

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
 MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE
 LA RECHERCHE SCIENTIFIQUE

Ecole Nationale Polytechnique
 Laboratoire Construction et Environnement



المدرسة الوطنية المتعددة التقنيات
 BIBLIOTHEQUE — المكتبة
 Ecole Nationale Polytechnique

المدرسة الوطنية المتعددة التقنيات
 Ecole Nationale Polytechnique

THESE

Présentée par

Abdelmalek BERMAD

Pour l'obtention du grade de

DOCTEUR D'ETAT

HYDRAULIQUE

**ELABORATION D'UNE METHODOLOGIE
 DE SIMULATION ET DE PREVISION DES
 VARIABLES CLIMATOLOGIQUES
 ALEATOIRES ET CYCLIQUES PAR
 L'ANALYSE EN COMPOSANTES
 PRINCIPALES**

Devant le jury :

Président :	A. AIDAOUÏ	Professeur	INA Alger
Rapporteur :	N. DECHEMI	Professeur	ENP Alger
Examinatrice :	B. TOUAÏBIA	M. de Conférence	ENSH Blida
Examineur :	M. MEDDÏ	M. de Conférence	C. U. K. MILIANA
Examineur :	M. LADJEL	M. de Conférence	U. F. ABBAS Sétif
Invité :	A. OUABDESSELAM	Directeur ENP	ENP Alger
Invité :	M. KADI	Dir. de Recherche	ONM Alger

Juin 2004

DEDICACES



Je dédie ce travail :

A ma mère

A ma mère

A ma mère

A mon père à qui je souhaite un prompt rétablissement

A mes frères et sœurs

A mes nièces et neveux

A mes enseignants

A tous mes collègues enseignants et travailleurs de l'ENP

A mes amis

A mes étudiants.

Abdelmalek BERMAD

REMERCIEMENTS

Je tiens d'abord à exprimer ma reconnaissance à mes parents pour avoir autant contribué à ce que je suis, et qui ont aussi accepté de me voir si longtemps sur les « bancs d'école » !

Ce travail a été réalisé au sein de l'équipe de recherche GESTOP du Laboratoire Construction et Environnement de l'Ecole Nationale Polytechnique d'Alger.

*Il a été dirigé avec tact et précision par le professeur **Noureddine DECHEMI**, qui m'a accueilli, formé, encouragé, accompagné, critiqué, supporté et finalement invité à manger chez lui. Je te remercie pour tout cela Noureddine, ainsi que pour la somme de choses que j'ai appris depuis que je suis avec Toi à GESTOP. Tu es le principal responsable du plaisir que j'ai eu à « chercher » et à mener à bien cette recherche.*

*J'adresse mes sincères remerciements à Monsieur **Abdellah AIDAOUI** Professeur à l'Institut National d'Agronomie (INA) qui m'a fait l'honneur de présider le jury de cette thèse.*

*Je remercie Mme **B. TOUAIBIA** et Mr **M. CHERRARED** qui ont accepté de juger ce travail, ainsi que MM. **M. MEDDI** et **M. LADJEL** pour avoir bien voulu participer au jury de thèse.*

*Je dois dire que je me sens très honoré de la présence du Directeur de l'ENP **Mr A. OUABDESSELAM** et du Directeur de l'ONM Mr **M. KADI** et je suis très fière de l'intérêt qu'il ont porté à mon travail.*

*Je garderais un très agréable souvenir de mon Directeur de « labo », le Pr **A. BALI** qui m'a toujours soutenu dans les moments difficiles qu'il trouve ici l'expression de ma profonde gratitude.*

Je remercie l'ensemble du personnel de la bibliothèque et du centre de calcul.

J'aimerais remercier enfin tous ceux qui m'ont soutenu durant les moments difficiles.

ملخص

إن المحاكاة والتنبؤ للمتغيرات المناخية تستعمل غالبا لإيجاد طريقة عقلانية لتسيير الموارد المائية . النماذج المستعملة حاليا لها تطبيقات خاصة ولا يمكن تعميمها. فتنوع المتغيرة ،خطوة الزمن، ثبوتهما، حجمها وتزامنها هي أهم مستلزمات تطبيق هذه النماذج . لدا تم مقارنة نموذج فريد من نوعه يركز على تقنية التحليل بالمركبات الرئيسية يسمح بإزالة كل أنواع القيود المفروضة على النماذج السابقة كما يسمح بمحاكاة كل المتغيرات المناخية الصدفوية والدورية على اختلاف الخطوات الزمنية المتغيرة بدون قيد أو شرط . كما يسمح النموذج التنبؤ بالحصول على نتائج حد حسنة و خاصة بالنسبة للمتغيرة الدورية . و قوة نتائجها . تتميز هذه الطريقة المتكثرة بسهولة استعمالها ، مرونتها .
كلمات الرئيسية : المحاكاة - التنبؤ - المتغيرات المناخية - الصدفوية - الدورية - التحليل بالمركبات الرئيسية .

RESUME

La simulation et la prévision des variables climatologiques sont utilisées pour l'optimisation de la gestion des ressources en eau. Les modèles actuels sont utilisés chacun dans des conditions spécifiques. Ils ont des exigences telles, le pas de temps, le type de variable, la normalité, la stationnarité de l'échantillon, ainsi que la taille et la chronologie de la série historique.

Un modèle original basé sur l'analyse en composantes principales a été élaboré. Il a permis de lever toutes les contraintes précitées, en effet par le biais de ce modèle nous arrivons à simuler et à prévoir les variables climatologiques aléatoires et cycliques, à différents pas de temps et sans aucune contrainte.

Le logiciel élaboré dans le cadre de cette recherche permet à partir d'un fichier historique quels que soient ses caractéristiques de ressortir les options adéquates de simulation parmi un très grand nombre de possibilités. Les résultats du modèle de prévision élaboré sont plus que satisfaisant surtout pour une variable cyclique.

La méthodologie élaborée est caractérisée par sa simplicité, son adaptabilité, son universalité, et sa robustesse.

Mots clés : Simulation – prévision – variables climatologiques – aléatoires – cycliques – ACP.

ABSTRACT

The simulation and the prediction of climatological variables are used for the optimisation of water resource management. The existing models are used in particular conditions. Time step, variable type, normality, stationnarity, size and the chronology of historical data together are especially needed.

An original model based on principal components analysis has been elaborated. It allowed to overtake all the problems. In fact, by using this model, we can simulated and predicted the random and cyclical climatological variables at any time step without any constraints.

The programm elaborated in this study allows to show the necessary simulation options among a large number of possibilities, using an historical data whatever its characteristics. The results of the elaborated model of the prediction are satisfactory especially for a cyclic variable.

The elaborated methodology is very simple, adaptable, universal and very strong.

Key words : Simulation – prediction – climatological variables – random – cyclical – PCA.

SOMMAIRE



INTRODUCTION GENERALE.....	1
----------------------------	---

Chapitre I : Revue Bibliographique

I. 1 LA SIMULATION.....	5
I-1-1 La méthode de Monté-Carlo.....	6
I-1-2 Le modèle de Thomas et Fiering.....	7
I-1-3 Les modèles FGN (Fractional Gaussian Noise)	8
I-1-4 Le modèle BL (Broken Line)	8
I-1-5 Les modèles ARMA (Auto Regressive Moving Average).....	8
I-1-6 Les modèles de désagrégation.....	10
I-1-7 Les modèles Markoviens.....	11
I-1-8 Le modèle Neyman-Scott.....	13
I-1-9 Le modèle DSH (Décomposition de la Série Historique).....	13
I-1-10 Le modèle SVD (Singular Value Decomposition)	15
I. 2 LA PREVISION.....	15
I-2-1 La prévision météorologique.....	17
I-2-1-1 Prévision par la SST (Surface Sea Temperature).....	17
I-2-1-2 Prévision par l'NOA (North Atlantic Oscillation).....	17
I-2-1-3 Prévision par l'imagerie satellitale.....	18
I-2-2 La prévision hydrologique.....	18
I-2-2-1 Prévision par la famille des modèles ARMA.....	18
I-2-2-2 Prévision par les modèles Markoviens.....	19
I-2-2-3 Prévision par les réseaux de neurones.....	19
I-2-2-4 Prévision par la logique floue.....	20
I. 3 CONCLUSION.....	20

Chapitre II : Collecte et Critique des Données

II. 1 INTRODUCTION.....	22
II. 2 INVENTAIRE DES DONNEES.....	22
II. 3 CRITIQUE DES DONNEES.....	23
II-3-1 Méthodes de comblement des lacunes.....	24
II-3-2 Détection des valeurs singulières.....	26
II. 4 TRAITEMENT STATISTIQUE.....	27
II-4-1 Test d'homogénéité.....	27
II-4-2 Test d'indépendance.....	28
II. 5 MODELE DE CALCUL DE L'EVAPOTRANSPIRATION POTENTIELLE (ETP)..	29

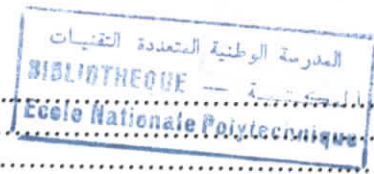
Chapitre III : Théories de l'Analyse en Composantes Principales et des Chaînes de Markov

III. 1	ANALYSE EN COMPOSANTES PRINCIPALES (ACP).....	32
	III-1-1 Introduction.....	32
	III-1-2 Définitions.....	33
	III-1-2-1 Notion d'individu et de caractère.....	33
	III-1-2-2 Définition algébrique.....	33
	III-1-2-3 Définition géométrique.....	35
	III-1-3 Formulation mathématique du problème.....	35
	III-1-4 Procédé d'application de l'ACP.....	36
	III-1-4-1 Calcul de la matrice de covariance.....	36
	III-1-4-2 Recherche des axes principaux.....	37
	III-1-4-3 Calcul des composantes principales.....	39
	III-1-5 Principaux résultats.....	39
	III-1-6 Représentation graphique.....	40
	III-1-7 ACP programmation.....	40
	III-1-7-1 Fichier de données INPUT.....	40
	III-1-7-2 Organigramme de calcul.....	41
	III-1-8 Application de l'ACP.....	43
	III-1-8-1 Influence de la transformation des variables.....	43
	III-1-8-2 Exemple d'application.....	47
III. 2	LES CHAINES DE MARKOV.....	50
	III-2-1 Introduction.....	50
	III-2-2 Ordre des chaînes de Markov.....	50
	III-2-2-1 Processus d'ordre 1.....	50
	III-2-2-2 Processus d'ordre 2.....	52
	III-2-3 Les modèles multiclassés.....	52

Chapitre IV : Elaboration du Modèle de Simulation

IV. 1	INTRODUCTION.....	55
IV. 2	MODELE DE SIMULATION.....	55
IV. 3	METHODES DE SIMULATION DES PARAMETRES DU MODELE.....	58
	IV-3-1 Simulation par les fonctions de répartition.....	58
	IV-3-2 Simulation par les lois d'ajustement.....	60
	IV-3-2-1 Simulation par la loi de Gauss ou loi Normale.....	61
	IV-3-2-2 Simulation par la loi de Gumbel.....	61
	IV-3-2-3 Simulation par la loi de Goodrich.....	62
	IV-3-2-4 Simulation par la loi de GEV (General Extrem Value).....	63
	IV-3-3 Simulation par les chaînes de Markov.....	65
	IV-3-3-1 Construction de la chaîne de Markov.....	65
	IV-3-3-2 Génération par les séries de Markov.....	66
IV. 4	TESTS DE VALIDATION.....	66

IV-4-1 Simulation au pas de temps mensuel.....	66
IV-4-2 Simulation au pas de temps décadaire.....	66
IV-4-3 Simulation au pas de temps journalier.....	67



Chapitre V : Application du Modèle de Simulation

V. 1	<i>DONNEES UTILISEES.....</i>	72
V. 2	<i>TRANSFORMATION DES DONNEES.....</i>	73
V. 3	<i>SIMULATION DES VARIABLES CLIMATOLOGIQUES.....</i>	73
	V-3-1 Simulation de la pluviométrie.....	73
	V-3-1-1 Simulation au pas de temps mensuel.....	73
	V-3-1-2 Simulation au pas de temps décadaire.....	75
	V-3-1-3 Simulation au pas de temps journalier.....	78
	V-3-2 Simulation de l'ETP.....	83
	V-3-2-1 Simulation au pas de temps mensuel.....	83
	V-3-2-2 Simulation au pas de temps décadaire.....	85
	V-3-2-3 Simulation au pas de temps journalier.....	88
V. 4	<i>AUTOMATISATION DE LA PROCEDURE DE SIMULATION.....</i>	93
V. 5	<i>EFFET DU NOMBRE DE COMPOSANTES PRINCIPALES.....</i>	99
V. 6	<i>EFFET DE LA TAILLE DE L'ECHANTILLON.....</i>	99
V. 7	<i>EFFET DE LA CHRONOLOGIE.....</i>	100
V. 8	<i>EFFET DE LA STATIONNARITE.....</i>	100

Chapitre VI : Elaboration du Modèle de Prévision

VI. 1	<i>INTRODUCTION.....</i>	101
VI. 2	<i>MODELE DE PREVISION.....</i>	101
	VI-2-1 Calage du modèle.....	103
	VI-2-2 Test du modèle.....	103
	VI-2-3 Exploitation du modèle.....	103
VI. 3	<i>CRITERES DE VALIDATION.....</i>	103
VI. 4	<i>PREVISION DES DIFFERENTES VARIABLES CLIMATOLOGIQUES.....</i>	104
	VI-4-1 Prévision de la pluviométrie.....	105
	VI-4-2 Prévision de l'ETP.....	118

<i>CONCLUSION GENERALE.....</i>	122
<i>BIBLIOGRAPHIE.....</i>	125
<i>ANNEXES.....</i>	133

LISTE DES TABLEAUX

Chapitre II : Collecte et Critique des Données

II-1 Caractéristiques de la station d'Alger.....	22
II-2 Caractéristiques statistiques de la pluviométrie de la station d'Alger.....	23
II-3 Résultats du Test de Grubbs et Beck.....	26
II-4 Résultats du Test d'homogénéité.....	28
II-5 Résultats du Test d'indépendance.....	29

Chapitre III : Théories de l'Analyse en Composantes Principales et des Chaînes de Markov

III-1 Codification des variables étudiées.....	47
III-2 Paramètres statistiques des variables étudiées.....	47
III-3 Matrice des corrélations.....	47
III-4 Vecteurs propres (matrice de passage).....	48
III-5 Contribution de chaque composante principale.....	48
III-6 Projection des variables étudiées sur les axes principaux.....	48
III-7 Matrice de passage d'ordre 1.....	51
III-8 Matrice de passage d'ordre 2.....	52

Chapitre IV : Elaboration du Modèle de Simulation

IV-1 Valeurs de n en fonction de C_s	63
IV-2 Valeurs de $\Gamma(n)$ en fonction de n	63
IV-3 Relation entre les différents paramètres de la loi GEV pour la population Z	65

Chapitre V : Application du Modèle de Simulation

V-1 Contribution de chaque CP à la variance totale.....	74
V-2 Contribution de chaque CP à la variance totale.....	75
V-3 Contribution de chaque CP à la variance totale.....	76
V-4 Contribution de chaque CP à la variance totale.....	77
V-5 Contribution de chaque CP à la variance totale.....	79
V-6 Contribution de chaque CP à la variance totale.....	81
V-7 Contribution de chaque CP à la variance totale.....	83

V-8 Contribution de chaque CP à la variance totale.....	85
V-9 Contribution de chaque CP à la variance totale.....	86
V-10 Contribution de chaque CP à la variance totale.....	87
V-11 Contribution de chaque CP à la variance totale.....	89
V-12 Contribution des CP	91



Chapitre VI : Elaboration du Modèle de Prévision

VI-1 Qualité du critère Nash.....	104
VI-2 Valeurs des β_{ij} pour chaque mois.....	106
VI-3 Résultats des tests de validation sur les variables brutes.....	108
VI-4 Valeurs de β_0 pour chaque mois.....	110
VI-5 Valeurs des β_{ij} correspondants à chaque mois.....	111
VI-6 Résultats de la modélisation pour les variables normées.....	112
VI-7 Résultats de la modélisation pour les variables racines brutes.....	114
VI-8 Résultats de la modélisation pour les variables racines normées.....	115
VI-9 Résultats de la modélisation pour les variables brutes.....	117
VI-10 Résultats de la modélisation pour les variables normées.....	118
VI-11 Résultats de la modélisation pour les variables brutes	119
VI-12 Résultats de la modélisation pour les variables normées.....	120
VI-13 Résultats de la modélisation pour les variables normées.....	121

LISTE DES FIGURES

Chapitre II : Collecte et Critique des Données

II-1 Représentation de la station d'Alger.....	23
II-2 Variation de la pluie annuelle par rapport à la moyenne.....	23
II-3 Reconstitution des valeurs manquantes.....	25

Chapitre III : Théories de l'Analyse en Composantes Principales et des Chaînes de Markov

III-1 Organigramme de l'Analyse en Composantes Principales.....	42
III-2 Projection des variables étudiées sur le plan d'axe 1x2.....	49
III-3 Projection des variables étudiées sur le plan d'axe 2x3.....	49

Chapitre IV : Elaboration du Modèle de Simulation

IV-1 Passage d'une fonction de probabilité F à une loi uniforme.....	58
IV-2 Tirage aléatoire à partir de la fonction de répartition.....	59
IV-3 Organigramme principale du modèle de simulation.....	68
IV-4 Sous programme TRACP.....	69
IV-5 Sous programme PSACP.....	70
IV-6 Sous programme SFACP.....	71

Chapitre V : Application du Modèle de Simulation

V-1 Evolution de la pluviométrie journalière.....	73
V-2 Ajustement par la loi de Gumbel $CP1$	74
V-3 Evolution des résidus des pluies mensuelles.....	74
V-4 Ajustement par la loi Normale $EPS2$	75
V-5 Ajustement par la loi de Goodrich $CP8$	76
V-6 Ajustement par la loi de Goodrich $BETA14$	76
V-7 Ajustement par la loi de Goodrich $EPS2$	77

V-8	Ajustement par la loi de Goodrich COV11.....	77
V-9	Ajustement par la loi de Goodrich EPS26.....	78
V-10	Ajustement par la loi de Goodrich CP2 Déc.....	80
V-11	Ajustement par la loi de Goodrich BETA1 Mars.....	80
V-12	Ajustement par la loi de Goodrich EPS10 Jan.....	80
V-13	Ajustement par la loi de Goodrich CP4 Fev.....	82
V-14	Ajustement par la loi de Goodrich COV1 Avr.....	82
V-15	Ajustement par la loi de Goodrich EPS20 Oct.....	83
V-16	Evolution de l'évapotranspiration journalière.....	83
V-17	Ajustement par la loi Normale CP4.....	84
V-18	Ajustement par la loi Normale EPS5.....	84
V-19	Ajustement par la loi Normale BETA01.....	85
V-20	Ajustement par la loi Normale EPS1.....	85
V-21	Ajustement par la loi de Goodrich CP4.....	86
V-22	Ajustement par la loi de Goodrich BETA14.....	86
V-23	Ajustement par la loi de Goodrich EPS20.....	87
V-24	Ajustement par la loi de Goodrich BETA011.....	87
V-25	Ajustement par la loi de Goodrich CP1.....	88
V-26	Ajustement par la loi de Goodrich COV26.....	88
V-27	Ajustement par la loi de Goodrich EPS26.....	88
V-28	Ajustement par la loi de Goodrich CP1 Nov.....	90
V-29	Ajustement par la loi de Goodrich BETA01 Avr.....	90
V-30	Ajustement par la loi de Goodrich CP4 Déc.....	92
V-31	Ajustement par la loi de Goodrich EPS10 Oct.....	92
V-32	Pluie moyenne mensuelle interannuelle simulée.....	97
V-33	Ecart-type de la pluie moyenne mensuelle interannuelle simulée.....	98
V-34	ETP moyenne mensuelle interannuelle simulée.....	98
V-35	Ecart-type de l'ETP moyenne mensuelle interannuelle simulée.....	98
V-36	Evolution de la pluviométrie journalière (fichier tronqué).....	100

Chapitre VI : Elaboration du Modèle de Prévision

VI-1	Evolution des coefficients de régression en fonction du temps.....	105
VI-2	Evolution de la 1 ^{ère} composante principale.....	106
VI-3	Evolution des variables résiduelles.....	107
VI-4	Prévision de l'année 1992.....	108
VI-5	Prévision des pluies mensuelles.....	109
VI-6	Evolution des moyennes mensuelles interannuelles.....	109
VI-7	Evolution des β_{ij} pour chaque mois.....	110
VI-8	Evolution de la 1 ^{ère} composante principale.....	111
VI-9	Evolution des variables résiduelles.....	112
VI-10	Prévision de l'année 1992.....	112
VI-11	Prévision des pluies mensuelles.....	113
VI-12	Prévision de l'année 1997.....	113
VI-13	Prévision de l'année 1992.....	114
VI-14	Prévision des pluies mensuelles.....	114
VI-15	Prévision de l'année 1992.....	115

Introduction Générale

A l'aube du XXIème siècle, la question sensible concernant les réserves en eau douce et la manière d'en optimiser l'utilisation à l'avenir apparaît désormais incontournable et récurrente.

Dès 1992, la déclaration de Dublin sur l'eau dans la perspective d'un développement durable (Conférence Internationale sur l'Eau et l'Environnement, Dublin, janvier 1992) rappelait que :

" La rareté de l'eau douce et son emploi inconsidéré compromettent de plus en plus gravement la possibilité d'un développement écologiquement rationnel et durable. Santé et bien-être de l'homme, sécurité alimentaire et industrialisation sont autant de domaines menacés, de même que les écosystèmes dont ils dépendent, sauf à opter dans la présente décennie et au-delà, pour une gestion plus efficace de l'eau et des sols ".

" Les problèmes n'ont pas un caractère théorique et ce n'est pas dans un avenir lointain qu'ils affecteront notre planète, ils sont bien réels et nous en ressentons les effets aujourd'hui. Il y va de la vie de millions d'êtres humains et une action efficace s'impose dans l'immédiat ".

Pour répondre à ces problèmes et à ces besoins, les hydrologues ont développé des outils que l'on appelle modèles hydrologiques. Ces modèles ont évolué très rapidement depuis quelques décades avec l'avènement de l'informatique et d'ordinateurs de plus en plus puissants.

Depuis, les chercheurs ont apporté plus d'attention à ces modèles en s'intéressant aux techniques de simulation et de prévision, en hydrologie. En effet plusieurs travaux ont été consacrés à ces domaines en vu :

- De mieux gérer les ressources en eau ;
- D'optimiser le dimensionnement des ouvrages d'assainissement urbains et agricoles ;
- D'étudier les comportements des bassins versants, et des ouvrages hydrauliques ;
- De dimensionner les ouvrages de retenue ;
- D'étudier d'éventuels changements climatiques ;
- D'optimiser les irrigations ;
- D'éviter les catastrophes naturelles.

La connaissance exacte des ressources en eau, et principalement des variables climatologiques auxquelles on peut s'attendre, sont des facteurs importants pour la planification des besoins en eau (consommation domestique, agricole, industrielle), et le développement des régions.

Dans ce travail, notre première tâche fut de relever dans la bibliographie, les différentes approches utilisées dans la modélisation en hydrologie, et plus particulièrement les modèles mathématiques utilisés pour l'étude des techniques de simulation et de prévision.

Dans ce contexte, rappelons que l'apparition de ces nombreux modèles mathématiques appliqués à la simulation et à la prévision hydrologique est un des effets de la préoccupation

principale des hydrologues qui consiste à mieux gérer la ressource en eau. Certains de ces modèles sont difficilement exploitables en raison de leur complexité, et des moyens d'acquisition des données à mettre en œuvre.

De plus, la plupart des modèles de simulation et de prévision des variables climatologiques élaborés exigent :

- La chronologie des données ;
- La stationnarité du phénomène étudié ;
- La normalité de l'échantillon ;
- La dépendance entre les valeurs observées ;
- La consistance des données ;
- Type de variable (aléatoire ou cyclique) ;
- Pas de temps spécifique.

C'est la raison pour laquelle nous avons veillé à élaborer une méthodologie, basée sur l'analyse en composantes principales (ACP), pour la simulation et la prévision des variables climatologiques facilement exploitables avec un nombre de paramètres du modèle le plus restreint possible en s'affranchissant de toutes les contraintes exigées par les autres modèles.

L'élaboration du modèle s'est faite selon plusieurs étapes :

La première étape consiste, grâce à l'analyse en composantes principales, à condenser l'information contenue dans les valeurs de l'échantillon, à mettre ainsi en évidence les effets prépondérants du phénomène étudié, d'élaborer la structure de base du modèle de simulation et de prévision, et de calculer ensuite les différents paramètres du modèle élaboré.

Dans la deuxième étape nous avons simulé les différents paramètres du modèle élaboré par le biais de diverses techniques de simulation selon le type, et le pas de temps de la variable étudiée.

Une fois toutes les possibilités de simulation testées, intervient alors la troisième étape dont le but est de choisir le meilleur modèle de simulation retenu selon le type et le pas de temps du phénomène étudié après avoir été validé par une série de tests statistiques.

Dans la dernière étape nous nous sommes intéressés à la prévision des paramètres du modèle élaboré selon plusieurs approches, des tests de validation graphiques et numériques ont été utilisés, pour tester la fiabilité des résultats de ce modèle.

Les données utilisées dans ce travail proviennent de l'Office National de la Météorologie (ONM) station d'Alger.

Ce travail est articulé autour :

- Un premier chapitre consacré à une revue bibliographique sur la modélisation en hydrologie ;
- L'analyse et la critique des données recueillies a fait l'objet du deuxième chapitre ;
- Les théories de l'analyse en composantes principales et des chaînes de Markov sont

traitées dans le troisième chapitre ;

- Le concept général et les bases théoriques de notre modèle de simulation et de prévision sont développés au chapitre quatre ;
- Le modèle élaboré est ensuite appliqué pour la simulation des données recueillies, et est testé et analysé au chapitre cinq ;
- Le chapitre six présente les différents résultats de prévision obtenus à partir du modèle développé pour les deux types de variables (aléatoire et cyclique), au pas de temps mensuel.
- Enfin une conclusion générale soulignant l'originalité du travail accompli.

En outre, une attention particulière, tout au long de cette thèse, a été portée sur le développement de logiciels informatiques, en effet au niveau de chaque phase un programme en fortran 90 a été développé ce qui nous a grandement facilité la tâche.

Chapitre I : Revue Bibliographique

Dans ce chapitre nous présentons une synthèse bibliographique des différentes approches de la modélisation hydrologique et des travaux publiés qui traitent de l'utilisation de la simulation et de la prévision de variables hydroclimatologiques aléatoires et cycliques à différents pas de temps tout en indiquant les limites de chaque modèle.

Avec le développement des méthodes numériques et des outils informatiques, les modèles hydrologiques se sont multipliés au cours de ces dernières années, se compliquant davantage à chaque fois que leur concepteurs ont cherché à décrire les processus réels du cycle hydrologique d'une manière complète et rigoureuse. Certains de ces modèles nécessitent souvent une masse importante de données relatives au bassin versant et aux processus physiques et hydrologiques, qui ne sont pas toujours facilement mesurables sur le terrain, les procédures d'estimation deviennent complexes. De plus, ces modèles sont parfois confrontés à d'autres difficultés : Validation des hypothèses qui ont permis leur construction, maîtrise des conditions initiales et aux limites, etc. En attendant de pouvoir exploiter ces modèles et face aux demandes immédiates des hydrologues pour résoudre les problèmes liés aux applications hydrologiques, on continue à développer des modèles simples et fiables permettant de répondre à une partie des attentes des hydrologues.

Les modèles hydrologiques sont nés bien avant l'avènement de l'informatique, mais les ordinateurs ont permis aux modèles de se développer et les hydrologues ont pu mettre au point des logiciels de plus en plus performants.

La littérature scientifique présente une très grande diversité de modèles dont très peu ont trouvé une utilisation opérationnelle. La question de l'utilisation de modèles en hydrologie, (pour effectuer des simulations, des prévisions, de dimensionner des ouvrages d'art, ou délimiter des zones inondables, etc.), se pose en particulier pour les bassins versants pour lesquels nous disposons de séries de données mesurées.

Ceci dit, nous pouvons distinguer trois types d'utilisation des modèles mathématiques en hydrologie :

- La modélisation comme outil de recherche : Cela reste aujourd'hui encore le champ d'application privilégié de la modélisation hydrologique. La modélisation peut être utilisée pour interpréter des données mesurées. Différents scénarii de fonctionnement hydrologique des bassins versants peuvent être confrontés aux mesures.
- La modélisation comme outil de prévision : Anticipation des évolutions futures du débit d'un cours d'eau. Il s'agit de l'utilisation opérationnelle la plus courante des modèles hydrologiques. Dans la plupart des cas cependant, les modèles développés sont basés sur des régressions linéaires entre les variables indépendantes (*pluie*, débits amont) et les variables dépendantes (débits aval), et ne s'intéressent que très peu aux processus hydrologiques.
- La modélisation comme outil d'extrapolation : Reconstitution de séries de débits plausibles. Dans certains cas, comme par exemple, le dimensionnement des évacuateurs de crue de barrages ou la délimitation des zones inondables dans le

cadre d'un plan de prévention des risques, il est nécessaire de proposer des scénarii de crues ou éventuellement d'étiages, de période de retour nettement supérieure à la durée d'observation des débits sur le site étudié. L'utilisation de longues séries de pluie-éventuellement générées à l'aide d'un modèle stochastique couplé à un modèle hydrologique permet de reconstituer des scénarii rares.

Les approches habituellement utilisées en modélisation sont :

1 – Le modèle déterministe : modèle qui associe à chaque jeu de variables d'entrée, de variables d'état et de paramètres une valeur réalisation unique des variables de sortie. Ils sont caractérisés donc par l'unicité de leur réponse.

2 – Le modèle stochastique : l'une au moins des variables d'entrée ou des variables d'état ou des paramètres est une variable aléatoire. Par voies de conséquence, la ou les variables de sortie sont des variables aléatoires. La reconstitution de la distribution des variables de sortie nécessite des simulations répétées en tirant aléatoirement la valeur de la variable d'entrée.

Dans ce genre de modèles, la notion de probabilité d'apparition d'un événement est introduite, si celle-ci est fonction du temps, on dit que le modèle est stochastique, sinon, on parle alors de modèle probabiliste.

Il existe une bibliographie très riche sur la modélisation hydrologique (Sing, 1988; Lee, 1993; Buchetele et al., 1996) et un grand nombre de critique ont été faites sur le choix des types de modèles à utiliser (Deschesnes et al., 1985), leur application (Clarke, 1973 ; Bathurst et al., 1992 ; De Marsily, 1994) leur efficacité ainsi que les problèmes rencontrés dans leur calage (Comer, 1976 ; Diskin et al., 1977; Garrick et al., 1978 ; Gupta et al., 1982 ; Bates, 1992 ; Chapell et al., 1992 ; Duan et al., 1992 ; Beckie, 1996).

I-1 LA SIMULATION :

Les modèles de simulation ne cherchent pas à reproduire une partie de la réalisation d'un processus, mais à générer des scénarii «possibles» d'un processus dont on connaît une réalisation. D'un point de vue pratique ces derniers, par exemple s'il s'agit de débits, peuvent être utiles comme données d'entrée pour vérifier le bien fondé d'une gestion de crue, pour calculer le bilan économique de l'exploitation d'un barrage réservoir, etc.

Les modèles de simulation sont souvent sollicités pour décrire les évolutions temporelles, autant que spatiales, en générant des séries synthétiques. Par exemple générer plusieurs centaines d'années de débits journaliers.

Leur principe est d'utiliser des générateurs de variables aléatoires qui devront respecter la structure statistique des processus à reproduire (MIQUEL, 2001).

Différents types de modèles stochastiques ont été proposés pour la modélisation des séries hydrologiques en général, et des séries d'apports en particulier. Ces modèles comprennent le modèle de Monté-Carlo (Métropolis 1953), le modèle autorégressif (Thomas et al., 1962), le modèle du bruit gaussien fractionnel FGN (Mandelbrot et al., 1969), les modèles autorégressifs à moyenne mobile ARMA (Carlson et al., 1970, O'Connel, 1971, Hipel et al., 1977), le modèle de la ligne brisée BL (Rodriguez et al., 1972), le modèle de désagrégation (Mejia et al., 1976), Les modèles ARMA-Markov (Lettenmaier et al., 1977).

En général ces modèles sont formulés tel que la valeur courante d'une variable est la somme pondérée des valeurs antécédentes et des nombres aléatoires représentant des effets inconnus.

Les modèles de simulation sont basés sur la génération de nombres aléatoires, l'aléatoire est un domaine souvent méprisé, il regroupe le domaine de ce qui est possible sans être sûr, ce que l'on ne peut maîtriser mais pas forcément ce que l'on ne peut utiliser.

I-1-1 La méthode de Monté-Carlo :

Celle-ci a été utilisée pour simuler des processus aléatoires tels que le mouvement thermique des particules ou pour obtenir une approximation d'une opération mathématique. Et ainsi, des résultats obtenus lors de ces simulations, on a pu déduire, des solutions aux problèmes les plus complexes.

Les principes de cette méthode ont été exposés par Métropolis dès 1953. Elle a été, entre autre, introduite en physique de la matière condensée et dans bien d'autres domaines du fait de ses facultés à contourner certains des problèmes rencontrés par les chercheurs.

Elle peut aussi être appliquée comme méthode de simulation, ou comme méthode d'échantillonnage.

Pour la génération de nombre aléatoire, on est souvent amené à considérer une fonction $g(X)$.

Dans le cas de la méthode de Monté-Carlo, cette fonction génère le nombre aléatoire suivant l'algorithme $X_{k+1} = g(X_k)$, en initialisant la série par une valeur X_0 .

La forme générale de cette méthode s'écrit :

$$[aX_0+c][\text{mod}m] \quad (\text{I-1})$$

X_0 : Valeur initiale positive ;

a : Multiple positif ;

c : Incrément positif ;

m : Modulo.

La séquence de nombre aléatoire est obtenue par :

$$X_{n+1} = [aX_n + c] \cdot [\text{mod}m] \quad (\text{I-2})$$

Pour générer des séquences de périodes relativement importantes (ce qui est souvent nécessaire), il est proposé de choisir des valeurs assez grandes de m .

La méthode de Monté-Carlo présente l'avantage de n'être que très peu limitative. En effet, elle permet d'envisager la résolution de problèmes par voies numériques quelle que soit la complexité des équations.

Le défaut majeur de la méthode de Monté-Carlo reste le temps de calcul. En effet, elle présente l'inconvénient d'accorder exactement le même poids au tirage des événements rares et fréquents. On se voit donc dans l'obligation d'imposer de très grands nombres de tirages

pour obtenir une barre d'erreur acceptable, par conséquent les temps de calculs s'allongent jusqu'à devenir dissuasifs.

En hydrologie, d'autres approches ont été élaborées, les modèles saisonniers appliqués aux séries mensuelles étaient utilisés à l'origine par Thomas et al., (1962) et par Roesner et al., (1966).

1-1-2 Le modèle de Thomas et Fiering :

Ce modèle est basé sur 12 équations de régression, où chaque mois j est régressé sur le mois précédent $j-1$, dans lesquelles on tient compte de la structure corrélacionnelle d'un mois à un autre non stationnaire. Dans sa formulation la plus simple il s'écrit comme suit :

$$q_{j+1} = \bar{q}_{j+1} + b_j (q_j - \bar{q}_j) + Z_j \cdot S_{j+1} \cdot \sqrt{1 - r_{(j,j+1)}^2} \quad (I-3)$$

q_j : Débit du mois j ;

\bar{q}_j : Débit moyen mensuel interannuel du mois j ;

b_j : Coefficient de régression pour l'estimation du débit du mois $j+1$ à partir du mois j

$$b_j = r_{(j,j+1)} \cdot \frac{S_{j+1}}{S_j} \quad (I-4)$$

Z_j : Ecart-type normale aléatoire $N(0,1)$;

S_{j+1} : Ecart-type entre les débits au pas j et $j+1$;

$r_{(j,j+1)}$: Coefficient de corrélation entre les mois j et $j+1$.

Si on considère la variable débit centrée réduite $Q = \frac{q - \bar{q}}{S}$, la formulation du modèle devient :

$$Q_{j+1} = r_{(j,j+1)} Q_j + Z_j \sqrt{1 - r_{(j,j+1)}^2} \quad (I-5)$$

Ce modèle peut être généralisé de deux manières :

- Faire une régression multiple au lieu d'une régression simple, ce qui revient à agrandir le temps de réponse :

Soit par exemple la régression multiple du second degré :

$$Q_{j+1} = r_{(j,j+1)} Q_j + r_{(j+1,j-1)} Q_{j-1} + Z_j \sqrt{1 - R_{(j,j+1)}^2} \quad (I-6)$$

R : Coefficient de corrélation multiple.

- Introduire d'autres variables comme la hauteur de la pluie.

Le modèle de Thomas et al., mis au point en 1962 pour générer des séries mensuelles, présente l'avantage de résoudre le problème de la non stationnarité, et ce en introduisant des paramètres

saisonniers au modèle autorégressif (AR), mais l'inconvénient de son utilisation reste la multiplication des paramètres, et par conséquent les problèmes d'estimation de ces derniers.

I-1-3 Les modèles FGN (Fractional Gaussian Noise) :

Mandelbrot (1965) a construit une classe de processus aléatoires appelés « Fractional Gaussien Noise » (FGN) ou modèles fractals ou multifractals, (Mandelbrot et al., (1968), Mandelbrot, (1975) et Schertzer (1993)), qui ont connu un développement récent. Ils offrent un cadre formel prenant en compte directement une forme d'invariance d'échelle du phénomène permettant ainsi une réduction du nombre de paramètres et une extrapolation de l'information entre les échelles, Chaouche (2001).

Pour la génération hydrologique, les FGN tentent de préserver les caractéristiques naturelles et par la même occasion les longues périodes d'étiages ou de crues. Ceci est appelé effet Joseph, qui découle directement de la propriété de la persistance infinie des processus FGN, ces derniers tentent également de reproduire les valeurs extrêmes (élevées ou faibles), cette caractéristique est dite effet Noah.

Le modèle fractal ne conduit cependant pas toujours à des résultats fiables (Chaouche, 2001). En effet, le critère de choix du facteur d'échelle n'est pas établi. De plus, la forte intermittence des séries pluviométriques rend inutile l'exploration d'un comportement fractal en dessous d'un pas de temps minimal, en outre, et dans le but d'accéder à une quantification de la qualité d'estimation, ce processus a été étudié dans un cadre probabiliste (Falconer 1990, Harte 1997).

I-1-4 Le modèle BL (Broken Line) :

Le processus ligne brisée (Broken Line) fut introduit par Ditlevsen (1971) pour vérifier par la méthode de Monté-Carlo quelques-uns de ses résultats. Il a été utilisé principalement dans le domaine de la géophysique, son apport comme un modèle possible pour la simulation de séries temporelles est basé sur deux propriétés :

- Existence de la dérivée secondaire à l'origine de sa fonction de corrélation.
- La possibilité de modéliser de longues interdépendances parmi ces composantes de telle sorte que la simulation du procédé soit plus rapide que celle de n'importe quel autre modèle connu.

Le modèle a été jugé utile au moins dans certains cas, et flexible, le grand inconvénient reste toujours l'estimation de nombreux paramètres.

I-1-5 Les modèles ARMA (Auto Regressive Moving Average) :

Les modèles autorégressifs ont été utilisés depuis le début des années soixante, leur développement historique en hydrologie peut être divisé en deux étapes :

- La décennie de 1960 a été initiée principalement par les travaux de Thomas et al., (1962) et Yevjevich (1963), durant cette première période, la procédure utilisée pour l'estimation des paramètres des modèles était basée sur la méthode des moments, et les tests d'adéquation étaient basés sur l'analyse du corrélogramme.

- La décennie de 1970 motivée par la publication du livre de Box et al., (1970), depuis les chercheurs utilisent des méthodologies plus raffinées, spécifiquement pour la vérification des hypothèses du modèle et la sélection du modèle adéquat.

La formulation mathématique du modèle autorégressif et à moyenne mobile (ARMA) d'ordre (p,q) est la combinaison de deux modèles, l'un autorégressif (AR) d'ordre p, et l'autre linéaire dit de moyenne mobile (MA) d'ordre q, et est défini par (Box et al., 1970) :

$$Z_t = \sum_{j=1}^p \phi_j * Z_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j * \varepsilon_{t-j} \quad (I-7)$$

ϕ_1, \dots, ϕ_p : Paramètres autorégressifs ;
 $\theta_1, \dots, \theta_q$: Paramètres de moyenne mobile ;
 ε_t : Variables aléatoires indépendantes avec une moyenne nulle, et une variance σ_ε^2 ;
 Z_t : ($t=1, 2, \dots, N$) est la série à modéliser, elle est de moyenne nulle ;
 N : Taille de la série à modéliser.

Les paramètres du modèle ARMA sont donc : $\mu, \sigma_\varepsilon^2, \theta_1, \theta_2, \dots, \theta_q, \phi_1, \phi_2, \dots, \phi_p$, soit au total $p+q+2$ paramètres à estimer à partir des données.

La série à modéliser, au moyen d'un processus ARMA, doit être stationnaire et symétrique :

$$Z_t = Y_t - \mu \quad (I-8)$$

Tao et al., (1976), Delleur et al., (1976) avaient utilisé un modèle ARMA (p,q) à paramètres périodiques pour la génération des débits des rivières. Par contre Salas et al., (1981) ont utilisé pour la modélisation de séries de débits annuels et mensuels des rivières, les modèles ARMA à coefficients constants.

L'application de ces modèles présente un intérêt particulier dans la modélisation des séries chronologiques en hydrologie Bouziane (1989), car la forme autorégressive est une caractéristique des séries étudiées (séries autocorrélées) ainsi que pour leur simplicité d'utilisation.

Afin de comparer les résultats obtenus par les deux modèles ARMA à coefficients constants et périodiques, Bouziane (1989) a utilisé les débits moyens mensuels de sept séries marocaines, d'une série européenne (Meuse) et d'une série africaine (fleuve du Sénégal). Ce choix a été motivé par la différence de climat qui existe entre les différentes séries.

La modélisation par le modèle ARMA à coefficients constants, pour les séries marocaines, a permis de mettre en évidence une certaine instabilité du point de vue ressemblance des caractéristiques statistiques. Le contraire a été observé pour les séries de la Meuse et du fleuve du Sénégal.

La modélisation par le modèle ARMA à coefficients périodiques, a montré que les caractéristiques statistiques (moyennes, variances, ...) se reproduisent de manière acceptable pour les séries marocaines et se reproduisent mieux pour les séries de la Meuse et du fleuve du Sénégal. Ceci est peut être dû à la régularité de l'écoulement de ces deux dernières rivières.

En outre les résultats montrent que les écarts obtenus entre les séries historiques et les séries générées sont plus importants pour le cas des modèles ARMA périodiques que pour les modèles ARMA à coefficients constants. Il est donc conseillé d'utiliser les modèles ARMA à coefficients constants et d'éviter dans la mesure du possible les modèles ARMA périodiques car ces derniers ne conduisent à aucune amélioration.

Malgré leur très large utilisation dans la simulation des séries hydrologiques, il n'en demeure pas moins que ces modèles présentent les principaux inconvénients suivants :

- Un nombre important de paramètres à estimer.
- La normalité et la stationnarité de la série étudiée sont exigées.
- La dépendance entre les variables étudiées (problème d'autocorrélation).

I-1-6 Les modèles de désagrégation :

Les modèles de désagrégation ont été introduit pour la première fois en hydrologie par Harms et al., (1966), mais le premier modèle qui fut accepté est celui présenté par Valencia et al., (1973), ce dernier est la base des techniques de désagrégation dans la simulation des séries hydrologiques.

Ce modèle est un processus par lequel des séries temporelles sont générées, dépendant des séries déjà disponibles. Les séries de départ sont générées par un modèle stochastique choisi, puis désagrégées en sous séries.

Généralement dans le cas de la désagrégation saisonnière (données saisonnières) la somme des sous séries doit donner exactement le série originale (données annuelles) ou série clé.

Il est à noter que la désagrégation n'est pas limitée à un seul niveau, autrement dit, une série annuelle peut être désagrégée en série semi-annuelle en première étape, qui elle-même peut être désagrégée en séries mensuelles.

Les deux formes de bases des modèles de désagrégation sont temporelles (désagrégation des données annuelles en données mensuelles et décadaires) et spatiales (désagrégation des débits d'une rivière en débit individuel des affluents).

La plupart des applications de la désagrégation sont dans le domaine temporel, Lane (1979) a appliqué les mêmes principes dans le domaine spatial. Il existe trois formes du modèle de désagrégation :

- Le modèle de base : proposé par Valencia et al., (1973), il est de la forme :

$$Y = A \cdot X + B\varepsilon \quad (I-9)$$

- X : Série mère ;
 Y : Série des valeurs saisonnières simulées ;
 A, B : Paramètres du modèle ;
 ε : Terme aléatoire.

C'est le modèle le plus simple, sa structure permet de préserver la covariance entre les valeurs annuelles, et les valeurs saisonnières, et de préserver la variance et la covariance entre valeurs

saisonniers.

- Le modèle étendu : développé par Mejia et al., (1976), est une extension du modèle de base, un terme supplémentaire a été introduit pour préserver les covariances saisonnières entre saisons successives, il s'écrit :

$$Y = A \cdot X + B \cdot \varepsilon + C \cdot Z \tag{I-10}$$

Z : Matrice colonne contenant des valeurs de l'année précédente ;

C : Matrice des paramètres supplémentaires.

Le nombre de paramètres de ce modèle est excessif, ce qui rend leur estimation plus complexe.

- Le modèle condensé : Lane (1979) a développé une approche qui annule chaque paramètre du modèle étendu qui n'est pas important, c'est le modèle condensé, le nombre de paramètres estimés, ainsi que le nombre de moments préservés est réduit, le modèle utilisé est le modèle étendu mais sur une base saisonnière, il est donné par :

$$Y = A_{\tau} X + B_{\tau} \varepsilon + C_{\tau} Y_{\tau-1} \tag{I-11}$$

τ : Saison à générer, donc il y a W saisons, donc W équations individuelles suivant la forme de l'équation et donc W ensembles de paramètres $A_{\tau}, B_{\tau}, C_{\tau}$.

Le principal avantage de ce modèle est la réduction du nombre de paramètres.

D'autres auteurs ont utilisé la désagrégation temporelle (Olsson, 1998). Dans ce cas, la hauteur totale de pluie mesurée sur un pas de temps (la journée ou plus) est répartie de manière aléatoire entre les sous-unités de ce pas de temps.

Ce modèle connaît un certain succès. Il bénéficie d'un corpus théorique important. Il existe en effet un lien étroit entre les processus multifractals et certains modèles de désagrégation (Schertzer et al., 1992). Par ailleurs, ce type de modèle permet de tirer profit des nombreuses données de pluie disponibles au pas de temps journalier pour produire des séries à des pas de temps plus fins. Cependant, les recherches les plus récentes sur des séries pluviométriques de climats tempérés (Olsson, 1998, Lima 1998) concluent que les modèles classiques de désagrégation ne permettent pas de rendre compte de la répartition temporelle des pluies à des pas de temps inférieurs à l'heure. Cet échec est en partie lié au fait que la répartition des intensités de pluie au sein d'un pas de temps n'est pas indépendante des valeurs d'intensité des pas de temps précédents et suivants. Ceci est particulièrement vrai lorsqu'on s'intéresse aux petites échelles temporelles. Les modèles de désagrégation classiques ne peuvent tenir compte de ce phénomène.

I-1-7 Les modèles Markoviens :

Une classe particulière de modèles dont l'évolution n'est pas déterministe (les chaînes de Markov), se différencient des autres par le fait que leur évolution ne dépend que de l'instant présent et du passé le plus récent. Ces modèles sont le plus souvent retenus pour représenter l'occurrence des pluies journalières (Chaouche et al., 1997, Smith et al., 1985).

Le précurseur de cette modélisation est Quetelet en 1852 (d'après Woolhiser, 1991). Van – Thanh (1984) détermine la loi de probabilité des cumuls journaliers appartenant à une période pluvieuse de n jours, en décrivant l'occurrence de pluie par une chaîne de Markov stationnaire d'ordre 1. Mais selon le degré de persistance désiré, les chaînes utilisées peuvent être d'ordre supérieur à 1 et même non homogène. Stern et al., (1984) représentent l'alternance des jours secs et pluvieux par une chaîne de Markov non homogène en se plaçant à l'échelle du jour.

