

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLICQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

8/94

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE - المكتبة
Ecole Nationale Polytechnique

وزارة التربية الوطنية

MINISTERE DE L'EDUCATION NATIONALE

المدرسة الوطنية المتعددة التقنيات

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT HYDRAULIQUE

PROJET DE FIN D'ETUDES

SUJET

CONTRIBUTION A LA SIMULATION DES DEBITS PAR LE
MODELE DE DESAGREGATION

PROPOSE PAR

BERMAD

ETUDIE PAR

N. CHENITI

DIRIGE PAR

BERMAD
D. SOUAG

PROMOTION

SEPTEMBRE 1994

E.N.P 10, AVENUE HACEN BADI EL-HARRACH ALGER

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

DEDICACES

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

A MA MERE

A MON PERE

A MES FRERES ET SOEURS

ملخص :

يهدف هذا العمل إلى التمثيل النموذجي
للتدفق الشهري، وذلك باستعمال التمثيل
التفكيكي ويختلف أشكاله ، ودراسة
قدرة هذا النموذج على تعثيل التدفق الشهري
لسلسلة بني بهدل .

Resumé :

Le but de cette étude est la simulation des débits mensuels par le modèle de désagrégation , sous ses trois formes; Et la verification de la capacite du modèle á reproduire les propriétés statistiques des données, et ce en utilisant la série du site de benibehdel. nous avons utilisé deux transfomations de données et deux types de désagrégation :

- 1 : annuel-mensuel.
- 2 : annuel-saisonnier , saisonnier-mensuel.

Abstract :

The object of the study is the simulation of mensual flow with disaggregation model under the the three type ; And the verification of the capacity of the model to reproduce the statistical characteristics of the historical data (Benibehdel serie) . we have used two transformation of the data and two type of disaggregation :

- 1 : annual-mensual .
- 2 : annual-saisonnier , saisonnier-mensual .

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

INTRODUCTION

INTRODUCTION

Il existe plusieurs formes et modèles de simulation en hydrologie , l' amélioration et le développement de ces derniers n'ont cessés jusqu'à ce jour; Bien que quelques uns sont appliqués avec succès , ces modèles ont des limites d'application vu la difficulté d' estimation des paramètres , le nombre important des paramètres et surtout vu le manque de la base physique.

Un des modèles utilisés pour la simulation des débits est le modèle de Désagrégation , appliqué avec succès par:

Valencia et Schake (1973) [11], pour:

- La génération des débits pour le bassin du Colorado en Argentine.

- La génération des données pluviométriques pour Puerto Rico.

- La simulation des demandes en eau pour Boston.

Curry et Bras (1976) [12], pour la génération des débits mensuels pour le Nile.

Tao et Delleur (1976) [10], pour la génération des données pour le Mississipi.

Ce modèle est alors l'objectif de notre étude qui consiste à vérifier la capacite de ce dernier , sous ses trois formes, à reproduire les caractéristiques statistiques des données historiques que ce soit du niveau annuel ou du niveau mensuel et ce pour la série des débits mensuels du site de Benibehdel.

Nous présentons dans ce travail trois grands chapitres:

- Nous donnons un aperçus général sur la simulation, puis nous présentons quelques modèles utilisés en hydrologie et quelques techniques de base de la simulation.

- Nous présentons ensuite le modèle de désagrégation sous ses différentes formes .

- Enfin nous exposons les résultats obtenus lors de la mise en oeuvre des programmes.

CHAPITRE UN

GENERALITES ET DEFINITIONS

I.1 - GENERALITES

I.1.1 - Définition de la simulation

La simulation est une technique qui a été largement utilisée lors des dernières décennies dans plusieurs domaines de la recherche scientifique, économique et notamment dans les systèmes de ressources hydrauliques.

Cette large utilisation est due à sa simplicité et à sa souplesse mathématique, Et surtout au fait qu'elle repose en grande partie sur des méthodes numériques qui ont connues de grandes évolutions.

On définit la simulation comme étant l'action de reproduire le comportement d'un certain système ou phénomène tout en gardant ses principales caractéristiques.

Les difficultés de cette méthode résident dans la partie numérique et surtout dans les cas multidimensionnels.

