le rang moyen correspondant. On calcule la somme W_x des rangs des éléments du premier échantillon dans la série commune :

$$W_x = \sum | \text{Rang} |.$$

WILCOXON a montré que dans le cas où les deux échantillons X , Y constituent une série homogène, la quantité W_x est comprise entre deux bornes W_{\max} et W_{\min} , données par les formules suivantes :

$$W_{\min} = \frac{(N_1 + N_2 + 1)N_1 - 1}{2} - U_{1-\alpha/2} \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}}.$$

$$W_{\max} = (N_1 + N_2 + 1) N_1 - W_{\min}.$$

$U_{1-\alpha/2}$: représente la valeur de la variable centrée réduite de la loi normale correspondante à $1-\alpha/2$.

Si $W_{\min} < W_x < W_{\max}$, l'hypothèse de l'homogénéité est acceptée.

II.4.2 Test d'indépendance : [15]

L'application des méthodes statistiques repose sur la validité de certaines hypothèses de base relatives à la constitution de l'échantillon observé. En effet le fichier recueilli doit être constitué par une série d'observations indépendantes formant ce que l'on appelle un échantillon extrait au hasard de la population statistique. En d'autre terme l'échantillon $X = (x_1, \dots, x_j, \dots, x_n)$ est dit (au hasard), si le test permet de vérifier cette hypothèse H_0 d'indépendance des observations.

Dans cette étude, on a appliqué le teste de Wald- Walfowitz (couramment utilisé) sur les valeurs mensuelles.

Test de Wald-Walfowitz :

Il est basé sur le critère qu'aucune observation appartenant à l'échantillon ne devrait influencer l'observation suivante :

La statistique du test est :

$$R = \sum_{j=1}^n X_j X_{j+1}$$

Si l'hypothèse H_0 selon la quelle l'échantillon est extrait au hasard est vraie, la statistique suit une distribution normale.

$$N\left(\frac{m_1^2 - m_2^2}{N-1}, \frac{m_1^2 - m_4}{N-1} + \frac{m_1^4 - 4m_1^2 m_3 + 4m_1 m_3 + m_2 - 2m_4}{(N-1)(N-2)}, \left[\frac{m_1^2 - m_2^2}{N-1}\right]\right)$$

Avec : $m_r = \sum_{i=1}^n x_i^r = N \times S_r$

S_r : moment d'ordre r de l'échantillon

La valeur statistique $U = \frac{R - \bar{R}}{\sqrt{\text{var}(R)}}$ suit une distribution normale standardisée $N(0,1)$, et est

utilisée pour tester à un niveau de signification, l'hypothèse de l'indépendance en comparant la valeur absolue de U avec la valeur normale standard $U_{\alpha/2}$ correspondant à la probabilité de dépassement de $\alpha/2$.

- Si $|U|$ est inférieure à $U_{\alpha/2}$ l'hypothèse d'indépendance est vérifiée.
- Si $|U|$ est supérieure à $U_{\alpha/2}$ l'hypothèse d'indépendance est rejetée.

II.4.3 Analyse des résultats :

Le résultat du test de WILCOXON illustré dans le tableau II.3 et appliqué aux différents seuils, montre que les observations pluviométriques sont homogènes.

L'hypothèse H_0 d'indépendance pour le pas du temps mensuel est rejetée pour les différents seuils utilisés (tableau II.3).

Tableau II.3 : résultats des tests statistiques.

Test	$\alpha=1\%$	$\alpha=5\%$	$\alpha=10\%$
Homogénéité	1	1	1
Indépendance	0	0	0

1: H_0 acceptée

0: H_0 rejetée

Ces résultats peuvent être expliqués par la tendance du comportement dans le temps des données mensuelles, où par le fait que l'année comporte des mois humides et secs dont la corrélation entre les mois de la même saison peut être importante.

Chapitre III

Théories de l'ACP et des chaînes de Markov

De nombreux modèles sont exposés pour la représentation des phénomènes naturels, mais malheureusement, leur amélioration conduit à l'introduction de paramètres souvent difficiles à estimer et à l'exploitation d'un volume de données de plus en plus important, de ce fait, l'utilisation des modèles hydrologiques simples est privilégiée par les hydrologues car elle fait intervenir le moins possible d'hypothèses et de paramètres

Le choix d'utilisation d'un de ces modèles dépend des facteurs suivant [17] :

- Jugement, expérience et préférence personnelle du modélisateur.
- Le processus physique du modèle à étudier.
- Les caractéristiques statistiques des séries chronologiques.

Dans cette étude, on s'est intéressé à la modélisation par le biais de l'analyse en composantes principales et les chaînes de Markov.

III.2 Analyse en composantes principales (ACP) :

III.2.1 Théorie :

1) Introduction [18]

Conçue pour la première fois par Karl Pearson en 1901, intégrée à la statistique mathématique par Harold Hotelling en 1933, l'analyse en composante principale n'est vraiment utilisée que depuis l'avènement et la diffusion des moyens de calculs actuels.

La technique d'analyse en composantes principales peut être présentée de divers points de vue.

Pour le statisticien classique, il s'agit de la recherche des axes principaux de l'ellipsoïde indicateur d'une distribution normale multidimensionnelle, ces axes étant estimés à partir d'un échantillon. C'est la présentation initiale de Hotelling (1933), puis celle des manuels classiques d'analyse multivariée.

Pour les factoralistes classiques, il s'agit d'un cas particulier de la méthode d'analyse factorielle des psychométriciens

Enfin, du point de vue plus récent des analystes de données, il s'agit d'une technique de représentation des données, ayant un caractère optimal selon certains critères algébriques et géométriques et que l'on utilise en général sans référence à des hypothèses de nature statistique ni à un modèle particulier.

L'analyse en composantes principales présente de nombreuses variantes selon les transformations apportées au tableau de données : le nuage des points- individus, variables- peut être centré ou non, réduit ou non.

2) Définition [18]

On appelle aussi l'analyse des axes principaux. Une variable de coordonnée (x, y) peut être représentée sur un plan, un point de trois coordonnées (x, y, z) peut aussi être représenté graphiquement dans l'espace. Mais lorsqu'il s'agit des n individus à P coordonnées ; ces n

individus forment un nuage de points dans un espace à P dimensions bien difficiles à présenter.

L'A.C.P permet de trouver un nouvel espace de dimension plus faible qui pourra présenter le nuage de point en perdant le moins d'information possible.

Elle est basée sur l'étude de la covariance ou de la corrélation entre variable, les résultats sont donnés sous forme de représentation graphique planes et d'indices permettant de les interpréter.

Il existe deux types de métrique, pour effectuer une A.C.P

- La métrique canonique : Les résultats considérés entre variable sont mesurés par les covariances
- La métrique normée : On appelle aussi A.C.P canonique des variables réduites, les relations considérées entre variables sont mesurées par les coefficients de corrélation.

Ce type de métrique est utilisé lorsque les unités de variables étudiées sont hétérogènes, étant donné que la covariance est un paramètre qui dépend des unités choisies.

Calculer la première C.P revient à chercher l'axe tel que la somme des carrés des distances des n points à cet axe soit minimale ou encore que la somme des carrés des projections des points sur cet axe soit maximale.

On détermine le second axe - après projection des n points dans un hyperplan orthogonal au premier axe - tel que la somme des carrés des distances des points (les observations) soit minimale à celui-ci. Ce calcul se réitère p fois, la détermination des p composantes principales revient à effectuer une rotation du système d'axes origine.

3) Calcul de l'analyse en composante principale [19]

a) Calcul de la matrice des covariances :

la matrice des covariances, notée [R] est la base de l'ACP, elle est obtenue en appliquant la relation suivante :

$$[R] = (1/n)[X]^t[M][X] \quad (III.1)$$

[R] : matrice de covariance.

[X] : matrice de données.

[X]^t : matrice transposée de [X].

[M] : métrique.

b) Recherche des axes principaux :

Le but est de construire un nouveau système d'axes avec un minimum de variables assurant un maximum de variance.

Recherche du premier axe :

La contribution maximale des CP est donnée par le premier axe principal. On doit donc maximiser la variance relative à celui-ci, la recherche du premier axe principal consiste à résoudre le problème :

$$\begin{cases} \text{MaxVar}(C_1) \\ C_1' \cdot [M] \cdot V_1 = 1 \end{cases}$$

On peut déterminer la variance de C_1 à l'aide de la matrice des covariances $[R]$ du vecteur : $X=(X_1, X_2, \dots, X_j, \dots, X_p)$.

$$\text{Var}(C_1) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 \quad (\text{III.2})$$

En utilisant la méthode du multiplicateur de LAGRANGE on peut écrire :

$$L = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 - \lambda_1 (V_1' \cdot [M] \cdot V_1 - 1) \quad (\text{III.3})$$

La dérivée par rapport à V est nécessairement nulle :

$$\partial L / \partial V_1 = 2 \cdot [M] \cdot [R] \cdot [M] \cdot V_1 - 2 \lambda_1 \cdot [M] \cdot V_1 = 0 \quad (\text{III.4})$$

Puisque la matrice $[M]$ est inversible :

$$[R] \cdot [M] \cdot V_1 = \lambda_1 \cdot V_1 \quad (\text{III.5})$$

Donc V_1 est le vecteur propre de la matrice $[R] \cdot [M]$. Il suffit de choisir comme vecteur V_1 le vecteur propre associé à la plus grande valeur propre λ_1 de la matrice $[R] \cdot [M]$ pour maximiser la variance de C_1 .

Recherche du second axe :

On cherche à déterminer le vecteur unitaire V_2 de telle façon que la composante C_2 soit de variance maximale et non corrélée à C_1 .

Sachant que :

$$\text{Var}(C_2) = V_2' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III.6})$$

$$V_2' \cdot [M] \cdot V_2 = 1 \quad (\text{III.7})$$

$$\text{Cov}(C_1, C_2) = 0 \quad (\text{III.8})$$

L'expression de $\text{Cov}(C_1, C_2)$ est donnée par :

$$\text{Cov}(C_1, C_2) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III.9})$$

Comme la variance ne tient pas compte de l'ordre, on a :

$$\text{Cov}(C_1, C_2) = \text{Cov}(C_2, C_1) = V_2' [M] [R] [M] V_1 \quad (\text{III.10})$$

Or on sait que V_1 est un vecteur propre de $[R].[M]$ associé à la valeur propre λ_1 . On en déduit que :

$$\text{Cov}(C_1, C_2) = \lambda_1 V_2' V_1 = 0 \quad (\text{III.11})$$

Une covariance nulle entre C_1 et C_2 est équivalente à l'orthogonalité des vecteurs V_1 et V_2 :

$$\text{Cov}(C_1, C_2) = 0 \Leftrightarrow V_1 \cdot V_2 = 0 \quad (\text{III.12})$$

En appliquant la même méthode pour la recherche du deuxième axe, on aura :

$$L = V_2' [M] [R] [M] V_2 - \lambda_2 (V_2' [M] V_2 - 1) - \mu (V_2' [M] V_1) \quad (\text{III.13})$$

$$\partial L / \partial V_2 = 2 [M] [R] [M] V_2 - 2 \lambda_2 [M] V_2 - \mu ([M] V_1) = 0 \quad (\text{III.14})$$

En simplifiant par $[M]$, on obtient :

$$2 [R] [M] V_2 - \lambda_2 V_2 - \mu V_1 = 0 \quad (\text{III.15})$$

On multiplie à gauche par V_1' , on obtient :

$$2 V_1' [M] [R] [M] V_2 - 2 \lambda_2 V_1' [M] V_2 - \mu V_1' [M] V_1 = 0 \quad (\text{III.16})$$

Par hypothèse :

$$V_1' [M] V_2 = 0$$

(III.17)

$$V_1' [M] [R] V_2 - V_1' [M] [R] [M] V_1 = \lambda_1 V_1' [M] V_1 = 0 \quad (\text{III.18})$$

Puisque le vecteur V_1 est unitaire :

$$V_1' [M] V_1 = 1 \quad (\text{III.19})$$

Le multiplicateur de LAGRANGE μ est donc nul et l'on est ramené au problème précédent. On peut donc énoncer la définition suivante :

Le second axe est défini par le vecteur V_2 , vecteur propre unitaire de la matrice $[M].[R]$ orthogonal V_1 à et associé à la plus grande valeur propre λ_2 inférieure ou égale à λ_1 .

Recherche des autres axes :

En itérant le procédé on détermine donc les valeurs propres de la matrice $[R].[M]$ pour obtenir la 1^{ème} composante principale C_1 .

Le vecteur propre unitaire de la matrice $[R].[M]$ définit le 1^{ème} axe orthogonal à V_1, V_2, \dots, V_{1-1} ; est associé à la 1^{ème} plus grande valeur propre λ_1 .

4) Application de l'ACP [17]

En général, le domaine d'utilisation de l'ACP est très large, sans inconvénients, sans dangers et ne nécessite pas au préalable d'hypothèses restrictives.

On distingue deux aspects principaux de l'ACP :

- a) **L'analyse descriptive** : s'intéresse à la structure de la matrice d'observation, basée principalement sur la représentation graphique des résultats trouvés par l'ACP et ce, afin de visualiser certains problèmes non décelables à l'état brut, vu l'indépendance des variables.
- b) **L'analyse opérationnelle** : l'ACP fournit un outil mathématique en apparence mais très bien adapté aux problèmes rencontrés en hydrologie a savoir :
 - Optimisation d'un réseau de mesures.
 - Critique et reconstitution des données prévision.
 - Prévision.
 - Simulation.

Dans le domaine descriptif, deux types de phénomènes peuvent être étudiés : temporel et spatial.

Le premier type met en évidence des relations entre des observations à deux époques distinctes de l'année alors que le second fait ressortir la notion de proximité inter-stations.

L'application de l'ACP portera sur deux types de variables :

- L'une homogène, caractérisant un phénomène spatial (précipitations relevées sur les six stations du nord algérien).
- L'autre hétérogène : par exemple, l'ensemble des paramètres intervenant dans la détermination de l'évapotranspiration.

L'ACP peut être effectuée sur quatre variables :

1-Variable centrée :

$$Y_j = X_j - \overline{X_j} \quad (\text{III.20})$$

Avec : X_j : Variable initiale ;

$\overline{X_j}$: Moyenne de la variable initiale ;

Y_j : Variable transformée (centrée).

2-Variable réduite :

$$Y_j = \frac{X_j}{\sigma_{X_j}} \quad (\text{III.21})$$

Avec : X_j : Variable initiale ;

σ_{X_j} : Ecart type de la variable initiale ;

Y_j : Variable transformée (réduite).

3-Variable centrée réduite :

$$Y_j = \frac{X_j - \bar{X}_j}{\sigma_{X_j}} \quad (\text{III.22})$$

Avec : X_j : Variable initiale ;

\bar{X}_j : Moyenne de la variable initiale ;

σ_{X_j} : Ecart type de la variable initiale ;

Y_j : Variable transformée (centrée réduite).

Le traitement des variables brutes et des variables centrées donne lieu à la même matrice de covariance, et par conséquent, les résultats de la diagonalisation sont identiques.

En effet, considérons la variable brute X et la variable centrée Y résultant de la transformation (III.20), et montrons que la covariance des deux variables brutes (X_j, X_k) est équivalente à celle des deux variables centrées respectives (Y_j, Y_k).

$$Y_j = X_j - \bar{X}_j \quad \Rightarrow \quad Y_j = 0$$

$$Y_k = X_k - \bar{X}_k \quad \Rightarrow \quad Y_k = 0$$

Or :

$$\text{Cov}(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \bar{Y}_j) * (Y_{ik} - \bar{Y}_k) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik}$$

En remplaçant chaque variable par son expression on obtient :

$$\begin{aligned} \text{Cov}(Y_j, Y_k) &= \frac{1}{N} \sum_{i=1}^N Y_{ji} * Y_{ki} = \frac{1}{N} \sum_{i=1}^N (X_{ji} - \bar{X}_j) * (X_{ki} - \bar{X}_k) \\ &= \text{Cov}(X_j, X_k) \end{aligned}$$

d'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(X_j, X_k)$$

Le traitement des variables réduites et centrées réduites se fait de la manière suivante :

Soient Y_j et Z_j les variables obtenues par les transformations (III.21) et (III.22), montrons que la covariance des deux variables réduites (Y_j, Y_k) est égale à celle des deux variables centrées réduites (Z_j, Z_k) , qui n'est autre que la corrélation entre les deux variables brutes correspondantes.

$$\begin{aligned} Y_j &= \frac{X_j}{\sigma_{X_j}} & \Rightarrow & \quad Y_j = \frac{\bar{X}_j}{\sigma_{X_j}} \\ Y_k &= \frac{X_k}{\sigma_{X_k}} & \Rightarrow & \quad Y_k = \frac{\bar{X}_k}{\sigma_{X_k}} \\ Z_j &= \frac{X_j - \bar{X}_j}{\sigma_{X_j}} & \Rightarrow & \quad Z_j = 0 \\ Z_k &= \frac{X_k - \bar{X}_k}{\sigma_{X_k}} & \Rightarrow & \quad Z_k = 0 \end{aligned}$$

Evaluons maintenant les covariances des variables (Y_j, Y_k) et (Z_j, Z_k) :

$$\begin{aligned} \text{Cov}(Y_j, Y_k) &= \frac{1}{N} \sum_{i=1}^N (Y_{ji} - \bar{Y}_j) * (Y_{ki} - \bar{Y}_k) \\ &= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ji}}{\sigma_{X_j}} - \frac{\bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ki}}{\sigma_{X_k}} - \frac{\bar{X}_k}{\sigma_{X_k}} \right) \\ &= \frac{1}{N} \sum_{i=1}^N (X_{ji} - \bar{X}_j) * (X_{ki} - \bar{X}_k) \\ &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \text{Cor}(X_j, X_k) \end{aligned}$$

Donc :

$$\begin{aligned}
 \text{Cov}(Y_j, Y_k) &= \text{Cor}(X_j, X_k) \\
 \text{Cov}(Z_j, Z_k) &= \frac{1}{N} \sum_{i=1}^N (Z_{ij} - \bar{Z}_j) * (Z_{ik} - \bar{Z}_k) \\
 &= \frac{1}{N} \sum_{i=1}^N Z_{ij} * Z_{ik} \\
 &= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik} - \bar{X}_k}{\sigma_{X_k}} \right) \\
 &= \frac{\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\sigma_{X_j} * \sigma_{X_k}} \\
 &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\
 &= \text{Cor}(X_j, X_k)
 \end{aligned}$$

Donc :

$$\text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k)$$

D'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k)$$

En conclusion on aura deux types de variantes à appliquer :

ACP brute = ACP centrée.

ACP centrée réduite (ou normée) = ACP réduite.

L'ACP brute peut être utilisée pour tous les types de variables homogènes c'est à dire possédant les mêmes unités de mesure alors que l'ACP normée peut être utilisée pour des variables homogènes et hétérogènes. En effet en centrant et en réduisant la variable, on élimine le problème de l'hétérogénéité de celle-ci.

III.2.2 Application :

Une ACP normée a été appliquée sur les données pluviométriques mensuelles de la station d'Alger sur la période d'étude.

L'application de cette méthode sur les données pluviométriques calendaires constituants 66 individus et 12 variables donne les résultats suivants :

-----Matrice de corrélation-----

1.00												
-.02	1.00											
-.06	.13	1.00										
.17	.07	.17	1.00									
-.05	-.23	.02	.15	1.00								
.18	.35	.08	.02	-.10	1.00							
.09	.16	-.18	-.16	-.05	.03	1.00						
.04	-.03	.00	-.03	.05	.10	.08	1.00					
-.01	.34	.31	-.13	-.19	.22	.01	.02	1.00				
-.03	-.02	.02	-.01	.22	.21	-.07	-.10	.07	1.00			
-.04	-.03	.06	-.06	-.15	.00	-.06	-.10	-.12	.09	1.00		
-.05	.05	.09	-.05	.13	.11	-.06	.03	.15	.18	-.24	1.00	

Les résultats montrent que les coefficients de corrélations ne sont pas significatifs, même entre les mois de la même saison.

La matrice de corrélation permet après diagonalisation de calculer les valeurs et les vecteurs propres .

Dans le processus de calcul, les vecteurs propres sont déterminés dans l'ordre des valeurs propres décroissantes (tableau 1).

Les principales propriétés d'un modèle normé et que la somme des valeurs propres est égale au nombre de variables et les CP sont centrées .

Tableau 1 : Les valeurs propres pour chaque composante principale.

N° de CP	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12	
Valeur propre	1.88	1.51	1.34	1.26	1.18	.97	.89	.79	.68	.57	.50	.41	$\Sigma=12$

En outre , les valeurs propres permettent de connaître la contribution de chacune des composantes principales à la variance totale du système.

Tableau .2 : Variance expliquée par chaque CP.

N°des CP	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12
CP normé	15.65	12.77	11.15	10.58	9.81	8.12	7.33	6.59	5.63	4.73	4.2	3.44
Cumul des CP (%)	15.65	28.42	39.57	50.15	59.96	68.08	75.41	82	87.63	92.36	96.56	100

Généralement, la première composante donne plus d'information sur la répartition des précipitations.

Le tableau 2 montre que les six premières CP expliquent plus de 68% de la variance totale.

III.3 les chaînes de Markov [20]

III.3.1 Introduction :

L'évolution, au cours du temps, des phénomènes où intervient le hasard ou dont les causes qui les régissent sont si complexes, sont bien représentés par des modèles aléatoires.

Un modèle très fréquent de dépendance est celui imaginé vers 1910 par le mathématicien russe A. Markov pour rendre compte de la succession des voyelles et consonnes utilisées dans un poème qu'il étudiait alors.

Le processus $X(t)$ est markovien si quelle que soit la suite croissante d'instants t_1, t_2, \dots, t_n inférieurs à t :

$$P [X(t)=x / X(t_1)=x_1, \dots, X(t_n)=x_n] = P [X(t)=x / X(t_n)=x_n]$$

C'est à dire que quelles que soient les informations recueillies au passé, la loi conditionnelle de $X(t)$ ne dépend que du passé le plus récent. Ce processus est dit stochastique s'il traite des événements aléatoires non indépendants.

L'utilisation des chaînes de Markov en météorologie consiste dans l'analyse des précipitations et l'étude des paramètres climatologiques qui suivent un processus aléatoire.

Dans une série chronologique d'événements aléatoires caractérisée par une variable aléatoire continue, on cherche souvent une relation entre les valeurs successives de celle-ci :

- Si $\lambda = 0$: on dira que les événements sont parfaitement indépendants
- Si $\lambda = 1$: ils seront dits liés linéairement.

Si l'état d'un jour k dépend de celui des jours précédents, on dira que le phénomène étudié suit une chaîne de Markov.

Celle-ci sera du premier ordre si l'événement au jour k ne dépend que celui de jour précédent, elle sera du $n^{\text{ème}}$ ordre si ce même événement dépend des jours $k-1, k-2, \dots, k-n$.