Les processus DARMA (p,q) créés par Buishand (1978) tiennent mieux compte des persistances à long terme que les chaînes de Markov d'ordre plus élevé mais la structure du modèle est difficile à interpréter physiquement.

On distingue les modèles d'occurrence de pluie et les modèles de hauteurs non nulles. Ces deux processus peuvent être modélisés séparément, puis recombinaison, soit modélisés simultanément.

1) Modèles d'occurrence de pluie :

Pour modéliser l'occurrence de pluie en un point fixé, on considère la variable aléatoire discrétisée à un certain pas de temps (généralement la journée) $X_t = 0$ ou 1 selon qu'il pleuve ou non sur ce pas de temps. Deux approches peuvent être considérées :

- la première consiste à s'intéresser aux séquences de jours de pluie consécutifs :

On ajuste une loi à la durée des périodes humides et sèches. Cette voie a été largement exploitée par des auteurs tels que Green (1964) ou Eagleson (1978) qui ont utilisé respectivement une loi exponentielle et une loi de Weibull pour les durées des périodes sèches et humides. Un autre modèle créé par Galloy et al., (1982) utilise la loi binomiale négative pour les durées des périodes sèches et humides. Cependant ces méthodes reposent sur l'hypothèse d'indépendance des périodes pluvieuses entre elles, rarement vérifiée à petit pas de temps (Wilks, 1989).

- La seconde approche utilise les séries à temps discret, série binaire de 0 ou 1 (Dechemi, 1997).

2) Modèle de hauteur non nulles :

L'approche la plus courante est de décrire la distribution des hauteurs $Y(t)$ indépendamment de celle de l'occurrence $X(t)$. Plusieurs types de dépendance caractérisent ces modèles :

- Les $Y(t)$ sont indépendantes : divers lois sont utilisées telles que la loi exponentielle, Gamma ou Weibull (Zucchini, 1992).
- Les $Y(t)$ sont indépendantes mais dépendent de l'occurrence du jour précédent $X(t-1)$: Ce sont les processus dépendants en chaînes, introduits par Katz (1977) qui conditionnent $Y(t)$ et $X(t-1)$. Ces modèles nécessitent un grand nombre de paramètres.

On peut modéliser l'occurrence $X(t)$ et les hauteurs $Y(t)$ séparément ou bien simultanément $Z(t) = (X(t), Y(t))$. La plupart des modèles mixtes s'appuient sur le cadre mathématique construit par le Cam (1961) qui adapte à l'hydrologie un processus d'agrégation de type Neyman-Scott.

I-1-8 Le modèle de Neyman-Scott :

Ce dernier créé en astronomie pour modéliser la formation d'étoiles, présente l'avantage de prendre en compte la dynamique des champs de pluie. De nombreux auteurs se sont attachés à analyser formellement la qualité des estimateurs (Coppertwait, 1991) ou à simuler des données à des pas de temps fins, notamment en hydrologie urbaine (Thauvin, 1998).

Dans ce type de modèle, la durée de temps sec et de temps pluvieux alternent. Les origines des périodes pluvieuses arrivent selon un processus de Poisson. A l'intérieur d'un intervalle pluvieux, la pluie se décompose en une succession d'averses. Chacune d'elles est caractérisée par sa durée et son intensité.

Les hauteurs de pluie successives sont considérées dans la plupart des modèles comme indépendantes. Pour représenter les fluctuations temporelles d'intensités de pluie à des pas de temps inférieurs à la journée, certains auteurs (Coppertwait et al., 1996) ont fait appel au processus de Neyman-Scott. Le hyétogramme d'un événement pluvieux est dans ce cas supposé être constitué par l'agrégation d'impulsions élémentaires de forme rectangulaire ou non (Onof et al., 1994).

Les processus Neyman-Scott présentent des variantes définies par la loi des intensités attribuées à chaque cellule de pluie. Citons les modèles, le Neyman-Scott White Noise (NSWN) de Rodriguez et al., (1984), le Neyman-Scott Rectangular Pulse (NSRP) de Rodriguez et al., (1987), ou le NSRPM (Overney, 1995). Ces derniers sont basés sur une représentation temporelle des averses, chaque averse possède une origine à partir de laquelle est positionné un ensemble de cellules de pluie ayant chacune une durée, et un volume propre.

Cette famille de modèles offre la possibilité de générer des séries pluviométriques à différents pas de temps, compatibles avec des caractéristiques de la série de référence, notamment avec les courbes IDF (Intensité-Durée-fréquence). Le modèle Neyman-Scott (NSRPM) a fait l'objet d'études plus approfondies. Il permet de générer de longues séries de précipitations qui peuvent servir d'entrée à un modèle hydrologique. Cette approche semble prometteuse puisqu'elle constitue une technique valable d'extrapolation du comportement des débits extrêmes à partir de celui des précipitations.

L'application de ces processus au pas de temps horaire se sont avérées peu convaincantes, Coppertwait et al., (1996), constatent que ces modèles permettent difficilement de simuler des événements pluvieux longs et peu intenses, et produisent des valeurs extrêmes d'intensité trop élevées.

I-1-9 Le modèle DSH (Décomposition de la Série Historique) :

Pour simuler des variables hydrologiques cycliques, l'approche basée sur la décomposition de la série historique a été la plus suivie (Dechemi, 1997). En effet, lorsqu'on examine la représentation graphique des séries hydrologiques cycliques, nous constatons la vraisemblance de certaines composantes fondamentales à savoir :

- Un mouvement cyclique qui désigne des oscillations de périodes et d'amplitudes plus ou moins irrégulières.
- Un élément aléatoire qui échappe à toute précision, et qui entraîne des variations accidentelles plus ou moins sporadiques qualifiées de variations résiduelles.

Les données brutes peuvent être considérées comme la superposition de deux séries indépendantes, la première prenant en compte les variations saisonnières (porteuse), la seconde les variations résiduelles ou fluctuations (composante fluctuante).

Pour la séparation de ces deux composantes, on utilise une technique de lissage, la série sera décomposée de la manière suivante (Arlery, 1973) :

$$X_{it} = V_{it} + W_{it} \quad (\text{I-12})$$

X_{it} : Données brutes ;

V_{it} : Composante saisonnière ou porteuse ;

W_{it} : Composante fluctuante ;

$t = 1$ à n (n étant le nombre d'années de la série historique).

La décomposition en séries de Fourier de la série initiale permet d'extraire la composante saisonnière, la composante fluctuante est donnée par :

$$V_{it} = X_{it} - W_{it} \quad (\text{I-13})$$

La simulation des composantes saisonnières et fluctuantes permet d'obtenir des séries synthétiques par superposition de ces dernières.

Sachant que la composante saisonnière peut se mettre sous la forme :

$$V_t = A_0 + a_1 \sin t + b_1 \cos t + a_2 \sin t + b_2 \cos t \dots \quad (\text{I-14})$$

En considérant m harmoniques le lissage sera donné par :

$$V_t = A_0 + \sum_{j=1}^m \sum_{t=1}^T a_j \cos \frac{2\pi j t}{T} + b_j \sin \frac{2\pi j t}{T} \quad (\text{I-15})$$

A_0 : Moyenne annuelle de la série brute ;

a_j, b_j : Coefficient de la série de Fourier ;

M : Nombre d'harmoniques ;

T : Nombre de jours par année.

Les fluctuations sont obtenues en retranchant à la série des données brutes, la composante saisonnière (obtenue par la décomposition en série de Fourier de la série historique).

Les fluctuations sont de moyenne nulle, et leurs distributions suivent une loi normale.

I-1-10 Le modèle SVD (Singular Value Decomposition) :

Les modèles utilisés usuellement pour la simulation des débits requièrent un nombre élevé de paramètres à estimer. Le modèle SVD (Singular Value decomposition) est apparu en alternative à ces derniers (Cavadias 1980, 1985; Haché et al., 1994) il est basé sur le théorème de décomposition en valeurs singulières. Dans sa forme originale le modèle SVD permet de générer des séries de débits. Le but de cette technique est la préservation des moyennes,

variances et corrélations entre les paires de débits de différentes périodes (semaine, mois) de la même année.

$$[X] = [C] \cdot [\lambda]^{1/2} [A]' \quad (I-16)$$

C : Composantes principales standardisées de X ;

λ : Matrice diagonale des valeurs propres ;

A : Matrice des vecteurs propres.

$$[C] = [X][A][\lambda]^{1/2} \quad (I-17)$$

La procédure de calcul se fait en deux étapes : la calibration et la simulation.

- La calibration consiste en un premier lieu à associer à la variable d'origine Y, une variable X tel que $x_{ij} = f_j(y_{ij})$, où f est une fonction destinée à normaliser les débits saisonniers, en seconde étape à calculer la matrice de corrélation de X qu'on notera R_X . En troisième étape à déterminer la matrice C_s des r composantes principales standardisées. En dernier lieu, à ajuster des distributions de probabilités aux éléments de chaque colonne de C_s .
- La simulation par le modèle SVD.

Le modèle SVD dans sa version originale reproduit d'une manière presque identique (en terme de moyennes, écarts types, coefficients de corrélation) l'historique, cette option ne satisfait pas l'une des caractéristiques naturelles des données qui est l'aléa, d'où la nécessité de réexaminer l'algorithme de calcul afin d'introduire une variabilité dans les caractéristiques de l'échantillonnage des débits simulés, pour cela trois variantes ont été étudiées :

- Réduction du nombre de composantes principales de 50% dans la phase de calibration.
- Elimination de la standardisation des composantes principales.
- Elimination de la standardisation des composantes principales et des débits simulés.

Avec le modèle SVD une variabilité a pu être introduite dans les caractéristiques statistiques, pour améliorer la qualité des simulations, l'application de cette variante aux débits observés a permis d'examiner de près le problème de la transformation des données qui influent directement sur la qualité de la simulation.

I-2 LA PREVISION :

Nous avons relevé dans la bibliographie consultée, qu'il existe deux types de prévision :

- Météorologique, faisant appel aux différents réseaux de mesures terrestre et satellitale, disposant d'une base de données climatiques importante qui se caractérise par le nombre important de points de mesure qui l'alimentent, par le nombre de paramètres météorologiques pris en compte, et par la durée des séries climatiques.
- Hydrologique, disposant de forte densité des points de mesure, constituant ainsi des fichiers historiques importants. Ces chroniques d'informations recueillies sont largement utilisées pour l'évaluation des débits des cours d'eau. Suivant les méthodes

utilisées, les données hydrologiques sont alors utilisées soit " seulement " comme variable exogène pour améliorer la qualité des estimations à partir des chroniques des débits observés, soit directement dans la modélisation des écoulements à partir notamment de la pluie (modèles pluie-débit).

L'observation météorologique est la base incontournable de la prévision météorologique et de la climatologie. L'existence, la connaissance et l'étude de ces données météorologiques et plus particulièrement celles concernant les précipitations, sont très importantes en hydrologie que ce soit dans l'étude et la compréhension des régimes des cours d'eau ou dans la prévision des crues et des étiages.

Il y a les prévisions à courte échéance (< 5 jours) qui sont directement données par les modèles dynamiques, et les prévisions à longue échéance (≥ 1 mois) qui peuvent être déterminées par les modèles statistiques (ou numériques), et qui sont jusqu'à présent mal définies.

Il est normale qu'au fur et à mesure que l'intervalle de temps augmente, la prévision devient plus difficile, plus incertaine et prend progressivement une forme moins précise. En effet, la qualité d'une prévision avec les modèles dynamiques diminue quant s'allonge l'échéance. Le délai d'une semaine paraît jusqu'à présent un maximum.

A partir d'un état initial de l'atmosphère, la prévision classique ne peut dépasser quelques jours, mais à partir de l'étude de la dynamique du climat et des traitements statistiques des séries de données historiques, il est possible d'élaborer des modèles de prévision, à caractère général, de la situation pluviométrique mensuelle ou saisonnière sur une région donnée.

Pour prévoir les précipitations à plus longue échéance, une méthode réalisable, mais délicate, consisterait à rechercher dans les annales du temps passé une situation aussi semblable que possible à celle du moment et en fonction du temps qu'il a fait dans le passé, en déduire le temps qu'il fera les prochains jours.

Pour les longues échéances les recherches se basent sur la prévision numérique qui devient sous ses multiples aspects l'un des pôles de la recherche en météorologie.

Un modèle numérique de prévision permet de calculer les paramètres caractérisant l'état de l'atmosphère (pression, température, humidité et vent) à un instant futur connaissant la valeur de ces paramètres à un instant initial.

Le problème de la prévision numérique est un problème mathématique, qui consiste à réunir un jeu suffisant d'équations reliant les paramètres dont on veut fixer l'évolution, notamment la pression, la température, le vent, la précipitation. Les trois premiers paramètres sont déterminés à plusieurs niveaux dont le nombre s'élève avec le degré de perfection de la méthode utilisée.

L'objectif de la prévision à longue échéance des précipitations est de fournir à l'utilisateur, plusieurs mois à l'avance, une prévision saisonnière la plus fiable possible.

L'intérêt de la réalisation d'un tel objectif est d'apporter une aide à la décision pour la gestion rationnelle de l'eau, et la planification des travaux des différents secteurs socio-économiques, principalement ceux qui sont liés à l'agriculture et aux ressources en eau.

I-2-1 La prévision météorologique :

La prévision météorologique à longue échéance est constituée de concepts initialement différents :

I-2-1-1 Prévision par la SST (Surface Sea Temperature) :

Ce premier concept repose sur l'exploitation de l'effet de mémoire qu'exercent les anomalies de température de surface de la mer (SST) sur la circulation atmosphérique générale, et par suite sur les indices des précipitations régionales.

Au début de chaque mois, on recueille les SST de la NOAA (l'Oscillation Atmosphérique Nord Atlantique) observées pour les dernières semaines ; on utilise les modèles statistiques autorégressifs qui permettent de prévoir les SST pour les mois à venir puis on injecte ces températures observées et prévues dans un modèle dynamique, qui génère, à partir d'un état initial, une simulation de l'état atmosphérique global pour chaque pas de temps sur la période de 4 mois. Enfin, un processus de traitement statistique, utilisant les données climatologiques, permet de transformer les sorties du modèle dynamique, en état pluviométrique régionale synthétique (sec, normal ou humide) avec d'autres indicateurs sur les quantiles, les écarts relatifs, le pourcentage par rapport à la moyenne.

En pratique, on injecte les SST observées des mois appropriés aux modèles statistiques qui permettent de fournir une prévision de l'indice de précipitation par région. Ces indices seront interprétés par la suite.

Le paramètre température de surface de la mer (SST) est parmi les conditions aux limites les plus importantes qui influent sur les changements atmosphériques et par suite sur les variabilités climatiques à long terme à cause de sa persistance et de l'importance de l'impact de l'interaction océan-atmosphère sur la circulation générale atmosphérique.

Des études relatives à l'interprétation de la liaison entre SST et la circulation générale atmosphérique ont montré que la SST explique une bonne partie de la variabilité des champs météorologiques sur le Maghreb et le nord atlantique (Driouech et al., 1999).

I-2-1-2 Prévision par la NOA (North Atlantic Oscillation) :

Ce second concept repose sur l'exploitation de l'Oscillation Nord Atlantique (NAO) ainsi que d'autres facteurs du système climatique global. La méthodologie est basée sur l'étude du comportement de NAO et des précipitations, l'étude de la prévisibilité de NAO, la recherche des liaisons possibles entre NAO et les précipitations et enfin, si les scores statistiques le justifient, le développement des modèles de prévision à longue échéance de précipitation en fonction de NAO prévue.

Par définition, l'Oscillation Atmosphérique Nord Atlantique (NOAA) est l'alternance à grande échelle des masses atmosphériques entre deux régions du Nord Atlantique celle de haute pression subtropicale centrée près des Açores et celle de basse pression sub-polaire centrée près de l'Islande. Cette oscillation fluctue entre deux états extrêmes :

Extrême 1: La pression de surface subtropicale sur les Açores est anormalement élevée et la pression de surface sub-polaire près de l'Islande est anormalement basse. Cet état induit un

déficit de précipitation au Maghreb (Driouech et al., 1999).

Extrême 2 : La pression de surface subtropicale sur les Açores est anormalement basse et la pression de surface sub-polaire près de l'Islande est anormalement élevée. Cet état, par contre favorise des précipitations abondantes au Maghreb.

La prévision peut s'étendre à des échéances parfois plus longues, dans le cas où le système couplé océan-atmosphère posséderait des modes de variabilité temporelle de périodes caractéristiques de quelques années. La prévision climatique saisonnière est très souvent construite à partir de simulations de type Monté-Carlo, avec des ensembles de réalisations utilisant des conditions initiales légèrement différentes. Dans l'état actuel de ces prévisions, il est possible de prévoir environ six mois à l'avance un certain nombre de phénomènes climatiques. Il existe, par ailleurs, une prévisibilité à encore plus longue échéance (inter-annuelle à décennale), mais seulement pour certains paramètres et certaines régions. (André et al., 2002).

I-2-1-3 Prévision par l'imagerie satellitale :

Pendant de nombreuses années, l'imagerie satellitale a été interprétée visuellement par les prévisionnistes pour déterminer la position des fronts et des dépressions, pour aider à la compréhension de la structure dynamique de l'atmosphère, et à l'interprétation des résultats des modèles de prévision numérique. Au Royaume Uni, des techniques de reconnaissance sont en cours d'évaluation pour estimer comment cette information peut être extraite de l'imagerie pour l'utilisation dans les systèmes de prévision météorologique automatiques. L'une d'elle, est utilisée avec succès, c'est le réseau neural, basé sur des modèles extrêmement simplifiés du fonctionnement du cerveau humain. Une fois éduqué, le réseau neural est capable de reconnaître les types de nuages dans d'autres images. Un réseau neural a été éduqué avec succès pour différencier les types de masses d'air (dynamiques ou convectives) nécessaires pour un système de prévision d'inondation. (Pankiewicz, 1997).

I-2-2 La prévision hydrologique :

Les modèles hydrologiques de prévision cherchent, à un instant donné t , à «prévoir» pour des instants futurs ($t+\Delta t$) le devenir d'une réalisation connue jusqu'à t . Par exemple, connaissant jusqu'à aujourd'hui le débit d'un cours d'eau en un point donné ainsi que les précipitations en quelques endroits du bassin versant, comment prévoir le débit du même cours d'eau, au même endroit, demain, après demain, ect.?

Leur principe est finalement proche des modèles de simulation, à chaque pas de temps t , prévoir revient à générer un ou plusieurs scénarii pour en tirer le comportement futur le plus probable conditionné à l'information connue. Pratiquement tous les modèles passés en revue pour simuler sont utilisables en prévision. (MIQUEL 2001).

I-2-2-1 Prévision par la famille des modèles ARMA :

Parmi les méthodes de prévision, les modèles linéaires sont de loin les plus employés, en hydrologie, en particulier par les services d'annonce des crues. Ces modèles appelés ARMA (Auto Regressive Moving Average), développés initialement par Box et al., (1970) et (1976), consistent à mettre en relation linéaire le débit que l'on souhaite prévoir avec des variables

explicatives (débits au pas de temps précédents mesurés à la même station, débit amont, pluie, évaporation) :

$$Q_t = \sum_{j=k}^n a_j Q_{t-j} + \sum_{j=k}^n \sum_{i=1}^m b_{i,j} X_{i,t-j} + \varepsilon_t \quad (I-18)$$

Q : Débit mesuré à la station où l'on souhaite faire les prévisions ;

X_i : Autres variables explicatives retenues ;

k : Horizon de prévision ;

ε : Bruit blanc ;

t : Indice de temps.

Ce modèle est capable de modéliser la relation mathématique pluie-débit sans prendre en considération le processus physique (Hsu, et al., 1995).

Son principal inconvénient, est le nombre élevé de paramètres à estimer.

Les modèles linéaires ARMAX (Auto Regressive Moving Average with eXternal data), qui supposent que les pluies et les débits sont liés par une relation linéaire invariante dans le temps, n'ont pas abouti aux attentes des hydrologues, en effet d'après Shamseldin (1997), ces modèles ne sont pas considérés comme étant fiables et opérationnels, mais seulement pour offrir une assise de comparaison pour les autres modèles plus performants.

Les modèles ARMA et leurs dérivés (ARIMA, PARMA, etc.) sont généralement utilisés pour la prévision de l'occurrence des pluies. Ils ne donnent aucune indication sur les volumes des précipitations. De plus, ils ne peuvent être utilisés qu'à un seul pas de temps, généralement journalier (Overney et al., 1995).

I-2-2-2 Prévision par les modèles Markoviens :

Seidou et al., (2002) ont proposé une méthode pour modéliser l'incertitude sur les apports à court terme en fonction de la période de l'année. Pour cela, un modèle paramétrique markovien à temps continu, basé sur le processus de Poisson filtré, est construit et estimé grâce à des observations régulièrement espacées dans le temps puis validé sur trois stations hydrométriques situées au Québec (Canada). Les points étudiés sont sa capacité à représenter correctement les transitions de débits et ses performances en tant que prédicteur de la valeur à venir du débit. Ce dernier aspect est comparé à celui du modèle ARMA saisonnier.

I-2-2-3 Prévision par les réseaux de neurones :

Ces dernières années une nouvelle génération de modèles est apparue, basée sur le concept des réseaux de neurones humains (Jodouin, 1994). French et al., (1992) l'ont utilisé pour la première fois dans la prévision spatiale et temporelle des précipitations, alors que Zealand et al., (1999), Campolo et al., (1999) l'ont appliqué pour la prévision de l'écoulement des rivières.

Ces modèles ont trouvé une vaste utilisation dans la simulation et la prévision des séries hydrologiques et notamment dans la modélisation pluie-débit (Lek et al., 1996, Hsu et al., 1995, Minns et al., 1996, Sajikumar et al., 1999, Shamseldin, 1997, Coulibaly et al., 1999).

Les réseaux de neurones artificiels constituent une nouvelle méthode d'approximation des systèmes complexes particulièrement utiles lorsque ces systèmes sont difficiles à modéliser à l'aide des méthodes statistiques classiques.

Pour augmenter la fiabilité d'un modèle réseau de neurone il faut choisir la bonne configuration. Le choix de la configuration a une influence sur la qualité de la prévision, en général il est préférable d'utiliser une seule couche cachée et un nombre de neurones dans cette couche du même ordre de grandeur que le nombre de neurones dans la couche d'entrée (Davallo et al., 1993).

I-2-2-4 Prévision par la logique floue :

La régression floue est un autre outil récent issu des théories des sous-ensembles, elle a été étudié d'une manière détaillée par Bardossy en 1990. En effet, ce dernier l'a utilisée, pour quantifier la prévision de la demande en eau en tenant compte des prévisions météorologiques, et pour prévoir la température moyenne journalière.

La régression linéaire floue repose sur l'hypothèse d'un modèle linéaire dont les coefficients sont flous :

$$Y = A_0 + \sum_{k=1}^K A_k (x_k - \bar{x}_k) \quad (\text{I-19})$$

A_k , ($k = 1, 2, \dots, K$) : Est le nombre flou ;

$\bar{x} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$: Est un point de l'espace choisi, pour lequel on estime que le modèle est plus précis.

Au cours de ces dix dernières années, les systèmes à base de règles floues se sont avérés des outils souples et puissants pour la modélisation de l'incertain dans un très grand nombre de domaines, et surtout dans la prévision des séries temporelles (Fiordaliso, 1999).

L'estimation des paramètres d'une régression linéaire floue est un problème de programmation linéaire. Les avantages de ce type de modèle proviennent du fait qu'il est basé sur des hypothèses moins restrictives que celles de la régression linéaire classique.

La régression linéaire floue est cependant très sensible aux valeurs extrêmes, qui doivent être validées avec précaution. Un autre inconvénient de cette méthode est la complexité de son algorithme d'estimation des paramètres ; pour résoudre 2^K systèmes d'inéquations à $2(K+1)$ inconnues et $2n$ contraintes, où n est le nombre d'observations, la méthode devient impraticable lorsque K et n sont grands mais avantageuse quand ces derniers sont petits.

Malgré ses avantages la méthode et ses fondements de base ont été remis en question par plusieurs scientifiques.

I-3 CONCLUSION :

Pour conclure, nous relevons, que la plupart des modèles mathématiques passés en revue pour simuler ou prévoir des variables climatologiques sont difficilement exploitables en raison de leur complexité et des conditions spécifiques de leur application (chronologie des données, stationnarité du phénomène étudié, normalité de l'échantillon, dépendance entre les valeurs

observées, consistance des données, type de variable, pas de temps spécifique, etc....).

En outre, ces modèles montrent que l'on peut affiner à volonté la représentation schématique des phénomènes naturels. Malheureusement, nous constatons que toute 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. Ainsi, l'utilisation des modèles hydrologiques simples est privilégiée par les hydrologues, car elle fait intervenir le minimum d'hypothèses et de paramètres, et c'est la tâche à laquelle nous nous sommes attelés dans cette thèse pour élaborer un modèle qui s'affranchit de toutes les contraintes exigées par les autres modèles, et surtout facilement exploitable.

Chapitre II : Collecte et Critique des Données

II-1 INTRODUCTION :

L'étude et le contrôle de fiabilité des données utilisées par les modèles de simulation et de prévision élaborés dans cette thèse constitue le principal objectif de ce chapitre. En effet le seul outil dont dispose l'hydrologue pour étudier et comprendre le comportement d'un phénomène naturel c'est les données observées, il est donc impératif de connaître à fond les techniques nécessaires d'analyse et de critiques de ces données afin de tirer le maximum d'informations de l'échantillon et de garantir leur qualité, de ce fait, il est indispensable d'effectuer une analyse et un traitement statistique des séries chronologiques recueillies aux niveaux des stations de mesure.

La démarche à suivre consiste, à combler les lacunes des fichiers recueillis, à détecter d'éventuelles anomalies des observations, et puis à vérifier l'homogénéité et l'indépendance de ces dernières.

II-2 INVENTAIRE DES DONNEES :

Les données utilisées nous ont été aimablement fournies par l'Office National de la Météorologie (O.N.M). La station retenue dans le cadre de cette étude est celle d'Alger (Dar El Beida) toujours fonctionnelle, et qui a été installée par l'administration coloniale en 1935.

Nous disposons de relevés complets qui s'étale :

- Pour la pluie (mm) :
1936 – 1995 (60 ans) au pas de temps journalier.
1936 – 2001 (66 ans) au pas de temps mensuel.
- Pour la température moyenne (°C) :
1958 – 1991 (34 ans) au pas de temps journalier.
- Pour l'humidité relative (%) :
1958 – 1991 (34 ans) au pas de temps journalier.
- Pour l'insolation (h et 1/10h) :
1958 – 1991 (34 ans) au pas de temps journalier.
- Pour la vitesse du vent (m/s) :
1958 – 1991 (34 ans) au pas de temps journalier.

Les caractéristiques de la station d'Alger sont présentées dans le tableau II-1.

Tableau II-1 : Caractéristiques de la station d'Alger.

Station	Numéro	Code	Lat.(Nord)	Long.(E/W)	Alt.(mètre)
Alger	6	390	36° 43	03°15E	24

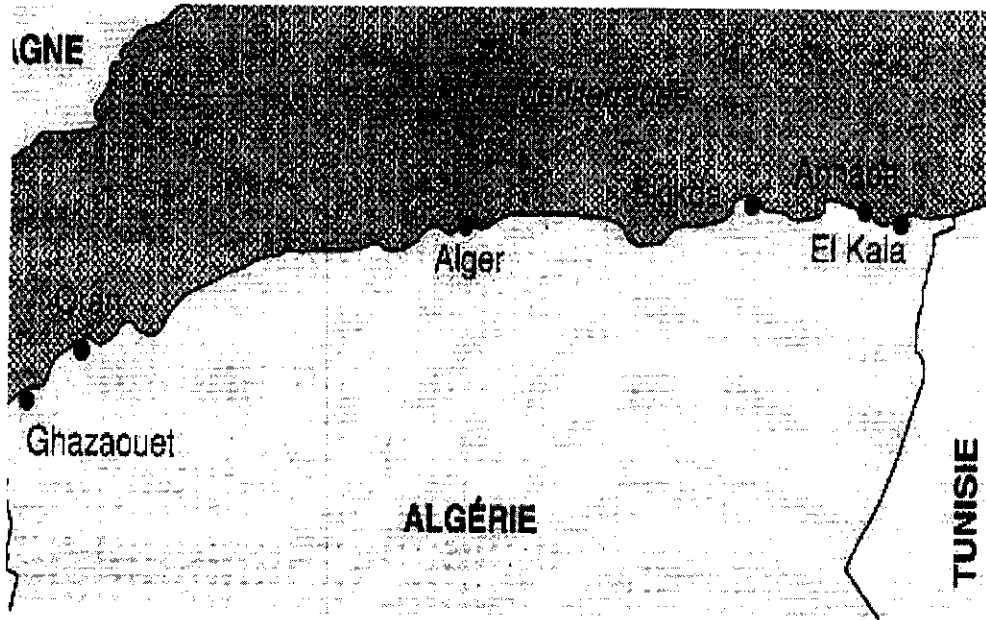


Figure II-1 Représentation de la station d'Alger.

Les paramètres statistiques des pluies annuelles sont donnés dans le tableau II-2.

Tableau II-2 : Caractéristiques statistiques de la pluviométrie à la station d'Alger.

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

L'analyse de la figure II-2 montre qu'il y a autant d'années déficitaires que d'années pluvieuses pour la station d'Alger.

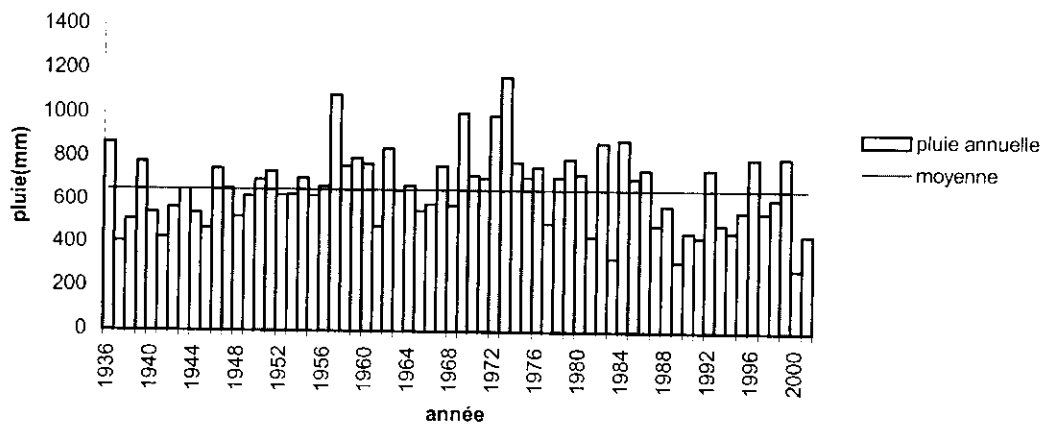


Figure II-2 Variation de la pluie annuelle par rapport à la moyenne.

II-3 CRITIQUE DES DONNEES :

Les types de problèmes le plus souvent rencontrés lors du traitement des données en hydrologie sont (Llamas, 1985) :

- Le manque de données : station non existante,

- Non-continuité des données : manque d'enregistrements à une station,
- La mauvaise qualité des données, qui peut dépendre de deux causes :
 - Données inconsistantes : lorsqu'elles sont entachées d'une erreur systématique due à des instruments de mauvaise qualité ou non calibrés.
 - Données non homogènes : lorsqu'il se produit à un moment donné, un changement de l'emplacement de l'appareil qui affecte la mesure (déplacement de la station de mesure ou changement d'appareil).

Afin de garantir la qualité des données, nous devons suivre les étapes suivantes :

- Comblent les lacunes des séries observées si elles existent.
- Détecter et corriger les valeurs douteuses.
- Vérifier l'homogénéité et l'indépendance des données.

II-3-1 Méthodes de comblement des lacunes :

Dans la pratique courante, de nombreuses causes concourent pour entraîner l'absence de certaines informations (absence de mesures, défaut de transmission, erreur de stockage, manipulation erronée, ...), ce qui fait que dans les fichiers disponibles certaines données sont manquantes.

Nous pouvons estimer les valeurs manquantes ou erronées à une station à partir des valeurs provenant des stations voisines soumises aux mêmes conditions climatiques et situées dans la même zone géographique. Plusieurs méthodes sont proposées pour cela, (El jabi et al., 1987) :

- Remplacer la valeur manquante par celle de la station la plus proche.
- Remplacer la valeur manquante par la moyenne des stations voisines. Cette méthode est utilisée lorsque les valeurs moyennes annuelles de la station à compléter ne diffèrent pas plus de 10 % des valeurs moyennes annuelles des stations de référence.
- Remplacer la valeur manquante par une moyenne pondérée par la tendance annuelle des stations pluviométriques par exemple, soit :

$$P_x = \frac{1}{n} \sum_{i=1}^n \left[\frac{\overline{P}_i}{\overline{P}} P \right] \quad (\text{II-1})$$

P_x : Donnée manquante de précipitation (par exemple) à estimer ;

n : Nombre de stations de référence ;

P : Précipitation à la station de référence ;

\overline{P}_x : Précipitation moyenne de la station ;

\overline{P} : Précipitation moyenne de la station de référence.

- Remplacer la valeur manquante par la moyenne inter-annuelle de la série considérée :

Une donnée journalière manquante, par exemple, 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 inter-annuelle du jour considéré. S'il s'agit d'une valeur mensuelle, nous

estimons la valeur manquante soit par la moyenne mensuelle des mois disponibles ou bien par la moyenne mensuelle inter-annuelle du mois considéré.

- Les méthodes basées sur l'analyse de régressions et corrélations :

Le calcul de régression se fait en déterminant les valeurs des paramètres de la relation entre la variable expliquée et la (ou les) variable(s) explicative(s). Cette régression peut être simple ou multiple.

Exemple : Soient Y une variable climatologique dont nous connaissons n valeurs : y_1, y_2, \dots, y_n , et X une autre variable observée de façon concomitante, avec les observations x_1, x_2, \dots, x_n , mais nous possédons par ailleurs sur une période d'observation antérieure par exemple N-n réalisations complémentaires $x_{n+1}, x_{n+2}, \dots, x_N$.

Il s'agit de reconstituer les valeurs manquantes $y_{n+1}, y_{n+2}, \dots, y_N$, concomitantes aux valeurs de la série complémentaire $x_{n+1}, x_{n+2}, \dots, x_N$ comme illustré par la figure ci-dessous :

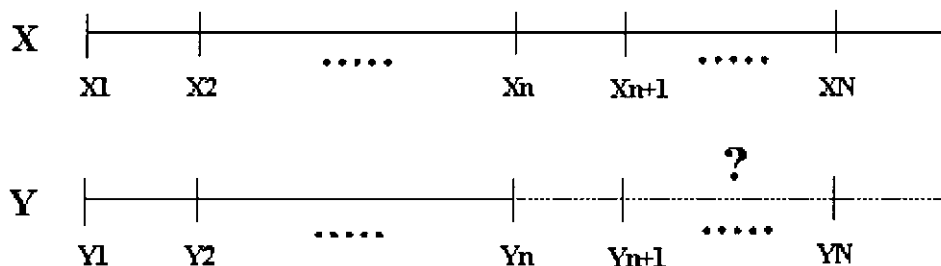


Figure II- 3 Reconstitution des séries manquantes.

Une méthode générale consiste à utiliser un modèle linéaire de régression du type :

$$y = a + bx + \varepsilon \quad (\text{II-2})$$

Nous estimons les paramètres a et b sur la série de couples connus.

Nous pouvons alors estimer les valeurs complémentaires y_i pour chaque x_i connu ($i = n+1, \dots, N$), à partir de la relation estimée :

$$y_i = \hat{a} + \hat{b}x_i \quad (\text{II-3})$$

- Méthode basée sur la simulation : (voir Chap.VI-3-1).

Pour étudier d'éventuelles anomalies au niveau des séries climatologiques prises en considération, nous avons élaboré un logiciel en fortran 90 (CRIDON) permettant de vérifier les valeurs douteuses et d'étudier l'homogénéité et l'indépendance des données collectées.

L'examen des fichiers fournis par l'O.N.M. pour les différents paramètres ne présentent pas de valeurs manquantes. Ceux-ci peuvent être considérés comme étant fiables car ils font l'objet de contrôles réguliers de la part des services gestionnaires de l'O.N.M. vue leur importance (communication de bulletins météorologiques pour le ministère des transports).

II-3-2 Détection des valeurs singulières :

Il arrive fréquemment que des erreurs se glissent dans les données de base et en dénaturent complètement le contenu. Cela peut se produire au cours de la mesure, de la transmission, du codage, du stockage ou du décodage des observations, ou à cause des valeurs intégrées après le comblement des lacunes.

Il est important donc avant tout traitement de détecter les erreurs grossières de ce genre, 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 :**

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

$$X_h = \exp(\bar{x} + K_n * s_x) \quad (\text{II-4})$$

$$X_l = \exp(\bar{x} - K_n * s_x) \quad (\text{II-5})$$

\bar{x} et s_x : Sont respectivement, la moyenne et l'écart type des logarithmes naturels des éléments constituant l'échantillon.

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

Au niveau de risque de 10%, l'approximation polynomiale suivante a été proposée par Bobée (1991) :

$$K_n = 3.62201 + 6.28446 * n^{1/4} - 2.49835 * n^{1/2} + 0.491436 * n^{3/4} - 0.037911 * n \quad (\text{II-6})$$

n : Taille de l'échantillon.

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

Le tableau II-3 récapitule les résultats obtenus après l'application du test de Grubbs et Beck.

Tableau II-3 Résultats du test de Grubbs et Beck

	n	\bar{x}	s_x	K_n	X_l	X_h
Plj (mm)	21915	0.26	0.90	-235.42	0	1.4E+92
Temp.(°C)	12784	2.80	0.34	-105.85	0	1.2E+17
Hum.(%)	12784	4.28	0.14	-113.09	0	8.9E+08
Inso.(h+0.1h)	12784	1.79	0.90	-113.09	0	1.0E+50
Vv (m/s)	12784	0.98	0.58	-113.09	0	1.7E+29
Pl.men(mm)	792	3.08	1.77	2.75	0.17	2800.8

Plj : Pluie journalière à la station d'Alger ;
 Temp : Température journalière à la station d'Alger ;
 Hum : Humidité relative journalière à la station d'Alger ;
 Inso : Insolation journalière à la station d'Alger ;
 Vv : Vitesse du vent journalière à la station d'Alger ;
 Pl.men : Pluie mensuelle à la station d'Alger ;

L'application de ce test au pas du temps journalier et mensuel a détecté les valeurs nulles comme étant singulières, mais nous ne pouvons rejeter ces valeurs car elles correspondent aux jours ou aux mois secs.

II-4 TRAITEMENT STATISTIQUE DES DONNEES :

Après avoir analysée et critiquée les données recueillies, il est indispensable de vérifier leur homogénéité et leur indépendance.

II-4-1 Test d'homogénéité :

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, nous avons utilisé dans ce cas celui de Wilcoxon-Mann-Withney.

Il permet de tester si une série statistique est homogène, c'est à dire si les éléments qui la constituent proviennent de la même population. En hydrologie, cela veut dire que les conditions qui ont prévalu lors de la collecte des données, ou de l'avènement du phénomène considéré (pluie, évaporation, etc.) n'ont pas changé pendant toute la durée de la collecte, ou en d'autres termes qu'il n'y a pas eu un phénomène extraordinaire qui aurait pu modifier les données hydrologiques considérées (changement de site de la station de mesure, construction d'un barrage à l'amont, urbanisation etc...).

Pour appliquer ce test nous avons procédé comme suit : l'échantillon est subdivisé en deux sous-ensembles de tailles respectives p et q avec $p \geq q$.

$$\begin{array}{l}
 x_1, x_2, \dots, x_i, \dots, x_q \\
 y_1, y_2, \dots, y_i, \dots, y_p
 \end{array}$$

La taille de l'échantillon original est $n=p+q$.

Ensuite nous classons les valeurs par ordre croissant de 1 à n et nous notons les rangs $R(x_i)$ des éléments du premier sous-ensemble et $R(y_i)$ des éléments du second sous-ensemble dans l'échantillon original.

Nous définissons deux statistiques K et F comme suit :

$$K = L - \frac{q(q+1)}{2} \text{ et } F = pq - K \quad (\text{II-7})$$

$$L = \sum_{i=1}^q R(x_i) \quad (\text{II-8})$$

L : Somme des rangs des éléments de l'échantillon 1 dans l'échantillon original.

K : Somme des nombres de dépassements de chaque élément du second échantillon par ceux du premier.

F : Somme des nombres de dépassements des éléments du premier échantillon par ceux du second.

Lorsque n, p et q son suffisamment grand, K et F suivent une distribution normale :

- de moyenne $\bar{K} = \bar{F} = \frac{pq}{2}$,
- et d'écart type $s_K = s_F = \left(\frac{pq(p+q+1)}{12} \right)^{1/2}$

Et nous pouvons alors tester si les deux échantillons proviennent de la même population au niveau de signification α en comparant la grandeur statistique :

$$T = \left| \frac{K - \bar{K}}{s_K} \right| \quad (\text{II-9})$$

Avec la variable normale centrée réduite U de probabilité de dépassement $\alpha/2$.

Si $T \leq U_{\alpha/2}$ alors les deux échantillons proviennent de la même population, dans le cas contraire l'échantillon étudié est considéré comme non homogène ($T \geq U_{\alpha/2}$).

Le tableau II-4 résume les résultats de ce test.

Tableau II-4 Résultats du test de Wilcoxon-Mann-Withney.

	Plj.(mm)	Temp.(°C)	Hum.(%)	Inso.(h+0.1h)	Vv.(m/s)	Pl.men(mm)
T	.3153	.0071	.0943	.0409	.0251	.0003
$U_{\alpha/2} (\alpha=5\%)$	1.96					

Les résultats de ce test, montrent que les observations de tous les paramètres étudiés de la station d'Alger à différents pas de temps sont homogènes.

II-4-2 Test d'indépendance :

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'autres termes, l'échantillon $X=(x_1, \dots, x_j, \dots, x_n)$ est dit « au hasard », si le test permet de vérifier l'hypothèse H_0 d'indépendance des observations.

Dans cette étude, nous avons appliqué le test de Wald-Wolfowitz à toutes les séries recueillies sur des pas de temps journaliers et mensuels.

A la suite des travaux d'Anderson (Dagnelie, 1998), Wald et Wolfowitz ont développé un test non paramétrique du coefficient d'autocorrélation, celui-ci est basé sur le critère qu'aucune observation appartenant à l'échantillon ne devrait influencer l'observation suivante.

La statistique du test est calculée comme suit :

$$R = \sum_{i=1}^{n-1} x_i x_{i+1} + x_n x_1 \tag{II-10}$$

Pour n suffisamment grand, cette statistique suit une distribution normale de moyenne et de variance :

$$E[R] = \frac{(S_1^2 - S_2)}{n-1} \tag{II-11}$$

$$Var[R] = \frac{(S_2^2 - S_4)}{n-1} + \frac{(S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4)}{(n-1)(n-2)} - \{E[R]\}^2 \tag{II-12}$$

$$S_k = \sum_{i=1}^n x_i^k \tag{II-13}$$

S_k : Moment d'ordre k de l'échantillon.

La valeur statistique $T = \frac{R - \bar{R}}{\sqrt{Var(R)}}$ suit une distribution normale centrée réduite $N(0,1)$, et est utilisée pour tester à un niveau de signification, l'hypothèse d'indépendance en comparant la valeur absolue de T avec la valeur normale centrée réduite $U_{\alpha/2}$ correspondant à la probabilité de dépassement de $\alpha/2$.

Si $|T|$ est inférieure à $U_{\alpha/2}$ l'hypothèse d'indépendance est vérifiée, celle-ci est rejetée dans le cas contraire.

Le tableau II-5 présente les résultats de ce test.

Tableau II-5 Résultats du test de Wald-Wolfowitz

	Plj.(mm)	Temp.(°C)	Hum.(%)	Inso.(h+0.1h)	Vv.(m/s)	Pl.men(mm)
T	49.05	156.45	155.75	133.72	125.48	97.40
$U_{\alpha/2} (\alpha=5\%)$	1.96					

Les résultats illustrés dans le tableau II-5, montrent que les observations de tous les paramètres de la station d'Alger à différents pas de temps ne sont pas indépendantes.

II-5 MODELE DE CALCUL DE L'EVAPOTRANSPIRATION POTENTIELLE (ETP) :

L'approche élaborée dans cette thèse est basée sur l'étude de variables aléatoire (la pluviométrie) et cyclique (l'EvapoTranspiration Potentielle - ETP), le choix de l'ETP a été

dicté par le fait que cette variable climatologique est prépondérante dans la détermination des besoins en eau des plantes.

L'évapotranspiration potentielle (ETP) est théoriquement la quantité d'eau que cède une culture en pleine croissance couvrant totalement un sol largement pourvu d'eau par transpiration des plantes et évaporation du sol. Elle dépend donc des conditions climatiques.

L'ETP est aussi une mesure climatique, dont la valeur est établie à partir de relevés météorologiques selon des formules plus ou moins complexes impliquant des variables météorologiques mesurables. Il est à noter que l'ETP correspond à une valeur moyenne tous couverts végétaux confondus.

La formule de Penman qui intègre quatre facteurs climatiques (températures, insolation, humidité relative et vitesse du vent), approche le mieux l'évapotranspiration potentielle du fait qu'elle prend en compte les différents paramètres intervenants dans ce phénomène.

La méthode de calcul de l'ETP utilisée dans cette thèse est celle proposée par Brochet et al., (1972, 1974).

Penman (1956) a estimé le flux de chaleur latente au niveau du sol et donc le flux d'eau évaporé à partir d'un simple bilan énergétique, l'expression mathématique de cette formule s'écrit :

$$R = \frac{R_n - A - S}{L} \quad (\text{II-14})$$

R_n : Radiation nette reçue au sol ;
 A : Flux de chaleur au bénéfice de l'atmosphère ;
 S : Flux de chaleur sensible au bénéfice du sol ;
 R : Flux d'eau évaporé ;
 L : Chaleur latente.

La formule développée prend la forme suivante :

$$ETP = \frac{1}{1 + \frac{F_T}{\gamma}} \left[\frac{I_g a \cdot F_T}{59 \cdot \gamma} (1 - a) \left(0.18 + 0.62 \frac{h}{H} \right) - \frac{\sigma \cdot F_T}{59 \cdot \gamma} \cdot T^4 \left(0.56 - 0.08 \sqrt{e} \right) \left(0.10 + 0.9 \frac{h}{H} \right) \right] + \frac{0.26}{1 + \frac{F_T}{\gamma}} (e_w - e) (1 + 0.54 \cdot V) \quad (\text{II-15})$$

ETP : Evapotranspiration potentielle en mm/j ;
 $I_g a$: Radiation directe solaire en l'absence d'atmosphère ($\text{cal/cm}^2/\text{j}$) ;
 a : Albédo de la surface évaporante (0.2 pour la végétation) ;
 h : Durée réelle d'insolation en heures du jour considéré ;
 H : Durée maximale possible d'insolation pour ce jour en heures ;
 σ : Constante de Boltzman ($1.18 \cdot 10^7 \text{ cal/cm}^2/\text{j/K}$) ;
 T : Température moyenne sous abri en degré Kelvin ;
 e : Tension moyenne de vapeur d'eau de l'air en millibars ;
 F_T : Pente de la courbe de tension de vapeur saturante à la température T ;
 γ : Constante psychométrique ;

e_w : Tension de vapeur saturante à la température T ;
 V : Vitesse moyenne journalière du vent à deux mètres du sol en m/s.

Pour le calcul de cette ETP Penman nous avons élaboré un programme informatique en Fortran 90 (PENMAN) qui nous a grandement facilité la tâche.

Du fait que l'ETP Penman est issue des séries de valeurs journalières des 34 années des fichiers de paramètres climatologiques déjà critiqués (température, insolation, vitesse du vent et humidité relative), nous nous sommes dispensés à appliquer la procédure de critique des données à cette série chronologique calculée.

Chapitre III : Théories de l'Analyse en Composantes Principales et des Chaînes de Markov

III-1 ANALYSE EN COMPOSANTES PRINCIPALES (ACP) :

III-1-1 Introduction :

L'analyse des données a vu au cours de ces dernières années son domaine d'application s'étendre considérablement, et une effervescence bien légitime règne dans le champs de recherche qu'elle a couvert.

Les premières analyses à plusieurs variables remontent au tout début du siècle dernier (Pearson 1901, Spearman 1904, Fisher 1915). Les bases théoriques de presque toutes celles utilisées actuellement ont été établies, de façon quasi définitive vers 1930 (Hotelling 1933 et 1936, Mahalanobis 1936) avec des compléments importants après 1950 (Burt 1950, Anderson 1958, Rao 1964).