Enfin nous pouvons distinguer deux types de simulation:

- # La simulation déterministe
- # La simulation stochastique

Une simulation est dite stochastique si le système fait appel à des phénomènes aléatoires. Dans le cas contraire elle est dite déterministe.

I.1.2 - Génération des débits

Les débits générés sont appelés débits synthétiques ou débits opérationnels pour les distinguer des données historiques, ces débits sont utilisés dans la gestion des réservoirs. Leur distribution est dite distribution synthétique.

La génération des débits exige un processus stochastique stationnaire. Dans ces conditions le modèle Stochastique peut générer des séquences synthétiques qui reproduisent les caractéristiques voulues des débits historiques.

Notons que la stationnarité d'un processus n'est pas toujours vraie, car un changement des caractéristiques physiques peut engendrer un changements des caractéristiques du modèle.

I.1.3 - Modèles statistiques de génération des débits

La première étape dans la construction d'un modèle est d'extraire, à partir de données historiques, certaines caractéristiques statistiques et lois de distributions dans le temps et dans l'espace. Ce modèle à construire devra satisfaire la ressemblance statistique. C'est à dire que ce modèle doit fidèlement reproduire les caractéristiques fondamentales de la distribution (la moyenne, la variance, le coefficient d'asymétrie et l'autocorrelation) Cette qualité rend ces modèles attractifs.

La deuxième étape est de choisir les caractéristiques dont on aura besoin dans la modélisation. Cette décision dépend de la destination des modèles, de la disponibilité et validité des données et du temps nécessaire pour la construction du modèle stochastique.

I.1.4 - Application simple de la simulation

Le premier modèle stochastique qui vient à l'esprit est le modèle Autorégressif, dit aussi modèle MARKOVIEN, ce modèle peut représenter une simple application de la simulation. Sa forme qui est très simple permet d'illustrer le fondamental des plus compliqués des modèles.

La première application de ce modèle a été faite par THOMAS et FIERING en 1962. Jusqu' à ce jour le modèle et ses applications n'ont cessés de se développer.

Principe du modèle autorégressif

On suppose que les débits suivent un processus Markovien, c'est à dire que chaque valeur du processus dépend de la valeur qui la précède dans le temps et que ces valeurs suivent une loi normale.

Soit une distribution normale de deux débits Y_t et $Y_{t'}$ correspondant respectivement à l'an t et à l'an t' , de moyenne μ , de variance σ^2 et de corrélation ρ .

La fonction de distribution est donnée par:

$$f_{Y_{t'}, Y_t}(y_{t'}, y_t) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[\frac{(y_t - \mu)^2 - 2\rho(y_t - \mu)(y_{t'} - \mu) - (y_{t'} - \mu)^2}{2\sigma^2(1-\rho^2)} \right]$$

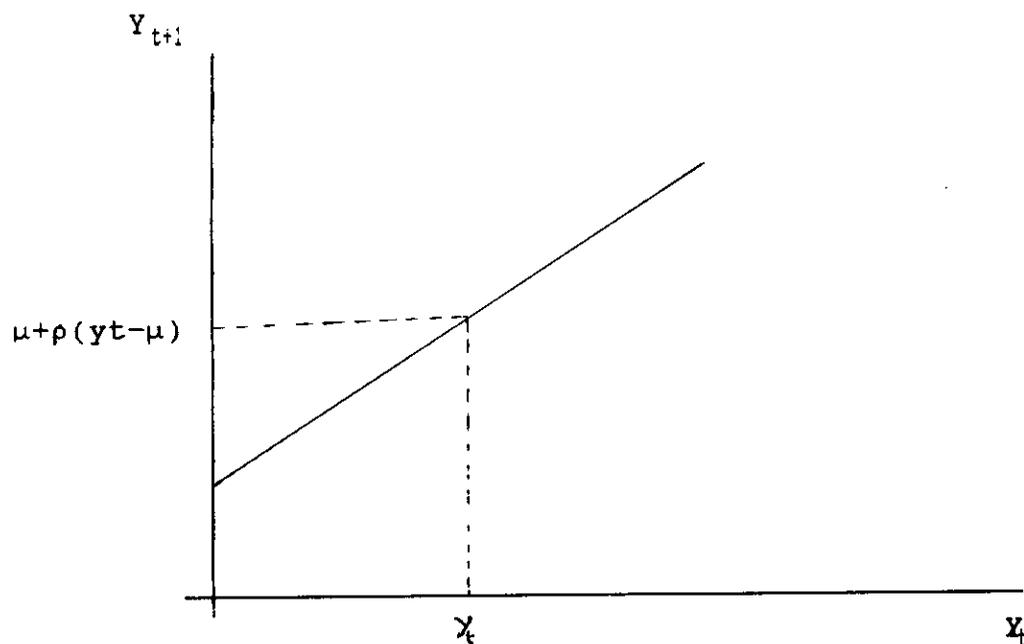
La distribution des deux débits dépend seulement de leurs moyenne μ , de la variance σ^2 et de leur corrélation ρ .