III.3.2 Ordre des chaînes de Markov :

En pluviométrie, on considère un système binaire , 0 pour le temps sec (par rapport à un seuil bien déterminé), et 1 pour le temps pluvieux

a) Processus d'ordre 1 :

Le processus d'ordre 1 est caractérisé par la probabilité conditionnelle de passage de l'état de la veille, à l'état d'aujourd'hui, l'équation caractéristique est notée comme suit :

$$P [X_k=j / x_{k-1}=i]$$

X_k : Variable aléatoire caractérisant l'état (X prend la valeur 0 ou 1).

X_{k-1} : Valeur du jour k-1 (i=0 ou 1) et j l'état réalisé le jour k (j=0 ou 1).

P_{ij} : Probabilité conditionnelle de l'état j au jour k, sachant qu'étant réalisé état i au jour k-1.

Avec ces deux cas , on peut obtenir quatre possibilités par la matrice suivante :

Tableau 3 : matrice de passage d'ordre 1 .

Etat au jour k-1	Etat au jour k	
	0	1
0	P_{00}	P_{01}
1	P_{10}	P_{11}

Exemple : appliquant le modèle Markovien d'ordre 1 sur les données pluviométriques journalières de l'année 2000.

Ces coefficients sont obtenus par le décompte de jour K caractérisé par l'état 0(jour sec) ou 1(jour pluvieux) précédé par le jour K-1 d'état 0 ou 1 :

	Nombre de jour K		
	0	1	
Nombre de jour K-1	0	34	= 303
	1	27	= 62

En divisant chaque ligne par sont total, on obtient la matrice de transition :

	0	1	
0	0.887	0.112	=1
1	0.564	0.435	=1

La probabilité d'avoir deux jours successifs secs est de 88%, par contre celle d'avoir deux jours successifs humides est de 43%.

b) Processus d'ordre 2 :

On cherche la probabilité de l'état d'aujourd'hui en fonction des états connus hier et avant-hier.

Le calcul statistique fournit les probabilités conditionnelles du tableau 4.

Tableau 4 : matrice de passage d'ordre 2

Couple avant-hier et hier (i, j)	Couple hier et aujourd'hui (j, k)			
	00	01	10	11
00	B_{000}	B_{001}	----	----
01	----	----	B_{010}	B_{011}
10	B_{100}	B_{101}	----	----
11	----	----	B_{110}	B_{111}

Pour cet ordre, l'état de jour k dépend de l'état au jour k-1 et k-2.

B_{ijk} : représente la probabilité conditionnelle d'obtenir un doublet de classe (j, k) succédant à une autre classe (i, j).

c) Processus d'ordre 3 :

Pour obtenir plus d'information sur l'occurrence des pluies, on fait appel à un ordre supérieur.

L'ordre 3 permet de déterminer les probabilités conditionnelles de l'état au jour k, en faisant intervenir les états des trois jours précédents (tableau 5).

Tableau 5: matrice de passage d'ordre 3

Couple avant-avant hier, avant hier et hier (i, j, l)	Couple avant hier, hier et aujourd'hui (j, l, K)							
	000	001	010	011	100	101	110	111
000	γ_{0000}	γ_{0001}	----	----	----	----	----	----
001	----	----	γ_{0010}	γ_{0011}	----	----	----	----
010	----	----	----	----	γ_{0100}	γ_{0101}	----	----
011	----	----	----	----	----	----	γ_{0110}	γ_{0111}
100	γ_{1000}	γ_{1001}	----	----	----	----	----	----
101	----	----	γ_{1010}	γ_{1011}	----	----	----	----
110	----	----	----	----	γ_{1100}	γ_{1101}	----	----
111	----	----	----	----	----	----	γ_{1110}	γ_{1111}

γ_{jlk} : représente la probabilité conditionnelle d'obtenir un triplet de classe (j, l, k) succédant à une autre classe (i, j, l).

III.3.3 Les modèles multiclassés :

Au lieu d'utiliser un système binaire, on peut aussi appliquer un modèle Markovien multiclassé qui comme son nom l'indique consiste à subdiviser la série historique en plusieurs classes.

Dans ce cadre là, on transforme l'état (système binaire) en classe (système multiclassé), le problème qui se posera est le choix des seuils de ces classes.

Pour un processus multiclassé d'ordre N constitué de N classes la matrice de transition obtenue dans ce cas est carrée et aura pour dimension $N \times N$.

Le découpage en classes peut se faire de différentes manières et ceci dans le but d'avoir une répartition homogène des données aux niveaux de toutes les classes afin de ne pas favoriser l'une d'elles par rapport à une autre.

Ce découpage en classe doit tenir compte du phénomène physique, il peut se faire :

- A) aléatoirement : on choisit les seuils et on suit l'évolution des coefficients de la matrice de passage.
- B) En utilisant la fonction de répartition de la série historique, on considère les classes d'égale probabilité marginale sauf pour la classe sèche
- C) En se basant sur la moyenne et l'écart type de la série historique : on procède au découpage à l'aide des seuils suivants :
 - ($m - \sigma$, $m + \sigma$)
 - ($m - 2 \sigma$, $m + 2 \sigma$)
 - ($m - 3 \sigma$, $m + 3 \sigma$)

m : moyenne de la série historique

σ : son écart type

Plus on augmente le nombre de classes, et plus on a des difficultés à choisir leurs seuils. L'augmentation du nombre de classes permet le passage d'un modèle discret à un modèle continu, le problème est l'homogénéité du nombre de valeur dans chaque classe.

Exemple :

Sur la base des données pluviométriques mensuelles de la station d'Alger (1936-2001). On a appliqué un modèle markovien multiclassé d'ordre 1

On choisit :

Un nombre de classe = 3

Les bornes sont choisies aléatoire : 30 - 70mm

La matrice de passage obtenu est :

	Classe 0	classe 1	classe 2
0	0.643	0.185	0.170
1	0.335	0.304	0.360
2	0.237	0.299	0.463

On remarque que la plus forte probabilité conditionnelle est celle d'avoir deux mois successifs secs ($< 30\text{mm}$), par contre la plus faible est d'avoir un mois humide ($> 70\text{mm}$) précédé d'un mois sec.

III.3.3 Le modèle multiclasse saisonnier :

Vu la disparité des saisons, il est parfois indispensable de scinder les années en saisons afin d'affiner l'étude et de ressortir les caractéristiques de chaque saison pour que le modèle soit le plus près possible du phénomène physique étudié.

Exemple :

Une répartition saisonnière est effectuée sur les données pluviométriques mensuelles de la station d'Alger.

Les saisons sont choisies de la manière suivante :

- Saison 1 : mois d'octobre- mois d'avril (saison humide)
Bornes des classes (3 classes) : 30.0mm - 50.0mm
- Saison 2 : mois de mai- mois de septembre
Bornes des classes (2 classes) : 30.0mm

Les matrices de passage obtenues sont :

Pour la saison 1 :

	0	1	2
0	.2328	.1897	.5776
1	.2333	.1556	.6111
2	.1563	.1797	.6641

Pour la saison 2 :

	0	1
0	.8468	.1532
1	.7667	.2333

Les résultats montrent que la plus forte probabilité conditionnelle dans la saison humide, est celle d'avoir deux mois successifs pluvieux ($> 50\text{mm}$), par contre, dans la deuxième saison, elle est celle d'avoir deux mois successifs secs ($< 30\text{mm}$)

Chapitre IV

Analyse du régime pluviométrique

Analyse du régime pluviométrique

IV.1 Introduction :

Sous ses différents aspects (agricole, hydrologique et météorologique), la sécheresse constitue un élément fondamental de l'environnement qui influe directement sur les termes du bilan besoin-ressource. Ce phénomène est perçu par son seul effet directement observable, c'est à dire le manque de précipitations (pluies) et son corollaire sur la disponibilité en eau pour subvenir aux besoins essentiels des êtres humains, des animaux et des plantes. De ce fait, l'étude chronologique des précipitations à l'échelle du jour, du mois et de l'année, constitue un élément essentiel, qui permet de suivre l'évolution du phénomène et déterminer les années de sécheresses caractérisées par des indices temporels.

L'objectif de ce chapitre est d'analyser les précipitations à différents pas de temps pour la région centre qui n'est représentée que par la station d'Alger pour la période 1936-2001.

IV.2 Evolution de la pluviométrie :

Lorsqu'il s'agit de déterminer l'étendue et l'intensité de la sécheresse et suivre son évolution, les climatologues utilisent plusieurs indices représentatifs qui prennent comme référence [21] :

- La moyenne climatologique : pour les ressources en eau, par exemple, l'indice se rapporte aux niveaux médians ou inférieurs à la moyenne de l'eau dans les réservoirs.
- L'analyse fréquentielle : peut se faire à partir des séries observées par leurs ajustements à une loi de probabilité, ce qui permet d'estimer plus correctement les différentes quantités et calculer leurs durées de retour.

IV.2.1 Pas de temps annuel :

1) Par années :

Les paramètres statistiques des pluies annuelles sont donnés dans le tableau IV.1.

L'analyse des figures IV.1 et IV.2 montre qu'il y a autant d'années sèches que d'années pluvieuses pour la série historique étudiée.

En général, il existe deux types de sécheresse, le cas isolé et le cas persistant, le premier cas est très vite compensé par le retour de la pluie, l'autre a un impact plus important.

On remarque que sur la période d'observation (66 ans), il y a 13 cas isolés (s'étalant de 1 à 2 années) et 3 cas persistants (s'étalant de 3 à 6 années)

En outre, on remarque que les plus longs déficits s'étalent sur les deux premières et les deux dernières décennies.

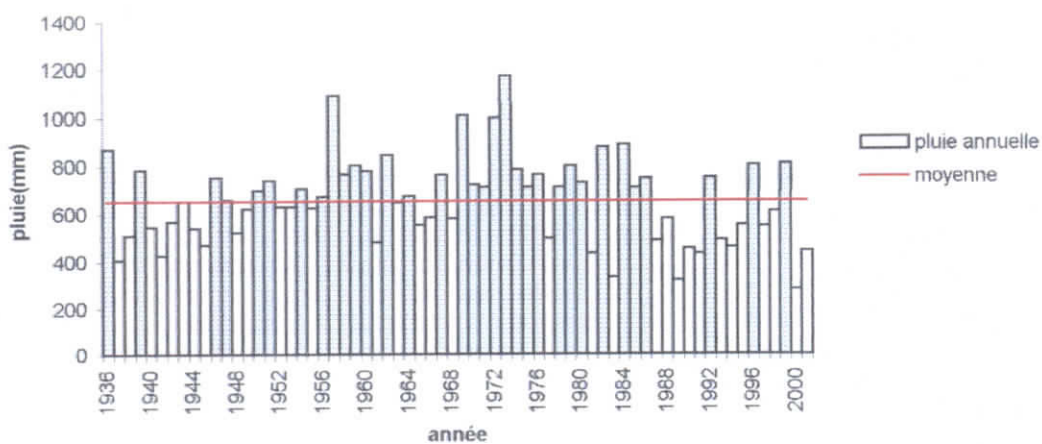


Fig: IV.1 Variation de la pluie annuelle par rapport à la moyenne

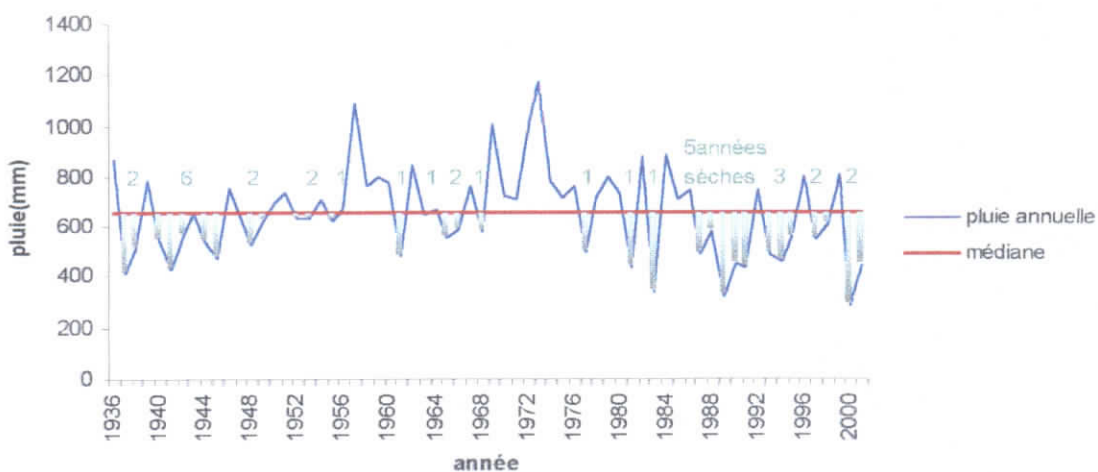


Fig: IV. 2 Variation de la pluie annuelle par rapport à la médiane

Tableau IV.1 : les caractéristiques statistiques de la station d'Alger.

Moyenne (mm)	Médiane (mm)	Ecart- type (mm)	Coefficient de variation
653.8	655.2	178.2	0.27

a) Variabilité pluviométrique :

Le coefficient de variation (CV) est le rapport entre l'écart type et la moyenne. Il permet de comparer la dispersion des variables autour de la celle-ci et caractérise aussi la pluviométrie propre à chaque année (annexe IV).

La figure IV.3. montre que la variabilité pluviométrique est comprise entre 1.57 et 0.71.

La valeur obtenue pour chaque CV met en évidence l'irrégularité du régime pluviométrique.

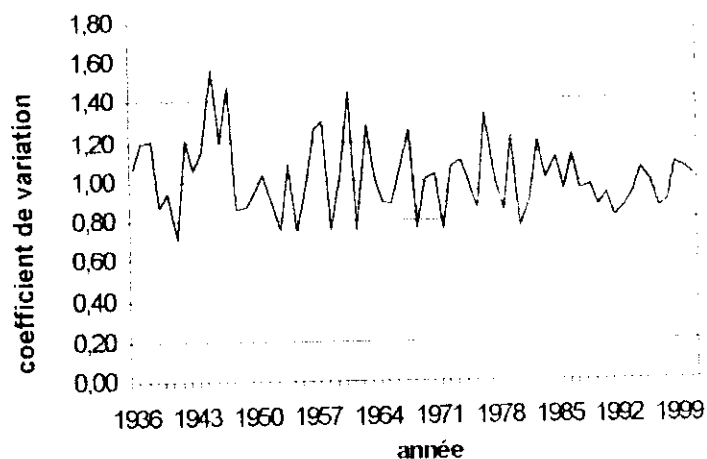


Fig: IV.3 La variabilité pluviométrique

b) La distribution des pluies par rapport à la moyenne :

La figure IV.4. montre une répartition symétrique par rapport à la moyenne entre le nombre d'années sèches et humides.

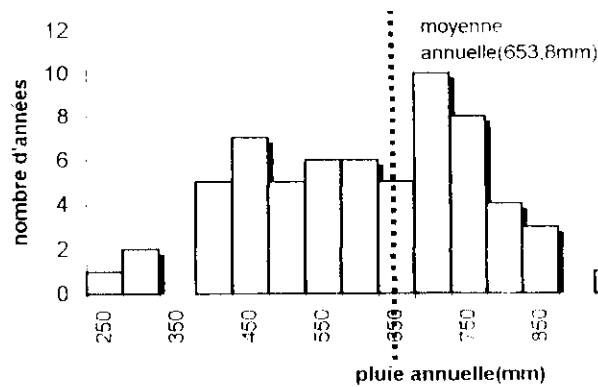


Fig:IV.4 Distribution pluviométrique

2) Par décennies :

Afin de mieux représenter l'évolution du phénomène, on a subdivisé la période total en 6 décennies représentées comme suit :

- 1^{ère} décennie : 1936-1946.
- 2^{ème} décennie : 1947-1957.
- 3^{ème} décennie : 1958-1968.
- 4^{ème} décennie : 1969-1979.
- 5^{ème} décennie : 1980-1990.
- 6^{ème} décennie : 1991-2001.

Les paramètres statistiques pour les différentes décennies sont donnés par le tableau IV.2.

L'évolution des pluies moyennes annuelles pour les différentes décennies est donnée dans la figure IV.5.

L'analyse de celle-ci montre que les quantités de pluie ont diminué dans les deux dernières décennies par rapport à la décennie la plus pluvieuse avec un déficit de 35.09% pour la 5^{ème} décennie et 43.37% pour la 6^{ème}. elles présentent aussi un déficit de 9.87% et 14.24% respectivement par rapport à la moyenne interannuelle qui est de l'ordre de 653.3mm.

En outre, on remarque que le plus faible écart est noté entre la 2^{ème} et la 3^{ème} décennie, et que le plus grand écart pluviométrique est noté entre la 4^{ème} et 6^{ème} décennie.

Tableau IV.2 : les paramètres statistiques des différentes périodes.

Decennies	Moyenne (mm)	Ecart- type (mm)	Coefficient de variation
1	592.8	151.5	0.25
2	689.8	143.2	0.20
3	679.1	118.6	0.17
4	803.2	201.4	0.25
5	594.6	204.0	0.34
6	560.2	167.6	0.30

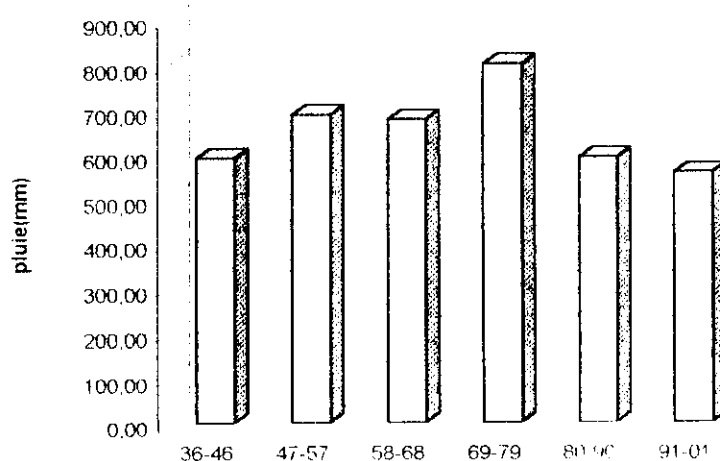


Fig: IV.5 Evolution de la moyenne des pluies annuelles pour les 6 décennies

IV.2.2 Pas de temps mensuel :

Les paramètres statistiques des pluies mensuelles interannuelles sont donnés dans le tableau IV.3.

Tableau IV.3 : la moyenne des pluies mensuelles interannuelles

Mois	Période totale	1ère décennie	2ème décennie	3ème décennie	4ème décennie	5ème décennie	6ème décennie
Janvier	84.3	75.0	113.1	87.4	82.4	70.8	77.8
Février	72.6	59.5	71.4	70.8	132.3	59.3	64.4
Mars	61.0	60.6	44.4	57.6	86.9	66.8	45.4
Avril	61.3	55.5	70.1	66.0	76.9	39.1	61.3
Mai	37.5	40.3	29.0	31.2	47.4	41.4	36.4
Juin	14.8	9.4	15.4	24.8	19.0	12.3	8.8
Juillet	3.1	1.0	2.6	2.0	6.9	5.7	2.2
Août	4.6	2.2	2.1	3.3	8.4	3.6	9.5
Septembre	27.6	19.2	28.3	17.8	47.2	23.8	28.6
Octobre	72.3	55.3	104.5	69.7	87.0	59.3	50.0
Novembre	85.9	77.6	79.9	104.2	110.4	65.7	82.2
décembre	104.3	119.1	103.1	114.1	79.6	119.5	69.9

L'analyse de la variation des pluies moyennes mensuelles interannuelles de la période étudiée montre que le mois de décembre est le plus pluvieux avec une moyenne des précipitations supérieure à **100mm**. et les mois de juillet et août sont les plus secs avec une moyenne qui ne dépasse pas les **5mm** (figure IV.6).

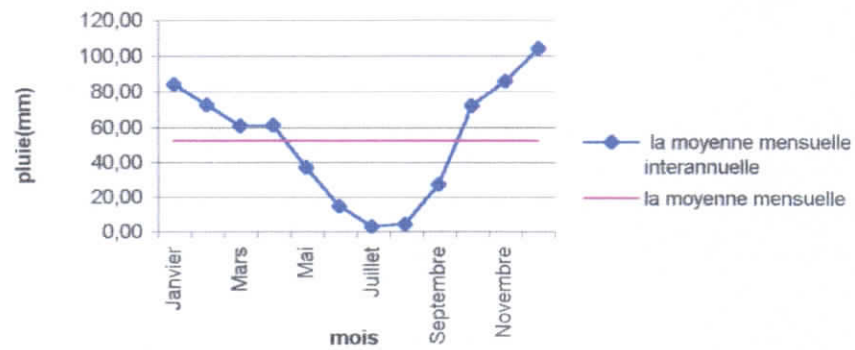


Fig: IV.6 Variation des pluies moyennes mensuelles interannuelles

L'analyse de la figure IV.7 montre qu'il y a deux périodes distinctes, l'une humide (octobre – avril) et l'autre sèche (mai – septembre). A partir de là, on peut dire que l'année peut être scinder à deux saisons.

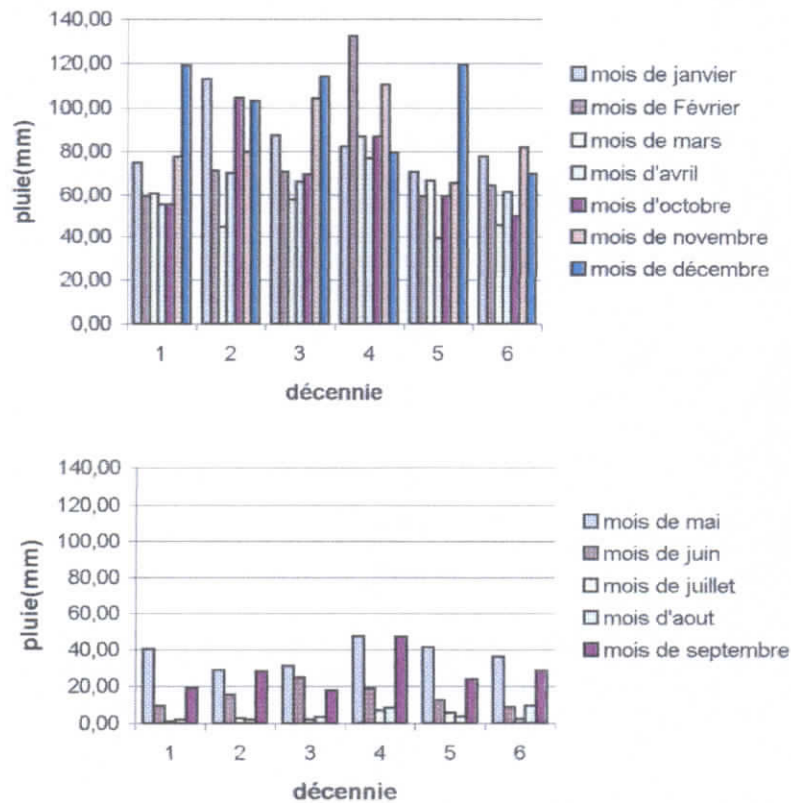


Fig IV.7 : évolution de la moyenne mensuelle interannuelle des précipitations pour les différentes décennies.