Cependant faute de moyens de calcul suffisants, ces méthodes n'ont pu être utilisées que de façon extrêmement limitée.

Jusqu'aux années soixante, elles étaient perfectionnées et s'enrichissaient de variantes, mais toutes restaient inabordables pour les praticiens, car elles nécessitaient une masse considérable de calcul (Benzecri 1982).

Elles ne sont devenues d'une application courante, que grâce au développement de l'informatique. Dès leur apparition en effet, les ordinateurs ont permis l'existence et l'exploitation de grands fichiers de données ainsi que la résolution de problèmes numériques devant lesquels l'homme seul est impuissant.

Ces analyses ont fréquemment le caractère d'une démarche exploratoire du statisticien et du chercheur. Il s'agit le plus souvent d'obtenir une description statistique pour un certain phénomène qui a donné lieu au recueil de mesures ou d'observations trop nombreuses et dépendantes les unes des autres pour être interprétables en première lecture. La simplification de grands tableaux d'informations ou de données s'avère un outil de synthèse remarquable (Jambu, 2000).

Enfin, cette analyse est devenue une technique de représentation des données, ayant un caractère optimal selon certains critères algébriques et géométriques spécifiés, que l'on utilise en général sans référence à des hypothèses statistiques ou à un modèle particulier (Delagarde 1983).

La plus simple et la plus connue des techniques d'analyse de données, c'est l'Analyse en Composantes Principales (ACP), qui a pour objectif de remplacer les variables initiales (numériques) par de nouvelles variables (également numériques) dites "Composantes Principales".

Dans ce chapitre nous nous intéressons essentiellement aux théories de l'analyse en composantes principales et les chaînes de Markov qui sont à la base des modèles de simulation et de prévision élaborés.

III-1-2 Définitions :

L'étude des phénomènes climatologiques est une tâche très laborieuse vue la masse importante de données à traiter, la démarche basée sur l'analyse en composantes principales (ACP), qui est une technique d'analyse des données, permet d'extraire l'essentiel de l'information contenue dans les tableaux des données, et fournit une représentation imagée se prêtant plus aisément à l'interprétation. Celle-ci a trois objectifs principaux :

- Le premier consiste à remplacer les variables initiales qui sont généralement plus ou moins liées entre elles, par des variables non corrélées, appelées Composantes Principales, (CP).
- Le second revient à réduire autant que possible le nombre des nouvelles variables (CP) à prendre en considération, tout en conservant la quasi totalité de la variance de l'ensemble des variables initiales (Duband 1989).
- Le troisième est de donner une interprétation particulière aux CP les plus importantes parmi les variables initiales.

III-1-2-1 Notions d'individu et de caractère :

Dans une série de données on distingue généralement deux ensembles : les individus et les caractères relatifs à ces individus. Le terme "individu" peut désigner selon les cas : une année d'observation ou une autre unité de temps. L'ensemble des individus peut provenir d'un échantillonnage dans une population ou il peut s'agir de la population toute entière (cas rare surtout en hydrologie).

- L'individu "i" est décrit par le vecteur appartenant à R^P :

$$X_i = \{X_{ij} / j = 1 \text{ à } P\}$$

Sur un individu, on relève un certain nombre de caractères (dits aussi variables) désignant en général un paramètre intervenant dans un phénomène complexe à étudier (exemple, la répartition spatiale ou temporelle des précipitations dans un bassin versant).

- Le caractère (ou variable) "j" est décrit par le vecteur de R^N :

$$X_j = \{X_{ij} / i = 1 \text{ à } N\}$$

Le terme X_{ij} est un nombre réel qui représente la mesure de la variable X_j sur l'individu i .

Si l'ensemble des individus doit être homogène, l'ensemble des variables peut être hétérogène.

III-1-2-2 Définition algébrique :

L'ACP considère P variables pour lesquelles on dispose de N individus, donc la matrice $[X]$ résultant du croisement « $N \times P$ » est la matrice de données $[X]$ tel que :

$$[X] = \begin{pmatrix} X_{11} & X_{12} & \dots & \dots & X_{1j} & \dots & \dots & X_{1P} \\ X_{21} & X_{22} & \dots & \dots & X_{2j} & \dots & \dots & X_{2P} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & \dots & X_{ij} & \dots & \dots & X_{iP} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{M1} & X_{M2} & \dots & \dots & X_{Mj} & \dots & \dots & X_{MP} \end{pmatrix}$$

Pour chaque variable, nous pouvons évaluer sa moyenne \bar{X}_j et son écart type S_j :

$$\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij} \tag{III-1}$$

$$S_j = \left[\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j)^2 \right]^{1/2} \tag{III-2}$$

Le coefficient de covariance entre les variables X_j et X_k est donné par :

$$Cov(X_j, X_k) = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k) \tag{III-3}$$

Le coefficient de corrélation entre les variables X_j et X_k est donné par :

$$Cor(X_j, X_k) = \frac{Cov(X_j, X_k)}{S_j * S_k} = \frac{\sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\left[\sum_{i=1}^N (X_{ij} - \bar{X}_j)^2 * \sum_{i=1}^N (X_{ik} - \bar{X}_k)^2 \right]^{1/2}} \tag{III-4}$$

Le tableau [X] est remplacé par un tableau [Y] (individus x nouvelles variables) en réduisant le nombre de variables nécessaires pour décrire les individus, avec une perte minimale d'informations, ces nouvelles variables sont appelées composantes principales (ou CP).

Calculer les composantes principales notées C_j revient à déterminer N relations linéaires entre les variables X_j :

$$\begin{cases} C_1 = a_{10} + a_{11} X_1 + \dots + a_{1j} X_j + \dots + a_{1N} X_N \\ C_2 = a_{20} + a_{21} X_1 + \dots + a_{2j} X_j + \dots + a_{2N} X_N \\ \dots \dots \dots \\ C_j = a_{j0} + a_{j1} X_1 + \dots + a_{jj} X_j + \dots + a_{jN} X_N \\ \dots \dots \dots \\ C_N = a_{N0} + a_{N1} X_1 + \dots + a_{Nj} X_j + \dots + a_{NN} X_N \end{cases}$$

C_j : j^{ème} Composante Principale ;

X_j : Vecteur variable initiale ;

a_{ji} : Coefficient du système.

Notons au passage que les termes a_{j0} désignent le vecteur permettant la translation de l'origine de l'ancien repère vers le centre de gravité du nuage de points. Un centrage des données initiales annule les coefficients a_{j0} .

III-1-2-3 Définition géométrique :

L'Analyse en Composantes Principales est puissante par son support géométrique, la méthode consiste à rechercher un premier axe qui soit le plus proche possible de tous les points au sens des moindres carrés, tel que la somme des carrés des distances des N points à cet axe serait minimale, ou encore la projection de ces derniers sur cet axe possède une dispersion maximale. Cet axe est appelé "axe factoriel".

Un second axe est obtenu après projection des N points sur un hyperplan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci serait toujours maximale, et le procédé se réitère P fois.

Nous obtenons ainsi un nouveau système d'axes défini par les nouvelles variables dites composantes principales.

III-1-3 Formulation mathématique du problème :

La recherche des composantes principales est faite sous deux contraintes :

- Elles doivent être indépendantes, c'est à dire, prise deux à deux, elles présentent obligatoirement des corrélations nulles.
- Les axes factoriels doivent être déterminés par ordre d'importance décroissante, le premier axe expliquera le maximum de la variance totale tandis que le second expliquera le maximum de la variance résiduelle non expliquée par le premier, jusqu'au dernier axe. Mais l'expérience a montré qu'un nombre Q d'axes nettement inférieur à P suffit pour donner le maximum d'informations.

a) Formulation matricielle :

L'objectif de l'ACP étant de maximiser la variance, la formulation matricielle du problème est la suivante :

Soit les matrices colonnes V_1, V_2, \dots, V_q de dimension $(N \times 1)$ représentant l'hyperplan formé par les axes principaux vérifiant les conditions de normalité et d'orthogonalité :

$$\begin{cases} \vec{V}_i \cdot \vec{V}_i = 1 \\ \vec{V}_i \cdot \vec{V}_j = 0 \end{cases} \rightarrow \begin{cases} i = 1 \rightarrow Q \\ j = 1 \rightarrow Q \end{cases} \quad (\text{III-5})$$

On veut maximiser la quantité :

$$\sum_{j=1}^Q \text{Var}(C_j)$$

$$\text{Sachant que : } \sum_{j=1}^Q \text{Var}(C_j) = V_j' [M] [R] [M] V_j \quad (\text{III-6})$$

C_j : Composante principale d'ordre j ;

$[R]$: Matrice des covariances des variables (X_1, X_2, \dots, X_p) ;

$[M]$: Métrique définissant le produit scalaire sur l'espace \mathbb{R}^p .

b) Choix de la métrique :

La métrique $[M]$ possède deux options classiques :

* $[M]=1$: Matrice identité.

La covariance sera utilisée afin de quantifier les relations inter variables, on parlera alors d'une *ACP canonique*.

* $[M]=D_{\sigma^2}$

Tel que :

$$D_{\sigma^2} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & & & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & & \\ & 0 & \cdot & & \\ & & 0 & \frac{1}{\sigma_i^2} & 0 \\ & & & \cdot & \\ & & & \cdot & 0 \\ 0 & & & 0 & \frac{1}{\sigma_p^2} \end{pmatrix}$$

On utilise généralement cette métrique pour pallier au problème de l'hétérogénéité des caractères (variables), et éviter l'influence du choix d'unité des variables. Dans ce cas on parlera d'ACP normée, elle est équivalente à une ACP canonique effectuée sur des variables centrées réduites.

Les données ainsi transformées se présentent sous forme d'une matrice dont toutes les variables sont de moyenne nulle et d'écart type unité.

III-1-4 Procédé d'application de l'ACP :

III-1-4-1 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 :

$$[R] = (1/n)[X]'. [M]. [X] \quad (\text{III-7})$$

$[R]$: Matrice de covariance de dimension $(p \times p)$;

$[X]$: Matrice de données ;

$[X]'$: Matrice transposée de $[X]$;

$[M]$: Métrique.

III-1-4-2 Recherche des axes principaux :

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

a) Recherche du premier axe :

Telle que mentionnée précédemment, la contribution maximale est donnée par le premier axe principal, nous devons donc maximiser la variance relative à celui-ci. La recherche du premier axe principal consiste à résoudre le problème

$$\begin{cases} \text{MaxVar}(C_1) \\ V_1' \cdot [M] \cdot V_1 = 1 \end{cases} \quad (\text{III-8})$$

Nous pouvons exprimer la variance de C_1 à l'aide de la matrice des covariances $[R]$ du vecteur aléatoire $X = (X_1, X_2, \dots, X_j, \dots, X_p)$:

$$\text{Var}(C_1) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 \quad (\text{III-9})$$

En utilisant la méthode du multiplicateur de LAGRANGE nous pouvons écrire :

$$L = V_1' [M] \cdot [R] \cdot [M] \cdot V_1 - \lambda_1 (V_1' \cdot [M] \cdot V_1 - 1) \quad (\text{III-10})$$

La dérivée par rapport à V est nécessairement nulle :

$$\partial L / \partial V_1 = 2 \cdot [M] \cdot [R] \cdot [M] \cdot V_1 - 2\lambda_1 \cdot [M] \cdot V_1 = 0 \quad (\text{III-11})$$

Puisque la matrice $[M]$ est inversible :

$$[R] \cdot [M] \cdot V_1 = \lambda_1 \cdot V_1 \quad (\text{III-12})$$

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 .

b) Recherche du second axe :

Nous cherchons à déterminer le vecteur unitaire V_2 tel que la composante C_2 soit de variance maximale et non corrélée à C_1 .

Sachant que :

$$\text{Var}(C_2) = V_2' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III-13})$$

$$V_2' \cdot [M] \cdot V_2 = 1 \quad (\text{III-14})$$

$$COV(C_1, C_2) = 0 \quad (\text{III-15})$$

L'expression de $COV(C_1, C_2)$ est donnée par :

$$COV(C_1, C_2) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III-16})$$

Comme la covariance ne tient pas compte de l'ordre nous pouvons écrire :

$$COV(C_1, C_2) = COV(C_2, C_1) = V_2' [M] [R] [M] V_1 = 0 \quad (\text{III-17})$$

Or nous savons que V_1 est un vecteur propre de $[R] \cdot [M]$ associé à la valeur propre λ_1 . Nous en déduisons que :

$$COV(C_1, C_2) = \lambda_1 V_2' \cdot V_1 = 0 \quad (\text{III-18})$$

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

$$COV(C_1, C_2) = 0 \Leftrightarrow V_1 \cdot V_2 = 0 \quad (\text{III-19})$$

En appliquant la même méthode pour la recherche du deuxième axe, nous aurons :

$$L = V_2' [M] [R] [M] V_2 - \lambda_2 (V_2' [M] V_2 - 1) - \mu_2 (V_2' [M] V_1) \quad (\text{III-20})$$

$$\partial L / \partial V_2 = 2 [M] [R] [M] V_2 - 2 \lambda_2 [M] V_2 - \mu_2 [M] V_1 = 0 \quad (\text{III-21})$$

En simplifiant par $[M]$ nous obtenons :

$$2 [R] [M] V_2 - 2 \lambda_2 V_2 - \mu_2 V_1 = 0 \quad (\text{III-22})$$

Nous multiplions à gauche par $V_1' \cdot [M]$ l'équation (III-22) s'écrit alors :

$$2 V_1' [M] [R] [M] V_2 - 2 \lambda_2 V_1' [M] V_2 - \mu_2 V_1' [M] V_1 = 0 \quad (\text{III-23})$$

$$\text{Or (par hypothèse)} \quad V_1' \cdot [M] \cdot V_2 = 0 \quad (\text{III-24})$$

Donc :

$$V_1' \cdot [M] \cdot [R] \cdot V_2 = V_2' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 = \lambda_1 \cdot V_2' \cdot [M] \cdot V_1 = 0 \quad (\text{III-25})$$

Puisque le vecteur V_1 est unitaire:

$$V_1' \cdot [M] V_1 = 1 \quad (\text{III-26})$$

Le multiplicateur de Lagrange μ_2 est donc nul, et nous sommes ramené au problème précédent. Nous pouvons donc énoncer la définition suivante : le second axe est défini par le

vecteur V_2 , vecteur propre unitaire de la matrice $[R] \cdot [M]$ orthogonal à V_1 et associé à la plus grande valeur propre λ_2 inférieure ou égale à λ_1 .

c) Recherche des autres axes :

En itérant le procédé, nous déterminons donc les valeurs propres et les vecteurs propres de la matrice $[R] \cdot [M]$ pour obtenir la $i^{\text{ème}}$ composante principale C_i .

Le vecteur propre unitaire de la matrice $[R] \cdot [M]$ définit le $i^{\text{ème}}$ axe orthogonal à $(V_1, V_2, \dots, V_{i-1})$ et associé à la $i^{\text{ème}}$ plus grande valeur propre λ_i . Nous constatons que la mise en équation de ces règles aboutit aux résultats suivants :

- Nous appelons $i^{\text{ème}}$ vecteur principal : le vecteur propre unitaire V_i de la matrice $[R] \cdot [M]$ associée, qui fournit les coefficients qui pondèrent les variables initiales pour le calcul des composantes principales.
- Nous appelons $i^{\text{ème}}$ axe principal, la droite engendrée par le $i^{\text{ème}}$ vecteur principal.
- Chaque composante C_k est portée par le $k^{\text{ème}}$ axe principal.
- La dispersion des projections des variables sur la composante C_k est mesurée par la valeur propre λ_k .
- Les valeurs sont rangées par ordre décroissant : $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_Q$.
- Les CP sont rangées de 1 à Q dans l'ordre des valeurs propres.
- La moyenne de chaque CP est nulle.

Si nous voulons normer les CP, c'est à dire, imposer à chacune d'entre elles d'avoir un écart type unité il suffit de diviser chacune d'elles par la racine de la variance expliquée $(\lambda_k)^{1/2}$ correspondante.

III-1-4-3 Calcul des composantes principales :

On désigne par CP, la projection du nuage de points initiale sur le nouveau système d'axes fournit par les vecteurs propres. Notons $[C]$ la matrice des CP.

$$[C] = [A]' [X] \quad (\text{III-27})$$

$[X]$: Matrice des données initiales ;

$[A]'$: Matrice transposée de la matrice composée par les vecteurs propres.

III-1-5 Principaux résultats :

En général une analyse en composantes principales fournit trois sources de renseignements, toutes nécessaires à l'interprétation :

- Un tableau de vecteurs et valeurs propres.
- Un tableau des corrélations des individus avec les axes factoriels.
- Un tableau des corrélations des variables aux axes principaux.

III-1-6 Représentation graphique :

Une fois les résultats numériques obtenus nous passons à la représentation graphique :

- Nous pouvons représenter chaque variable par un point dont les coordonnées sont les covariances avec les CP réduites.

$$COV(X_j, C_k) = (\lambda)^{1/2} * V_k(j) \tag{III-28}$$

Si nous considérons une variable centrée réduite Y_j au lieu de la variable initiale X_j , les coordonnées de cette nouvelle variable dans le plan factoriel sont calculées par :

$$COV(Y_j, C_i) = COV\left(\frac{X_j}{S_j}, C_i\right) = Cor(X_j, C_i) \tag{III-29}$$

C_1 : Composante principale normée ;
 S_j : Ecart type de la $j^{ème}$ variable.

Comme la variance de Y_j est égale à 1, cela signifie que son point représentatif se trouve sur une sphère de rayon égal à 1. C'est pourquoi le cercle unité tracé sur chacun des plan étudié s'appelle cercle de corrélation. La variable sera d'autant mieux expliquée, que son point représentatif se rapproche du cercle et inversement.

- Nous pouvons aussi observer la répartition des individus selon les principaux axes choisis. Ainsi deux individus seront proches dans l'espace R^P , s'ils sont proches dans le plan factoriel. Dans le graphique des individus on s'intéresse aux distances inter-individus qu'on peut interpréter comme étant des ressemblances.

III-1-7 ACP Programmation :

Le logiciel que nous avons mis au point est composé de deux programmes, l'un en Fortran 90 (TACP), constitué d'une suite de sous programmes traitant toute la partie numérique de l'ACP, l'autre en Turbo-Pascal (DESS), et permet de visualiser graphiquement tous les résultats numériques de l'ACP. Le logiciel élaboré est très simple à utiliser et affranchit l'utilisateur du format des fichiers INPUT, vu que la lecture de ceux-ci se fait sous format libre.

III-1-7-1 Fichier de données INPUT :

Le fichier des données doit être constitué comme suit :

			N		
			P		
X_{11}	X_{12}	X_{1j}	X_{1p}
X_{21}	X_{11}	X_{2j}	X_{11}
.....
X_{N1}	X_{N2}	X_{Nj}	X_{Np}

N : Entier désignant le nombre d'individus (nombre d'années) ;
 P : Entier désignant le nombre de variables (nombre de valeurs) ;
 X_{ij} : Données à introduire.

L'introduction des données se fait ligne par ligne, dont chacune d'elle représente un individu.

III-1-7-2 Organigramme de calcul :

Le logiciel que nous avons élaboré est constitué de cinq sous programmes principaux (figure III-1) effectuant les tâches suivantes :

- 1- Transformation des données observées.
- 2- Calcul de la matrice de covariance ou de corrélation $[R]$ ainsi que les paramètres statistiques.
- 3- Diagonalisation de $[R]$ (recherche des vecteurs propres).
- 4- Calcul des composantes principales.
- 5- Visualisation graphique.

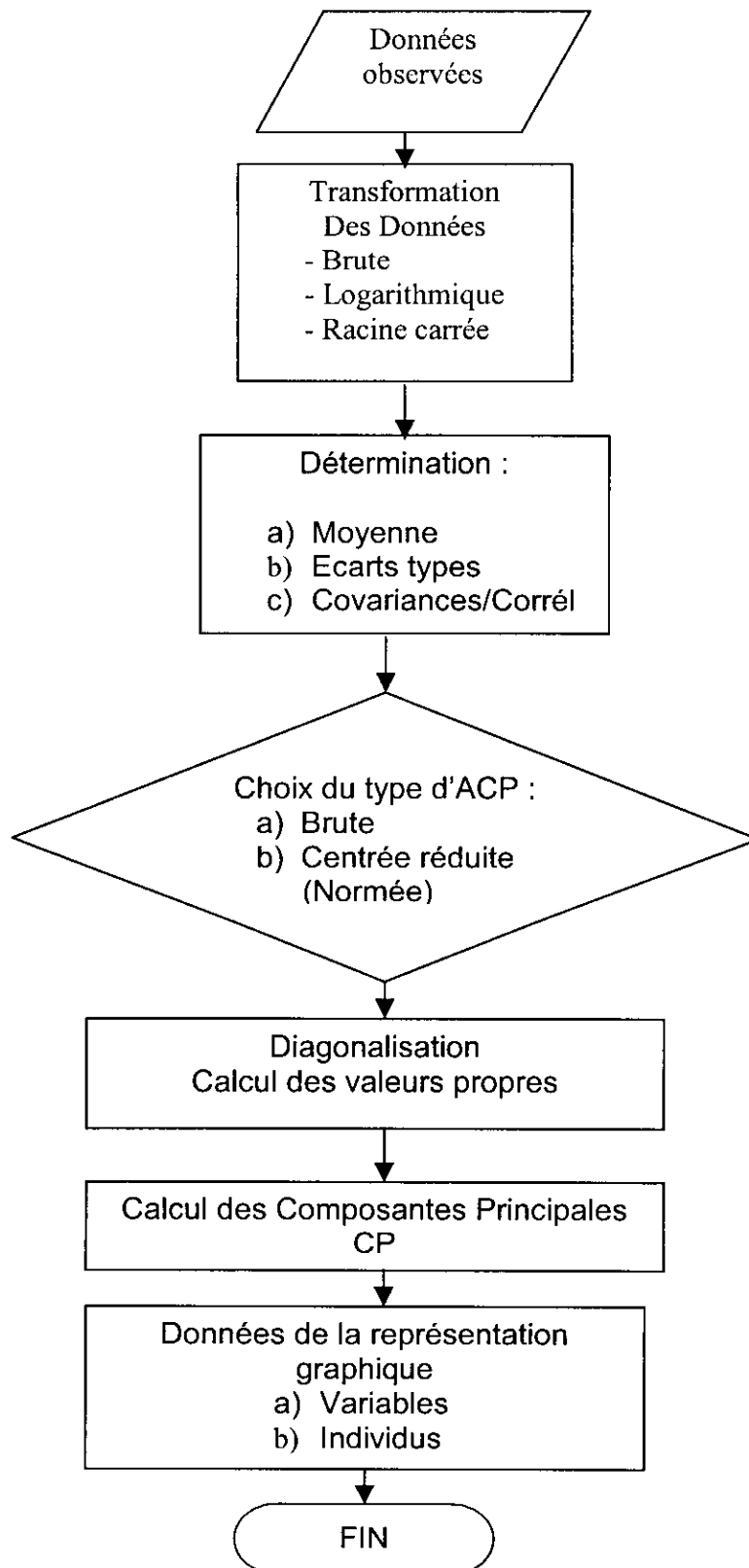


Figure III-1 Organigramme de l'Analyse en Composantes Principales.

En conclusion, nous pouvons dire que l'analyse en composantes principales, dans son aspect qualitative, permet d'établir deux sortes de bilans :

- Un bilan de liaisons entre les variables pouvant nous renseigner sur les points suivants :
 - Quelles sont les variables qui sont liées positivement, et celles qui s'opposent (liées négativement)?
 - Existe-t-il une répartition en groupes des variables inter-corrélées ?
- Un bilan de ressemblance entre individus répondant aux questions suivantes :
 - Quels sont les individus qui se ressemblent et ceux qui diffèrent ?
 - Existe-t-il des groupes homogènes d'individus ?

III-1-8 Application de l'ACP :

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.

Si les conditions de normalité ne sont ni impératives ni nécessaires pour pratiquer l'ACP, le type de variables utilisé influe par contre sur les résultats obtenus.

Nous distinguons deux aspects principaux de l'ACP :

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

III-1-8-1 Influence de la transformation des variables :

La matrice de corrélation (variable centrée réduite) ou de covariance (variable brute) constitue le point de départ de toute analyse en composantes principales, c'est l'une ou l'autre en effet qui permet après diagonalisation de calculer les valeurs et les vecteurs propres, et par conséquent les composantes principales (CP).

Or la valeur du coefficient de corrélation (ou de covariance) entre les variables X_j et X_k dépend essentiellement des couples d'observations X_{ij} et X_{ik} et par conséquent toute transformation appliquée aux p variables X_j modifiera les coefficients de corrélation (ou de covariance), et de ce fait la structure même des composantes principales.

Il existe des transformations usuelles pour l'analyse des données à savoir :

1-Variable centrée :

$$Y_j = X_j - \overline{X_j} \quad (\text{III-30})$$

Avec : X_j : Variable initiale ;
 $\overline{X_j}$: Moyenne de la variable initiale ;
 Y_j : Variable transformée (centrée).

2-Variable réduite :

$$Y_j = \frac{X_j}{\sigma_{X_j}} \quad (\text{III-31})$$

Avec : X_j : Variable initiale ;
 σ_{X_j} : Ecart type de la variable initiale ;
 Y_j : Variable transformée (réduite).

3-Variable centrée réduite :

$$Y_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}} \quad (\text{III-32})$$

Avec : X_j : Variable initiale ;
 $\overline{X_j}$: Moyenne de la variable initiale ;
 σ_{X_j} : Ecart type de la variable initiale ;
 Y_j : Variable transformée (centrée réduite).

L'analyse en composantes principales peut donc être effectuée sur quatre types de variables :

- ACP sur variables brutes (sans transformation) ;
- ACP sur variables centrées ;
- ACP sur variables réduites ;
- ACP sur variables centrées réduites.

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-30), 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).

$$Y_j = X_j - \overline{X_j} \Rightarrow \overline{Y_j} = 0 \tag{III-33}$$

$$Y_k = X_k - \overline{X_k} \Rightarrow \overline{Y_k} = 0 \tag{III-34}$$

Or :

$$Cov(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \overline{Y_j}) * (Y_{ik} - \overline{Y_k}) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} \tag{III-35}$$

En remplaçant chaque variable par son expression nous obtenons :

$$Cov(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k}) \tag{III-36}$$

$$= Cov(X_j, X_k) \tag{III-37}$$

d'où :

$$Cov(Y_j, Y_k) = Cov(X_j, X_k) \tag{III-38}$$

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-31) et (III-32), 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.

$$Y_j = \frac{X_j}{\sigma_{X_j}} \Rightarrow \overline{Y_j} = \frac{\overline{X_j}}{\sigma_{X_j}} \tag{III-39}$$

$$Y_k = \frac{X_k}{\sigma_{X_k}} \Rightarrow \overline{Y_k} = \frac{\overline{X_k}}{\sigma_{X_k}} \tag{III-40}$$

$$Z_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}} \Rightarrow \overline{Z_j} = 0 \tag{III-41}$$

$$Z_k = \frac{X_k - \overline{X_k}}{\sigma_{X_k}} \Rightarrow \overline{Z_k} = 0 \tag{III-42}$$

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

$$\begin{aligned}
Cov(Y_j, Y_k) &= \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \bar{Y}_j) * (Y_{ik} - \bar{Y}_k) \\
&= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij}}{\sigma_{X_j}} - \frac{\bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik}}{\sigma_{X_k}} - \frac{\bar{X}_k}{\sigma_{X_k}} \right) \\
&= \frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k) \\
&= \frac{\sigma_{X_j} * \sigma_{X_k}}{\sigma_{X_j} * \sigma_{X_k}} \\
&= \frac{Cov(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\
&= Cor(X_j, X_k)
\end{aligned} \tag{III-43}$$

Donc :

$$Cov(Y_j, Y_k) = Cor(X_j, X_k) \tag{III-44}$$

$$Cov(Z_j, Z_k) = \frac{1}{N} \sum_{i=1}^N (Z_{ij} - \bar{Z}_j) * (Z_{ik} - \bar{Z}_k) \tag{III-45}$$

$$\begin{aligned}
&= \frac{1}{N} \sum_{i=1}^N Z_{ij} * Z_{ik} \\
&= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik} - \bar{X}_k}{\sigma_{X_k}} \right) \\
&= \frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k) \\
&= \frac{\sigma_{X_j} * \sigma_{X_k}}{\sigma_{X_j} * \sigma_{X_k}} \\
&= \frac{Cov(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\
&= Cor(X_j, X_k)
\end{aligned} \tag{III-46}$$

Donc :

$$Cov(Z_j, Z_k) = Cor(X_j, X_k) \tag{III-47}$$

$$D'où : Cov(Y_j, Y_k) = Cov(Z_j, Z_k) = Cor(X_j, X_k) \tag{III-48}$$

En conclusion nous aurons 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-1-8-2 Exemple d'application :

Le modèle de calcul de l'ETP Penman développé au chapitre II, fait intervenir quatre paramètres climatologiques à savoir la température, la durée d'insolation, la vitesse du vent et l'humidité relative, nous proposons de discerner la part d'explication de la variation de l'ETP qui revient à chacun de ces paramètres, en considérant comme base de données, la série des valeurs journalières d'ETP calculées (d'après le modèle de Penman) sur les 34 années du fichier climatique de la station d'Alger.

Vu l'hétérogénéité des paramètres étudiés, nous avons effectué une ACP normée sur le fichier historique constitué d'une matrice de [12418 x 5]. Pour une question pratique, chaque variable a été affectée d'un code qui l'identifie.

Tableau III-1 Codification des variables étudiées.

Variable	Code affecté
Température	01
Vitesse du vent	02
Insolation	03
Humidité relative	04
ETP	05

Pour chaque colonne (variable) de cette matrice, nous calculons la moyenne et l'écart type, ensuite nous transformons les données initiales en données centrées réduites.

Tableau III-2 Paramètres statistiques des variables étudiées.

Variable	Moyenne	Ecart type
01(°C)	17.4	5.6
02(m/s)	3.1	1.7
03(h+.1)	7.7	3.9
04(%)	73.5	9.7
05(mm)	3.5	1.7

Tableau III-3 Matrice des corrélations.

	01	02	03	04	05
01	1.00				
02	-0.02	1.00			
03	0.42	-0.22	1.00		
04	-0.38	-0.17	-0.34	1.00	
05	0.84	0.26	0.59	-0.56	1.00

De la lecture du tableau III-3, ressort l'influence de la température qui a un coefficient de corrélation de 0.84 avec l'ETP calculée par le modèle de Penman.

L'influence de l'insolation et de l'humidité relative est moins forte, mais non négligeable, puisqu'elles ont respectivement un coefficient de corrélation de 0.59 et -0.56 avec l'ETP, la vitesse du vent ne semble pas avoir une grande influence sur l'ETP Penman au pas de temps

journalier pour la zone d'étude puisque le coefficient de corrélation n'est que de 0.26 et est encore moins corrélée avec les autres paramètres climatiques intervenant dans la formule de calcul de l'ETP Penman.

L'ACP est une technique qui consiste à mettre en évidence les redondances éventuelles et les groupements de variables (par les vecteurs propres) qui expliquent le mieux la variance de tout l'ensemble d'informations (la part critère de variance expliquée est le rapport de la valeur propre au nombre de variables).

La matrice de corrélation permet après diagonalisation de calculer les valeurs et vecteurs propres. En tenant compte des cinq variables (température, vitesse du vent, insolation, humidité relative, et ETP), le calcul des valeurs propres conduit aux parts de variances globales consignées dans le tableau ci-dessous :

Tableau III-4 Vecteurs propres (matrice de passage).

	Axe1	Axe2	Axe3	Axe4	Axe5
01	.521	-.081	.529	-.418	.521
02	.076	.867	.108	.400	.263
03	.435	-.428	-.196	.720	.271
04	-.430	-.211	.785	.378	-.094
05	.591	.117	.233	.062	-.760

Tableau III-5 Contribution de chaque composante principale.

	Valeurs propres	Contribution (%)	Contribution cumulée (%)
1	2.61	52.20	52.20
2	1.19	23.74	75.94
3	.64	12.88	88.82
4	.50	10.02	98.84
5	.06	1.06	100.00

En examinant l'apport de chaque composante, nous remarquons que pour avoir près de 76% de la variance totale, il suffit de ne considérer que les deux premières CP, mais vu l'apport assez consistant de la troisième CP (environ 12%), nous ne pouvons négliger cette dernière, les trois premières CP reconstituent à elles seules près de 89% de la variance totale du système.

La projection dans le plan des variables permet de donner leurs caractéristiques statistiques en fonction des positions de leurs points représentatifs.

Tableau III-6 Projection des variables étudiées sur les axes principaux.

	Axe1	Axe2	Axe3	Axe4	Axe5
01	.842	-.092	.425	-.300	.126
02	.122	.944	.081	.282	.063
03	.703	-.469	-.159	.511	.065
04	-.693	-.228	.629	.263	-.023
05	.955	.124	.185	.042	-.183

POINT CACHE

PAR

PROJECTION SUR LE PLAN FACTORIEL (1 * 2)

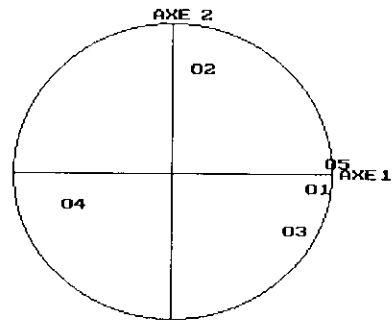


Figure- III-2 Projection des variables étudiées sur le plan d'axe 1 x 2

La figure III-2 montre un regroupement des variables ETP, température et à un degré moindre l'insolation autour de l'axe 1, cela signifie que ces deux derniers paramètres sont prépondérants dans la détermination de l'ETP Penman journalier.

L'humidité relative est opposée au groupe formé par l'ETP, la température et l'insolation par rapport à l'axe 1, une humidité relative importante est un apport pour le couvert végétal.

POINT CACHE

PAR

PROJECTION SUR LE PLAN FACTORIEL (2 * 3)

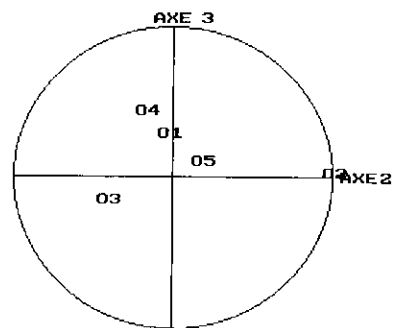


Figure III-3 Projection des variables étudiées sur le plan d'axe 2 x 3

La figure III-3 montre que la vitesse du vent n'est corrélée qu'avec le second axe, le tableau III-3 montre une indépendance de cette variable avec les autres, l'axe 2 est défini par l'effet du vent jouant le rôle de transport des masses d'air saturées.

Sur la base des résultats obtenus, nous remarquons que la température représente le facteur prépondérant dans la détermination de l'ETP, la part d'explication de la variation de l'ETP qui revient à chacun des autres paramètres est faible voire même négligeable.

III-2 LES CHAINES DE MARKOV :

III-2-1 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 par le mathématicien russe Markov (1907) 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 antérieurs à t :

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

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 climatologie consiste en 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 $\rho = 0$: on dira que les événements sont parfaitement indépendants.
- Si $\rho = 1$: 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 de celui du jour précédent, elle sera du $n^{\text{ème}}$ ordre si ce même événement dépend des jours $k-1, k-2, \dots, k-n$.

III-2-2 Ordre des chaînes de Markov :

III-2-2-1 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 c'est à dire qu'on ne tient compte que des événements qui se

sont produits la veille.

Soient X la variable aléatoire caractérisant l'état ($x=0$ ou 1) et X_k , la valeur du jour k, soient i l'état réalisé le jour k-1 ($i=0$ ou 1), et j l'état réalisé le jour k ($j=0$ ou 1) la probabilité conditionnelle est la probabilité de l'état j au jour k sachant qu'était réalisé l'état i au jour k-1.

$$P [X_k=j / X_{k-1}=i] = \alpha_{ij} \tag{III-49}$$

α_{ij} : Probabilité d'obtenir l'état j au jour k, sachant qu'au jour k-1, était réalisé l'état i.

En considérant deux états possibles, nous obtenons une matrice de passage présentée dans le tableau III-7.

Tableau III-7 : Matrice de passage d'ordre 1

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

Avec :

$$\begin{cases} \alpha_{00} + \alpha_{01} = 1 \\ \alpha_{10} + \alpha_{11} = 1 \end{cases} \quad \text{et } \alpha_{ij} \geq 0 \tag{III-50}$$

Ces probabilités sont obtenues par le décompte des jours caractérisés par les états i, j.

Exemple : nous avons appliqué le modèle Markovien d'ordre 1 sur les données pluviométriques journalières de l'année 1995 de la station d'Alger.

Le décompte des jours k caractérisés par l'état 0 (jour sec) ou 1 (jour pluvieux) précédés par les jours k-1 d'état 0 ou 1 a donné les résultats suivants :

Nombre de jour k-1	Nombre de jour k		
	0	1	
0	269	34	= 303
1	35	27	= 62

En divisant chaque ligne par son total, nous obtenons la matrice de passage :

0	0.887	0.112	= 1
1	0.564	0.435	= 1

Nous pouvons dire, par exemple, que pour l'année 1995, la probabilité d'avoir deux jours successifs secs était de 88%, par contre celle d'avoir deux jours successifs humides était de 43%.

III-2-2-2 Processus d'ordre 2 :

Il est caractérisé par la probabilité conditionnelle de passage des états d'aujourd'hui et d'hier en fonction des états connus d'hier et d'avant-hier. Le dénombrement statistique fournit le tableau suivant :

Tableau III-8 : Matrice de passage d'ordre 2

Etats au jours k-1 et k-2	Etats au jours k-1 et k			
	00	01	10	11
00	β_{000}	β_{001}	-	-
01	-	-	β_{010}	β_{011}
10	β_{100}	β_{101}	-	-
11	-	-	β_{110}	β_{111}

β_{ijk} : Probabilité conditionnelle d'obtenir un doublet de classe jk succédant à un autre de classe ij .

Du fait qu'il y a chevauchement de deux couples sur la veille, il est nécessaire d'avoir l'égalité de la classe de transition dans les deux couples. Pour cet ordre, l'état au jour k , dépend de l'état aux jours $k-1$ et $k-2$. Lors de la définition de la matrice de passage exprimant les probabilités de couples successifs, il y a impossibilité de certaines combinaisons, vue la succession de certains doublets.

III-2-3 Les modèles multiclassés :

Au lieu d'utiliser un système binaire, nous pouvons 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, nous transformons 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 $N^r \times N^r$.

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

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

- Aléatoirement : nous choisissons les seuils et nous suivons l'évolution des coefficients de la matrice de passage.
- En utilisant la fonction de répartition de la série historique, nous considérons les classes d'égale probabilité sauf pour la classe sèche.
- En se basant sur la moyenne et l'écart type de la série historique : nous procédons au découpage à l'aide des seuils suivants :

$$\begin{aligned} & (m - \sigma, m + \sigma) \\ & (m - 2\sigma, m + 2\sigma) \\ & (m - 3\sigma, m + 3\sigma) \end{aligned}$$

m : Moyenne de la série historique ;
 σ : Ecart type.

Nous avons élaboré un programme en Fortran 90 MARBIM (Modèle MARKovien Binaire et Multiclasse), qui permet d'automatiser toutes ces tâches fastidieuses, il est subdivisé en deux sous-programmes :

- Le premier permettant de traiter les modèles binaires ;
- Le second s'intéresse aux modèles multiclassés.

Un fichier spécifique d'entrée des données (DONNEA.dat) a été développé afin d'étudier toutes les variantes possibles en agissant non seulement sur les ordres des chaînes de Markov, mais aussi sur le nombre de classes à prendre en considération et sur le choix des saisons.

Exemple 1 :

Sur la base des données pluviométriques mensuelles de la station d'Alger (1936-2001). Nous avons appliqué un modèle Markovien multiclasse d'ordre 1. Pour cela nous avons choisi un nombre de classe égale à 3. Les bornes des classes étant choisies aléatoirement comme suit :

$$[0 \quad 30[\qquad [30 \quad 70[\qquad [70 \quad \infty[$$

La matrice de passage obtenue s'écrit :

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

Nous remarquons que la plus forte probabilité conditionnelle est celle d'avoir deux mois successifs secs (<30mm), par contre la plus faible est d'avoir un mois humide (≥ 70 mm) précédé d'un mois sec .

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 représente le plus fidèlement possible le phénomène physique étudié.

Exemple 2 :

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).

Nombre de classes : 03

Bornes des classes : [0 30[[30 50[[50 ∞[

- Saison 2 : mois de mai- mois de septembre (saison sèche).

Nombre de classes : 02

Bornes des classes : [0 30[[30 ∞[

Les matrices de passage obtenues sont :

Pour la saison 1 :

	0	1	2
0	.233	.189	.577
1	.233	.155	.611
2	.156	.179	.664

Pour la saison 2 :

	0	1
0	.846	.153
1	.766	.233

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

Chapitre IV : Elaboration du Modèle de Simulation

IV-1 INTRODUCTION :

Par définition, simuler c'est faire paraître comme réel une chose qui ne l'est point, et toutes les sciences utilisent la simulation afin de créer des modèles pour tester ou représenter une théorie quelconque, en s'appuyant sur les ordinateurs. Elle permet dans bien des cas d'éviter les déboires de l'expérimentation.

La simulation s'appuie sur la construction d'un modèle dont toutes ses caractéristiques sont connues, pour identifier le comportement d'un procédé ou d'un système et interpréter son mécanisme, ce qui aide à entreprendre de meilleures décisions. Nous pouvons dire donc que la simulation est un très bon planificateur.

En hydrologie, par exemple, celle-ci permet de créer à partir des séries observées, d'autres séries générées qui ont les mêmes caractéristiques statistiques, afin de pouvoir prendre des décisions et anticiper sur le problème de la gestion de la ressource en eau.

Nous utilisons souvent aussi la simulation pour évaluer une politique d'exploitation. Cette méthode permet en effet, en reproduisant le comportement du barrage, d'examiner les conséquences d'une politique d'exploitation sur l'évolution de ce barrage (l'état des côtes d'eau).

L'objectif de ce chapitre est de présenter le modèle élaboré dans cette thèse ainsi que les différentes techniques de simulation utilisées. Celui-ci est basé sur la technique de l'analyse en composantes principales qui est généralement utilisée dans le domaine spatial.

IV-2 MODELE DE SIMULATION :

Les composantes principales étant des combinaisons linéaires des variables initiales, cela se traduit par (Dechemi et al., 1994) :

$$C_l = \sum_{j=1}^p a_{lj} * X_j \quad (IV-1)$$

C_l : l^{ième} composante principale ;

a_{lj} : Cosinus directeur, éléments de la matrice des vecteurs propres ;

X_j : Variable d'ordre j.

Soit la composante principale normée C_l' telle que :

$$C_l' = \frac{C_l}{\lambda_l^{1/2}} \quad (IV-2)$$

λ_l : Valeur propre correspondant à C_l .

En divisant l'équation (IV-1) par $\lambda_1^{1/2}$, nous obtenons :

$$C'_i = \frac{C_i}{\lambda_1^{1/2}} = \frac{1}{\sqrt{\lambda_1}} \sum_{j=1}^p a_{ij} * X_j \quad (IV-3)$$

Sous forme matricielle cette expression devient :

$$[C'] = \lambda^{-1/2} [A][X] \Rightarrow [X] = \lambda^{1/2} [A]' [C'] \quad (IV-4)$$

- λ : Vecteur dont les composantes sont les valeurs propres ;
- $[A]$: Matrice des cosinus directeurs, formée par les vecteurs propres ;
- $[A]'$: Matrice transposée de $[A]$;
- $[X]$: Matrice des variables initiales.

Soit donc :

$$[X] = \lambda^{1/2} [A]' [C'] \quad (IV-5)$$

Si nous ne considérons que les Q premières composantes principales, et que nous ne perdons pas de vue la variance non expliquée prise en compte par le terme résiduel noté ε_j , et en explicitant la relation (IV-5) nous aurons :

$$X = \lambda_1^{1/2} \sum_{i=1}^Q a_{ji} * C'_i + \varepsilon_j \quad (IV-6)$$

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

$$\sigma_{\varepsilon_j} = \sqrt{1 - \sum_{i=1}^Q a_{ji}^2 * \lambda_i} \quad (IV-7)$$

Or :

$$\lambda_1^{1/2} * a_{ji} = Cov(C'_i, X_j) \quad (IV-8)$$

En remplaçant ce résultat dans l'équation (IV-6), nous obtenons l'expression finale du modèle brute (Dechemi et al. 1997) :

$$\hat{X}_j = \sum_{i=1}^Q Cov(C'_i, X_j) * C'_i + \varepsilon_j \quad (IV-9)$$

Pour une variable centrée réduite :

$$Y_j = \frac{X_j - \bar{X}_j}{\sigma_{X_j}} \quad (IV-10)$$

L'équation (IV-4) s'écrit :

$$[C'] = \lambda^{-1/2} [A][Y] \quad (IV-11)$$

[Y] : Matrice des variables initiales centrées réduites.

Réciproquement nous pouvons exprimer la variable Y en fonction des composantes principales normées en transposant la matrice des vecteurs propres [A] :

$$[Y] = \lambda^{1/2} [A]^T [C'] \quad (IV-12)$$

En explicitant la formule (IV-12) nous aurons :

$$Y_j = \lambda^{1/2} \sum_{l=1}^q a_{jl} * C'_l + \varepsilon_j \quad (IV-13)$$

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

$$\sigma_{\varepsilon_j} = \sqrt{1 - \sum_{l=1}^q a_{jl}^2 * \lambda_l} \quad (IV-14)$$

Et sachant que : $\lambda_l^{1/2} * a_{jl} = Cov(C'_l, Y_j)$ (IV-15)

Or : $Cor(C'_l, Y_j) = \frac{Cov(C'_l, Y_j)}{\sigma_{C'_l} * \sigma_{Y_j}}$ (IV-16)

$\sigma_{C'_l} = 1$ et $\sigma_{Y_j} = 1$ (pour une ACP normée et une variable centrée réduite)

$$Cov(C'_l, Y_j) = Cor(C'_l, Y_j) \quad (IV-17)$$

En tenant compte de ce résultat l'équation (IV-9) devient :

$$Y_j = \sum_{l=1}^q Cor(C'_l, Y_j) * C'_l + \varepsilon_j \quad (IV-18)$$

Substituons l'équation (IV-10) dans (IV-18) :

$$\begin{aligned} \frac{X_j - \overline{X_j}}{\sigma_{X_j}} &= \sum_{l=1}^q Cor(C'_l, Y_j) * C'_l + \varepsilon_j \\ X_j &= \overline{X_j} + \sum_{l=1}^q \sigma_{X_j} * Cor(C'_l, Y_j) * C'_l + \sigma_{X_j} * \varepsilon_j \end{aligned} \quad (IV-19)$$

En posant :

$$\begin{aligned} \beta_{j0} &= \overline{X_j} \\ \beta_{jl} &= \sigma_{X_j} * Cor(C'_l, Y_j) \\ E_{.j} &= \sigma_{X_j} \varepsilon_j \end{aligned}$$

L'expression finale du modèle normé est donnée par :

$$\hat{X}_j = \beta_{j0} + \sum_{i=1}^q \beta_{ji} * C_i + E_{.j} \quad (\text{IV-20})$$

En se basant sur la structure des deux modèles obtenus brutes (à trois paramètres) et normés (à quatre paramètres), différentes possibilités de simulation peuvent être envisagées.

IV-3 METHODES DE SIMULATION DES PARAMETRES DU MODELE :

Toutes les méthodes de simulation se basent sur un tirage de nombres au hasard, en effet disposant de l'ensemble des valeurs d'un phénomène, tirer au hasard un échantillon de n valeurs parmi cette population signifie que chaque individu a la même chance d'être choisi, et que tous les échantillons possibles d'effectif n , ont la même chance d'être sélectionnés. Pour réaliser cette opération indépendamment de toute influence humaine ou physique, nous utilisons les nombres au hasard, plus connus sous le nom des méthodes de Monté-Carlo.