Si le processus Markovien est à pas -1, le débit de l'an $t+1$ dépend du débit de l'an t qui est supposé égal à la valeur historique y_t .

Avec cette condition sur la distribution ($Y_t = y_t$), le débit Y_{t+1} suit la loi de distribution de moyenne et de variance suivantes:

$$E[Y_{t+1}/Y_t=y_t] = \mu + \rho(y_t - \mu)$$

$$\text{Var}(Y_{t+1}/Y_t=y_t) = \sigma^2(1-\rho^2)$$



— Distribution conditionnelle de Y_{t+1} ($Y_t = y_t$)
 — pour une loi normale à deux variables —

La génération des débits normalement distribués de moyenne μ , de variance σ^2 et de corrélation ρ est produite par le modèle suivant:

$$Y_{t+1} = \mu + \rho(Y_t - \mu) + \varepsilon_t \sigma \sqrt{1 - \rho^2}$$

où ε_t est une variable aléatoire normalement distribuée de variance unité et de moyenne nulle ($\sigma^2 = 1$, $\mu = 0$)

De plus ε_t est indépendante des débits $Y_{t'}$ ($t' \leq t$) et indépendante de $\varepsilon_{t'}$ ($t' \neq t$). Ce qui se traduit par :

$$E[e_t, e_{t'}] = 0 \quad t' \neq t$$

$$E[(Y_t - \mu), e_t] = 0 \quad t' \leq t$$

Il est bien évident que Y suit une distribution normale si Y et ε_t sont normalement distribués. Les caractéristiques produites par le modèle sont aussi préservées.

Le modèle Markovien (pas nécessairement à pas -1) peut ne pas être capable de reproduire ou de générer des données , mais il permet d'analyser le comportement des cours d'eau . A ce stade là de l'étude, l'analyste peut donner une décision et trancher pour tel ou tel modèle .

I.1.5 - La modélisation en Hydrologie

Les premières études ont été faites par HAZEN en 1914 et SUDLER en 1927 . Ces études ont montrés la possibilité d'utiliser la théorie des probabilités et des statistiques dans l'analyse des séquences des débits fluviaux .

BARNES (1954) a étendu les premières études de HAZEN et SUDLER et a introduit l'idée de la génération synthétique des débits en utilisant les tables de loi normale. Cependant , le développement formel de la simulation n'a commencé que vers le début des années soixante avec l'introduction des modèles autorégressifs (THOMAS - FIERING 1962 , YEVJEVICH 1963).

Depuis, de grands efforts ont été consacrés dans le but :

- D'améliorer et de développer les premiers modèles.
- De donner les justifications physiques des modèles.
- D'introduire de différentes formes alternatives et d'étudier leur impact sur le système des ressources hydrauliques .

Plusieurs modèles stochastiques ont été proposés :

- # Modèle autorégressif (AR): THOMAS-FIERING 1962, YEVJEVICH 1963, MATALAS 1967
- # Modèle bruit gaussien fractionnaire (FGN):
MANDELBORT-WALLIS 1968, MATALAS-WALLIS 1971
- # Modèle ARMA: CARLSON et.al 1970, O'CONNELL 1971
- # Modèle de la ligne brisée (BL): MEDJIA 1971
- # Modèle shot noise: WEISS 1973
- # Processus intermittent: YAKOWITZ 1973, KELMAN 1977
- # Désagrégation: VALENCIA et SCHAAKE 1973
- # Markov mixture: JACKSON 1975
- # ARMA-Markov: LETTENMAIER-BURGES 1977
- # General mixture models: BOES-SALAS 1978

Tout ces modèles ont été développés dans le but de reproduire les caractéristiques statistiques principales observées ou identifiées dans les séries chronologiques , et ce, selon le système d'étude (objectif de l'étude).

