

12/97

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département d'Hydraulique

MEMOIRE DE FIN D'ETUDES
en vue de l'obtention du diplôme
d'ingénieur d'état en Génie Hydraulique

Thème

**Contribution à la simulation des
paramètres hydrométéorologiques
aléatoires et cycliques au pas de temps journalier**

Proposé et dirigé par :
M^r. N. DECHEMI

Etudié par :
M^r. M. TOUATI
M^{lle}. A. BOUCHERIH

ERRATUM

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Erreur	Correction	Page	Ligne
β_{100}	β_{ijk}	09	
Ref. [15]	Ref. [7]	13	
Et	Est	20	13
Matrice des covariances (2 ^{ème} rep.)	Matrice des vecteurs propres		
Ref. [7]	Ref. [12]	28	
Et	Est	31	19
Ref. [12]	Ref. [7]	32	
$x \rightarrow \alpha$	$n \rightarrow \infty$	32	7
$[A]^t$	$[B]^t$	38	
Une série	D'une série	42	3
(365x50)	(50x365)	43	2
(365x22)	(22x365)	62	2
Moyennes simulées	Ecart-types simulés	73	3
K : rapport de forme	K : parametre de forme	85	
Gumble	Gumbel	85	
[15] SOUAG D. (1993) – ARMA. These d'ingenieur, ENP Alger	[15] SOUAG D. (1993) – Contribution à la gestion d'un reservoir par les modeles LDR et simulation des debits par les modeles ARMA	86	

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département d'Hydraulique

MEMOIRE DE FIN D'ETUDES
en vue de l'obtention du diplôme
d'ingénieur d'état en Génie Hydraulique

Thème

***Contribution à la simulation des
paramètres hydrométéorologiques
aléatoires et cycliques au pas de temps journalier***

Proposé et dirigé par :
M^r. N. DECHEMI

Etudié par :
M^r. M. TOUATI
M^{lle}. A. BOUCHERIH

Promotion 1997



DEDICACES

*A mes parents.
A la mémoire de mon frère Omar.
A mes sœurs et frères.
A mon beau-frère et belle-sœur.
A mes neveux et nièces.
A tous mes amis(es) que j'ai côtoyé.*

Je dédie ce mémoire.

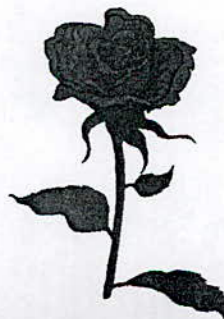
Asma

Je dédie ce modeste travail en guise de reconnaissance à :

- mes chères parents,*
- mon sacré petit frère,*
- toute ma famille,*
- mon meilleur ami Amer,*
- mes tendres amis(es),*

Et sans oublier mon fidèle compagnon Scorpi.

Moh



REMERCEMENTS

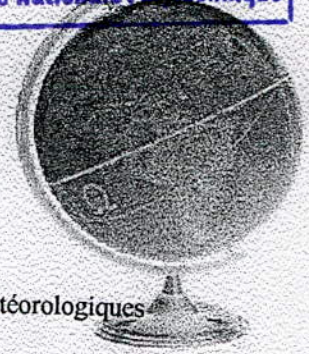


المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Nous tenons à remercier vivement nos chères parents, pour leurs soutient moral et d'avoir mis en notre disposition tous les moyens nécessaires.

Nous remercions chaleureusement les perles de l'Ecole Nationale Polytechnique, Krimo, Salah et leurs collègues de la bibliothèque centrale, pour leur gentillesse inouïe.

Enfin, nous offrons une rose à chaque personne ayant participé de loin ou de prêt à l'élaboration de ce mémoire de fin d'études.



RESUME

L'objectif de notre travail est la simulation des paramètres hydrométéorologiques cycliques (températures) et aléatoires (pluies) au pas de temps journalier.

Pour cela, on a utilisé le modèle de simulation GESTOP (brute et normé) basé sur l'analyse en composantes principales (ACP), afin de générer des séries synthétiques qui sont utilisées comme entrées aux modèles d'optimisation de la gestion de la ressource en eau.

Le modèle de simulation GESTOP développé, permet d'atténuer les inconvénients des autres modèles, à savoir le pas de temps, la stationnarité et la chronologie.

ABSTRACT

The aim of our project is the simulation of the hydroclimatologicals cyclical parameters (temperature) and random parameters (rainfall) on daily step time.

We have used for that, the simulation model GESTOP (brut and normed) based on the principal components analysis (PCA) to generate a synthetical series, who are used as input to management models of water resources.

The GESTOP model of simulation developed, permit to mitigate the draw back of other models that is the step time, stationnarity and chronology.

ملخص

الهدف من عملنا هذا هو تمثيل العوامل الهيدروجوية الدورية (درجة الحرارة) و العشوائية (الأمطار) و ذلك في مجال الزمن اليومي.

استعملنا لذلك، النموذج GESTOP (الطبيعي و الموحد) أساسه التحليل بالمركبات الأساسية ليتم به تركيب سلاسل جديدة مستعملة كمدخل للنماذج المثالية لتسيير موارد المياه.

نموذج التمثيل GESTOP الموسع للدراسة، تستطيع التخفيف من عيوب النماذج الأخرى الموجودة وذلك فيما يخص المجال الزمني، الإستقرارية و الأحداث التاريخية.

Sommaire

PARTIE THEORIQUE

INTRODUCTION

I. MODELES DE SIMULATION EN HYDROLOGIE	2
I.1. INTRODUCTION	2
I.2. LES MODÈLES SAISONNIERS DE THOMAS - FIERING	3
I.3. LE MODÈLE DU BRUIT GAUSSIEN FRACTIONNEL (FGN) ^[7]	3
I.4. LE MODÈLE DE LA LIGNE BRISÉE (BROKEN LINE) ^[7]	5
I.5. LES MODÈLES AUTOREGRESSIFS	6
I.5.1. Formulation mathématique du modèle	6
I.5.1.a. Les modèles autorégressifs à paramètres constants	6
I.5.1.b. Les modèles autorégressifs à paramètres périodiques	7
I.6. LES MODÈLES MARKOVIENS	8
I.6.1. Définition	8
I.6.2. Dédution analytique des paramètres	8
I.6.2.a. Processus d'ordre 1	8
I.6.2.b. Processus d'ordre 2	9
I.6.3. Principe de la simulation	10
I.7. LES MODÈLES DE DÉSAGRÉGATION	10
I.7.1. Descriptions du modèle de désagrégation	11
I.7.1.a. Modèle général de désagrégation	11
I.7.1.b. Modèle temporel de désagrégation à site unique	11
I.7.1.b.i. Le modèle de base	11
I.7.1.b.ii. Le modèle étendu	12
I.7.1.b.iii. Le modèle condensé	12
I.8. LES MODÈLES AUTOREGRESSIFS A MOYENNES MOBILE ARMA ^[15]	13
I.8.1. Formulation mathématique	13
II. ANALYSE EN COMPOSANTES PRINCIPALES	15
II.1. DÉFINITIONS	15
II.2. INTERPRÉTATIONS	17
II.3. CALCUL DES COMPOSANTES PRINCIPALES	18
II.3.1. Calcul de la matrice des covariances	18
II.3.2. Calculs des valeurs et vecteurs propres	18
II.3.2.a. Théorie de la méthode itérative de la puissance	18
II.3.2.a.i. Calcul de la première valeur propre et le premier vecteur propre	18
II.3.2.a.ii. Calcul de λ_2 et b_2	19
II.3.2.a.iii. Calcul de λ_3 et b_3	20
II.3.3. Composantes principales	20
II.4. REPRÉSENTATIONS GRAPHIQUES	21
II.5. DIAGRAMME DE CALCUL	22
II.6. EXEMPLE D'APPLICATION	23

III. LE MODÈLE GESTOP DE SIMULATION	25
III.1. DÉFINITION.	25
III.2. TECHNIQUES DE SIMULATION.	26
III.2.1. Simulation par la fonction de répartition ^[6]	26
III.2.2. Simulation par les chaînes de Markov ^[6]	27
III.2.2.a. Construction de la chaîne de Markov.	27
III.2.2.b. Génération par les séries de Markov.	28
III.2.3. Simulation par les lois d'ajustement.	28
III.2.3.a. Tests statistiques ^[7]	28
III.2.3.a.i. Test d'indépendance.	28
III.2.3.a.ii. L'homogénéité.	29
III.2.3.a.iii. La stationnarité.	30
III.2.3.b. Simulation par la loi normale ^[6]	31
III.2.3.c. Simulation par la loi G.E.V ^[12]	32
III.3. MODÈLE GESTOP BRUT.	36
III.4. MODÈLE GESTOP NORMÉ.	38
III.5. DIAGRAMME DE SIMULATION.	40

PARTIE APPLICATION

IV. SIMULATION DE LA PLUVIOMÉTRIE AU PAS DE TEMPS JOUR. PAR LE MODÈLE GESTOP	42
IV.1. DONNÉES UTILISÉES.	42
IV.2. ANALYSE DES DONNÉES PAR L'A.C.P.	43
RECONSTITUTION.	48
IV.4. AJUSTEMENTS.	49
IV.5. ETUDE DES RÉSULTATS.	57
IV.5.1. Cumul annuel.	57
IV.5.2. Moyennes annuelles.	58
IV.5.3. Ecart-types annuels.	59
IV.5.4. Maximums annuels.	60
V. SIMULATION DES TEMPERATURE AU PAS DE TEMPS JOUR. PAR LE MODÈLE GESTOP	61
V.1. DONNÉES UTILISÉES.	61
V.2. ANALYSE DES DONNÉES PAR L'A.C.P.	62
V.3. RECONSTITUTION.	65
V.4. AJUSTEMENTS.	66
V.5. ETUDE DES RÉSULTATS.	72
V.5.1. Moyennes annuelles.	72
V.5.2. Ecart-types annuels.	73
V.5.3. Maximums Annuels.	74

CONCLUSIONS ET PERSPECTIVES

ANNEXE I : Eléments statistiques de base

ANNEXE II : Calcul matriciel

ANNEXE III : Ecriture matricielle des éléments statistiques

APPENDICE : Symboles et abréviations

BIBLIOGRAPHIE

Introduction générale

L'optimisation de la gestion des barrages réservoirs nécessite comme entrées la modélisation des séries synthétiques de débits, ces débits sont générés par des pluies, par le biais des fonctions de transfert, s'il y a des séries simulées de pluie, on peut aboutir à des séries synthétiques de débits.

Les barrages sont conçus pour satisfaire les demandes domestiques (A.E.P.), industrielles et agricoles, cette dernière est quantifiée par le biais de la demande climatique, elle-même obtenue à partir de l'étude chronologique des évapotranspirations.

Les caractéristiques des séries chronologiques des paramètres étudiés (pluviométrie, température) sont totalement distincts, en effet la pluviométrie suit un phénomène totalement aléatoire contrairement à la température qui est un phénomène cyclique.

Dans la première partie de ce mémoire, et dans le premier chapitre, on a présenté quelques modèles de simulation existants en hydrologie, la théorie de l'analyse en composantes principales (A.C.P.) est présentée dans le deuxième chapitre, et enfin dans le troisième chapitre, on a présenté les techniques de simulation, et le modèle GESTOP permettant la simulation des séries chronologiques.

La deuxième partie est consacrée à l'application et la présentation des résultats du modèle de simulation GESTOP au pas de temps journalier des variables aléatoires (pluies) et cycliques (températures).

Chapitre 1

Les modèles de simulation en hydrologie

CHAPITRE I

Modèles de simulation en hydrologie

La plupart des modèles de simulation en hydrologie contenue dans ce chapitre sont connus, nous ferons qu'apporter une brève description sur ces différents modèles.

1.1. Introduction.

La simulation a fait l'objet de nombreux sujets de recherche, elle répond à la nécessité de générer de séries chronologiques similaires aux séries historiques, dans le sens où elle préserve les caractéristiques statistiques de ces dernières.

La génération des séries synthétiques, trouve son application dans la recherche des méthodes d'optimisation de la gestion, en général la simulation est destinée au design, la planification et les études opérationnelles des schémas de gestion de la ressource en eau.

Différents types de modèles ont été proposés pour la modélisation des séries hydrométéorologiques, parmi cette panoplie de modèles, on trouve le modèle autoregressif de Thomas et Fiering (1962), le modèle du bruit Gaussien fractionnel [FGN] de Mandelbrot et Wallis (1965), le modèle autoregressif à moyenne mobile [ARMA] de Carlson et Al (1970), de O'connel (1971) et de Hipel et Al (1977), le modèle de la ligne brisée de Ditlevson (1969) et les modèles ARMA-Markov de Lettenmaier et Burges (1977).

1.2. Les modèles saisonniers de Thomas - Fiering.

En hydrologie, les modèles saisonniers appliqués aux séries mensuelles sont à l'origine utilisés par Thomas et Fiering (1962) et par Roesner et Yevjevich (1966) en utilisant la méthode de Yevjevich (1966).

Les $X_{t, \tau}$ représente le débit au mois t , $t = 1, 2, 3, \dots, 12N$ (N : le nombre d'années) correspondant aux mois τ , $\tau = 1, 2, 3, \dots, N$.

Les modèles saisonniers de Thomas -Fiering prennent la forme suivante :

$$X_{t, \tau} = \mu_{\tau} + \sigma_{\tau} \rho_{\tau} (X_{t-1, \tau-1} - \mu_{\tau-1}) / \sigma_{\tau-1} + \sigma_{\tau} \eta_t$$

avec :

$$\rho_{\tau} = E\{(X_{t, \tau} - \mu_{\tau})(X_{t-1, \tau-1} - \mu_{\tau-1})\} / \sigma_{\tau} \sigma_{\tau-1}$$

C'est le coefficient de corrélation entre les mois τ et $\tau-1$.

Pour la composante aléatoire $E(\eta_t) = E(\eta_t \eta_{t-k}) = 0$ pour $k \neq 0$ et $\text{var}(\eta_t) = 1 - \rho_{\tau}^2$

Le coefficient d'inclinaison appliqué à la composante aléatoire au mois τ est donné par :

$$\gamma_{\eta, \tau} = (\gamma_{\tau} - \rho_{\tau}^3 \gamma_{\tau-1}) / (1 - \rho_{\tau}^2)^{3/2}$$

Où γ_{τ} est le coefficient d'inclinaison.

L'inconvénient ici, c'est que pour les séries d'observations assez courtes les valeurs de ρ_{τ} sont très différentes, et on n'arrive pas à les estimer quelques fois, mais ρ_{τ} peut prendre la valeur nulle pour des mois particuliers de l'année.

1.3. Le modèle du bruit Gaussien fractionnel (FGN) [7].

Après que les résultats des études de Hurst ont été connus, Mandelbrot (1965) a construit une classe de processus aléatoires connus sous le nom de Fractionnel Gaussien Noise (FGN).

Les processus FGN diffèrent dans leur structures autocorrélatives des processus ARMA, la différence de base consiste dans leurs mesures relatives de persistance, qui, si on suit la définition de Taylor (1938), peut être défini comme l'intégral de la fonction d'autocorrélation entre les bornes 0 et ∞ , la persistance est finie pour tous types de processus ARMA, au contraire l'intégral de la fonction d'autocorrélation du processus FGN est infinie.

1.4. Le modèle de la ligne brisée (Broken Line) [7].

Le processus simple de la ligne brisée était mis au point par Ditlevson (1969), il est dérivé de l'interpolation linéaire au temps t entre les points A, B, C,... J, K, L,... d'ordonnées normalement distribuées, de moyenne nulle et de variance σ^2_i , ces points ont des projections égales sur l'axe de temps de distance a_i .

Le processus simple de la ligne brisée $\beta_i(t)$ est représenté sur la figure ci-dessous.

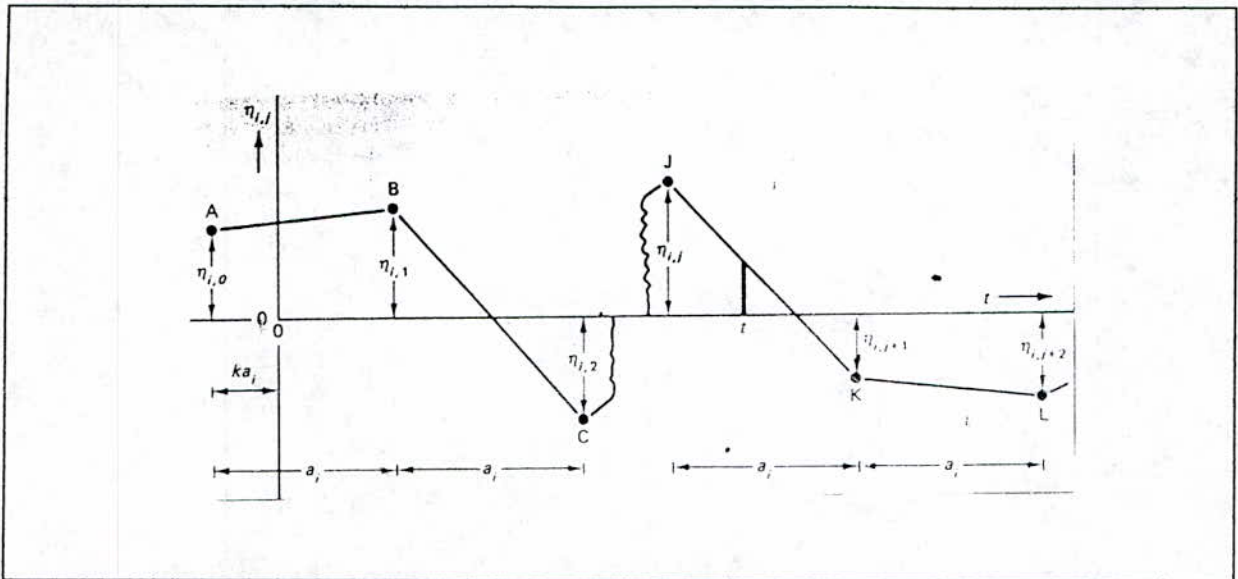


Figure I-1 : Représentation du Processus simple de la ligne brisée.

La figure I-1 représente le processus simple de la ligne brisée, il est formé par l'interpolation au temps t entre les points A, B, C,... J, K, L, d'abscisses $-ka_i$, a_i-ka_i , $2a_i-ka_i$,... etc.

Le point A est placé avant l'origine de l'axe des temps d'une quantité ka_i où k est uniformément distribué dans l'intervalle $(0, 1)$.

Il s'ensuit en général,

$$\beta_i(t) = \eta_{i,j} + \frac{t + ka_i - ja_i}{a_i} (\eta_{i,j+1} - \eta_{i,j})$$

$$ja_i - ka_i \leq t \leq (j+1)a_i - ka_i$$

généralement, on indique que $\text{var}\{\beta_i(t)\} = 2\sigma_i^2/3$

Le processus de ligne brisée ξ_t est formé par la somme de n processus simples de lignes brisées $\beta_i(t)$, comme suit :

$$\xi_t = \sum_{i=1}^n \beta_i(t)$$

Pour appliquer ce modèle, on a besoin de déterminer n , a_i , σ_i^2 , $i=1, 2, \dots, n$, ça fait au total $2.n + 1$ paramètres à déterminer, et ceci est un inconvénient à cause du nombre élevé de paramètres à estimer.

1.5. Les modèles autoregressifs.

Les modèles autorégressifs ont été utilisés depuis le début des années 1960 en hydrologie pour la modélisation des séries temporelles annuelles et périodiques, l'application de ces modèles a été intéressante en hydrologie à cause de :

- La forme autorégressive qui est une caractéristique des séries étudiées.
- La simplicité d'utilisation.

La procédure utilisée pour l'estimation des paramètres des modèles était basée sur la méthode des moments, et la méthode aux moindres carrés.

1.5.1. Formulation mathématique du modèle.

1.5.1.a. Les modèles autorégressifs à paramètres constants

Soit Y une série temporelle normalement distribuée de moyenne μ et d'écart type σ avec une structure de dépendance. Le modèle autorégressif d'ordre p noté $AR(p)$ représente la variable Y_t s'écrit généralement comme suit :

$$Y_t = \mu + \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + \dots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t$$

$$Y_t = \mu + \sum_{j=1}^p \phi_j (Y_{t-j} - \mu) + \varepsilon_t$$

où :

- ε_t : Série indépendante normalement distribuée de moyenne nulle et d'écart type σ_ε .
 - ϕ_1, \dots, ϕ_p : Paramètres autorégressifs.
- Les paramètres du modèle estimés à partir des données sont $\mu, \sigma_\varepsilon, \phi_1, \dots, \phi_p$.

I.5.1.b. Les modèles autorégressifs à paramètres périodiques

Le modèle autorégressif AR(p) à moyenne et variance périodique et de coefficients autorégressifs constants, peut être représenté par :

$$Y_{v, \tau} = \mu_{\tau} + \sigma_{\tau} Z_{v, \tau}$$

$$Z_{v, \tau} = \sum_{j=1}^p \phi_j Z_{v, \tau-j} + \varepsilon_{v, \tau}$$

où :

$Y_{v, \tau}$: Variable dépendante pour l'année et le temps (mois, semaine, ...) avec $\tau = 1, 2, \dots, \omega$ et ω est le nombre d'intervalles de temps.

$Y_{v, \tau}$: Supposé normalement distribué de moyenne μ_{τ} et de variance σ_{τ}^2 et a une structure autorégressive d'ordre p.

$Z_{v, \tau}$: Variable dépendante normalement distribuée de moyenne nulle et de variance unité.

$\varepsilon_{v, \tau}$: Variable indépendante de moyenne nulle et de variance σ_{ε}^2 .

Les paramètres du modèle sont : μ_{τ} , $\phi_1 \dots \phi_p$ et σ_{ε}^2 .

Un modèle autorégressif d'ordre p à coefficients périodiques est donné par :

$$Z_{v, \tau} = \sum_{j=1}^p \phi_{j, \tau} Z_{v, \tau-j} + \sigma_{\varepsilon\tau} \xi_{v, \tau}$$

où :

$\phi_{j, \tau}$: Coefficient autorégressif périodique au temps.

$\sigma_{\varepsilon\tau}$: Coefficient périodique.

$\xi_{v, \tau}$: Variable indépendante standardisée.

Les paramètres du modèle sont : μ_{τ} , σ_{τ}^2 , $\phi_{1\tau}, \dots, \phi_{p\tau}$ et $\sigma_{\varepsilon\tau}^2$, $\tau = 1, 2, \dots, \omega$.

I.6. Les modèles Markoviens.

I.6.1. Définition.

Une chaîne de Markov est un système qui subit au cours du temps des changements d'états à transition aléatoire, et qui sans être dénué de mémoire ne garde de son passé que le souvenir le plus récent.

Une chaîne de Markov est dite d'ordre k si l'équation ci - dessous est vérifiée

$$P\{X_t, X_{t-1}, \dots, X_{t-k}, X_{t-k-1}\} = P\{X_t, X_{t-1}, \dots, X_{t-k+1}, X_{t-k}\}$$

Où X est une variable aléatoire.

Pour l'ordre k , la réalisation d'un état donné ne dépend que des k réalisations précédentes, dans ce qui suit la pluviométrie (phénomène aléatoire) est prise comme exemple.

I.6.2. Déduction analytique des paramètres

Les précipitations sont décrites par le processus stochastique suivant :

$$Z_t = X_t Y_t$$

X_t : Caractérise l'état au jour t ($X_t = 0$: jour sec, $X_t = 1$: jour pluvieux)

Y_t : Hauteur de la pluie (si $X_t = 1$).

Pour la détermination de la matrice de passage, on s'intéresse seulement à la variable X , qui décrit le passage d'un état E_i à un état E_j .

I.6.2.a. Processus d'ordre 1.

Soit X la variable aléatoire caractérisant l'état du jour, X peut prendre deux valeurs (système binaire), 0 pour l'état sec et 1 pour l'état pluvieux.

Ce processus va être caractérisé par les probabilités conditionnelles de passage de l'état de la veille à l'état d'aujourd'hui.

Ainsi si l'état au jour k ne dépend que de l'état au jour $k-1$, on note :

$$\text{Prob}(X_k = j / X_{k-1} = i) = \alpha_{ij}$$

α_{ij} : Représente la probabilité d'obtenir au jour k l'état j , sachant qu'au jour $k-1$, on avait l'état i .

En considérant deux états possibles, on obtient une matrice de passage P présentée ci-contre :

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

La matrice P ayant les propriétés d'une matrice stochastique implique la relation :

$$\alpha_{i0} + \alpha_{i1} = 1$$

Les coefficients de la matrice P sont obtenus par le décompte des jours caractérisés par les états (0,1).

La probabilité marginale (ou inconditionnelle) P_0 peut s'écrire comme suit :

$$P_0 = \frac{\alpha_{10}}{1 - \alpha_{00} + \alpha_{10}}$$

Compte tenu du caractère booléen de la variable aléatoire, le coefficient d'autocorrélation s'écrit :

$$\rho = \frac{\alpha_{00} - P_0}{1 - P_0} = \frac{\alpha_{11} - (1 - P_0)}{P_0}$$

1.6.2.b. Processus d'ordre 2.

On cherche la probabilité de l'état de la classe 0 ou 1 d'aujourd'hui, en fonction des états connus d'hier et d'avant hier.

La matrice de transition est présentée par le tableau suivant :

Etats au jours k-1 et k-2	Etats au jour 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_{100} : Représente la probabilité conditionnelle d'obtenir un doublet de classe (i, j), 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 vu la succession de certains doublets.

1.6.3. Principe de la simulation.

La simulation aboutit à l'obtention de séries synthétiques à partir d'une série historique plus au moins longue.

Comme on a vu précédemment la variable s'écrit de la manière suivante :

$$Z_t = X_t Y_t$$

Pour l'ordre 1, à partir des probabilités conditionnelles α_{00} et α_{10} , on génère l'état journalier des séries synthétiques après tirage d'un nombre aléatoire compris entre 0 et 1 (et constituant une probabilité) qu'on compare à α_{i0} ($i=0$ ou 1), le jour simulé est sec, sinon il est pluvieux.

Le principe de simulation est similaire pour les ordres supérieurs dans ce cas la probabilité obtenue par tirage aléatoire sera comparée aux probabilités conditionnelles de la matrice de passage correspondant à l'ordre du processus de simulation.

S'il s'avère que le jour simulé est pluvieux, on lui attribue une hauteur de pluie qu'on tire des fonctions de répartition conditionnelles.

En tenant compte de la position du jour humide dans un épisode donné, on donne à chaque jour simulé pluvieux une hauteur de pluie, cela se fait par le biais d'un deuxième tirage d'un nombre aléatoire qui sera confronté à la fonction de répartition considérée.

1.7. Les modèles de désagrégation.

Les modèles de désagrégation ont été introduit pour la première fois pour la modélisation en hydrologie par Harms et Cambdell (1969), mais le premier modèle qui fut accepté est celui présenté par Valencia et Shaake (1973) ce modèle est de base de technique de désagrégation dans la simulation des séries hydrologiques.

Medjia et Rousselle ont proposé des modifications du modèle de base pour surmonter les défauts de ce dernier, Lane (1979) a aussi proposé quelques améliorations supplémentaires.

La plupart des applications de la désagrégation ont été dans le domaine temporel, Lane (1979) a appliqué les mêmes principes dans le domaine spatial.

1.7.1. Descriptions du modèle de désagrégation.

Le modèle de désagrégation 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 désagrégé en sous séries.