IV-3-1 Simulation par les fonctions de répartition :

Si nous considérons la variable aléatoire X de fonction de répartition $F(x)$, la nouvelle variable $Y = F(x)$ est uniformément répartie sur le segment $]0,1[$:

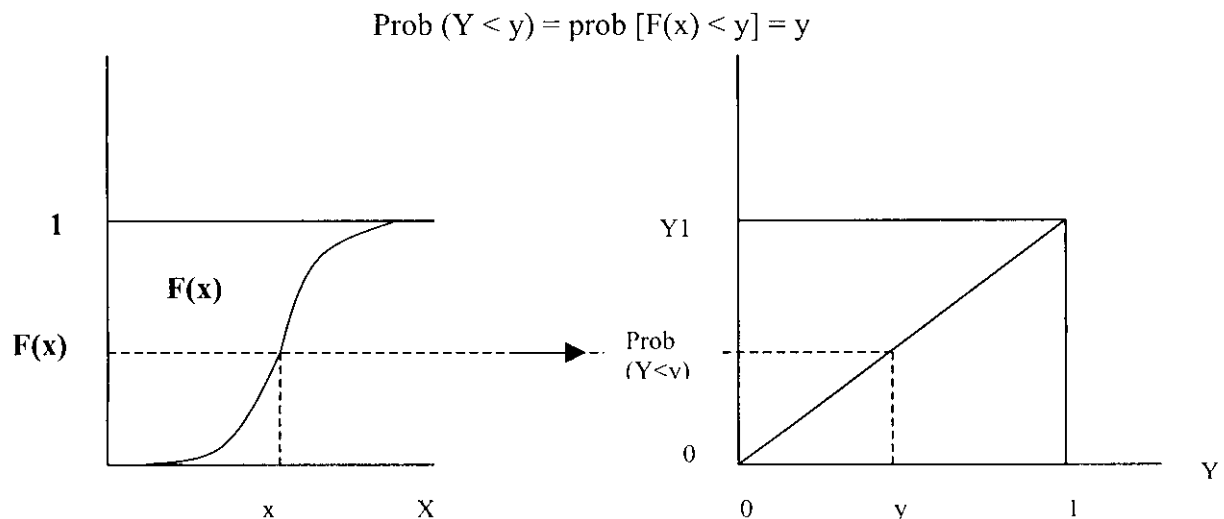


Figure IV-1 : Passage d'une fonction de probabilité F à une loi uniforme

Quelle que soit la variable aléatoire donnée, nous pouvons toujours la transformer en une variable uniformément distribuée sur le segment $]0,1[$.

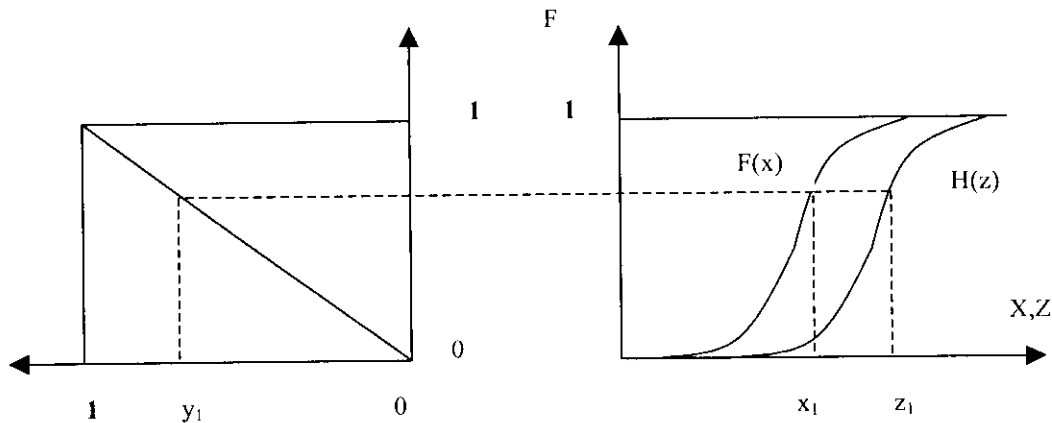


Figure IV-2 Tirage aléatoire à partir de la fonction de répartition

La figure IV-2 représente le tirage aléatoire :

- D'une valeur de la variable aléatoire X définie par la loi de probabilité $F(x)$.
- D'une valeur de la variable aléatoire Z définie par la loi de probabilité $H(z)$.

Nous mettons la loi de probabilité F ou H dans l'urne des nombres au hasard, et nous tirons un nombre aléatoire y_1 qui, après transformation, fournit la valeur x_1 ou z_1 .

En pratique, la génération de la fréquence y_1 est remplacée par une loi discrète dont les sauts valent 10^{-k} ($k=3$ ou 4), nous associons des ensembles de k chiffres extraits des tables de nombres au hasard.

Pour la simulation par la fonction de répartition, la projection de la fréquence simulée sur cette dernière ne peut se faire que si nous disposons de l'expression analytique de la fonction elle-même, pour cela nous faisons appel aux méthodes de lissage.

L'une des techniques les plus couramment utilisées, c'est le cubique spline qui est une méthode de lissage par intervalles utilisant un polynôme du 3^{ème} degré.

Ce polynôme du 3^{ème} degré possède quatre constantes lui conférant une flexibilité suffisante pour assurer la continuité de la dérivée seconde sur le même intervalle.

Soit f une fonction à approximer, définie sur l'intervalle $[a, b]$, et un ensemble de nombres dits nœuds tels que :

$$a = x_0 < x_1 < \dots < x_N = b$$

On désigne par S la fonction d'interpolation spline satisfaisant les conditions suivantes :

- * S est un polynôme du 3^{ème} degré noté S_j , dans le sous intervalle $[X_j, X_{j+1}]$
- $\forall j, j = 0, N-1$
- * $S_j(X_j) = f(X_j)$
- $\forall j, j = 0, N$

$$* S_{j+1}(X_{j+1}) = S_j(X_{j+1}) \quad \forall j, j = 0, N-2$$

$$* S'_{j+1}(X_{j+1}) = S'_j(X_{j+1}) \quad \forall j, j = 0, N-2$$

$$* S''_{j+1}(X_{j+1}) = S''_j(X_{j+1}) \quad \forall j, j = 0, N-2$$

* L'une des deux conditions limites suivantes doit être satisfaite :

$$* S(X_0) = S(X_N) = 0 \quad \text{extrémité libre}$$

$$* S'(X_0) = f'(X_0) \quad \text{et} \quad S(X_n) = f(X_n)$$

Quand la condition limite de l'extrémité libre est satisfaite, le spline est dit naturel et le graphe résultant est une courbe flexible forcée de passer par tous les points :

$$\{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_N, f(x_N))\}$$

Pour construire la fonction d'interpolation du cubique spline par une fonction f donnée, les conditions citées auparavant peuvent être appliquées au polynôme de degré trois :

$$S_j(X) = a_j + b_j(X - X_j) + c_j(X - X_j)^2 + d_j(X - X_j)^3 \quad (\text{IV-21})$$

$$\forall j, j = 0, N-1$$

IV-3-2 Simulation par les lois d'ajustements :

La simulation par les lois d'ajustements n'est applicable que si les étapes suivantes sont satisfaites :

- L'échantillon de données doit satisfaire certains critères statistiques tels que, l'homogénéité et l'indépendance.
- Choisir la loi d'ajustement la plus adéquate pour l'échantillon considéré (avec le moins de paramètres possible à estimer).

La génération dans la loi uniforme est la base de la simulation dans la plupart des lois d'ajustements statistiques.

Un tirage aléatoire dans une loi Laplace-Gauss centrée réduite peut être effectué selon la méthode de Box et al., (1958), nous tirons deux variables aléatoires, uniformes sur $]0,1[$ indépendantes, grâce à la fonction random (aléatoire), alors Z_1 et Z_2 définies par :

$$Z_1 = (-2 * \text{Ln}U_1)^{1/2} * \text{Cos}(2\pi U_2) \quad (\text{IV-22})$$

$$Z_2 = (-2 * \text{Ln}U_1)^{1/2} * \text{Sin}(2\pi U_2)$$

sont deux variables aléatoires normales centrées réduites et indépendantes.

IV-3-2-1 Simulation par la loi de Gauss ou loi normale :

La loi normale est une loi très importante en hydrologie parce qu'une grande partie des phénomènes climatologiques (précipitations, ...) suivent une loi sensiblement semblable à celle-ci. Sa fonction de répartition s'écrit :

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma * (2 * \pi)^{1/2}} * \text{Exp} \left[-\frac{1}{2} \left(\frac{t - \mu}{\sigma} \right)^2 \right] dt \quad (\text{IV-23})$$

Cette loi d'ajustement de moyenne μ et d'écart type σ , notée $N(\mu, \sigma)$, peut toujours se ramener à une loi normale de moyenne nulle et d'écart type unité, $N(0,1)$, en posant :

$$u = \frac{x - \mu}{\sigma} \quad (\text{IV-24})$$

u : Variable réduite de Gauss ;
 x : Variable initiale ;
 μ : Moyenne de la variable x ;
 σ : écart type de la variable x .

La nouvelle fonction de distribution s'écrit alors :

$$F(u) = \frac{1}{(2 * \pi)^{1/2}} \int_{-\infty}^u \text{Exp} \left(-\frac{u^2}{2} \right) du \quad (\text{IV-25})$$

On peut toujours faire la transformation inverse par la suite :

$$x = \mu + \sigma * u \quad (\text{IV-26})$$

Pour générer dans la loi normale, nous tirons aléatoirement une fréquence notée u , que nous injectons dans la fonction de distribution de Gauss ou dans la transformation inverse.

IV-3-2-2 Simulation par la loi de Gumbel :

Tout phénomène hydrologique comporte des événements rares qui sont aux extrêmes de la fonction de densité de probabilité. Ces extrêmes peuvent être des maxima ou des minima. Les lois utilisées pour décrire les extrêmes prennent souvent des formes différentes des lois habituelles puisque ce ne sont que les extrémités des fonctions de densité de probabilité qui deviennent importantes. A ce jour les lois dites asymptotiques sont les plus acceptées à cette fin. Sous sa forme la plus simple, la fonction de distribution de Gumbel s'écrit :

$$F(u) = \text{Exp}(-\text{Exp}(-u)) \quad (\text{IV-27})$$

$$u = \alpha(x - \beta) \quad (\text{IV-28})$$

u : Variable réduite de Gumbel ;
 x : Variable initiale ;

α : Paramètre de position ;

β : Paramètre d'échelle.

La variable réduite de Gumbel peut s'écrire aussi :

$$u = -\text{Ln}[-\text{Ln}F(x)] \quad (\text{IV-29})$$

La transformation inverse s'écrit :

$$x = \frac{1}{\alpha}u + \beta \quad (\text{IV-30})$$

Pour générer dans la loi de Gumbel, nous tirons aléatoirement une fréquence notée u , que nous injectons dans la fonction de distribution de Gumbel ou dans la transformation inverse.

IV-3-2-3 Simulation par la loi de Goodrich :

Comme la fonction de distribution de Goodrich contient trois paramètres, ce type d'approximation est assez flexible ; et donne, en général, de très bons résultats surtout lorsque nous travaillons avec des valeurs moyennes.

La fonction de distribution de Goodrich (Llamas, 1985) s'écrit :

$$F(x) = 1 - \text{Exp}\left[-a*(x-b)^{1/n}\right] \quad (\text{IV-31})$$

$$b \leq x \leq +\infty$$

Où a , b , et n sont les paramètres de loi de Goodrich.

Ces paramètres sont déterminés comme suit :

$$m_1 = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}$$

$$m_2 = \frac{1}{N} \sum_{i=1}^N x_i^2$$

$$m_3 = \frac{1}{N} \sum_{i=1}^N x_i^3$$

$$s_2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$s_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3$$

$$m_3' = m_3 - (3\bar{x}s_2 + \bar{x}^3) \quad (\text{IV-32})$$

Or :

$$C_s = \frac{m_3'}{s^3} \quad (\text{IV-33})$$

Après avoir calculé C_s , nous pouvons obtenir directement la valeur de n à partir du tableau suivant :

Tableau IV-1 : Valeurs de n en fonction de C_s .

n	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65
C_s	0.069	0.217	0.359	0.496	0.631	0.764	0.895	1.028
n	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
C_s	1.160	1.294	1.430	1.567	1.708	1.852	2.000	

Nous pouvons calculer ainsi les valeurs de a et b , d'après le système d'équations suivant :

$$a = \left[\frac{1}{s^2} [\Gamma(2n+1) - \Gamma^2(n+1)] \right]^{1/2n} \tag{IV-34}$$

$$b = \bar{x} - \frac{\Gamma(n+1)}{a^n} \tag{IV-35}$$

La fonction $\Gamma(n)$ est donnée dans le tableau suivant pour des valeurs de n comprises entre 0 et 1 :

Tableau IV-2 : Valeurs de $\Gamma(n)$ en fonction de n .

N	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
$\Gamma(n)$	1	0.951	0.918	0.897	0.887	0.886	0.894	0.909	0.931	0.962	0.1

Pour d'autres valeurs de n , nous employons la formule récursive suivante :

$$\Gamma(n) = (n-1) * (n-2) * \dots * (m) * \Gamma(m) \tag{IV-36}$$

$1 \leq m \leq n$

Pour générer dans la loi de Goodrich, nous tirons aléatoirement une fréquence notée y , que nous injectons dans la fonction de distribution de Goodrich :

$$x = b + \left(\frac{\text{Ln}(1-y)}{a} \right)^n \tag{IV-37}$$

IV-3-2-4 Simulation par la loi G.E.V (General Extrem Value) :

La forme générale de cette loi de distribution a été suggérée par Jenkinson (1969) :

$$x = \alpha + \frac{1}{\beta} * \frac{1 - \exp(-k * y)}{k} \tag{IV-38}$$

Avec α , β et k sont respectivement les paramètres de position, d'échelle et de forme de cette loi.

Afin d'évaluer la période de retour T d'un événement par la méthode de Jenkinson, l'équation (IV-38) peut s'écrire comme suit :

$$x(T) = \alpha + \frac{z(T)}{\beta} \quad (\text{IV-39})$$

$z(T)$: Valeur centrée réduite de la variable Z donnée par :

$$z(T) = \frac{1 - \exp(-k * y)}{k} \quad (\text{IV-40})$$

Sachant que y et T sont liées par l'équation qui donne la relation entre la variable réduite de Gumbel et la période de retour T :

$$y = -Ln[\ln(T) - Ln(T - 1)] \quad (\text{IV-41})$$

La méthode des sextiles appliquée par Jenkinson (1969) donne une estimation approximative des paramètres α , β et k comme suit :

La fonction de densité des probabilités, de la population infinie des valeurs de Z , peut être donnée comme fonction de k et y à partir de l'équation (IV-40), est considérée comme étant arrangée dans un ordre croissant, et est divisée en six (06) groupes de tailles égales. Soient $\mu_{z,i}$ la moyenne des variables du $i^{\text{ème}}$ sextile, et $i=1$ représente le plus grand sextile, et soient μ_z et σ_z la moyenne et l'écart type des six groupes de valeurs respectivement.

De la même façon, nous divisons les valeurs correspondantes de X en six groupes. Ceci est effectué dans une série ordonnée des valeurs de l'échantillon et en les divisant en six groupes égaux.

Soit $\mu_{x,i}$ la moyenne de la population des variables du $i^{\text{ème}}$ groupe sextile. La moyenne et l'écart type de ces six valeurs sont respectivement :

$$\mu = \frac{1}{6} \sum_{i=1}^6 \mu_{x,i} \quad (\text{IV-42})$$

$$\sigma = \left[\frac{1}{6} (\mu_{x,i} - \mu)^2 \right]^{1/2} \quad (\text{IV-43})$$

De l'équation (IV-39), et en tenant compte des équations des moyennes et des variances :

$$\mu = \alpha + \frac{\mu_z}{\beta} \quad (\text{IV-44})$$

$$\sigma = \frac{\sigma_z}{\beta} \quad (\text{IV-45})$$

Les relations entre les paramètres de forme k (qui est commun aux deux populations X et Z), μ_z , σ_z et le rapport de forme $r = \frac{(\mu_{z,5} - \mu_{z,6})}{(\mu_{z,1} - \mu_{z,2})}$ sont données dans le tableau IV-3.

En tenant compte de l'équation (IV-40) et de la fonction des moments généralisée nous pouvons déterminer μ_Z :

$$\mu_Z = \frac{1 - \Gamma(1+k)}{k} \quad (\text{IV-46})$$

Cependant pour calculer σ_Z et r , on doit utiliser la fonction Gamma inverse.

Le rapport de forme r , qui est commun aux populations X et Z de l'équation (IV-39), est estimé par :

$$\hat{r} = \frac{\hat{\mu}_{x,5} - \hat{\mu}_{x,6}}{\hat{\mu}_{x,1} - \hat{\mu}_{x,2}} \quad (\text{IV-47})$$

Par interpolation, les estimations correspondantes de \hat{k} , $\hat{\mu}_Z$ et $\hat{\sigma}_Z$ sont données dans le tableau IV-3, après avoir utilisée les équations (IV-43) et (IV-44) pour calculer $\hat{\alpha}$ et $\hat{\beta}$.

Tableau IV-3 Relations entre les différents paramètres de la loi GEV pour la population Z.

k	μ_Z	σ_Z	r
-0.5	1.54	2.85	0.08
-0.4	1.22	2.24	0.11
-0.3	0.99	1.83	0.16
-0.2	0.82	1.55	0.23
-0.1	0.69	1.35	0.32
0.0	0.58	1.20	0.44
0.1	0.49	1.09	0.59
0.2	0.41	1.01	0.79
0.3	0.34	0.95	1.05
0.4	0.28	0.92	1.39
0.5	0.23	0.89	1.83
0.6	0.18	0.88	2.39
0.7	0.13	0.87	3.13

IV-3-3 Simulation par les chaînes de Markov :

Pour simuler une série donnée par les chaînes de Markov, il existe deux types de modèles : les modèles binaires dont le principe est basé sur l'existence de deux états, et les modèles multiclassés ou la série est subdivisée en différentes classes (§ III-2).

Nous utiliserons dans ce cadre là un modèle binaire, basé sur le calcul des probabilités conditionnelles caractérisées par les deux états zéro (0) et un (1).

IV-3-3-1 Construction de la chaîne de Markov :

Nous définissons en premier lieu l'état 0 et l'état 1, sur lesquels nous nous basons pour transformer la série donnée en système binaire (00111010..). Pour obtenir une telle suite, il suffit d'effectuer un comptage des éléments appartenant respectivement à l'état 0 et 1. Nous calculons ensuite les probabilités conditionnelles P_0 et P_1 .

IV-3-3-2 Génération par les séries de Markov :

Après avoir calculé les probabilités conditionnelles des deux états (0) et (1), nous déterminons les fonctions de répartition des valeurs appartenant à ces deux états.

La simulation se fait en générant uniformément un nombre aléatoire y dans l'intervalle $]0,1[$, que nous comparons aux probabilités conditionnelles.

Sachant que $P_0 + P_1 = 1$

alors $P_1 = 1 - P_0$

Donc si y est inférieur à la probabilité conditionnelle P_0 , la projection de y se fait sur la fonction de répartition correspondant à l'état 0, sinon elle se fera dans celle de l'état 1.

Afin de nous faciliter la tâche, nous avons élaboré le programme SimVac (Simulation de Variables aléatoires et cycliques) en fortran 90 (Fig IV-3), qui est constitué de trois sous-programmes :

- TRACP (fig. IV - 4) : Permet la transformation des fichiers de données recueillies à différents pas de temps (journalier, décadaire, etc.) nécessaires au modèle de simulation.
- PSACP (fig. IV - 5) : Permet la préparation des données pour la simulation de tout type de variables par l'analyse en composantes principales, et calcule les différents paramètres de simulation (β_{ij} , ϵ_j , etc.).
- SFACP (fig. IV - 6) : Permet de simuler les différents paramètres des modèles brutes et normés par les différentes méthodes de simulation.

IV-4 TESTS DE VALIDATION DU MODELE :

Nous faisons subir aux séries générées différents tests de validation afin de décider de l'aptitude du modèle à simuler le phénomène climatologique en question.

Les tests sur lesquels nous nous basons pour examiner les résultats sont :

IV-4-1 Simulation au pas de temps mensuel :

- Moyennes et écart types des totaux mensuels interannuels.
- Moyennes et écart types des totaux annuels.
- Moyennes et écart types mensuels.
- Ajustement des valeurs annuelles.
- Normalité des résidus.

IV-4-2 Simulation au pas de temps décadaire :

- Moyennes et écart types des totaux annuels.
- Moyennes et écart types des totaux mensuels.
- Moyennes et écart types des totaux décadaires.
- Moyennes et écart types mensuels interannuels.

- Ajustement des valeurs annuelles.
- Normalité des résidus.

IV-4-3 Simulation au pas de temps journalier :

- Moyennes et écart types des totaux mensuels.
- Moyennes et écart types des totaux journaliers interannuels.
- Moyennes et écart types mensuels interannuels.
- Ajustement des valeurs annuelles.
- Ajustement des maxima journaliers annuels.
- Normalité des résidus.

La détermination des intervalles de confiance pour tous les pas de temps, a été faite sur la base d'un risque d'erreur $\alpha = 5\%$.

Nous avons élaboré le programme ALS (Ajustement des Lois Statistiques) en fortran 90 pour l'ajustement graphique des différents paramètres du modèle développé.

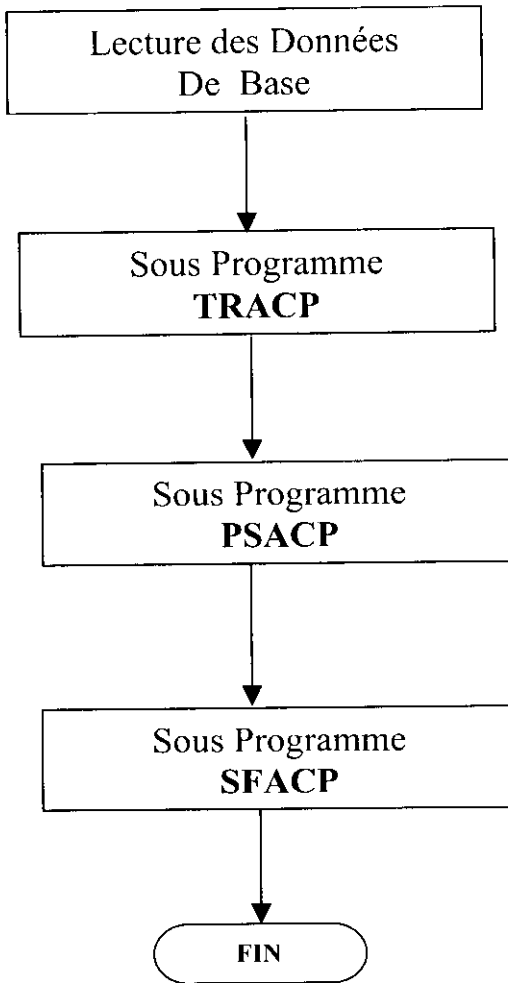


Figure IV-3 Organigramme principale du modèle de simulation (SimVac).

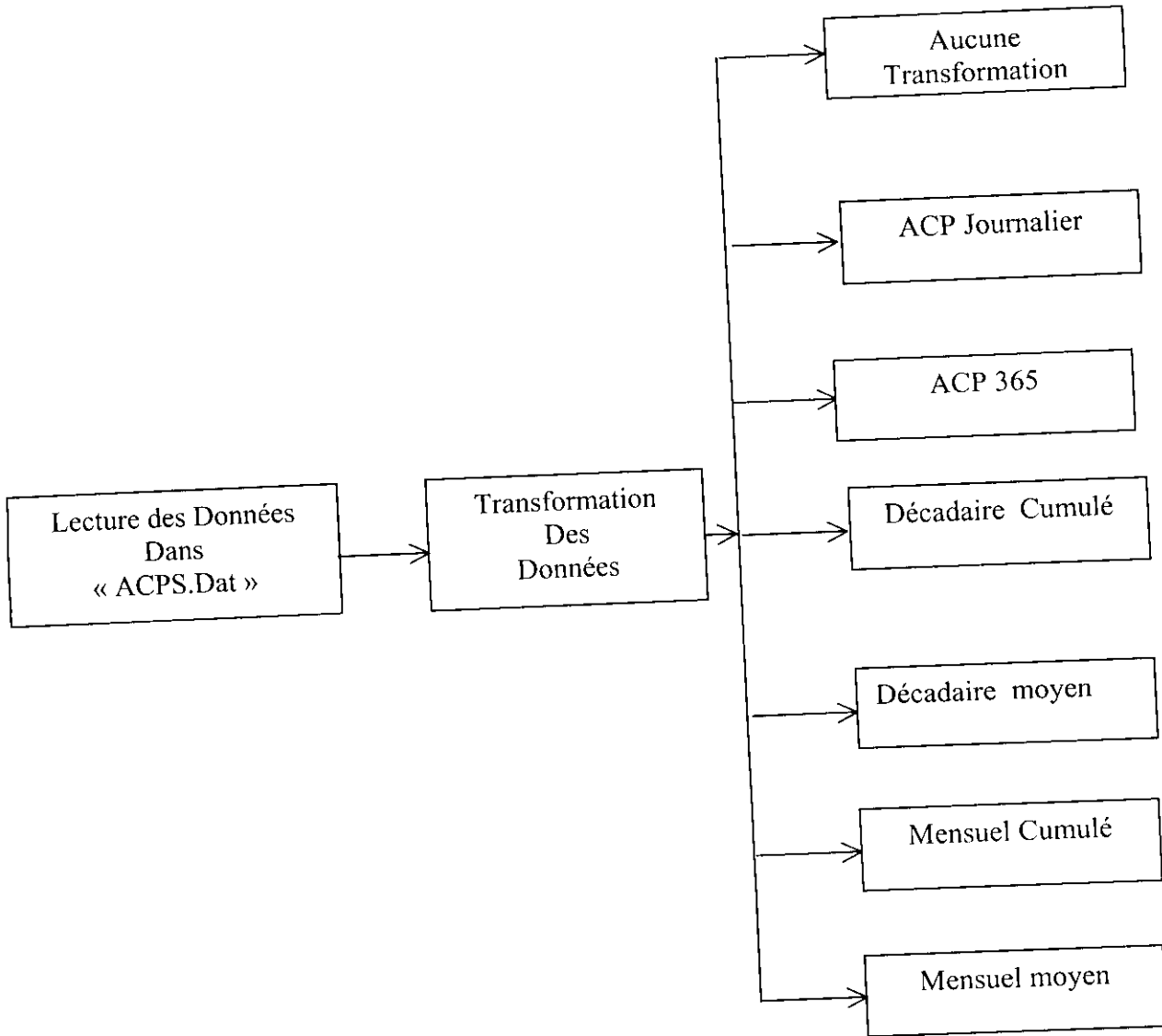


Figure IV-4 Sous programme TRACP.

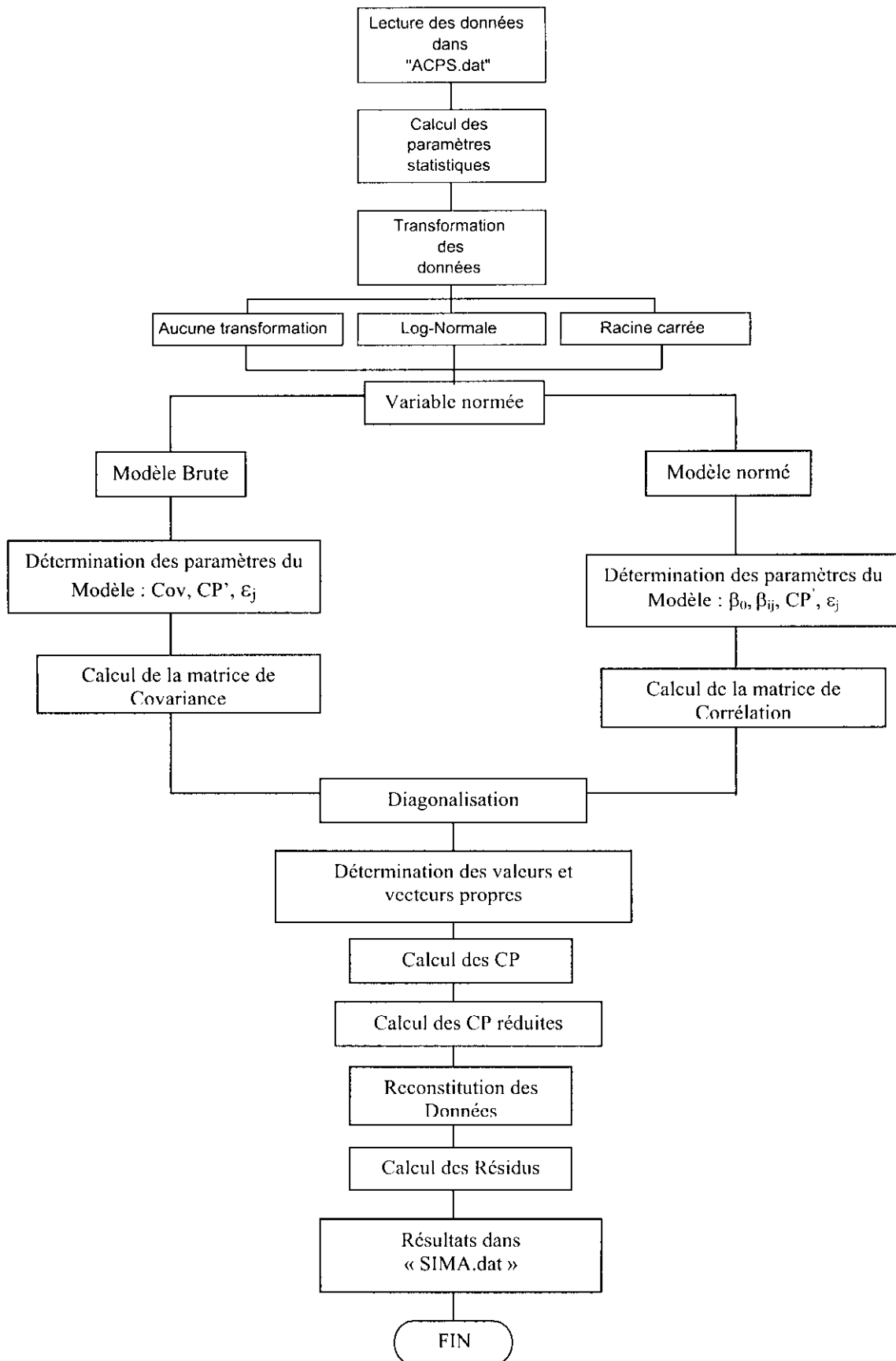


Figure IV-5 Sous programme PSACP.

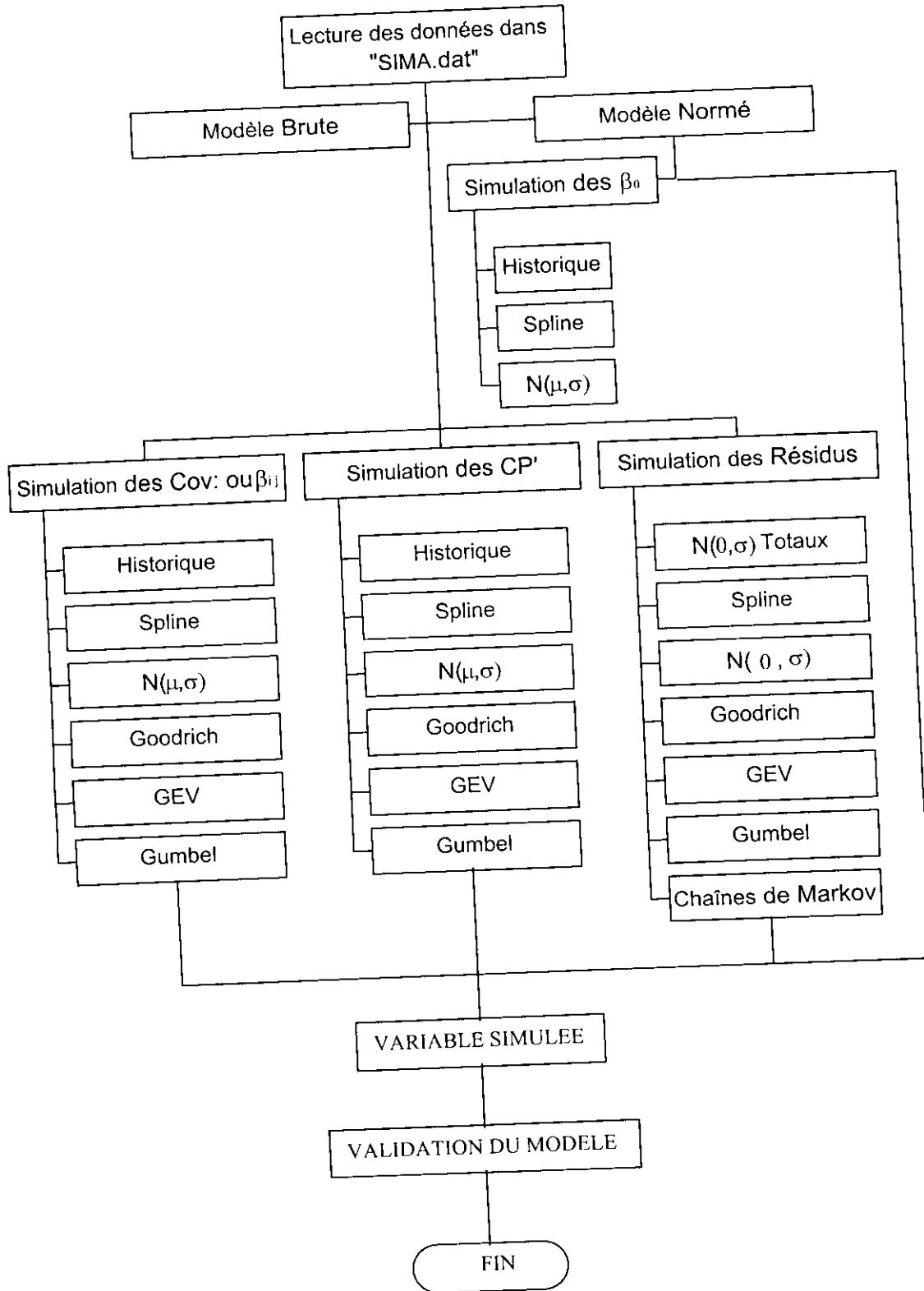


Figure IV-6 Sous programme SFACP.

Chapitre V : Application du Modèle de Simulation

V-1 DONNEES UTILISEES :

Les fichiers de données de base recueillis auprès de l'ONM sont au pas de temps journalier et mensuel pour la pluviométrie, alors que le fichier de l'évapotranspiration a été obtenu par calcul par la formule de Penman.

Afin d'étudier la fiabilité du modèle élaboré, différents pas de temps ont été utilisés :

1- Le fichier de données ne subit aucune transformation, et est utilisé tel qu'il a été recueilli (c'est à dire sous forme matricielle).

2- Pas de temps journalier 365 :

L'introduction des données dans ce fichier se fait ligne par ligne dont chacune d'elle représente une année d'observation (individu). Pour s'affranchir du problème de la colonne incomplète que présente les années bissextiles (366 valeurs au lieu de 365), on a préféré éliminer la valeur du 29 février mais cette élimination n'est pas faite de facto ; En ce qui concerne les variables cycliques telles que la température, l'ETP, etc...., on peut enlever ces valeurs sans grande incidence, alors que pour les variables aléatoires (la pluviométrie), ces valeurs sont reportées sur les jours d'avants, et cela n'avait aucune influence sur le fichier de données du fait que les valeurs n'étaient pas importantes.

3- Pas de temps journalier ACP :

Le pas de temps proposé précédemment présente l'inconvénient de ne pas tenir compte de l'effet saisonnier que présente une année hydrologique, en plus de la difficulté à expliquer l'apport des CP du fait de leur faible contribution, c'est pour cela qu'on a opté pour un fichier où l'effet saisonnier est bien mis en évidence, en effet le fichier proposé est composé de douze (12) sous fichiers représentant chacun un mois de l'année, l'introduction des données est faite ligne par ligne dont chacune d'elle représente les valeurs journalières du mois considéré de chaque année observée.

4- Pas de temps décadaire :

Le passage au pas de temps décadaire permet de formuler deux sortes de fichiers selon le type de variable ou du phénomène étudié. Les fichiers des données décadaires cumulées se prêtent bien aux phénomènes aléatoires.

L'introduction des données est faite ligne par ligne, chacune d'elle étant constituée de trente six (36) décades, et chaque décade représente le cumul de dix (10) valeurs journalières observées sauf pour la dernière décade du mois considéré où l'on cumule les valeurs du nombre de jours restants (8, 9, 10 ou 11 jours). Pour le fichier décadaire moyen qui se prête bien au phénomène cyclique (ETP), chaque décade représente la moyenne décadaire journalière. Chaque ligne du fichier décadaire représente une année d'observation.

5- Pas de temps mensuel :

De même que pour le pas de temps décadaire, il est possible d'avoir deux (2) sortes de fichiers selon le type de variable ou de phénomène étudié. Le fichier mensuel cumulé se prête bien aux phénomènes aléatoires alors que le fichier mensuel moyen se prête bien lui aux phénomènes cycliques.

On introduit les données ligne par ligne dont chacune contient douze (12) valeurs, dans le premier fichier, chaque valeur représente le cumul des valeurs journalières de chaque mois, alors que dans le second fichier, chaque valeur représente la moyenne mensuelle journalière du mois considéré. Chaque ligne du fichier mensuel représente une année d'observation.

V-2 TRANSFORMATION DES DONNEES :

Les fichiers des données brutes pour des pas de temps différents, utilisés dans le modèle élaboré, donnent lieu parfois à des valeurs simulées négatives ce qui est physiquement impossible pour une variable météorologique, c'est pour cela que l'on propose deux (2) possibilités de transformation des données brutes X, en une nouvelle variable Y :

1 - Transformation Log-Néperien (LN) : $Y = \text{LN } X$

2 - Transformation racine carrée : $Y = \sqrt{X}$

V-3 SIMULATION DES VARIABLES CLIMATOLOGIQUES :

Afin de pouvoir apprécier, la fiabilité du modèle de simulation élaboré, nous avons considéré deux variables climatologiques, l'une aléatoire (la pluviométrie), et l'autre cyclique (l'ETP).

V-3-1 Simulation de la pluviométrie :

Nous simulons les fichiers de la pluviométrie de la station d'Alger à différents pas de temps.

La figure V-1 représentant l'évolution de la pluviométrie journalière à cette station, montre l'aspect aléatoire du phénomène étudié.

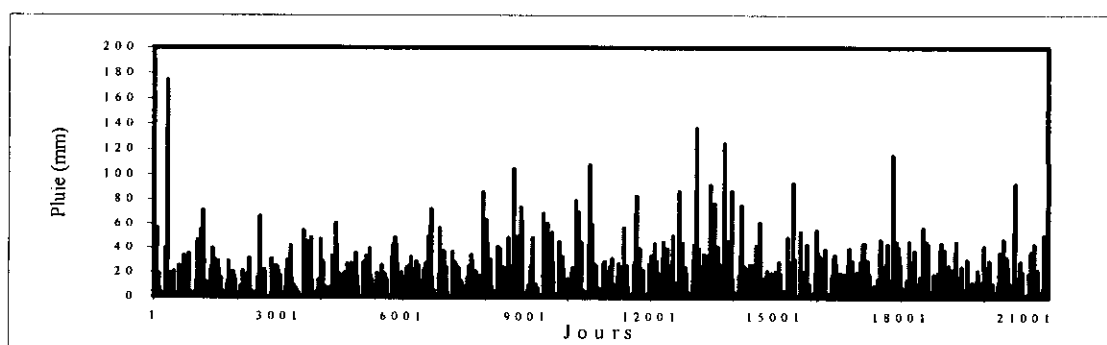


Figure V-1 Evolution de la pluviométrie journalière.

V-3-1-1 Simulation au pas de temps mensuel :

a/- Modèle normé :

En effectuant une analyse en composantes principales sur le fichier des pluies mensuelles cumulées, nous remarquons qu'avec quatre (04) CP seulement , nous arrivons à expliquer 52,1 % de la variance totale, (Tableau V-1).

Tableau V-1 : Contribution de chaque CP à la variance totale.

N° des CP	1	2	3	4	5	6	7	8	9	10	11	12
Valeurs propres	1.88	1.66	1.44	1.27	1.22	0.96	0.85	0.74	0.58	0.57	0.48	0.34
% des CP	15.7	13.8	12.0	10.6	10.1	8.00	7.10	6.20	4.90	4.70	4.00	2.90
% des CP cumul.	15.7	29.5	41.5	52.1	62.2	70.2	77.3	83.5	88.4	93.1	97.1	100.

L'ajustement des différents paramètres constituant le modèle normé à savoir $\beta_0, \beta_{ij}, CP'_j$ et ϵ_{ij} montre que :

- Les quatre composantes principales (CP) prises en considération suivent les lois de Gumbel, et de la General Extrem Value (GEV).

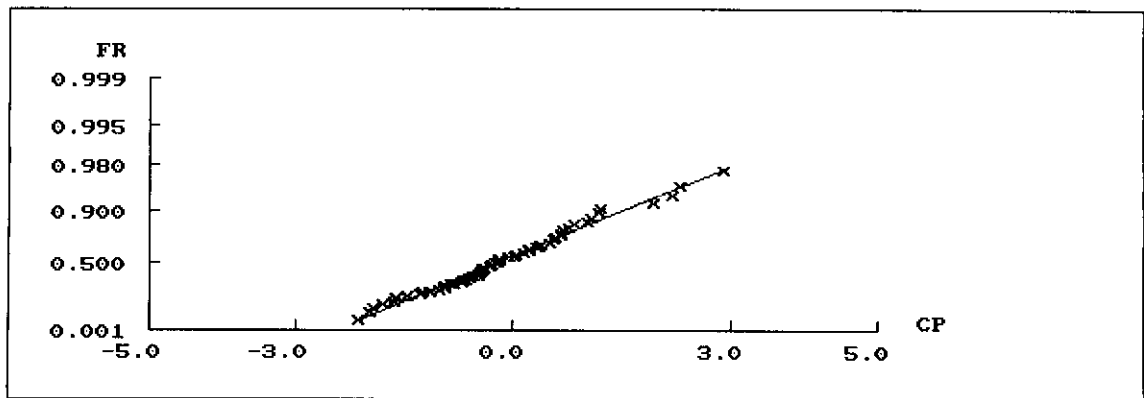


Figure V-2 Ajustement par la loi de Gumbel : CP1

- Les coefficients de régression (β_{ij}), ne suivent aucune loi en période humide, et sont presque nuls en période sèche, ils seront simulés par les fonctions splines ou à partir de l'historique.

- Les résidus (ϵ_{ij}) suivent la loi normale (figure V-3) en période humide et sont presque nuls en été, c'est pour cela qu'il faut les considérer sous forme vectorielle pendant cette période, et les simuler soit par les fonctions splines soit par les chaînes de Markov.

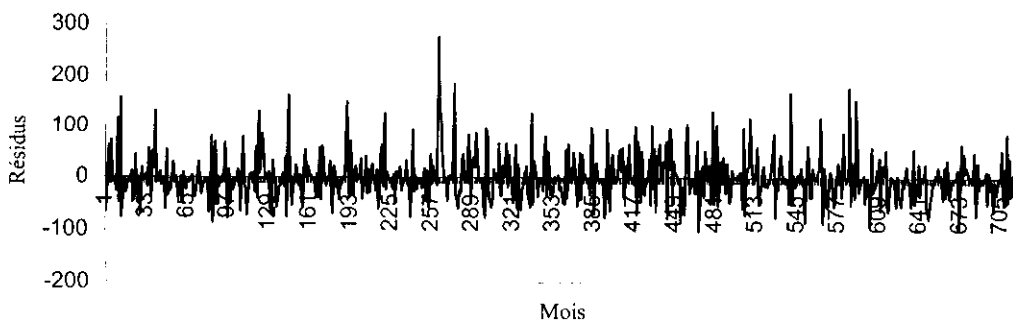


Figure V-3 Evolution des résidus des pluies mensuelles normées.

- Les pluies moyennes mensuelles cumulées (β_0) suivent la loi de Gumbel en période humide, et ne suivent aucune loi en période sèche (juillet et août), elles seront représentées dans ce cas soit par les valeurs de l'historique, soit simulées dans les fonctions splines.

Ces ajustements permettent de trouver le meilleur modèle de simulation pour la pluviométrie mensuelle .

b/- Modèle brute :

La même démarche est initiée pour le modèle brute qui n'est constitué que de trois paramètres (les covariances (β_{ij}), les composantes principales (CP) et les résidus (ϵ_{ij})).

Le tableau V-2 montre que les trois premières composantes principales représentent 56,6 % de la variance totale.

Tableau V-2 : Contribution de chaque CP à la variance totale.

N° des CP	1	2	3	4	5	6	7	8	9	10	11	12
Valeurs propres	6192	4840	3904	2994	2574	2416	1674	887	567	240	57	45
% des CP	23.5	18.3	14.8	11.3	9.8	9.2	6.3	3.4	2.1	0.9	0.2	0.20
% des CP cumul.	23.5	41.8	56.6	67.9	77.7	86.9	93.2	96.6	98.7	99.6	99.8	100.

L'ajustement des trois paramètres constituant le modèle brute permet de noter que :

- Les trois (03) CP suivent la loi de Gumbel.
- Les covariances ne suivent aucune loi, elles seront simulées soit par les fonctions splines ou directement prises égales à l'historique.
- Les résidus suivent la loi normale en période humide, et ne suivent aucune loi en période sèche (juin-juillet-août-septembre), ils seront simulés de la même manière que les covariances, ainsi que par les chaînes de Markov.

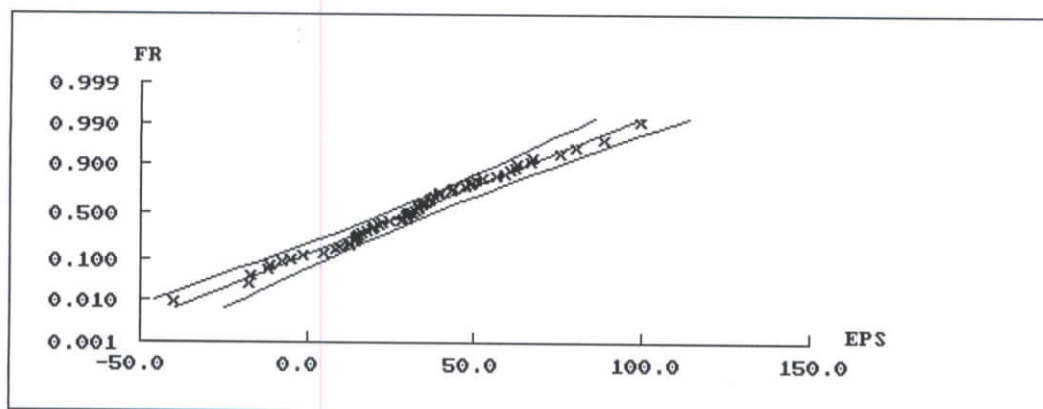


Figure V-4 Ajustement par la loi Normale : EPS2

V-3-1-2 Simulation au pas de temps décadaire :

a/- Modèle normé :

En procédant à l'analyse en composantes principales du fichier des pluies décadaires cumulées nous remarquons qu'avec neuf (09) CP sur un total de trente six (36), nous approchons les 54 % de la variance totale.

Tableau V-3 : Contribution de chaque CP à la variance totale.

N° des CP	1	2	3	4	5	6	7	8	9
Valeurs propres	3.4	2.6	2.4	2.2	1.9	1.9	1.8	1.6	1.5
% des CP	9.5	7.3	6.7	6.2	5.4	5.2	5.0	4.4	4.2
% des CP cumul.	9.5	16.8	23.5	29.7	35.1	40.3	45.3	49.7	53.9

L'ajustement des différents paramètres constituant le modèle normé montre que :

- Les neuf CP suivent la loi de Goodrich.

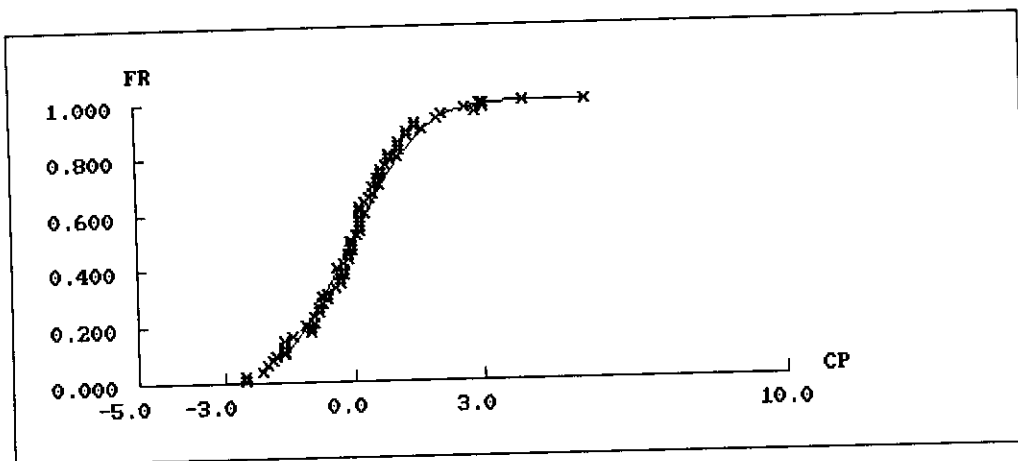


Figure V-5 Ajustement par la loi de Goodrich : CP8

- Les β_{ij} suivent la loi de Goodrich en période humide (figure V-6), et ne suivent aucune loi en période sèche (valeurs faibles), ils seront simulés dans ce cas par les fonctions splines.