Bien que chaque modèle ait ses propres mérites, et bien que quelques uns peuvent être appliqués avec succès, ces modèles ont cependant des limites d'application ; En effet ils présentent plusieurs inconvénients, à savoir :

- Difficulté dans l'estimation des paramètres
- Nombre important de paramètres
- Manque de bases physiques

Les modèles les plus utilisés dans la pratique pour l'estimation et la désagrégation sont les modèles AR et ARMA.

Les étapes de la modélisation

Les modèles proposés sont seulement des approximations , la forme mathématique exacte du modèle n'est jamais connue et les valeurs des paramètres ne peuvent qu'être estimés.

Plusieurs contraintes interviennent lors du choix du modèle pouvant reproduire avec le plus d'exactitude les caractéristiques statistiques historiques.

Un schéma simplifié (page) composé de six (06) phases donné par SALAS et SMITH 1980 [] montre le chemin suivi dans la modélisation des séries chronologiques en hydrologie

PHASE UNE : Dans cette phase on décide si le modèle doit être :

- univariable ou multivariable
- Une combinaison d'un modèle univariable et d'un modèle de désagrégation.

- Une combinaison d'un modèle multivariable et d'un modèle de désagrégation.

PHASE DEUX : Consiste à choisir un des modèles disponible en hydrologie stochastique .

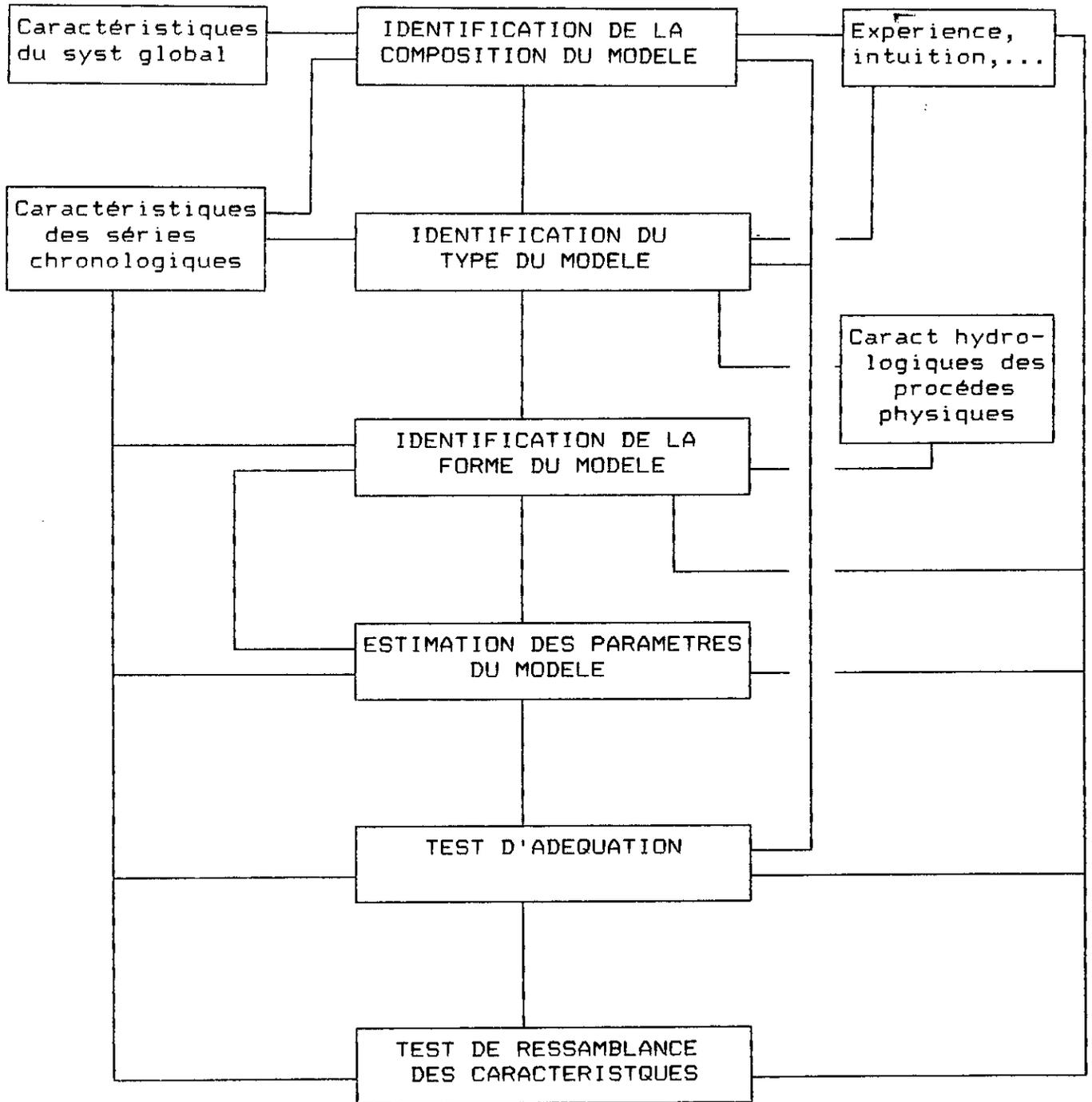
PHASE TROIS : On identifie l' ordre du modèle. Pour un modèle autorégressifs par exemple on détermine le nombre des paramètres autorégressifs p .