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

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

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.

1.7.1.a. Modèle général de désagrégation.

Le modèle de dépendance linéaire est donné par

$$Y = A.X + B.\varepsilon$$

Y : Séries d'observations à simuler (sous séries).

X : Série clé.

ε : Terme stochastique.

1.7.1.b. Modèle temporel de désagrégation à site unique.

Trois formes du modèle de désagrégation temporel à site unique.

1.7.1.b.i. Le modèle de base.

Celui de Valencia et Shaake proposé en (1979), il est de la forme :

$$Y = AX + B\varepsilon$$

X : Série mère.

Y : Série des valeurs saisonnières.

A, B : Paramètres du modèle.

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

1.7.1.b.ii. Le modèle étendu.

Développé par Medjia et Rousselle (1976), c'est une extension du modèle de base, un terme supplémentaire a été inclus pour préserver les covariances saisonnières entre saisons successives, il a la forme :

$$Y = A.X + B.Y + C.Z$$

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

C : Matrice de paramètres supplémentaires.

Le nombre de paramètres du modèle est excessif, et ceci rend l'estimation de ces paramètres, plus compliqué.

1.7.1.b.iii. 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

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

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

Le principal avantage de ce modèle est la réduction du nombre de paramètres, mais il a deux inconvénients majeurs : le premier, est que le modèle n'est pas aussi clair et direct que les deux précédents, le second, puisque les saisons ne sont pas générées successivement la somme ne donne pas exactement les valeurs annuelles.

Il est à noter que le problème est comme à toutes les formes, si les données sont soumises à des transformations.

1.8. Les modèles autoregressifs a moyennes mobile ARMA [15].

Le groupe de modèles ARMA mis au point par Box et Jenkins (1970), sont des modèles stochastiques appliqués généralement aux séries temporelles stationnaires.

Il en existe trois types : les modèles autoregressifs (AR), les modèles à moyenne mobile (MA), et les modèles mixtes (ARMA).

1.8.1. Formulation mathématique.

Un modèle autoregressif d'ordre p est décrit tel que la valeur ξ_t au temps t est constituée de p valeurs aux temps $t-1, t-2, t-3, \dots, t-p$ en lui ajoutant un nombre aléatoire $t = 0, 1, 2, \dots$ il prend la forme suivante :

$$\begin{aligned}\xi_t &= \phi_{p,1}\xi_{t-1} + \phi_{p,2}\xi_{t-2} + \phi_{p,3}\xi_{t-3} + \dots + \phi_{p,p}\xi_{t-p} + \eta_t \\ &= \sum_{i=1}^p \phi_{p,i}\xi_{t-i} + \eta_t\end{aligned}\quad (\text{Equ. I-1})$$

Où $\phi_{p,i}, i = 1, 2, 3, \dots, p$ sont les paramètres autoregressifs, les propriétés de ξ_t et η_t sont :

$$\begin{aligned}E(\xi_t) &= E(\eta_t) = 0. \\ \text{Var}(\xi_t) &= E(\xi_t^2) = \sigma_\xi^2. \\ \text{Var}(\eta_t) &= E(\eta_t^2) = \sigma_\eta^2.\end{aligned}$$

Les modèles à moyenne mobile MA(q) d'ordre q , prennent la forme suivante :

$$\begin{aligned}\xi_t &= \eta_t - \theta_{q,1}\eta_{t-1} - \theta_{q,2}\eta_{t-2} - \dots - \theta_{q,q}\eta_{t-q} \\ &= \eta_t - \sum_{j=1}^q \theta_{q,j}\eta_{t-j}\end{aligned}\quad (\text{Equ. I-2})$$

$$E(\xi_t) = 0 \quad \text{et} \quad E(\xi_t^2) = 1$$

E : Représente l'espérance, ou bien la moyenne de la composante stochastique.

La variance de la composante aléatoire est donnée par :

$$\sigma_\eta^2 = 1 / (1 + \theta_{q,1}^2 + \theta_{q,2}^2 + \dots + \theta_{q,q}^2)$$

Les modèles du troisième type (ARMA) sont des combinaisons de modèles des deux premiers types, comportant les termes non répétitifs des équations (I-1) et (I-2) comme suit :

$$\begin{aligned}\xi_t &= \phi_{p,1} \xi_{t-1} + \phi_{p,2} \xi_{t-2} + \phi_{p,3} \xi_{t-3} + \dots + \phi_{p,p} \xi_{t-p} + \eta_t - \theta_{q,1} \eta_{t-1} - \theta_{q,2} \eta_{t-2} - \dots - \theta_{q,q} \eta_{t-q} \\ &= \sum_{i=1}^p \phi_{p,i} \xi_{t-i} + \eta_t - \sum_{j=1}^q \theta_{q,j} \eta_{t-j}\end{aligned}$$

$\phi_{p,i}, \theta_{q,j}$: Paramètres du modèle estimés par la méthode des moindres carrés.



Chapitre II

Analyse en composantes principales

C
H
A
P
I
T
R
E
I
I

CHAPITRE II

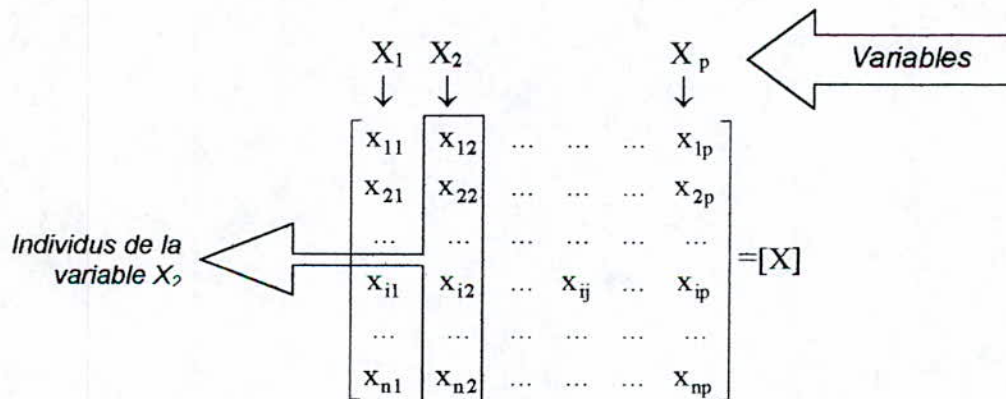
Analyse en composantes principales

Dans ce chapitre, nous nous efforcerons de décrire la théorie de l'analyse en composantes principales, et son mode d'utilisation.

II.1. Définitions.

L'analyse en composante principale est basée sur l'étude de la covariance ou de la corrélation entre variable, ses résultats sont représenter sous forme graphique.

Soit la matrice d'observations suivante :



$X_1, X_2, X_3, \dots, X_p$: Vecteurs colonnes de la matrice d'observations.

Les composantes principales sont des combinaisons linéaires des variables de la matrice d'observations $X = (X_1, X_2, X_3, \dots, X_p)$.

Donc : $C_j = b_{j1} X_1 + b_{j2} X_2 + b_{j3} X_3 + \dots + b_{jp} X_p$, dont la variance est maximale.

Supposant que les variables sont centrées réduites, les C_j le sont également.

$$\begin{array}{cccc}
 C_j & X_1 & X_2 & X_p \\
 \downarrow & \downarrow & \downarrow & \downarrow \\
 \begin{pmatrix} C_{1j} \\ C_{2j} \\ \cdot \\ \cdot \\ C_{nj} \end{pmatrix} & = b_{j1} \begin{pmatrix} x_{11} \\ x_{21} \\ \cdot \\ \cdot \\ x_{n1} \end{pmatrix} & + b_{j2} \begin{pmatrix} x_{12} \\ x_{22} \\ \cdot \\ \cdot \\ x_{n2} \end{pmatrix} & + \dots + b_{jp} \begin{pmatrix} x_{1p} \\ x_{2p} \\ \cdot \\ \cdot \\ x_{np} \end{pmatrix}
 \end{array}$$

$$\text{et } C_j = X \cdot b_j \quad \text{tel que : } b_j = \begin{pmatrix} b_{j1} \\ b_{j2} \\ \cdot \\ \cdot \\ b_{jp} \end{pmatrix}$$

La variance de la $j^{\text{ème}}$ composante principale est :

$$\frac{1}{n} C_j' C_j = \frac{1}{n} (X \cdot b_j)' (X \cdot b_j) \Rightarrow C_j' C_j = (b_j' X') \cdot X \cdot b_j$$

$$C_j' C_j = b_j' X' X b_j = b_j' V b_j$$

tel que V représente la matrice de covariance de X .

Afin de conserver la plus grande partie de l'information apportée sur la variable d'origine, on cherchera $b_{j1}, b_{j2}, b_{j3}, \dots, b_{jp}$ de façon que la variance de C_j calculée sur les données soit maximale.

Le problème est de trouver $b_j = [b_{j1}, b_{j2}, \dots, b_{jp}]$, tel que $b_j' V b_j$ soit $b_j' V b_j$, soit maximum, avec $b_j' b_j = 1$, rappelons que $b_j' V b_j$ est une fonction quadratique.

Il faut annuler la dérivée par rapport à b_j' du Lagrangien L :

$$L = b_j' V b_j - k \cdot (b_j' b_j - 1)$$

avec $b_j' b_j - 1 = y(b) = 0$ (cela veut dire que b est un vecteur normé $|b| = 1$)

$$\frac{dL}{db_j} = 2 \cdot V \cdot b_j - 2 \cdot k \cdot b_j = 0 \quad (\text{Annexe II})$$

$$\Rightarrow V \cdot b_j - k \cdot b_j = 0 \Rightarrow (V - kI) \cdot b_j = 0$$

Donc les vecteurs b_j recherchés sont les vecteurs propres de la matrice de covariance $[V]$ de X , et k multiplicateur de Lagrange, c'est la valeur propre.

Résultat :

On a pu démontrer que b_j sont les vecteurs propres de la matrice des covariances de X (matrice des observations initiales), et λ_p sont les valeurs propres associées, tel que $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$.

D'où l'écriture matricielle généralisée :

$$C_j = X.B_j$$

II.2. Interprétations.

Pour chaque observation, on effectue la transformation $C_1 = b_1'.x$. Si on projette chaque observation sur le premier vecteur propre b_1 , on obtient N points dont la variance est maximale (λ_1 variance de C_1), et ceci quelle que soit la forme du nuage des N points dans l'espace à n dimensions.

On peut dire que la première composante principale rend compte d'une proportion $\lambda_1 / \text{tr}(V)$ de la variance totale (ce pourcentage dans certains cas dépasse 90%), cela veut dire qu'une seule composante permet d'expliquer les variables d'origines avec une très forte proportion, mais ceci ne suffira pas, c'est pour cela que dans la plupart des cas, on fait la projection sur le plan des deux premières composantes principales pour bien expliquer la variabilité des observations (variables d'origine) ;

(C) est alors représentée par un point de coordonnées

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} b_1' \\ b_2' \end{pmatrix} . x$$

(2x1) (2xn) (nx1)

L'expérience montre que très souvent la troisième composante C_3 , même si le pourcentage de variation expliqué par λ_3 est faible, complète très utilement la description, c'est pour cela qu'on est ramené à réaliser les projections (C_1, C_3) et où (C_1, C_3) en même temps que (C_1, C_2) .

En général, les interprétations varient avec le problème étudié, on ne peut généraliser une règle donnée sur les interprétations.

II.3. Calcul des composantes principales.

Le calcul des composantes principales se ramenant à celui des valeurs et vecteurs propres de la matrice des covariances.

II.3.1. Calcul de la matrice des covariances.

La matrice de covariance de X est :

$$V = \frac{1}{n} (X'.X - \frac{1}{n} S.S') = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1p} \\ V_{21} & V_{22} & \dots & V_{2p} \\ \dots & \dots & \dots & \dots \\ V_{p1} & V_{p2} & \dots & V_{pp} \end{bmatrix} \quad (\text{Annexe III}).$$

II.3.2. Calculs des valeurs et vecteurs propres.

Pour les matrices d'ordre 2 et 3, les valeurs propres peuvent être calculées directement à partir des équations caractéristiques qui sont du degré 2 et 3 en λ

$$\lambda^2 - (v_{11} + v_{22}) \lambda + (v_{11} v_{22} - v_{12} v_{21}) = 0$$

$$\lambda^3 + (v_{11} + v_{22} + v_{33}) \lambda^2 - (V_{11} + V_{22} + V_{33}) \lambda + |V| = 0$$

Mais à partir du cinquième degré, les racines ne peuvent être calculées que par des méthodes itératives, il en existe plusieurs, dans notre approche, on a choisit une méthode dite : *itérative de la puissance*.

II.3.2.a. Théorie de la méthode itérative de la puissance.

II.3.2.a.i. Calcul de la première valeur propre et le premier vecteur propre.

Supposons que la matrice des covariances est de dimension 3, pour déterminer son premier vecteur propre et sa première valeur propre, partant d'un vecteur $x^0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, en

effectuant le produit $V.x^0$ on aura le vecteur x^1 , en divisant ce vecteur par son plus grand terme (pour le normer), et en multipliant encore la matrice V par ce dernier, on obtient le vecteur x^2 , en le normant, et en appliquant le processus suivant :

$$\begin{aligned} V.x^0 &= x^1 \\ V.x^1 &= x^2 \\ \dots & \\ \dots & \\ V.x^{k-1} &= x^k \\ V.x^k &= x^{k+1} \end{aligned}$$

On arrête l'opération lorsque le vecteur produit sera stabilisé ça signifie que $x^k = x^{k+1}$, alors ce dernier sera le premier vecteur propre de la matrice de covariance, et la première valeur propre associée à ce vecteur, est le plus grand terme du vecteur correspondant non normé.

II.3.2.a.ii. Calcul de λ_2 et b_2 .

On peut calculer λ_2 et b_2 par la méthode de déflation ; elle est basée sur la propriété suivante :

Si b_1 et b_1' sont tels que $(b_1.b_1' = 1)$, la matrice $V^{(1)} = V - \lambda_1.b_1.b_1'$ a les mêmes valeurs propres et vecteurs propres que V à l'exception de λ_1 qui est remplacée par une valeur nulle, les lignes où colonnes de $V - \lambda_1.b_1.b_1'$ définissent un sous espace de dimension $(n-1)$ déduit de l'espace de dimension n défini par V .

Si on a une matrice de covariance d'ordre 3, ses vecteurs propres définissent un espace de 3 dimensions.

Le sous espace $V - \lambda_1.b_1.b_1'$ correspond au plan défini par b_2 et b_3 .

Le produit $(V - \lambda_1.b_1.b_1').x^0$, est un vecteur x^1 du plan $b_2.b_3$ la suite de l'opération c'est à dire :

$$\begin{aligned} (V - \lambda_1.b_1.b_1') x^0 &= x^1 \\ (V - \lambda_1.b_1.b_1') x^1 &= x^2 \\ \dots & \\ \dots & \\ (V - \lambda_1.b_1.b_1') x^{k-1} &= x^k \\ (V - \lambda_1.b_1.b_1') x^k &= x^{k+1} \end{aligned}$$

Jusqu'à ce que la stabilisation $(x^k = x^{k+1})$, tende vers le deuxième vecteur propre et la deuxième valeur propre.

II.3.2.a.iii. Calcul de λ_3 et b_3 .

De même on obtient λ_3 et b_3 après une seconde déflation tel que

$$V^{(2)} = V - \lambda_1.b_1.b_1' - \lambda_2.b_2.b_2' \quad \text{avec } b_2.b_2' = 1$$

Cette nouvelle matrice $V^{(2)}$ a les mêmes valeurs et vecteurs propres que V à l'exception λ_1 et λ_2 remplacées par des valeurs nulles, elle définit un sous espace d'ordre $(n-2)$ qui ne contient pas les vecteurs b_1 et b_2 .

Le produit

$$\begin{aligned} (V - \lambda_1.b_1.b_1' - \lambda_2.b_2.b_2') x^0 &= x^1 \\ (V - \lambda_1.b_1.b_1' - \lambda_2.b_2.b_2') x^1 &= x^2 \\ \dots & \\ \dots & \\ \dots & \\ (V - \lambda_1.b_1.b_1' - \lambda_2.b_2.b_2') x^{k-1} &= x^k \\ (V - \lambda_1.b_1.b_1' - \lambda_2.b_2.b_2') x^k &= x^{k+1} \end{aligned}$$

Jusqu'à la stabilisation, c'est à dire $x^k = x^{k+1}$ tendra vers le troisième vecteur propre et la troisième valeur propre, et la plus grande valeur du vecteur carré correspondant non normé.

Remarque : on peut déterminer λ_n et b_n après $n-1$ déflations, et ça donne une matrice diagonale des valeurs propres, et une matrice des vecteurs propres défini par les colonnes des vecteurs propres calculés à chaque étape.

II.3.3. Composantes principales.

Après avoir déterminé la matrice des vecteurs propres, on peut déduire la matrice des composantes principales comme suit :

$$[C]_{n,p} = [X]_{n,p} \cdot [B]_{p,p}$$

tel que :

[X] : Matrice d'observations.

[B] : Matrice des vecteurs propres de la matrice des covariances.

[C] : Matrice des composantes principales.

II.4. Représentations graphiques.

Les résultats d'une A.C.P. sont représentés dans un cercle unité s'appelle cercle de corrélations, l'axe (1) représente la première composante principale (C_1) et l'axe (2) représente la deuxième composante principale (C_2).

Pour cela on calcul la matrice des corrélations entre les composantes principales et les variables de la série d'observation X , dont la corrélation est donnée par la formulation suivante :

$$R = \frac{\text{Covariance}(C_j, X_i)}{\sqrt{\text{var}(C_j)}\sqrt{\text{var}(X_i)}} = D_\lambda^{1/2} \cdot B' \cdot D_V^{-1/2} \quad (\text{annexe III})$$

Les variables sont représentées sur un cercle par des points dont l'abscisse est la corrélation entre la première composante principale et la première variable d'observation, et l'ordonnée est la corrélation entre la première composante principale et la deuxième variables d'observation.

Remarque :

La distance d'une variable à l'origine des axes est égale à sa norme (son écart-type). Le cosinus de l'angle formé par deux variables est égale a leurs coefficients de corrélations.

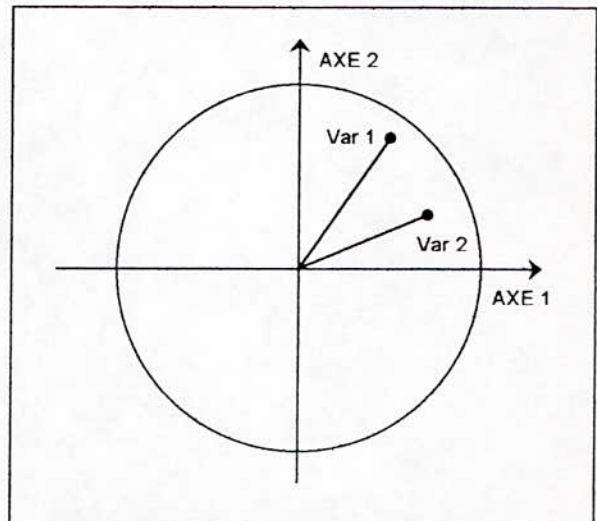


Fig II-1 : Cercle de corrélation

II.5. Diagramme de calcul.

Le diagramme de la figure II-2 reflète un aperçu sur la technique et la méthodologie de l'analyse en composantes principales.

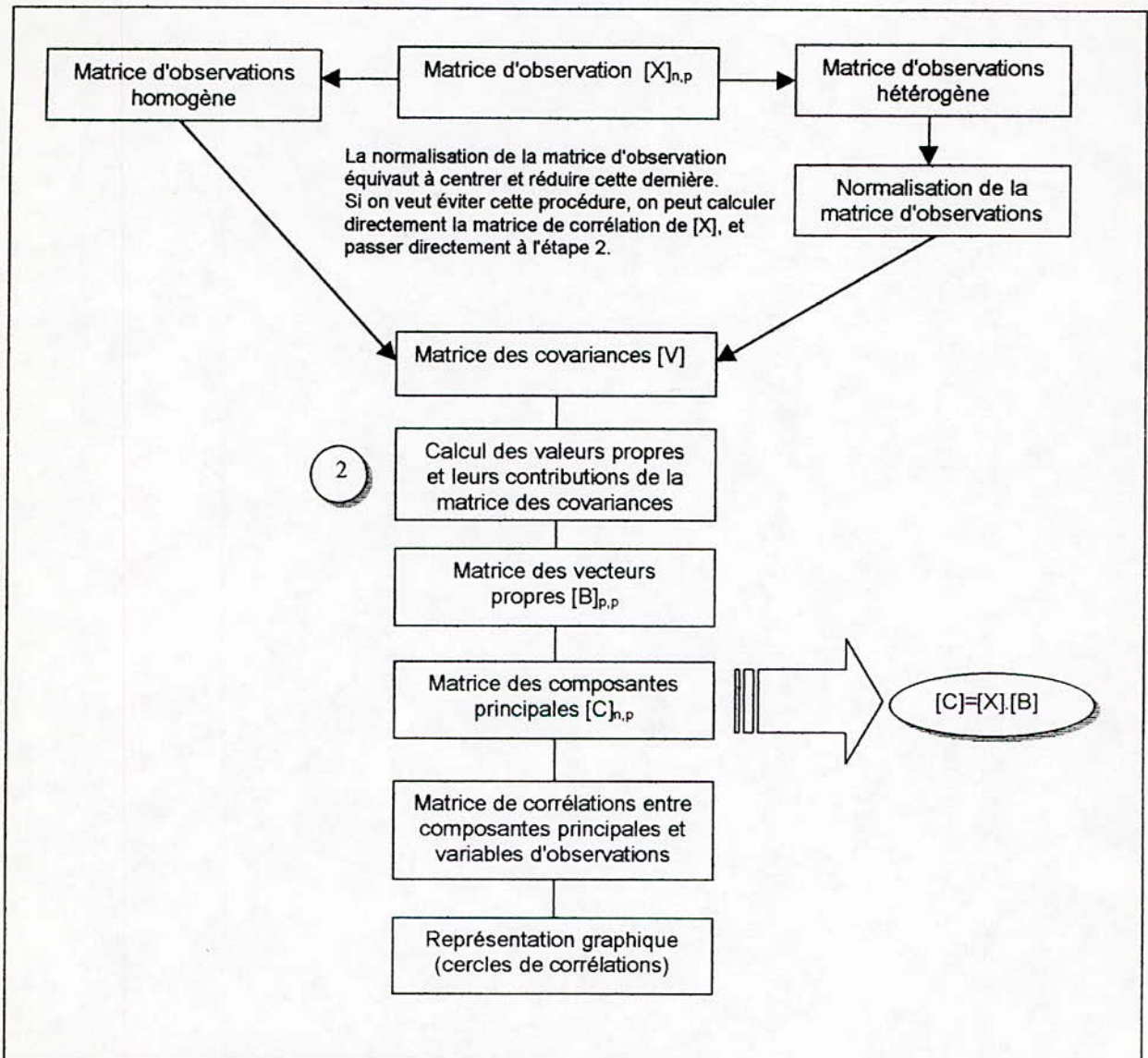


Fig. II-2 : Diagramme de calcul des composantes principales

II.6. Exemple d'application.

On se propose d'étudier et d'analyser une classe d'étudiants de 10 individus dans 5 matières différentes qui représentent les variables de notre série d'observations par le biais de l'A.C.P.

	Maths	Physique	Chimie	Philosophie	Littérature
Individu 1	15.00	15.50	15.00	10.50	10.25
Individu 2	16.00	15.75	16.50	09.75	10.00
Individu 3	14.50	14.00	15.00	10.00	10.00
Individu 4	16.00	15.00	16.00	11.25	11.00
Individu 5	16.75	16.50	15.50	10.00	10.50
Individu 6	16.50	16.00	16.00	09.00	09.25
Individu 7	14.75	15.00	14.50	10.25	10.00
Individu 8	14.00	14.50	15.00	09.50	09.50
Individu 9	15.00	15.00	14.75	10.00	09.50
Individu 10	16.50	16.75	16.00	10.00	10.00

Résultats de l'analyse :

Moyenne	Ecart-type
15.50	00.96
15.40	00.87
15.42	00.67
10.25	00.59
10.00	00.51

Matrice des corrélations

1.00	0.86	0.76	-0.04	0.28
0.86	1.00	0.56	-0.17	0.09
0.76	0.56	1.00	-0.10	0.20
-0.04	-0.17	-0.10	1.00	0.20
0.28	0.09	0.20	0.86	1.00

Matrice des covariances

0.93	0.72	0.49	-0.02	0.14
0.72	0.75	0.32	-0.09	0.04
0.49	0.32	0.44	-0.04	0.07
-0.02	-0.09	0.04	0.35	0.26
0.14	0.04	0.07	0.26	0.26

Matrice des covariances

0.70	0.08	0.03	0.70	-0.11
0.59	-0.15	-0.57	-0.54	0.03
0.39	0.02	0.81	-0.44	-0.08
-0.04	0.75	-0.14	-0.14	-0.62
0.09	0.63	-0.01	-0.04	0.77

Valeurs propres Contribution Contribution cumulée

00001.83	66.71	066.71 (1)
00000.59	21.44	088.15 (2)
00000.24	08.67	096.82 (3)
00000.06	02.32	099.14 (4)
00000.02	00.86	100.00 (5)

On remarque que les deux premiers axes arrivent à expliquer plus de 85% de la variance totale.

Matrice des composantes principales

025.99	013.57	002.14	-006.33	-001.10
027.43	012.92	003.34	-006.31	-001.05
024.75	013.23	003.05	-005.78	-000.97
026.82	014.79	003.15	-005.93	-001.20
028.04	013.35	002.08	-005.80	-000.81
027.70	011.87	002.91	-005.74	-001.17
025.31	013.27	002.05	-005.97	-001.09
024.67	012.42	002.82	-006.32	-000.97
025.55	012.79	002.30	-005.85	-001.36
028.17	012.99	002.34	-006.31	-001.20

Corrélation entre CP & X

0.98	0.06	0.02	0.18	-0.02
0.92	-0.13	-0.32	-0.16	0.00
0.79	0.03	0.59	-0.17	-0.02
-0.09	0.97	-0.11	-0.06	-0.16
0.23	0.94	-0.01	-0.02	0.23

On remarque dans la figure II-3, que toutes les variables sont bien représentées puisqu'elles sont proches du cercle de corrélation.

Les variables Maths, Physique et chimie sont orienté vers l'axe 1, ceci s'explique par la faible variation de la moyenne de chaque matière, il en est de même pour les deux autres variables orientées vers l'axe 2.

On notera que les variables proches dans le plan sont proche dans la réalité, c'est-à-dire qui sont bien corrélées.