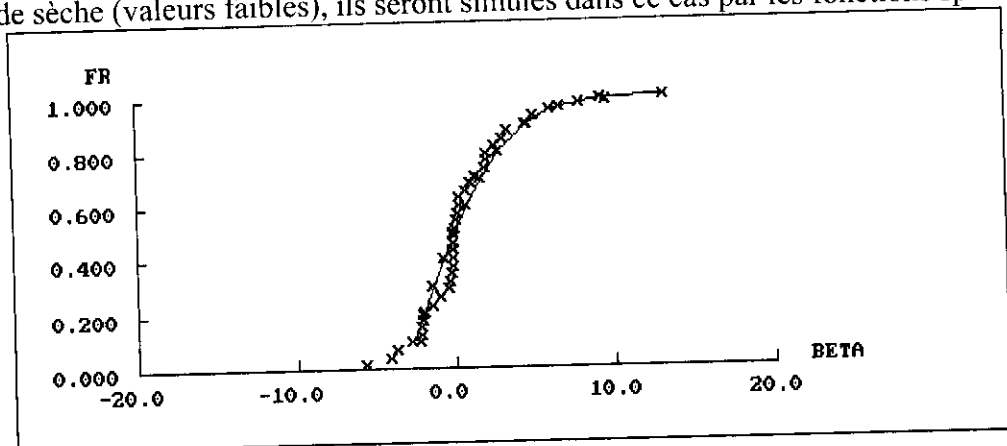


Figure V-6 Ajustement par la loi de Goodrich : BETA14

- Les ϵ_{ij} suivent la loi de Goodrich en période humide (figure V-7). En été, ils sont presque nuls et ne suivent aucune loi, ils seront simulés soit dans la fonction spline, soit par les chaînes de Markov.

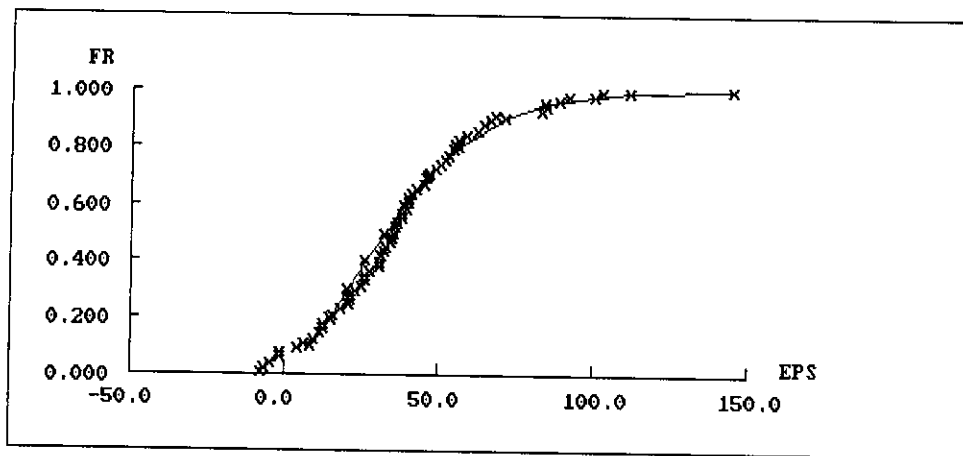


Figure V-7 Ajustement par la loi de Goodrich : EPS2

- Les pluies moyennes décadaires cumulées (β_0) ne suivent aucune loi, elles seront simulées soit dans les fonctions splines, soit prises directement dans l'historique.

b/- Modèle brute :

En effectuant une analyse en composantes principales sur le fichier de pluie nous avons noté qu'avec six (06) CP seulement, nous pouvons expliquer 52,2 % de la variance totale.

Tableau V-4 : Contribution de chaque CP à la variance totale.

N° des CP	1	2	3	4	5	6
Valeurs propres	2614	2402	2242	1795	1571	1432
% des CP	11.3	10.4	9.7	7.8	6.8	6.2
% des CP cumul.	11.3	21.7	31.4	39.2	46.0	52.2

L'ajustement des trois paramètres constituant le modèle montre que :

- Les six (CP) suivent les lois de Goodrich, et de Gumbel .

- Les covariances suivent la loi de Goodrich en période humide (figure V-8) et ne suivent aucune loi en période sèche, elles seront simulées alors par les fonctions splines ou prises directement dans l'historique.

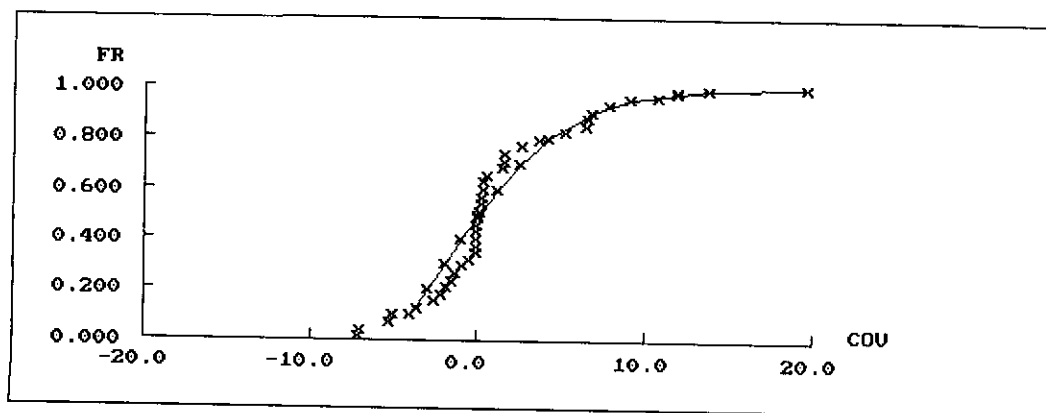


Figure V-8 Ajustement par la loi de Goodrich : COV11

- Les résidus suivent la loi de Goodrich en période humide (figure V-9), et sont presque nuls en été, ils seront simulés dans ce cas soit par les fonctions splines, soit par les chaînes de Markov.

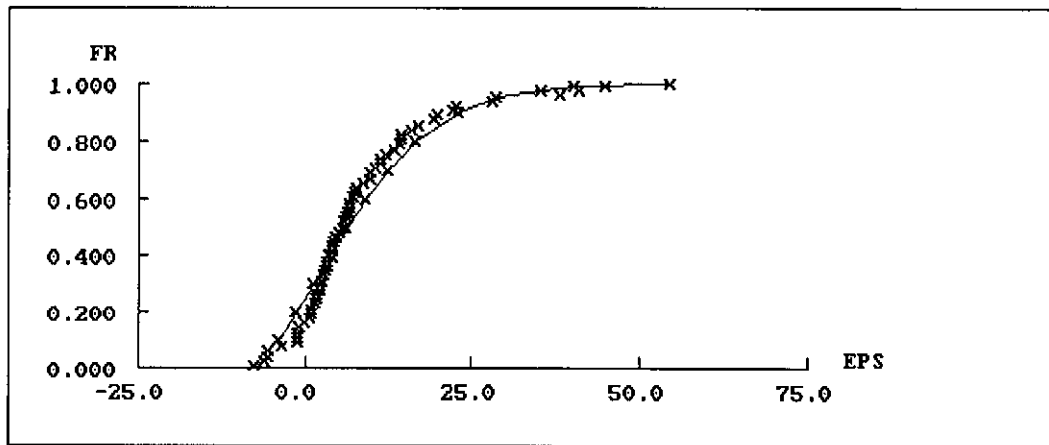


Figure V-9 Ajustement par la loi de Goodrich : EPS26

V-3-1-3 Simulation au pas de temps journalier :

Lorsque le pas de temps diminue le nombre de tests à effectuer sur les différents paramètres constituant les modèles normé et brute devient très vite très important.

a/- Modèle normé :

Le passage du pas de temps décadaire au pas de temps journalier ne se passe pas sans incidence sur le nombre de tâches à effectuer, en effet cette étape nécessite de procéder à l'analyse en composantes principales pour chaque mois, de choisir le nombre de CP avec lequel nous arrivons à expliquer plus de 50 % de la variance totale, ensuite d'ajuster les différents paramètres constituant le modèle normé.

Vu le nombre très élevé de tests effectués, les résultats de la variable étudiée sont synthétisés.

D'après le tableau V-5, nous remarquons que le nombre de CP à prendre en considération pour expliquer plus de 50 % de la variance totale varie de 5 à 8 pour tous les mois de l'année.

Tableau V-5 Contribution de chaque CP à la variance totale

	N° des CP	1	2	3	4	5	6	7	8
Janvier	Val. Propres	3.5	2.9	2.4	2.1	2.0	1.8	1.7	1.7
	%	11.3	9.4	7.8	6.7	6.4	6.0	5.6	5.5
	% cumulé	11.3	20.7	28.5	35.2	41.6	47.6	53.2	58.7
Février	Val. Propres	4.3	3.2	2.7	2.1	1.8	1.7	1.4	1.3
	%	15.5	11.5	9.5	7.4	6.5	6.0	5.0	4.7
	% cumulé	15.5	27.0	36.5	43.9	50.4	56.4	61.4	66.1
Mars	Val. Propres	3.5	2.6	2.5	2.2	2.1	1.8	1.7	1.5
	%	11.3	8.3	8.0	7.1	6.7	5.8	5.4	5.0
	% cumulé	11.3	19.6	27.6	34.7	41.4	47.2	52.6	57.6
Avril	Val. Propres	2.7	2.6	2.4	2.1	2.1	2.0	1.7	1.5
	%	9.0	8.7	7.9	7.0	6.9	6.7	5.7	5.1
	% cumulé	9.0	17.7	25.6	32.6	39.5	46.2	51.9	57.0
Mai	Val. Propres	4.0	3.4	2.6	2.2	2.1	1.8	1.7	1.5
	%	12.9	10.9	8.4	7.2	6.6	5.8	5.4	4.8
	% cumulé	12.9	23.8	32.2	39.4	46.0	51.8	57.2	62.0
Juin	Val. Propres	3.2	2.7	2.5	2.0	1.8	1.5	1.3	1.3
	%	10.7	8.9	8.3	6.7	5.9	5.0	4.4	4.3
	% cumulé	10.7	19.6	27.9	34.6	40.5	45.5	49.9	54.2
Juillet	Val. Propres	4.9	2.9	2.5	2.0	2.0	2.0	2.0	1.5
	%	15.7	9.2	8.1	6.6	6.6	6.6	6.6	4.8
	% cumulé	15.7	24.9	33.0	39.6	46.2	52.8	59.4	64.2
Août	Val. Propres	3.7	3.0	2.9	2.2	2.2	2.0	1.6	1.4
	%	12.0	9.8	9.3	7.2	7.0	6.4	5.1	4.6
	% cumulé	12.0	21.8	31.1	38.3	45.3	51.7	56.8	61.4
Sep.	Val. Propres	2.8	2.6	2.3	2.1	1.9	1.8	1.8	1.6
	%	9.2	8.5	7.5	6.9	6.5	6.1	6.1	5.5
	% cumulé	9.2	17.7	25.2	32.1	38.6	44.7	50.8	56.3
Oct.	Val. Propres	4.1	2.9	2.7	2.3	2.2	1.8	1.6	1.4
	%	13.2	9.5	8.6	7.3	7.1	6.0	5.2	4.4
	% cumulé	13.2	22.7	31.3	38.6	45.7	51.7	56.9	61.3
Nov.	Val. Propres	3.3	3.1	2.6	2.5	1.9	1.9	1.8	1.5
	%	11.0	10.3	8.5	8.2	6.4	6.3	6.1	4.9
	% cumulé	11.0	21.3	29.8	38.0	44.4	50.7	56.8	61.7
Déc.	Val. Propres	3.7	2.8	2.6	2.3	2.2	2.0	1.6	1.5
	%	11.8	9.1	8.5	7.5	7.1	6.6	5.2	5.0
	% cumulé	11.8	20.9	29.4	36.9	44.0	50.6	55.8	60.8

Les résultats des ajustements des différents paramètres constituant le modèle normé montrent que :

- Les CP prises en considération séparément pour chaque mois, suivent la loi de Goodrich pour la période humide (figure V-10), et ne suivent aucune loi durant la période sèche, elles seront simulées dans ce cas par les fonctions splines.

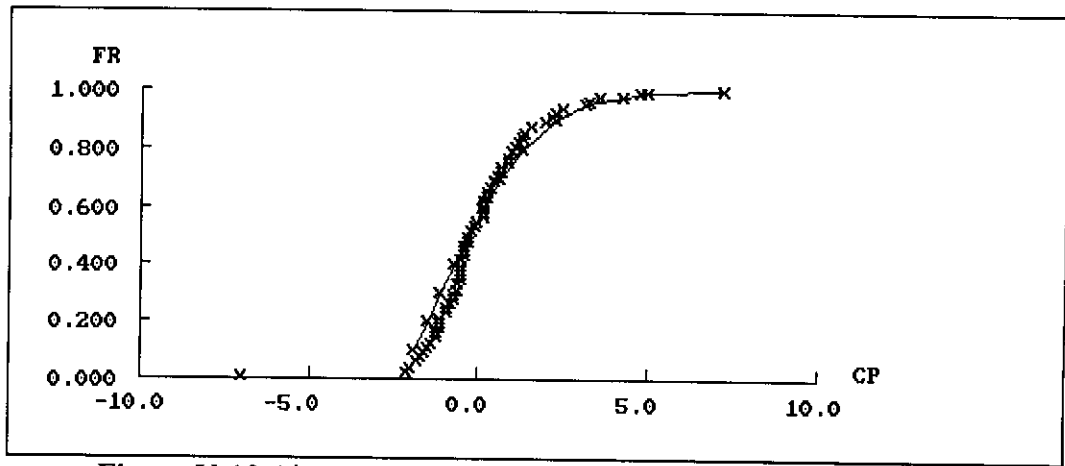


Figure V-10 Ajustement par la loi de Goodrich : CP2 Déc.

- Les β_{ij} suivent la loi de Goodrich pour la période humide (figure V-11), et sont presque nuls en été et ne suivent par conséquent aucune loi, ils seront simulés dans ce cas par les fonctions splines.

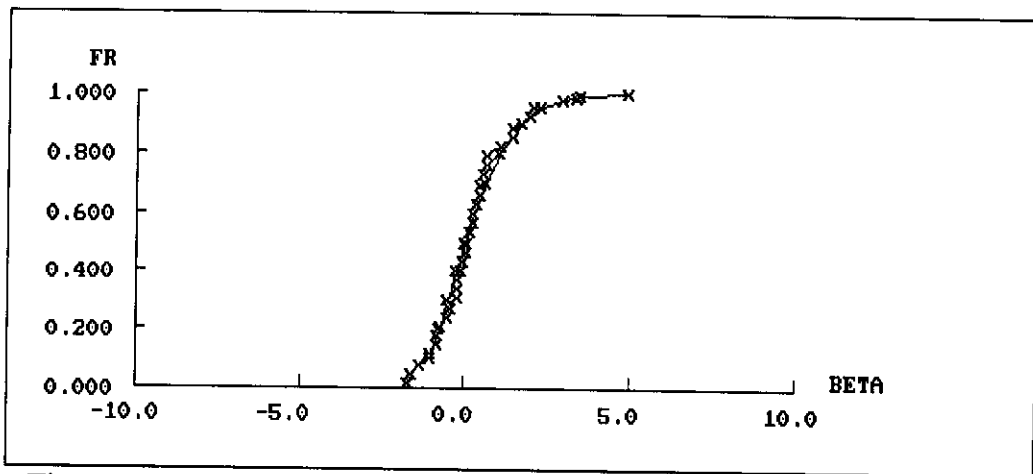


Figure V-11 Ajustement par la loi de Goodrich : BETA1 Mars.

- Les ε_{ij} suivent la loi de Goodrich (figure V-12), sauf pour les mois d'été où ils ne suivent aucune loi (valeurs très faibles), ils seront simulés soit par les fonctions splines, soit par les chaînes de Markov.

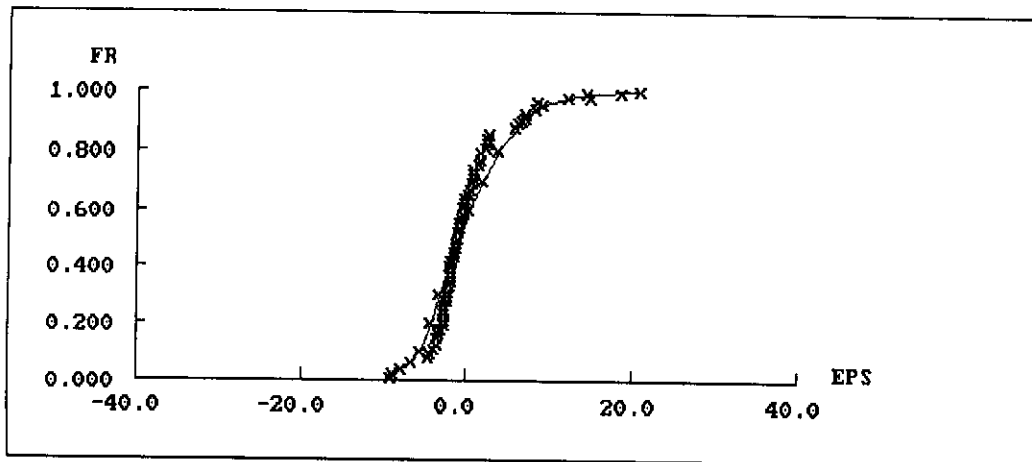


Figure V-12 Ajustement par la loi de Goodrich : EPS10 Jan.

- Les pluies moyennes journalières (β_0), étant presque nulles, ne suivent aucune loi, elles seront prises directement dans l'historique ou simulées par les fonctions splines.

b/- Modèle brute :

Les tâches réalisées pour ce type de modèle restent toutes aussi laborieuses que celles effectuées pour un modèle normé lorsque nous passons du pas de temps décadaire au pas de temps journalier, car le nombre de paramètres les séparant n'est que de un.

Le nombre de CP à prendre en considération pour expliquer plus de 50 % de la variance totale varie entre 2 et 6 pour tous les mois de l'année. (Tableau V-6).

Tableau V-6. Contribution de chaque CP à la variance totale

	N° des CP	1	2	3	4	5	6
Janvier	Val. Propres	188.	153.	132.	125.	116.	105.
	%	12.9	10.5	9.0	8.5	7.9	7.2
	% cumulé	12.9	23.4	32.4	40.9	48.8	56.0
Février	Val. Propres	252.	138.	120.	98.	87.	80.
	%	20.2	11.0	9.6	7.8	7.0	6.4
	% cumulé	20.2	31.2	40.8	48.6	55.6	62.0
Mars	Val. Propres	200.	139.	95.	72.	61.	48.
	%	20.5	14.3	9.7	7.4	6.3	4.9
	% cumulé	20.5	34.8	44.5	51.9	58.2	63.1
Avril	Val. Propres	269.	223.	133.	115.	71.	67.
	%	19.9	16.5	9.8	8.5	5.3	4.9
	% cumulé	19.9	36.4	46.2	54.7	60.0	64.9
Mai	Val. Propres	106.	97.	57.	48.	45.	42.
	%	17.8	16.3	9.6	8.1	7.6	7.1
	% cumulé	17.8	34.1	43.7	51.8	59.4	66.5
Juin	Val. Propres	76.	27.	16.	12.	11.	10.
	%	38.	13.5	8.0	6.1	5.6	5.0
	% cumulé	38.	51.5	59.5	65.6	71.2	76.2
Juillet	Val. Propres	11.	7.	5.	5.	2.	2.
	%	33.9	20.0	15.0	13.6	7.0	5.3
	% cumulé	33.9	53.9	68.9	82.5	89.5	94.8
Août	Val. Propres	20.	12.	12.	7.	1.	1.
	%	36.8	23.0	21.4	12.2	1.4	1.1
	% cumulé	36.8	59.8	81.2	93.4	94.8	95.9
Sep.	Val. Propres	288.	124.	61.	51.	42.	40.
	%	36.7	15.7	7.8	6.5	5.4	5.1
	% cumulé	36.7	52.4	60.2	66.7	72.1	77.2
Oct.	Val. Propres	326.	307.	229.	207.	170.	116.
	%	15.5	14.5	10.9	9.8	8.1	5.5
	% cumulé	15.5	30.0	40.9	50.7	58.8	64.3
Nov.	Val. Propres	423.	293.	218.	197.	160.	153
	%	18.9	13.1	9.7	8.8	7.1	6.8
	% cumulé	18.9	32.0	41.7	50.5	57.6	64.4
Déc.	Val. Propres	697.	254.	220.	211.	169.	130.
	%	27.0	9.8	8.5	8.2	6.5	5.0
	% cumulé	27.0	36.8	45.3	53.5	60.0	65.0

L'ajustement des différents paramètres constituant le modèle montre que :

- Les CP suivent la loi de Goodrich en période humide (figure V-13), et sont presque nulles en été, elles seront simulées dans ce cas par les fonctions splines.

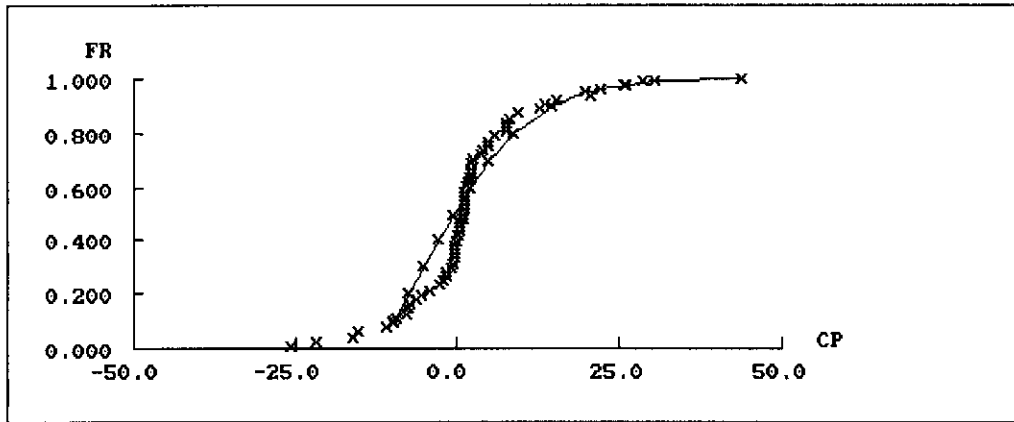


Figure V-13 Ajustement par la loi de Goodrich : CP4 Fév.

- Les covariances suivent la loi de Goodrich (figure V-14) sauf pour les mois secs ou ils ne suivent aucune loi, ils seront simulés par les fonctions splines.

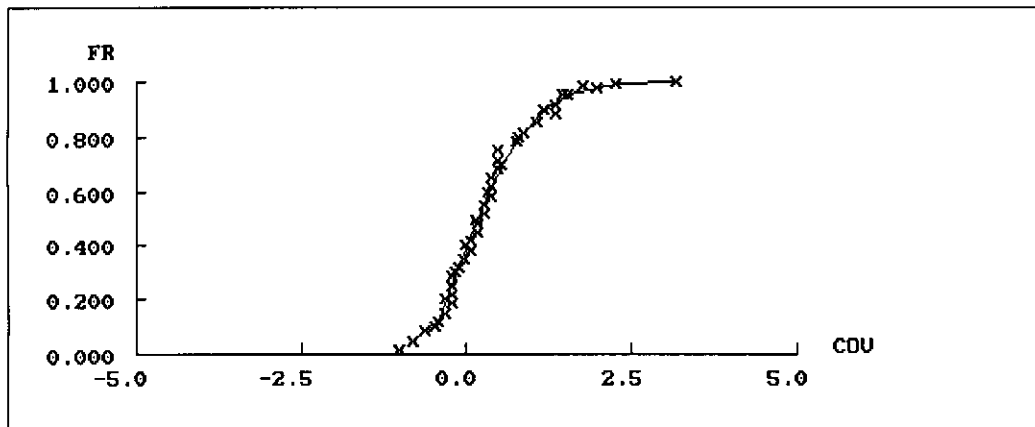


Figure V-14 Ajustement par la loi de Goodrich : COV1 Avr.

- Les ϵ_{ij} suivent la loi de Goodrich (figure V-15) sauf pour la période sèche, ils seront simulés soit par les fonctions splines soit par les chaînes de Markov.

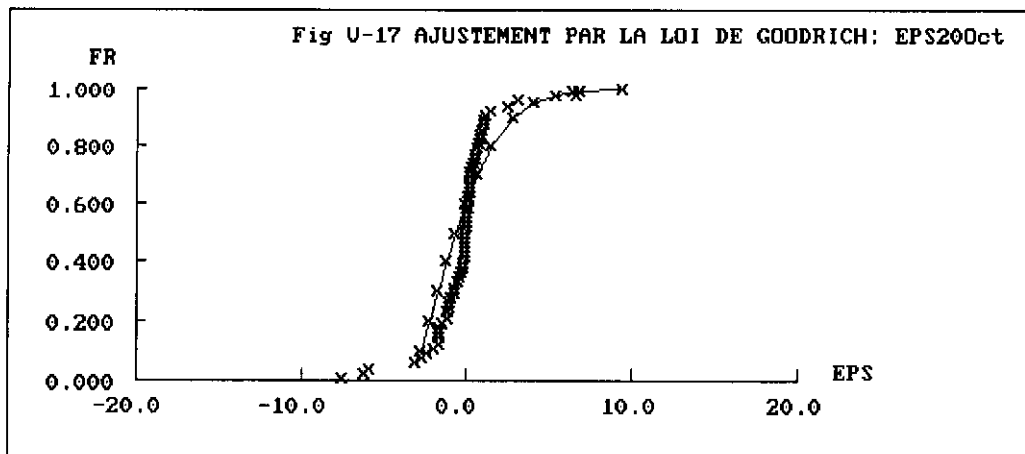


Figure V-15 Ajustement par la loi de Goodrich : EPS20 Oct.

V-3-2 Simulation de l'ETP :

Le fichier des évapotranspirations potentielles (ETP) calculé par la formule de Penman nous a permis de simuler ces dernières à différents pas de temps.

V-3-2-1 Simulation au pas de temps mensuel :

La figure V-16 , représentant l'évolution de l'ETP journalière à la station d'Alger, montre l'aspect cyclique de cette variable.

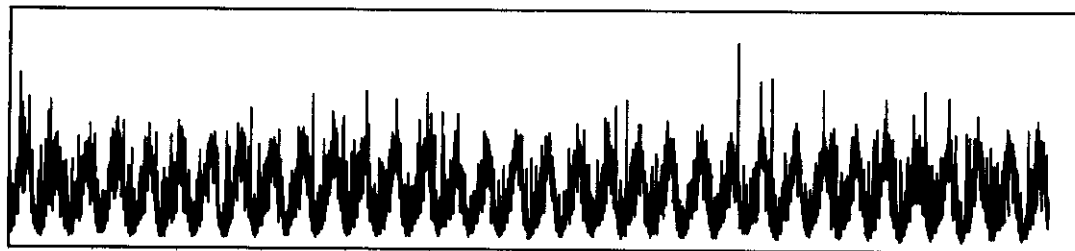


Figure V-16 Evolution de l'évapotranspiration journalière.

a/- Modèle normé :

En procédant à l'analyse en composantes principales du fichier des ETP mensuelles, nous remarquons qu'on approche les 57.4 % de la variance totale avec seulement quatre (04) CP.

Tableau V-7 Contribution de chaque CP à la variance totale

N° des CP	1	2	3	4
Valeurs propres	2.35	1.95	1.36	1.24
% des CP	19.6	16.2	11.3	10.3
% des CP cumul.	19.6	35.8	47.1	57.4

Les résultats des ajustements des différents paramètres du modèle montrent que :

- Les quatre CP suivent la loi normale (figure V-17), et la loi GEV.

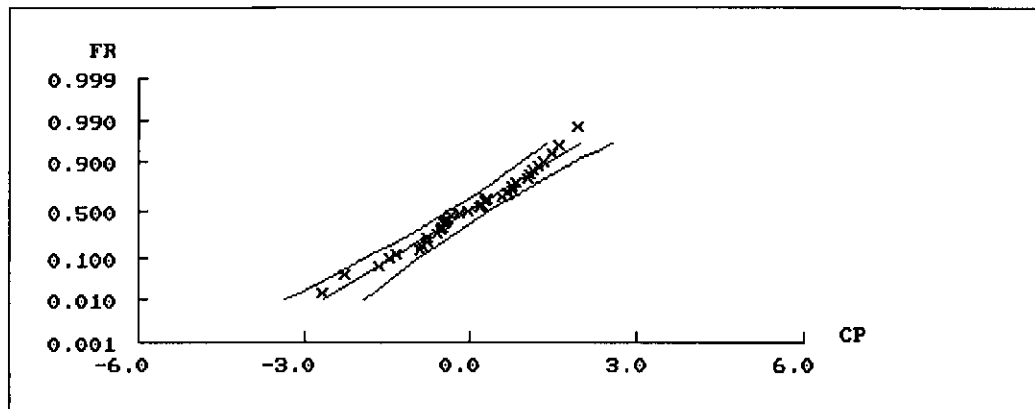


Figure V-17 Ajustement par la loi Normale : CP4.

- Les coefficients de régression (β_{ij}), ne suivent aucune loi en période humide, et sont presque nuls en période sèche, ils seront simulés soit par les fonctions splines soit à partir de l'historique.
- Les ε_{ij} suivent la loi GEV ou la loi normale (figure V-18).

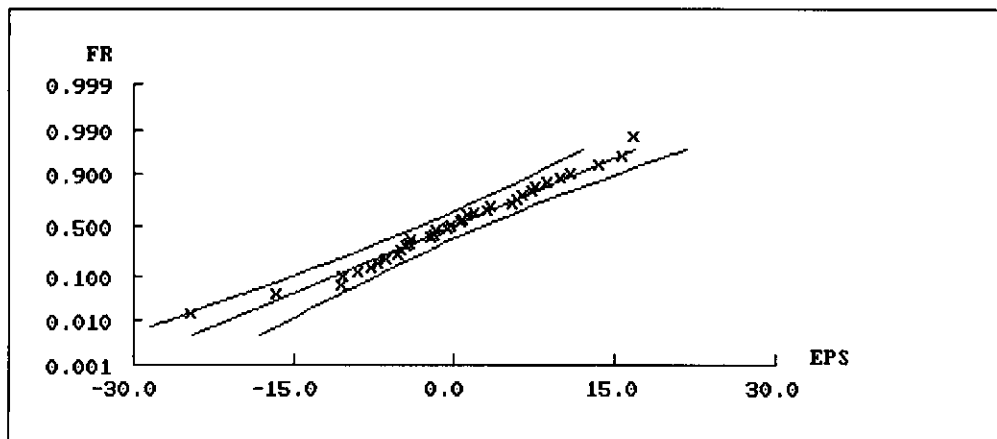
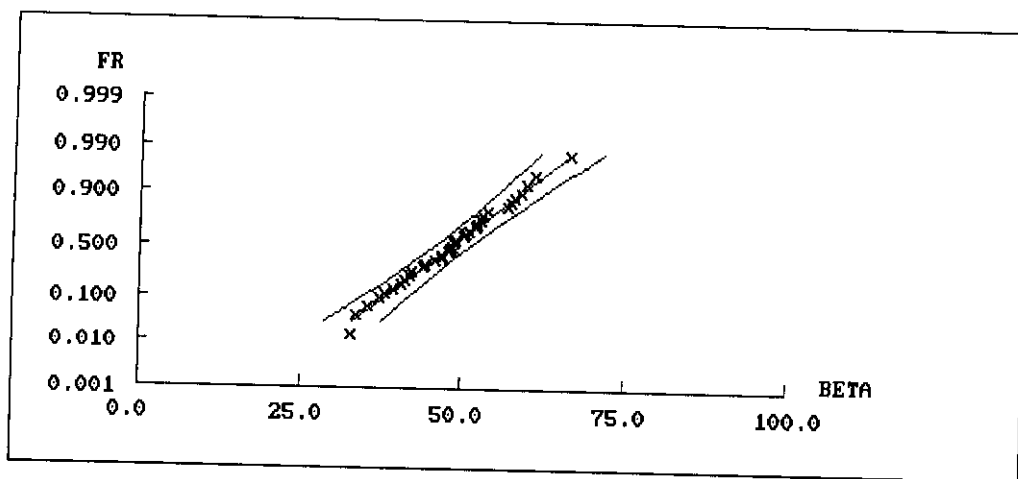


Figure V-18 Ajustement par la loi Normale : EPS5.

- Les β_0 suivent la loi GEV, ou la loi normale (figure V-19).



FigureV-19 Ajustement par la loi Normale : BETA01.

b/- Modèle brute :

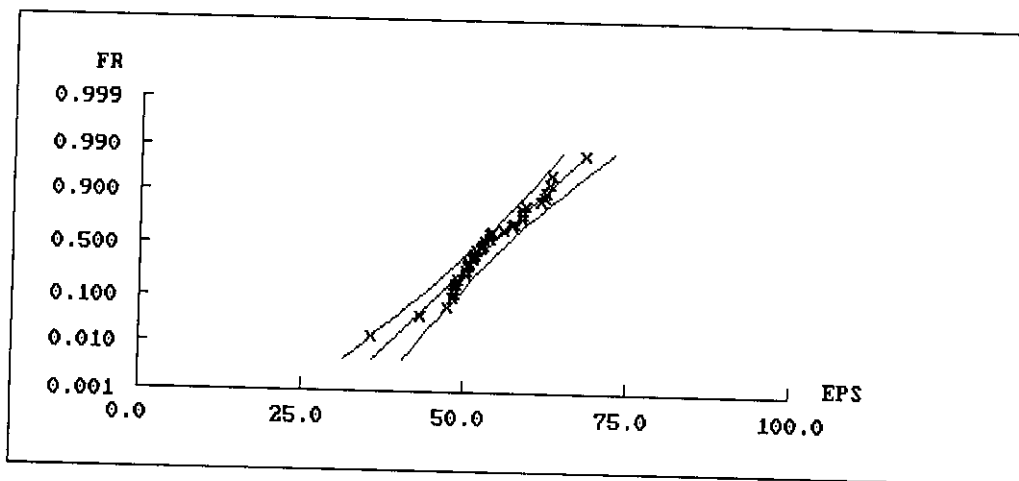
Avec trois (03) CP nous arrivons à expliquer 50,8 % de la variance totale.

Tableau V-8 Contribution de chaque CP à la variance totale

N° des CP	1	2	3
Valeurs propres	269.8	199.8	142.6
% des CP	22.4	16.6	11.8
% des CP cumul.	22.4	39.0	50.8

L'ajustement des paramètres du modèle montre que :

- Les trois CP suivent la loi GEV.
- Les covariances ne suivent aucune loi, elles seront simulées par les fonctions splines.
- Les ϵ_{ij} suivent les lois normales (figure V-20) et GEV.



FigureV-20 Ajustement par la loi Normale : EPS1.

V-3-2-2 Simulation au pas de temps décadaire :

a/- Modèle normé :

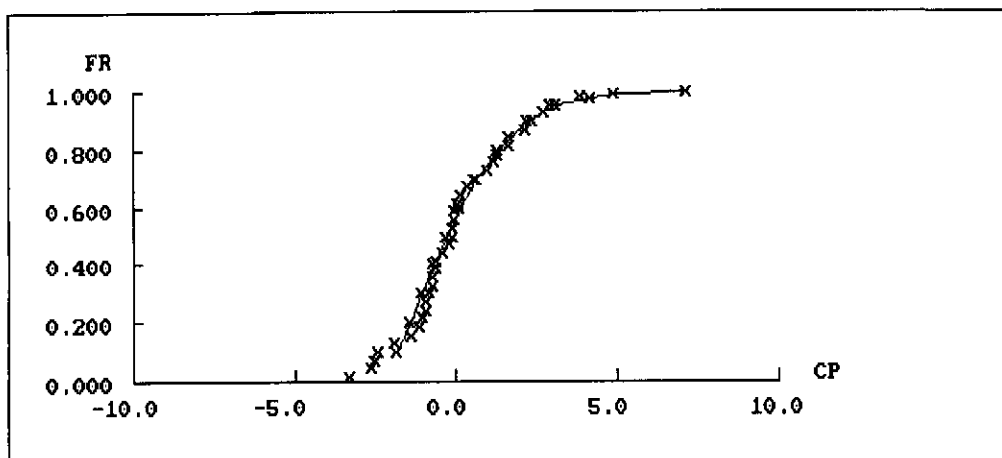
Pour cette variable, nous procédons à l'analyse en composantes principales du fichier des ETP décennaires. D'après le tableau V-9, six (06) CP seulement d'un total de trente six (36), permettent d'expliquer 52,3 % de la variance totale.

Tableau V-9 Contribution de chaque CP à la variance totale

N° des CP	1	2	3	4	5	6
Valeurs propres	4.53	3.65	3.04	2.90	2.42	2.29
% des CP	12.6	10.1	8.5	8.0	6.7	6.4
% des CP cumul.	12.6	22.7	31.2	39.2	45.9	52.3

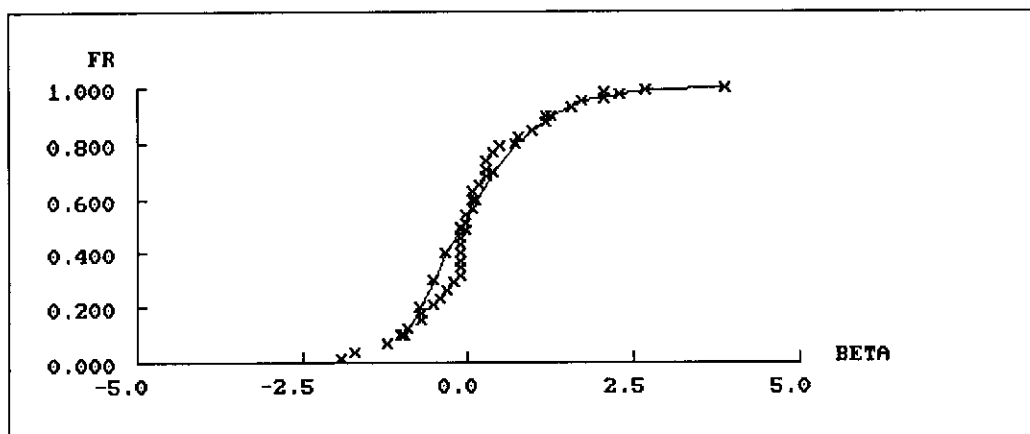
D'après les ajustements effectués nous avons noté que :

- Les six CP suivent la loi de Goodrich (figure V-21).



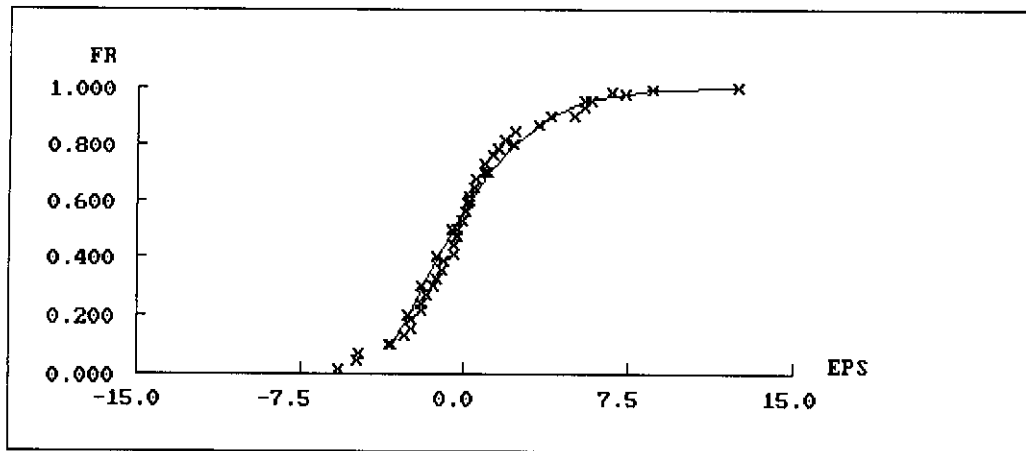
FigureV-21 Ajustement par la loi de Goodrich : CP4.

- Les β_{ij} suivent la loi de Goodrich (figure V-22).



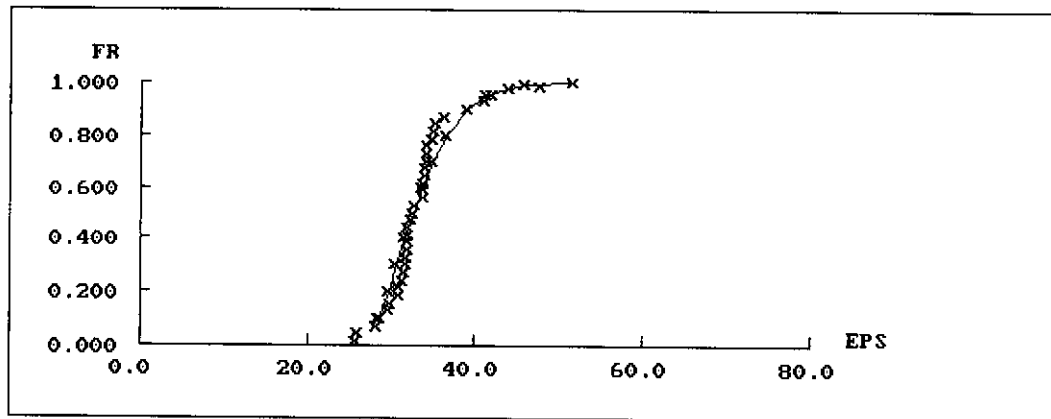
FigureV-22 Ajustement par la loi de Goodrich : BETA14.

- Les ϵ_{ij} suivent la loi de Goodrich (figure V-23).



FigureV-23 Ajustement par la loi de Goodrich : EPS20.

- Les β_0 suivent la loi de Goodrich (figure V-24).



FigureV-24 Ajustement par la loi de Goodrich : BETA011.

b/- Modèle brute :

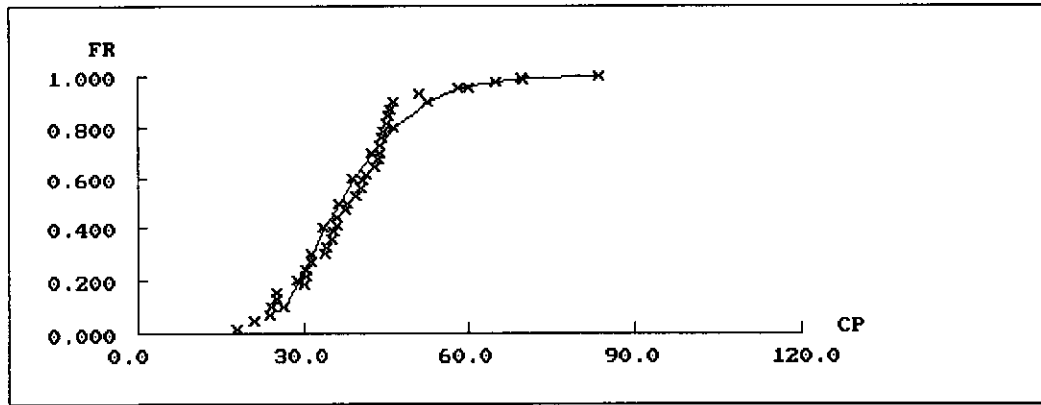
En effectuant une ACP sur la variable ETP, nous obtenons 54,9 % de la variance totale avec seulement six (06) CP (Tableau V-10).

Tableau V-10 Contribution de chaque CP à la variance totale

N° des CP	1	2	3	4	5	6
Valeurs propres	112.6	76.9	68.8	64.0	53.9	45.3
% des CP	14.6	10.0	8.9	8.4	7.0	5.9
% des CP cumul.	14.7	24.7	33.6	42.0	49.0	54.9

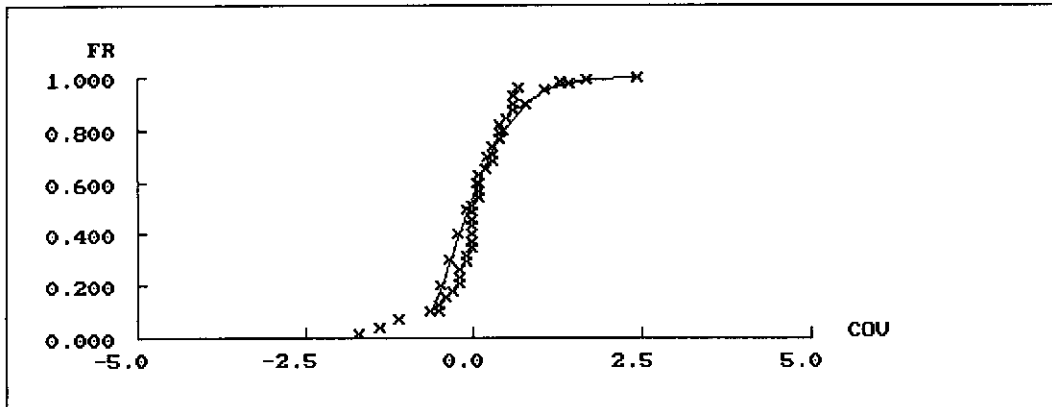
L'ajustement des paramètres du modèle montre que :

- Les six CP suivent la loi de Goodrich (figure V-25).



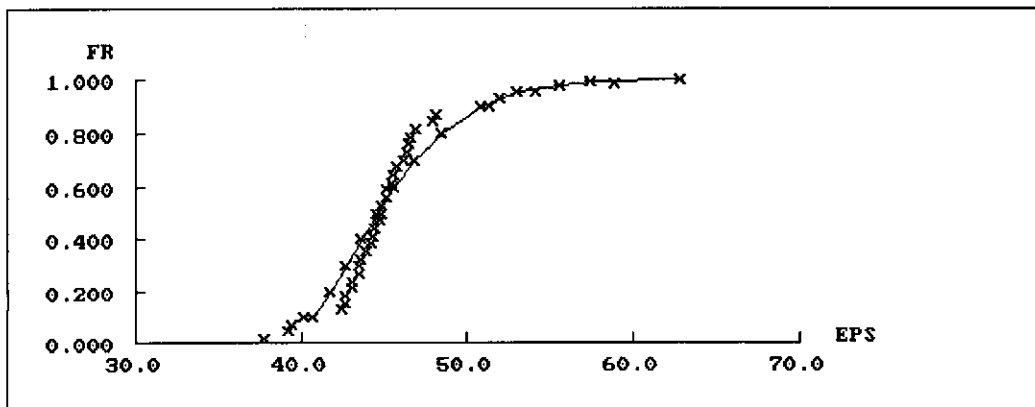
FigureV-25 Ajustement par la loi de Goodrich : CP1.

- Les covariances suivent la loi de Goodrich (figure V-26).



FigureV-26 Ajustement par la loi de Goodrich : COV26.

- Les résidus suivent la loi de Goodrich (figure V-27).



FigureV-27 Ajustement par la loi de Goodrich : EPS26.

V-3-2-3 Simulation au pas de temps journalier :

a/- Modèle normé :

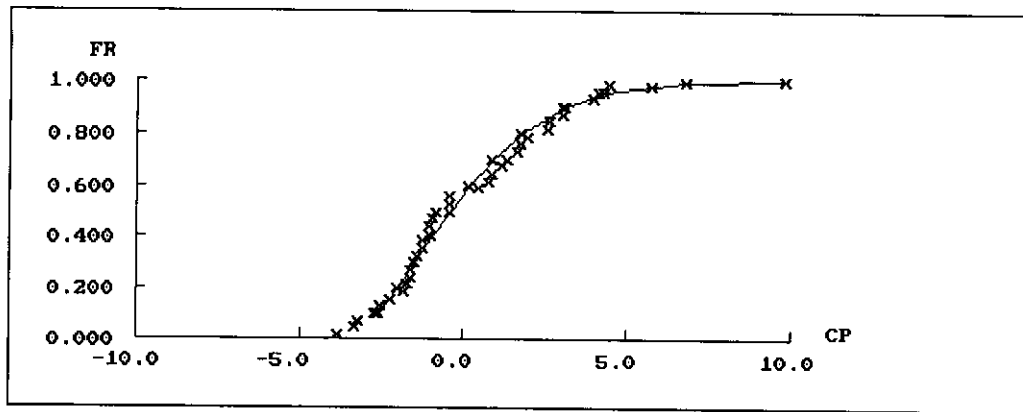
D'après le tableau V-11, le nombre de CP à prendre en considération pour expliquer plus de 50 % de la variance totale varie entre 4 et 6 pour tous les mois de l'année.