IV.2.3 Pas de temps journalier :

Le tracé de l'évolution des pluies journalières montre le caractère aléatoire du phénomène étudié, ce qui rend leur étude très complexe (figure IV.8)

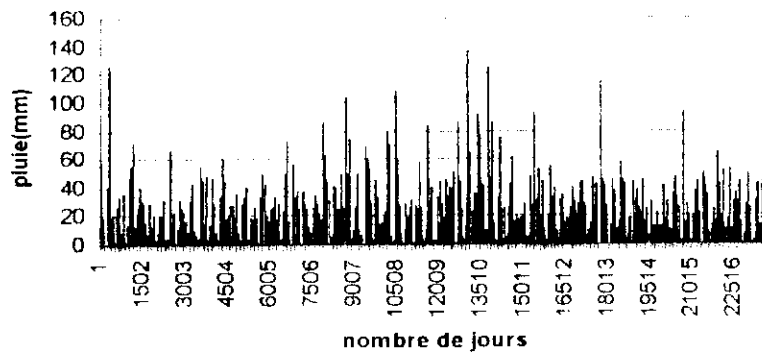


Fig: VI.8 Evolution des pluies journalières

a) Etude des pluies journalières maximales :

L'analyse de la figure IV.9 montre que la quantité minimum des pluies enregistrée est inférieure à 22mm (1941), et la plus grande quantité recueillie au niveau de la station d'Alger a été enregistrée en 1971 et dépasse les 135mm.

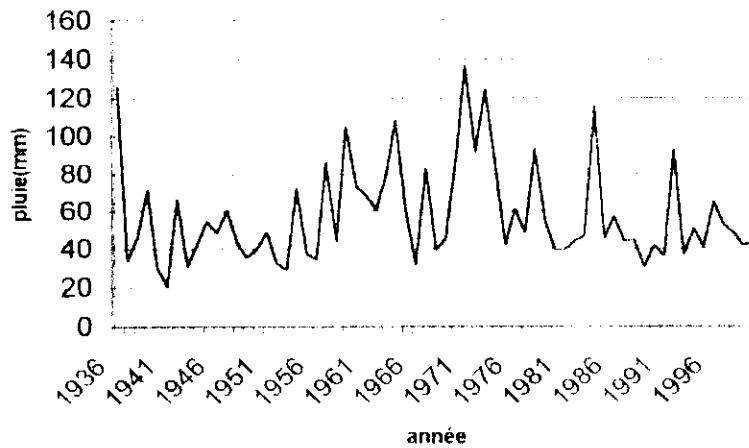


Fig : IV. 9 Evolution des pluies journalières maximales (1936-2000)

b) Nombre des jours de pluie :

Une donnée intéressante et complémentaire à la quantité de pluie est le nombre de jours de pluie. En effet un total de pluie aura des effets divers selon qu'il ait été recueilli en une ou en plusieurs journées.

L'analyse des jours de pluie nous permet de connaître si le déficit pluviométrique constaté est dû à une diminution du nombre de jours de pluie ou plutôt à une diminution des fortes quantités de pluies (figure IV.10).

On considère un jour pluvieux si la quantité des pluies tombée dépasse la moyenne des pluies journalières pour la période sèche (mai- septembre) qui est de l'ordre de 6 mm.

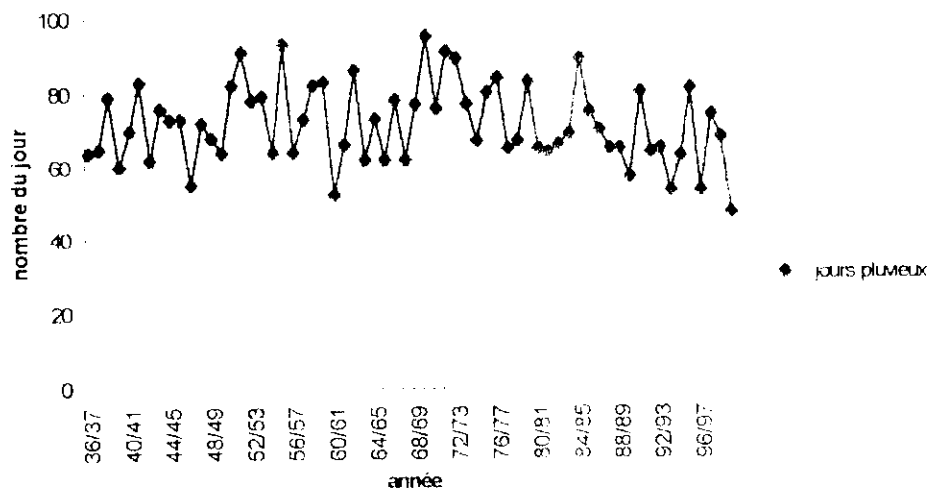


Fig : IV.10 Evolution du nombre des jours pluvieux

L'analyse de la figure IV 10 montre que le plus grand nombre de jours pluvieux est observé en 1969 (95 jours) et le plus faible a été enregistré en 1999 avec seulement 47 jours pluvieux.

D'après ces résultats, on peut dire que le déficit pluviométrique des deux dernières décennies à la station d'Alger est essentiellement dû à une baisse des fortes précipitations, car le nombre de jours pluvieux est identique à ceux des décennies précédentes.

c) La fonction de répartition des maximums journaliers :

En classant la série d'observation par ordre décroissant, on attribue à chaque valeur un numéro d'ordre n compté à partir de 1.

Les fréquences expérimentales de non dépassement sont calculées par la formule de «Weibull » donnée par :

$$F(x) = \frac{n}{N+1}$$

Tels que :

n : le rang de chaque valeur.
N : la taille de l'échantillon.

Le tableau IV.4 résume les fréquences qui donnent une valeur entière de la période de retours.

Le reste des fréquences expérimentales ainsi que leurs périodes de retour sont présentées dans l'annexe IV.1.

Tableau IV.4 : les fréquences expérimentales des pluies journalières maximales.

La pluie maximale (mm)	Fréquence expérimentale	Période de retour
136.6	0.02	66
48.5	0.03	33
42.4	0.05	22
36.2	0.09	11
31.5	0.17	6
30.4	0.33	3
29.1	0.5	2
21.14	0.98	1

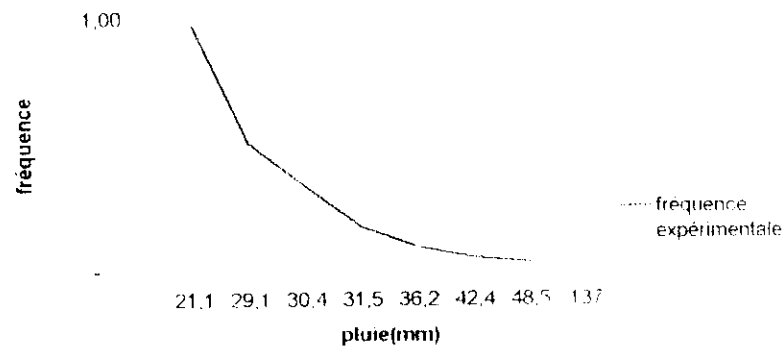


Fig IV.11 La distribution expérimentale des pluies annuelles

On remarque que la plus grande quantité de pluie enregistrée à Alger a une probabilité d'apparition de 2%, alors que la plus faible a une probabilité d'apparition de 98%.

IV.3 Caractérisations des années sèches :

Le concept de sécheresse est généralement mal défini. pour Larousse, c'est l'état de ce qui est sec. Les autres définitions se rapportent plutôt aux effets ou à l'impact des cumuls de quantités de précipitations inférieures à des valeurs normales sur l'agriculture (déséquilibre hydrique) et aux ressources en eau (pénurie d'eau).

IV.3.1 Définition de la période sèche :

En ce qui concerne la définition d'une période sèche, deux approches ont été proposées selon les paramètres utilisés [22]

- A partir des mesures de pluies et de températures : les diagrammes ombrothermiques de Bagnouls et Gausson permettent de fixer le début et la fin d'une période sèche aux intersections des courbes des valeurs mensuelles de la pluviométrie et de la température
- A partir de l'évapotranspiration potentielle : cette approche est plus exacte, elle consiste à combiner la pluviométrie et les mesures d'évapotranspiration potentielle, le diagramme obtenu permet de définir la longueur de la période sèche, elle correspond au nombre de mois, de décades ou de jours où la pluie est soit inférieure à l'évapotranspiration potentielle, soit égale à 50% ou 35% de l'évapotranspiration potentielle.

Pour étudier l'évolution de la longueur des saisons seches à la station d'Alger , on a utilisé un fichier des températures mensuelles disponibles à l'ONM et qui s'étalent de 1952 à 2001.

Les diagrammes ombrothermiques représentés par la figure IV 12 montre que la durée de la saison sèche varie de 7 mois (avril- octobre) à 9 mois (mars- novembre) .

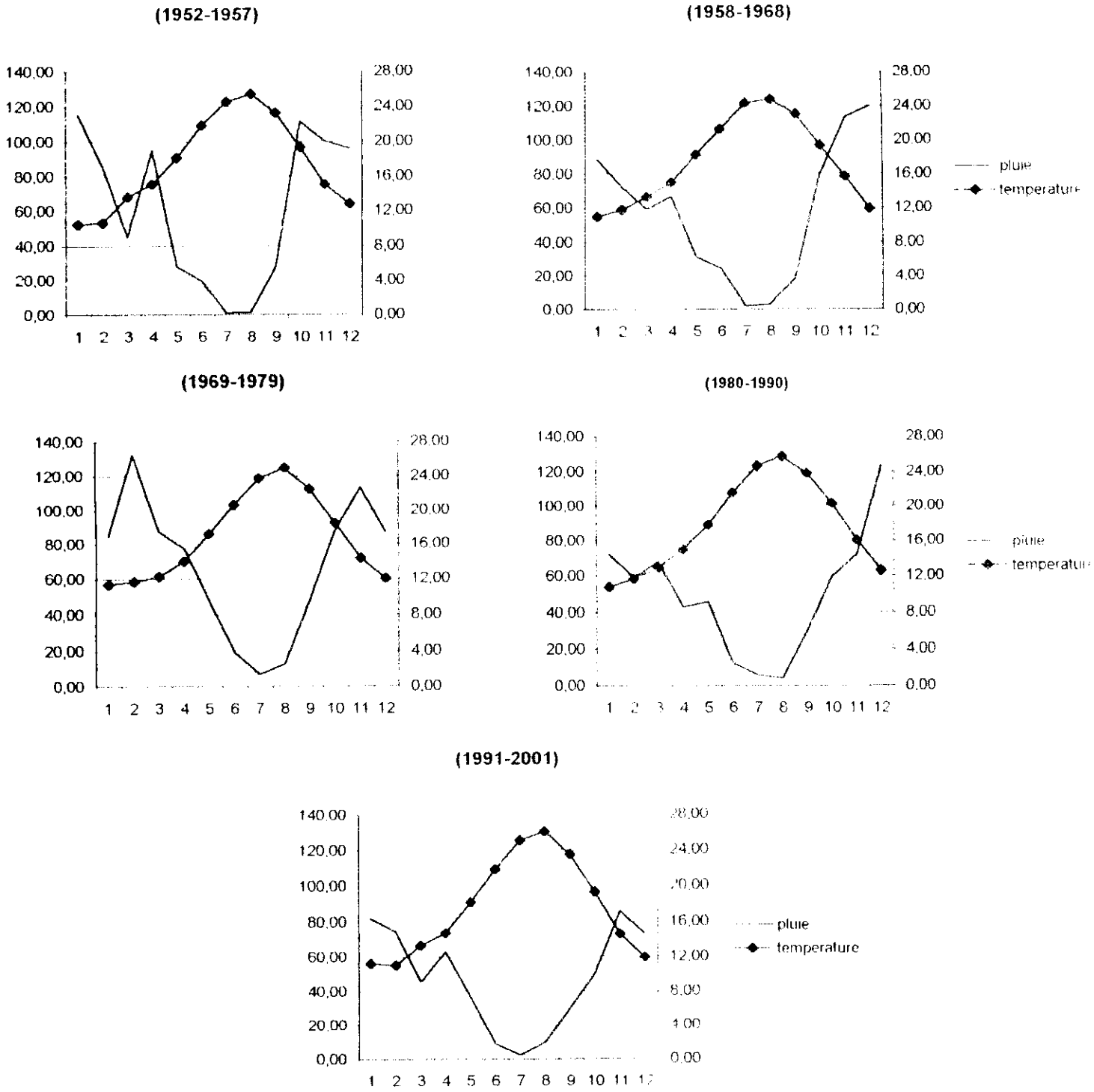


Fig.IV.12 Caractérisation d'une période sèche à partir des mesures de pluies et de température.

IV.3.2 Variation de la séquence sèche :

La figure IV.13 représente la variation du nombre des mois secs (pluie < 30mm) [23], ainsi que la longueur de la saison sèche pour chaque année.

On constate qu'il n'existe aucune tendance à un assèchement climatique pour la station d'Alger et mais qu'il existe une saison sèche de durée variable en été, avec parfois des interruptions et des raccourcissements de la saison pluvieuse.

En outre, on remarque que la plus longue séquence sèche enregistrée est de 8 mois en une année et la plus courte n'est que deux mois (saison d'été).

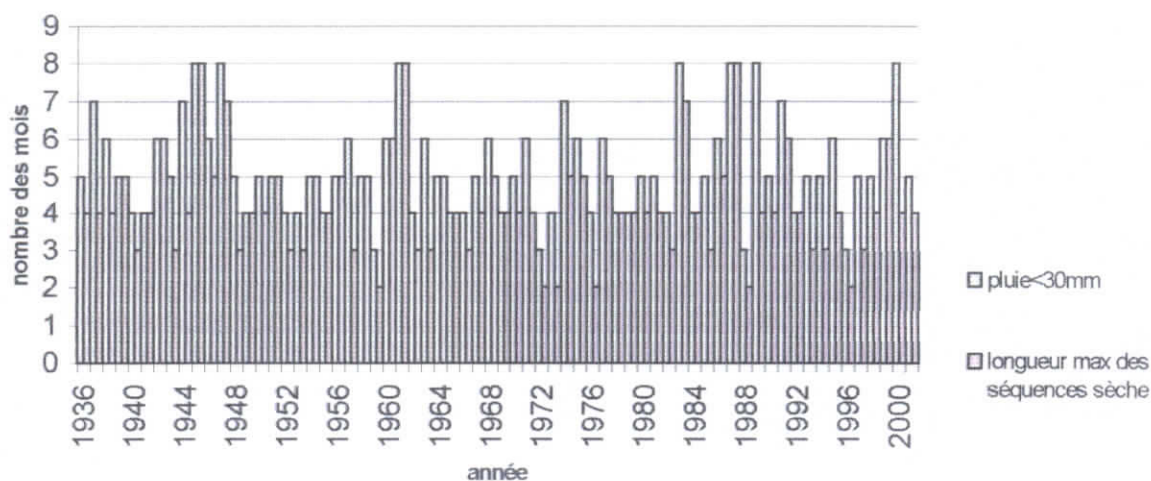


Fig: IV.13 séquences de sécheresses

IV.3.3 Détermination des années sèches :

La définition de la sécheresse est relative au domaine considéré, mais elle est généralement fondée sur un concept essentiel, à savoir : le déficit pluviométrique.

Pour caractériser les années sèches, il faut déterminer les seuils de ce déficit qui sont généralement choisis selon la sévérité désirée puisqu'il n'existe pas de règles préalables.

D'une manière générale, on distingue : [2]

- Un seuil tolérable : c'est le déficit maximum atteint sans qu'il y ait une sécheresse.
- Un seuil de sécurité : c'est le déficit atteint sans crainte de sécheresse.
- Un seuil d'alerte : qui correspond au déficit minimum enregistré pendant une année sèche.
- Un seuil critique : c'est le déficit à partir duquel une sécheresse est certaine.
- Un seuil maximum : c'est le déficit maximum atteint durant la période étudiée.

Le niveau de déficit pluviométrique à partir duquel on peut dire qu'il y a sécheresse ou non constitue souvent une difficulté majeure.

Dans ce cas, on calcule l'écart par rapport à la moyenne, E_i , qui correspond à l'excédent ou au déficit pluviométrique de l'année considérée par rapport à la moyenne des années d'observations

$$E_i (\%) = \frac{P_i - P}{P} \times 100$$

On a utilisé ensuite un indice pluviométrique I pour différents seuils (10, 15, 20 et 30%) ; afin de caractériser les années sèches, normales (moyennes) et humides tels que :

- $I > 1$ $E_i > \alpha$ \rightarrow année humide.
- $I = 0$ $0 < E_i < \alpha$ \rightarrow année normale
- $I < 0$ $- \alpha < E_i < 0$ \rightarrow année normale.
- $I = -1$ $E_i < - \alpha$ \rightarrow année sèche.

Dans cette étude, On a considéré les données annuelles hydrologiques (Sep- Août) et les données annuelles calendaires (Jan- Déc).

Les résultats obtenus sont résumés dans le tableau IV 5 et IV 6.

L'analyse des résultats obtenus montre que :

Année	Nombre d'années sèches	Nombre d'années humides	Nombre d'années sèches durant les 2 dernières décennies	Nombre d'années humides durant les 2 dernières décennies
Calendaire	10	7	8	2
hydrologique	8	11	4	1

L'utilisation des données calendaires donne un nombre d'années sèches plus élevé que les données hydrologiques.

Les années sèches sont pratiquement isolées en utilisant les données hydrologiques, alors que les années calendaires indiquent une séquence de 3 années consécutives à partir de 1989.

Pour l'année hydrologique , le plus grand déficit a été enregistré en 1989 avec 48.74% , alors que l'année 2000 est la plus déficitaire avec 57.04% pour les années calendaires.

Calcul des fréquences des années de secheresses :

Le calcul de cette fréquence permet de déterminer la probabilité d'apparition d'une année ou le seuil de déficit pluviométrique est supérieur à 30%. Celle-ci est donnée par la relation suivante :

$$F = \frac{\text{nombre d'années sèches}}{\text{nombre d'années totales}}$$

Le nombre d'années sèches observées en utilisant les données annuelles calendaires est de 10 sur 66 ans, ce qui donne une probabilité d'apparition de 15.2%, soit en moyenne une année sèche tous les six à sept ans.

Pour les années hydrologiques, la probabilité d'apparition est de 12.3% (8 années sèches sur 65), soit en moyenne une année sur huit.

**Tableau IV.4 : l'écart et l'indice pluviométrique
(pour les années hydrologiques)**

années	E _i (%)	L'indice pluviométrique I				Etat de l'année (pour α = 30%)
		α 10%	α 15%	α 20%	α 30%	
36/37	-5.28	0-	0-	0-	0	normale
37/38	-32.84	1-	1-	1-	1	sèche
38/39	33.42	1+	1+	1+	1+	humide
39/40	-12.56	1-	0-	0-	0-	normale
40/41	-31.06	1-	1-	1-	1	sèche
41/42	-24.46	1-	1-	1-	0	normale
42/43	-18.87	1-	1-	0-	0	normale
43/44	-8.26	0-	0-	0-	0-	normale
44/45	-15.1	1-	1-	0-	0-	normale
45/46	28.09	1+	1+	1+	0-	humide
46/47	-42.29	1-	1-	1-	1-	sèche
47/48	18.12	1+	1+	0+	0-	normale
48/49	-6.56	0-	0-	0-	0-	normale
49/50	-3.95	0-	0-	0-	0-	normale
50/51	-20.06	1-	1-	1-	0-	sèche
51/52	31.55	1+	1+	1+	1-	humide
52/53	-1.74	0-	0-	0-	0-	normale
53/54	12.97	1+	0+	0+	0+	normale
54/55	-14.88	1-	0-	0-	0	normale
55/56	-0.94	0-	0-	0-	0-	normale
56/57	-11.8	1-	0-	0-	0	normale
57/58	69.32	1+	1+	1+	1-	humide
58/59	20.45	1+	1+	1+	0-	normale
59/60	53.13	1+	1+	1+	1-	humide
60/61	-29.76	1-	1-	1-	0-	normale
61/62	20.59	1+	1+	1+	0+	normale
62/63	-4.76	0-	0-	0-	0-	normale
63/64	-2.62	0-	0-	0-	0-	normale
64/65	5.82	0+	0+	0+	0-	normale
65/66	-19.98	1-	1-	0-	0-	normale
66/67	-2.32	0-	0-	0-	0-	normale
67/68	10.56	1+	0+	0+	0+	normale
68/69	6.65	0+	0+	0+	0+	normale
69/70	42.65	1+	1+	1+	1+	humide
70/71	7.84	0+	0+	0+	0-	normale
71/72	40.93	1+	1+	1+	1+	humide
72/73	81.03	1+	1+	1+	1+	humide
73/74	37.14	1+	1+	1+	1-	humide
74/75	5.2	0+	0+	0+	0+	normale
75/76	25.38	1+	1+	1+	0+	humide
76/77	-15.37	1-	1-	0-	0-	normale
77/78	5.99	0+	0+	0+	0-	normale
78/79	-9.29	0-	0-	0-	0-	normale
79/80	21.99	1+	1+	1+	0+	normale
80/81	-3.03	0-	0-	0-	0-	normale
81/82	-14.17	1-	0-	0-	0-	normale
82/83	-2.19	0-	0-	0-	0-	normale
83/84	-8.42	0-	0-	0-	0-	normale
84/85	42.18	1+	1+	1+	1+	humide
85/86	-13.83	1-	0-	0-	0-	normale
86/87	5.1	0+	0+	0+	0+	normale
87/88	-26.84	1-	1-	1-	0	normale

88/89	-20.75	1-	1-	1-	0-	normale
89/90	-48.74	1-	1-	1-	1-	sèche
90/91	-26.49	1-	1-	1-	0-	normale
91/92	-2.67	0-	0-	0-	0-	normale
92/93	-17.69	1-	1-	0-	0-	normale
93/94	-35.8	1-	1-	1-	1-	sèche
94/95	5.76	0+	0+	0+	0+	normale
95/96	14.01	1+	0+	0+	0+	normale
96/97	-34.45	1-	1-	1-	1-	sèche
97/98	0.69	0+	0+	0+	0-	normale
98/99	-0.91	0-	0-	0-	0-	normale
99/00	-19.71	1-	1-	0-	0-	normale
00/01	-36.11	1-	1-	1-	1-	sèche