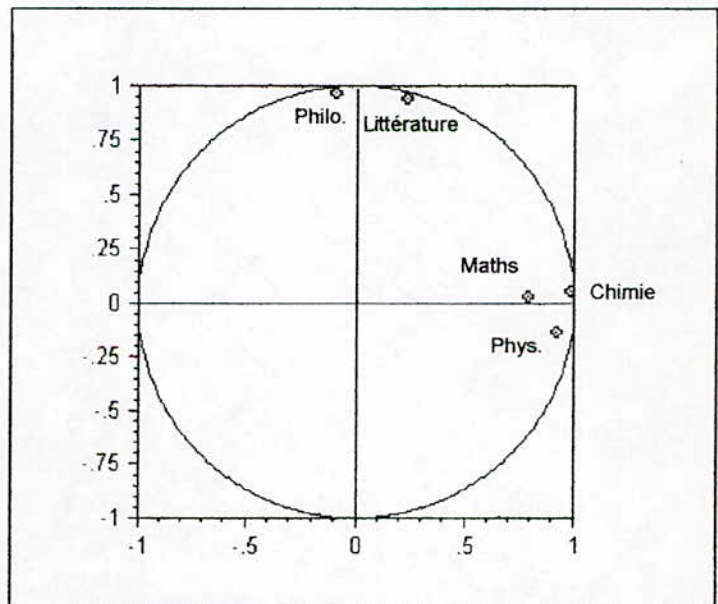


Fig. II-3 : Analyse des variables



Chapitre III

Le modèle GÉSTOP de simulation

C
H
A
P
I
T
R
E
I
I
I

CHAPITRE III

Le modèle GESTOP de simulation

Le modèle GESTOP en deux versions, brute et normé, basé sur la technique de l'analyse en composante principale sera le cœur de ce chapitre.

III.1. Définition.

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 calculateurs numériques ou les ordinateurs.

La simulation permet dans certain cas d'éviter les déboires de l'expérimentation.

La simulation s'appuie sur la construction d'un modèle dont toutes ces caractéristiques sont connues, qui vise à identifier le comportement d'un procédé ou d'un système et interpréter son mécanisme, ce qui aide à entreprendre de meilleures décisions. Dans tous cela, on peut dire que la simulation est un très bon planificateur.

En hydrologie, la simulation nous permet de créer à partir des séries d'observations, d'autres séries qui ont les mêmes caractéristiques statistiques, afin de pouvoir faire des études statistiques et anticiper la gestion de la ressource en eau.

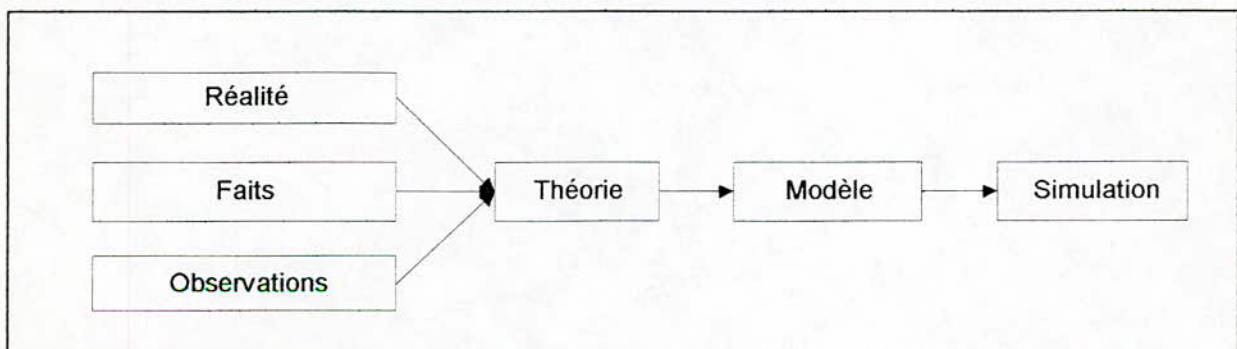


Figure III-1 : Mécanisme de simulation.

III.2. Techniques de simulation.

La simulation est basée essentiellement sur le concept du hasard, c'est-à-dire le tirage de nombres aléatoires, qui n'a de sens scientifique que si l'on se réfère au mécanisme de génération qui diffère d'une méthode de simulation à une autre.

III.2.1. Simulation par la fonction de répartition ^[6].

Si l'on considère la variable aléatoire x de fonction de répartition $F(x)$, la nouvelle variable $y=F(x)$ est uniformément répartie sur le segment $[0,1]$.

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

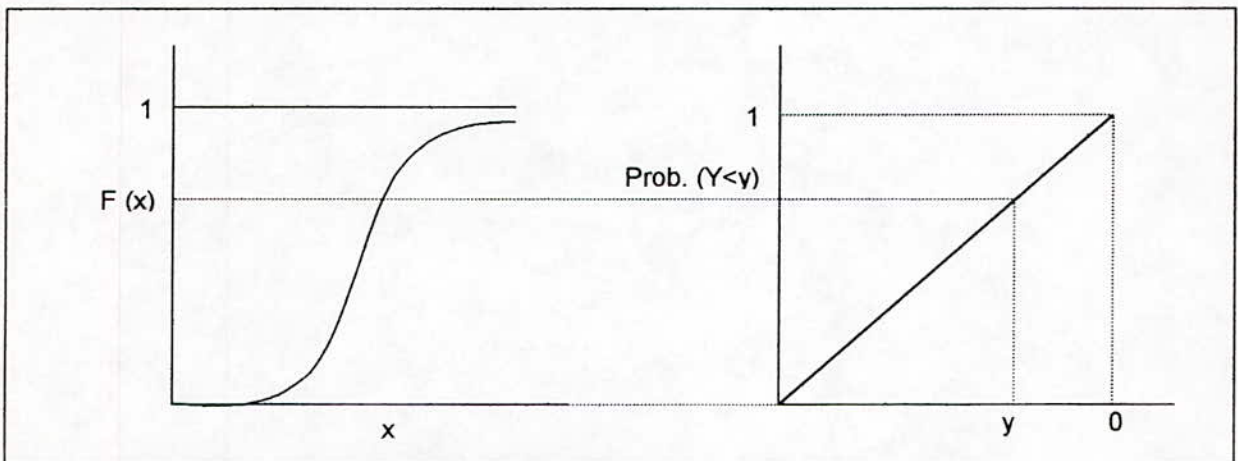


Figure III-2 : Passage d'une fonction de probabilité F à une loi uniforme.

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

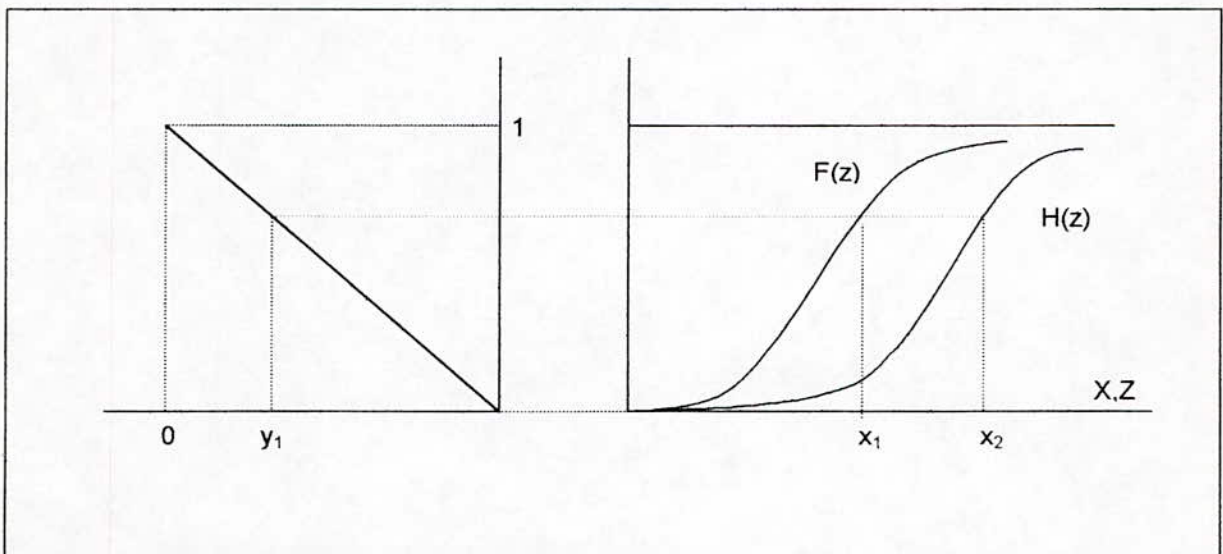


Figure III-3 : Tirage aléatoire à partir de la fonction de répartition.

La figure (III-3) représente le tirage au hasard :

- 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 $H(z)$.

On met la loi de probabilité F et/ou H dans l'urne des nombres aléatoires et on tire au hasard un nombre y_1 , qui après transformation fournit la valeur x_1 et/ou z_1 .

III.2.2. Simulation par les chaînes de Markov ^[6].

Un processus stochastique est un processus dont la valeur à l'instant t dépend des valeurs antérieures, celui-ci est dit Markovien si la valeur à l'instant t ne dépend en fait que de la valeur la plus récente, c'est-à-dire :

$$P\left\{X(t) = \frac{x}{X(t_1)} = x_1, X(t_2) = x_2, \dots, X(t_n) = x_n\right\} = x_n = P\left\{X(t) = \frac{x}{X(t_n)} = x_n\right\}$$

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.
- les modèles multiclassés dont la série est subdivisée en différentes classes.

III.2.2.a. Construction de la chaîne de Markov.

On définit en premier lieu l'état 0 et l'état 1, sur lesquels on se base pour transformer la série donnée en système binaire (00111010...). Pour obtenir une telle série, il suffit d'effectuer un comptage des éléments appartenant respectivement à l'état 0 et 1. On calcule ensuite les probabilités inconditionnelles P_0 et P_1 .

$$P_0 = \frac{\text{Nombre d'éléments appartenant à l'état 0}}{\text{Nombre total d'éléments}}$$

$$P_1 = \frac{\text{Nombre d'éléments appartenant à l'état 1}}{\text{Nombre total d'éléments}}$$

III.2.2.b. Génération par les séries de Markov.

Après avoir construit la matrice de passage d'ordre zéro (0), on détermine les fonctions de répartition des valeurs appartenant aux états 0 et 1.

La simulation se fait en générant uniformément un nombre aléatoire y dans l'intervalle $[0,1]$, qu'on compare aux probabilités inconditionnelles.

Sachant que $P_0 + P_1 = 1$, alors $P_1 = 1 - P_0$.

Donc si y est inférieur à 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.

III.2.3. Simulation par les lois d'ajustement.

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

- sélection d'un échantillon de données satisfaisant certains critères statistiques tel que l'indépendance, l'homogénéité et la stationnarité.
- choisir la loi d'ajustement la plus adéquate pour l'échantillon considéré.

III.2.3.a. Tests statistiques [7].

III.2.3.a.i. Test d'indépendance.

Wald et Wolfowitz, ont émis un critère d'indépendance signifiant qu'aucune observation appartenant à l'échantillon ne devrait influencer l'observation suivante; en d'autres termes, les observations ne devraient pas être corrélées entre elles.

Pour un échantillon de taille n (x_1, x_2, \dots, x_n) le test de Wald-Wolfowitz considère la valeur statistique :

$$R = \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + x_1 \cdot x_n.$$

Au cas où les éléments de l'échantillon seraient indépendants, R suit une distribution normale, de moyenne et de variance respectivement :

$$R = \frac{S_1^2 - S_4}{n - 1}$$

$$\text{Var}(R) = \frac{S_2^2 - S_4}{n-1} - \bar{R}^2 + \frac{S_1^4 - 4.S_1^2.S_2 + 4.S_1.S_3 + S_2^2 - 2.S_4}{(n-1).(n-2)}$$

Avec :

$$S_r = n.m'_r$$

m'_r le moment d'ordre r de l'échantillon.

La valeur $U = \frac{R - \bar{R}}{\sqrt{\text{Var}(R)}}$ suit une distribution normale standardisée (moyenne 0 et variance 1) et peut être utilisée pour tester, à un niveau de signification, l'hypothèse d'indépendance en comparant $|U|$ avec la valeur normale $U^{\alpha/2}$ correspondant à la probabilité de dépassement de $\alpha/2$.

III.2.3.a.ii. L'homogénéité.

Les méthodes statistiques d'analyse des séries pluviométriques exigent de celles-ci une homogénéité de leurs composantes car nous ne pouvons pas faire une analyse statistique d'un échantillon composé de n observations d'une variable hydrologique que si certaines de ces n observations ne présentent pas d'erreurs systématiques rendant l'échantillon hétérogène.

Soit deux échantillons de taille p et q ($p \leq q$) ; leur combinaison donne un échantillon de taille $n=p+q$, rangés par ordre croissant.

Le test de Mann-Witney (1947) considère les quantités :

$$V = R - \frac{P.(P+1)}{2}$$

$$W = P.Q - V$$

Où R est la somme des rangs des éléments du premier échantillon (taille p) dans la série considérée.

V et W sont calculés à partir de R, P et Q.

V représente le nombre de fois où un élément de l'échantillon 1 suit dans l'ordre un élément de l'échantillon 2.

La valeur statistique de Mann-Witney est U et elle est définie par la plus petite des deux valeurs de V et W.

Quand $n > 20$ et $p, q > 3$, et quand les deux échantillons ne proviennent pas d'une même population, U est approximativement normalement distribuée avec une moyenne:

$$U = \frac{P.Q}{2}$$

Et une variance :

$$\text{Var}(U) = \left[\frac{P.Q}{n.(n-1)} \right] \left[\frac{n^3 - n}{12 - \sum T} \right]$$

Sachant que :

$$T = \frac{(J^3 - J)}{12}$$

Où J est le nombre d'observations associées à une classe donnée. Pour un test à un niveau de signification α , la quantité:

$$|U| = |(U - \bar{U})[\text{Var}(U)]^{1/2}|$$

est comparée avec la variable normale standardisée correspondant à une probabilité de dépassement de $\alpha/2$.

III.2.3.a.iii. La stationnarité.

Ce test a pour but de mettre en évidence les points singuliers de la série. Pour appliquer ce test, nous devons calculer tout d'abord :

$$X_H = e^{(\bar{X} + K_N.S)}$$

$$X_L = e^{(\bar{X} - K_N.S)}$$

X et S sont respectivement la moyenne et la variance des logarithmes naturels des éléments constituant l'échantillon.

K_N est la valeur statistique du test de Grubbs et Beck, tabulée, pour différentes tailles d'échantillon de niveau de signification, par Grubbs et Beck en 1972.

Au niveau de 10%, l'approximation polynomiale suivante a été proposée par Pilon (1985).

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

Où n est la taille de l'échantillon.

En appliquant le test de Grubbs et Beck, toutes les valeurs de l'échantillon supérieures à X_H ou inférieures à X_L sont considérées comme points singuliers qu'il faut comparer avec les autres valeurs des autres variables.

III.2.3.b. Simulation par la loi normale ^[6].

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

Un tirage aléatoire dans une loi Laplace-Gauss centrée réduite peut être effectué selon la méthode Box-Muller (1958), on tire deux nombres U_1 et U_2 grâce à la fonction Random qui effectue un tirage aléatoire dans la loi uniforme su $[0, 1]$.

On calcul :

$$\begin{cases} Z_1 = (-2.\text{Ln}U_1)^{\frac{1}{2}}.\text{Cos}(2.\pi.U_2) \\ Z_2 = (-2.\text{Ln}U_1)^{\frac{1}{2}}.\text{Sin}(2.\pi.U_2) \end{cases}$$

Ainsi définies, Z_1 et Z_2 sont des variables aléatoires de Laplace-Gauss centrées réduites et indépendantes.

Si la loi d'ajustement est une normale (m, σ) , on peut toujours se ramener à une normale de moyenne nulle et d'écart type unité $N(0, 1)$, et faire une transformation inverse par la suite.

$$X = m + \sigma.Z$$

Z : Variable de Gauss.

X : Variable initiale.

m : Moyenne de la variable X .

σ : Ecart type de la variable X .

III.2.3.c. Simulation par la loi G.E.V ^[12].

La théorie des valeurs extrêmes qui date depuis 1927, revient à Frechet (1927), Fisher et Tipeh (1928). Il existe trois types, le type I développé et appliqué par Gumbel (1958), c'est une distribution particulière des valeurs extrêmes possédant deux paramètres, ses applications sont limitées, ce qui explique la considération d'autres types : type II et type III de distribution des valeurs extrêmes, possédant trois (03) paramètres.

Le type II, de distributions asymptotiques de valeurs extrêmes est donnée par :

$$\lim_{x \rightarrow \infty} [\Pr\{W_{(n)} \leq x\}] = \exp \left[- \left\{ \frac{u-x}{x-\varepsilon} \right\}^\alpha \right]$$

Pr : Probabilité au non dépassement.

$W_{(n)}$: Variable aléatoire qui représente la valeur maximale.

x : Valeur particulière ; $x \geq \varepsilon$ et $\mu \geq \varepsilon$.

Elle est connue généralement sous le nom de distribution de Frechet

$$x = \varepsilon + (u - \varepsilon) \cdot \exp \left(\frac{y}{\alpha} \right) \quad (\text{Equ. III-1})$$

y : Variable réduite de Gumbel type I de distribution des valeurs extrêmes.

Pour une forme exponentielle positive ($\alpha > 0$), x augmente lorsque y augmente. Alors la distribution est représentée par une courbe concave vers le haut sur un papier probabilité Gumbel.

Le type III, de distribution des valeurs extrêmes est donnée par :

$$\lim_{n \rightarrow \infty} [\Pr\{W_n \leq x\}] = \exp \left[- \left\{ \frac{w-x}{w-u} \right\}^2 \right]$$

$$x \leq \omega ; u \leq \omega$$

la relation entre x et y est donné par :

$$x = \omega - (\omega - u) \cdot \exp \left(- \frac{y}{\alpha} \right) \quad (\text{Equ. III-2})$$

C'est un type exponentiel négatif, et il est représenté par une courbe concave vers le bas sur un papier de probabilité Gumbel.

La formule générale de distribution des valeurs extrêmes correspondant à l'équation (III-1) et l'équation (III-2), a été suggéré par Jenkinson (1967) :

$$x = u + \frac{1}{\alpha} \cdot \frac{1 - \exp(-K.y)}{K} \quad (\text{Equ. III-3})$$

Cette relation représente la relation entre x et y des trois (03) types de distribution des valeurs extrêmes, avec u , α et k , paramètres de localisation, d'étendue, et de forme respectivement.

Elle est appelée *general extreme values (G.E.V.) distribution*.

Pour ($k = 0$), la formule mène à la relation linéaire de x en fonction de y , où elle caractérise le type I de distribution des valeurs extrêmes donné par l'équation $y = \alpha.(x - u)$, c'est la variable réduite de Gumbel.

Le type II de distribution des valeurs extrêmes est appliqué lorsque $k < 0$, et si $k > 0$, ça signifie le type III de distribution. Ces deux types sont représentés par l'équation (III-1) et l'équation (III-2) respectivement.

Afin d'évaluer la variable de l'année T par la méthode de Jenkinson, l'équation III-3 est écrite comme suit :

$$x(T) = u + \frac{z(T)}{\alpha} \quad (\text{Equ. III-4})$$

$z(T)$: valeur de la variable standardisée z est donnée par :

$$z(T) = \frac{1 - \exp(-ky)}{k} \quad (\text{Equ. III-5})$$

tel que y et T sont reliés par l'équation qui donne la relation entre la variable réduite de Gumbel y et la période de retour comme suit :

$$y = -\text{Ln}(\text{Ln}(T) - \text{Ln}(T-1))$$

la méthode de sextiles appliquée par Jenkinson (1969) donne l'estimation approximative des paramètres k , u et α .

La population infinie des valeurs de Z , la distribution peut être donnée par une fonction de k et y à travers l'équation (III-5), est considéré pour arranger dans un ordre croissant et en divisant en six (06) groupes de tailles égale. On note la moyenne des variables dans le sextile par $\mu_{Z,i}$, où $i=1$ représente le sextile large, et on note la moyenne et l'écart-type de ces six (06) valeurs moyennes par μ_z et σ_z respectivement.

De la même façon, on divise les valeurs correspondantes de X en six (06) groupes. Ceci est donné dans une série, en ordonnant les valeurs de l'échantillon et en les divisant en six (06) groupes égaux ou presque égaux.

$\mu_{x,i}$ indique la moyenne de la population des variables dans le $i^{\text{ème}}$ sextile groupe.

La moyenne et l'écart-type de ces six (06) valeurs sont :

$$\mu = \frac{1}{6} \sum_{i=1}^6 \mu_{x,i}$$

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

De l'équation (III-4), et en prenant les espérances et en égalisant les variances :

$$\mu = u + \frac{u_z}{\alpha} \quad (\text{Equ. III-6})$$

$$\sigma = \frac{\sigma_z}{\alpha} \quad (\text{Equ. III-7})$$

La relation entre les paramètres de forme k (qui est commun aux deux populations (X et Z), μ_z , σ_z et le rapport de forme $r = (\mu_{z,5} - \mu_{z,6}) / (\mu_{z,1} - \mu_{z,2})$ sont donnés dans le tableau (III-1).

$$\mu_z = \frac{1 - \Gamma(1+k)}{k}$$

et pour calculer σ_z et r, on a besoin d'utiliser la fonction inverse de Gamma.

Le rapport de forme r, qui est commun pour les deux populations Z et X, de l'équation (III.4) est estimé par :

$$\hat{r} = \frac{\hat{\mu}_{x,5} - \hat{\mu}_{x,6}}{\hat{\mu}_{x,1} - \hat{\mu}_{x,2}}$$

Par interpolation de \hat{k} , \hat{u}_z et $\hat{\sigma}_z$ sont données dans le tableau (III-1) après utilisation de l'équation (III-6) et (III-7) pour calculer \hat{u} et $\hat{\alpha}$.

Tableau (III-1) : Rapport entre les différents paramètres de la G.E.V.

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	0.58	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

Afin de faciliter la programmation, on a déterminé les différentes équations reliant les paramètres représentés dans le tableau ci-contre.

[1] $k = 0.29 + 0.32.Ln(r)$
 [2] $\mu_z = 0.11 + \frac{1}{0.45} \cdot \sqrt{r}$
 [3] $\sigma_z = 0.81 + \frac{0.16}{r}$

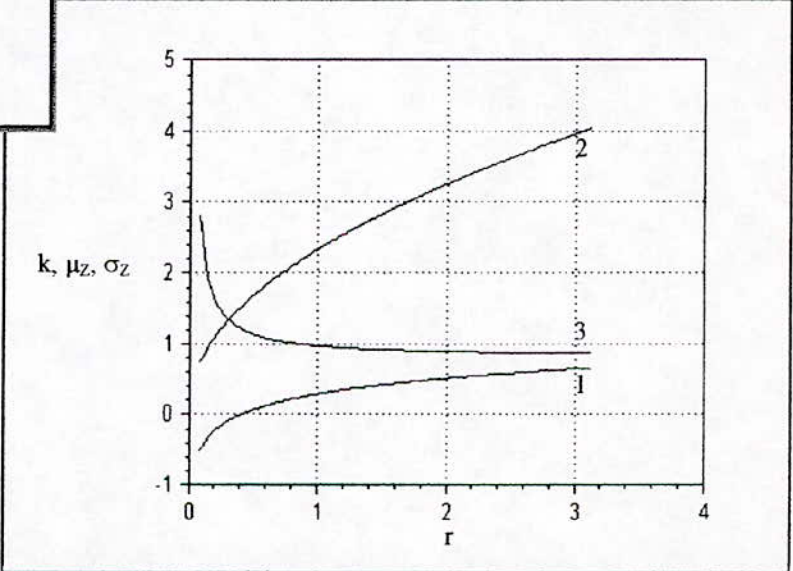


Fig. III-4 : courbes reliant les différents paramètres de la G.E.V.

III.3. Modèle GESTOP brut.

Dans le cas d'une A.C.P. brute, les composantes principales sont tirées par le produit de la matrice d'observation brute [X] et la matrice des vecteurs propres [B] de la matrice de covariance de [X] :

$$[C]_{n \times p} = [X]_{n \times p} \cdot [B]_{p \times p} \quad (\text{Equ. III-8})$$

[C] : Matrice des composantes principales.

[X] : Matrice d'observations brutes.

[B] : Matrice des vecteurs propres de la matrice des covariances de [X].

Soit la matrice des composantes principales centrée réduite [C'] tel que :

$$[C'] = ([C] - [C^*]) \cdot [\sigma_c]^{-1} \quad (\text{Equ. III-9})$$

[C'] : Matrice des composantes principales centrée réduite.

[C*] : Matrice des moyennes des composantes principales (C.P.), tel que chaque moyenne d'un vecteur est disposée en colonne.

[\sigma_c] : Matrice diagonale des écart-types des C.P.

En introduisant l'équation (III-8) dans l'équation (III-9), on obtient :

$$[C'] = ([X] \cdot [B] - [C^*]) \cdot [\sigma_c]^{-1} \quad (\text{Equ. III-10})$$

A partir de cette dernière, on tire la matrice d'observations :

$$[X] \cdot [B] = [C'] \cdot [\sigma_c] + [C^*]$$

$$[X] = [C'] \cdot [\sigma_c] \cdot [B]^t + [C^*] \cdot [B]^t$$

Sachant que :

$$[\sigma_c] \cdot [B]^t = [\text{COV}] = \text{COV}(c'_i, x_j) \quad , \quad [i = \overline{1, p}; j = \overline{1, p}]$$

$$[C^*] \cdot [B]^t = [X^*]$$

L'expression finale de la matrice d'observations devient :

$$[X] = [C'] \cdot [\text{COV}] + [X^*]$$

$$x_{ij} = \sum_{k=1}^p c'_{ik} \cdot \text{cov}_{kj} + \bar{x}_j = \sum_{k=1}^p c'_{ik} \cdot \text{cov}(c'_k, x_j) + \bar{x}_j$$

Sachant qu'on ne considère que les q premiers CP, et qu'on ne perd pas de vue la variance non expliquée prise en compte par le terme résiduel ε_{ij} , l'expression finale du modèle prend la forme suivante.

$$x_{ij} = \sum_{k=1}^q c'_{ik} \cdot \text{cov}(c'_k, x_j) + \bar{x}_j + \varepsilon_{ij}$$

$$[X] = [C'] \cdot [COV] + [X^*] + [E]$$

- [X] : Matrice des variables simulées.
- [C'] : Matrice des C.P. centrées réduites.
- [COV]: Matrice des covariances entre c' et x .
- [X*] : Matrice des moyennes des x .
- [E] : Matrice des résidus.

III.4. Modèle GESTOP normé.

Dans le cas d'une A.C.P. normée, les composantes principales sont le résultat du produit de la matrice d'observations $[X]$ centrée réduite par la matrice des vecteurs propres de la matrice des covariances de $[X]$.

$$[C]_{n \times p} = [X']_{n \times p} \cdot [B]_{p \times p} \quad (\text{Equ. III-11})$$

- $[C]$: Matrice des composantes principales.
 $[X']$: Matrice d'observations centrée réduite.
 $[B]$: Matrice des vecteurs propres.

Soit la matrice des composantes principales réduite $[C']$ tel que :

$$[C'] = [C] \cdot [\sigma_C]^{-1} \quad (\text{Equ. III-12})$$

- $[C']$: Matrice des composantes principales réduite.
 $[\sigma_C]$: Matrice diagonale des écart-types des composantes principales.