Tableau V-11 Contribution de chaque CP à la variance totale

	N° des CP	1	2	3	4	5	6
Janvier	Val. Propres	5.0	4.0	3.0	3.0	2.0	2.0
	%	15.2	12.9	10.5	8.2	7.7	6.2
	% cumulé	15.2	28.1	38.6	46.8	54.5	60.7
Février	Val. Propres	5.0	4.0	3.0	2.0	2.0	2.0
	%	18.9	12.8	10.1	8.7	8.1	7.0
	% cumulé	18.9	31.7	41.8	50.5	58.6	65.6
Mars	Val. Propres	5.0	4.0	3.0	3.0	2.0	2.0
	%	15.7	12.2	10.0	8.8	7.7	6.0
	% cumulé	15.7	27.9	37.9	46.7	54.4	60.4
Avril	Val. Propres	4.0	3.0	3.0	2.0	2.0	2.0
	%	12.5	11.1	9.4	8.0	8.0	6.8
	% cumulé	12.5	23.6	33.0	41.0	49.0	55.8
Mai	Val. Propres	6.0	4.0	3.0	2.0	2.0	2.0
	%	19.7	11.7	9.7	7.2	6.4	5.6
	% cumulé	19.7	31.4	41.1	48.3	54.7	60.3
Juin	Val. Propres	5.0	4.0	3.0	3.0	2.0	2.0
	%	15.6	11.9	10.0	9.0	8.2	6.3
	% cumulé	15.6	27.5	37.5	46.5	54.7	61.0
Juillet	Val. Propres	5.0	3.0	3.0	2.0	2.0	2.0
	%	16.4	10.2	8.4	7.4	7.2	6.0
	% cumulé	16.4	26.6	35.0	42.4	49.6	55.6
Août	Val. Propres	5.0	3.0	3.0	2.0	2.0	2.0
	%	16.5	10.8	8.6	7.2	6.5	6.0
	% cumulé	16.5	27.3	35.9	43.1	49.6	55.6
Sep.	Val. Propres	4.0	4.0	3.0	2.0	2.0	2.0
	%	14.8	11.8	9.1	7.7	6.9	6.3
	% cumulé	14.8	26.6	35.7	43.4	50.3	56.6
Oct.	Val. Propres	6.0	4.0	3.0	3.0	2.0	2.0
	%	19.2	12.1	9.5	8.2	7.9	6.2
	% cumulé	19.2	31.3	40.8	49.0	56.9	63.1
Nov.	Val. Propres	5.0	4.0	3.0	3.0	2.0	2.0
	%	18.1	12.0	9.3	8.9	7.6	6.2
	% cumulé	18.1	30.1	39.4	48.3	55.9	62.1
Déc.	Val. Propres	6.0	4.0	3.0	3.0	2.0	2.0
	%	19.0	12.6	10.7	8.7	7.2	6.9
	% cumulé	19.0	31.6	42.3	51.0	58.2	65.1

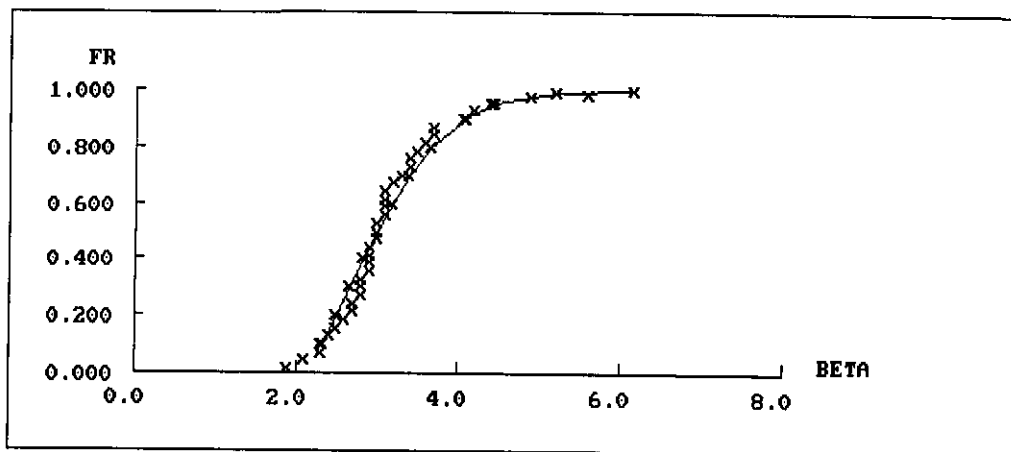
Nous avons noté lors des ajustements que :

- Les CP suivent la loi de Goodrich (figure V-28).



FigureV-28 Ajustement par la loi de Goodrich : CP1Nov.

- Les β_{ij} sont presque nuls et par conséquent ne suivent aucune loi, ils seront simulés soit par les fonctions splines, soit pris directement dans l'historique.
- Les ε_{ij} sont aussi presque nuls et ne suivent aucune loi, ils seront simulés soit par les fonctions splines, soit par les chaînes de Markov.
- Les β_0 suivent la loi de Goodrich (figure V-29).



FigureV-29 Ajustement par la loi de Goodrich : BETA01Avr.

b/- Modèle brute :

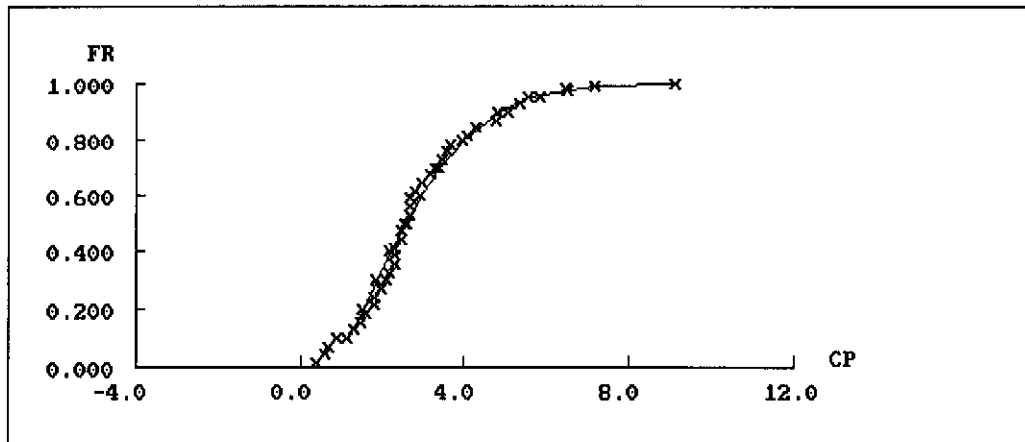
D'après le tableau V-12, le nombre de CP à prendre en considération pour expliquer plus de 50% de la variance totale varie entre 3 et 5 pour tous les mois de l'année.

Tableau V-12 Contribution des CP

	N° des CP	1	2	3	4	5
Janvier	Val. Propres	3.0	3.0	3.0	2.0	2.0
	%	16.4	13.5	12.7	9.0	7.8
	% cumulé	16.4	29.6	42.3	51.3	59.1
Février	Val. Propres	6.0	4.0	3.0	2.0	2.0
	%	24.3	15.5	11.7	9.0	6.8
	% cumulé	24.3	39.8	51.5	60.5	67.3
Mars	Val. Propres	4.0	3.0	2.0	2.0	2.0
	%	16.5	14.6	11.0	8.7	7.0
	% cumulé	16.5	31.1	42.1	50.8	57.8
Avril	Val. Propres	3.0	3.0	2.0	2.0	2.0
	%	14.2	13.5	9.4	8.5	7.8
	% cumulé	14.2	27.7	37.1	45.6	53.4
Mai	Val. Propres	6.0	3.0	3.0	2.0	2.0
	%	21.0	12.4	9.9	7.0	6.9
	% cumulé	21.0	33.4	43.3	50.3	57.2
Juin	Val. Propres	5.0	3.0	3.0	3.0	2.0
	%	16.7	11.7	10.0	9.3	8.0
	% cumulé	16.7	28.4	38.4	47.7	55.7
Juillet	Val. Propres	4.0	3.0	2.0	2.0	1.0
	%	19.7	13.1	8.4	7.0	5.9
	% cumulé	19.7	32.8	41.2	48.2	54.1
Août	Val. Propres	3.0	3.0	2.0	2.0	1.0
	%	15.3	13.3	8.9	8.4	7.2
	% cumulé	15.3	28.6	37.5	45.9	53.1
Sep.	Val. Propres	3.0	3.0	2.0	2.0	2.0
	%	16.7	12.7	9.3	8.7	7.9
	% cumulé	16.7	29.4	38.7	47.4	55.3
Oct.	Val. Propres	7.0	4.0	3.0	3.0	2.0
	%	21.8	12.7	10.9	8.8	7.6
	% cumulé	21.8	34.5	45.4	54.2	61.8
Nov.	Val. Propres	6.0	6.0	3.0	3.0	2.0
	%	20.6	19.5	11.0	9.0	7.2
	% cumulé	20.6	40.1	51.1	60.1	67.3
Déc.	Val. Propres	5.0	3.0	3.0	2.0	1.0
	%	21.8	14.6	11.9	9.7	5.4
	% cumulé	21.8	36.4	48.3	58.0	63.4

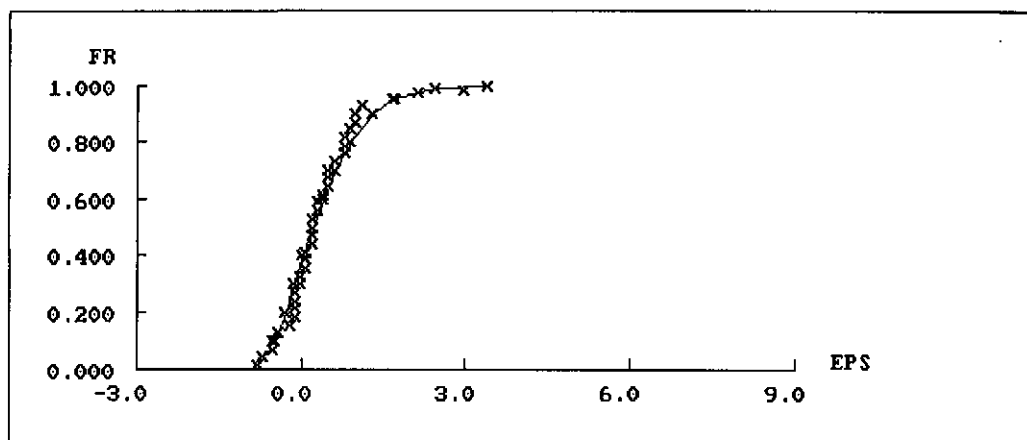
Les résultats des ajustements montrent que :

- Les CP suivent la loi de Goodrich (figure V-30).



FigureV-30 Ajustement par la loi de Goodrich : CP4Déc.

- Les covariances étant presque nuls, ne suivent aucune loi, ils seront simulés dans ce cas par les fonctions splines.
- Les ε_{ij} suivent la loi de Goodrich (figure V-31).



FigureV-31 Ajustement par la loi de Goodrich : EPS10Oct.

En conclusion, nous remarquons que le nombre de possibilités de simulation, ainsi que le nombre d'ajustements devient pléthorique.

Au pas de temps mensuel, le nombre de possibilités de simulation est calculé comme suit :

Pour un modèle normé constitué de quatre (04) paramètres (β_0 , β_{ij} , CP_j et ε_j), il faut effectuer 756 possibilités de simulation pour chaque CP, et pour chaque transformation de la variable brute, soit au total, si nous considérons onze (11) CP, et trois (03) transformations, 24948 possibilités de simulation.

Pour un modèle brute constitué de trois (03) paramètres le nombre de possibilités de simulation requis, et de 252 pour chaque CP, soit au total 8316 possibilités.

Il faut donc, au pas de temps mensuel, 33264 possibilités de simulation pour chaque variable météorologique étudiée.

Dans le cas du pas de temps décadaire le nombre de possibilités de simulation augmente davantage en effet :

Pour le modèle normé constitué de quatre (04) paramètres, il y a lieu d'effectuer 756 possibilités de simulation pour chaque CP, et pour chaque transformation de la variable brute comme pour le pas de temps mensuel, mais en considérant trente cinq (35) CP, et trois (03) transformations, soit au total 79380.

Pour un modèle brute constitué de trois (03) paramètres le nombre de possibilités de simulation requis est de 26460.

Il faut donc effectuer 105840 possibilités de simulation, au pas de temps décadaire pour chaque variable météorologique étudiée.

Nous remarquons que pour le pas de temps journalier le nombre de possibilités de simulation à effectuer augmente de manière exponentielle. En effet, les tests d'ajustements sont effectués pour chaque mois de l'année ce qui rend la tâche fastidieuse et rebutante voire même impossible à réaliser.

Le nombre de possibilités de simulation à effectuer pour le modèle normé est de 734832 et pour le modèle brute de 244944, soit un total de 979776 possibilités à réaliser pour chaque variable météorologique étudiée.

Nous constatons donc qu'en diminuant le pas de temps, le nombre d'ajustements devient pléthorique, et la tâche de l'utilisateur très laborieuse voir même impossible à réaliser d'où l'idée de simplifier la démarche en automatisant la procédure de simulation.

V-4 AUTOMATISATION DE LA PROCEDURE DE SIMULATION :

Le fait d'ajuster les paramètres des deux modèles normé et brute colonne par colonne, pour chaque transformation de la variable étudiée et pour chaque CP, augmente le nombre de possibilités de simulation surtout lorsque le pas de temps est réduit. Cette façon de procéder a constitué un handicap majeur pour la suite du travail entamé. Dans le but de simplifier cette démarche, nous avons amélioré le modèle élaboré en l'automatisant, et ce par l'introduction d'un fichier contenant toutes les informations nécessaires à la réalisation de toutes les possibilités de simulation.

Ce fichier appelé ACPS.DAT constitue le moteur de la procédure de simulation. En effet, après lecture des données brutes de la variable étudiée, celles-ci sont transformées à différents pas de temps selon le choix de l'utilisateur (Chap.V-1). Les composantes principales du fichier ainsi transformé sont calculées et présentées dans le fichier CP.dat. A chaque simulation, ACPS.DAT permet de créer un fichier appelé SIMA.dat, dans lequel tous les résultats nécessaires à la simulation sont consignés. Enfin les valeurs simulées de la variable étudiée sont présentées dans le fichier SIM.RES.

En outre, ACPS.DAT présente une panoplie de choix permettant de tester les différentes possibilités de simulation à savoir :

- Le nombre d'années de simulation désirées, l'année de début de simulation et le code de la station étudiée.
- Le choix de l'option utilisée (1-automatique ou 2-sinon) ainsi que la valeur du risque d'erreur (alpha) permettant de déterminer l'intervalle de confiance.
- Le type d'ACP (1-brute ou 2-normée).
- Le pas de temps (0-aucune transformation, 1-journalier ACP, 2-journalier 365, 3-décadaire cumulé, 4-décadaire moyen, 5-mensuel cumulé et 6-mensuel moyen) ainsi que le type de transformation (1-variable brute, 2-Log Néperien et 3-racine carrée).
- Le type de simulation de β_0 : 1-Historique, 2-Fonction spline et 3-Loi Normale,
- Le type de fréquence : 1-Hazen, 2-Weibull, 3-Chegodayev, 4-Cunane, 5-Gringorten, 6-Bloom et 7-Adamovsky, ainsi que le nombre de valeurs nécessaires au calcul de β_0 .
- Le nombre de composantes à prendre en considération : 27 CP max pour le fichier journalier, 35 pour le décadaire et 11 pour le mensuel.
- Le type de simulation des β : 1-Historique, 2-Spline, 3-Loi normale, 4-Goodrich, 5-GEV et 6-Gumbel.
- Le type de simulation des CP : 1-Historique, 2-Spline, 3-Loi normale, 4-Goodrich, 5-GEV et 6-Gumbel.
- Le type de simulation des résidus : 1-Historique, 2-Spline, 3-Loi normale, 4-Goodrich, 5-GEV, 6-Gumbel et 7-Chaînes de Markov.

Exemple : ACPS.DAT pour un fichier mensuel (pluviométrie station d'Alger).

```
Algp.dat
Algp.men
CP.dat
SIMA.dat
SIM.res
30 2000 300
1 1.96
1 5 1
2 1 10
1
4
5
7
```

Le fichier SIMA.dat dans lequel sont consignés tous les résultats nécessaires à la simulation devient un fichier de données du sous-programme SFACP, sa première partie est pratiquement identique à celle du fichier ACPS.dat, sa deuxième partie est obtenue à partir du sous-programme PSACP.

Exemple : SIMA.dat pour le même fichier mensuel.

```
Algp.men
Cp.dat
Sim.res
30 2000 300
1 1.960000
1 5 1
1
```

2 1 10
 1
 4
 5
 7

Les résultats obtenus à partir du sous-programme PSACP sont :

- Les moyennes et les écarts type de la variable étudiée.

12
 86.6525 76.0153 64.4136 65.4203 36.3458 15.3915 3.2034 4.1034 30.5000
 78.0491 90.8593 112.8441
 51.9146 57.7008 45.8813 47.5064 32.3426 17.6911 7.4140 7.7879 34.6688
 66.9939 63.2773 69.5439

- Les limites inférieures et supérieures des moyennes de la variable étudiée.

12
 57.2791 43.3680 38.4538 38.5411 18.0462 5.3818 -.9915 -.3030 10.8843
 40.1438 55.0568 73.4959
 116.0260 108.6625 90.3733 92.2996 54.6453 25.4012 7.3982 8.5098 50.1157
 115.9545 126.6618 152.1922

- Les limites inférieures et supérieures des écarts types de la variable étudiée.

12
 31.1445 34.6157 27.5250 28.4999 19.4028 10.6132 4.4478 4.6721 20.7984
 40.1907 37.9611 41.7205
 72.6848 80.7859 64.2377 66.5129 45.2823 24.7690 10.3802 10.9036 48.5392
 93.7970 88.5935 97.3672

- Matrice des Béta.

12
 -4.6843 -2.4899 -19.9264 36.9352 -6.4569 24.9262 15.0674 2.5235 -2.8941
 -1.2448 .0098 -.1863
 4.9403 -10.3950 49.4712 19.2091 -13.9527 3.8091 -10.7548 7.1525 -1.3771
 -1.6217 -.1488 -.1645
 3.3993 .9650 15.9342 8.6393 29.3397 -19.0655 20.9122 5.5249 -8.4926
 -1.1077 .1594 .0662
 -2.0124 -4.4695 -9.5752 30.2414 22.2908 -11.7827 -22.2474 -8.9058 4.0437
 .2518 .1237 .2423
 9.1763 5.4295 -11.9575 3.3521 4.2748 -6.3873 -1.6928 23.5463 12.6964
 .4714 -.4529 .1011
 2.8581 2.1851 5.0957 4.9574 -2.3839 2.9499 1.1624 -.2260 -.5055
 15.2712 -.3435 .3369
 -.3499 -.8884 .4747 .0481 -2.1223 1.0846 .3493 .9654 -.1881
 -4.847 3.0645 6.1136
 1.2976 -.6830 .2264 -.0852 -.0610 .1135 -.3662 1.2169 1.1642
 .9811 6.8617 -2.7323
 8.3195 -2.0566 16.9934 4.9166 -1.1654 -2.3264 19.1130 -12.4035 17.0138
 -.5380 -.0620 .1139
 24.9097 58.2197 -1.0950 10.5489 -14.1948 -12.2650 -.5365 -2.6408 -2.4056

-.8434 .1268 -.0038
 -37.5647 35.6646 15.7622 -6.4336 20.9822 23.8176 -3.4923 1.0069 3.1128
 -.2046 .0532 .0486
 62.7296 -1.9364 2.5803 -7.6605 17.5707 22.4619 -4.0837 -1.1172 -.8143
 -.4554 -.0454 .0793

- Moyennes et écarts type des Béta.

12
 3.4650 3.8473 4.7729 -.1499 3.2131 2.6131 .6722 .6611 3.9932
 4.9850 4.3961 7.4424
 15.2307 16.9271 12.9052 14.3224 9.1553 4.5828 2.1211 2.2454 9.5857
 19.5169 18.5165 19.4737

- Les coefficients de Goodrich des Béta.

1
 2.628904E-03 -22.239980 5.680707E-01

- Moyennes et écarts type des CP.

12
 1.0178 1.3415 1.0856 2.3113 1.7820 1.6302 .0590 .8758 .6033
 -.4960 .5597 .5601
 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
 1.0000 1.0000 1.0000

- Les coefficients de Goodrich des CP.

1
 1.895482E-01 -9.502043E-01 4.799977E-01

- Moyennes et écarts type par colonne des résidus.

12
 91.4202 70.9871 60.9538 67.4685 27.0062 12.4825 3.5595 2.7827 22.0325
 52.6962 129.0925 48.9982
 51.7029 57.4889 45.7552 47.4638 31.0135 17.4587 7.4057 7.6790 33.6558
 62.1907 50.9207 30.0225

- Moyenne et écart type des totaux des résidus.

0.000000E+00 73.511120

- Les coefficients de Goodrich des résidus pour chaque colonne.

12
 1.177287E-06 -38.616050 3.651134E-01
 9.610380E-03 1.184502 9.251803E-01
 1.336415E-03 -6.463953 6.548159E-01
 1.155260E-03 -3.328450 6.476362E-01
 2.322231E-03 -18.587670 6.512486E-01
 5.317958E-02 -10.529400 1.065934
 -9999.000000 -9999.000000 -9999.000000
 1.091900E-01 -12.322830 1.187878
 4.582017E-03 -24.366870 7.293007E-01
 1.656069E-03 -34.226430 7.125218E-01

1.356886E-04 37.074390 5.223277E-01
 8.100967E-03 11.477750 7.804570E-01

- Les coefficients de Markov.

1.271186E-01 8.728814E-01

- Les moyennes et les écarts types des résidus inférieurs à 0 ou supérieurs à 1.

-6.268685 57.190130
 8.559944 53.805280

Une fois vérifiées par les différents tests de validation du modèle, toutes les possibilités de simulation sont présentées dans le fichier OPT avec appréciation des résultats obtenus.

En effet, nous considérons un modèle comme étant fiable lorsque 24 des 26 valeurs simulées (12 moyennes mensuelles + 12 écart types mensuels + 1 moyenne annuelle + 1 écart type annuel) au moins soient à l'intérieur de l'intervalle de confiance.

Exemples de simulation des variables pluie et ETP :

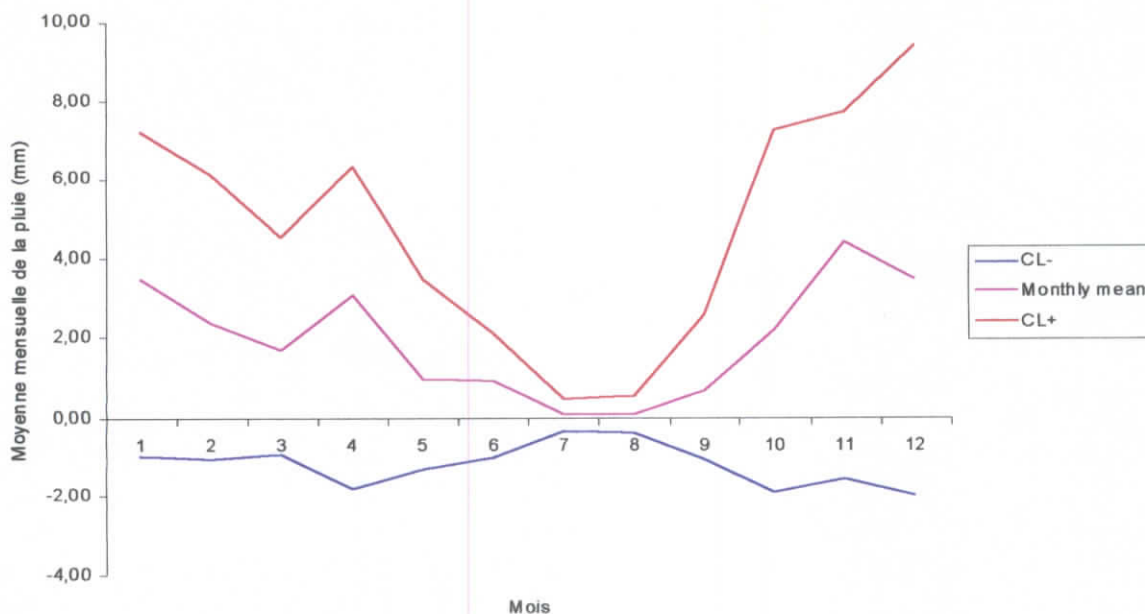


Figure V-32 Pluie moyenne mensuelle interannuelle simulée

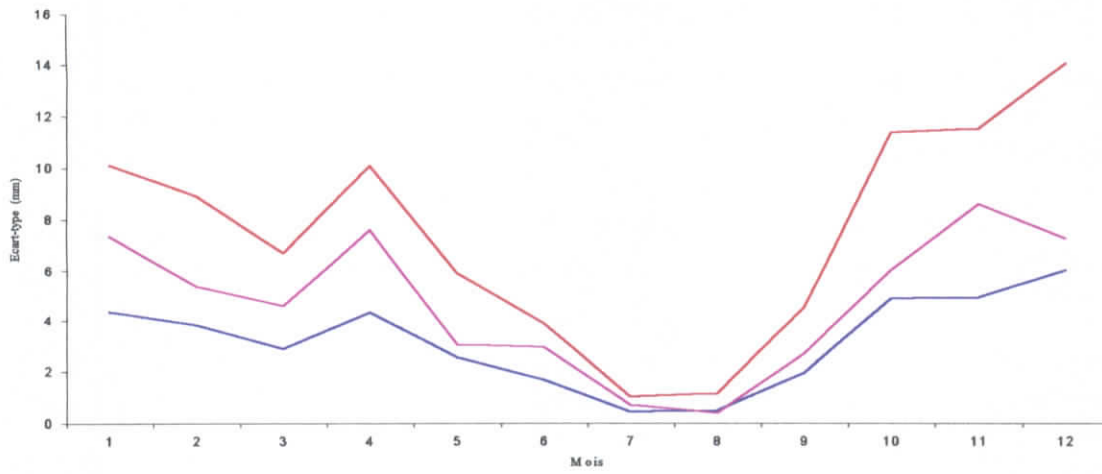


Figure V- 33 –Ecart type de la pluie moyenne mensuelle inter annuelle simulée.

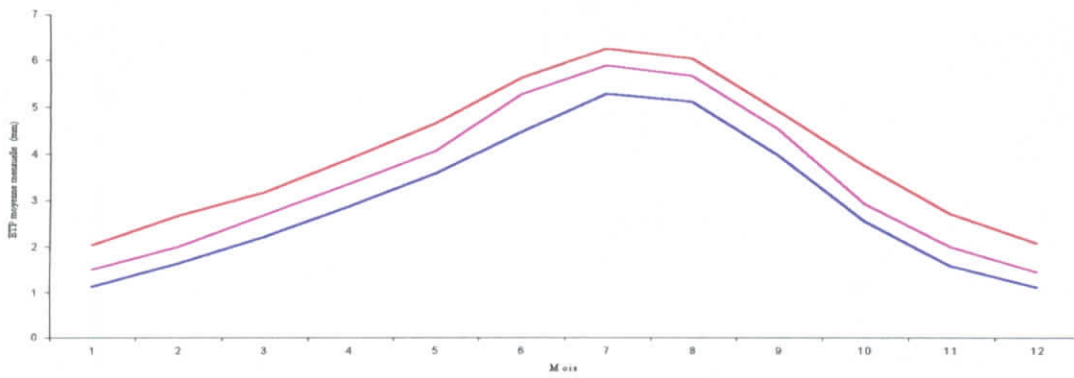


Figure V-34 - ETP moyenne mensuelle interannuelle simulée.

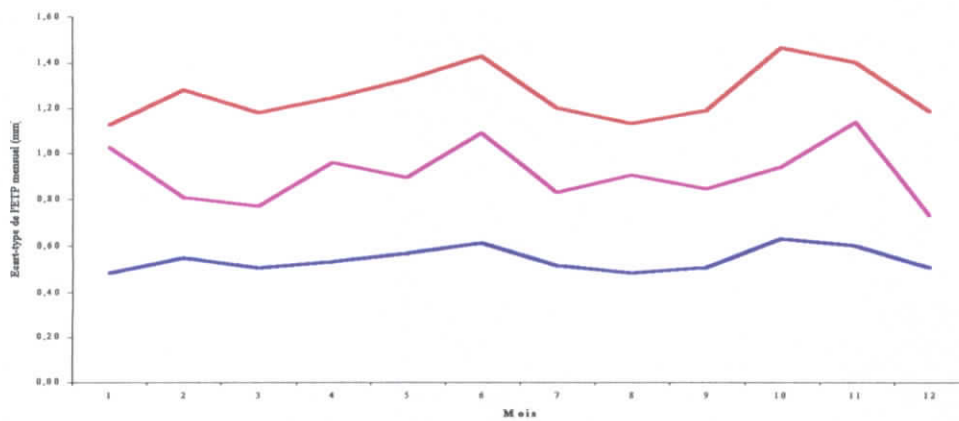


Figure V-35 - Ecart-type de l'ETP mensuelle interannuelle simulée.

V-5 EFFET DU NOMBRE DE COMPOSANTES PRINCIPALES :

La procédure automatique permet d'étudier l'effet du nombre de CP sur les simulations obtenues. Nous remarquons en effet, que l'apport du nombre de CP diffère, et qu'il varie en fonction du phénomène étudié.

Il s'est avéré que lorsque nous prenons une seule CP pour un modèle adéquat, les résultats sont très satisfaisants, cela s'explique par le fait que la variance résiduelle reproduit très bien le reste de l'information.

Pour la variable pluviométrique (aléatoire), au pas de temps journalier, les résultats montrent qu'avec une seule CP, le critère d'acceptation des possibilités de simulation est vérifié plusieurs fois à 100 %.

Pour la variable ETP (cyclique), au pas de temps journalier, nous avons constaté que le critère d'acceptation des possibilités de simulation est vérifié plusieurs fois à plus de 96 %.

V-6 EFFET DE LA TAILLE DE L'ECHANTILLON :

La brièveté des séries de données climatologiques dont dispose les chercheurs constitue un handicap pour toute étude fiable, et sont souvent une source d'erreurs au calcul de dimensionnement d'aménagements hydrauliques, ainsi que dans l'établissement des plans de protection et de gestion des ressources hydriques.

Comme un échantillon contient l'extrême majorité de l'information du phénomène étudié, nous avons voulu savoir à partir du modèle de simulation élaboré quelle est la taille minimale à partir de laquelle les résultats ne deviennent plus fiables.

Pour cela nous avons subdivisé les fichiers de pluviométrie et d'ETP en plusieurs sous échantillons de 10, 20 et 30 ans. Les fichiers simulés ont montré que les résultats du modèle normé sont meilleurs que ceux du modèle brute, et ce pour les deux types de phénomènes aléatoire ou cyclique.

En outre, nous avons remarqué que, même en diminuant la taille des fichiers de l'historique, les résultats restent tout aussi satisfaisants que ceux de l'échantillon initial.

Pour des échantillons de moins de dix (10) ans le modèle donne des résultats probants.

V-7 EFFET DE LA CHRONOLOGIE :

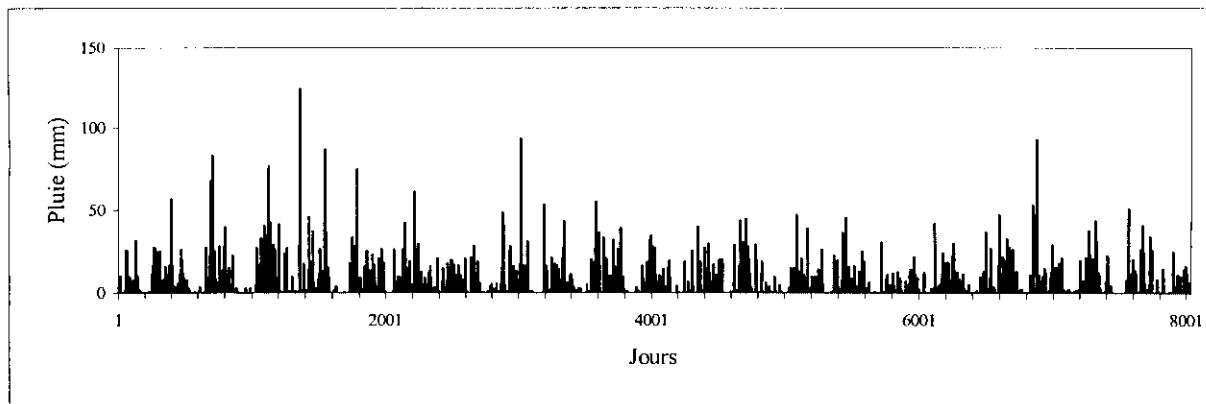


Figure V-36 Evolution de la pluviométrie journalière (fichier tronqué).

Le principal inconvénient des modèles utilisés en simulation des variables climatologiques est l'exigence de la chronologie de l'échantillon.

Comme la majorité des échantillons observés sont tronqués (contiennent des valeurs manquantes), leur utilisation reste limité.

Pour tester la fiabilité du modèle élaboré à différents pas de temps, le fichier historique de la pluviométrie a été tronqué à différents niveaux (figure V-36) et les deux types de modèles élaborés ont été appliqués avec les options minimales (nombre de CP égale à 1). Les résultats obtenus montrent que l'effet de la chronologie ne se pose même pas, d'où la robustesse et l'adaptabilité du modèle à ce type de problème.

V-8 EFFET DE LA STATIONNARITE :

Cette partie découle des développements réalisés précédemment, en effet lorsque le pas de temps diminue, le phénomène devient non stationnaire (passage du mensuel au décadaire et du décadaire au journalier), les résultats obtenus montrent que le modèle n'exige pas la stationnarité des variables étudiées, et que la série simulée reste conforme à celle de l'historique. Ce qui change dans ce cas c'est le type de simulation des paramètres du modèles.

Chapitre VI : Elaboration du Modèle de Prévision

VI-1 INTRODUCTION :

L'élaboration du modèle, les techniques de prévision utilisées pour prédire la pluviométrie et l'évapotranspiration (ETP) au pas de temps mensuel, et les résultats obtenus sont les objectifs de ce chapitre.

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. Les organismes qui ont en charge la sécurité civile, l'assainissement pluvial, l'annonce des crues, la protection des végétaux, la sécurité routière, attendent des services de la météorologie des informations qui leur permettent de mettre en œuvre, en temps opportun, les moyens de prévention et d'intervention dont ils ont la maîtrise.

Pour faire face à toutes ces situations et avoir une bonne gestion des ressources en eau, la prévision des précipitations et des ETP 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), les caractéristiques physiques du bassin versant, l'état du couvert végétal et le degré de saturation des sols.

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.
- A la longueur des séries étudiées.

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 (Kadi, 1995).

VI-2 LE MODELE DE PREVISION :

Pour faire une bonne prévision sur l'évolution future d'une variable :

- il faut un bon modèle ;
- il faut « nourrir » ce modèle avec des informations de qualité.

A supposer que le modèle soit de la forme :

$$Y_t = \alpha + \beta X_t + \varepsilon_t \quad (\text{VI-1})$$

Et que les coefficients α et β soient déjà connus, la qualité de la prévision sur Y est tributaire de la qualité des prévisions sur X. L'idéal, serait de pouvoir construire à la date t, une prévision sur Y_{t+k} qui serait fondée sur l'histoire de cette seule variable.

Nous pouvons parfois lire l'état de la variable Y à la date t comme étant le reflet :

- De ses valeurs antérieures.
- Des perturbations passées et présentes.

Ce qui nous intéresse, c'est de savoir s'il est possible de tirer dans l'histoire des valeurs passées de Y, et dans celle des perturbations qui ont affectées cette variable des informations utiles pour prévoir les valeurs ultérieures de cette dernière.

A supposer qu'on se soit fixé pour objectif la prévision des valeurs futures de Y, l'idée principale pour la construction d'un modèle est la suivante : la valeur de Y à la date t est la résultante des valeurs prises antérieurement par les variables : X_1, X_2, \dots, X_k , mais :

- L'identification exhaustive de ces variables est souvent difficile.
- En admettant qu'on puisse définir ces variables, l'appréciation de l'influence spécifique de chacune d'elles sur les Y_t est très délicate.
- Enfin, nous pouvons admettre que la chronique des valeurs passées de Y contient elle même toute l'information nécessaire pour une prédiction efficace de Y, en d'autres termes, les influences des valeurs passées de X_1, \dots, X_k sur Y_t se trouvent en quelque sorte résumées dans la chronique des valeurs passées de Y.

L'approche préconisée dans cette partie, pour la prévisibilité de la variable Y, ne tiendra donc compte que des variables étudiées à savoir la pluviométrie et l'ETP. Le modèle de prévision élaboré est basé sur l'analyse en composantes principales appliquée dans un contexte prévisionnel sur des variables brutes et normées, aléatoires et cycliques, et sur les chaînes de Markov.

La procédure d'élaboration du modèle de prévision est la même que celle adoptée pour le modèle de simulation (voir § IV-2). L'expression finale du modèle de prévision brute s'écrit :

$$\hat{X} = \sum_{l=1}^q \beta_{jl} C_l + \varepsilon_j \quad (\text{VI-2})$$

Celle du modèle normé est donnée par :

$$\hat{X}_j = \beta_{j0} + \sum_{l=1}^q \beta_{jl} C_l + E_j \quad (\text{VI-3})$$

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

$E_j = \sigma_{xj} \varepsilon_j$

En se basant sur la structure de ces deux modèles de prévision, nous déterminons :

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

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

VI-2-1 Calage du modèle :

Cette étape dénommée « calage » ou « calibration » est une phase purement quantitative qui se résume en terme mathématique à « minimiser » l'écart entre les valeurs calculées du modèle et celles observées du phénomène étudié.

VI-2-2 Test du modèle :

Cette étape consiste à tester la capacité du modèle, c'est à dire, à vérifier l'aptitude du modèle calé à prévoir la réalité, et à 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.

VI-2-3 Exploitation du modèle :

Cette étape ne pose théoriquement aucun problème dès lors que le calage et le test ont été bien conduits.

VI-3 CRITERES DE VALIDATION :

Pour juger la qualité d'une prévision, nous faisons appel à des fonctions critères qui permettent d'apprécier quantitativement et qualitativement les résultats mensuels, estimant globalement l'écart entre les sorties calculées (pluies ou ETP mensuelles prévues) et les variables historiques (Beldring, 2002), il existe deux types de critères :

- Les critères graphiques : l'analyse graphique est indispensable et primordiale, cela est obtenu en portant sur un graphique les valeurs observées et celles calculées sur une période test.
- Les critères numériques (Rognon, 1996, Motovilov et al., 1999) :

1- Le critère de Nash (NS), est défini par :

$$NS = \left[1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \right] * 100 \quad (VI-4)$$

Q : Valeurs des données observées ;

\hat{Q} : Valeurs prévues par le modèle ;

\bar{Q} : Moyenne des données observées.

Ce critère a beaucoup séduit les hydrologues, il a été utilisé au début pour juger un modèle fonctionnant par épisodes (Nash et al., 1970 ; O'Connell et al., 1970). Il exprime le pourcentage de la variance de la série mesurée expliqué par le modèle, ou en d'autres termes, il explique ce que l'on gagne par rapport à un modèle médiocre qui donnerait comme valeurs calculées la moyenne des valeurs observées (Nascimento, 1995).

Le critère de Nash peut prendre des valeurs de $-\infty$ à 100 %. Dans le cas général une « bonne » ou une « mauvaise » valeur du critère de Nash sont définies comme suit (Rakem, 1999) :

Tableau VI-1 Qualité du critère de Nash.

Valeurs de Nash (%)	Qualité
Nash < 70	très insuffisant
$70 \leq \text{Nash} < 80$	insuffisant
$80 \leq \text{Nash} < 85$	médiocre
$85 \leq \text{Nash} < 90$	correct
$90 \leq \text{Nash} < 95$	Bon
Nash ≥ 95	très bon

2- Moyenne biaisée : c'est la différence entre la moyenne des valeurs observées et prévues. Celle-ci est donnée par la relation suivante (Legates et al., 1999) :

$$\{B\} = \bar{x}_{\text{prév}} - \bar{x}_{\text{obs}} \tag{VI-5}$$

$\bar{x}_{\text{prév}}$: Moyenne des valeurs prévues durant la période test ;

\bar{x}_{obs} : Moyenne des données observées durant cette période.

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

3- Indice de coefficient de variation :

La validation du modèle peut être complétée en comparant les coefficients de variation des valeurs observées et celles calculées.

$$I = C_{v_c} / C_{v_o} \tag{VI-6}$$

C_{v_c} : Coefficient de variation des valeurs calculées pendant la période test ;

C_{v_o} : Coefficient de variation des données observées pendant la période test.

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

VI-4 PREVISION DES DIFFERENTES VARIABLES CLIMATOLOGIQUES :

Afin de mieux apprécier la fiabilité du modèle de prévision élaboré, nous avons comme pour la simulation considéré deux variables climatologiques, l'une aléatoire (la pluie) et l'autre cyclique (l'ETP).

VI-4-1 Pr evision de la pluviom etrie :

Le fichier de la pluviom etrie au pas de temps mensuel s' tend de 1936 – 2001, soit 66 ans. Le calage du mod le se fera sur 56 ans et la p riode de validation s' tendra sur 10 ans.

a) Mod le brute :

Plusieurs m ethodes de pr evision des param tres du mod le peuvent  tre envisag es :

- 1 re approche :

- Etude des coefficients β_{ij} :

Pour la d termination des β_{ij} , nous avons  tudi  leur  volution pour une seule CP sur diff rentes p riodes (40, 46, 48, 50, 52, 54, 56ans).

La figure VI-1 montre que les β_{ij} ont tendance   suivre la m me  volution quant la taille de l' chantillon augmente, et les courbes se stabilisent au fur et   mesure que le nombre d'ann es s'accro t.

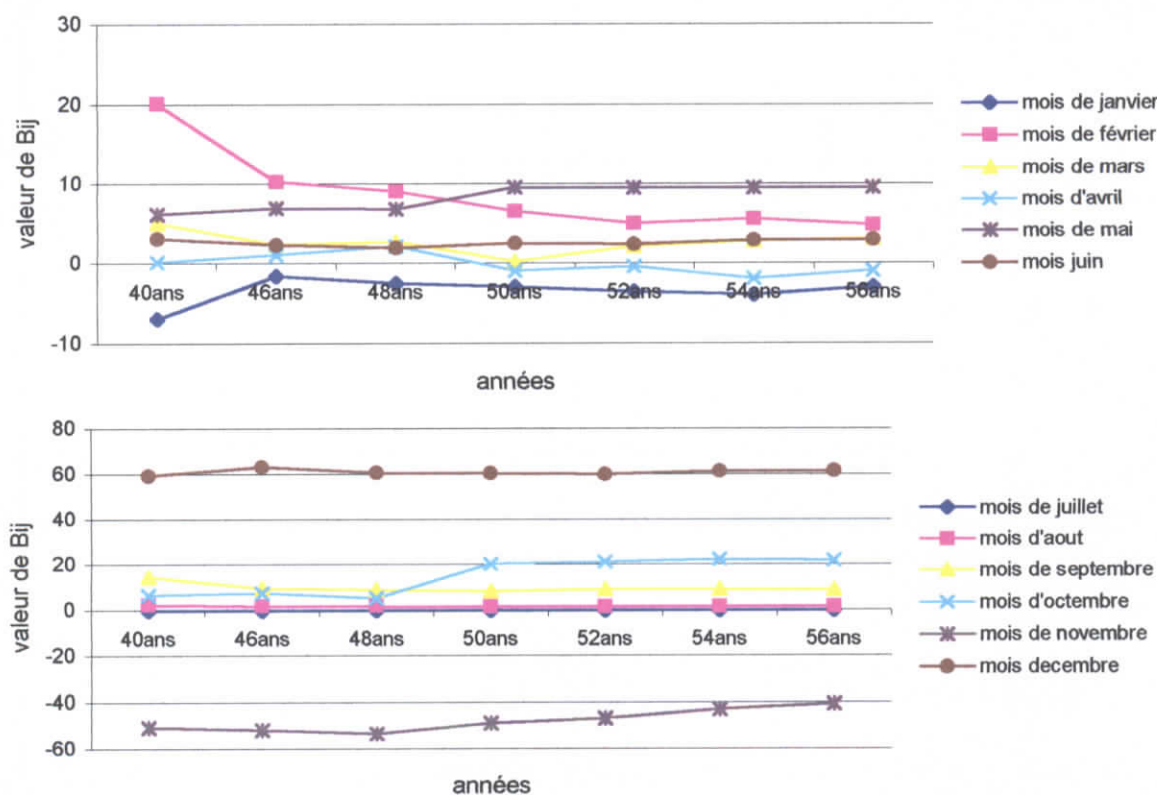


Figure VI-1 Evolution des coefficients de r gression en fonction du temps

Le tableau VI-2 donne le nombre d'ann es pour lequel β_{ij} est stable et la valeur correspondante   chaque mois.

Tableau VI-2 - Valeurs des β_{ij} pour chaque mois

Mois	Valeur de β_{ij}
	6ans 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 VI-2 représente l'évolution de la 1^{ère} composante principale en fonction du temps, et montre une variation aléatoire des CP.

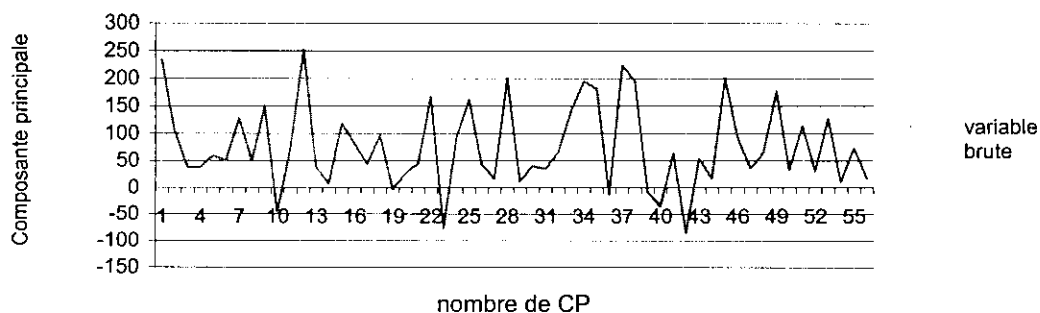


Figure VI-2 - Evolution de la 1ère composante principale.

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

- Estimation des résidus :

Le résidu ε_j est défini comme étant la perte d'informations dues aux CP négligées, il est donné par la différence entre la variable observée et celle prévue.

$$\varepsilon_j = \hat{X}_j - \left(\sum_{i=1}^q \beta_{ji} * C_i \right) \quad (\text{VI-7})$$

La figure VI-3 qui illustre l'évolution résiduelle en fonction du temps, montre le caractère aléatoire des ε_j .