**Tableau IV.5 : l'écart et l'indice pluviométrique
(pour les années calendaires)**

années	E _i (%)	L'indice pluviométrique I				Etat de l'année (pour α= 30%)
		α= 10%	α= 15%	α= 20%	α= 30%	
1936	32.87	1+	1+	1+	1+	humide
1937	-37.65	1-	1-	1-	1-	sèche
1938	-22.02	1-	1-	1-	0-	normale
1939	19.66	1+	1+	0+	0+	normale
1940	-16.51	1+	1+	0-	0+	normale
1941	-34.63	1-	1-	1-	1-	sèche
1942	-13.35	1-	0-	0-	0-	normale
1943	-0.24	0-	0-	0-	0-	normale
1944	-17.28	1-	1-	0-	0-	normale
1945	-28.04	1-	1-	1-	0-	normale
1946	14.84	1+	0+	0+	0+	normale
1947	0.65	0+	0+	0+	0+	normale
1948	-19.83	1-	1-	0-	0-	normale
1949	-4.81	0-	0-	0-	0-	normale
1950	6.77	0+	0+	0+	0-	normale
1951	12.91	1+	0+	0+	0-	normale
1952	-3.88	0-	0-	0-	0-	normale
1953	-3.57	0-	0-	0-	0-	normale
1954	8.02	0+	0+	0+	0+	normale
1955	-4.46	0-	0-	0-	0-	normale
1956	2.68	0+	0+	0+	0+	normale
1957	66.23	1+	1+	1+	1+	humide
1958	16.89	1+	1+	0+	0-	normale
1959	22.7	1+	1+	1+	0+	normale
1960	18.8	1+	1+	0+	0+	normale
1961	-26.44	1-	1-	1-	0-	normale
1962	29.03	1+	1+	1+	0+	normale
1963	-0.76	0-	0-	0-	0-	normale
1964	2.94	0+	0+	0+	0+	normale
1965	-15.29	1-	1-	0-	0-	normale
1966	-10.45	1-	0-	0-	0-	normale
1967	16.71	1+	1+	0+	0+	normale
1968	-11.42	1-	0-	0-	0-	normale
1969	53.89	1+	1+	1+	1+	humide
1970	10.19	1+	0+	0+	0+	normale
1971	8.74	0+	0+	0+	0+	normale
1972	52.27	1+	1+	1+	1+	humide
1973	78.82	1+	1+	1+	1+	humide
1974	19.7	1+	1+	0+	0+	normale
1975	8.83	0+	0+	0+	0-	normale
1976	16.52	1+	1+	0+	0+	normale
1977	-24.08	1-	1-	1-	0-	normale
1978	8.82	0+	0+	0+	0+	normale
1979	22.03	1+	1+	1+	0-	normale
1980	11.29	1+	0+	0+	0+	normale
1981	-33.59	1-	1-	1-	1-	sèche
1982	33.55	1+	1+	1+	1+	humide
1983	-49.09	1-	1-	1-	1-	sèche
1984	35.18	1+	1+	1+	1+	humide
1985	8.14	0+	0+	0+	0-	normale
1986	14.15	1+	0+	0+	0+	normale
1987	-25.53	1-	1-	1-	0-	normale
1988	-11.69	1-	0-	0-	0-	normale

1989	-51.09	1-	1-	1-	1-	sèche
1990	-30.81	1-	1-	1-	1-	sèche
1991	-34.05	1-	1-	1-	1-	sèche
1992	14.54	1+	0+	0+	0+	normale
1993	-25.09	1-	1-	1-	0-	normale
1994	-30.09	1-	1-	1-	1-	sèche
1995	-15.62	1-	1-	0-	0-	normale
1996	22.63	1+	1+	1+	0+	normale
1997	-16.4	1-	1-	0-	0-	normale
1998	-6.8	0-	0-	0-	0-	normale
1999	23.18	1+	1+	1+	0+	normale
2000	-57.04	1-	1-	1-	1-	sèche
2001	-32.52	1-	1-	1-	1-	sèche

IV.4 Etude de la pluviométrie par les chaînes de Markov :

Sur la base des données pluviométriques mensuelles de la station d'Alger, on a appliqué un modèle Markovien multiclassé saisonnier d'ordre 1 sur trois périodes distinctes :

1936-1957

1958-1979

1980-2001

Cependant, on a subdivisé l'année calendaire en :

1. Une saison
2. Deux saisons : - du mois d'octobre au mois de d'avril (Octobre- avril)
- du mois de mai au mois de septembre (mai- septembre)
3. Quatre saisons reparties comme suit : - décembre- février
- mars- mai
- juin- août
- septembre- novembre

Nombre des classes =2

Les seuils sont choisis comme suit :

- Pour une saison : 30mm
- Pour deux saisons : 50mm(1^{ère} saison)
5mm (2^{ème} saison)
- Pour quatre saisons : 50mm (1^{ère}, 2^{ème} et 4^{ème} saison)
5mm (3^{ème} saison).

Résultats et interprétations :

a) Pour une saison :

Les résultats obtenus sont donnés dans le tableau IV 7

La comparaison des trois périodes étudiées montre une différence non significative des probabilités conditionnelle, on peut dire donc que le régime pluviométrique est pratiquement stable à l'échelle annuelle.

Tableau IV.7 : Probabilité conditionnelle d'ordre 1 pour une saison.

période	Probabilité conditionnelle d'ordre 1			
	P ₀₀	P ₀₁	P ₁₀	P ₁₁
1936-1957	0.65	0.34	0.27	0.72
1958-1979	0.60	0.39	0.27	0.72
1980-2001	0.63	0.36	0.30	0.70

b) Pour deux saisons :

Les résultats du tableau IV.8 montre qu'il n'y a pas de changements significatifs des probabilités pour la saison sèche, par contre pour la saison humide, on remarque une augmentation des probabilités des mois successifs pluvieux (>30mm) dans la 2^{ème} période (de 65% à 72%) puis une diminution jusqu'à 53% dans la dernière période.

Tableau IV.8 : Probabilité conditionnelle d'ordre 1 pour deux saisons.

période	saison	Probabilité conditionnelle d'ordre 1			
		P ₀₀	P ₀₁	P ₁₀	P ₁₁
1936-1957	1	0.40	0.60	0.34	0.65
	2	0.50	0.50	0.48	0.51
1958-1979	1	0.34	0.65	0.27	0.72
	2	0.50	0.50	0.40	0.60
1980-2001	1	0.40	0.60	0.46	0.53
	2	0.50	0.50	0.44	0.56

c) Pour quatre saisons :

Les résultats du tableau IV.9 montrent que les probabilités de la 2^{ème} saison (mars- juin) sont plus ou moins proches de celles de la 4^{ème} saison (septembre- novembre), donc elles ont presque le même régime pluviométrique

En outre, On remarque une augmentation des probabilités des mois successifs humides dans la 2^{ème} période et une diminution des fortes précipitations a été observée dans la dernière période pour les quatre saisons par rapport aux deux périodes précédentes.

Tableau IV.9 : Probabilité conditionnelle d'ordre 1 pour quatre saisons.

période	saison	Probabilité conditionnelle d'ordre 1			
		P ₀₀	P ₀₁	P ₁₀	P ₁₁
1936-1957	1	0.17	0.82	0.26	0.73
	2	0.68	0.31	0.51	0.48
	3	0.70	0.30	0.83	0.16
	4	0.55	0.44	0.52	0.47
1958-1979	1	0.13	0.86	0.27	0.72
	2	0.52	0.47	0.41	0.58
	3	0.48	0.51	0.61	0.38
	4	0.52	0.47	0.28	0.71
1980-2001	1	0.40	0.6	0.34	0.65
	2	0.55	0.44	0.62	0.37
	3	0.70	0.30	0.65	0.34
	4	0.60	0.40	0.58	0.41

De point de vue, nombre des mois pluvieux, on remarque une diminution des probabilités conditionnelles (P_{11}) des mois pluvieux lors de dernière période (1980- 2001), ceci ne peut s'expliquer que par le fait que le nombre de mois pluvieux a diminué, ce qui sous entend une diminution des fortes précipitations durant cette période

IV.5 Conclusion :

L'analyse de la pluviométrie à différents pas du temps de la station d'Alger, illustre le caractère aléatoire du phénomène étudié, les résultats montrent une variation temporelle des précipitations et une apparition des groupes d'années excédentaires et déficitaires qui se répètent d'une manière aléatoire plutôt que régulière.

L'évolution du coefficient de variation en fonction du temps met en évidence l'irrégularité du régime pluviométrique.

L'étude comparative entre les 6 décennies, a révélé que la dernière décennie est la plus déficitaire (43.37% par rapport à la décennie la plus pluvieuse) et que ce déficit n'est que 14.24% par rapport à la période totale.

L'analyse des pluies mensuelles a montré l'existence de deux saisons, l'une humide (octobre- avril) et l'autre sèche (mai- septembre), de ce fait, l'année peut être caractérisée seulement par deux saisons dans la région d'Alger

Le calcul du nombre de jours pluvieux montre que le déficit pluviométrique des deux dernières décennies est dû à une baisse des fortes précipitations, du fait que ce nombre n'a pas changé par rapport à ceux des décennies précédentes.

La durée de la saison sèche sur la période s'étalent de 1952 à 2001 varie de 7 (avril-octobre) à 9 mois (mars - novembre) par année.

La distribution des pluies dans le temps par rapport au seuil considéré montre une alternance de séquences sèches et pluvieuses non périodiques, les sécheresses enregistrées non pas étaient nombreuses, rarement persistantes et de sévérité inégale.

Les années calendaires montre une succession de trois années sèches à partir de 1989 par rapport à un seuil de 30% et le plus grand déficit a été enregistré en l'an 2000 (plus de 57%).

L'analyse de la pluviométrie par les chaînes de Markov a montré une diminution des probabilités conditionnelles des mois pluvieux lors de la dernière période (1980- 2001). Ce résultat confirme que le déficit pluviométrique enregistré lors des deux dernières décennies est dû essentiellement à la diminution des fortes précipitations et non pas au nombre de jours pluvieux.

Chapitre V

Le modèle de prévision

Le modèle de prévision

V.1 Introduction :

La prévision est une estimation préalable de l'état futur d'une grandeur bien déterminée à différentes échéances.

Les déficits et les excédents de précipitations (sécheresse - inondations), les fortes pluies d'orage dans les régions très urbanisées ont, sur les activités humaines, des conséquences souvent dommageables. Pour éviter ce genre de situation et avoir une bonne gestion des ressources en eau, la prévision des précipitations demeure indispensable et ne serait efficace en dehors d'une vision globale des différents facteurs constituant le régime hydrologique, à savoir, les indicateurs de l'atmosphère (température, pression, vent, humidité) et même les caractéristiques physiques du bassin versant, l'état du couvert et le degré de saturation des sols.

La question qui persiste et qui mérite beaucoup d'attention est relative à la connaissance de l'événement (sécheresse ou manque d'eau) à des échéances allant d'un mois, une saison voir plus à l'avance, cette question d'importance capitale commence à connaître un début de réponse avec les récents progrès enregistrés par le projet de la prévision mensuelle et saisonnière des précipitations sur le bassin méditerranéen. [8]

Plusieurs tentatives de mise au point d'une méthode fiable pour une prévision à long terme (≥ 1 mois) n'ont pas abouti jusqu'à présent, la difficulté peut être liée au :

- Caractère aléatoire du phénomène.
- La longueur des séries étudiées.
- Les changements climatiques.

Le modèle de prévision étudié est basé sur l'analyse en composantes principales appliqué dans un contexte prévisionnel sur des variables brutes et normées.

L'approche préconisée ne tiendra compte que du paramètre étudié à savoir la pluviométrie, pour lequel, on dispose des mesures mensuelles historiques de la station d'Alger entre 1936-2001.

V.2 Le modèle de prévision :

Un modèle est toujours un compromis entre les deux nécessités opposées :

- Elaborer un modèle suffisamment complet et raffiné pour décrire le système étudié le plus fidèlement possible.
- Elaborer un modèle suffisamment simple pour que sa formulation symbolique (par exemple une équation) soit possible, et que son exploitation ne soit pas trop lourde.

V.2.1 Calage du modèle :

Une des procédures indispensables dans une modélisation est la tentative de calage, qui consiste à déterminer les paramètres d'ajustement (paramètres internes) à partir d'un jeu de données et de réponses mesurées à partir de la série historique.

La prévision est rarement assez fidèle pour garantir que l'on ait bien la réponse attendue à une sollicitation donnée du modèle. Il y a donc souvent lieu d'introduire, outre les paramètres objectifs, un certain nombre de paramètres d'ajustement, que l'on évaluera au cours de cette phase de développement.

V.2.2 Test du modèle :

Le test du modèle est une opération indispensable dans le processus d'élaboration. Il consiste à vérifier l'aptitude du modèle calé, à prévoir la réalité et permet d'évaluer les performances finales du modèle.

Bien que très fréquemment négligé, le test est le seul moyen objectif permettant de préciser l'ordre de grandeur des erreurs à attendre lors de l'exploitation du modèle.

Pour le modèle de prévision développé sur une période d'observation de 66ans, on choisit une période de calage de 56ans et une période test de 10ans.

V.3 Critère de validation :

Pour juger la qualité d'une prévision, on fait appel à des fonctions critères qui permettent d'apprécier quantitativement et qualitativement les résultats mensuels, estimants globalement l'écart entre les sorties calculées (pluies mensuelles prévues) et les variables historiques, il existe deux types de critères :

- Les critères graphiques : l'analyse graphique de n'importe quel modèle est indispensable et primordial, cela est obtenu en portant sur un graphique les valeurs observées et les valeurs calculées sur une période test.
- Les critères numériques : [23]

1. Le critère de Nash (NS) : ce critère introduit par Nash et Sutcliffe (1970), est défini par :

$$NS = \left[1 - \frac{\sum_{i=1}^n (P_i - \hat{P}_i)^2}{\sum_{i=1}^n (P_i - \bar{P})^2} \right] \cdot 100 \quad (V.1)$$

Avec :

P_i : la pluie observée.

\hat{P}_i : la valeur prévue par le modèle.

\bar{P} : la moyenne des pluies mensuelles observées.

Le critère de Nash peut être interprété comme étant la proportion de la valeur calculée par le modèle.

Si $NS=100\%$, la prévision est parfaite, par contre si $NS<0$, cette dernière est mauvaise.

2. Moyenne biaisée : elle définit la différence entre la moyenne des pluies observées et prévues. Ce paramètre est donné par la relation suivante :

$$\{B\} = M_{\text{prév}} - M_{\text{obs}} \quad (\text{V.2})$$

Avec :

$M_{\text{prév}}$: moyenne des pluies prévues durant la période test.

M_{obs} : moyenne des pluies observées durant cette période.

Lorsque $\{B\}$ tend vers zéro, les moyennes se coïncident.

3. Coefficient de variation :

La validation du modèle peut être complétée en comparant les coefficients de variation des pluies observées avec ceux calculées.

$$I = Cv_c / Cv_o \quad (\text{V.3})$$

Avec :

Cv_c : coefficient de variation des pluies calculées pendant la période test.

Cv_o : coefficient de variation des pluies observées pendant la période test.

Si l'indice I est proche de l'unité, la prévision est parfaite.

V.4 Elaboration du modèle de prévision :

Les composantes principales sont des combinaisons linéaires des variables initiales X_j , on écrit :

$$C_i = \sum_{j=1}^p a_{ij} X_j \quad (\text{V.4})$$

Avec : C_i : $i^{\text{ème}}$ composante principale

a_{ij} : cosinus directeurs, élément de la matrice $[A]$ des vecteurs propres.

X_j : variable d'ordre J .

La CP normée s'écrit : $\hat{C}_i = C_i / \lambda_i^{1/2}$ avec λ_i : $i^{\text{ème}}$ valeur propre.

L'équation (V.4) devient :

$$\hat{C}_i = \lambda_i^{-1/2} \sum_{j=1}^p a_{ij} X_j \quad (\text{V.5})$$

Sous la forme matricielle, l'équation (V.5) s'écrit :

$$[\hat{C}] = [\lambda]^{-1/2} [A][X] \quad (V.6)$$

D'où :

$$[X] = [\lambda]^{1/2} [A]^T [\hat{C}] \quad (V.7)$$

X : matrice des variables initiales.

λ : vecteur des valeurs propres.

\hat{C} : matrice des composantes principales normées.

A^T : transposée de la matrice des vecteurs propres.

Sachant qu'on ne considère que les Q premières C.P.

La perte d'information sera prise en compte par le terme résiduel noté ε_j

L'équation (V.7) devient :

$$X_j = \lambda_q^{1/2} \sum_{l=1}^Q a_{jl} \hat{C}_l + \varepsilon_j \quad (V.8)$$

ε_j : variable résiduelle de moyenne nulle et d'écart type :

$$\sigma_{\varepsilon_j} = \left[1 - \sum_{l=1}^Q a_{jl}^2 \lambda_l \right]^{1/2}$$

En détermine $\text{COV}(\hat{C}_l, X_j)$:

$$\text{COV}(\hat{C}_l, X_j) = \sum_{i=1}^n \hat{C}_{il} \left[\lambda_q^{1/2} \sum_{l=1}^Q a_{ij} \hat{C}_{il} \right] = \lambda_q^{1/2} a_{ij} \sum_{i=1}^n \hat{C}_{il} \hat{C}_{il}$$

La quantité $\sum_{i=1}^n \hat{C}_{il} \hat{C}_{il}$ représente la norme du vecteur \hat{C}_l et par conséquent elle est égale à un .

$$\text{Donc : } \text{COV}(\hat{C}_l, X_j) = \lambda_q^{1/2} a_{ij}$$

L'équation (V.8) s'écrit sous la forme :

$$X_j = \sum_{l=1}^Q \text{COV}(\hat{C}_l, X_j) \hat{C}_l + \varepsilon_j \quad (V.9)$$

L'expression finale du modèle de prévision est donnée par :

$$X_j = \sum_{l=1}^Q \beta_{jl} \hat{C}_l + \varepsilon_j \quad (V.10)$$

Avec :

β_{jl} : coefficient de régression entre la variable X_j et la CP \hat{C}_l

En pratique, les calculs de C.P s'effectuent sur des variables centrées réduites :

Soit la variable centrée réduite : $Y_j = \frac{X_j - \bar{X}_j}{\sigma_j}$

La composante principale normée \hat{C}_1 s'écrit :

$$\hat{C}_1 = \lambda^{1/2} \sum_{j=1}^p a_{1j} \frac{X_j - \bar{X}_j}{\sigma_j} \quad (V.11)$$

Sous la forme matricielle, l'expression (V.11) s'écrit :

$$[\hat{C}] = [\lambda]^{1/2} [A][Y] \quad (V.12)$$

d'où :

$$[Y] = [\lambda]^{1/2} [A]^T [\hat{C}] \quad (V.13)$$

sachant que :

$$\begin{aligned} \lambda^{1/2} a_{1j} &= \text{COV}(\hat{C}_1, Y_j) \\ \lambda^{1/2} a_{1j} &= \text{COV}(\hat{C}_1, \frac{X_j - \bar{X}_j}{\sigma_{x_j}}) = \text{COR}(\hat{C}_1, X_j) \end{aligned}$$

L'équation (V.13) devient :

$$Y_j = \sum_{l=1}^q \text{COR}(\hat{C}_l, X_j) \hat{C}_l + \varepsilon_j \quad (V.14)$$

D'où :

$$X_j = \bar{X}_j + \sigma_{x_j} \sum_{l=1}^m \text{COR}(\hat{C}_l, X_j) \hat{C}_l + \sigma_{x_j} \varepsilon_j \quad (V.15)$$

L'expression finale du modèle de prévision est donnée par :

$$X_j = \beta_{j0} + \sum_{l=1}^m \beta_{jl} \hat{C}_l + E_j \quad (V.16)$$

Avec :

$\beta_{j0} = \bar{X}_j$ (la moyenne de la variable d'ordre j).

$E_j = \sigma_{x_j} \varepsilon_j$

En se basant sur la structure de ces deux modèles de prévision, on détermine :

- Les éléments de la matrice $[\beta]$: les vecteurs de la matrice ont pour composantes les coefficients de régression entre les variables et les CP.

Dans le cas d'une ACP normée , celle-ci est constituée d'un premier vecteur β_0 , dont les éléments sont les moyennes mensuelles interannuelles de la série historique.

- Les éléments de la matrice des CP : $[\hat{C}]$.
- Les éléments de la matrice des résidus $[E]$.

L'estimation de ces paramètres peut se faire de différentes manières :

- Par les fonctions de répartition.
- Par les lois d'ajustement.
- Par les chaînes de Markov.

Pour le modèle de prévision développé, on s'est intéressé aux chaînes de Markov, le principe repose sur l'utilisation d'une seule composante principale afin d'avoir le maximum d'information sur les résidus.

V.4.1 1^{ère} approche :

a) ACP brute :

- Etude des β_{ij} :

Pour la détermination de la matrice β_{ij} , on a été amené à étudier l'évolution de ces coefficients pour une seule CP sur différentes périodes (40, 46, 48, 50, 52, 54, 56ans).

La figure V.1 montre que les β_{ij} ont tendance à suivre la même évolution quand la taille de l'échantillon augmente et les courbes deviennent presque identiques.

En outre, on constate, que les coefficients de régression se stabilisent au fur à mesure que les dimensions de l'échantillon (années) augmente.

Le tableau V.1 donne le nombre d'années pour lequel β_{ij} est stable et la valeur correspondante à chaque mois.

Tableau V.1 : les valeurs de β_{ij} pour chaque mois

Mois	Valeur de β_{ij} ans de stabilité
Janvier	-3.44
Février	5.40
Mars	1.96
Avril	-1.14
Mai	9.47
Juin	2.56
Juillet	-0.29
Août	1.41
Septembre	8.63
Octobre	21.37
Novembre	-44.68
Décembre	60.88

- Détermination des CP :

La figure V.2 représente l'évolution de la 1^{ère} composante principale en fonction du temps, et montre une variation aléatoire des CP.

Comme le modèle de prévision préconisé est basé sur l'utilisation d'une seule composante principale, on considère qu'il n'est pas nécessaire de faire appel à une nouvelle approche pour prévoir cette CP. On propose, tout simplement, de prendre à chaque fois la valeur de la CP historique précédente.