En introduisant l'équation (III-11) dans l'équation (III-12), on obtient :

$$[C'] = [X'] \cdot [B] \cdot [\sigma_C]^{-1} \quad (\text{Equ. III-13})$$

A partir de cette dernière, on tire la matrice d'observations centrée réduite :

$$[X'] = [C'] \cdot [\sigma_C] \cdot [B]^t$$

Sachant que :

$$[\sigma_C] \cdot [A]^t = \text{COR}(c'_i, x_j) = [\text{COR}], \quad [i = \overline{1, p}; j = \overline{1, p}]$$

on a :

$$[X'] = [C'] \cdot [\text{COR}]$$

$$x'_{ij} = \sum_{k=1}^p c'_{ik} \cdot \text{COR}_{kj} = \sum_{k=1}^p c'_{ik} \cdot \text{COR}(c'_k, x_j)$$

$$\frac{x_{ij} - \bar{x}_j}{\sigma_{xj}} = \sum_{k=1}^p c'_{ik} \cdot \text{COR}_{kj} \Rightarrow x_{ij} = \left(\sum_{k=1}^p c'_{ik} \cdot \text{COR}_{kj} \right) \cdot \sigma_{xj} + \bar{x}_j$$

$$[X] = [C'] \cdot [\text{COR}] \cdot [\sigma_{xj}] + [X^*]$$

Sachant que :

$$[\text{COR}].[\sigma_{xj}] = [\beta]$$

L'expression finale de la matrice d'observations devient :

$$[X] = [C'].[\beta] + [X^*]$$

$$x_{ij} = \sum_{k=1}^q c'_{ik} \beta_{kj} + \bar{x}_j$$

Le terme résiduel ε_{ij} prend en compte la variance non expliquée, en tenant compte que des q premières composantes principales, l'expression finale du modèle s'écrit :

$$x_{ij} = \sum_{k=1}^q c'_{ik} \beta_{kj} + \bar{x}_j + \varepsilon_{ij}$$

$$[X] = [C'].[\beta] + [X^*] + [E]$$

- [X] : Matrice des variables simulées.
- [C'] : Matrice des composantes principales centrée réduite.
- [β] : Matrice des coefficients de régressions.
- [X*] : Matrice des moyennes des x .
- [E] : Matrice des résidus.

Remarque : on parle de matrice de composantes principales réduites, mais en faite cette matrice est centrée réduite, sauf qu'on n'a pas besoins de la centrer, puisque les composantes principales qui découlent d'une A.C.P. normée, sont de moyenne nulle.

III.5. Diagramme de simulation.

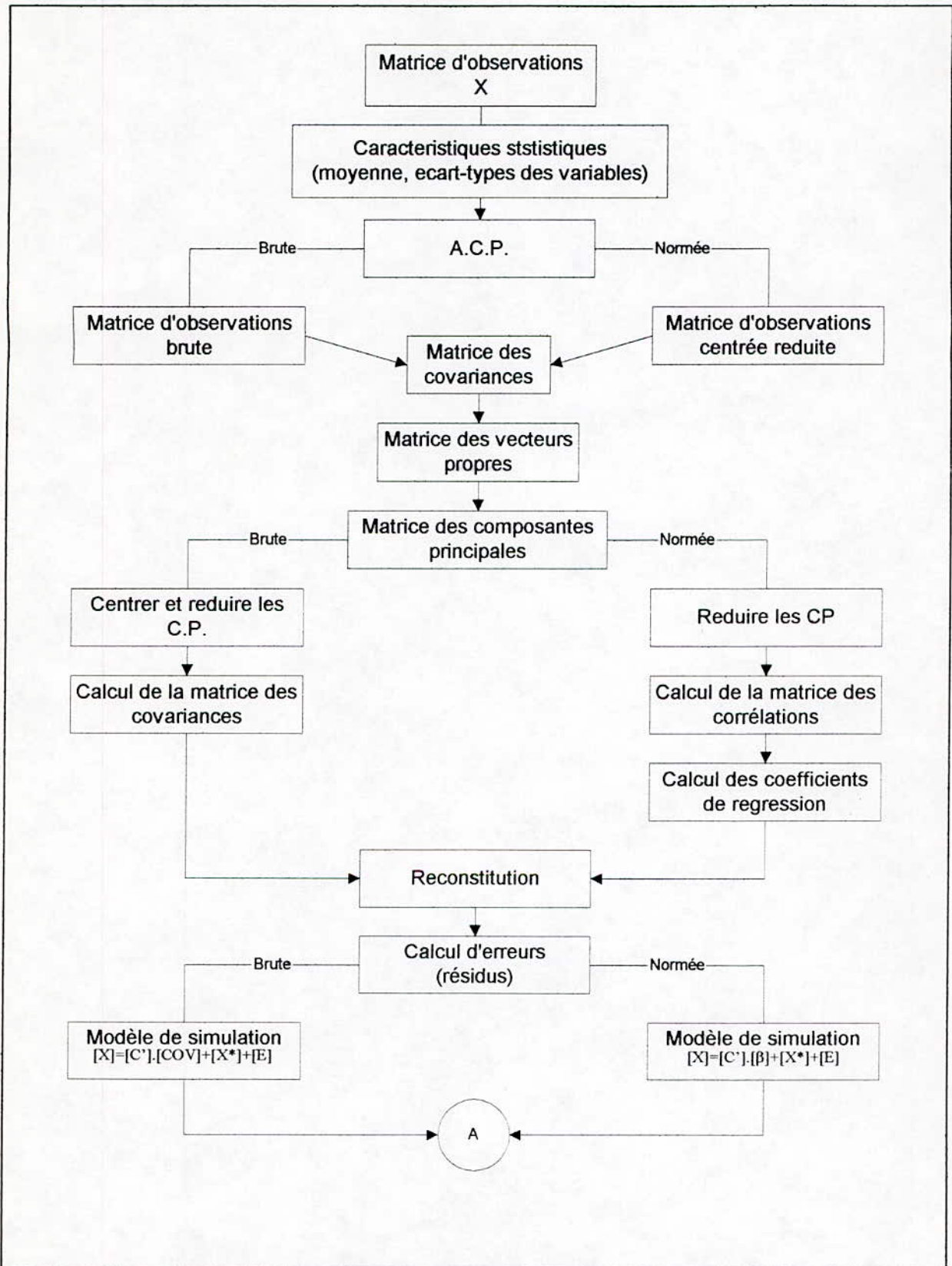


Fig. III-5 : Diagramme de simulation par le modèle GESTOP

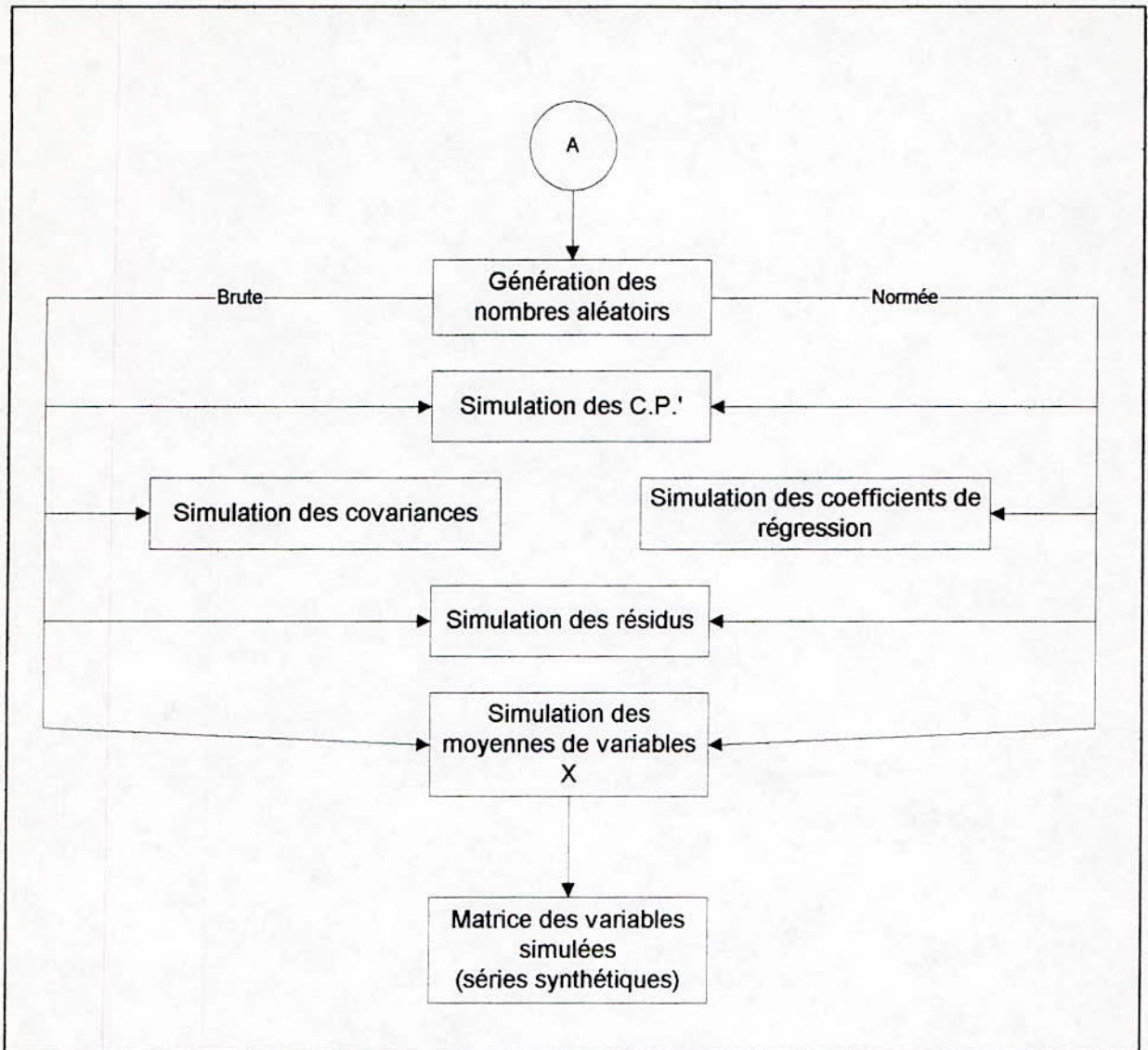


Fig. III-5a : Diagramme de simulation par le modèle GESTOP (suite)

La figure (III-5,5a) représente le processus de simulation par le modèle GESTOP avec les deux variantes, brute et normée.





PARTIAL APPLICATION

Chapitre IV

**Simulation de la pluviométrie au pas
de temps journalier par le modèle GASTOR**

CHAPITRE IV

Simulation de la pluviométrie au pas de temps journalier par le modèle GESTOP

Dans ce chapitre, on va simuler la pluviométrie au pas de temps journalier une série d'observation de la station d'Alger.

IV.1. Données utilisées.

On se propose d'étudier la pluviométrie journalière de la station d'Alger, pour cela on dispose d'un échantillon de 50 années d'observations pour une période allant de 1936 à 1985.

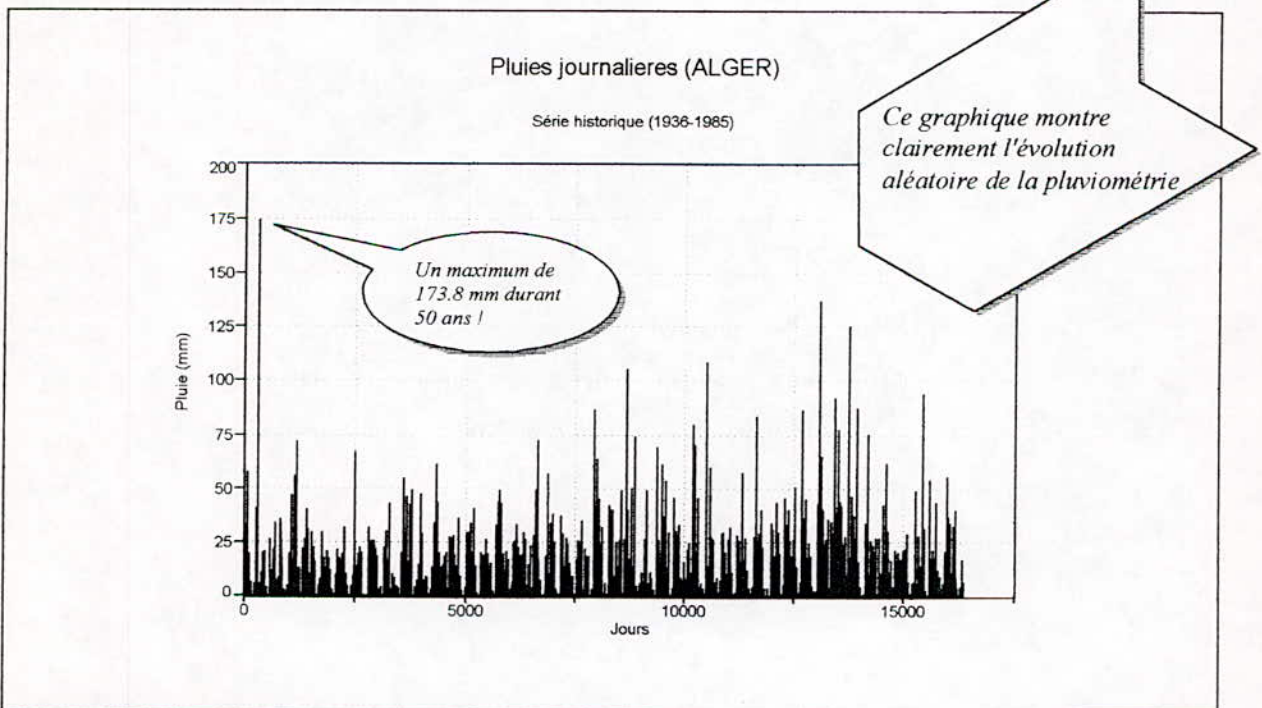


Figure IV-1 : Evolution des pluies historiques

IV.2. Analyse des données par l'A.C.P.

Les caractéristiques statistiques et les résultats de l'A.C.P. de la matrice d'observation (365x50) sont reportés sur le tableau suivant :

Type d'A.C.P. : brute.
 Nombre de variables : 365
 Nombre d'individus : 50

TableauIV-1 : Caractéristiques statistiques de la matrice d'observations pluviométrique

Jour	Moyenne (mm)	Ecart-type (mm)	Valeur propre	Contribution de chaque C.P.	Contribution cumulée
1	004.23	10.51	1389.32	9.0043	9.0043
2	004.11	06.28	982.7	6.3690	15.3733
3	002.63	05.43	812.96	5.2689	20.6421
4	003.75	08.68	791.53	5.1300	25.7721
5	003.74	09.19	639.26	4.1431	29.9152
6	003.56	06.75	508	3.2924	33.2076
7	002.82	05.58	469.33	3.0418	36.2493
8	001.74	04.18	466.05	3.0205	39.2699
9	003.53	08.13	461.77	2.9928	42.2626
10	002.51	06.55	444.6	2.8815	45.1441
11	002.74	05.95	443.96	2.8773	48.0215
12	002.74	05.71	417.85	2.7081	50.7296
13	004.16	09.26	409.43	2.6535	53.3831
14	002.96	07.42	395.34	2.5622	55.9454
15	002.61	04.46	356.48	2.3104	58.2557
16	001.96	03.87	328.38	2.1283	60.3840
17	003.62	09.26	318.18	2.0622	62.4461
18	004.94	08.97	306.23	1.9847	64.4308
19	002.56	04.85	287.56	1.8637	66.2945
20	003.89	09.93	286.97	1.8599	68.1544
21	002.24	05.78	281.71	1.8258	69.9802
22	004.61	08.71	271.73	1.7611	71.7413
23	002.42	04.91	269.38	1.7459	73.4872
24	001.86	06.79	256.1	1.6598	75.1470
25	001.80	05.74	221.93	1.4383	76.5853
26	002.78	08.89	219.86	1.4249	78.0103
27	002.01	05.41	214.48	1.3901	79.4003
28	002.22	05.06	193.63	1.2549	80.6553
29	002.31	05.20	192.37	1.2468	81.9020
30	001.46	05.38	171	1.1083	83.0103
31	001.67	03.01	169.36	1.0976	84.1079
32	003.50	08.31	167.85	1.0878	85.1958
33	002.65	05.37	163.82	1.0617	86.2575
34	003.87	11.63	148.59	0.9630	87.2205
35	003.83	09.26	135.02	0.8751	88.0956
36	001.82	04.65	131.91	0.8549	88.9505
37	003.59	11.03	128.25	0.8312	89.7817
38	002.12	04.94	126.52	0.8200	90.6017
39	002.76	10.55	117.21	0.7596	91.3614
40	002.58	05.49	115.87	0.7510	92.1123

41	003.55	07.92		114.23	0.7403	92.8527
42	002.73	04.22		112.41	0.7285	93.5812
43	003.24	06.35		110.23	0.7144	94.2956
44	001.99	03.91		100.61	0.6521	94.9477
45	002.94	04.39		98.41	0.6378	95.5855
46	003.67	06.91		84.21	0.5458	96.1312
47	003.35	05.85		74.79	0.4847	96.6160
48	002.49	08.58		62.37	0.4042	97.0202
49	002.73	05.09		55.19	0.3577	97.3779
50	003.09	05.81		52.1	0.3377	97.7155
51	002.72	05.90		26.57	0.1722	97.8878
52	001.98	04.41		21.94	0.1422	98.0299
53	001.67	03.56		20.04	0.1299	98.1598
54	002.61	07.05		18.34	0.1189	98.2787
55	003.33	05.86		18	0.1167	98.3953
56	002.25	04.49		17.85	0.1157	98.5110
57	002.40	05.89		17.31	0.1122	98.6232
58	003.05	06.75		15.54	0.1007	98.7239
59	003.15	07.69		14.31	0.0927	98.8167
60	002.57	06.69		14.11	0.0914	98.9081
61	002.27	05.03		13.62	0.0883	98.9964
62	001.97	06.55		10.89	0.0706	99.0670
63	002.21	04.21		10.73	0.0695	99.1365
64	003.16	07.99		10.36	0.0671	99.2037
65	002.70	06.89		9.09	0.0589	99.2626
66	001.79	04.40		8.31	0.0539	99.3164
67	002.16	05.58		8.25	0.0535	99.3699
68	002.84	05.80		8.11	0.0526	99.4225
69	001.99	04.64		7.5	0.0486	99.4711
70	002.12	04.98		6.47	0.0419	99.5130
71	002.51	05.55		6.27	0.0406	99.5536
72	001.93	03.75		5.92	0.0384	99.5920
73	002.63	04.56		5.59	0.0362	99.6282
74	000.65	01.81		5.25	0.0340	99.6623
75	002.61	06.83		4.32	0.0280	99.6903
76	000.94	02.75		4.15	0.0269	99.7172
77	000.94	02.28		3.75	0.0243	99.7415
78	001.66	03.86		3.49	0.0226	99.7641
79	001.68	03.77		3.45	0.0224	99.7864
80	001.36	02.80		3.41	0.0221	99.8085
81	000.63	01.54		2.89	0.0187	99.8273
82	001.45	04.44		2.84	0.0184	99.8457
83	002.81	08.39		2.22	0.0144	99.8601
84	001.85	03.77		2.05	0.0133	99.8734
85	001.76	04.51		2	0.0130	99.8863
86	001.20	02.98		1.89	0.0122	99.8986
87	004.47	08.33		1.88	0.0122	99.9108
88	004.64	13.67		1.37	0.0089	99.9196
89	003.09	05.99		1.25	0.0081	99.9277
90	001.85	05.07		0.96	0.0062	99.9340

91	001.53	04.20		0.94	0.0061	99.9400
92	001.51	03.57		0.88	0.0057	99.9458
93	001.93	06.08		0.85	0.0055	99.9513
94	001.91	04.04		0.84	0.0054	99.9567
95	001.34	03.83		0.69	0.0045	99.9612
96	002.98	13.40		0.56	0.0036	99.9648
97	003.68	11.68		0.44	0.0029	99.9677
98	003.17	09.80		0.41	0.0027	99.9703
99	001.39	04.13		0.34	0.0022	99.9725
100	002.86	07.15		0.33	0.0021	99.9747
101	001.52	04.16		0.29	0.0019	99.9765
102	004.08	12.22		0.26	0.0017	99.9782
103	003.83	12.76		0.25	0.0016	99.9798
104	002.52	05.36		0.24	0.0016	99.9814
105	002.67	07.39		0.22	0.0014	99.9828
106	001.86	04.08		0.22	0.0014	99.9843
107	001.60	03.51		0.19	0.0012	99.9855
108	001.21	03.39		0.18	0.0012	99.9866
109	001.68	04.70		0.18	0.0012	99.9878
110	002.86	07.32		0.15	0.0010	99.9888
111	003.20	09.26		0.14	0.0009	99.9897
112	003.74	08.04		0.14	0.0009	99.9906
113	001.59	04.39		0.13	0.0008	99.9914
114	001.28	03.43		0.13	0.0008	99.9923
115	000.97	03.50		0.12	0.0008	99.9931
116	001.62	04.05		0.11	0.0007	99.9938
117	002.43	05.94		0.1	0.0006	99.9944
118	002.79	05.92		0.09	0.0006	99.9950
119	002.04	06.55		0.09	0.0006	99.9956
120	002.60	06.42		0.07	0.0005	99.9960
121	002.45	06.18		0.06	0.0004	99.9964
122	001.82	04.48		0.06	0.0004	99.9968
123	002.72	09.25		0.06	0.0004	99.9972
124	001.82	04.34		0.05	0.0003	99.9975
125	001.64	05.76		0.05	0.0003	99.9979
126	001.06	02.85		0.04	0.0003	99.9981
127	001.30	03.16		0.04	0.0003	99.9984
128	002.35	06.57		0.02	0.0001	99.9985
129	002.44	05.70		0.02	0.0001	99.9986
130	000.80	03.29		0.02	0.0001	99.9988
131	000.38	01.42		0.02	0.0001	99.9989
132	000.97	03.13		0.02	0.0001	99.9990
133	001.24	05.05		0.02	0.0001	99.9992
134	000.93	04.28		0.02	0.0001	99.9993
135	000.91	02.68		0.01	0.0001	99.9994
136	000.69	02.00		0.01	0.0001	99.9994
137	002.13	06.52		0.01	0.0001	99.9995
138	002.45	07.59		0.01	0.0001	99.9995
139	001.69	06.87		0.01	0.0001	99.9996
140	000.77	02.93		0.01	0.0001	99.9997

141	000.31	01.61		0.01	0.0001	99.9997
142	000.88	03.01		0.01	0.0001	99.9998
143	000.17	00.71		0.01	0.0001	99.9999
144	000.47	01.42		0.01	0.0001	99.9999
145	001.40	05.68		0.01	0.0001	100.0000
146	000.37	01.83		0.00	0.0000	100.0000
147	000.91	02.74		0.00	0.0000	100.0000
148	000.70	02.71		0.00	0.0000	100.0000
149	000.61	01.94		0.00	0.0000	100.0000
150	001.16	06.10		0.00	0.0000	100.0000
⋮	⋮	⋮		0.00	0.0000	100.0000
356	004.01	09.00		0.00	0.0000	100.0000

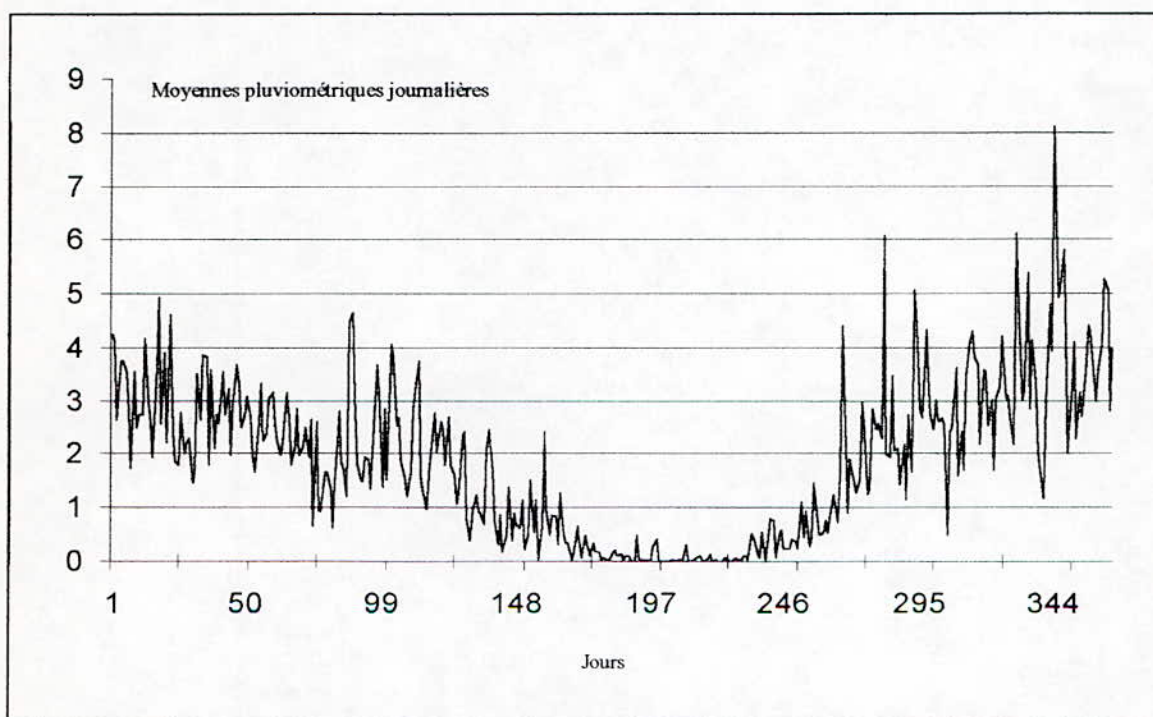


Fig. IV-2 : Variation des moyennes pluviométriques journalières interannuelles

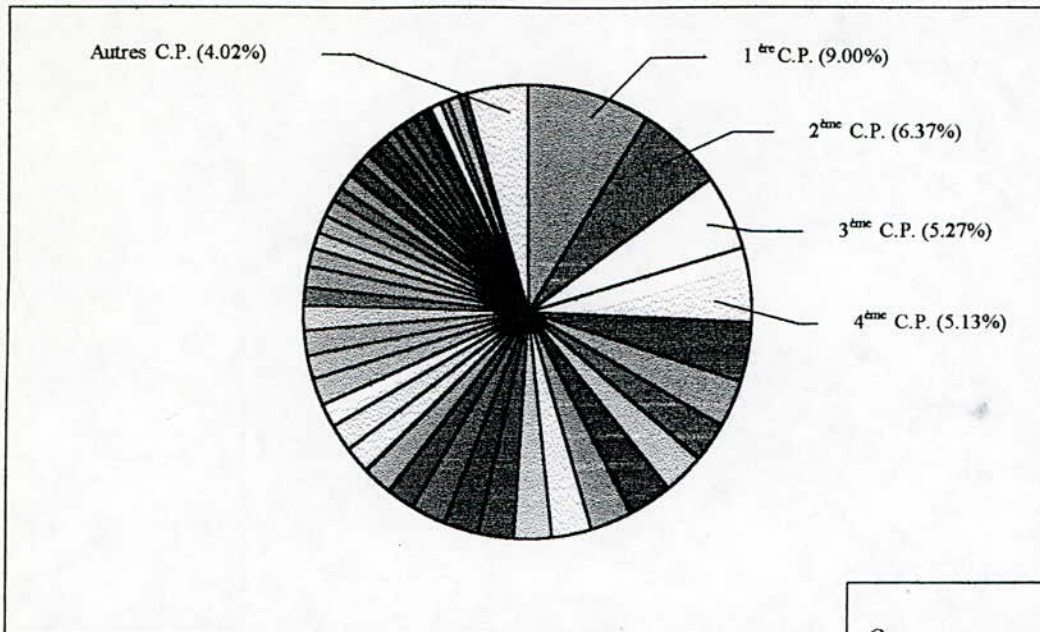


Fig. IV-3 : Pourcentage de la variance apportée par chaque C.P.