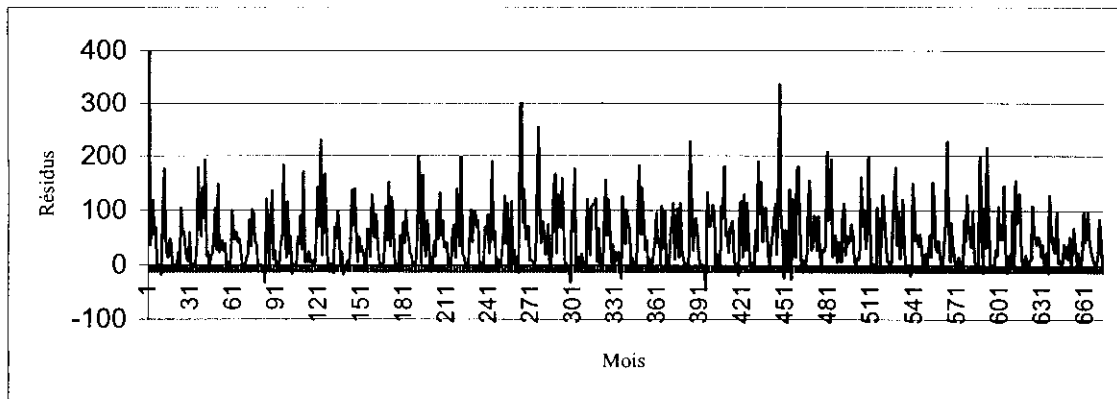


Figure VI-3 - Evolution des résidus des pluies mensuelles brutes

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, nous proposons :

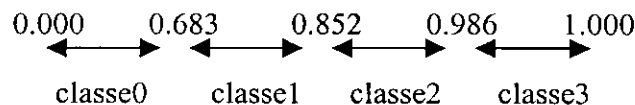
- Un nombre de classes égale à 4
- Les bornes de classes suivantes :

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

Pour prévoir l'année 1992, par exemple, le modèle Markovien multiclasse appliqué sur les résidus de 56 ans 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, nous lui attribuerons ensuite la valeur correspondante de la série historique des résidus.

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

$$\hat{X} = \sum_{l=1}^Q \beta_{jl} C_l + \varepsilon_j \quad (\text{VI-8})$$

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).

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

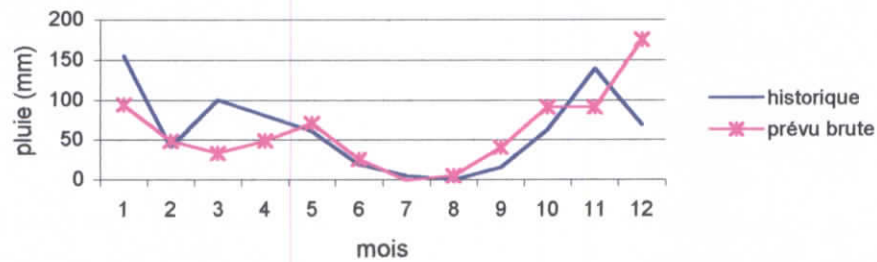


Figure VI-4 Prévision de l'année 1992.

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

Les résultats des critères numériques obtenus par application des tests de validation sur chaque année sont résumés dans le tableau VI-3.

Tableau VI-3 - Résultats des tests de validation sur les variables brutes.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	C.Nash (%)	{B}	I
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 valeurs du critère de Nash obtenues sont très insuffisants, et les moyennes des mois prévus ne coïncident globalement pas avec les moyennes de la série observée, ceci s'explique par l'écart parfois trop important entre les valeurs observées et celles calculées.

La figure VI-5 qui représente l'évolution des valeurs prévues et observées pour toute la période test (1992-2001) montre un décalage non significatif en général entre les deux courbes. En outre, nous remarquons que l'écart entre les valeurs observées, et celles calculées sur 69 mois ne dépasse pas 30mm, soit en moyenne 57% des mois prévus.

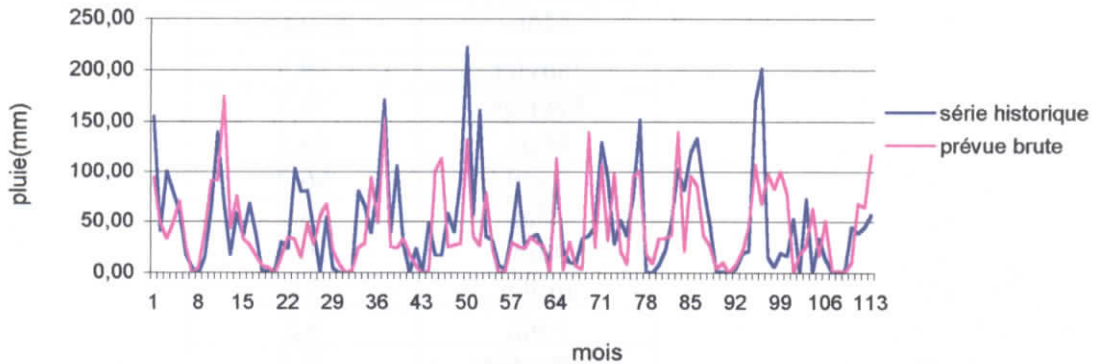


Figure VI- 5 - Prédiction des pluies mensuelles (variables brutes).

b) Modèle normé :

- Détermination des β_0 :

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

La figure VI-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.

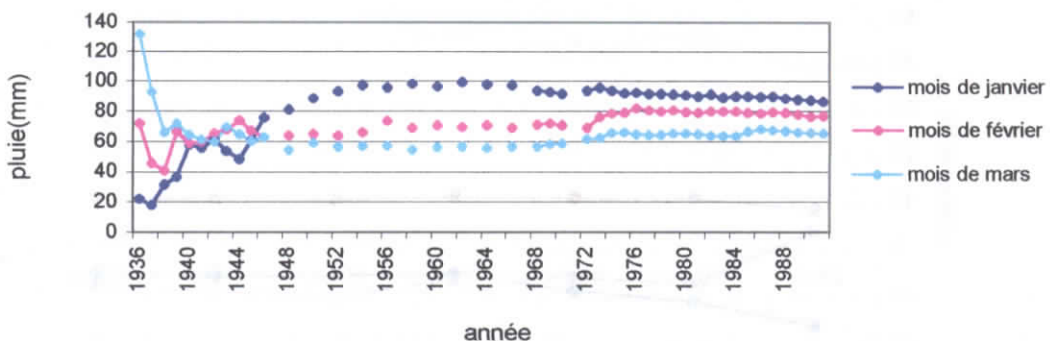


Figure VI- 6 - Evolution des moyennes mensuelles interannuelles.

L'évolution des moyennes mensuelles inter annuelles en fonction du temps des autres mois est représentée dans l'annexe VI-2.

Nous remarquons 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 VI-4).

Tableau VI - 4 Valeurs de β_0 pour chaque mois.

Mois	β_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 des β_{ij} est étudiée de la même manière que pour les β_{ij} du modèle brute.

Les mêmes constatations sont faites pour le modèle normé, en effet, la figure VI-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.

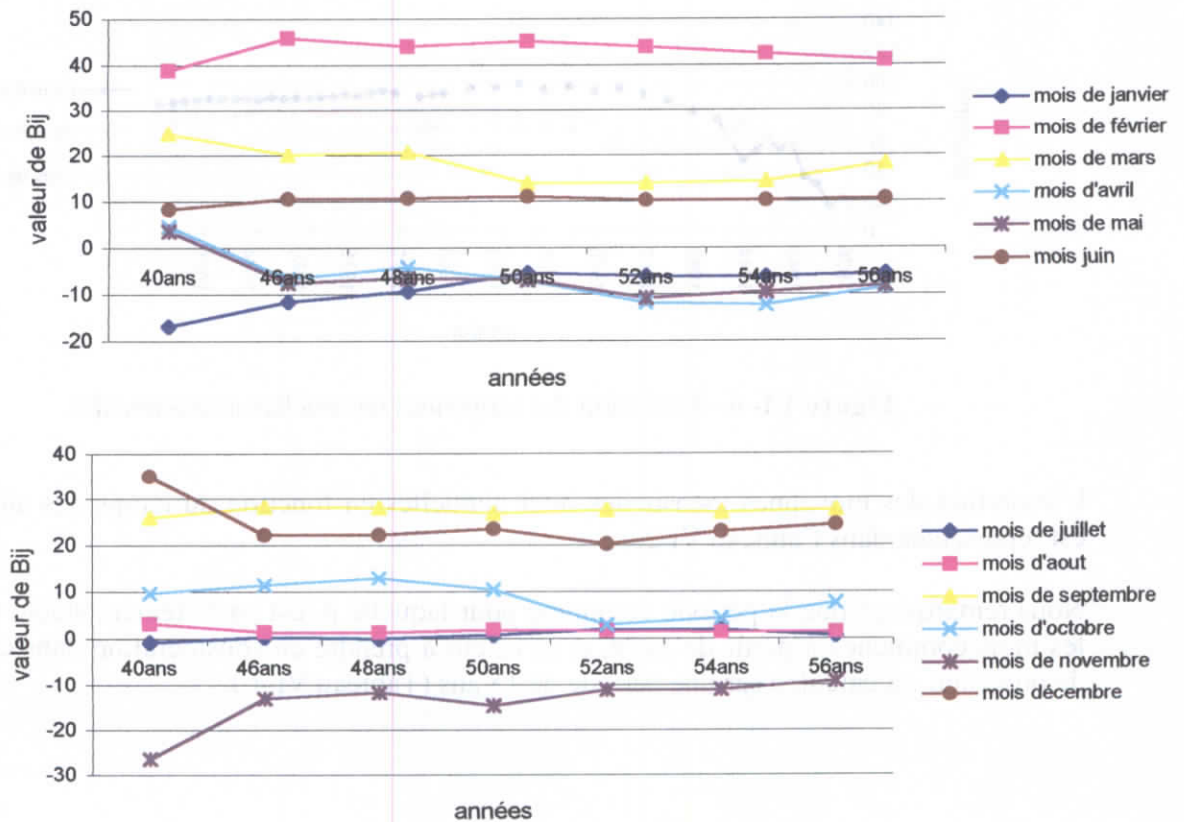


Figure VI-7 Evolution des β_{ij} pour chaque mois.

Le tableau VI- 5 donne les valeurs des β_{ij} pour chaque mois.

Tableau VI- 5 Valeurs des β_{ij} correspondants à chaque mois.

Mois	Valeurs des 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

- Détermination des CP :

Vu le caractère aléatoire des CP en fonction du temps (figure VI-8), nous avons appliqué la même démarche pour la détermination des CP que pour le modèle brute.

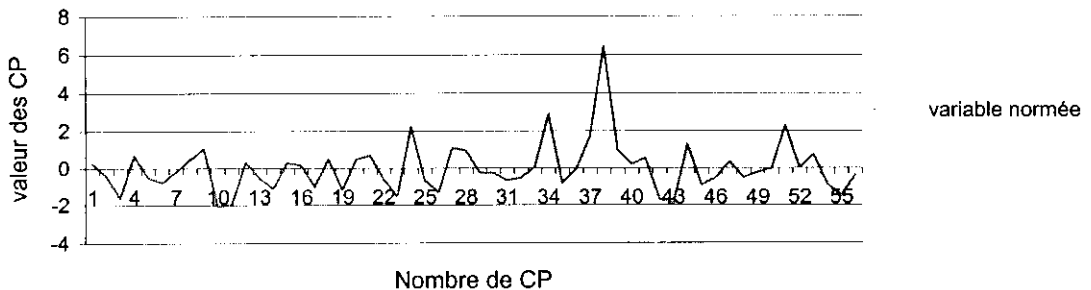


Figure VI- 8 - Evolution de la 1ère composante principale

- Estimation des résidus :

L'évolution aléatoire des ε_j est illustrée par la figure VI-9, la même procédure est utilisée pour la détermination des ε_j que pour le modèle brute.

$$\varepsilon_j = \hat{X}_j - \left(\sum_{i=1}^p \beta_{ji} C_i - \beta_{j0} \right) \tag{VI-9}$$

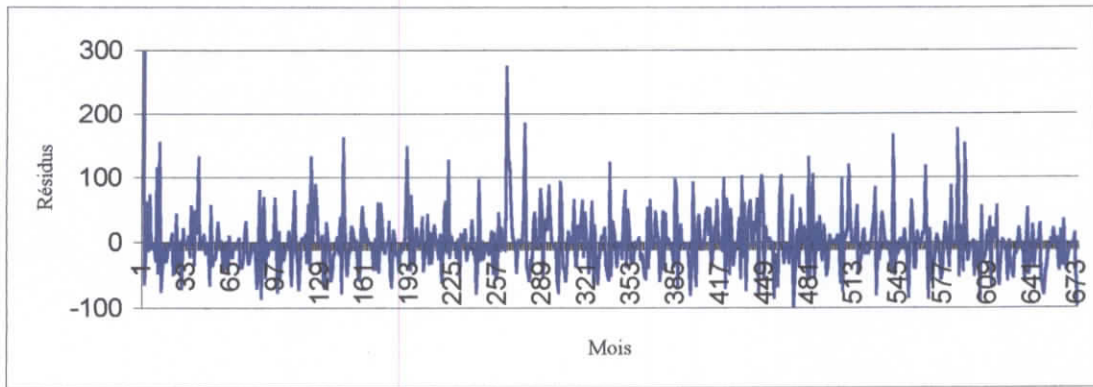


Figure VI-9 - Evolution des résidus des pluies mensuelles normées

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

classe 0 classe 1 classe 2 classe 3
 $]-\infty \ 0[$, $[0 \ 50[$, $[50 \ 100[$, $[100 \ \infty[$

Les séries mensuelles des pluies observées et prévues pour l'année 1992 sont représentées dans la figure VI-10, les autres années prévues sont données dans l'annexe VI-1.

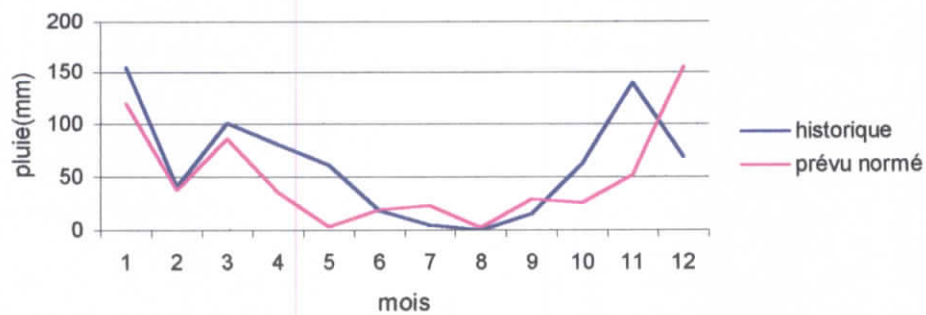


Figure VI-10 Prévission de l'année 1992

Les résultats obtenus, en appliquant les tests de validation sont résumés dans le tableau VI-6.

Tableau VI – 6 Résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	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 très insuffisantes. En outre, le modèle normé permet de prévoir 53% des mois avec une différence qui ne dépasse pas 30mm par rapport aux valeurs observées.

La figure VI-11 montre toujours un léger décalage des valeurs prévues par rapport à celles observées.

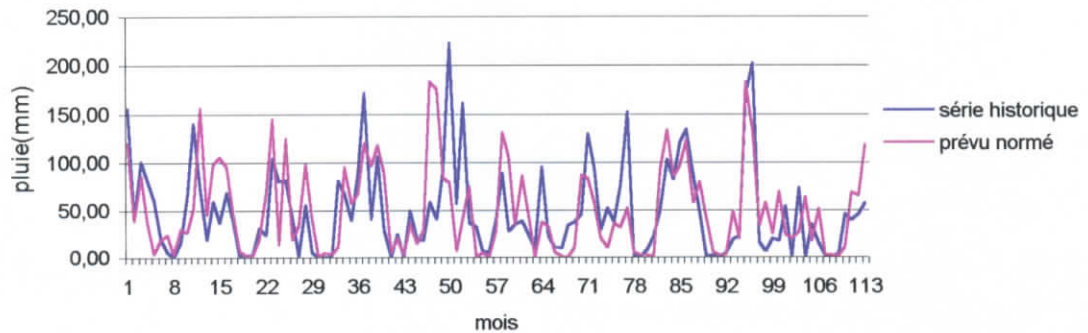


Figure VI-11- Prédiction des pluies mensuelles

L'inconvénient de cette première approche est que parfois le tirage d'un nombre aléatoire donne lieu à des valeurs de précipitation prévues négatives ce qui est physiquement impossible, celles-ci seront remplacées par des valeurs nulles (Figure VI-12).

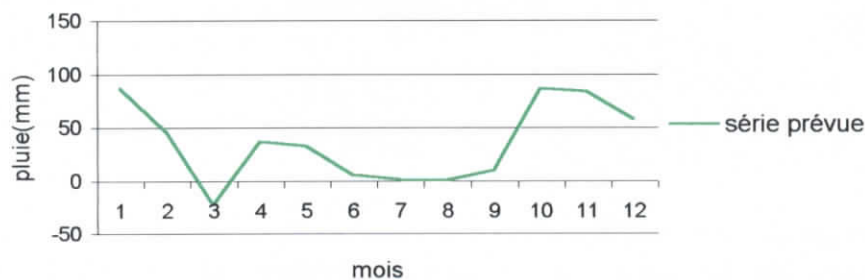


Figure VI-12 - Prédiction de l'année 1997 (variable normée).

- 2^{ème} approche :

Pour éviter ce genre de situation, à savoir la rectification des valeurs prévues des précipitations, nous avons transformés les données brutes observées en racine carrée ($y=\sqrt{x}$).

- Modèle brute :

L'évolution des coefficients de régression en fonction du temps est donnée en annexe VI-3, alors que les tracés de la première composante principale et des résidus sont représentés dans l'annexe VI-4.

La représentation graphique des séries observées et prévue pour l'année 1992 est donnée par la figure VI-13 et les graphiques des autres années sont présentés dans l'annexe VI-5.

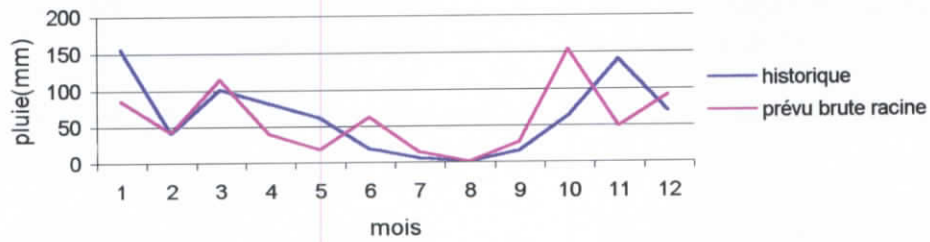


Figure VI-13 - Prédiction de l'année 1992.

Le tableau VI-7 résume les paramètres statistiques des séries historiques et prévues, ainsi que les résultats obtenus par application des tests de validation.

Tableau VI-7 - Résultats de la modélisation pour les variables racines brutes.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	C.Nash (%)	{B}	I
92	62.41	50.59	0.81	57.91	45.34	0.78	1.29	-4.5	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	4.1	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}, alors que pour Nash les valeurs restent toujours très insuffisantes.

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

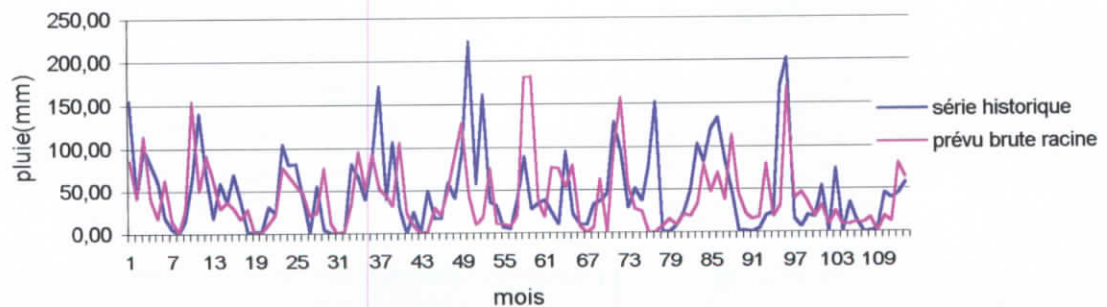


Figure VI-14 - Prédiction des pluies mensuelles.

b) Modèle normé:

L'évolution des coefficients de régression en fonction du temps est donnée en annexe VI-6, alors que les tracés de la première composante principale et des résidus sont représentés dans l'annexe VI-7.

La prévision de l'année 1992 est représentée par la figure VI-15, les autres années prévues sont données dans l'annexe VI-5.

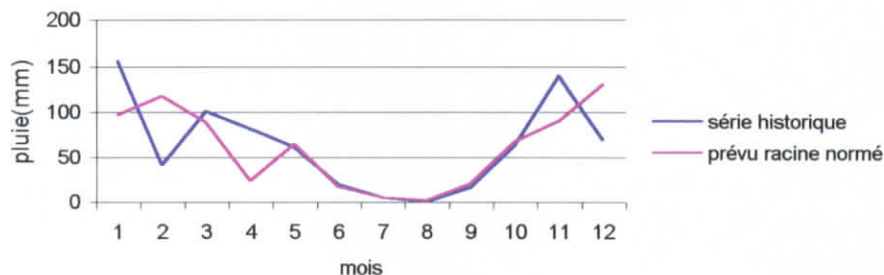


Figure VI-15 - Prévision de l'année 1992.

Les résultats obtenus, en appliquant les tests de validations sont résumés dans le tableau VI-8.

Tableau VI-8 - Résultats de la modélisation pour les variables racines normées.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	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 bien reproduite (Figure VI-16).

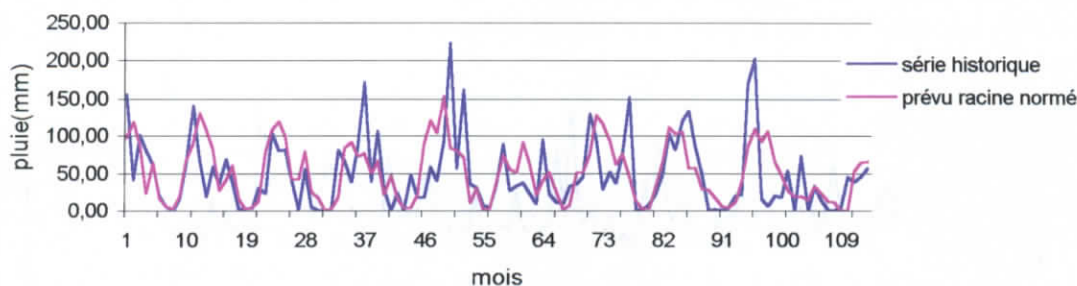


Figure VI-16 - Prévision des pluies mensuelles.

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

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

- 3^{ème} approche :

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

La classe du mois prévue sera celle correspondante à la plus forte probabilité, et ce en tenant compte de la classe précédente, ensuite nous tirons 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 classes suffisant, de manière à garder les caractéristiques de chaque saison.

a) Modèle brute :

Nous avons choisi la répartition suivante :

Saison 1 : décembre- février	$]-\infty$	25[[25	70[[70	100[[100	∞ [
Saison 2 : mars- mai	$]-\infty$	25[[25	70[[70	100[[100	∞ [
Saison 3 : juin- août	$]-\infty$	0[[0	25[[200	∞ [
Saison 4 : septembre- novembre	$]-\infty$	50[[50	100[[100	∞ [

Pour la détermination des β_{ij} et des CP elle se fait de la même manière que précédemment.

Le résidu est obtenu de la même manière que dans l'exemple cité ci-dessous pour le modèle normé. L'annexe VI-8 représente les séries prévues et celles observées pour chaque année.

Nous remarquons qu'en utilisant les variables brutes, les quantités de pluies prévues sont toujours sous-estimées par rapport à celles de l'historique (Figure VI-17).

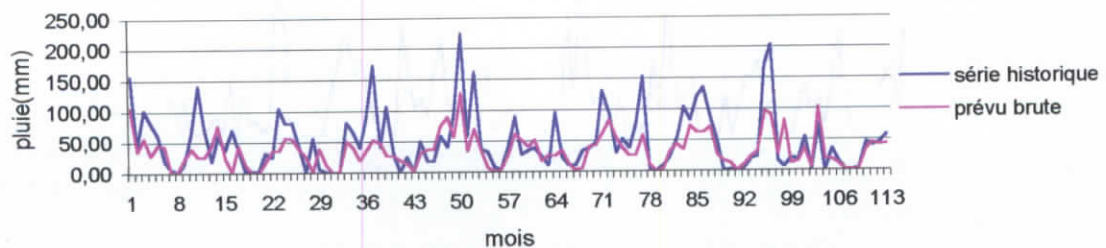


Figure VI-17- Prévission des pluies mensuelles.

Les résultats de la prévision mensuelle par le principe de la plus forte probabilité d'apparition par le modèle Markovien multiclasse sont résumés dans le tableau VI-9 qui montre une amélioration du modèle de prévision développé.

En effet, ce modèle permet de prévoir 71% des mois avec un écart de moins de 30 mm par rapport aux valeurs mensuelles observées.

Tableau VI- 9 - Résultats de la modélisation pour la variables brute.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	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

b) Modèle normé :

L'année est scindée en deux saisons :

Saison 1 : septembre- mai

]- ∞ -20[[-20 25[[25 50[[50 ∞ [

Saison 2 : juin- août

]- ∞ -5[[-5 0[[0 10[[10 ∞ [

La détermination des β_0 , β_{ij} et des CP se fait de la même manière que précédemment

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, nous lui attribuons la valeur correspondante.

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

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 VI-18).

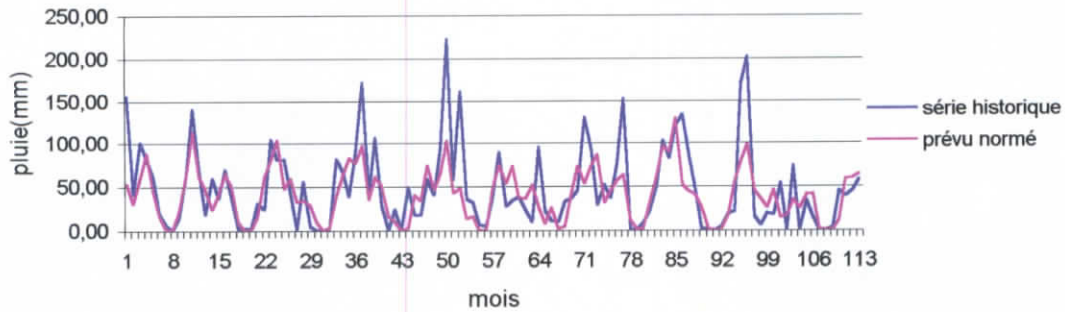


Figure VI-18 - Prévision des pluies mensuelles.

Les résultats obtenus, en appliquant les tests de validations sont résumés dans le tableau VI-10.

Tableau VI-10 - Résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	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	49.80	-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

Ces résultats montrent une amélioration significative la aussi du modèle de prévision développé.

En outre, les mois prévus avec un écart de moins de 30 mm par rapport aux valeurs observées sont de 71% dans ce cas.

VI-4-2 Prévision de l'ETP :

Pour la variable ETP, nous disposons d'une série d'observations s'étalant de 1958 – 1991, soit 34 ans. La période de calage sera de 27 ans, et la période test de 7 ans.

- 1^{ère} approche :

a) Modèle brute :

Pour la prévision des paramètres de ce modèle (β_{ij} , CP, ϵ_j) la même démarche a été appliquée que pour une variable aléatoire.

Les séries mensuelles observées et prévues des ETP pour la période test sont représentées par la figure VI-19.

Celle-ci montre qu'il n'y a pratiquement pas de décalage entre les valeurs prévues et celles observées.

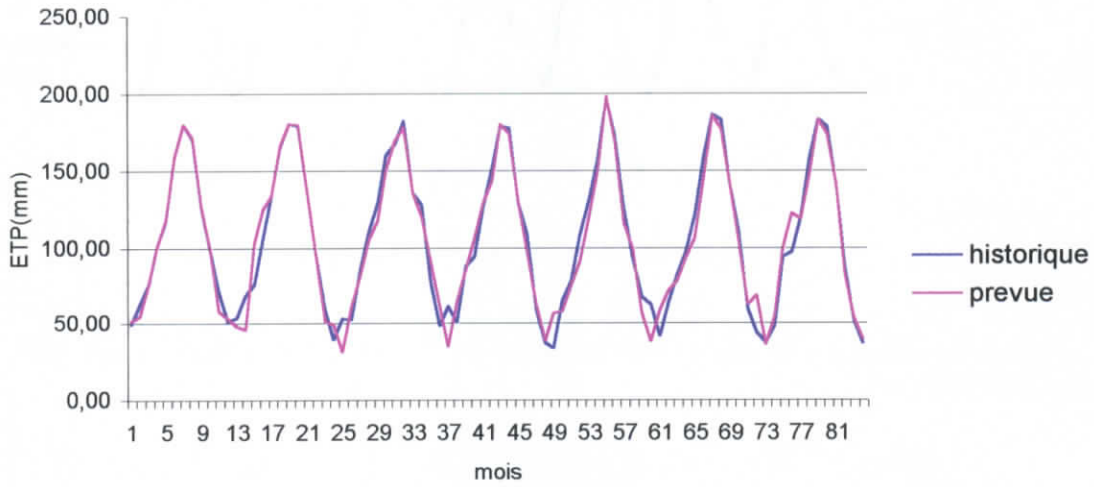


Figure VI-19 - Prévision des ETP mensuelles.

Les résultats obtenus par application des tests de validation sur chaque année sont résumés dans le tableau VI-11.

Tableau VI-11 - Résultats de la modélisation pour les variables brutes.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	C.Nash (%)	{B}	I
85	104.98	46.08	0.43	103.66	47.35	0.46	98.98	-1.32	1.06
86	107.51	50.53	0.47	109.29	52.1	0.48	93.94	1.78	1.01
87	110.53	47.13	0.42	108.40	45.61	0.42	94.44	-2.13	1.00
88	105.08	48.55	0.46	103.69	49.02	0.47	95.31	-1.39	1.03
89	107.28	49.91	0.46	102.08	50.06	0.49	93.12	-5.20	1.07
90	106.96	51.59	0.48	107.46	44.61	0.42	93.16	0.50	0.86
91	102.43	53.37	0.52	104.24	51.54	0.49	97.08	1.81	0.95

Les résultats de Nash obtenus sont très satisfaisants et les moyennes des années prévues coïncident avec celles des séries observées.

b) Modèle normée :

Nous adoptons toujours la même démarche pour la détermination des paramètres de ce modèle (β_0 , β_{ij} , CP, et ϵ_j) que pour une variable aléatoire.

La figure VI-20 montre l'évolution des mois prévus par rapport à ceux de l'historique.

Nous remarquons là aussi le décalage négligeable qui existe entre la série prévue et celle observée dans le cas d'une variable cyclique normée.

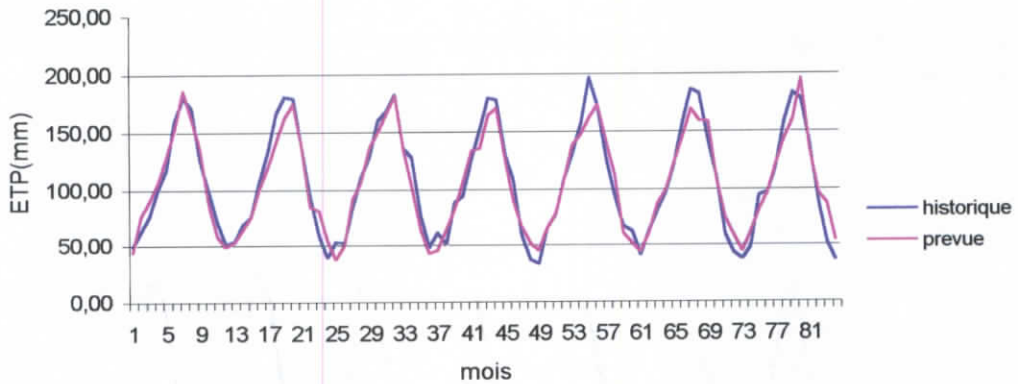


Figure VI-20 - Prévion des ETP mensuelles.

Les résultats obtenus par application des tests de validation sur chaque année sont résumés dans le tableau VI-12.

Tableau VI-12 - Résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	C.Nash (%)	{B}	I
85	104.98	46.08	0.43	105.73	46.60	0.44	95.42	0.75	1.02
86	107.51	50.53	0.47	104.10	42.07	0.40	89.76	-3.41	0.86
87	110.53	47.13	0.42	105.28	49.33	0.47	95.21	-5.25	1.12
88	105.08	48.55	0.46	102.69	42.78	0.42	91.29	-2.39	0.91
89	107.28	49.91	0.46	106.58	45.54	0.43	90.52	-0.70	0.93
90	106.96	51.59	0.48	107.17	43.59	0.41	91.61	0.21	0.85
91	102.43	53.37	0.52	106.30	45.73	0.43	86.94	3.87	0.83

Vu les résultats presque parfaits que nous avons obtenus pour les modèles brute et normé, (que ce soit graphiquement ou numériquement), nous nous sommes limités à l'étude du modèle normé pour une variable transformée en racine.

- 2^{ème} approche :

La figure VI-21 montre une parfaite similitude entre les séries prévue et observée.

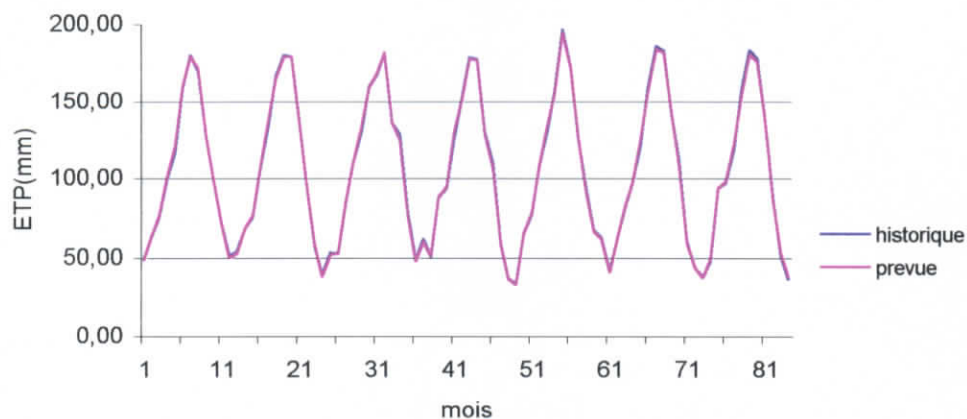


Figure V- 21 - Prévision des ETP mensuelles.

Les résultats obtenus pour les tests de validation sont résumés dans le tableau VI-13.

Tableau VI-13 - Résultats de la modélisation pour les variables normées.

Année	Variable observée			Variable prévue					
	\bar{x}	s_x	C_v	\bar{x}	s_x	C_v	C.Nash (%)	$\{B\}$	I
85	104.98	46.08	0.43	105.15	46.31	0.44	99.95	0.17	1.02
86	107.51	50.53	0.47	107.57	50.61	0.47	99.97	0.06	1.00
87	110.53	47.13	0.42	110.29	47.32	0.43	99.93	-0.24	1.02
88	105.08	48.55	0.46	104.99	48.50	0.46	99.95	-0.09	1.00
89	107.28	49.91	0.46	107.02	49.86	0.47	99.95	-0.26	1.01
90	106.96	51.59	0.48	106.60	51.21	0.48	99.93	-0.36	1.00
91	102.43	53.37	0.52	102.24	52.41	0.51	99.99	-0.19	0.99

Nous remarquons dans ce cas que tous les critères obtenus donnent des valeurs identiques (ou presque) et que les paramètres se trouvent inchangés (moyenne, écart-type, etc.).

En conclusion, nous pouvons dire que les résultats obtenus sont très satisfaisants du fait que la tendance des phénomènes étudiés est bien reproduite et que les tests de validation sont globalement vérifiés.

Conclusion Générale

L'eau devient de plus en plus rare, particulièrement dans les zones à forte densité de population des régions arides et semi-arides, ce phénomène menace non seulement l'économie des pays, mais également le bien-être des populations.

La gestion de cette denrée rare dépend, parmi d'autres variables hydroclimatologiques de la pluie, et de l'ETP. Dans l'impossibilité de prévoir le changement de ces variables dans le futur, nous adoptons fréquemment l'hypothèse de répétition du registre des variables observées dans le passé, appelée série historique, mais celle-ci n'est qu'une des possibilités de réalisations d'un processus stochastique, ou autrement dit, on peut imaginer que la nature « a choisi » la série historique selon un ensemble quelconque de lois probabilistiques. Un nouveau « tirage au sort » engendrerait une autre série, différente de l'historique mais également probable.

Nous nous sommes attelés pour notre part à rechercher une méthodologie susceptible d'apporter une contribution au problème de la gestion de la ressource en eau en Algérie qui accuse un déficit relatif dans les apports en eau, principalement dû aux fluctuations des conditions climatiques, et à l'irrégularité de la répartition spatiale de cette ressource, aggravé par les mauvaises options dans le domaine hydraulique prises par les décideurs (mauvais choix des sites d'implantation de barrages et gestion anarchique des réserves d'eau disponibles).

En effet, la simulation des variables climatiques (pluie, ETP, etc.) permet de mieux appréhender ces problèmes et fournit un outil précieux pour l'étude des comportements des bassins versants ainsi que l'établissement des schémas directeurs de gestion des ressources en eau.

Nous avons essayé, tout au long de ce travail d'apporter une contribution nouvelle parmi les différentes recherches réalisées sur les techniques de simulation et de prévision des variables climatiques, car les modèles élaborés à cette fin (Thomas et Fiering, ARMA, Markov, décomposition de la série historique, etc.), et qui sont proposés dans la littérature, sont difficilement exploitables, d'une part en raison :

- Du nombre trop élevé de paramètres à prendre en considération ;
- De la difficulté d'estimation de ces paramètres ;
- Du manque de base physique ;
- De la non reproduction de la dépendance à court et long terme ;
- Du manque de données relatives aux paramètres pris en compte.

Et d'autre part, de leur exigence :

- La chronologie des données ;
- La stationnarité du phénomène étudié ;
- La normalité de l'échantillon ;
- La dépendance entre les valeurs observées ;

- La taille de l'échantillon ;
- Type de variable (aléatoire ou cyclique) ;
- Pas de temps spécifique.

C'est pour cela que nous avons veillé à élaborer un modèle de simulation qui permet d'obtenir des séries générées pour tous les types de variables (aléatoire et cyclique), à différents pas de temps, et surtout en éliminant toutes les contraintes précitées. Il constitue ainsi un outil d'aide à la décision pour les gestionnaires des barrages réservoirs qui ne seront plus obligés de faire un choix du modèle à utiliser.

Le modèle élaboré est simple, avec peu de paramètres, et permet d'apporter des réponses satisfaisantes à certains problèmes de l'hydrologie opérationnelle, et peut se mesurer aux modèles plus complexes existants et utilisés par les ingénieurs hydrologues.

La démarche suivie dans le cadre de ce travail de recherche, est basée essentiellement sur la technique de l'analyse en composantes principales utilisée généralement en hydrologie dans le domaine spatial, et qui a permis dans le domaine temporel d'apporter des éléments de réponses à des problèmes qui n'avaient à ce jour aucune solution.

L'ajustement des paramètres des deux modèles normé et brute colonne par colonne, pour chaque transformation de la variable étudiée et pour chaque composante principale, augmente le nombre de possibilités de simulation surtout lorsque le pas de temps est réduit. L'automatisation de cette façon de procéder, à partir du logiciel élaboré à cette fin, simplifie énormément la tâche de l'utilisateur du modèle et ce grâce à l'introduction d'un fichier configurateur de tout le programme, et qui contient toutes les informations nécessaires à la réalisation de toutes les possibilités de simulation.

Cette procédure automatique permet de tester la fiabilité et la robustesse du modèle élaboré et ce en étudiant la variation de plusieurs paramètres sur la qualité des simulations obtenues.

Nous remarquons en effet, que l'apport du nombre de composantes principales diffère et qu'il varie en fonction du phénomène étudié.

Le problème de la brièveté des séries de données météorologiques a été traité et les fichiers simulés montrent que les résultats du modèle normé sont meilleurs que ceux du modèle brute et ce pour les deux types de phénomènes étudiés (aléatoire ou cyclique), cela s'explique par le fait que la variance résiduelle reproduit très bien le reste de l'information.

Pour des échantillons de moins de dix (10) ans le modèle donne des résultats probants.

Le principal inconvénient des modèles utilisés en simulation reste toujours l'exigence de la chronologie de l'échantillon ce qui limite leur domaine d'utilisation car la majorité des données météorologiques sont tronquées.

Les deux types de modèles élaborés ont été appliqués avec les options minimales (nombre de composantes principales égale à 1). Les résultats obtenus montrent que l'effet de la chronologie ne se pose même pas d'ou l'adaptabilité du modèle à ce type de problème.

Les résultats obtenus montrent que la stationnarité des variables étudiées n'est plus exigée par le modèle élaboré.

Pour le modèle de prévision au pas de temps mensuel élaboré, il est basé sur la combinaison de l'analyse en composantes principales et des chaînes de Markov.

Partant de la structure du modèle de base de l'analyse en composantes principales, il y a trois paramètres à estimer dans le modèle brute et quatre dans le modèle normé.

Pour la prévision des variables climatologiques, nous avons remarqué que la majorité des modèles proposés dans la littérature, sont basés soit sur l'observation satellitale de différents phénomènes météorologiques à court et moyen terme, soit sur l'utilisation de modèles mathématiques faisant intervenir deux, voir plusieurs variables en même temps (pluie-débit).

Pour notre part nous avons veillé à n'utiliser que la variable observée seulement.

Pour la prévision de la variable aléatoire par le modèle élaboré, deux approches ont été appliquées pour l'estimation des résidus, la première consiste à utiliser le modèle Markovien multiclasse d'ordre 1, mais les résultats obtenus n'ont pas été satisfaisants. L'introduction de la racine carrée aux valeurs observées, a permis d'améliorer les résultats du modèle de prévision, en effet une amélioration des tests numériques et graphiques appliqués a été remarquée.

Pour affiner le modèle développé, nous avons procédé dans la troisième 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 fidèlement le phénomène physique étudié.

En outre, nous remarquons que la variable normée donne des résultats meilleurs par rapport au modèle brute, cette amélioration peut être expliquée par l'intégration du paramètre stable, qui est la moyenne mensuelle interannuelle.

Pour la variable cyclique, les résultats obtenus pour toutes les approches sont amplement satisfaisants, ceci s'explique par la faible variabilité du phénomène d'évapotranspiration, l'approche basée sur le modèle normé permet une parfaite prévision (illustrée par le critère de Nash) du phénomène cyclique pris en considération.

Néanmoins les résultats trouvés restent à comparer avec d'autres techniques très récentes qui ont permis une évolution significative dans d'autres domaines scientifiques, entres autres, les réseaux de neurones et la logique floue.

Bibliographie

ANDERSON, T.W., (1958) : An Introduction to Multivariate Statistical Analysis (Second edition : 1984). J. WILEY, New York.

ANDRE, J.C., CANEILL, J.Y., DEQUE, M., ROGEL, P., TERRAY, L., TOURRE, Y., (2002) : La prévision du climat, de l'échelle saisonnière à l'échelle décennale. Comptes Rendus Géosciences, Vol., 334(16), pp. 1115-1127.

ARLERY, R., (1973) : Climatologie-méthodes et pratiques. Edition Gauthier-Villars.

BARDOSSY, A., (1990) : Note on fuzzy regression. Fuzzy sets and systems, 37, p 65-75.

BATES, B.B., (1992) : Improved methodology for parameter inference in nonlinear hydrologic regression models. Water Resources Research, 28(1), pp. 89-97.

BATHURST, J. C., O'CONNELL, P. E., (1992) : Futur of distibuted modelling. The European Hydrological System- système hydrologique européen," SHE", Hydological processes, 6, pp. 265-277.

BECKIE, R., (1996) : Measurement scale, network sampling scale, and groundwater model parameter. Water Resources Research, 32(1), pp. 65-76.

BELDRING, S., (2002) : Multi-criteria validation of a precipitation-runoff model. Journal of hydrology, 257, pp. 189-211.

BENZECRI, J.P. (1982). Histoire et préhistoire de l'analyse des données. DUNOD.

BOBEE, B. (1991). The Gamma family and drived distribution applied in hydrology. Water ressources application, Colorado, USA.

BOUZIANE, A. (1989). Contribution à la modélisation stochastique des débits mensuels et à la recherche de meilleures règles de gestion d'un réservoir. Thèse de doctorat. Université Catholique de Louvain, Belgique.

BOX, G.E.P., MÜLLER, M.E., (1958) : A note on the generation of random normal deviates. Ann. Math. Statist., 29, pp. 610-611.

BOX, G.E.P., JENKINS, G.M., (1970) : Time series analysis, forecasting and control, Holden Day, SAN FRANCISCO.

BOX, G.E.P., JENKINS, G.M., (1976) : Time series analysis, forecasting and control, Revised Edition. Holden Day, SAN FRANCISCO.

- BROCHET, P., GERBIER, N., (1972): Une méthode pratique de calcul de l'évapotranspiration potentielle. *Annales agronomiques*, 23(1), pp. 31-49.
- BROCHET, P., GERBIER, N., (1974): L'Evapotranspiration, Monographie n° 65 de la Météorologie Nationale Française. Paris.
- BUCHETELE, J., ELLIAS, V., TESAR, M., (1996): Runoff components simulated by rainfall-runoff models. *Hydrological sciences*, 41(1), pp. 49-60.
- BUIHAND, T.A., (1978): Some remarks on the use of daily rainfall data models. *Journal of hydrology*, 36, pp. 295-308.
- BURT, C., (1950): The factorial analysis of qualitative data. *British J. of Statist. Psychol.* 3, pp. 166-185.
- CAMPOLO, M., ANDREUSSI, P., SOLDATI, A., (1999): River flood forecasting with a neural network model. *Water Ressources Reasarch*, 35(4), pp. 1191-1197.
- CARLSON, R.F., MAC CORMICK, A.J.A., WATTS, D.J., (1970): Application of linear models to four annual streamflow series. *Water Ressources Reasarch*, 6(4), pp. 1070-1078.
- CAVADIAS, G., (1980): An approach to the stochastic simulation of river flows. Rapport n° 80-1, Dept. OF Civil Engineering and Applied Mechanics, Mc Gill University, Montréal;
- CAVADIAS, G., (1985): A multivariate seasonal model for streamflow simulation. Colloque sur la recherche en hydrologie au Québec. 53^{ème} congrès de l'ACFAS Chicoutimi, Canada.
- CHAOUCHE, A., PARENT, E., (1997): Modélisation de la pluie journalière en régime de mousson soudano-sahélienne. Communication personnelle.
- CHAOUCHE, K., (2001): Approche multifractale de la modélisation stochastique en hydrologie. Thèse de Doctorat. Ecole Nationale du Génie Rural, des eaux et des forêts, centre de Paris.
- CHAPELL, N., TERNAN, L., (1992): Flow path dimensionality and hydrological modelling. *Hydrological processes*, 6, pp. 327-345.
- CLARCKE, R.T., (1973): A rewiew of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of hydrology*, 19, pp. 1-20.
- COMER, G.H., (1976): An optimization technique to usdahl-74, revised model of watershed hydrology. *Water Ressources Bulletin*, 12(1), pp. 139-146.
- COULIBALY, P., ANCTIL, F., BOBEE, B., (1999): Prévigion hydrologique par réseaux de neurones artificiels. Etat de l'art. *Journal of Civi Engineering Canada* n° 26, pp. 293-301.
- COWPERTWAIT, P., (1991): Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Ressources Research*, 27(7), pp. 1431-1438.

- COWPERTWAIT, P., O'CONNELL, P., METCALFE, A., MAWDSLEY, J., (1996) : Stochastic point modelling of rainfall. *Journal of hydrology*, 175, pp. 17-65.
- DAGNELIE, P. (1998). *Statistique théorique et appliquée. Tome 1*, De Boeck Université, Bruxelles.
- DAVALO, E., NAÏM, P., (1993) : *Des réseaux de Neurones*, Edition Eyrolles, Paris.
- DECHEMI, N., BERMAD, A., HAMRICHE, A. (1994). Simulation des débits moyens mensuels en zone semi-aride par l'Analyse en Composantes Principales (ACP). *Hydrologie Continentale*, Vol. 9, N° 1, pp17-24.
- DECHEMI, N., BERMAD, A., TACHET, K. (1997). Etude et Simulation de l'Evapotranspiration (ETP) par l'Analyse en Composantes Principales (ACP). *ICID Journal*, Vol. 46, Number 2, October, pp45-60.
- DECHEMI, N. (1997). Modelisation of water balance by the simulation of potentiel evapotranspiration and rainfall for the region of Annaba. *Arab Golf Journal of scientific research*, 15 (1), pp. 63-89.
- DELAGARDE, J. (1983). *Initiation à l'analyse des données*. DUNOD.
- DELLEUR, J.W., TAO, P.C., (1976) : Seasonal and nonseasonal ARIMA models in hydrology. *ASCE, Journal of Hydrology Division*, 102, HY 10, pp. 1541-1559.
- DELLEUR, J.W., TAO, P.C., KAVVAS, M.L., (1976) : An evaluation of the practicality of some rainfall and runoff time series models. *Water Resources Reserch*, 12(5), pp. 953-970.
- DE MARSILY, G., (1994) : Quelques réflexions sur l'utilisation des modèles en hydrologie. *Revue des sciences de l'eau*, 7(1), pp. 219-234.
- DESCHESNES, J., VILLENEUVE, L.P., (1985) : Modelling the hydrological cycle : the MC model. *Nordic hydrology*, 16, pp. 257-272.
- DISKIN, M.H., SIMON, E., (1977) : A procedure for the selection of objective functions for hydrologic simulation models. *Journal of hydrology*, 34, pp. 129-149.
- DITLEVSEN, O., (1971) : *Extremes and first passage times with application in civil engineering*, Danmark Ingenioakademie, Copenhagen.
- DRIQUECH, F., EL HAMLY, M., EL KHATRI, S., MOKSSIT, A., SEBBARI, R., (1999) : *Le programme scientifique de recherche sur la prévision saisonnière des précipitations au Maroc*. La Direction de la Météorologie Nationale (DMN), Note synthétique. Maroc.
- DUAN, Q., SOROOSHIAN, S., GUPTA, V., (1992) : Effective and efficient global optimisation for conceptual rainfall-runoff models, *Water resources*, n° 28, p1015-1031.
- DUBAND, D. (1989). *Hydrologie statistique approfondie*. I.N.P.G.