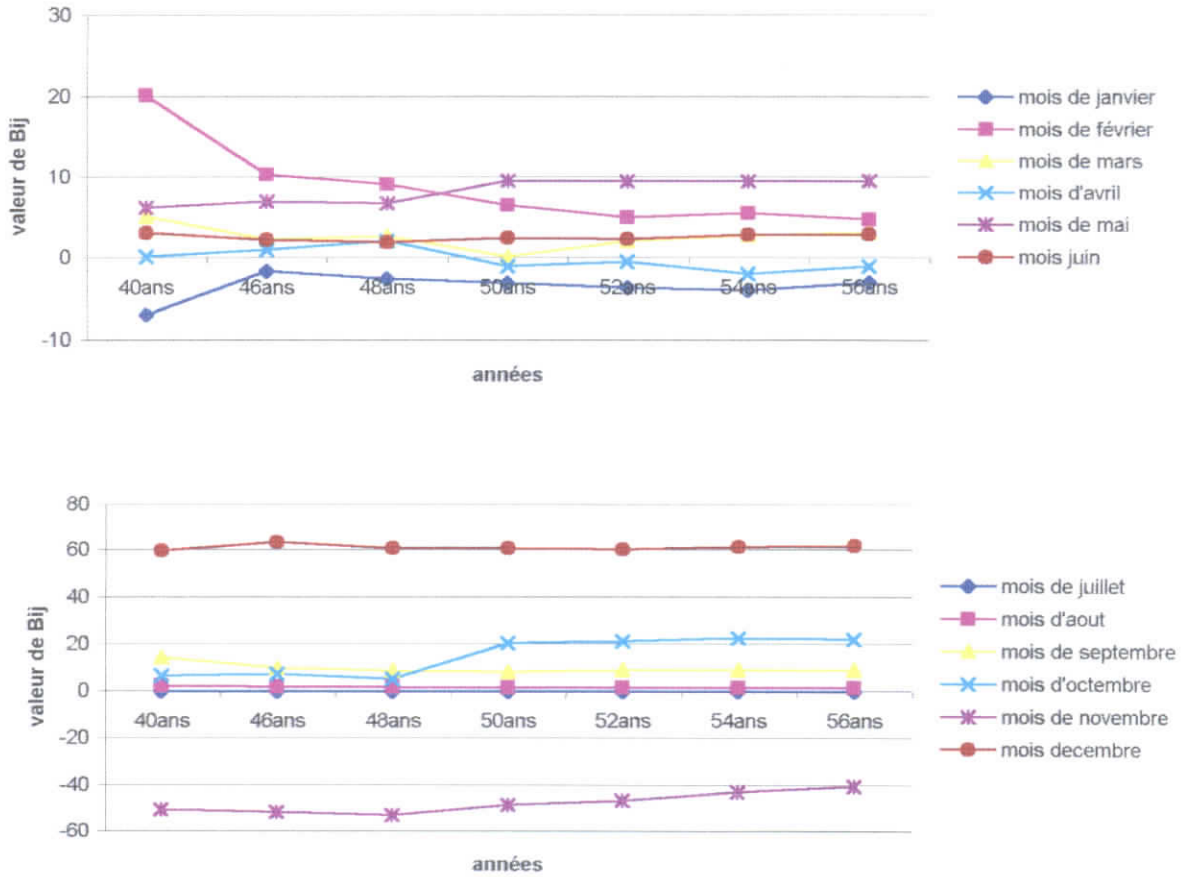


Fig :V.1 l'évolution des coefficients de régression en fonction du temps(variable brute)

- Estimation des variables résiduelles :

On définit le résidu \mathcal{E}_{ij} comme étant la perte d'informations dûs aux CP négligées, il est donné par la différence entre la variable observée et celle prévu.

$$\varepsilon_j = \bar{\lambda}_j - \sum_{l=1}^m \beta_{jl} \acute{C}_l$$

La figure V.3 qui illustre l'évolution résiduelle en fonction du temps, montre une variation aléatoire des \mathcal{E}_{ij} .

L'approche préconisée pour l'estimation de ce paramètre est basée sur l'application d'un modèle markovien multiclasse d'ordre 1.

Les bornes sont choisies de telle manière à avoir une bonne répartition des valeurs observées dans toutes les classes, pour cela , on propose :

- Un nombre de classe égale à 4
- Des bornes de classes :

Classe0	classe1	classe2	classe3
$]-\infty \ 50[$	$[50 \ 100[$	$[100 \ 200[$	$[200 \ \infty[$

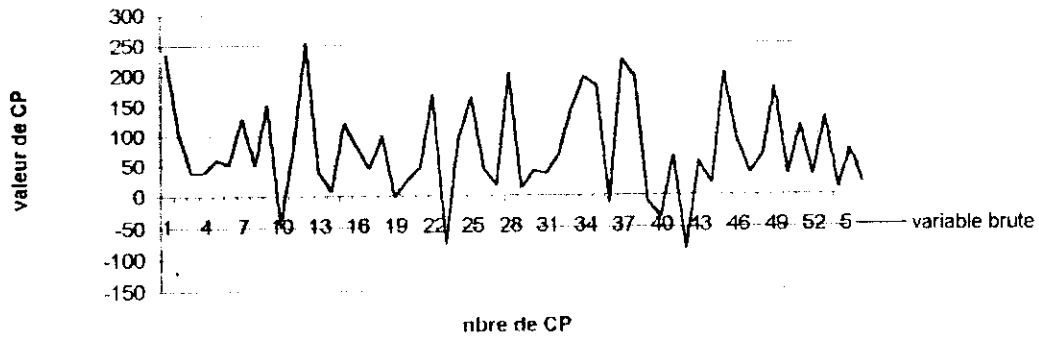


Fig : V. 2 Evolution de la 1ère composante principale

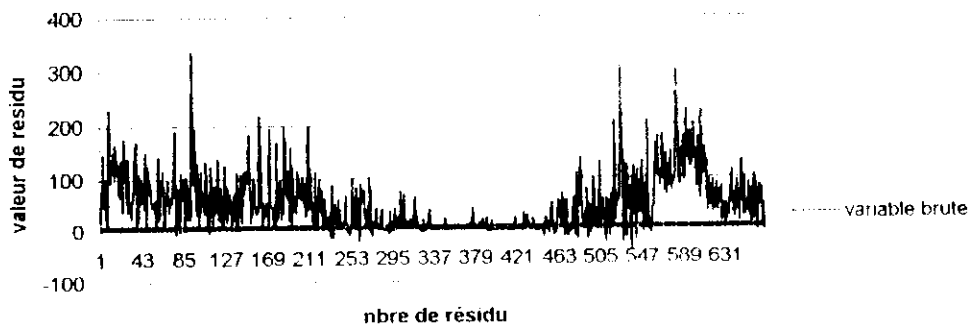


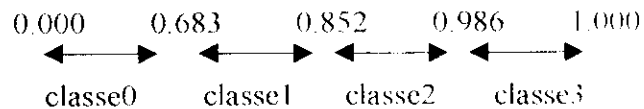
Fig V.3 Evolution des variables résiduelles

Exemple :

Pour prévoir l'année 1992 , le modèle multiclasse appliqué sur les résidus de 56ans donne la matrice de passage suivante :

		Classe du mois prévu			
		0	1	2	3
Classe du mois précédent	0	0.683	0.169	0.134	0.012
	1	0.523	0.238	0.218	0.019
	2	0.431	0.394	0.156	0.018
	3	0.090	0.363	0.454	0.090

Cette matrice sera représentée par des probabilités conditionnelles cumulées, par exemple, si la classe 0 est la classe initiale, les probabilités s'écrivent comme suit :



En effectuant un tirage d'un nombre aléatoire (compris entre 0 et 1), le résultat obtenu, qui est une probabilité, sera comparé aux seuils des fréquences cumulées et indiquera la classe de ε prévu, en tenant compte de la classe précédente, on lui attribue ensuite la valeur correspondante de la série historique des résidus.

Une fois que les paramètres du modèle sont déterminés, le calcul de la valeur de la pluie mensuelle prévue se fait comme suit :

$$X_j = \sum_{l=1}^m \beta_{jl} \hat{C}_l + \varepsilon_j$$

Pour la prévision du 2^{ème} mois, la même démarche est appliquée, en effectuant un tirage aléatoire mais le résidu dans ce cas là est représenté par le 1^{er} tirage.

Les mêmes opérations se répètent pour prévoir toutes les années (en gardant le même nombre de classes et les mêmes seuils).

Résultats :

La figure V.4 représente la prévision mensuelle de l'année 1992.

Les séries mensuelles observées et prévues des pluies pour les autres années sont représentées dans l'annexe V.1.

Les résultats obtenus par application des tests de validité sur chaque année sont résumés dans le tableau V.2.

Tableau V.2 : résultats des tests de validation sur les variables brutes.

Année	Variable observée			Variable prévue					
	moy	σ	Cv	moy	σ	Cv	C.nash (%)	{B}	l
92	62.41	50.59	0.81	60.42	48.3	0.80	13.00	-1.90	0.99
93	38.57	33.09	0.86	25.70	20.85	0.81	-5.38	-12.80	0.94
94	38.07	35.91	0.94	35.88	28.37	0.79	0.70	-2.19	0.84
95	45.98	48.71	1.06	44.42	50.01	1.13	-5.90	-1.56	1.07
96	66.82	65.74	0.98	37.23	35.57	0.96	22.63	-29.50	0.98
97	45.55	39.37	0.86	43.10	48.53	1.13	-3.14	-2.40	1.31
98	50.77	45.59	0.90	51.52	44.43	0.86	31.07	0.75	0.96
99	67.11	72.76	1.08	42.46	38.04	0.90	51.60	-24.60	0.83
00	23.41	24.75	1.06	47.19	45.67	0.97	-332.5	23.70	0.92
01	36.42	37.51	1.03	42.92	38.86	0.91	-76.09	6.50	0.88

Les résultats de Nash obtenus ne sont pas satisfaisants et les moyennes obtenues ne coïncident pas globalement avec la moyenne des séries observées, ceci peut être expliqué par le fait que les valeurs calculées sont parfois trop éloignées de celles observées.

La figure V.5 montre un léger décalage des valeurs prévues par rapport à l'historique. En outre, on remarque que l'écart entre les valeurs observées et celles calculées de 69 mois ne dépasse pas 30mm, soit en moyenne 57% des mois prévus.

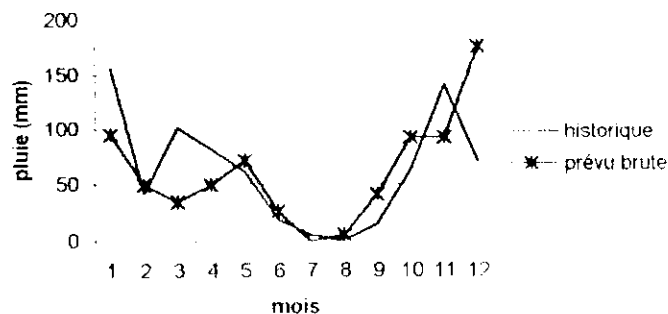


Fig V.4 la série prévue de l'année 1992

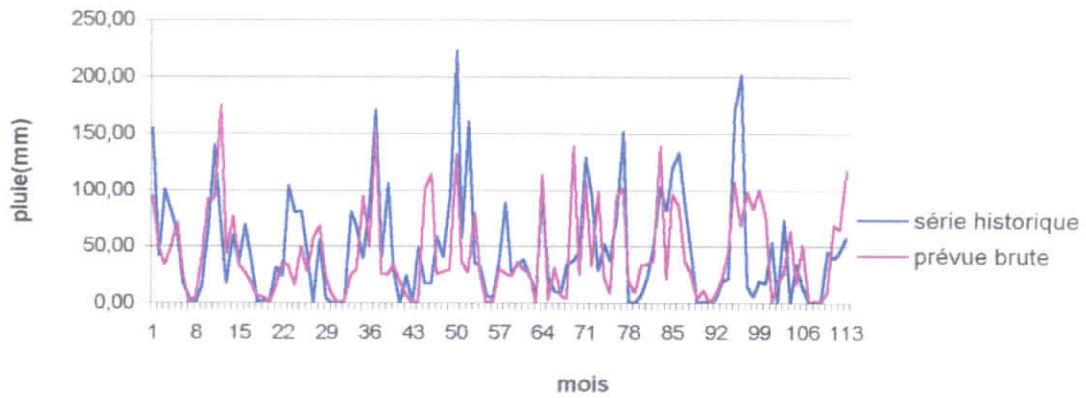


Fig : V.5 prévision des pluies mensuelles (variables brutes)

b) ACP normée :

- Détermination des β_0 :

La méthode repose sur l'analyse de l'évolution de la pluie moyenne mensuelle interannuelle (β_0). Ces moyennes ont été calculées sur différents pas de temps.

La figure V.6 représente l'évolution des valeurs de β_0 pour les mois (janvier, février et mars), celle-ci tend à se stabiliser au fur à mesure que la taille de la série augmente.

L'évolution des moyennes mensuelles interannuelles en fonction du temps des autres mois est représentée dans l'annexe V.2.

On remarque que la période commune pour laquelle β_0 est parfaitement stable pour tous les mois commence à partir de 1979 et la valeur à prendre en considération dans ce modèle de prévision est calculée sur une période de 15 ans (Tableau V.3).

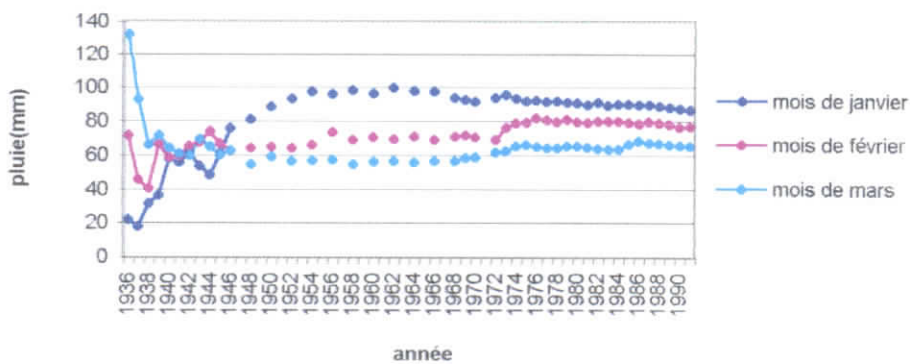


Fig: V.6 l'évolution des moyennes mensuelles interannuelles en fonction du temps

Tableau V.3 : les valeurs de β_0 pour chaque mois.

Mois	$\beta_{0(mm)}$
Janvier	89.7
Février	79.3
Mars	65.5
Avril	66.7
Mai	38.2
Juin	16.3
Juillet	2.8
Août	4.5
Septembre	29.5
Octobre	82.3
Novembre	93.7
Décembre	117.0

- Etude des β_{ij} :

La matrice β_{ij} est éventuellement étudiée de la même manière que pour β_0 . On reprend l'évolution de ces coefficients pour une seule CP sur une période de (40, 46, 48, 50, 52 et 56ans)

Les mêmes constatations sont faites pour le modèle normé, en effet, la figure V.7 qui représente l'évolution des β_{ij} pour chaque mois, montre que ces coefficients ont tendance à suivre la même évolution quand la taille de l'échantillon augmente.

Tableau V.4 : La valeur de β_{ij} correspondante à chaque mois.

Mois	Valeur de β_{ij} 6ans de stabilité
Janvier	-5.81
Février	43.12
Mars	15.15
Avril	-9.90
Mai	-8.84
Juin	10.53
Juillet	1.33
Août	1.57
Septembre	27.27
Octobre	6.34
Novembre	-11.57
Décembre	22.79

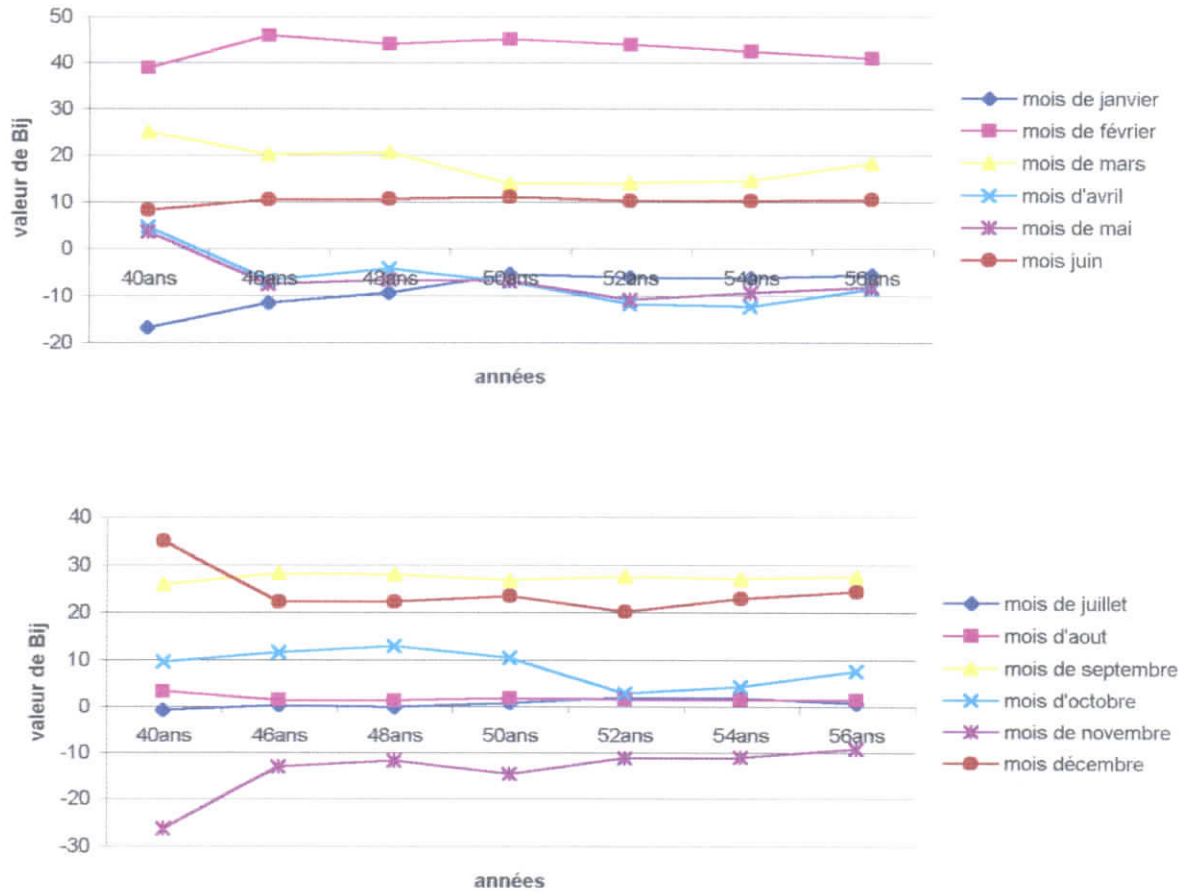


Fig V.7 : l'évolution des β_{ij} pour chaque mois.

- Détermination des CP :

La figure V.8 montre l'évolution aléatoire des CP en fonction du temps.

De ce fait, on applique la même démarche pour la détermination des CP.

- Estimation des variables résiduelles :

L'évolution aléatoire des ε_j est illustrée par la figure V.9. Pour la détermination de ces coefficients, on suit la procédure établie pour le modèle brute.

Le nombre des classes choisi est égal à 4 et les bornes sont :

classe0 classe1 classe2 classe3
 $]-\infty \ 0[$, $[0 \ 50[$, $[50 \ 100[$, $[100 \ \infty[$.

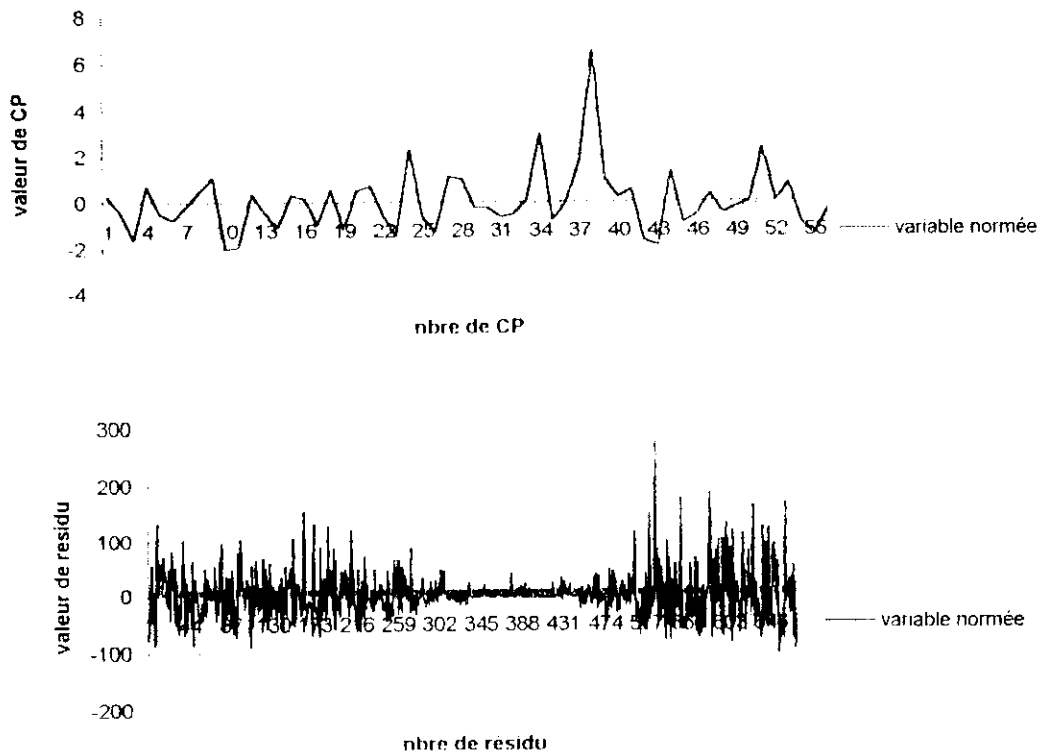


Fig : V.9 Evolution des variables résiduelles

Résultats :

Les séries mensuelles observées et prévues des pluies pour chaque année sont représentés dans la figure V.10 et l'annexe V.1.

La figure V.11 montre toujours un léger décalage des valeurs prévues.

Les résultats obtenus, en appliquant les tests de validations sont résumés dans le tableau V.5.

Tableau V.5 : résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable prévue						
	Moy	σ	Cv	moy	σ	Cv	C.nash (%)	$\{B\}$	I	
92	62.41	50.59	0.81	49.18	47.43	0.96	15.06	-13.23	1.19	
93	38.57	33.09	0.86	53.66	48.16	0.9	-34.82	15.09	1.05	
94	38.07	35.91	0.94	46.02	42.34	0.92	8.31	7.95	0.98	
95	45.98	48.71	1.06	73.81	65.37	0.89	-68.60	27.83	0.84	
96	66.82	65.74	0.98	49.02	44.74	0.91	-1.60	-17.80	0.93	
97	45.55	39.37	0.86	36.99	34.75	0.94	15.36	-8.56	1.09	
98	50.77	45.59	0.90	38.91	43.36	1.11	25.21	-11.86	1.23	
99	67.11	72.76	1.08	65.78	58.10	0.88	82.46	-1.33	0.81	
00	23.41	24.75	1.06	38.27	32.06	0.84	-45.70	14.86	0.79	
01	36.42	37.51	1.03	42.92	38.86	0.91	2.20	6.50	0.88	

L'écart entre les valeurs observées et celles prévues reste non négligeable, ce qui explique l'augmentation de la moyenne biaisée, et les valeurs de Nash obtenues sont généralement non satisfaisantes.

En outre, le modèle normé permet de prévoir 53% des mois avec une différence qui ne dépasse pas 30mm par rapport aux valeurs observées.