On remarque que la variance totale est expliquée par 145/365 composantes principales

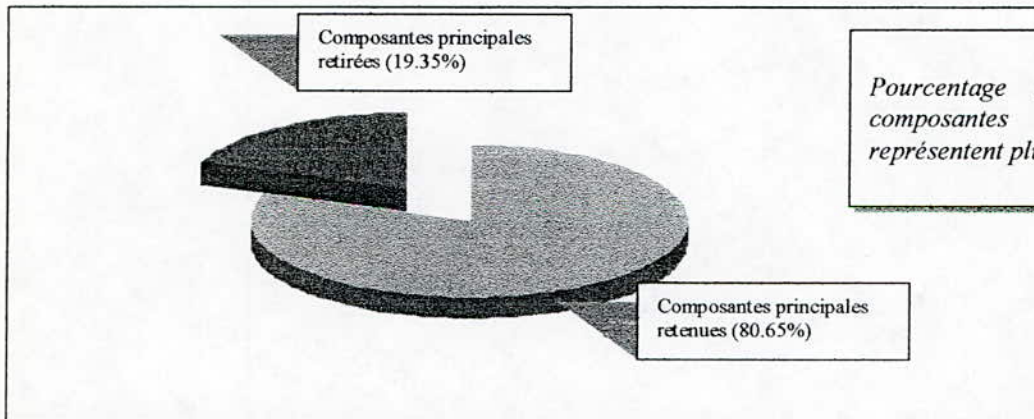


Fig. IV-4 : Choix du nombre des composantes principales à retenir

Pourcentage retenu par 28 composantes principales qui représentent plus de 80%

IV.3. Reconstitution.

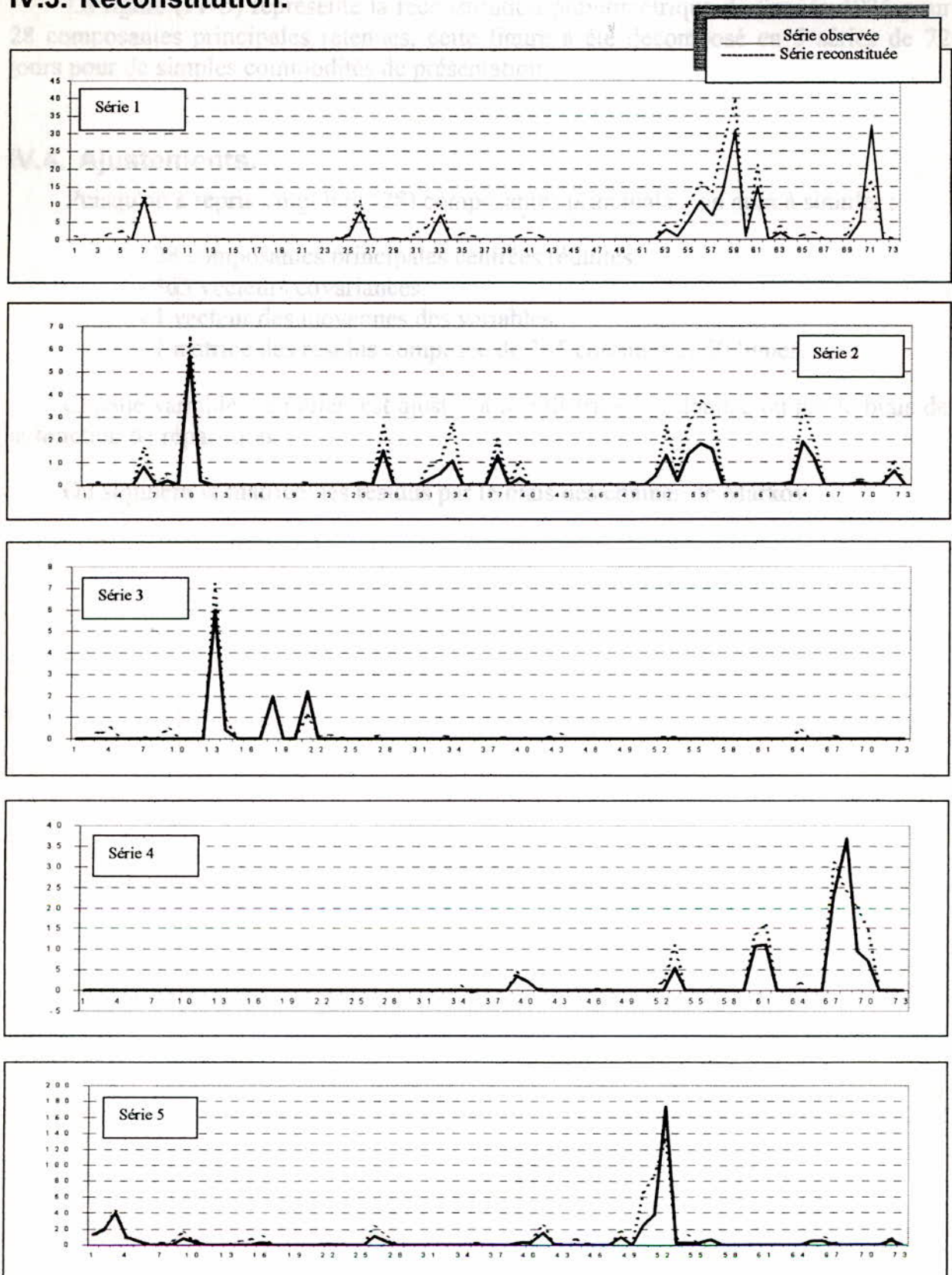
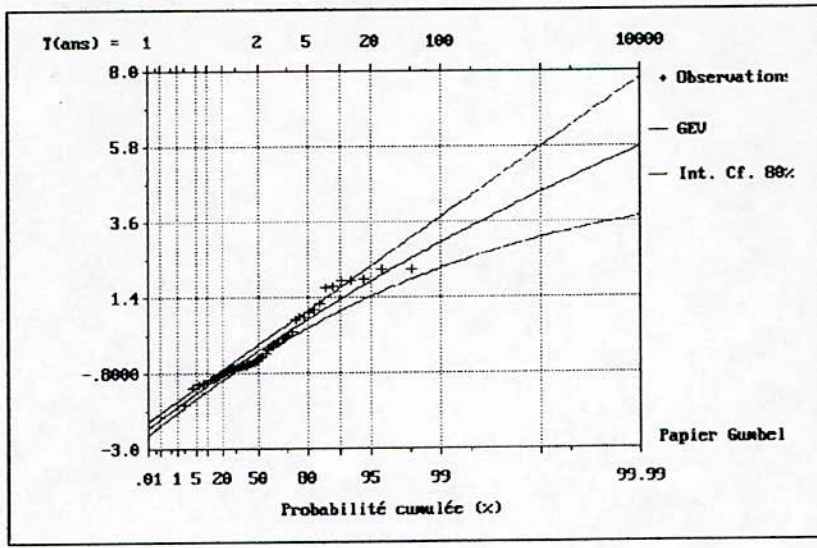
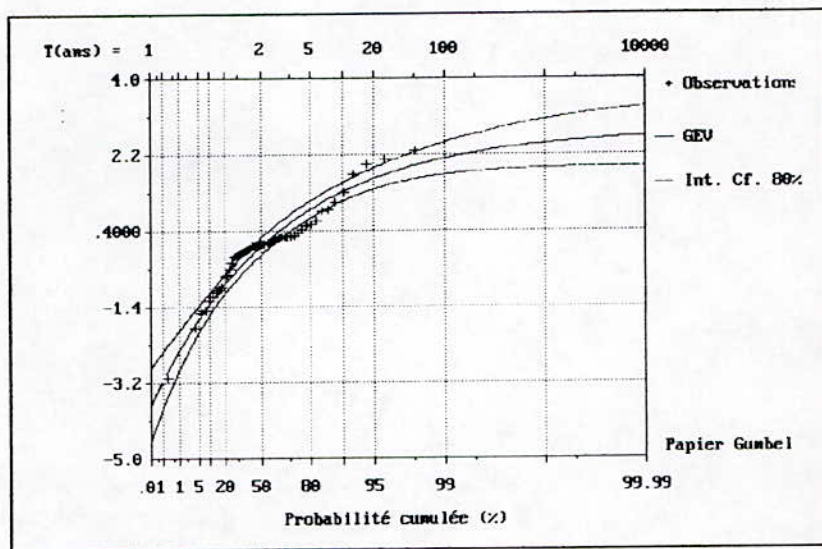


Fig. IV-5 : Reconstitution pluviométrique de l'année 1936 pour 28 C.P.

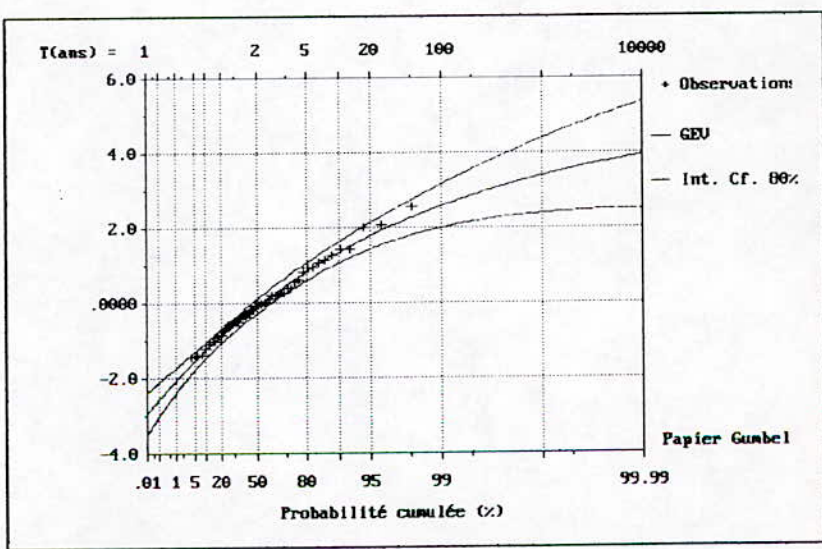
Ajustement : Loi G.E.V.
Variable : 1^{ère} C.P centrée réduite



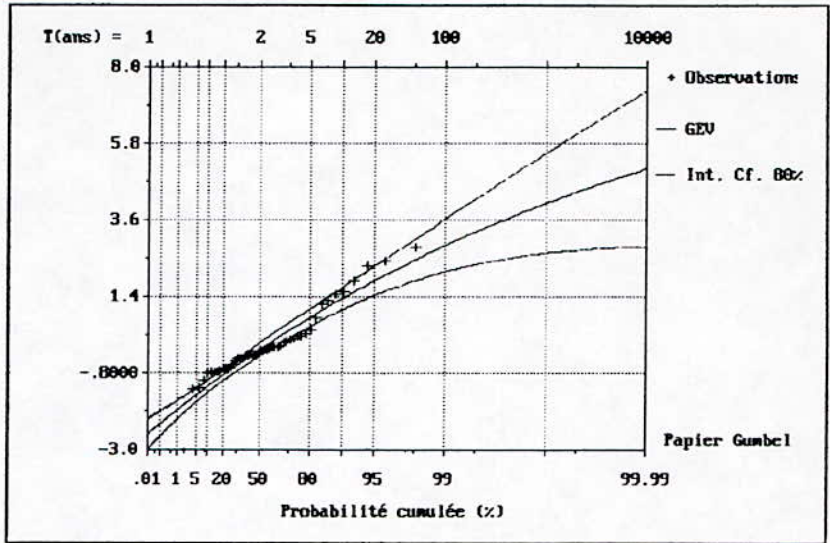
Ajustement : Loi G.E.V.
Variable : 2^{ème} C.P centrée réduite



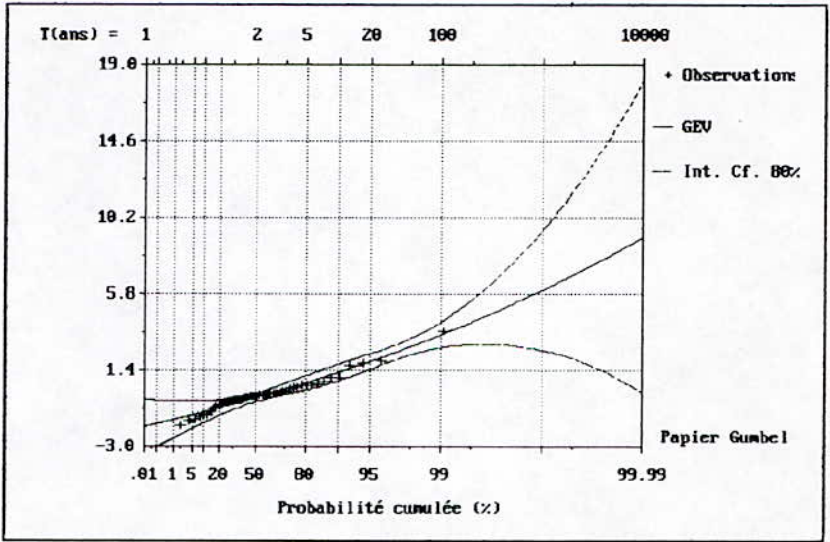
Ajustement : Loi G.E.V.
Variable : 3^{ème} C.P centrée réduite



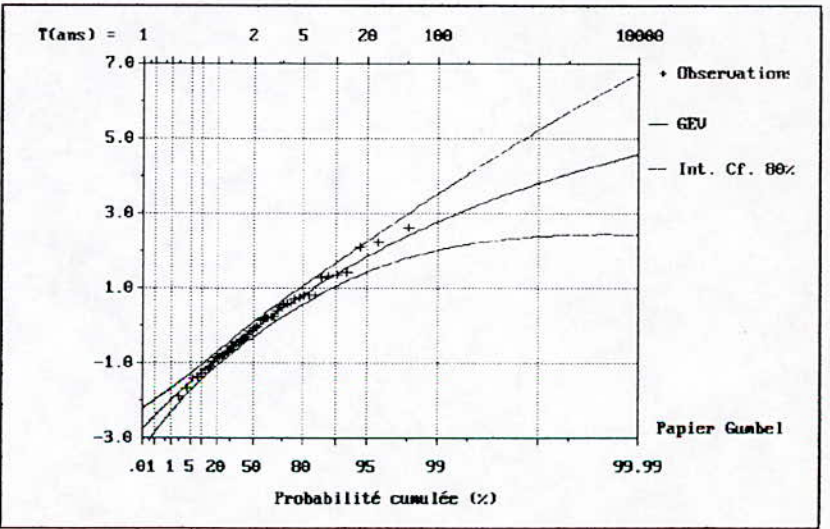
Ajustement : Loi G.E.V.
Variable : 4^{ème} C.P centrée réduite



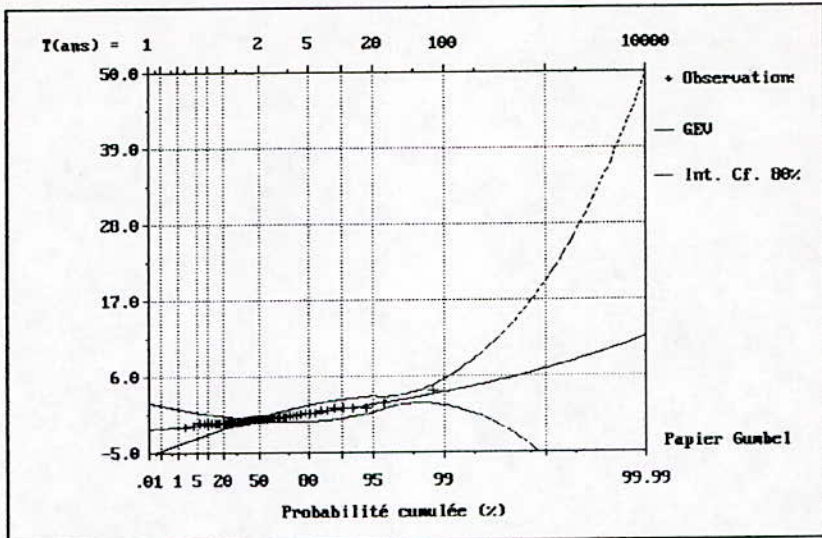
Ajustement : Loi G.E.V.
Variable : 5^{ème} C.P centrée réduite



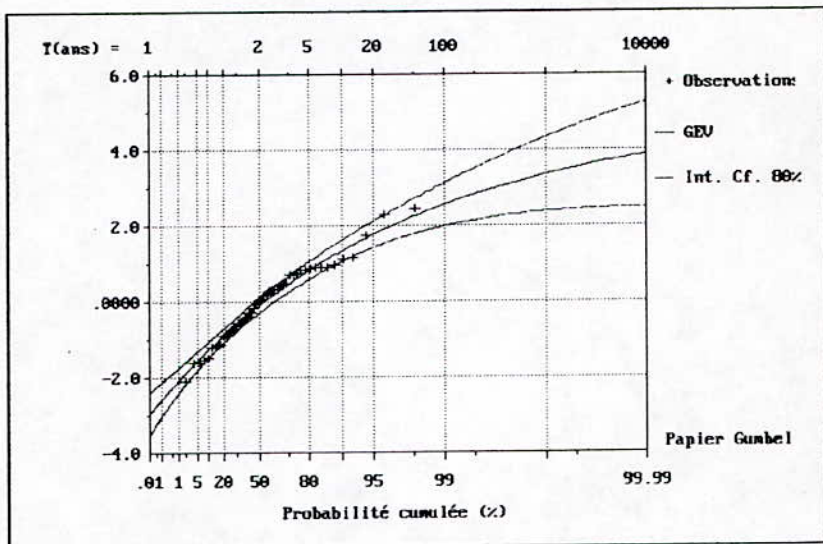
Ajustement : Loi G.E.V.
Variable : 6^{ème} C.P centrée réduite



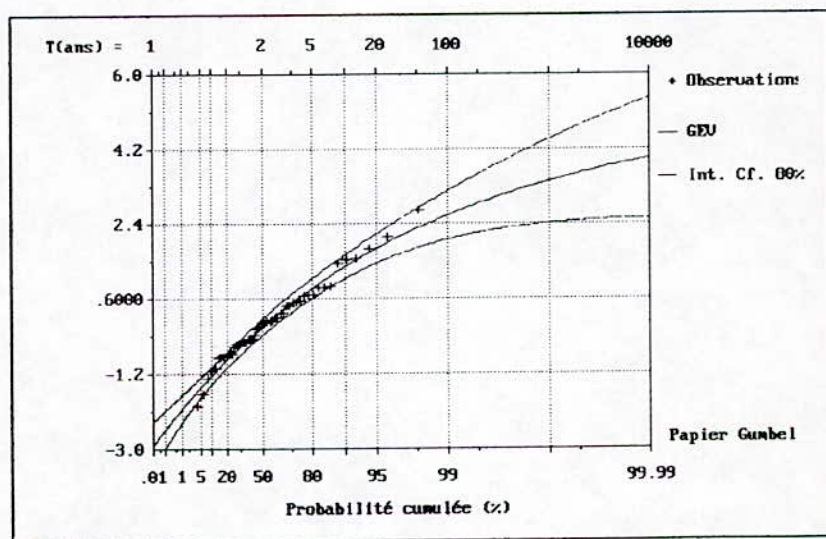
Ajustement : Loi G.E.V.
Variable : 7^{me} C.P centrée réduite



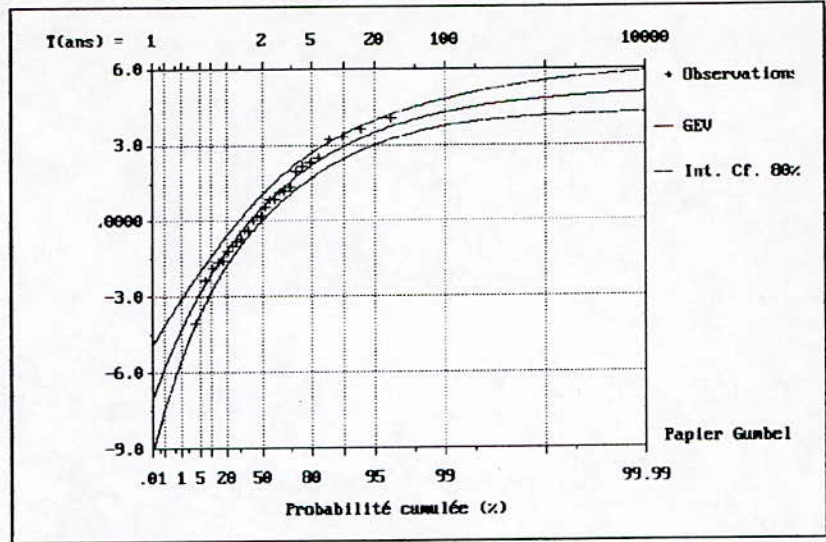
Ajustement : Loi G.E.V.
Variable : 8^{me} C.P centrée réduite



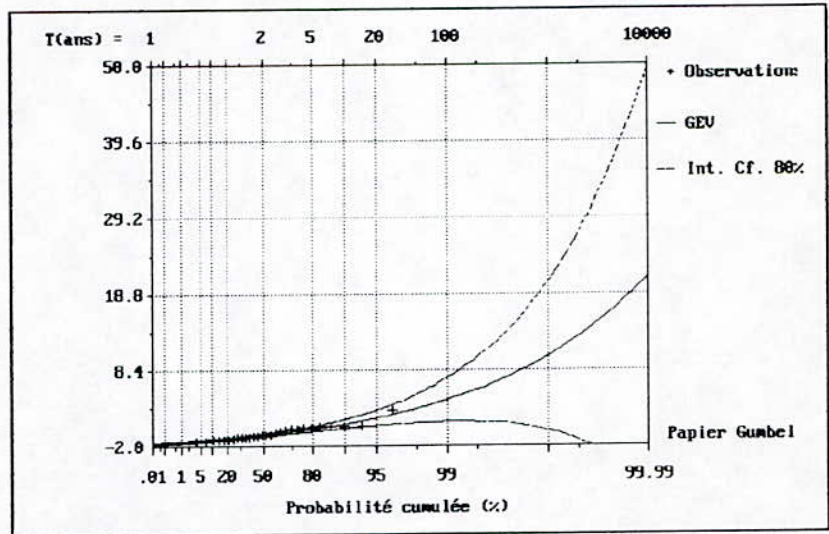
Ajustement : Loi G.E.V.
Variable : 9^{me} C.P centrée réduite



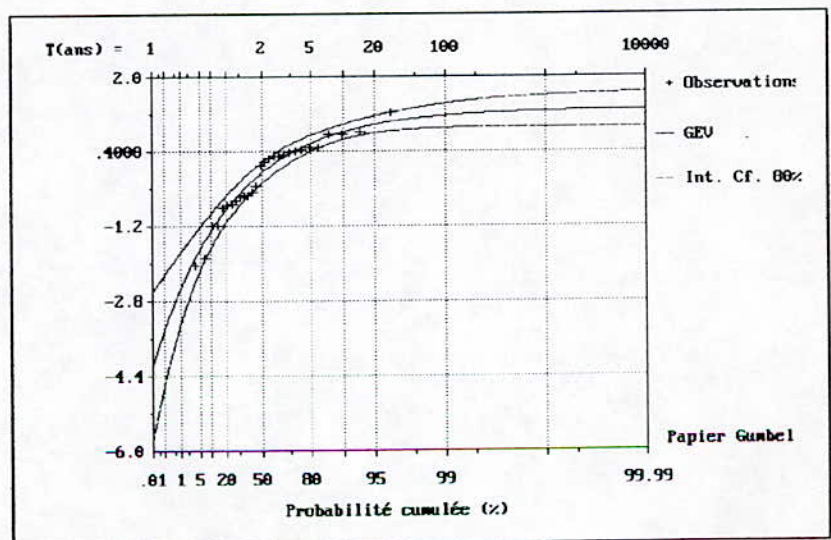
Ajustement : G.E.V.
Variable : 1^{ère} covariance



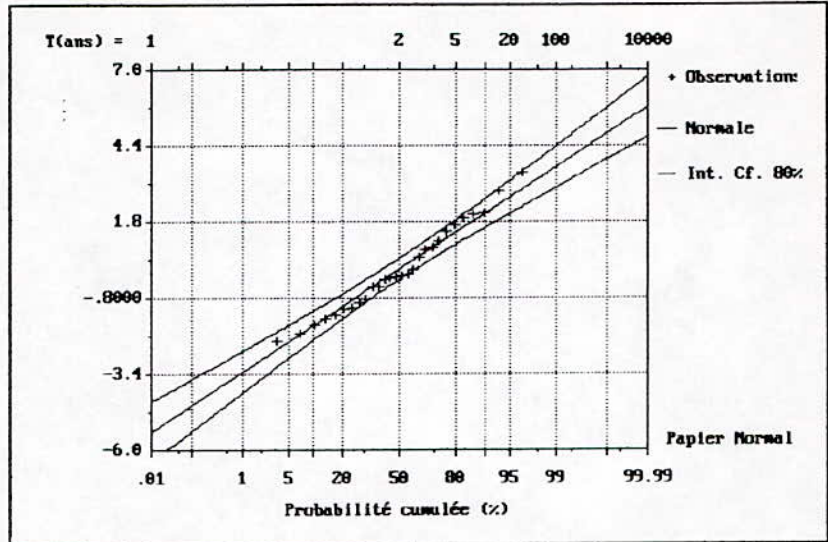
Ajustement : G.E.V.
Variable : 2^{ème} covariance



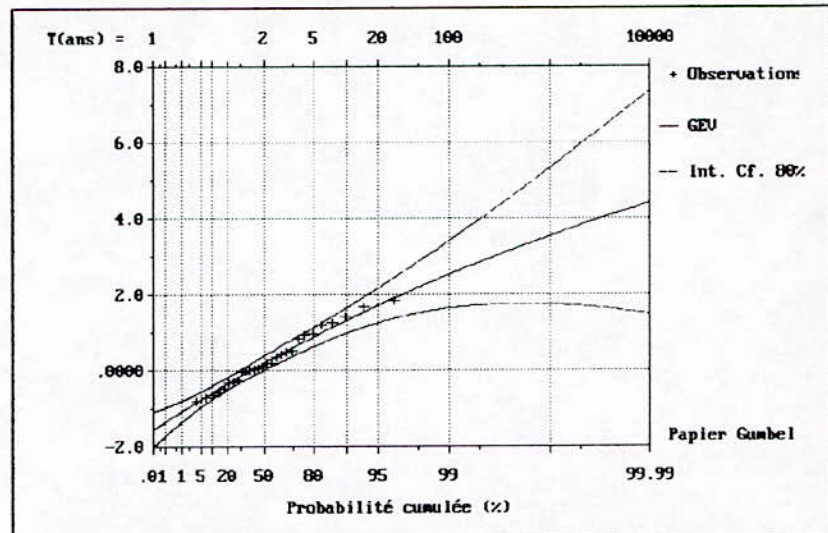
Ajustement : G.E.V.
Variable : 3^{ème} covariance