- EL JABI, N., ROUSSELLE, J., (1987) : Hydrologie fondamentale. Ecole Polytechnique de Montréal.
- EAGLESON, P.S., (1978) : Climate, soil and vegetation. 2- the distribution of annual precipitation derived from observed storm sequences. *Water Resources Research*, 14(5), pp. 713-719.
- FALCONER, K.J., (1990) : Fractal geometry : Mathematical foundations and applications. John Wiley and Sons, England.
- FIORDALISO, A., (1999) : Systèmes flous et prévision de séries temporelles. Edition HERMES.
- FISHER, R.A., (1915) : Frequency distribution of the value of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, pp. 507-521.
- FRENCH, M.N., KRAJEWSKI, W.F., CUYKENDALL, R.R., (1992) : Rainfall Forecasting in Space and Time using a Neural Network, *Journal of Hydrology*, Vol. 137, p. 1-31.
- GALLOY, E., MARTIN, S., LEBRETON, A., (1982) : Analyse des séquences de jours secs consécutifs, application à 31 postes du réseau météorologique Français. *La météorologie*, 28(6), pp.5-24.
- GARRICK, M., NASH, J.E., (1978) : A criterion of efficiency for rainfall-runoff models. *Journal of hydrology*, 36, pp. 375-381.
- GREEN, J.R., (1964) : A model for rainfall occurrence. *J. R. Statist. Soc*, B(26), pp. 345-353.
- GUPTA, V.K., SOROOSHIAN, S., (1982) : Parameter estimation problems caused by model structure : case of conceptual rainfall-runoff models. Rapport technique, department of systems engineering. Case western reserve university.
- HACHE, M., RASMUSSEN, P.F., CAVADIAS, G., (1994) : The multivariate seasonal flow simulation model SVD. Application to the Ottawa river system. Rapport interne I-133, INRS-EAU, Sainte Foy, Québec.
- HARMS, CAMPBELL (1966) : Analyzing multivariate data. The Dryden Press.
- HARTE, D.S., (1997) : Multifractals, theory and application. PhD Thesis, Victoria University of Wellington.
- HIPPEL, K.W., Mc LEOD, A.I., LENNOX, W.C., (1977) : Advances in Box-Jenkins modeling, 1- Model construction. *Water Resources Research*, 13(3), pp. 567-575.
- HOTELLING, H., (1933) : Analysis of a complex statistical variables into principal component. *J. Educ. Psy.*, 24, pp. 417-441 & pp. 498-520.
- HOTELLING, H., (1936) : Relation between two sets of variables. *Biometrika*, 28, pp. 129-149.

- HSU, K., GUPTA, H., SOROOSHIAN, S., (1995) : Artificial neural network modeling of rainfall-runoff process. *Water Resources Research*, 31(10), pp. 2517-2530.
- JAMBU, M., (2000) : Méthode de base de l'analyse des données, Edition Eyrolles, Paris p. 155-185.
- JENKINSON, A.F., (1969) : Estimation of maximum floods. *World Meteorol. Organ., Tech. Note*, n° 98, Chapter 5, pp. 183-257.
- JODOUIN, J. F., (1994) : Les réseaux de neurones, Edition Hermès, Paris .
- KADI, M., (1995) : Indicateurs météorologiques de la sécheresse en méditerranée et prévision saisonnière. *Revue Techniques et Sciences* n°3.
- KATZ, R.W., (1977) : Precipitation as a chain dependent process. *Journal of applied meteorology*, 16(7), pp. 671-676.
- LANE, M., (1979) : Conditional chance – constrained model for reservoir control. *Water Resources Research*, 9, pp. 937-948.
- Le CAM, L., (1961) : A stochastic description of precipitation. In IV-th Symposium on probability and statistics, Berkeley, University of California.
- LEE, J., (1993) : A formal approach to hydrological model conceptualization. *Hydrology sciences*, 38(5), pp. 391-401.
- LEGATES, D.R., McCABE, G.J., (1999) : Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35, pp. 233-241.
- LEK, S., DIMOPOULOS, I., DERRAZ, M., ELGHACHTOUL, Y., (1996) : Modélisation de la Relation Pluie - Débit à l'aide des Réseaux de Neurones Artificiels, *Revue des Sciences de l'eau*, Vol. 3, p. 319-331.
- LETTENMAIER, D.P., BURGESS, S.J., (1977) : Operational assessment of hydrologic models of long-term persistence. *Water Resources Research*, 13(1), pp. 113-124.
- LIMA, M.I.P., (1998) : Multifractals and the temporal structure of rainfall. PhD thesis, Wageningen agricultural university, Netherlands.
- LIAMAS, J. (1985). *Hydrologie générale : principes et applications*. GAETAN MORIN, Québec.
- MAHALANOBIS, P.C., (1936) : On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*, 12, pp. 49-55.
- MANDELROT, B.B., (1965) : Une classe de processus stochastiques homothétiques à soi : application à la loi climatologique de H.E., Hurst. *Comptes Rendus de l'Académie des Sciences de Paris*, 260, pp. 3274-3277.

- MANDELBROT, B.B., VAN NESS, J.W., (1968) : Fractional brownian motions, fractional noises and application. Soc. Ind. Appl. Math. Rev., 10, pp. 422- 437.
- MANDELBROT, B.B., WALLIS, J.R., (1969) : Computer Experiments with Tractional Gaussian Noises. Water Resources Research 5(1).
- MANDELBROT, B.B., (1975) : Les objets fractals : forme, hasard, et dimension. Flammarion, Paris.
- MARKOV, A.A., (1907) : Investigation of a noteworthy case of dependent trials. Izvestiya Rossiski Akademii Nauk, 1.
- MEJIA, J.M., ROUSSELLE, J., (1976) : Disaggregation Models in Hydrology Revisited. Water Resources Research 12(2), pp. 185-208.
- METROPOLIS, N., ROSENBLUTH, M., ROSENBLUTH, A.W., TELLER, A.H., TELLER, E. (1953). Equation of state calculations by fast computing machines. Journal of chemical physics, 21, pp. 1087-1092.
- MINNS, A., HALL, J., (1996) : Artificial neural network as rainfall-runoff models. Journal des sciences hydrologiques, 41(3), pp. 399-417.
- MIQUEL, J., (2001) : Hydrologie Statistique-introduction à l'étude des processus hydrométéorologiques. Application à la prédétermination des débits de crue. ENPC, Paris.
- MOTOVILOV, Y.G., GOTTSCHALK, L., ENGELAND, K., RODHE, A., (1999) : Validation of a distributed hydrological model against spatial observations. Agricultural and Forest Meteorology 98-99, pp. 257-277.
- NASCIMENTO, N.O., (1995) : Appréciation à l'aide d'un modèle empirique des effets d'actions anthropiques sur la relation pluie-débit à l'échelle d'un bassin versant. Thèse de doctorat, ENPC. CERGRENNE.
- NASH, J.E., SUTCLIFFE, J.V., (1970) : River flow forecasting through conceptual models : Part I – a discussion of principles. Journal of hydrology, 10, pp. 282-290.
- O'CONNELL, P.E., NASH, J.E., FARREL, J.P., (1970) : River flow forecasting through conceptual models : Part II – The brasna catchment at ferbzne. Journal of hydrology, 10, pp. 317-329.
- O'CONNELL, P.E., (1971) : A simple stochastic modelling of Hurst's Law. In Mathematical Models in Hydrology, Warsaw Symposium, (IAHS Publ. 100, 1974), 1, pp. 169-187.
- OLSSON, J., (1998) : Evaluation of a scaling cascade model for temporal rainfall disaggregation. Hydrology and earth system sciences, 2(1), pp. 19-30.
- ONOF, C., WHEATER, H., ISHAM, V., (1994) : Note on the analytical expression of the inter-event time characteristics for barlett-lewis type rainfall models. Journal of hydrology, 157, pp. 197-210.

OVERNEY, O., CONSUEGRA, D., FAVRE, A.C., SAULNIER, G.M., (1995) : Estimating expected flood values with Monte-Carlo methods, Proceedings of the International Conference on Statistical and Bayesian Methods in Hydrological Sciences, Unesco, Paris.

PANKIEWICZ, G.S., (1997) : Neural network classification of convective airmasses for a flood forecasting system, International Journal of Remote Sensing, (numéro spécial sur l'utilisation des réseaux neuraux en télédétection). Meteorological Office, Bracknell, UK.

PEARSON, K., (1901) : On lines and planes of closest fit to systems of points in space. Phil. Mag. 2, n° 11, pp. 559-572.

PENMAN, H.L., (1956) : Estimating evaporation. Trans Amer. Geoph. Union. Vol. 31 n°1, pp. 43-50.

RAKEM, Y., (1999) : Analyse critique et reformulation mathématique d'un modèle pluie-débit empirique (GR4J). Thèse de doctorat, ENPC. CERGRENE.

RAO, C.R., (1964) : The use and interpretation of principal components analysis in applied research. Sankhya serie A, 26, pp. 329-357.

RODRIGUEZ-ITURBE, I., MEJIA, J.M., DAWDY, D.R., (1972) : Streamflow simulation, 2-The broken line process as a potential model for hydrologic simulation. Water Resources Research, 8, pp. 931-941.

RODRIGUEZ-ITURBE, I., GUPTA, V., WAYMIRE, E., (1984) : Scale considerations in the modeling of temporal rainfall. Water Resources Research, 20(11), pp. 1611-1619.

RODRIGUEZ-ITURBE, I., COX, D.R., ISHAM, V., (1987) : Some models for rainfall based on stochastic point processes. Proc. R. Soc. Lond., A(410), pp. 269-288.

ROESNER, L.A., YEVJEVICH, V.M., (1966) : Mathematical models for time series of monthly precipitation and monthly runoff. Colo. St. Univ. Fort Collins, Hydrol. Papers, n° 15

ROGNON, P., (1996) : Sécheresse et aridité, leur impact sur la désertification au Maghreb. Cahier de sécheresse n°7, pp. 287-297.

SAJIKUMAR, N., THANDAVESWARA, B.S., (1999) : A Non-Linear Rainfall-Runoff model using an Artificial Neural Network, Journal of Hydrology, Vol. 216, p. 32-55.

SALAS, J.D., SMITH, R.A., (1981) : Physical basis of stochastic model of annual flows. Water Resources Ressearch, 17(2), pp. 428-430.

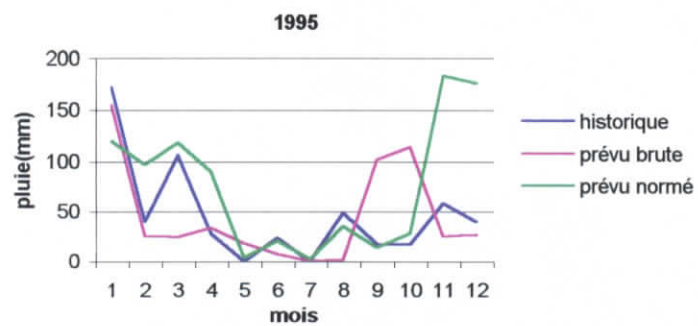
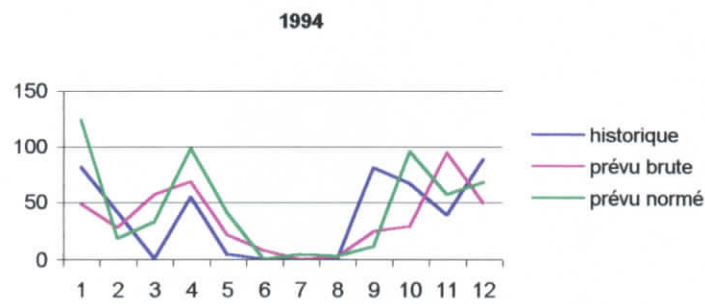
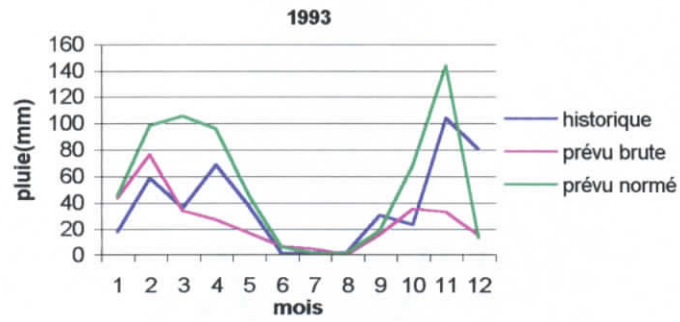
SCHERTZER, D., LOVEJOY, S., (1992) : Multifractals et géophysique, quelques aspects fondamentaux. Group for the analysis of nonlinear variability in geophysics, CNRS-URA 1367, Paris.

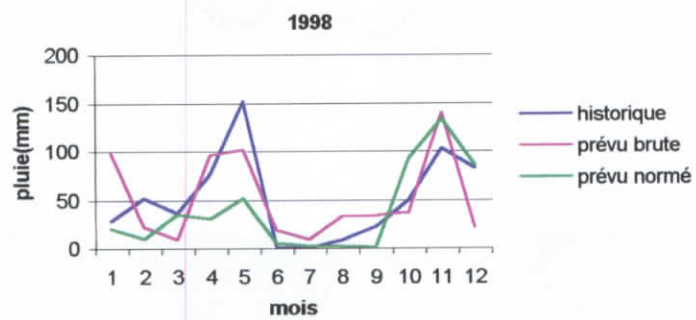
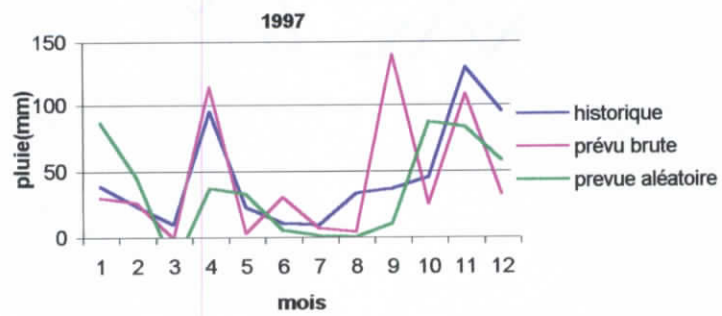
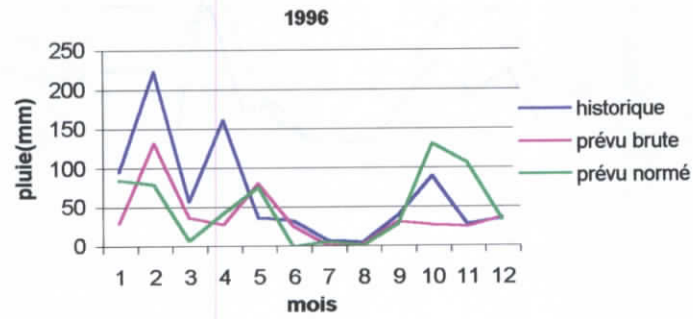
SCHERTZER, D., LOVEJOY, S., (1993) : Non linear variability in Geophysics : scaling and multifractal processes. Institut scientifique de Cargèse, Cargèse.

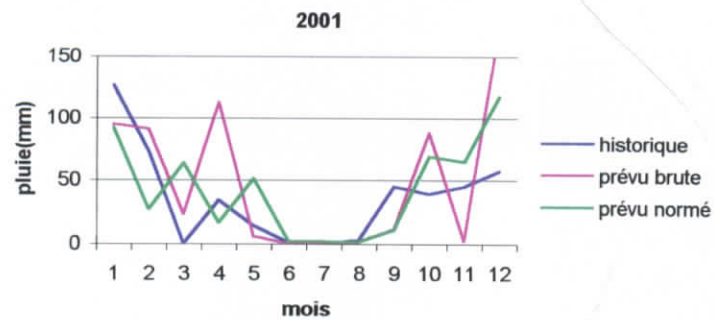
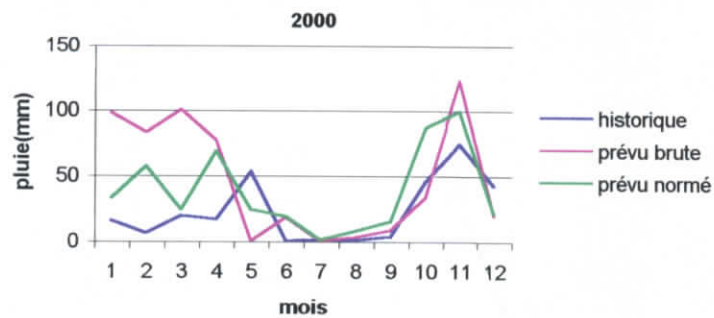
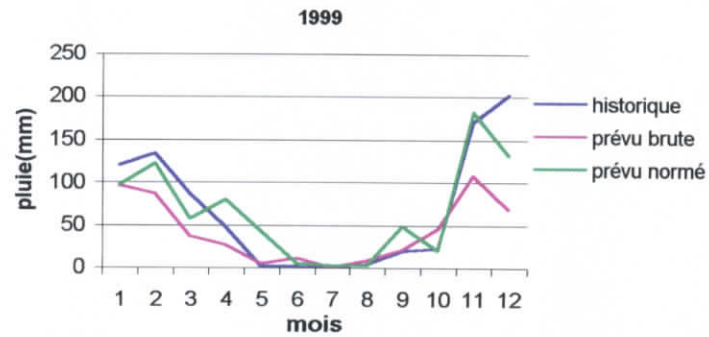
- SEIDOU, O., ROUSSELLE, J., LEFEBVRE, M., LAUZON, N., RIBEIRO, J., (2002) : Modélisation de l'incertitude sur les séquences futures de débits en rivière. *Journal des Sciences Hydrologiques*, 47(3), pp. 367–385.
- SHAMSELDIN, A. Y., (1997) : Application of a Neural Network Technique to Rainfall-Runoff Modelling, *Journal of Hydrology*, Vol. 199, p. 272-294.
- SINGH, V.P., (1988) : Hydrologic systems. Rainfall-runoff modelling. Prentice Hall.
- SMITH, J.A., KARR, A.F., (1985) : Parameter estimation for a model of space-time rainfall. *Water Resources Research*, 21(8), pp. 1251-1257.
- SPEARMAN, C., (1904) : General intelligence, objectively determined and measured. *Amer. Journal of Psychology*, 15, pp. 201- 293.
- STERN, R.D., COE, R., (1984) : A model fitting analysis of rainfall data. *J. Roy. Stat. Soc. Serie A*, 147, pp. 1-34.
- TAO, P.C., DELLEUR, J.W., (1976) : Multistation, Multiyear Synthesis of Hydrologic Time Series by Disaggregation. *Water Resources Research*, 12(6), pp. 1303-1312.
- THAUVIN, V., GAUME, E., ROUX, C., (1998) : A short time-step point rainfall stochastic model. *Water Science and Technology*, 37.
- THOMAS, H., FIERING, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In design of water resources systems (A. Mass et al., edition), pp. 459-493, Cambridge, Massachusetts, Harvard University Press.
- VALENCIA, R.D., SHAAKE, J.C., (1973) : A Disaggregation Model for Time Series Analysis and Synthesis.
- VAN – THANH, VAN - NGUYEN, (1984) : A stochastic description of temporal daily rainfall patterns. *Can. J. Civ. Eng.*, 11, pp. 234-238.
- WILKS, D.S., (1989) : Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research*, 25(6), pp. 1429-1439.
- WOOLHISER, D.A., (1991) : Modeling daily precipitations-progress and problems. In GUTTORP, P., WALDEN, A., Editors. *Statistics in the environmental and earth sciences*, pp. 71-89. GRIFFIN.
- YEVJEVICH, V.M., (1963) : fluctuations of wet and dry years. Part I, Research data assembly and mathematical models. Hydrology paper 1, Colorado State University, Fort Collins, Colorado.
- ZEALAND, C., BURN, H., SLOBODAN, SIMONOVIC, P., (1999) : Short term streamflow forecasting using artificial neural network . *J. of hydrology* 1999 N° 214. Pp 32 – 48.
- ZUCCHINI, W., ADAMSON, P., Mc NEIL, L., (1992) : A model of southern african rainfall. *Suid-Afrikaanse Tydskrif vir Wetenskap*, 88, pp. 103-109.

ANNEXE VI-1

Séries mensuelles observées et prévues de la pluviométrie
(Variables brute et normée)

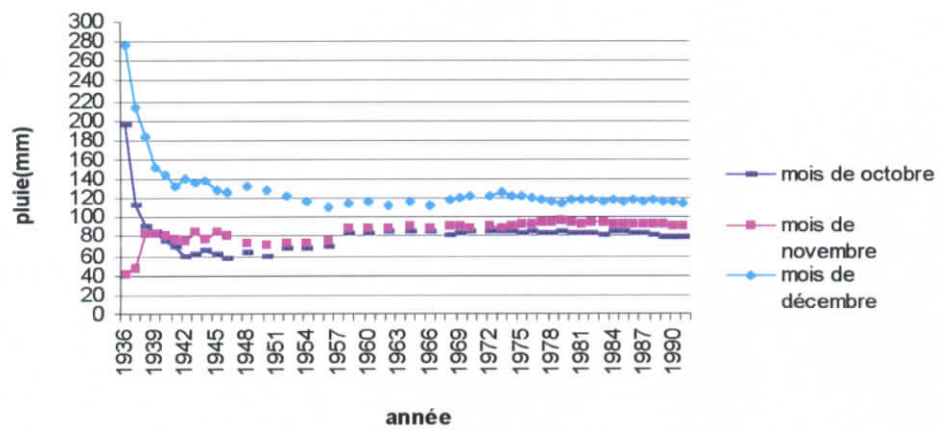
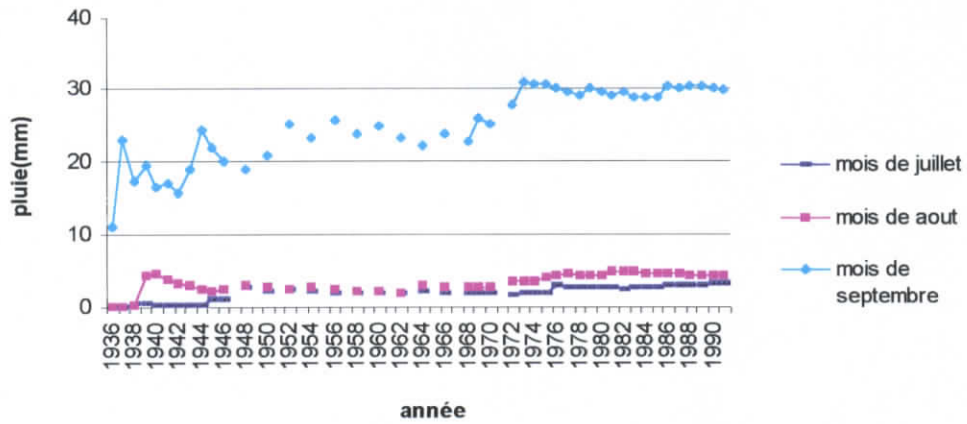
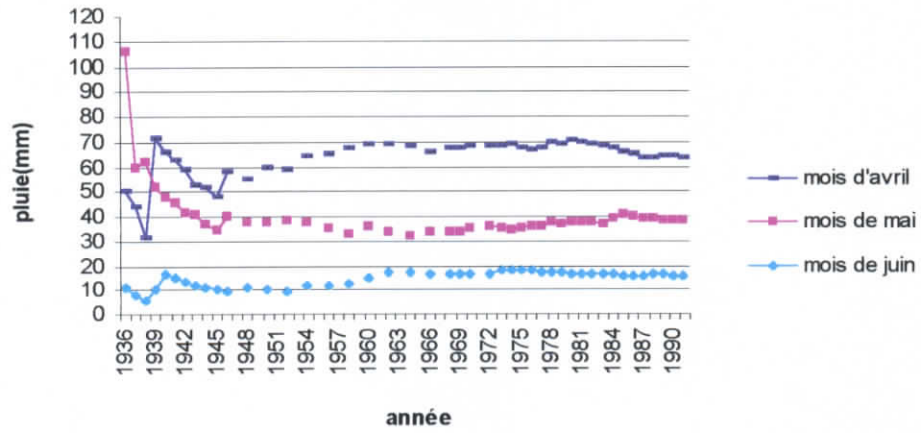






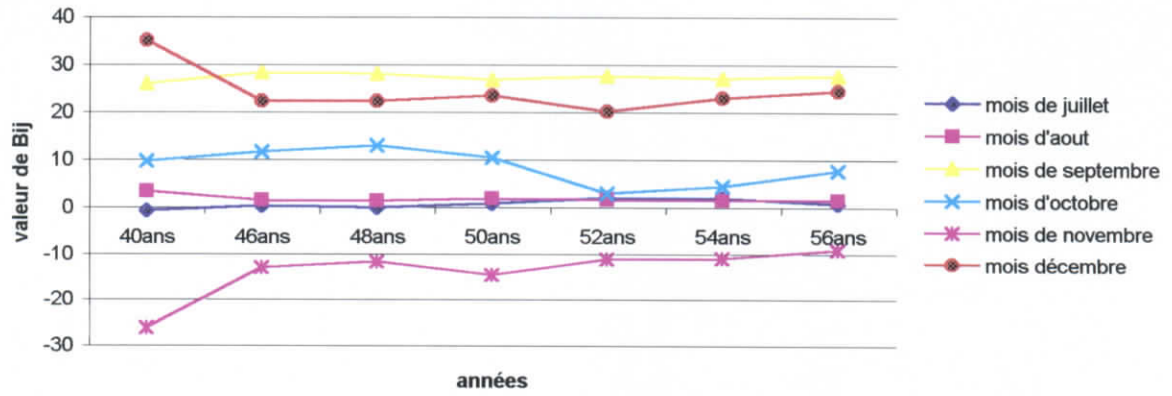
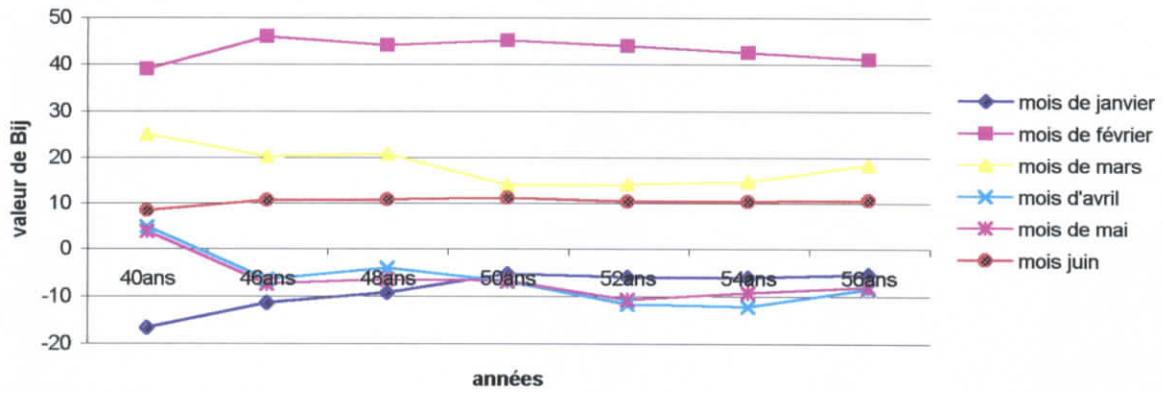
ANNEXE VI-2

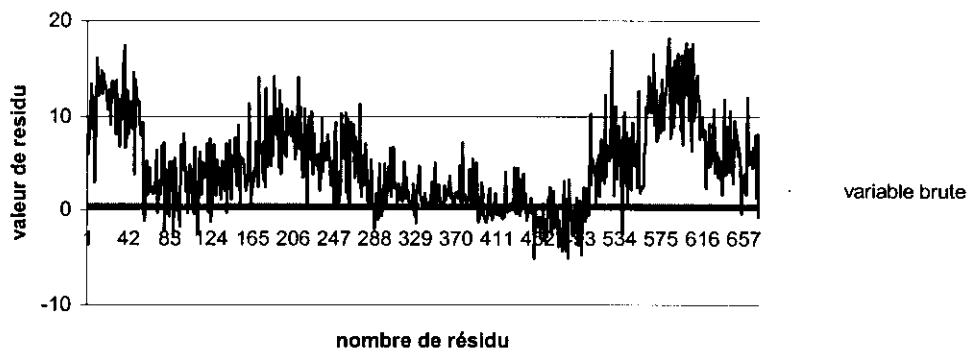
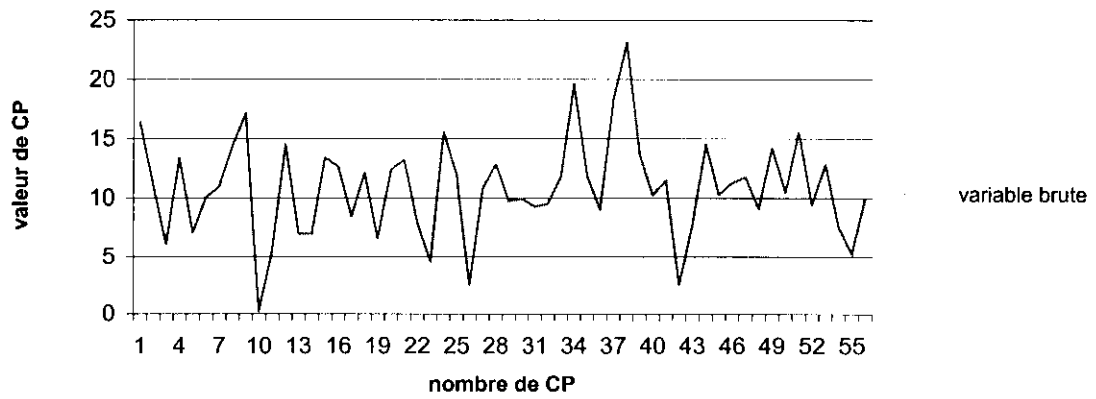
Evolution de la moyenne mensuelle interannuelle



ANNEXE VI-3

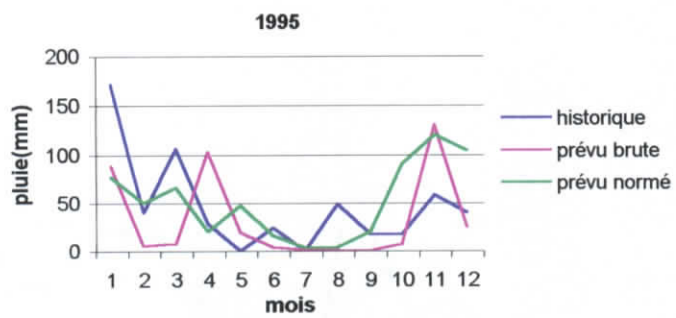
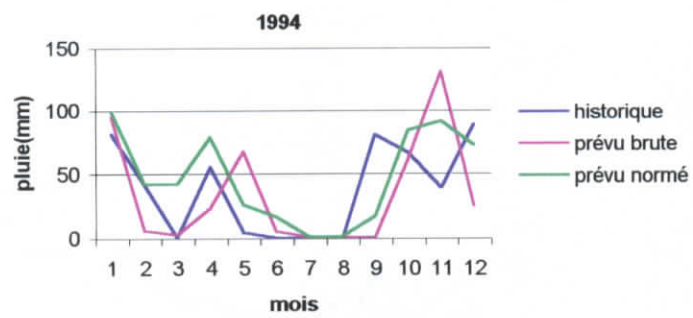
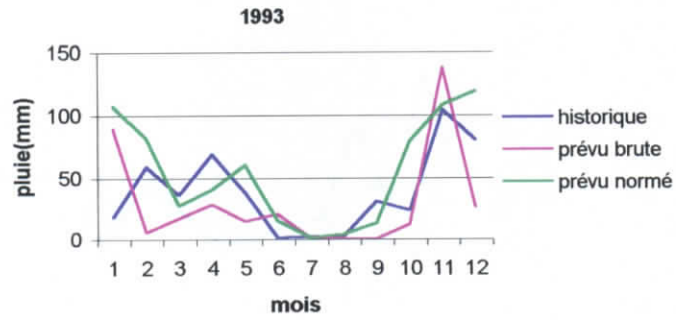
Evolution des β_{ij} mensuelle (variable racine brute)

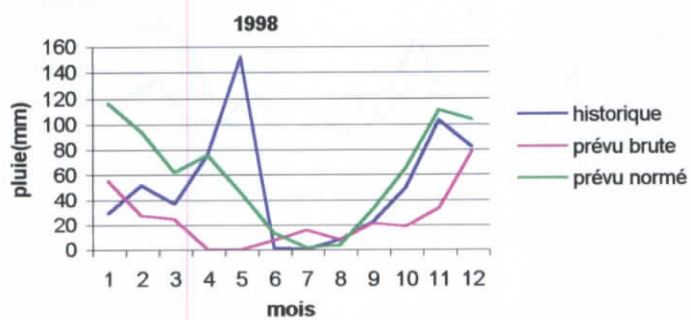
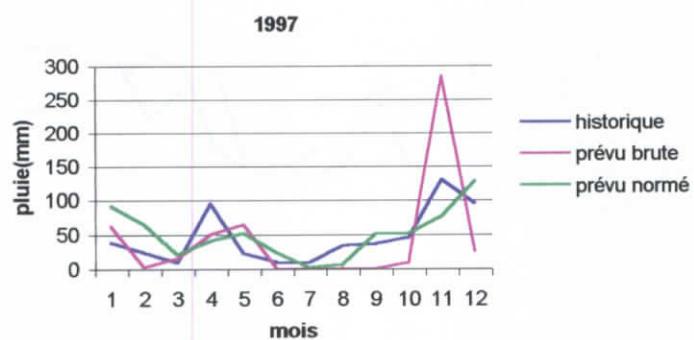
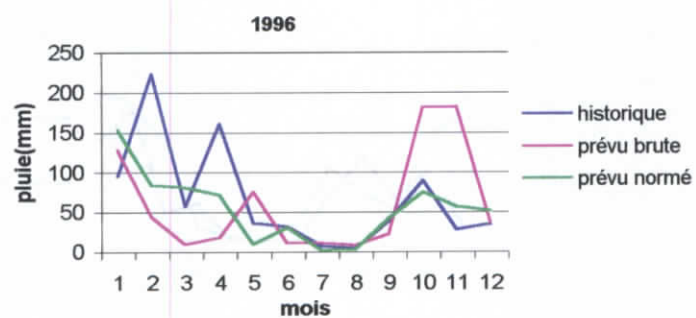


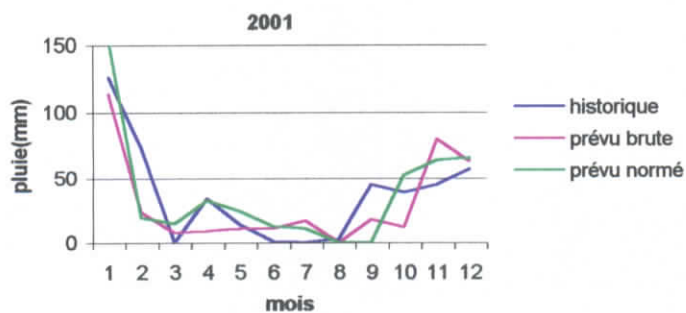
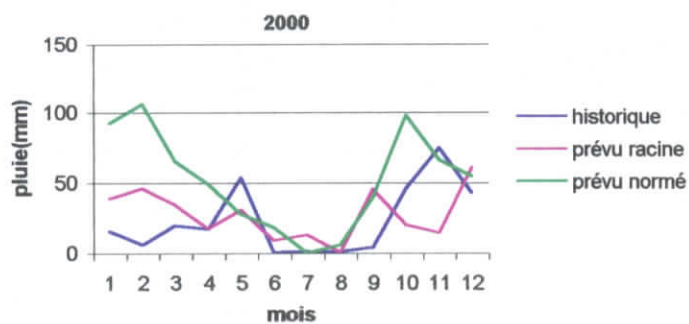
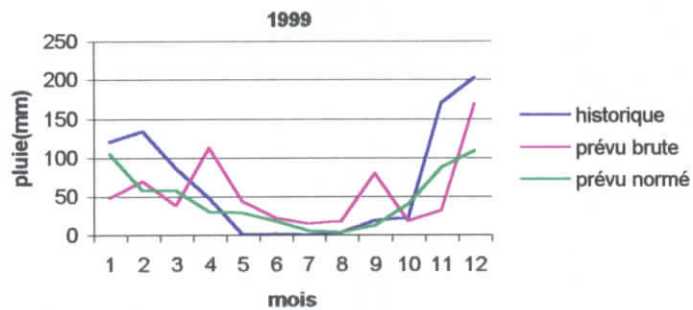


ANNEXE VI-5

Séries mensuelles observées et prévues de la pluviométrie
(Variables racines brute et normée)

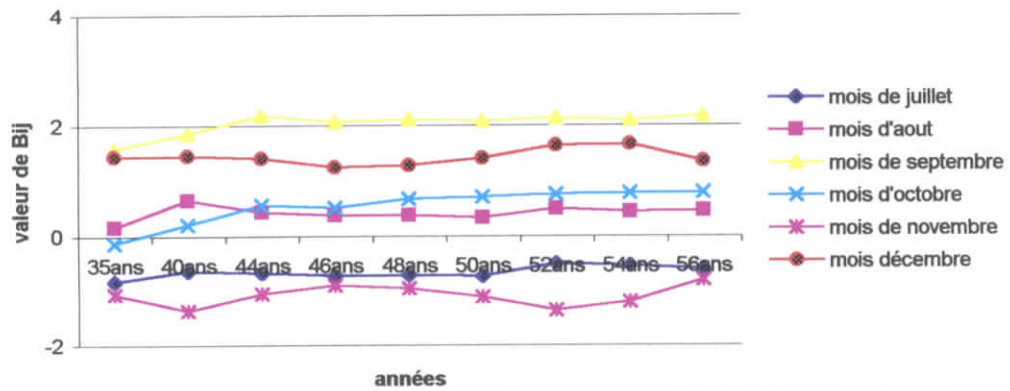
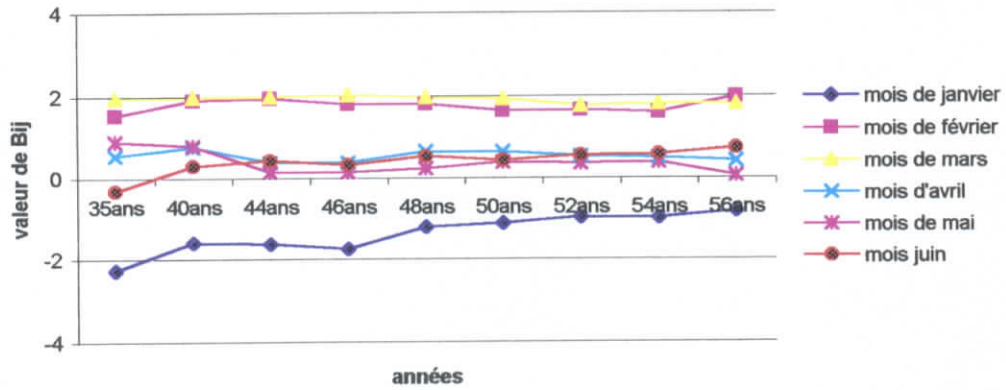






ANNEXE VI-6

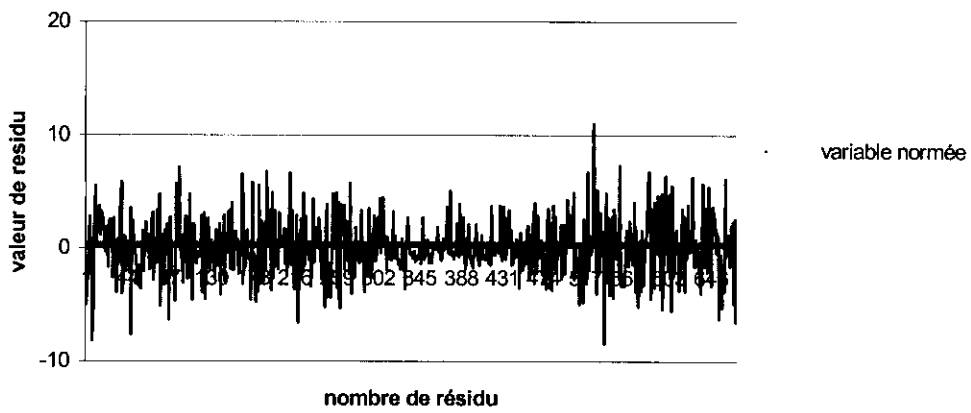
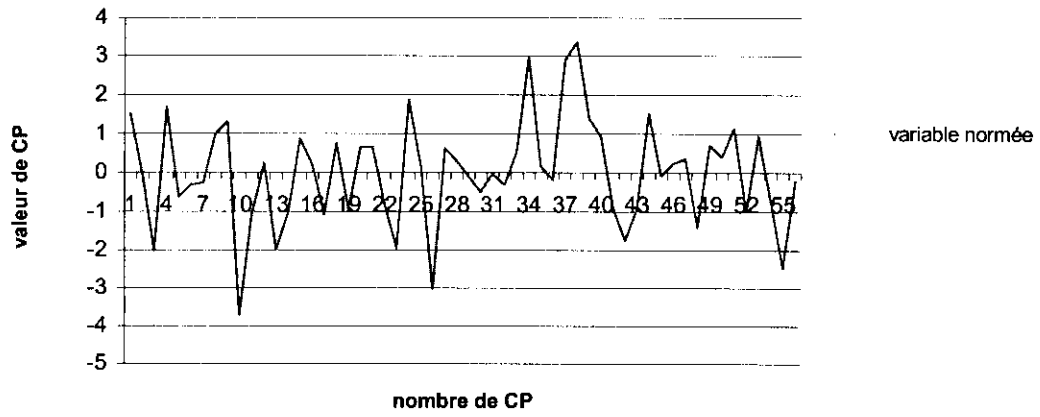
Evolution des β_{ij} mensuelle (variable racine normée)



ANNEXE VI-7

Variable racine normée

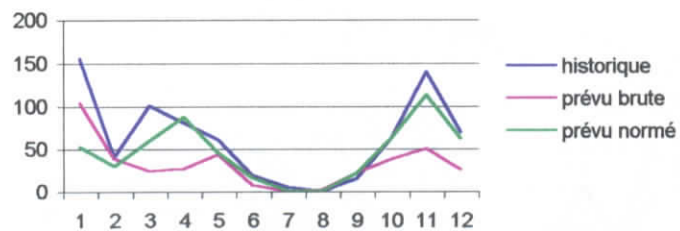
Tracés de la 1ère composante principale et des résidus



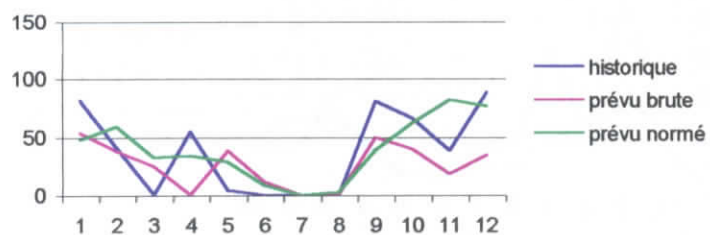
ANNEXE VI-8

Séries mensuelles observées et prévues de la pluviométrie
(Modèle multiclasse saisonnier)

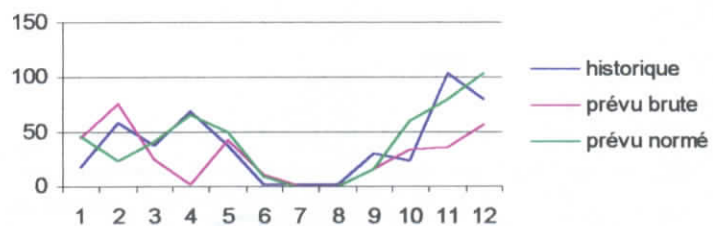
1992



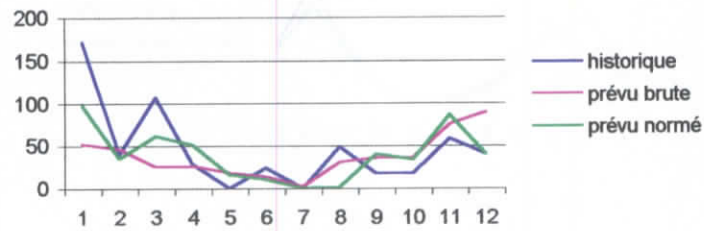
1994



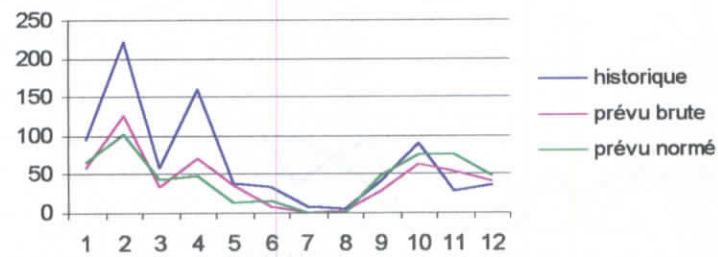
1993



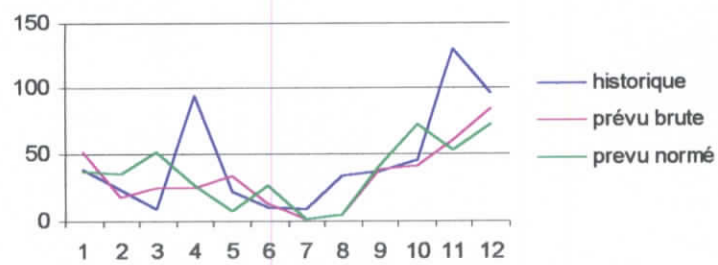
1995



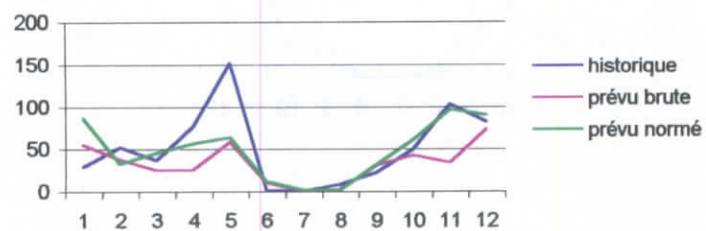
1996



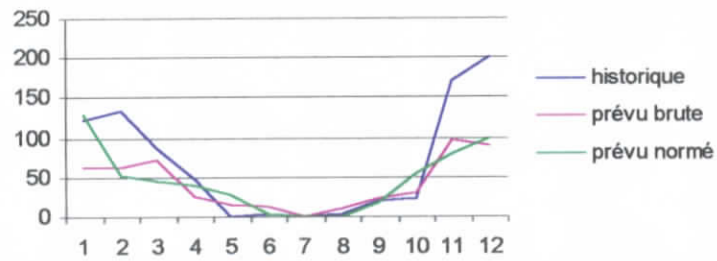
1997



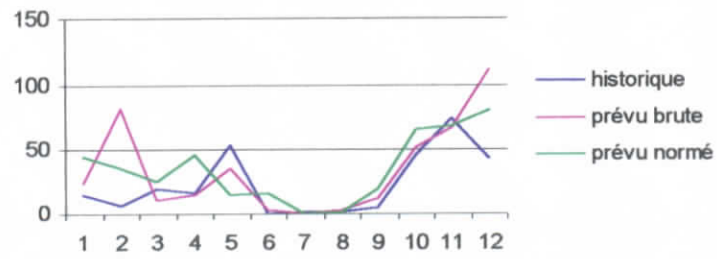
1998



1999



2000



2001