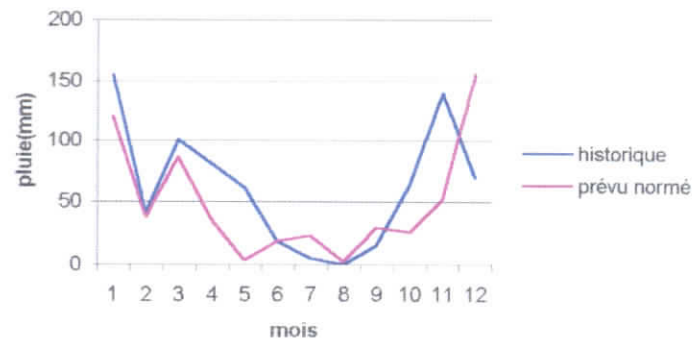


Fig V.10 la série prévue de l'année 1992

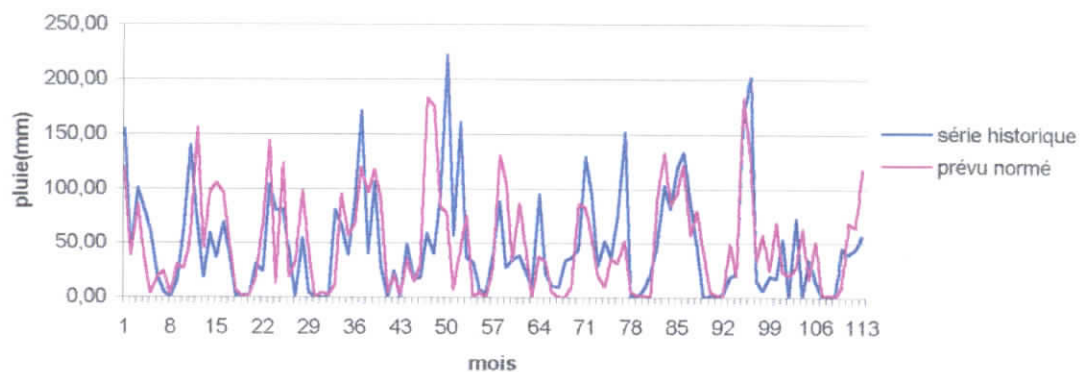


Fig:V.11 prévision des pluies mensuelles

Inconvénient de la technique :

En effectuant un tirage d'un nombre aléatoire, la valeur des résidus extraite de la série historique donne parfois une valeur prévue de précipitation négative ce qui est physiquement impossible, celle-ci est remplacée par une valeur nulle (figure V.12).

Pour éviter ce genre de situation, à savoir la rectification des valeurs prévues des précipitations, on est amené à utiliser la variable racine carrée des données brutes et normées.

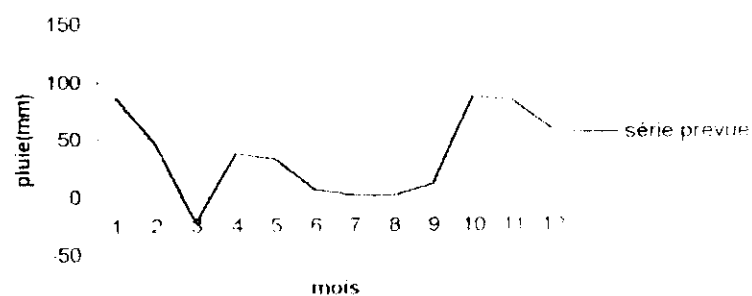


Fig.V.12 prévision de l'année 1997
(variable normée)

c) ACP brute :

L'évolution des coefficients de régression en fonction du temps est représentée par l'annexe V.3, la stabilité de ces derniers a été observée à partir de 1944.

Le comportement des composantes principales appliquées sur la série historique et les variables résiduelles est présenté dans les annexes V.4 et V.5 respectivement.

Résultats :

La présentation graphique des séries prévues est donnée par la figure V.13 et l'annexe V.6.

Le tableau V.6 résume les paramètres statistiques des séries historiques et prévues, ainsi que les résultats obtenus par application des tests de validation :

Tableau V.6 : résultats de la modélisation pour les variables brutes racines.

Année	Variable observée			Variable prévue						
	Moy	σ	Cv	moy	σ	Cv	C.nash (%)	{B}	I	
92	62.41	50.59	0.81	57.91	45.34	0.78	1.29	-4.50	0.96	
93	38.57	33.09	0.86	31.8	25.29	0.80	43.02	-6.77	0.93	
94	38.07	35.91	0.94	42.43	33.05	0.78	26.37	4.36	0.83	
95	45.98	48.71	1.06	38.24	32.91	0.86	-19.88	-7.74	0.81	
96	66.82	65.74	0.98	60.33	66.41	1.10	-90.14	-6.49	1.12	
97	45.55	39.37	0.86	52.61	48.20	0.92	-21.97	7.06	1.07	
98	50.77	45.59	0.9	24.01	23.00	0.96	-50.67	-26.76	1.07	
99	67.11	72.76	1.08	54.94	46.57	0.85	26.59	-12.17	0.79	
00	23.41	24.75	1.06	27.51	18.04	0.66	-40.22	4.10	0.62	
01	36.42	37.51	1.03	30.27	35.46	1.17	58.48	-6.15	1.14	

Les résultats obtenus pour les critères de validations ont été amélioré dans ce cas, surtout pour le critère de $\{B\}$.

Cette approche permet de prévoir 56% des mois avec un écart de 30 mm par rapport aux valeurs observées (figure V.14).

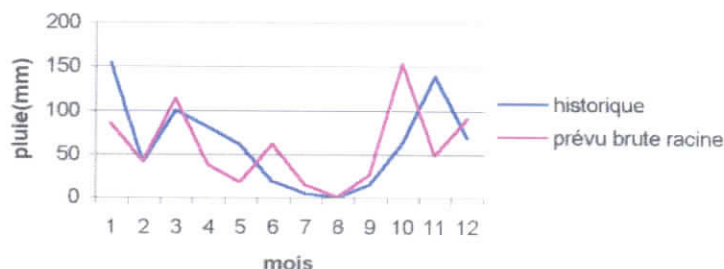


Fig V.13 la série prévue de l'année 1992

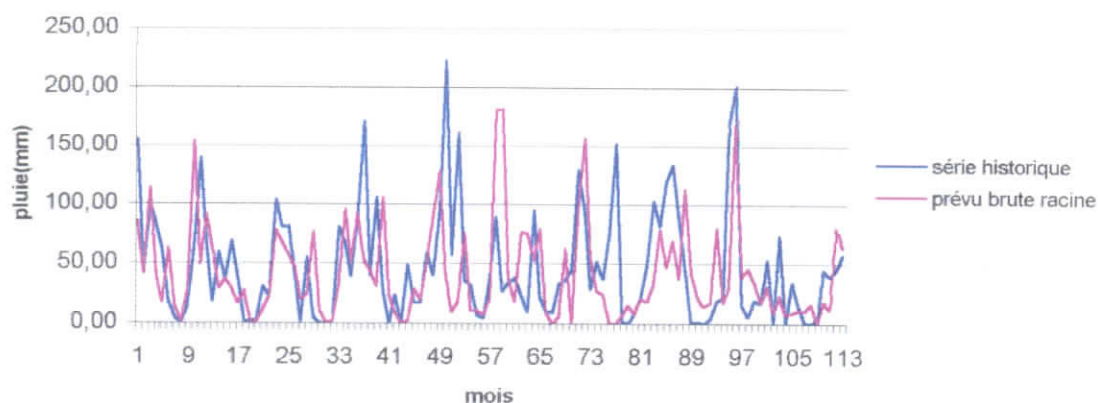


Fig: V.14 prévision des pluies mensuelles

d) ACP normée :

L'évolution des coefficients de régression est donnée par l'annexe V.7

Les tracés des CP et des résidus, sont données dans les annexes V.8 et V.9

Résultats :

La prévision de l'année 1992 est représentée par la figure V.15, les autres séries prévues sont données dans l'annexe V.6.

Les résultats obtenus, en appliquant les tests de validations sont résumés dans le tableau V.7.

Tableau V.7 : résultats de la modélisation pour les variables racines normées.

Année	Variable observée			Variable prévue					
	Moy	σ	Cv	moy	σ	Cv	C.nash (%)	{B}	I
92	62.41	50.59	0.81	59.84	45.38	0.76	32.53	-2.57	0.94
93	38.57	33.09	0.86	54.63	43.5	0.8	-24.08	16.06	0.93
94	38.07	35.91	0.94	47.54	35.97	0.76	24.77	9.47	0.81
95	45.98	48.71	1.06	51.54	40.1	0.78	-8.11	5.56	0.74
96	66.82	65.74	0.98	54.68	42.8	0.78	30.07	-12.14	0.80
97	45.55	39.37	0.86	50.62	36.21	0.72	21.54	5.07	0.84
98	50.77	45.59	0.9	60.13	41.39	0.69	2.12	9.36	0.77
99	67.11	72.76	1.08	45.79	37.52	0.82	58.54	-21.32	0.76
00	23.41	24.75	1.06	51.83	35.33	0.69	-258.43	28.42	0.65
01	36.42	37.51	1.03	37.47	43.17	1.15	55.75	1.05	1.12

L'utilisation de la variable racine normée a apporté des résultats plus ou moins meilleurs, et la tendance du phénomène a été généralement reproduit (figure V.16).

En outre, les mois prévus avec un écart de 30 mm par rapport aux valeurs observées sont de 62%.

Pour enlever le caractère aléatoire au modèle développé, on propose d'identifier les résidus d'une autre manière.

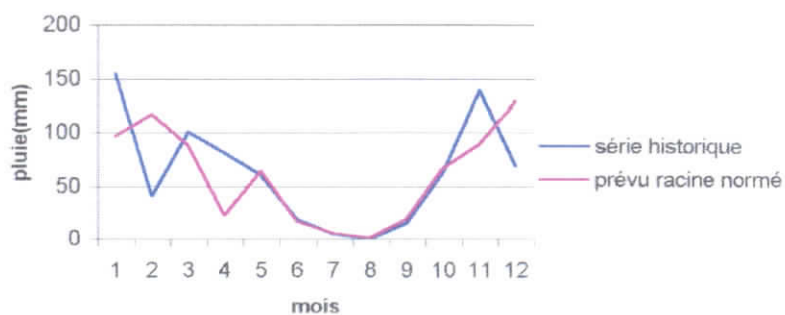


Fig: V. 15 les séries prévues de l'année 1992

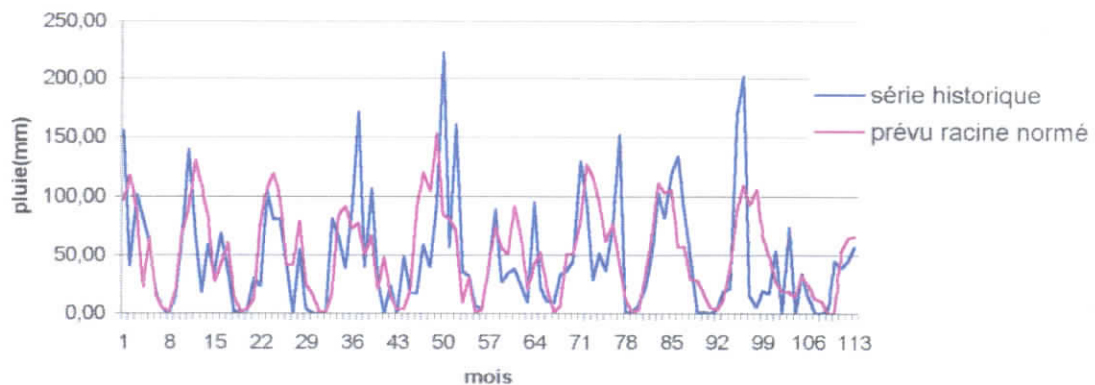


Fig:V.16 prévision des pluies mensuelles (variable racine)

V.4.2 2^{ème} approche :

Le principe est de faire une répartition saisonnière des résidus, en appliquant le modèle multiclasse d'ordre 1.

La classe du mois prévue sera celle correspondante à la plus forte probabilité. En tenant compte de la classe précédente, on tire la valeur du résidu de la série historique.

Pour tenir compte du maximum de cas possibles qui peuvent apparaître, il est important d'effectuer une répartition saisonnière convenable et de choisir un nombre de classe suffisant, de manière à garder les caractéristiques de chaque saison.

Pour les variables brutes, on choisit la répartition suivante :

Saison 1 : décembre – février

$]-\infty \ 25[, [25 \ 70[, [70 \ 100[, [100 \ \infty[.$

Saison 2 : mars - mai

$]-\infty \ 25[, [25 \ 70[, [70 \ 100[, [100 \ \infty[.$

Saison 3 : juin – août

$]-\infty \ 0[, [0 \ 25[, [25 \ \infty[.$

Saison 4 : septembre – novembre

$]-\infty \ 50[, [50 \ 100[, [100 \ \infty[.$

Et pour le normé :

Saison 1 : septembre – mai

$]-\infty \ -20[, [-20 \ 25[, [25 \ 50[, [50 \ \infty[.$

Saison 2 : juin – août

$]-\infty \ -5[, [-5 \ 0[, [0 \ 10[, [10 \ \infty[.$

Exemple :

Pour prévoir la pluie du mois de janvier 1992 (qui appartient à la saison 1), la matrice de passage pour le modèle normé est :

	0	1	2	3
0	0.374	0.340	0.111	0.173
1	0.376	0.402	0.097	0.123
2	0.327	0.448	0.034	0.189
3	0.493	0.246	0.164	0.095

La classe du résidu précédent (mois de décembre) est la classe 0, le résidu prévu appartient donc à la même classe P_{00} (correspondant à la plus forte probabilité) et à partir de la série historique, on lui attribue la valeur correspondante

Le calcul des pluies mensuelles se fait de la même manière pour toutes les années.

Résultats :

L'annexe V.10 représente les séries prévues par rapport à celles observées pour chaque année

Les résultats de la prévision mensuelle par le principe de la plus forte probabilité d'apparition dans les classes de Markov (tableaux V.8 et V.9) montrent une amélioration significative du modèle de prévision développé.

On remarque qu'en utilisant les variables brutes, les quantités de pluies prévues sont toujours sous-estimées par rapport à celles de l'historique (figure 17)

Les résultats se sont améliorés, pour le modèle normé et l'analyse graphique montre nettement une meilleure prévision des pluies mensuelles ou la reproduction du phénomène est très bien appréciée (figure 18).

Le modèle Markovien multiclasse permet de prévoir 71% des mois (pour les variables brutes ou normées) avec un écart de 30 mm par rapport aux valeurs observées.

Tableau V.8 : résultats de la modélisation pour les variables brutes.

Année	Variable observée			Variable brute						
	Moy	σ	Cv	moy	σ	Cv	C.nash (%)	{B}	I	
92	62.41	50.59	0.81	34.78	26.89	0.77	13.86	-27.63	0.95	
93	38.57	33.09	0.86	28.26	24.01	0.85	5.24	-10.31	0.99	
94	38.07	35.91	0.94	25.89	19.24	0.74	25.75	-12.18	0.79	
95	45.98	48.71	1.06	37.37	24.85	0.67	4.61	-8.61	0.63	
96	66.82	65.74	0.98	42.21	35.14	0.83	54.09	-24.61	0.85	
97	45.55	39.37	0.86	33.04	23.82	0.72	34.12	-12.51	0.84	
98	50.77	45.59	0.90	32.64	22.21	0.68	23.77	-18.13	0.76	
99	67.11	72.76	1.08	41.05	33.22	0.81	52.57	-26.06	0.75	
00	23.41	24.75	1.06	34.7	35.84	1.03	-65.43	11.29	0.97	
01	36.42	37.51	1.03	34.59	35.77	1.03	84.51	-1.83	1.00	

Tableau V.9 : résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable normée						
	Moy	σ	Cv	moy	σ	Cv	C.nash (%)	{B}	I	
92	62.41	50.59	0.81	45.88	34.19	0.75	52.15	-16.53	0.93	
93	38.57	33.09	0.86	41.02	32.74	0.80	58.97	2.45	0.93	
94	38.07	35.91	0.94	39.48	27.28	0.69	48.45	1.41	0.73	
95	45.98	48.71	1.06	39.25	30.87	0.79	52.77	-6.73	0.75	
96	66.82	65.74	0.98	44.2	32.02	0.72	33.37	-22.62	0.73	
97	45.55	39.37	0.86	35.72	24.35	0.68	11.59	-9.83	0.79	
98	50.77	45.59	0.90	47.78	33.24	0.70	45.64	-2.99	0.78	
99	67.11	72.76	1.08	45.36	40.70	0.90	49.8	-21.75	0.83	
00	23.41	24.75	1.06	34.81	26.53	0.76	4.60	11.4	0.72	
01	36.42	37.51	1.03	34.10	26.32	0.77	51.82	-2.32	0.75	

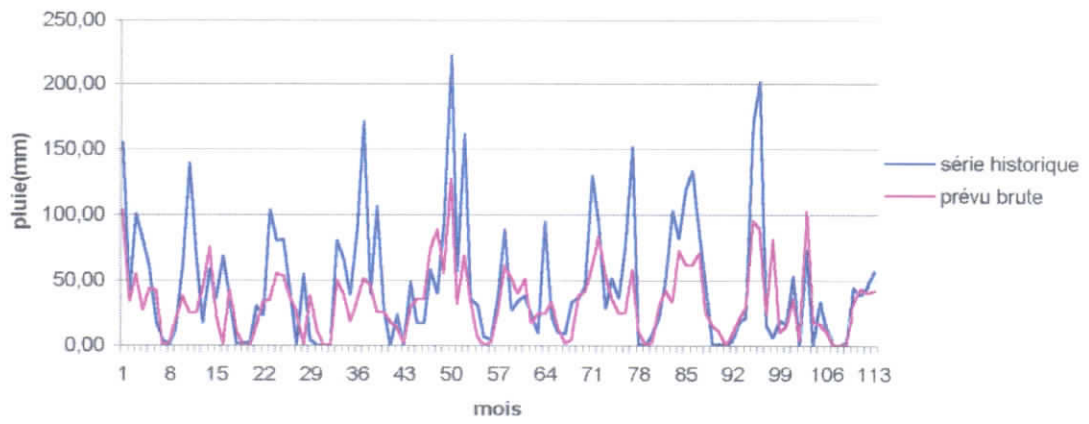


Fig:V.17 prévision des pluies mensuelles

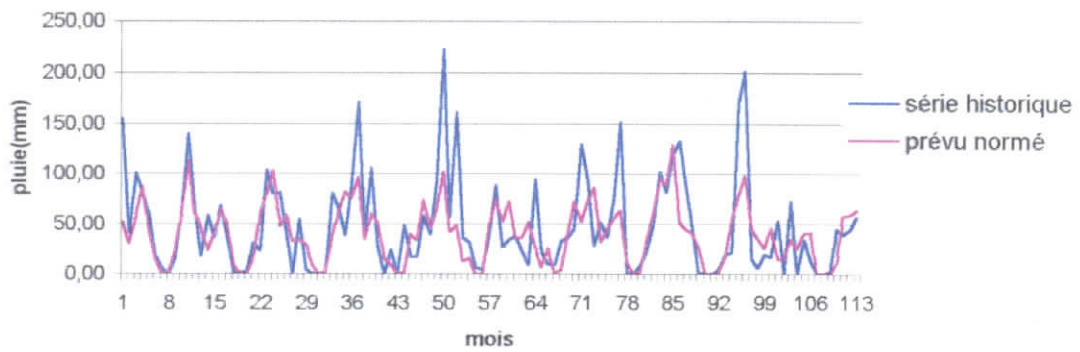


Fig:V.18 prévision des pluies mensuelles

V.5 Conclusion :

Un modèle de prévision des pluies mensuelles a été élaboré, celui-ci est basé sur la combinaison de l'analyse en composantes principales et les chaînes de Markov.

Partant de la structure du modèle de base de l'ACP, on remarque qu'il y a trois paramètres à estimer dans le modèle brute et quatre dans le modèle normé.

Pour l'estimation des résidus, deux approches ont été élaborées, la première consiste à effectuer des tirages de nombres aléatoires mais les résultats obtenus n'ont pas été satisfaisants.

L'introduction de la racine carrée a permis d'améliorer les résultats du modèle de prévision, mais afin d'affiner le modèle développé, on a procédé dans la deuxième approche à l'estimation des résidus selon la plus forte probabilité d'apparition dans les classes du modèle Markovien saisonnier, ce qui a permis d'améliorer nettement le modèle et d'approcher beaucoup plus le phénomène physique étudié.

Néanmoins, l'inconvénient de cette approche consiste à ne pas reproduire les valeurs extrêmes, car en attribuant à la valeur prévue la classe correspondant à la plus forte probabilité, on donne une importance aux cas fréquents et on occulte les classes de faibles probabilités.

En outre, on remarque que la variable normée donne des résultats meilleurs par rapport au modèle brute, cette amélioration peut être expliquée par l'intégration du paramètre stable β_n .

En conclusion, on peut dire que les résultats obtenus sont satisfaisants du fait que la tendance du phénomène étudié est bien reproduite, et que les résultats des tests de validation sont globalement vérifiés.

Conclusion générale

L'objectif principal de notre étude a été l'analyse de la pluviométrie à la station d'Alger et le développement d'un modèle de prévision de ce paramètre, dans le but d'apporter une contribution aux problèmes de gestion des ressources hydriques.

La première étape de ce travail consiste à collecter et à critiquer les données recueillies auprès de l'ONM, à savoir un fichier de pluie à différents pas de temps s'étalant de 1936 à 2001.

Le tracé de l'évolution des pluies journalières en fonction du temps a mis en évidence le caractère aléatoire du phénomène étudié.

L'analyse des caractéristiques statistiques a montré l'irrégularité du régime pluviométrique.

L'étude comparative entre les différentes décennies a permis de relever une légère tendance à la baisse du régime pluviométrique durant les deux dernières décennies avec un écart pluviométrique maximum qui ne dépasse pas les 15% par rapport à la moyenne lors de la dernière décennie. Ce déficit pluviométrique (appelé improprement sécheresse) est dû principalement à la baisse des fortes précipitations et non pas au nombre de jours pluvieux durant cette décennie.

En conclusion de la première partie, on peut dire que la situation hydrique actuellement déficitaire à Alger n'est pas le fait d'une sécheresse mais plutôt d'une succession d'années déficitaire (9 années sur 12). L'aggravation de cette situation a été accélérée par l'absence d'une politique claire en matière de la gestion des ressources hydriques.

Dans la deuxième partie de ce travail, on a développé un modèle de prévision des pluies au pas de temps mensuels, celui-ci est basé sur la combinaison de l'analyse en composantes principales et des chaînes de Markov.

En se basant sur la structure du modèle de l'analyse en composantes principales, deux approches ont été développées et comparées pour aboutir au meilleur modèle qui reproduit le plus fidèlement possible la série historique.