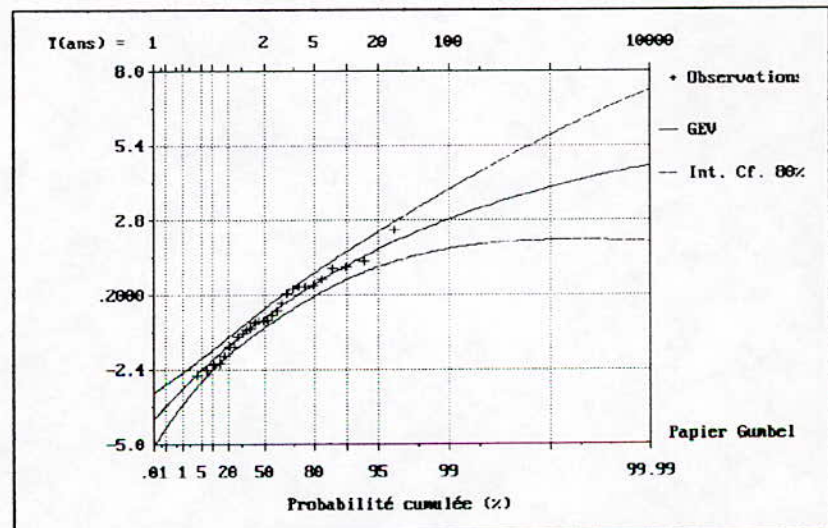
Ajustement : Loi normale
Variable : 4^{me} covariance



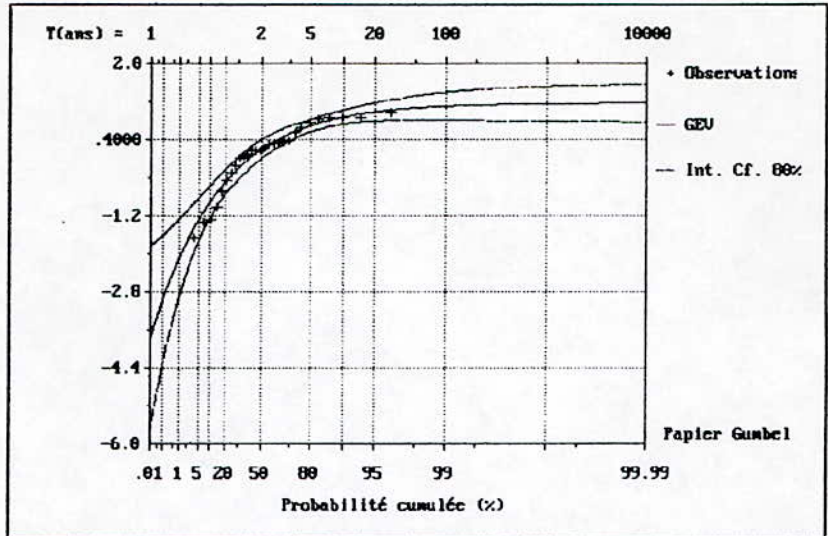
Ajustement : G.E.V.
Variable : 5^{me} covariance



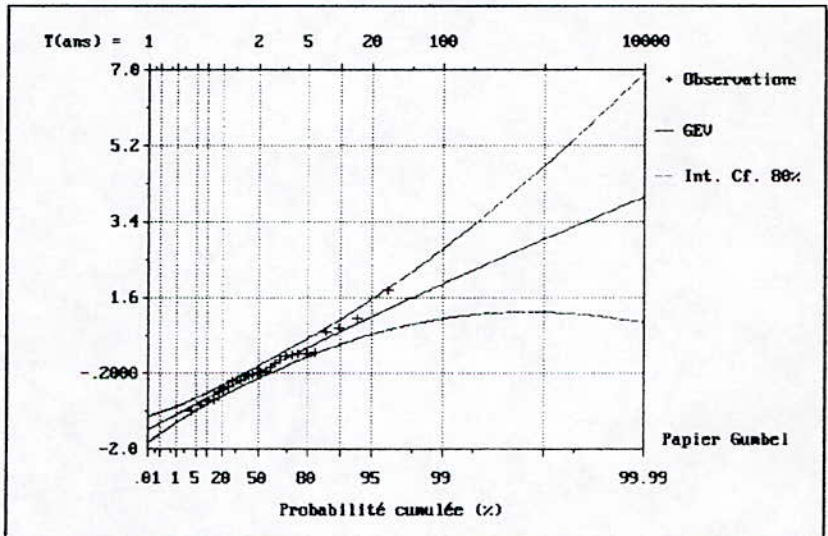
Ajustement : G.E.V.
Variable : 6^{me} covariance



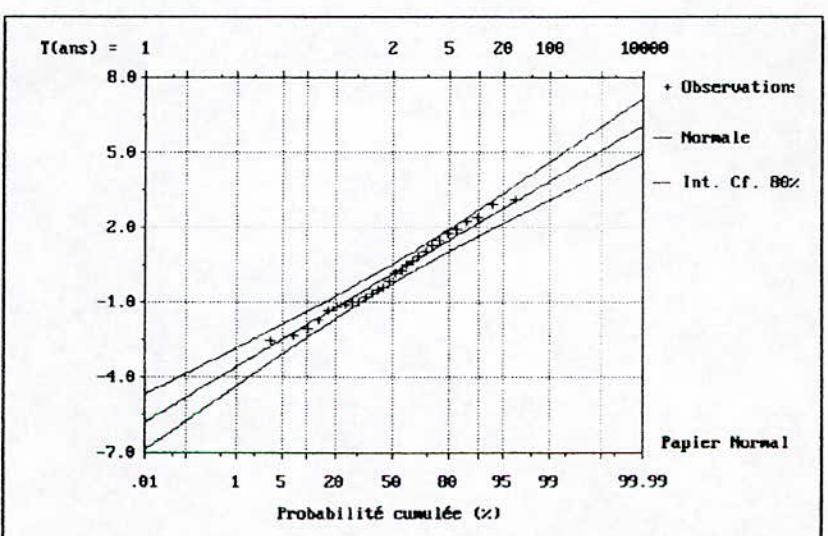
Ajustement : Loi G.E.V.
Variable : 7^{ème} covariance



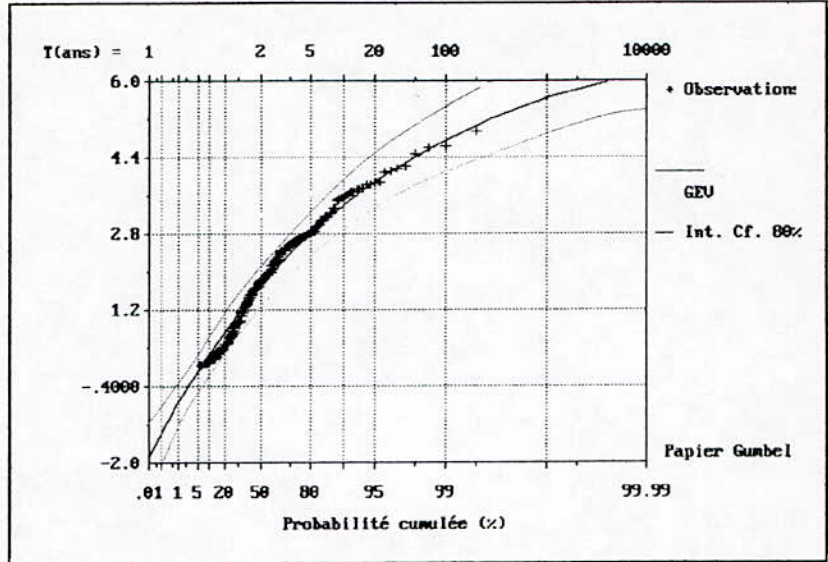
Ajustement : G.E.V.
Variable : 8^{ème} covariance



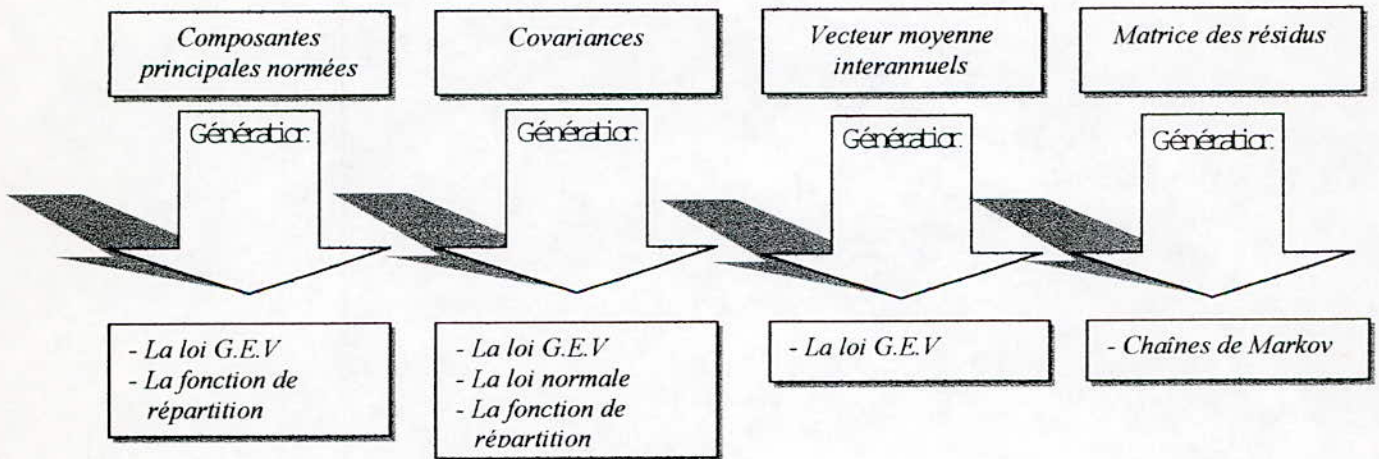
Ajustement : Loi normale.
Variable : 9^{ème} covariance



Ajustement : G.E.V.
Variable : Moyenne des variables



On a remarqué que pratiquement toutes les variables ont été ajustées par loi G.E.V., les variables ajustées à la loi normale étaient très rares, et quelque fois, on était obligé d'ajuster par le biais de la fonction de répartition.



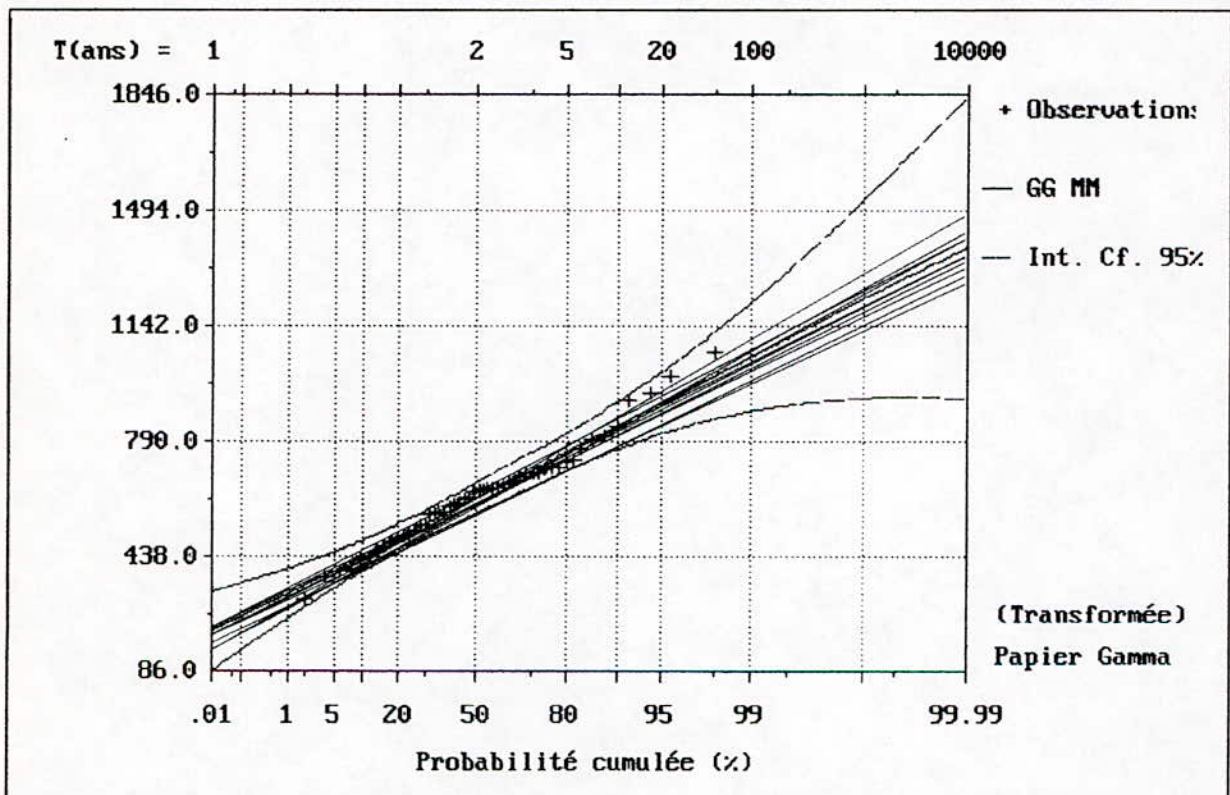
IV.5. Etude des résultats.

La question qui se pose maintenant est : " les séries générées appartiennent-elles à la même population que la série historique ?" Puisqu'on a estimé les paramètres statistiques de l'échantillon observé, on va essayer de les comparer avec la série simulée et donner des jugements sur l'ensemble.

Pour cela on va tirer dix séries simulées, et on les comparera au cumul annuel, moyennes annuelles, écart-types annuels et les maximums également annuels.

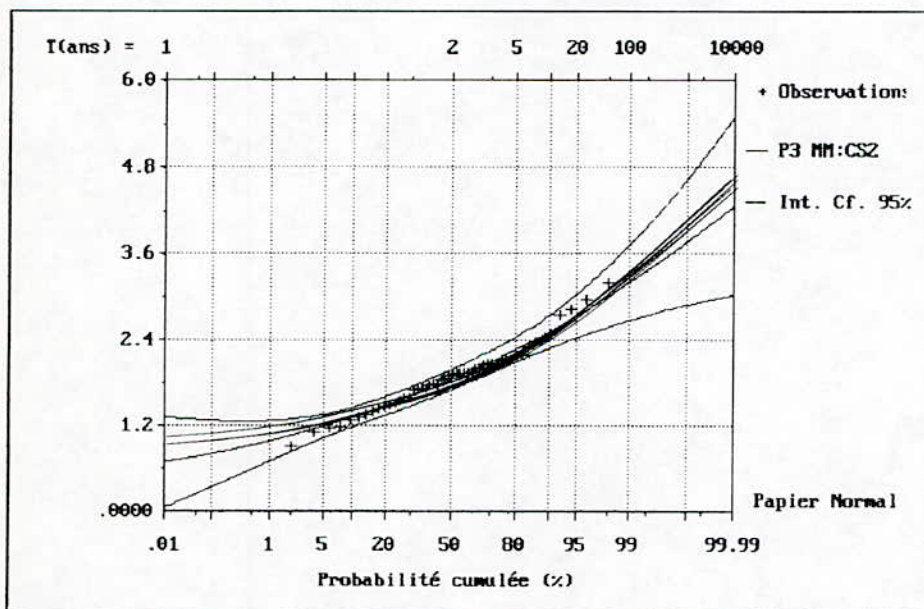
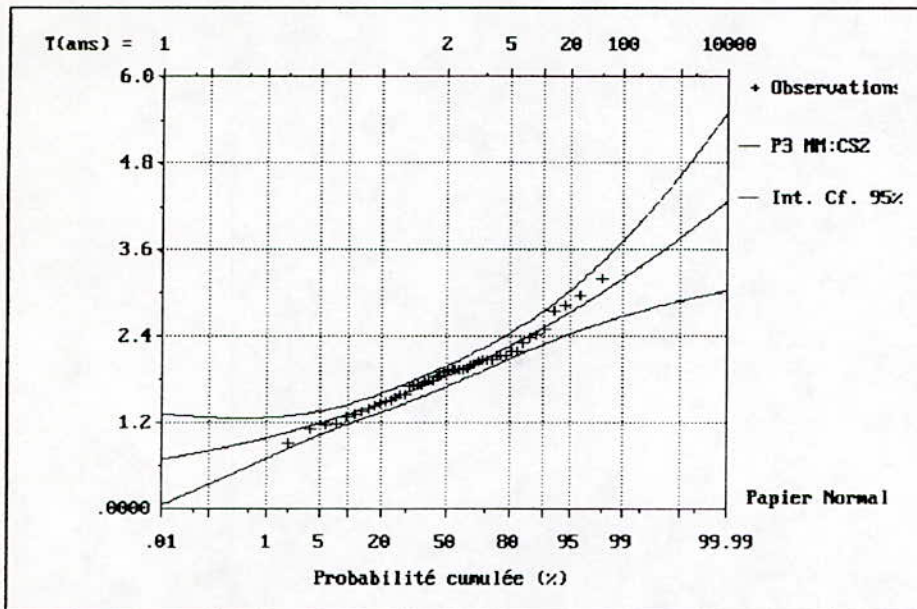
IV.5.1. Cumul annuel.

Le cumul annuel de la série historique s'ajuste à la loi Gamma généralisée, de même pour les 10 séries simulées, qui sont à l'intérieur de l'intervalle de confiance du cumul historique.



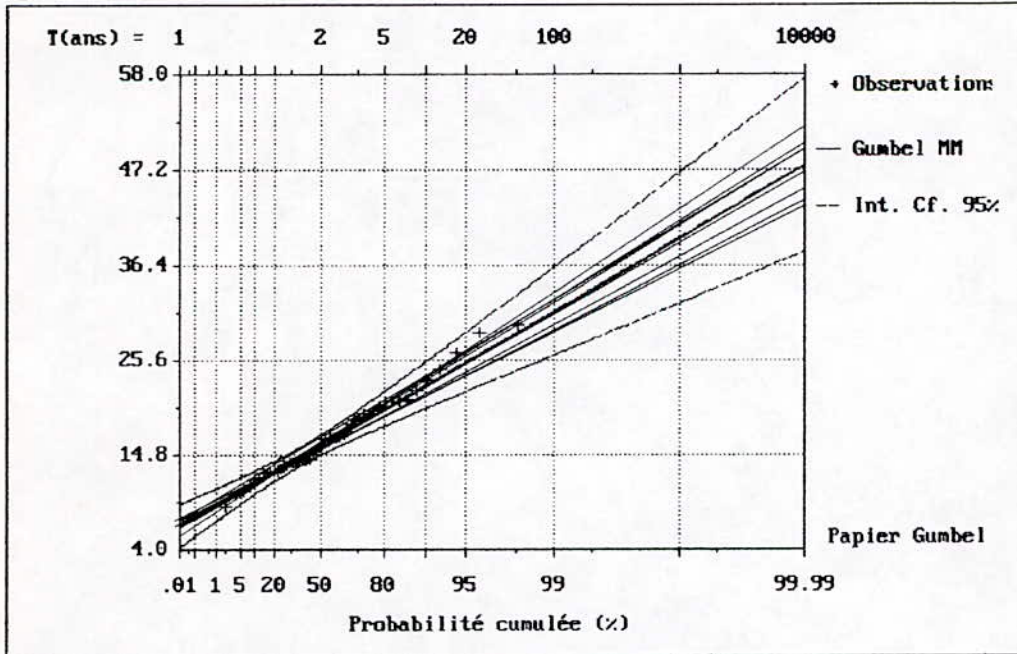
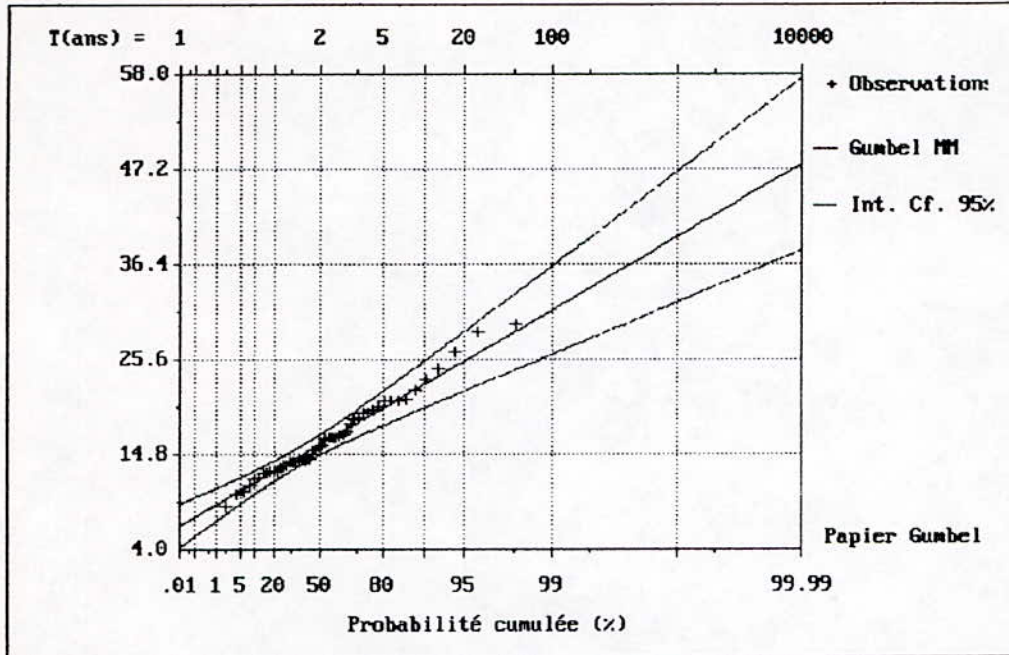
IV.5.2. Moyennes annuelles.

Les moyennes annuelles de la série historique s'ajuste à la loi Pearson, de même pour les moyennes annuelles simulées, qui sont à l'intérieur de l'intervalle de confiance de la variable historique.



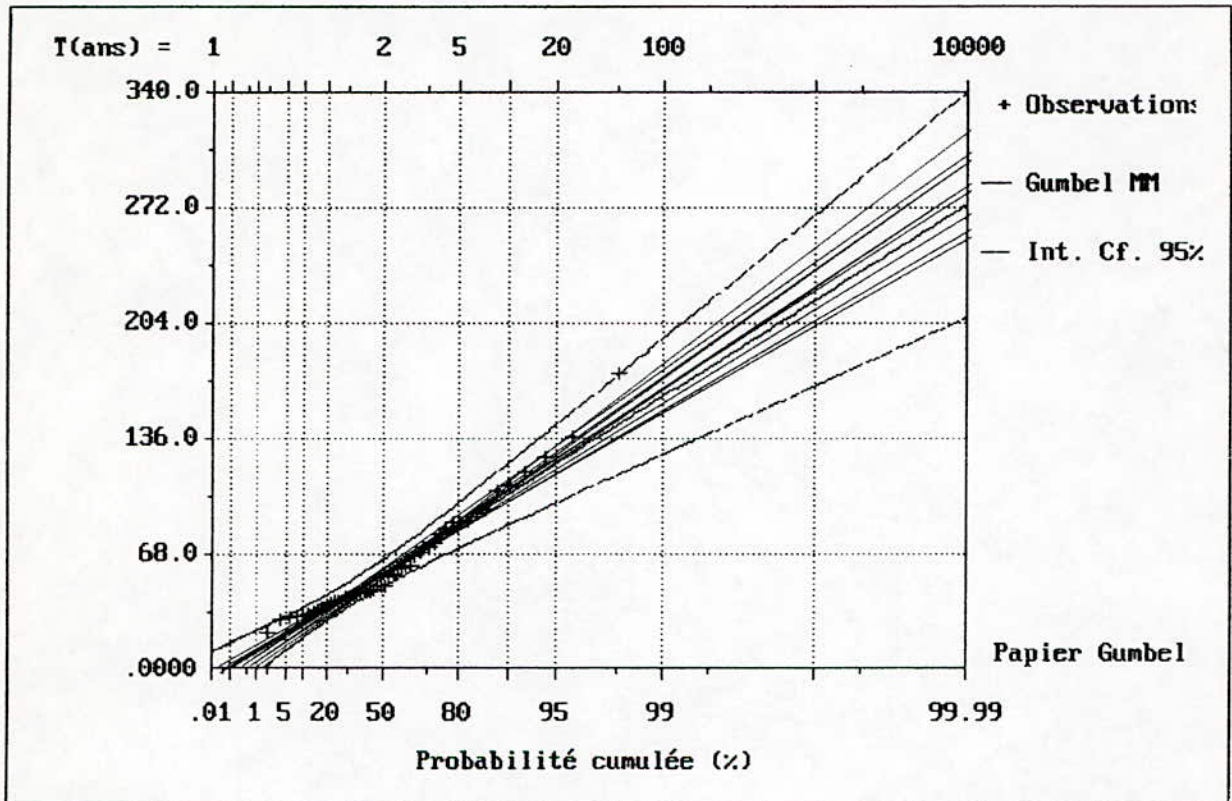
IV.5.3. Ecart-types annuels.

Les écart-types de la série historique s'ajuste à la loi Gumbel, de même pour les 10 séries simulées, qui sont à l'intérieur de l'intervalle de confiance des écart-types historiques.



IV.5.4. Maximums annuels.

Les maximums annuels de la série historique s'ajuste à la loi Gumbel, de même pour les maximums simulées, qui sont à l'intérieur de l'intervalle de confiance de la variable historique.



Chapitre V

**Simulation des températures au pas
de temps journalier par le modèle GRS'09**

CHAPITRE V

Simulation des températures au pas de temps journalier par le modèle GESTOP

Dans ce chapitre, on va simuler les températures au pas de temps journalier d'une série d'observation de la station d'Alger.

V.1. Données utilisées.

On se propose d'étudier les températures journalières au niveau de la station d'Alger, pour cela on dispose d'un échantillon de 22 années d'observations pour une période allant de 1967 à 1988.

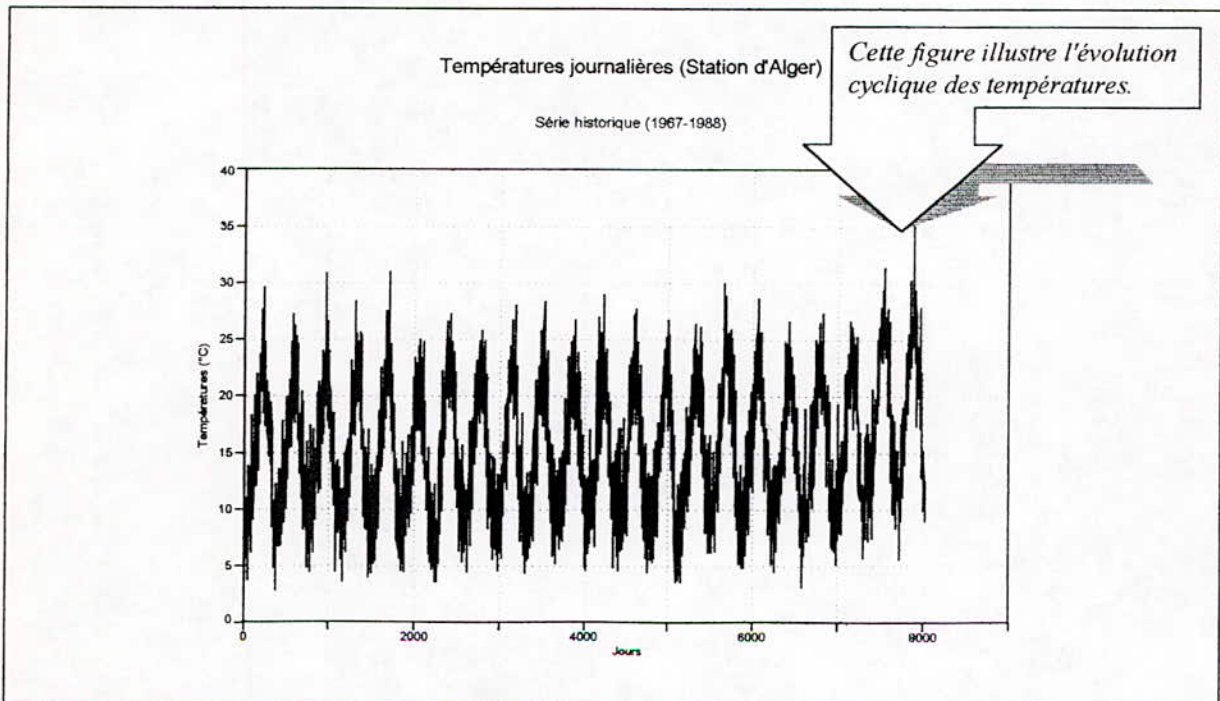


Figure V-1 : Evolution des températures historiques

V.2. Analyse des données par l'A.C.P.