Le tracé de l'évolution des différents paramètres a mis en évidence le caractère aléatoire des composantes principales et des résidus et une tendance à la stabilité de la moyenne mensuelle interannuelle $[\beta_0]$ et des coefficients de régression $[\beta_{ii}]$.

Dans les deux approches, on a pris en considération la moyenne des deux parties stables pour β_0 et β_{ii} , et la valeur historique de la CP.

La première approche consiste à introduire le modèle Markovien multiclasse et à effectuer des tirages de nombres aléatoires pour l'estimation des résidus.

D'après les tests de validation, les résultats obtenus pour le modèle de prévision brute et normé ne sont pas satisfaisants avec même des valeurs de pluie négatives, ce qui est physiquement impossible.

Le changement des variables brutes et normées a permis d'améliorer les résultats du modèle de prévision surtout dans le cas normé.

Afin d'affiner le modèle développé, on a opté dans la deuxième approche à l'élimination de la procédure de tirage de nombres aléatoires par l'introduction du modèle Markovien multiclasse saisonnier, qui permet de ressortir les caractéristiques de chaque saison et de s'approcher le plus possible du phénomène physique étudié.

Les résultats obtenus dans ce cas se sont nettement améliorés par rapport à la première approche, néanmoins le modèle de prévision développé ne permet pas de reproduire les valeurs extrêmes (les max et les min) de la variable étudiée.

En conclusion, on peut dire que l'utilisation de l'analyse en composante principale qui est une technique simple dans la prévision d'une variable aléatoire, a donné lieu à des résultats satisfaisants peuvent être améliorés en affinant mieux l'estimation des paramètres ou par l'introduction d'autres combinaisons possibles avec cette technique.

Référence bibliographique

- [1] Jean.A . Vergnes: Conférence , journée mondiale de l'eau, 2002.
- [2] Fatna Barakat, Abderrahim Handoufe: approche agroclimatique de la sécheresse agricole au Maroc, cahiers sécheresse, volume 9, numéro 3, Septembre1998 .
- [3] Mohamed Kadi: quelques aspects de la sécheresse en Afrique du Nord et dans le bassin Méditerranéen, article paru dans la revue techniques et sciences, Septembre1992.
- [4] Annick Douguedroit et Caroline Norrant: tendance quasi-seculaires des précipitations et des pressions de surface dans le bassin méditerranéen (PRESANOR-01) , 2002 .
- [5] Mohamed Kadi : sécheresse ,climat et environnement , article paru dans la revue techniques et sciences, Décembre1990.
- [6] Hirche Azziz . Addi Nacer et Boughani Abdelmadjid: contribution a l'étude de l'évolution de la pluviosité dans quelques stations steppiques Algériennes (PRESANOR-01). 2002 .
- [7] A. Matari, A. Douguedroit, Kerouche. M , Bouachria .M ., Bousid. H: la sécheresse en Algerie et ses causes possibles (PRESANOR-01), 2002 .
- [8] Mohamed Kadi: indicateurs météorologiques de la sécheresse en Méditerranée et prévision saisonnière [CONFID], article paru dans la revue techniques et sciences, Mars 1995.
- [9] Forum international sur la variabilité climatique et la prévision saisonnière en Afrique du Nord et en Méditerranée (PRESANOR-01) 2002.
- [10] F.Youcef Ettoumi: analyse statistique des précipitations pour le nord de l'Algérie (PRESANOR-01), 2002
- [11] R.Benzaghrou : quantification de l'influence des oscillations méditerranéennes sur les indices pluviométrique du nord de l'Algérie (PRESANOR-01) , 2002 .
- [12] Dj.Bouanoune , Dahmani Megherouche: application des réseaux neuronaux a la reconstitution des données pluviométrique(PRESANOR-01) ,2002 .
- [13] Lazreg Benaichata: régionalisation des précipitations du l'Algérie du nord (PRESANOR-01), 2002 .
- [14] Liams .j: hydrologie générale : principes et applications. Gaetan Morin, Québec, pp.38-60, 1985.
- [15] Bobee, B: The gamma family and drived distribution applied in hydrology. Water ressource application, colorado, U.S.A, pp 50-69,1999 .

[16] Chambaz, H: contribution a l'étude de la méthode DPFT(différence première de la fonction transfert).Thèse de magistère ENP Alger,1990 .

[17] N. Dechemi , A. Bermad, T. Merabtene, développement théorique et étude des variables aléatoires et cyclique par l'analyse en composantes principales, rapport de recherche.

[18] Marie Piron, statistique exploratoire multidimensionnelle ; 2 édition, 1998.

[19] Hamriche, A: contribution à la simulation des paramètres hydrométéorologiques aléatoires et cycliques à différents pas de temps mensuelle, journalier et horaire par le modèle GESTOP. Thèse de magistère ENP Alger,1998 .

[20] Bermad: élaboration d'un modèle de gestion d'un barrage réservoir. Thèse de magistère ENP Alger , 1990 .

[21] A.Matari , A.Douguedroit : les stratégies pour le contrôle des sécheresses, chronologie des précipitations et des sécheresses dans l'ouest Algérien, 9^e conférence régionale afro-asiatique des irrigations et du drainage,1995

[22] Pierre Rognon, sécheresse et aridité : leur impact sur la désertification au Maghreb, cahier sécherese1996 ;7 :287-1997.

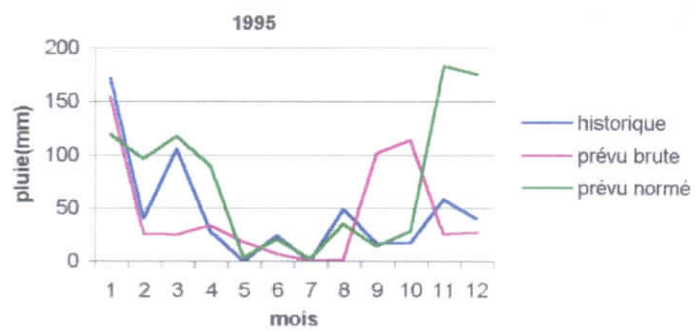
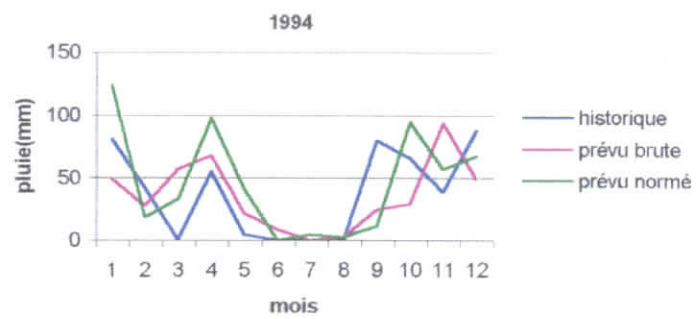
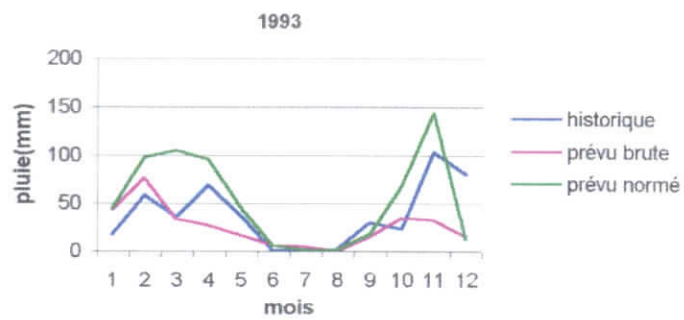
[23] R. Krzysztofowicz, J.S.G.McCulloch, G.Vachaud, P.Van Cappellen, Automatic calibration of conceptual rainfall- runoff models : sensitivity to calibration data, journal of hydrology , volum.181, numéro.1-4, juin 1996.

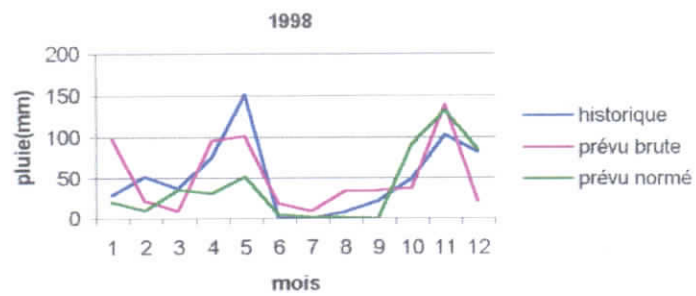
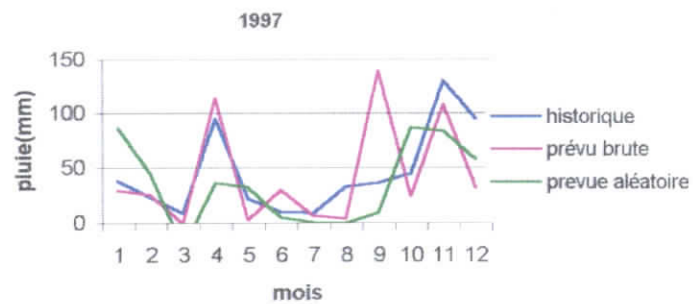
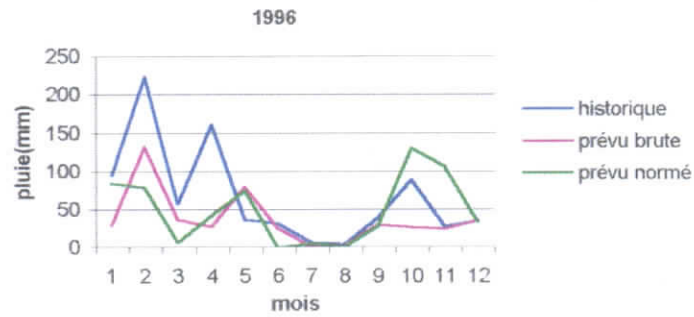
AmiExe

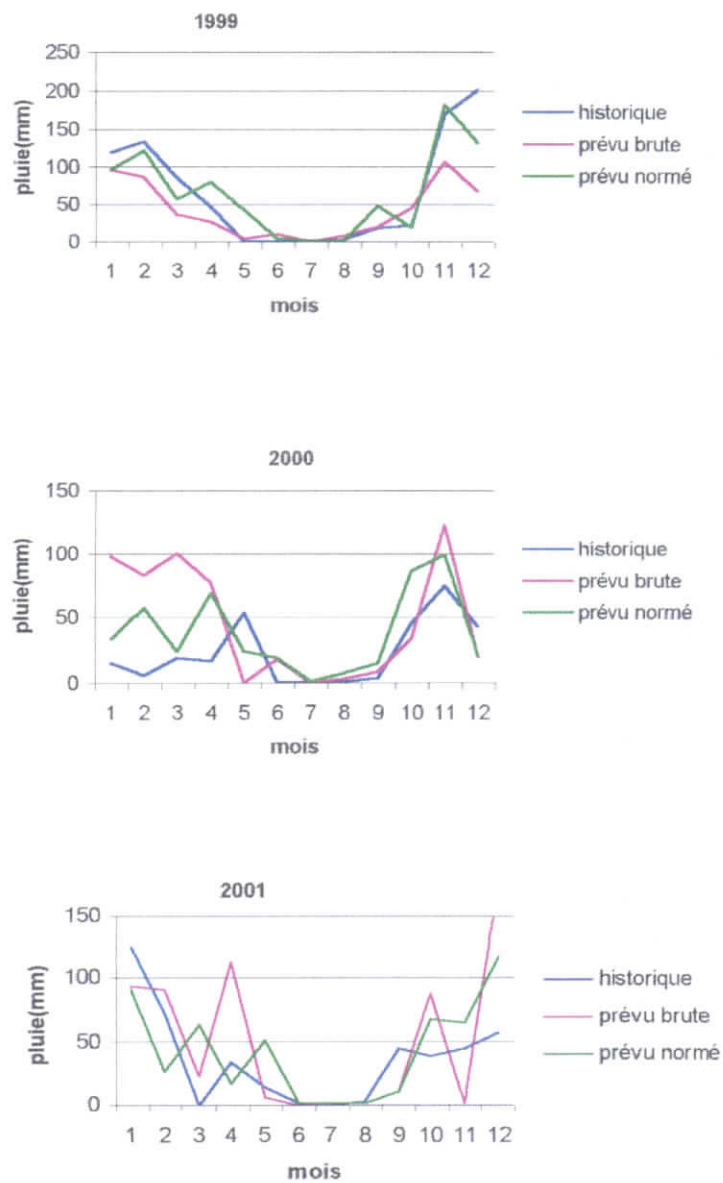
**Annexe IV.1 : les fréquences expérimentales des pluies
journalières maximales.**

La pluie maximale(mm)	Fréquence expérimentale	période de retour
136.6	0.02	66.00
125.5	0.03	33.00
124.3	0.05	22.00
115.2	0.06	16.50
108.0	0.08	13.20
104.4	0.09	11.00
93.3	0.11	9.43
92.5	0.12	8.25
91.6	0.14	7.33
86.7	0.15	6.60
86.2	0.17	6.00
85.9	0.18	5.50
83.0	0.20	5.08
79.1	0.21	4.71
73.6	0.23	4.40
71.9	0.24	4.13
71.3	0.26	3.88
68.5	0.27	3.67
66.2	0.29	3.47
64.9	0.30	3.30
61.0	0.32	3.14
60.5	0.33	3.00
60.3	0.35	2.87
59.3	0.36	2.75
57.1	0.38	2.64
55.0	0.39	2.54
54.6	0.41	2.44
52.9	0.42	2.36
50.3	0.44	2.28
48.7	0.45	2.20
48.7	0.47	2.13
48.5	0.48	2.06
48.5	0.50	2.00
46.9	0.52	1.94
46.0	0.53	1.89
45.2	0.55	1.83
45.1	0.56	1.78
44.9	0.58	1.74
44.3	0.59	1.69
44.3	0.61	1.65
44.0	0.62	1.61
43.3	0.64	1.57
43.1	0.65	1.53
42.4	0.67	1.50
42.0	0.68	1.47
41.8	0.70	1.43
41.5	0.71	1.40
40.7	0.73	1.38

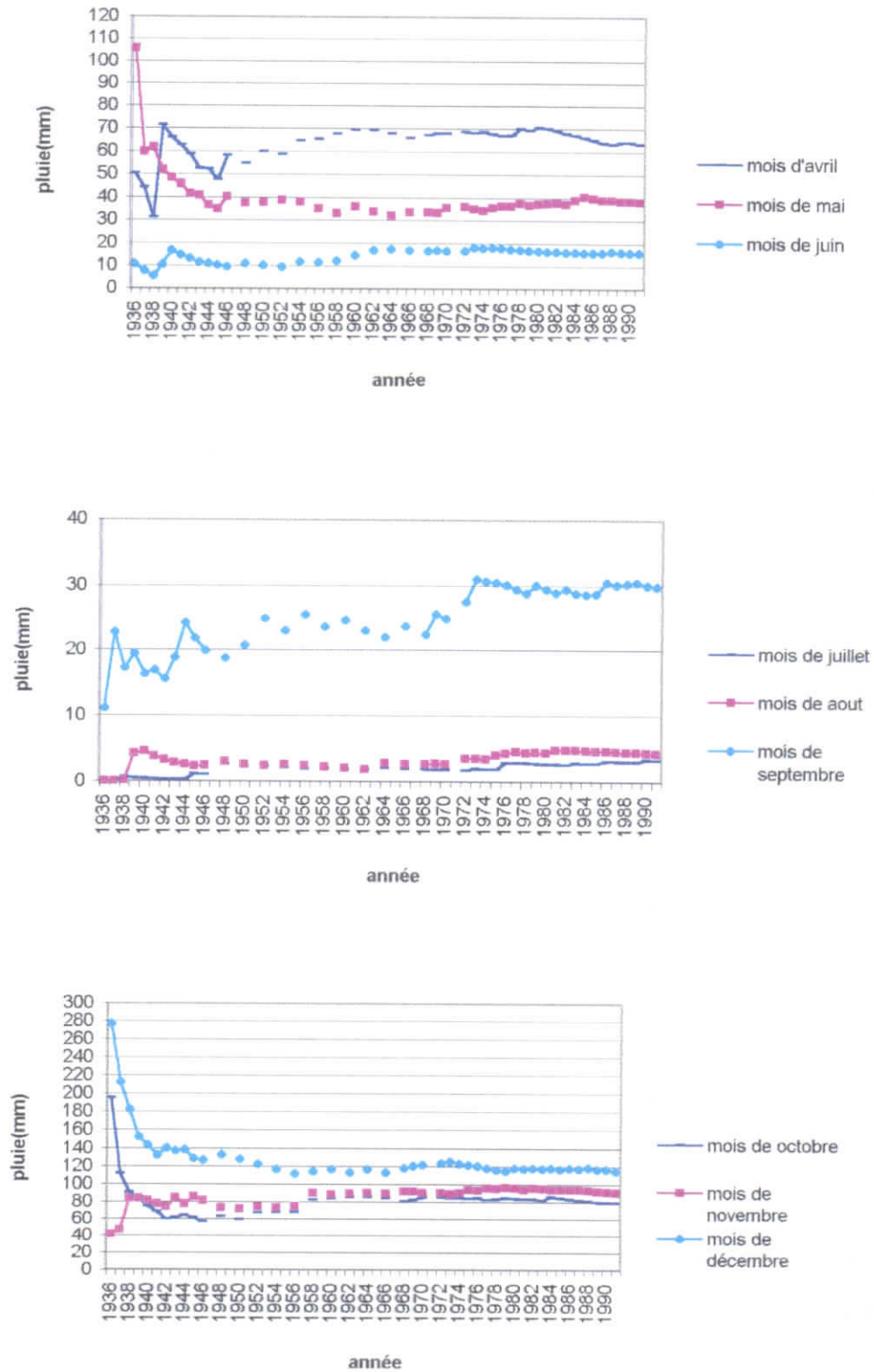
39.9	0.74	1.35
39.7	0.76	1.32
39.6	0.77	1.29
39.3	0.79	1.27
37.3	0.80	1.25
37.0	0.82	1.22
36.2	0.83	1.20
35.6	0.85	1.18
34.5	0.86	1.16
33.4	0.88	1.14
32.5	0.89	1.12
31.5	0.91	1.10
31.1	0.92	1.08
30.4	0.94	1.06
30.4	0.95	1.05
29.1	0.97	1.03
21.1	0.98	1.02



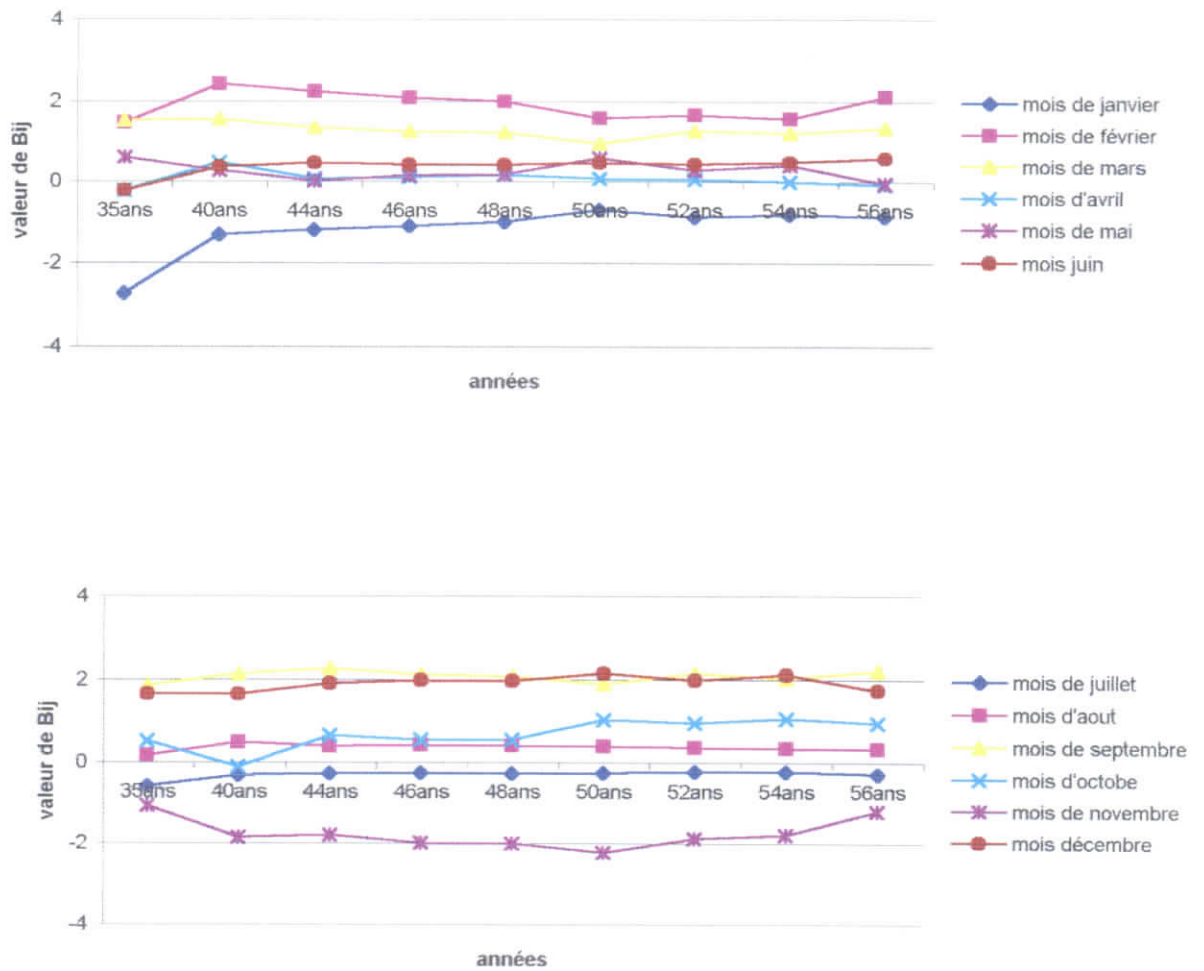




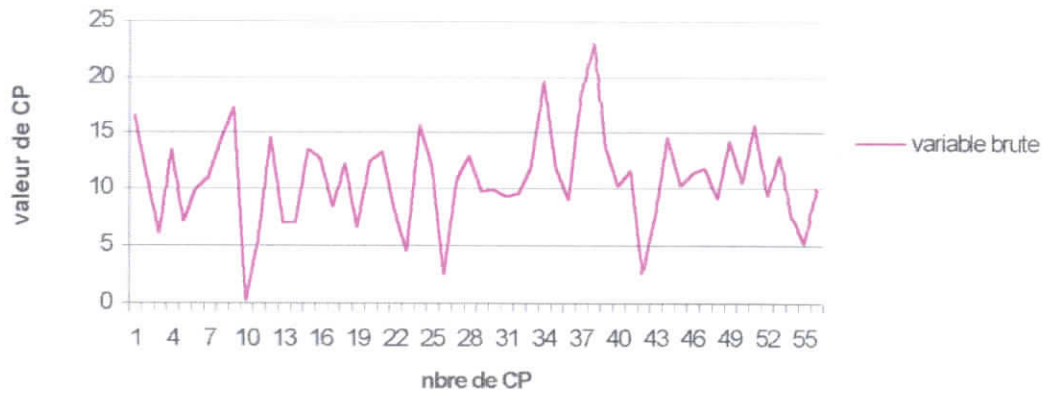
Annexe V.1 : les séries mensuelles prévues et observées des pluies pour les variables brutes et normées



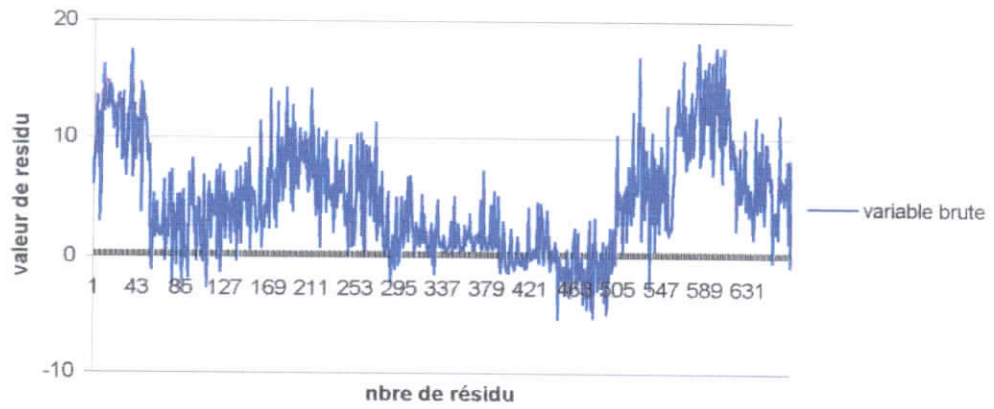
Annexe V.2 Evolution de la moyenne mensuelle interannuelle β_0 .