Les caractéristiques statistiques et les résultats de l'A.C.P. de la matrice d'observation (365x22) sont reportés sur le tableau suivant :

Type d'A.C.P. : brute.
 Nombre de variables : 365
 Nombre d'individus : 22

Tableau V-1 : Caractéristiques statistiques de la matrice d'observations des températures

Jour	Moyenne (°C)	Ecart-type (°C)	Valeur propre	Contribution de chaque C.P.	Contribution cumulée
1	9.8	2.32	164.650	17.295	17.295
2	9.4	2.08	97.560	10.248	27.542
3	9.1	1.87	76.540	8.040	35.582
4	9.68	2.06	62.320	6.546	42.128
5	9.79	2.03	56.088	5.891	48.019
6	9.55	2.63	50.479	5.302	53.322
7	9.48	2.48	45.431	4.772	58.094
8	9.51	2.5	40.888	4.295	62.389
9	9.45	2.83	36.799	3.865	66.254
10	9.28	2.51	33.119	3.479	69.733
11	8.73	2.68	29.807	3.131	72.864
12	8.61	2.25	26.827	2.818	75.682
13	9.23	2.76	24.144	2.536	78.218
14	9.05	3.04	21.730	2.282	80.500
15	9.69	2.52	19.557	2.054	82.554
16	9.85	2.47	17.601	1.849	84.403
17	9.06	2.06	15.841	1.664	86.067
18	8.48	2.66	14.257	1.498	87.565
19	9.03	2.04	12.831	1.348	88.912
20	9.27	2.21	11.548	1.213	90.125
21	9.42	1.83	10.393	1.092	91.217
22	9.61	2.45	9.354	0.983	92.200
23	9.75	2.55	8.419	0.884	93.084
24	9.41	1.99	7.577	0.796	93.880
25	9.92	2.12	6.819	0.716	94.596
26	9.25	2.31	6.137	0.645	95.241
27	9.54	2.8	5.523	0.580	95.821
28	9.52	3.31	4.971	0.522	96.343
29	10.01	2.87	4.474	0.470	96.813
30	9.52	2.67	4.027	0.423	97.236
31	9.58	2.4	3.624	0.381	97.616
32	9.32	2.06	3.262	0.343	97.959
33	9.4	2.26	2.935	0.308	98.267
34	9.52	2.57	2.642	0.277	98.545
35	9.75	2.07	2.378	0.250	98.795
36	9.71	2.78	2.140	0.225	99.019
37	9.94	2.82	1.926	0.202	99.222
38	10.01	2.36	1.541	0.162	99.384
39	10.19	2.79	1.233	0.129	99.513
40	9.36	2.58	0.986	0.104	99.617

41	9.89	2.64		0.789	0.083	99.699
42	9.74	2.84		0.631	0.066	99.766
43	9.56	2.61		0.505	0.053	99.819
44	9.52	2.43		0.404	0.042	99.861
45	9.35	2.18		0.323	0.034	99.895
46	9.09	2.35		0.258	0.027	99.922
47	8.95	2.78		0.207	0.022	99.944
48	9.6	3.46		0.165	0.017	99.961
49	9.61	2.85		0.116	0.012	99.974
50	9.65	2.97		0.081	0.009	99.982
51	9.56	2.7		0.057	0.006	99.988
52	9.98	2.23		0.040	0.004	99.992
53	9.52	2.11		0.028	0.003	99.995
54	10	2.42		0.019	0.002	99.997
55	10.22	2.59		0.014	0.001	99.999
56	9.57	1.82		0.007	0.001	99.999
57	9.19	2.52		0.003	0.000	100.000
58	8.8	2.4		0.002	0.000	100.000
59	11.87	4.69		0.001	0.000	100.000
60	9.47	3.42		0.000	0.000	100.000
61	9.32	3.32		0.000	0.000	100.000
62	9.77	2.5		0.000	0.000	100.000
.....		0.00	0.0000	100.000
356	9.69	2.24		0.00	0.0000	100.0000

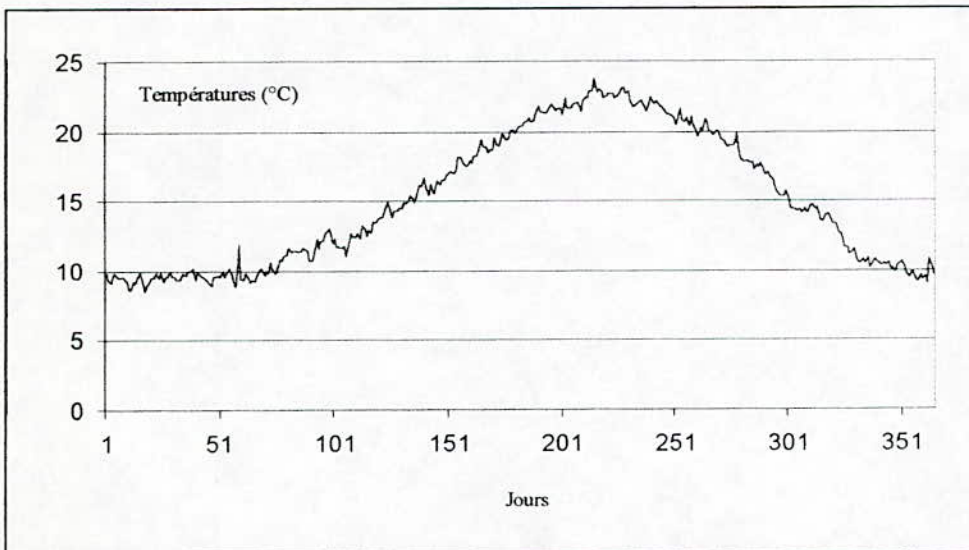


Fig. V-2 : Variation des températures moyennes journalières interannuelles

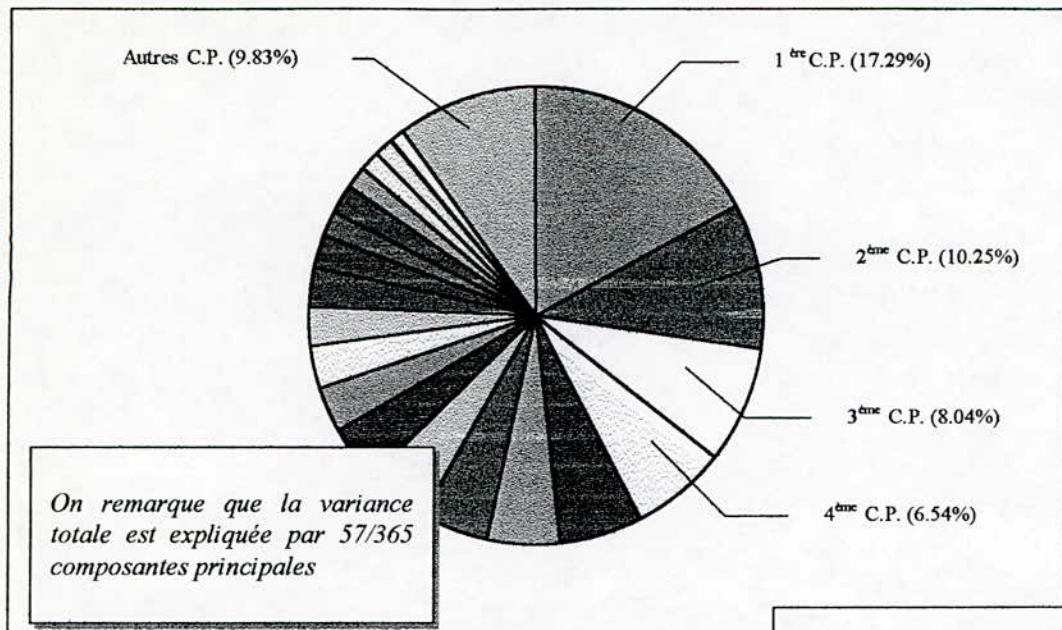


Fig. V-3 : Pourcentage de la variance apportée par chaque C.P.

Pourcentage retenu par 14 composantes principales qui représentent plus de 80%

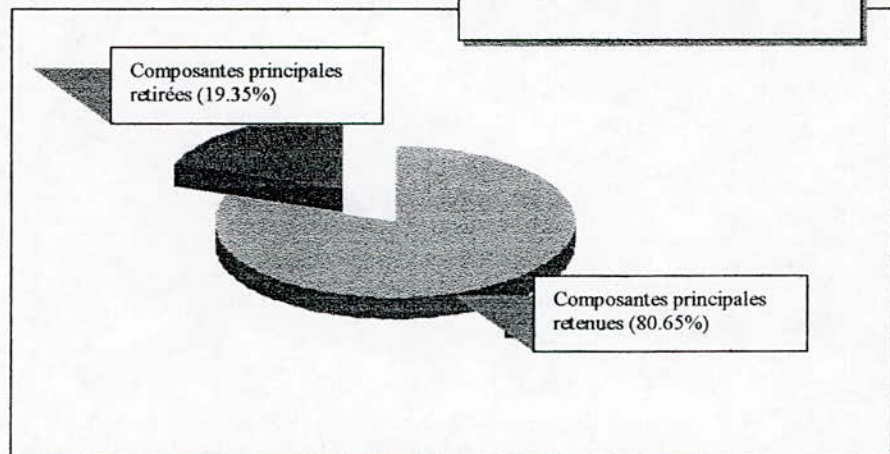


Fig. V-4 : Choix du nombre des composantes principales à retenir

V.3. Reconstitution.

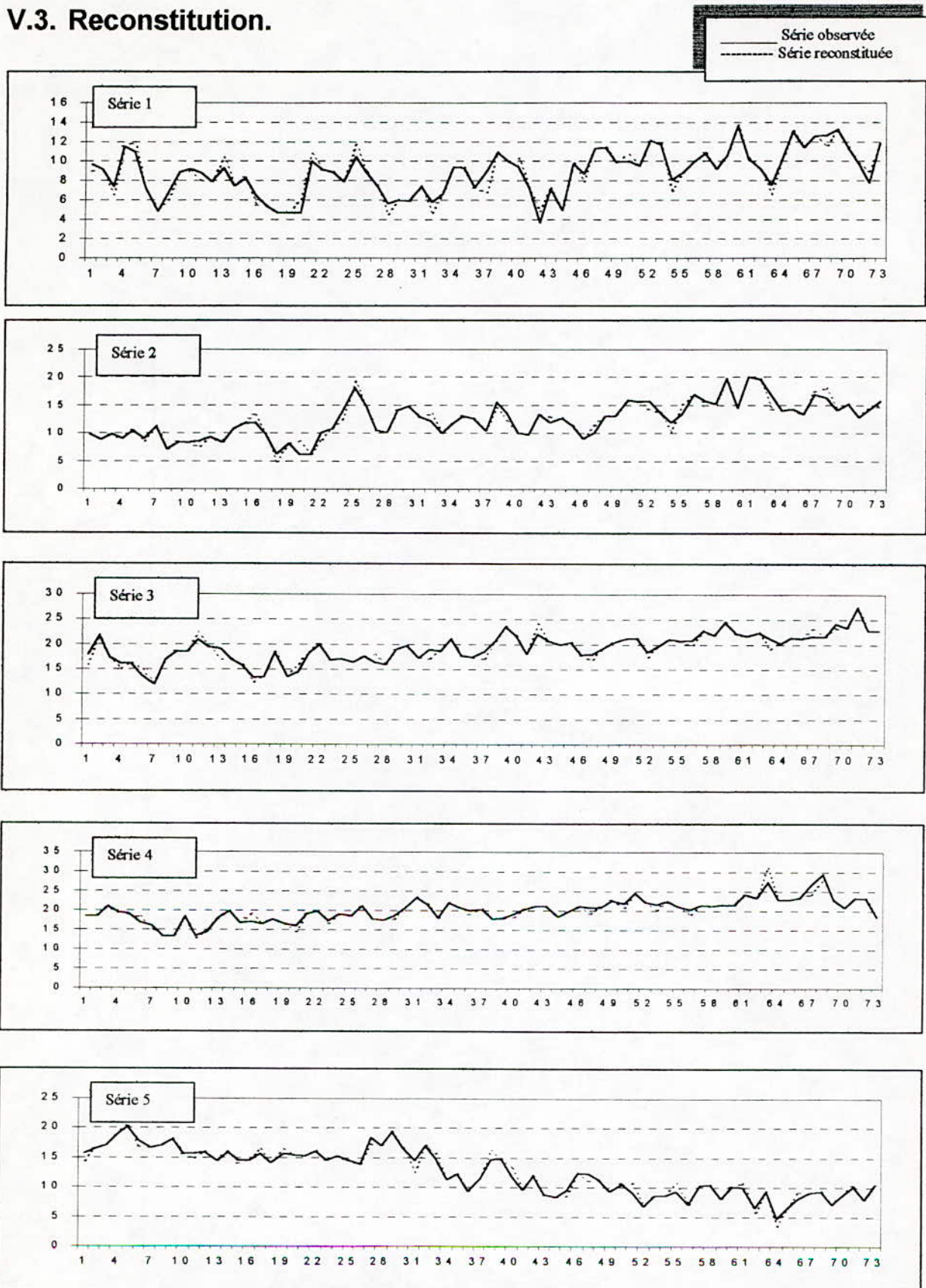


Fig. V-5 : Reconstitution des températures de l'année 1967 pour 14 C.P.

V.4. Ajustements.

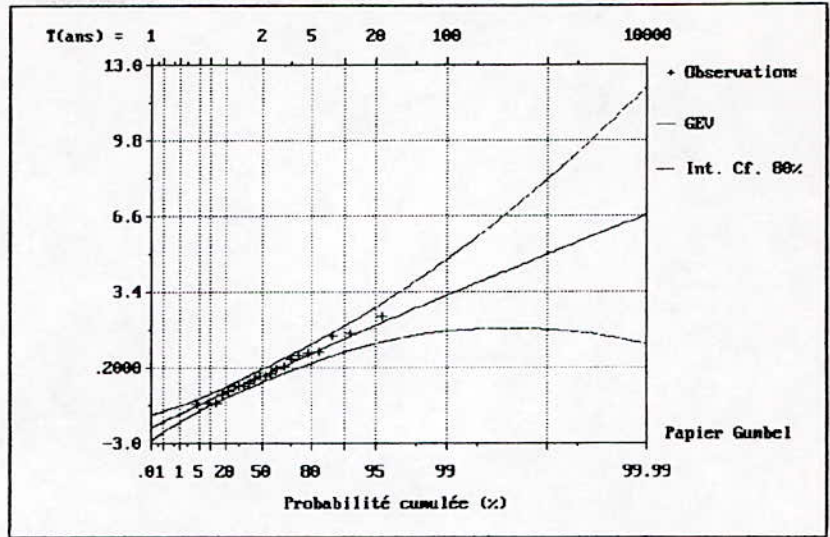
Puisqu'on a repris quatorze (14) composantes principales, on aura à simuler :

- 14 composantes principales centrées réduites.
- 365 vecteurs covariances.
- 1 vecteur des moyennes des variables.
- 1 matrice des résidus composée de 365 colonnes et 22 lignes.

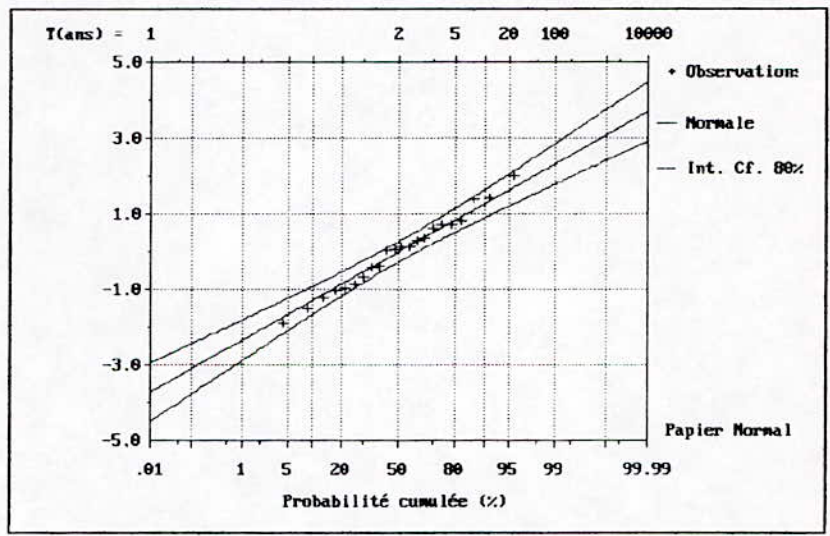
Chaque variable à simuler, est ajustée à une distribution donnée ou par le biais de sa fonction de répartition.

On simulera la matrice des résidus par le biais des chaînes de Markov.

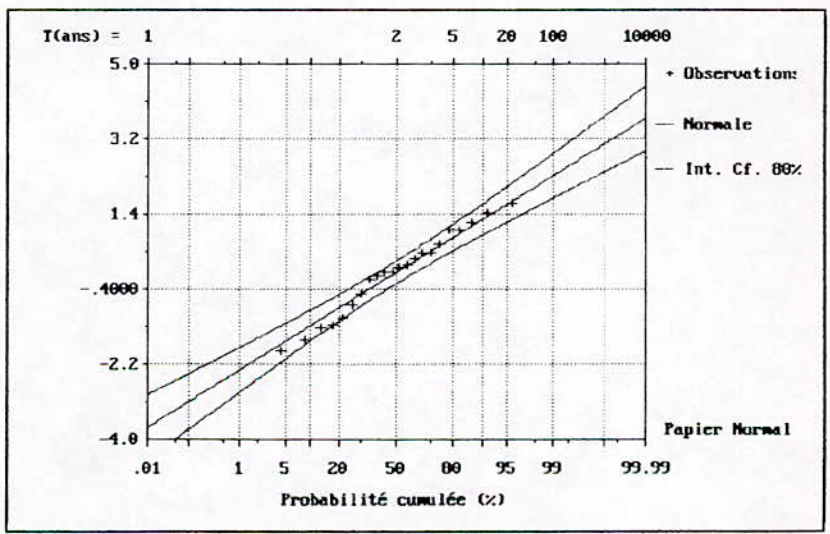
Ajustement : Loi G.E.V.
Variable : 1^{ère} C.P centrée réduite



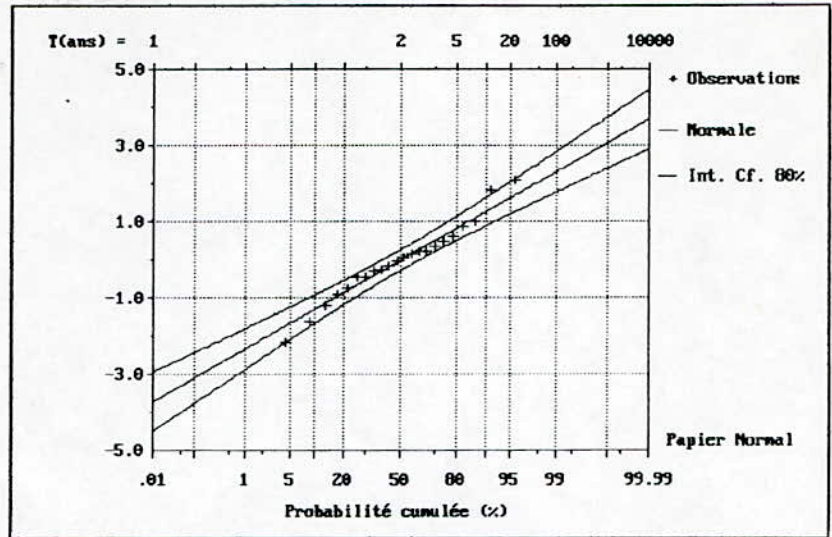
Ajustement : Loi normale
Variable : 2^{ème} C.P centrée réduite



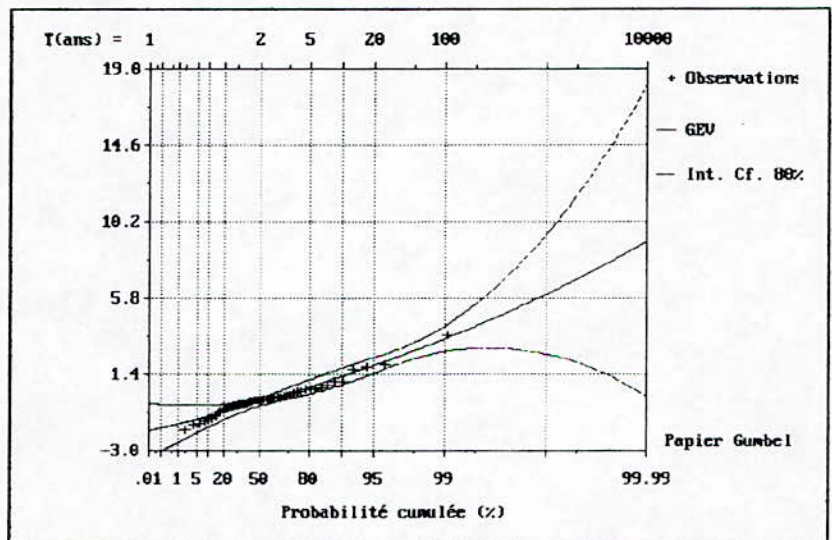
Ajustement : Loi normale
Variable : 3^{ème} C.P centrée réduite



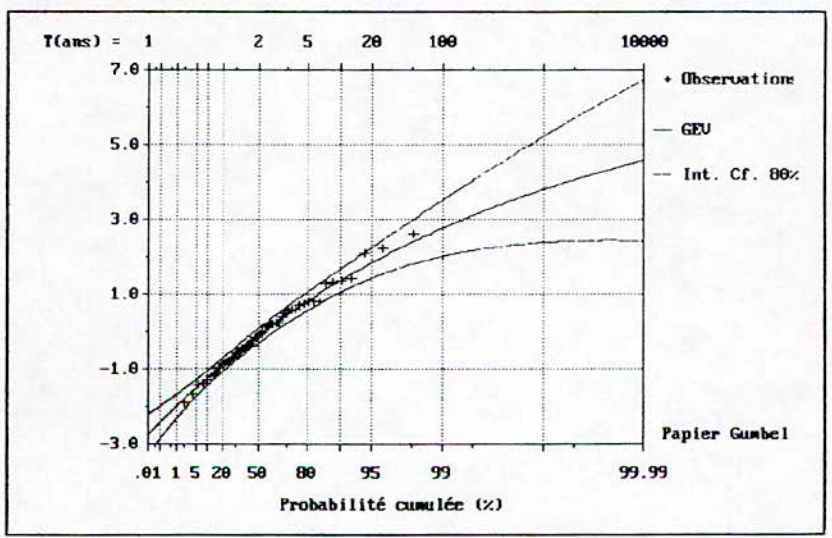
Ajustement : Loi G.E.V.
Variable : 4^{ème} C.P centrée réduite



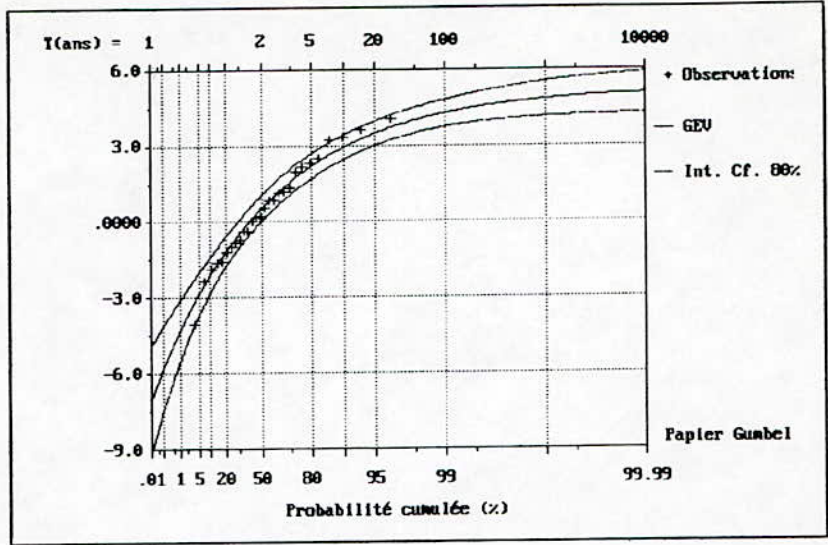
Ajustement : Loi G.E.V.
Variable : 5^{ème} C.P centrée réduite



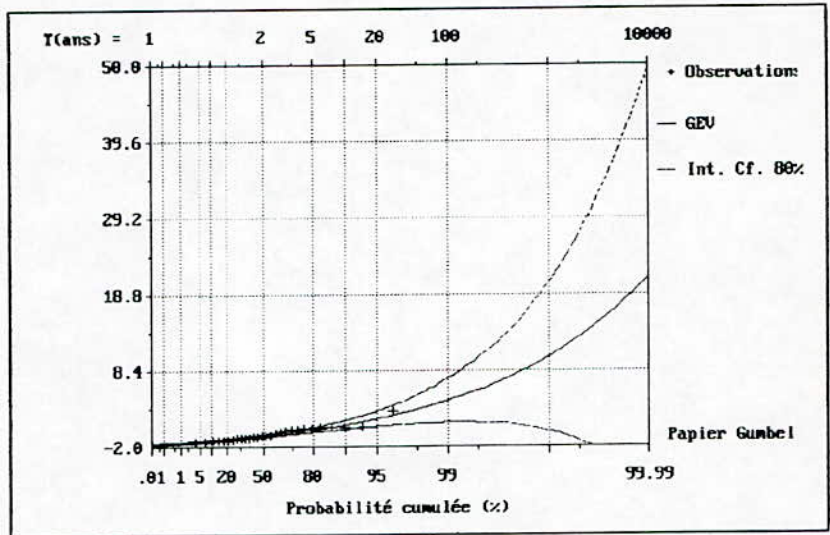
Ajustement : Loi G.E.V.
Variable : 6^{ème} C.P centrée réduite



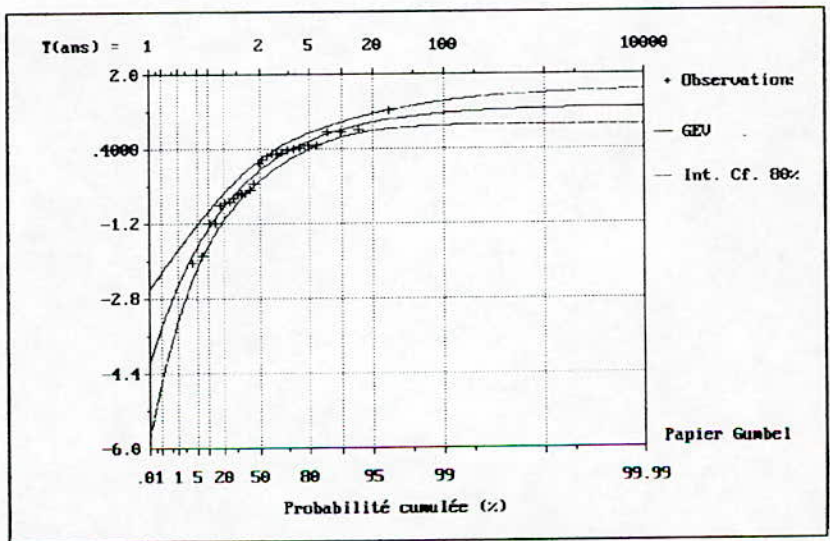
Ajustement : G.E.V.
Variable : 1^{ère} covariance



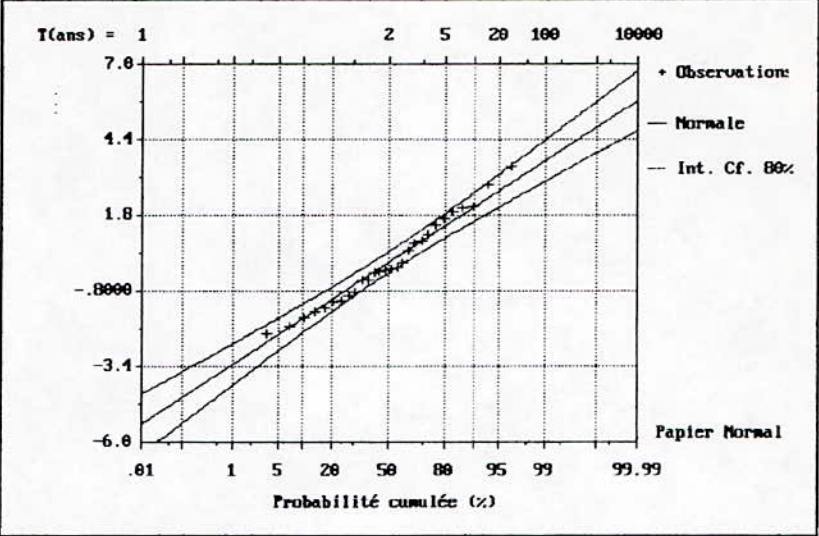
Ajustement : G.E.V.
Variable : 2^{ème} covariance



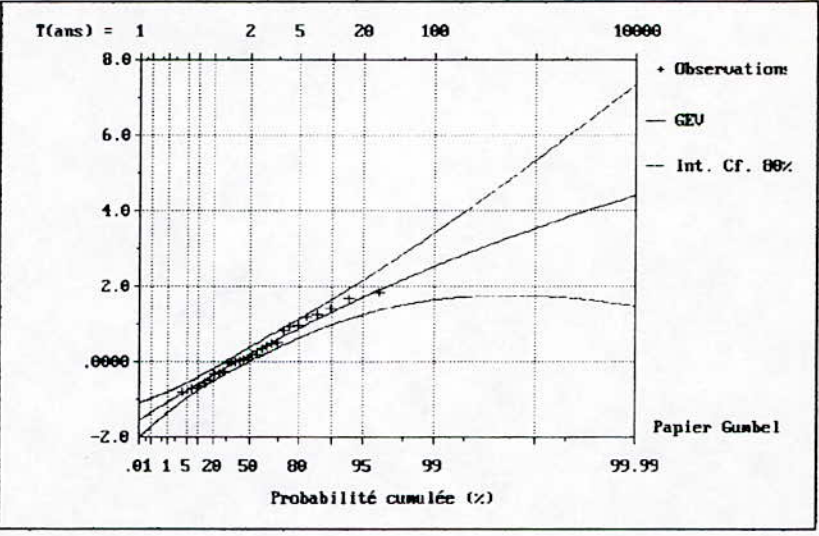
Ajustement : G.E.V.
Variable : 3^{ème} covariance



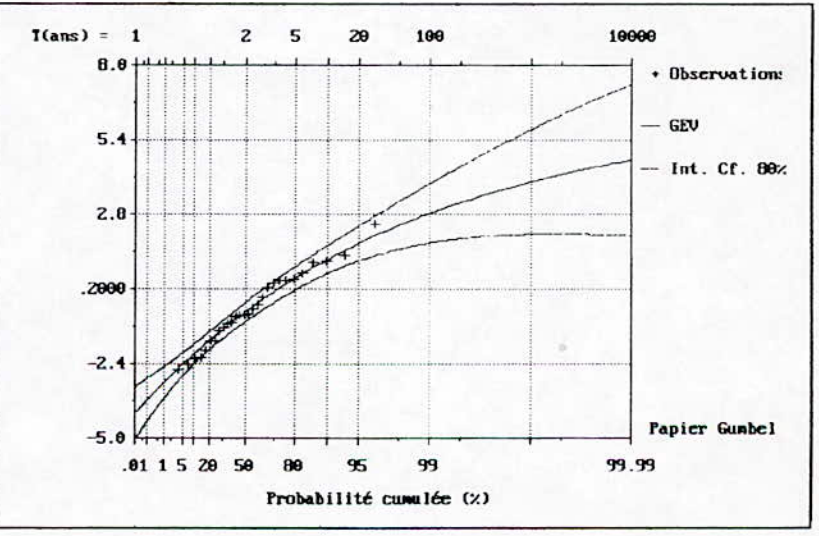
Ajustement : Loi normale
Variable : 4^{ème} covariance



Ajustement : G.E.V.
Variable : 5^{ème} covariance



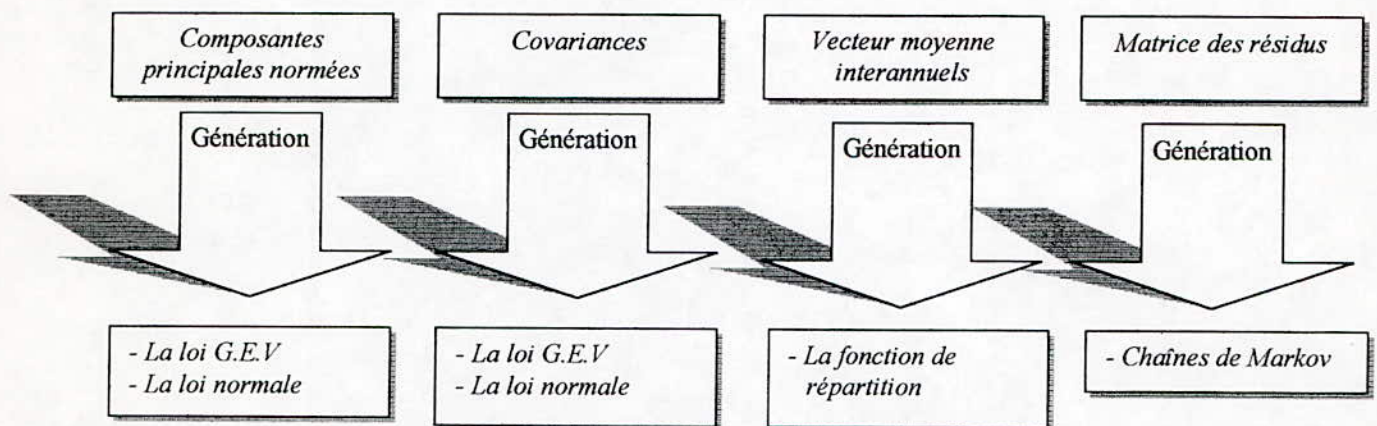
Ajustement : G.E.V.
Variable : 6^{ème} covariance



Ajustement : Fonction de repartition
Variable : Moyenne des variables

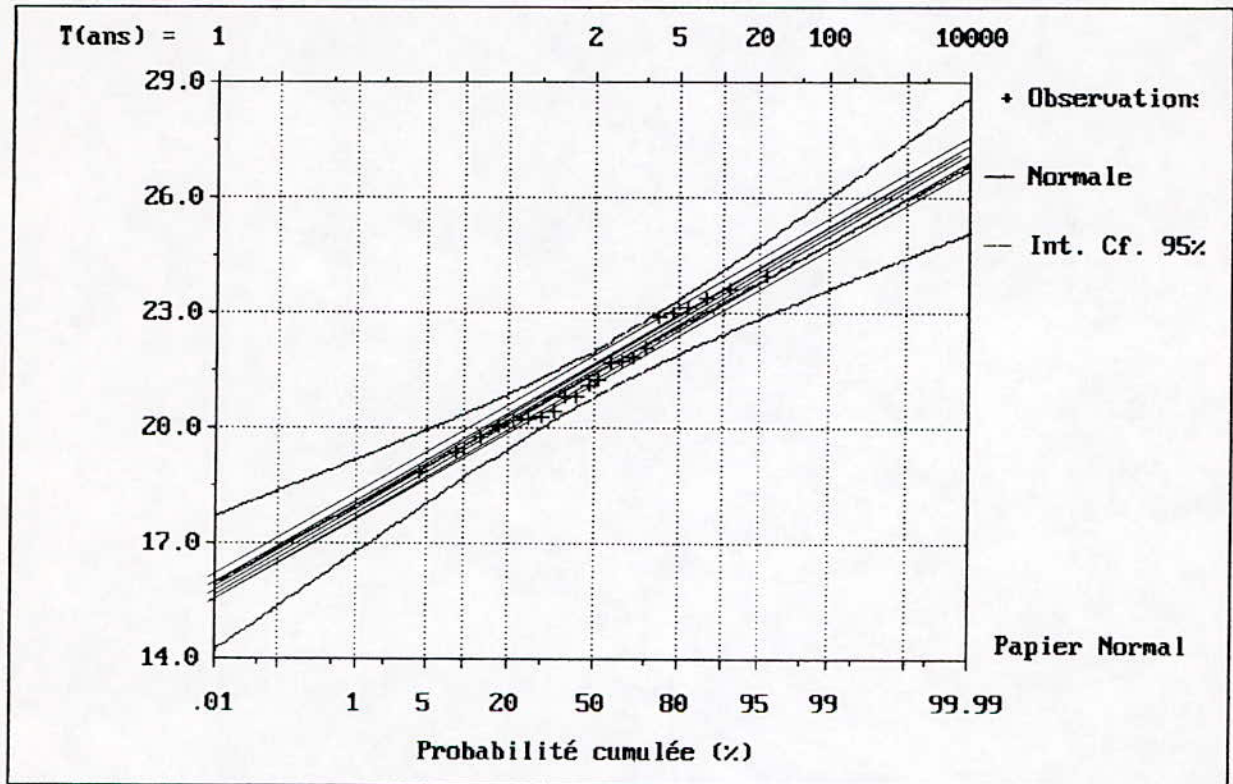
Le vecteur moyenne interannuel ne s'ajuste à aucune distribution statistique, ce qui oblige le passage par le biais de la fonction de répartition.

On a remarqué que tantôt les variables s'ajustent à loi G.E.V., et tantôt à la loi normale, par contre, on a utiliser la fonction de répartition pour le vecteur moyennes interannuels.



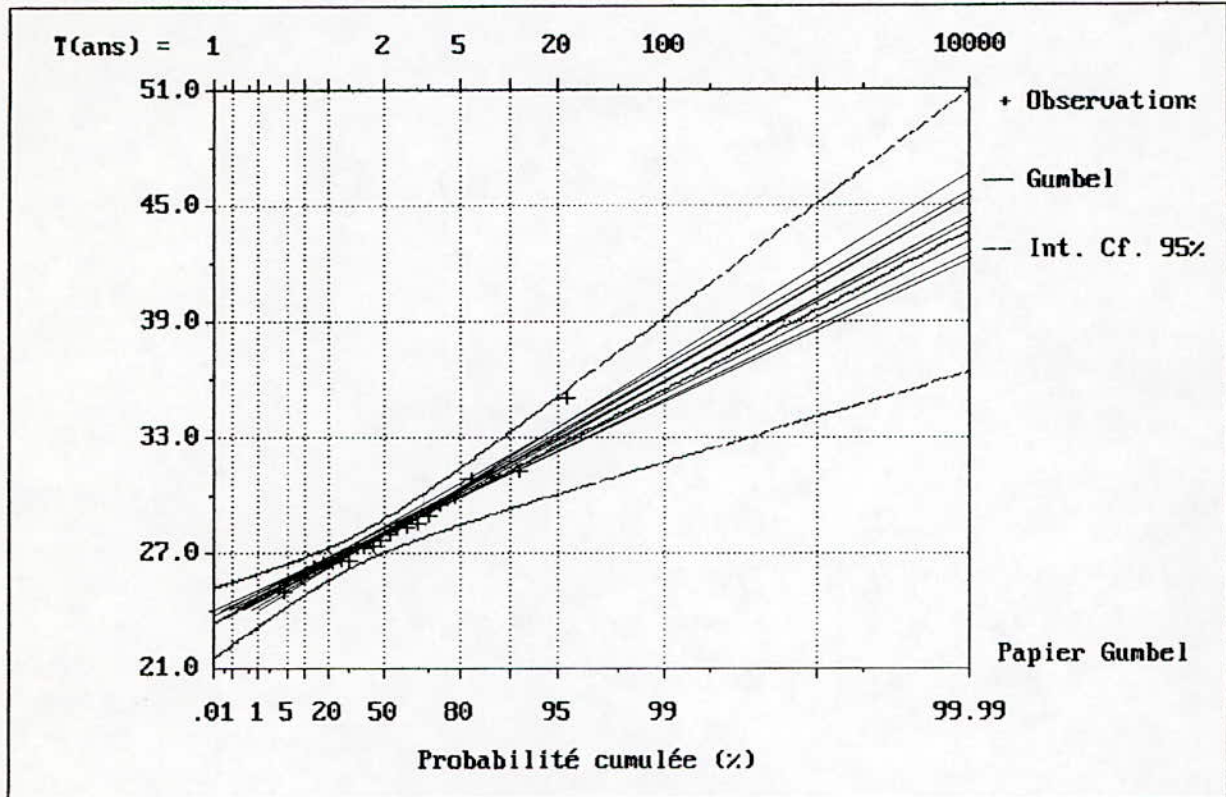
V.5.2. Ecart-types annuels.

Les écart-types annuelles de la série historique s'ajuste à la loi normale, de même pour les moyennes simulées, qui sont à l'intérieur de l'intervalle de confiance de la variable historique.



V.5.3. Maximums Annuels.

Les maximums annuels de la série historique s'ajuste à la loi Gumbel, de même pour les maximums simulées, qui sont à l'intérieur de l'intervalle de confiance des maximums historiques.



D'après ces résultats, on remarque que le modèle de simulation GESTOP est assez efficace au pas de temps journalier pour les phénomènes aléatoires et cycliques, n'empêche que phénomènes cycliques sont beaucoup plus aisés à manipuler au point de vue ajustement, car elles oscillent autour d'une certaine moyenne avec une faible variance relativement aux phénomènes aléatoires.

Conclusion générale et perspectives

L'étude développée dans ce mémoire, avait pour base les données hydrométéorologiques suivant des phénomènes aléatoires (pluviométrie) et cycliques (température).

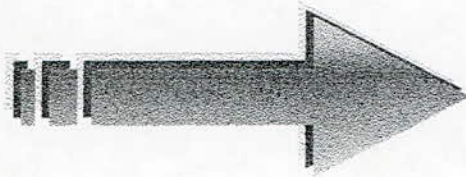
L'approche de la simulation de tels phénomènes devient très complexe en travaillant à des pas de temps très court.

Un modèle à plusieurs variantes a été développé, et a permis de faire la simulation de différentes variables (aléatoires et cycliques) au pas de temps journaliers, en éliminant les inconvénients majeurs des modèles existant, à savoir l'exigence de la stationnarité, de l'homogénéité et de la chronologie des séries étudiées.

La perspective à venir devrait s'intéresser à la simulation spatio-temporelle, pour résoudre le problème du manque ou de la mauvaise qualité des données hydrométéorologiques, ce problème se pose avec acuité en Algérie, du fait de la construction de nombreux barrages - réservoirs sur des sites ne disposant pas de séries de données (ou bien séries très courtes) nécessaire au dimensionnement de ces derniers.

ANNEXE I

Eléments Statistiques de base



Dans cette annexe, on présentera les éléments statistiques des séries chronologiques en hydrologie.

1. Moyenne.

Soit la série X_i , $i = 1, 2, \dots, N$
La moyenne de l'échantillon est donnée par

$$\bar{X} = 1/N \sum_{i=1}^N X_i \text{ tel que } N : \text{ est la taille de l'échantillon}$$

2. Variance

La variance d'un échantillon var est donnée par

$$\text{var} = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}_i)^2 \quad \text{si } N > 30,$$

$$\text{var} = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_i)^2 \quad \text{si } N \leq 30.$$

Et l'écart type est défini par la racine carrée de la variance

$$\sigma = \sqrt{\text{var}}$$

3. Covariance

La covariance est calculée pour deux variables

$$\text{Cov} = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}_i) (Y_i - \bar{Y}_i)$$

4. Coefficient d'autocorrélation

Le coefficient d'autocorrélation est donné par

$$\rho = \frac{\text{Cov}(X_i, Y_i)}{\sigma_x \sigma_y}$$

σ_x : écart type de La variable X_i

σ_y : écart type de La variable Y_i
généralement $-1 < \rho < 1$

5. Variable centrée

$$Z_i = X_i - \bar{X}_i$$

X_i : variable initiale

\bar{X}_i : moyenne de X_i

Z_i : variable transformée

C'est une translation de l'origine du repère au centre de gravité du nuage de points

6. Variable réduite

$$Z_i = X_i / \sigma_x$$

avec σ_x est l'écart type de la variable X_i

Cette transformation permet de standardiser les variables en considérant que leurs variations absolues.

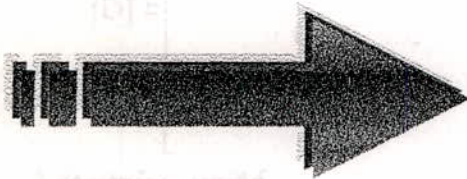
7. Variable centrée réduite

$$Z_i = \frac{X_i - \bar{X}_i}{\sigma_{xi}}$$

Cette dernière réunit les deux premières, permet d'avoir d'une part, l'origine du repère confondu avec le centre de gravité du nuage de points, et d'autre part, des Variables standards qui n'interviennent que par leurs variations absolues.

ANNEXE II

Calcul matriciel



Dans cette annexe, on évoquera quelques définitions en algèbre linéaire et les principales opérations sur les matrices.

1. Définitions

Soit une matrice A :

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$$

• Trace de $[A]$

$$\text{Tr}(A) = \sum_{i=1}^n a_{ii}$$

• Transposée de $[A]$

$$[A^t] = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{nn} \end{bmatrix}$$

et l'inverse de la matrice transposée c'est la matrice elle même.

$$(A^T)^{-1} = (A^{-1})^T = A$$

3. Formule de dérivation vectorielle

Définition d'une fonction quadratique

Soit x : un vecteur de p éléments

x^T : transposée de ce vecteur (vecteur ligne).

A : matrice de $p \times p$ éléments.

Q : est dite fonction quadratique si elle est écrite comme suit :

$Q = x^T A x$, A est appelée matrice de forme quadratique.

$$\frac{d(x^T A x)}{dx} = Ax + A^T x.$$

En particulier pour une matrice V symétrique (V c'est la matrice de covariance et elle est toujours symétrique), c'est à dire $V = V^T$

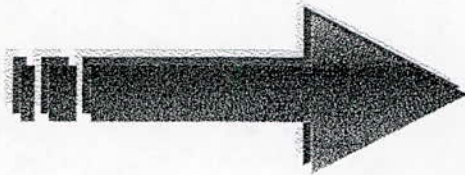
Donc

$$\frac{d(x^T V x)}{dx} = Vx + V^T x = 2Vx$$

$(V = V^T)$

ANNEXE III

Ecriture matricielle des éléments statistiques



Dans cette annexe, on évoquera quelques définitions en algèbre linéaire et les principales opérations sur les matrices.

Cas spécial : pour une variable

L'ensemble des valeurs prises par une variable X dans n observations $x_1, x_2, x_3, \dots, x_n$ est considéré comme un vecteur de n éléments.

$$X = (x_1, x_2, \dots, x_n)$$

• La somme des carrés de n valeurs de X s'écrit $C = X^2$, et sous forme matricielle s'écrit $C = X' X$

X : Vecteur colonne.

X' : Transposée du vecteur X (c'est un vecteur ligne)

• La somme des éléments de X s'écrit : $S = x_1 + x_2 + x_3 + \dots + x_n$, et sous forme matricielle s'écrit : $S = I' X = X' I$

tel que

I : Vecteur identité colonne.

I' : Transposée du vecteur I (vecteur ligne).

Cas général : pour N variables

Pour N variables la matrice d'observations est :

$$X_{(n,p)} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1p} \\ X_{21} & X_{22} & X_{23} & \dots & X_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & X_{n3} & \dots & X_{np} \end{pmatrix}$$

• Le vecteur somme est obtenu par la multiplication

$$X' I = S = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_p \end{pmatrix}$$

On peut écrire les éléments statistiques représentés dans l'annexe1, comme suit

• **Moyenne** $\bar{X}_j = \bar{X}_j = \frac{1}{n} \sum_{i=1}^N X_{ij} = \frac{1}{n} S = \frac{1}{n} I' X = X' I$

\bar{X}_j : moyenne de la $j^{\text{ème}}$ variable

• **Variance** $\text{var} = \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})^2 = \frac{1}{n} [(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2]$

$$= \frac{1}{n} [X_1^2 + X_2^2 + \dots + X_n^2 - 2\bar{X}(X_1 + X_2 + \dots + X_n) + n\bar{X}^2]$$

$$= \frac{1}{n} [X' X - 2 \frac{1}{n} I' X I' X + n(\frac{1}{n} I' X)^2]$$

$$= \frac{1}{n} [X' X - 2 \frac{1}{n} I' X I' X + \frac{1}{n} I' X I' X]$$

$$= \frac{1}{n} [X' X - \frac{1}{n} X' I \frac{1}{n} X' I I' X].$$

$$\text{Var} = \frac{1}{n} [X' X - \frac{1}{n} X' I I' X].$$

• **Ecart type** $\sigma = \sqrt{\text{var}} = \sqrt{1/n(X'X - 1/n(X'I I'X))}$

• **Covariance**
$$\begin{aligned} \text{Cov} &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = (X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + \\ &\dots + (X_n - \bar{X})(Y_n - \bar{Y}). \\ &= \frac{1}{n} [(X_1 Y_1 + X_2 Y_2 + \dots + X_n Y_n - \bar{X} (Y_1 + Y_2 + \dots + Y_n) + \\ &\bar{Y} (X_1 + X_2 + \dots + X_n) + \bar{X} \bar{Y})] = \frac{1}{n} [X'Y - \frac{1}{n} X'I Y'I \\ &- \frac{1}{n} X'I Y'I + \frac{1}{n} X'I Y'I] \\ &= \frac{1}{n} [(X'Y - \frac{1}{n} X'I Y'I)] \end{aligned}$$

$$\text{Cov} = \frac{1}{n} [(X'Y - \frac{1}{n} X'I Y'I)]$$

c'est la covariance pour deux variables

Pour N variables la matrice des covariances

$$V = \frac{1}{N} (X'X - \frac{1}{N} S S') = \begin{pmatrix} V_{11} & V_{12} & V_{13} & \dots & V_{1p} \\ V_{21} & V_{22} & V_{23} & \dots & V_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ V_{p1} & V_{p2} & V_{p3} & \dots & V_{pp} \end{pmatrix}$$

Les éléments diagonaux de la matrice de covariance constituent la matrice des variances

$$D = \begin{pmatrix} V_{11} & . & . & . & . & . \\ . & V_{22} & . & . & . & . \\ . & . & V_{33} & 0 & . & . \\ . & 0 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & V_{pp} \end{pmatrix}$$

$$\text{et on définit } D^{-1/2} = \begin{pmatrix} 1/\sqrt{v_{11}} & . & . & . & . & . \\ . & 1/\sqrt{v_{22}} & . & . & . & . \\ . & . & 1/\sqrt{v_{33}} & 0 & . & . \\ . & 0 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & 1/\sqrt{v_{pp}} \end{pmatrix}$$

• **Corrélations** : est une matrice symétrique et d'éléments diagonaux unité

$$R = D_v^{1/2} B' D_\lambda^{-1/2} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{pmatrix}$$

APPENDICE

Symboles et abréviations

A.C.P.	Analyse en composantes principales
σ	Ecart-type d'un échantillon
λ_p	Valeur propre
[B]	Matrice des vecteurs propres
[C]	Matrice des composantes principales
[C']	Matrice des composantes principales réduites
[COR]	Matrice des corrélations
[COV]	Matrice de covariance
[E]	Matrice des résidus
[V]	Matrice de covariances
[X]	Matrice d'observations
[X']	Transposée de la matrice d'observations
[X _p]	Vecteur colonne de la matrice d'observation
F(x)	Fonction de répartition
GEV	General extrêmes values
k	Rapport de forme
m	Moyenne d'un échantillon
n	Taille d'un échantillon
R	Corrélations
r	Rapport de forme
T	Période de retour
u	Paramètre de localisation
Var	Variance
X	Variable
y	Variable réduite de Gumble

Bibliographie

- [1] BENOIST A. (1974) - Notions sur les plans d'expériences. *Editions Technip*
- [2] CEHESSAT R. (1976) - Exercices commentés de statistiques et informatique appliquée. *Editions Dunod*
- [3] DECHEMI N., BERMAD A., HAMRICH A. (1994) - Simulation des débits moyens mensuels en zone semi-aride par l'analyse en composantes principales (ACP). *Hydrol. Continent.*, 9(1), pp. 17-24
- [4] FOUCART T. (1991) - Introduction aux tests statistiques. *Editions Technip*.
- [5] GILBERT N. (1983) - Statistiques. *Les éditions HRW Ltée, Montréal*.
- [6] HAMRICHE A., TACHET K. (1993) - Contribution à l'étude et à la simulation des paramètres hydrométéorologiques par l'analyse en composantes principales (ACP). *Thèse d'ingénieur, ENP Alger*.
- [7] KOTTEGODA N. T. (1980) - Stochastic water resources technology.
- [8] LAVOIE R. (1978) - Statistiques appliquées. *Presses de l'Université du Québec*.
- [9] LEBART M. (1975) - Statistiques et informatiques appliquées. *Editions Dunod*
- [10] LEFEBVRE J. P. (1983) - Introduction aux analyses statistiques multidimensionnelles. *Editions Masson*.
- [11] LIPSCHUTZ R. (1972) - Probabilités, cours et exercices. (*Série Schaum*) *Edition Française*.
- [12] MOULAI-HAMLAOUI L. (1995) - Etude des précipitations maximales journalières dans la région du centre algérien. *Thèse de magister, ENP Alger*
- [13] PAPOULIS M. (1984) Probability, Random variable and stochastic processus. *Student Edition*.
- [14] SAPORTA L. (1978) - Théorie et méthodes de la statistique. *Edition Technip*.
- [15] SOUAG D. (1993) - ARMA . *Thèse d'ingénieur, ENP Alger*