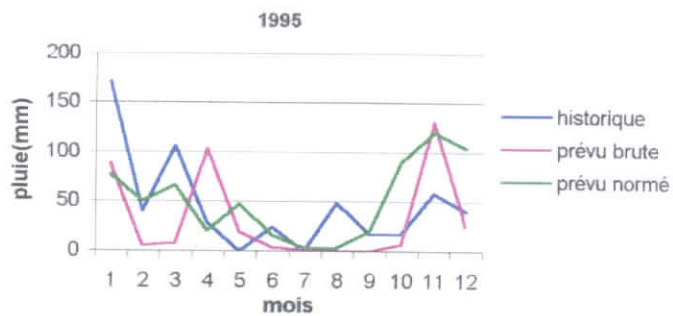
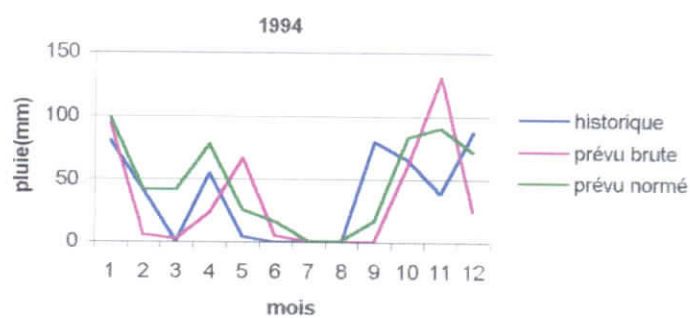
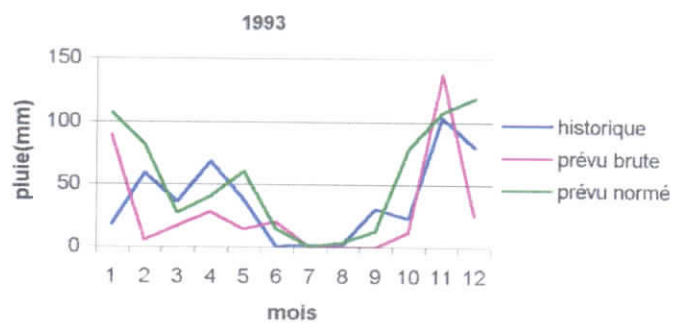
Annexe V.3 : l'évolution des β_{ij} pour chaque mois (variable racine brute).

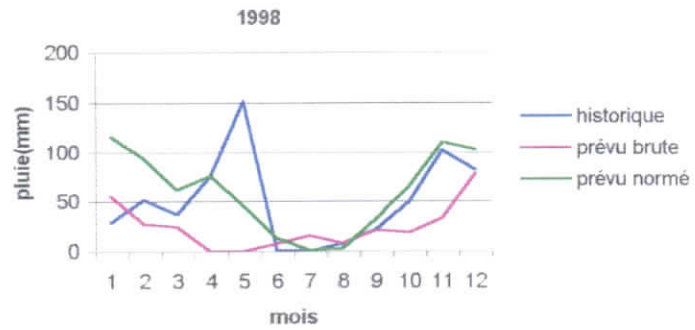
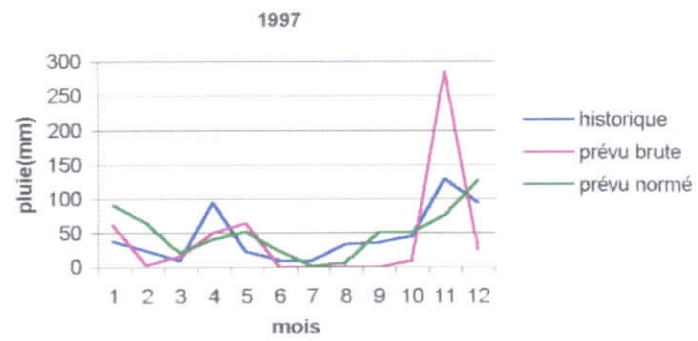
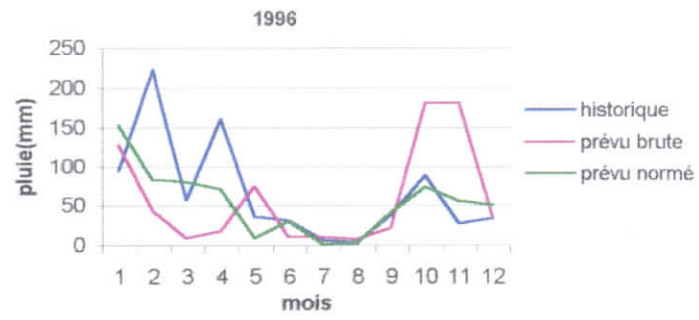


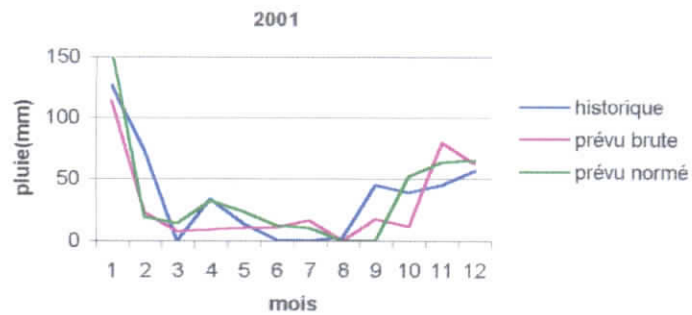
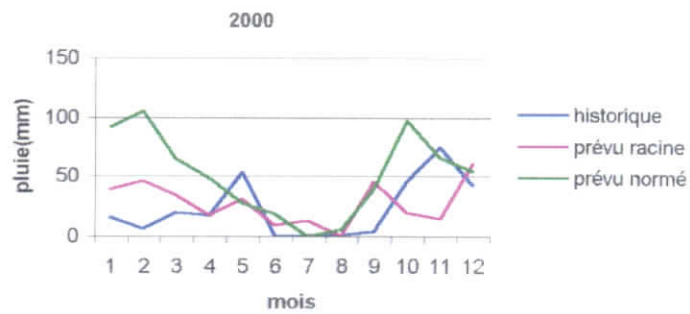
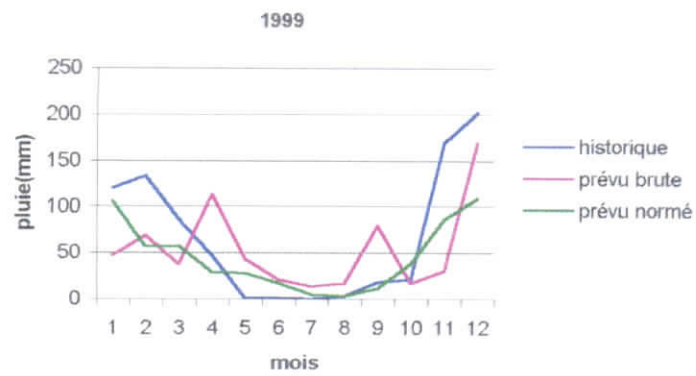
**Annexe V.4 Evolution de la 1^{ère} composante principale
(variable racine brute)**



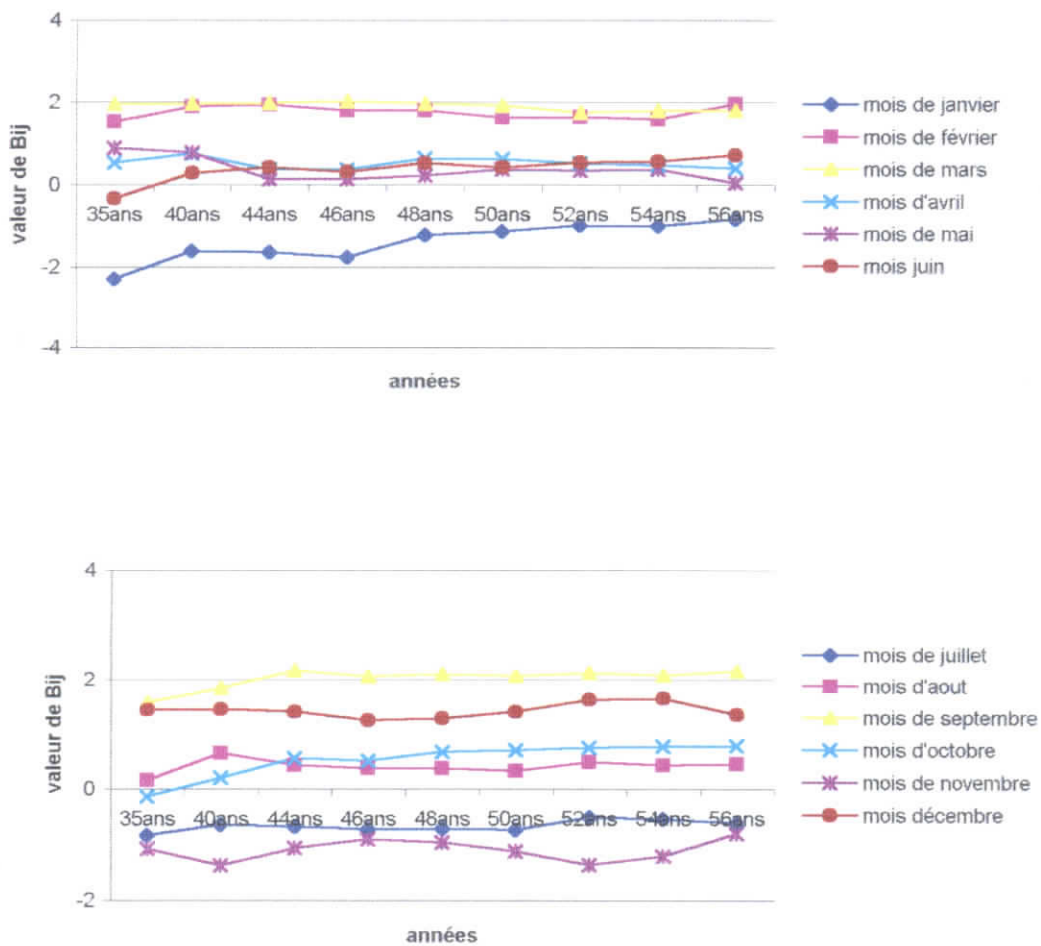
**Annexe V.5 Evolution des variables résiduelles
(variable racine brute)**



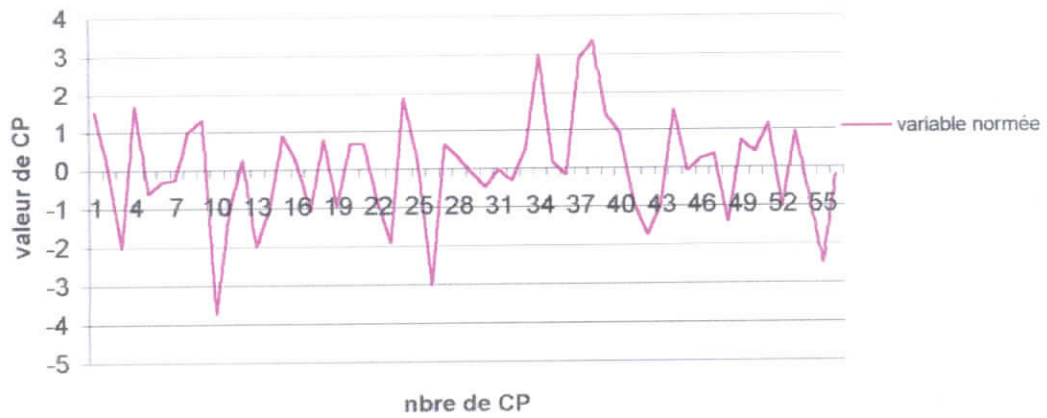




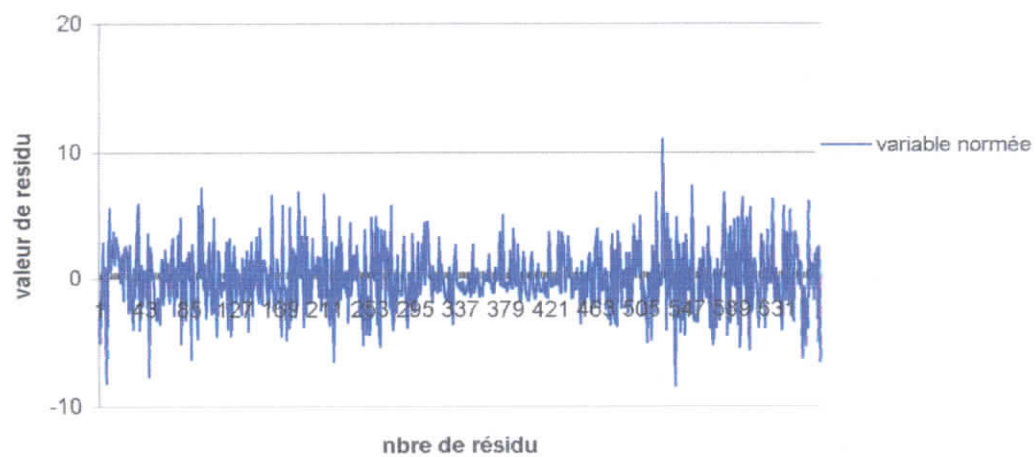
**Annexe V.6 : Les séries mensuelles prévues et observées des pluies
(variables racines brutes et normées)**



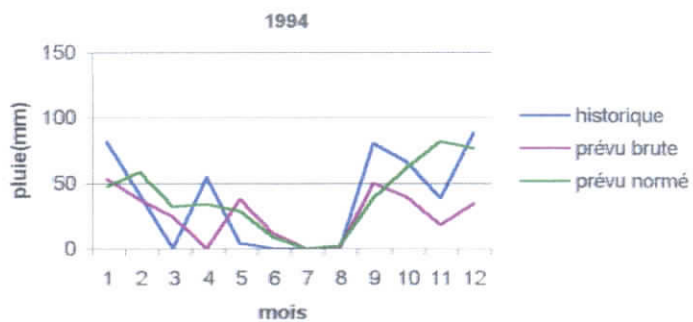
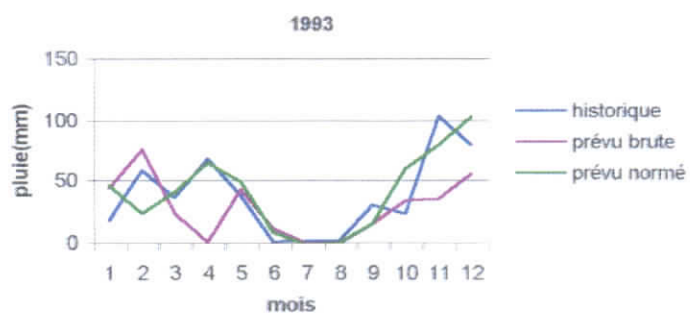
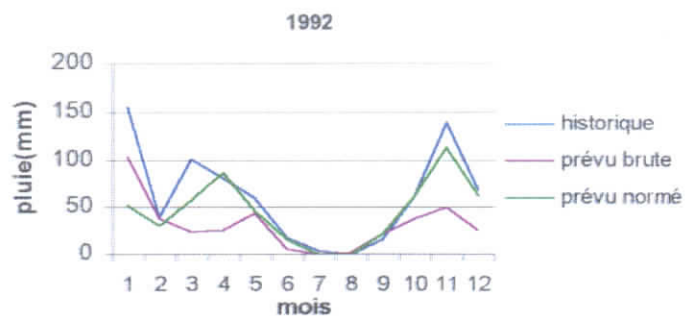
**Annexe V.7 : Evolution des coefficients de régression
(variable racine normée)**

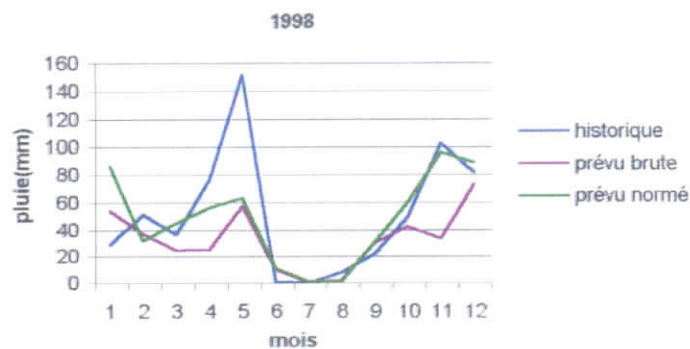
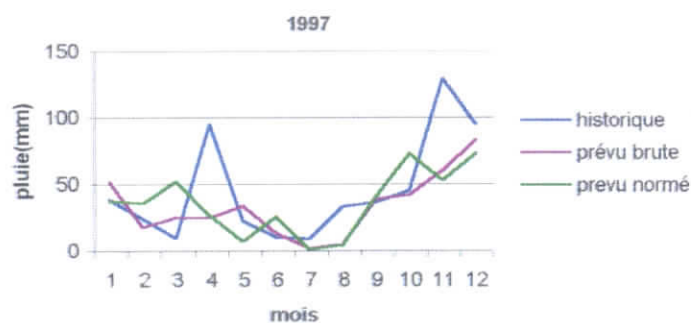
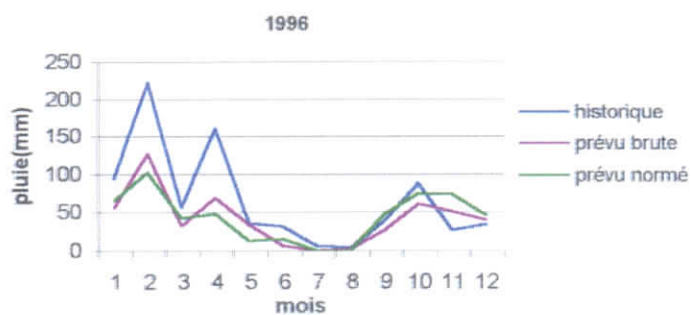
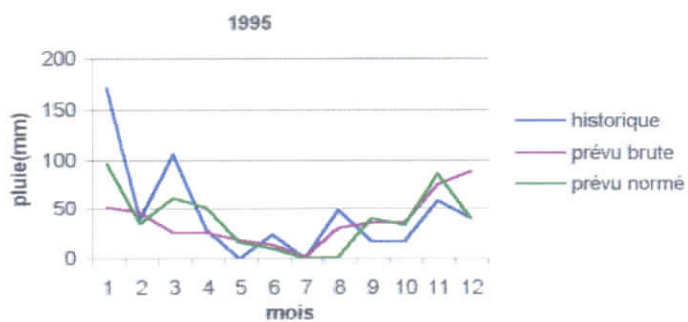


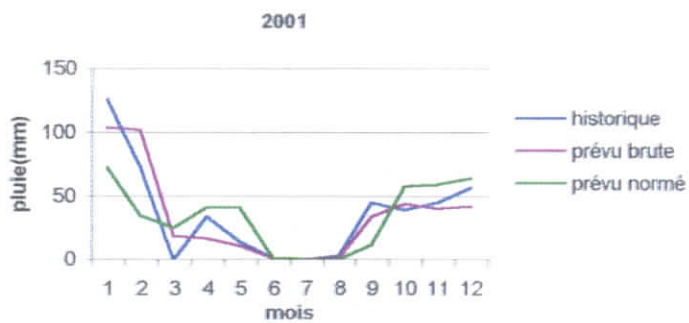
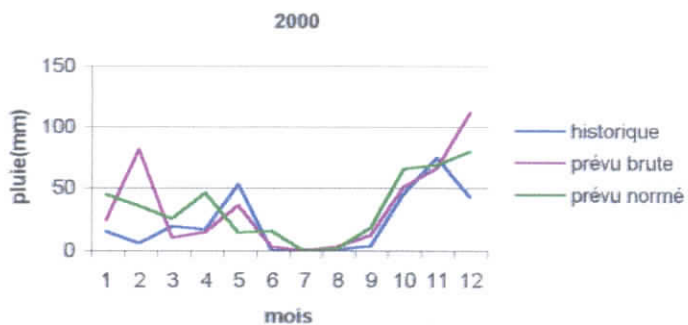
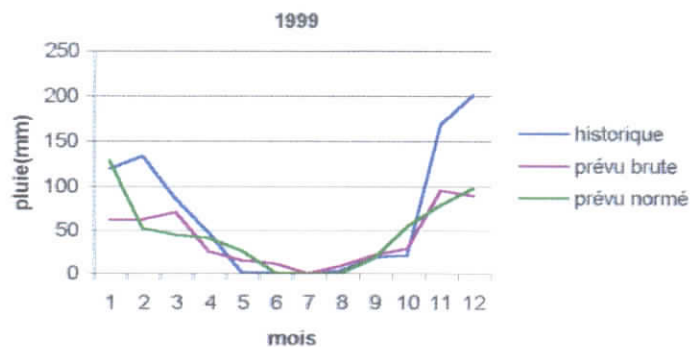
**Annexe V.8 Evolution de la 1^{ère} composante principale
(variable racine normée)**



**Annexe V.9 Evolution des variables résiduelles
(variable racine normée)**







Annexe V.10 Les séries prévues et observées pour chaque année