PHASE QUATRE : Une fois le nombre de paramètres identifié on estime alors ces derniers par une des méthodes présentées au chapitre (14).

PHASE CINQ : Dans cette étape on s' assure que les séries générées du modèle vérifient bien les tests d' adéquation.

PHASE SIX : Cette dernière étape désigne les tests de comparaison des caractéristiques statistiques des séries générées et des caractéristiques des séries de départ . Dans le cas du rejet des tests la procédure de modélisation est répétée jusqu'à la satisfaction.

ORGANIGRAMME



I.2 - MODELES DE SIMULATION

I.2.1 - Modèles autorégressifs

Déjà cités dans les sections précédentes ils sont de la forme suivante:

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

$$V_t = e_t \sigma_e$$

$$\sigma_v = \sigma^2 (1 - \sum_{j=1}^p \phi_j r_j)$$

où:

- μ : moyenne des données historiques
- Y_t : variable synthétique
- ϕ : coefficients autorégressifs
- e_t : variable aléatoire centrée réduite
- p : ordre du modèle
- r_j : coefficients d'autocorrélation

En multipliant les deux membres de l'équation par Y_{t+k} et en prenant l'espérance des deux cotés, on obtient une relation de récurrence entre les coefficients d'autocorrélation r_k :

$$r_k = \sum_{j=1}^p \phi_j r_{k-j}$$

En portant sur un graphe, et ce pour chaque ordre p les valeurs des couples (r_k, k) , on obtient ce qu'on appelle le corrélogramme spécifique du modèle.

Ainsi construits, ces corrélogrammes nous permettent de trancher sur l'ordre du modèle autorégressif représentant les données historiques.

Les coefficients autorégressifs sont estimés à partir de l'équation (4) et cela pour chaque ordre du modèle .

Les coefficients autorégressifs ϕ_j et d'autocorrélations r_k ont donc cette seconde utilité qui permet de choisir un modèle capable de générer des séquences synthétiques.

Prenons l'exemple d'un modèle d'ordre 1 :

$$r_k = \phi_1 r_{k-1}$$

D'où

$$\phi_1 = r_1$$

Dans le cas d'un modèle d'ordre 2 nous avons:

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2}$$

$$r_1 = \phi_1 r_0 + \phi_2 r_{-1}$$

$$r_2 = \phi_1 r_1 + \phi_2 r_0$$

Sachant que : $r_{-1} = r_1$

(par symétrie)

$$r_0 = 1$$

On obtient :

$$\phi_1 = \frac{r_2 - 1}{r_1^2 - 1} r_1$$

$$\phi_2 = \frac{r_1^2 - r_2}{r_1^2 - 1}$$

I.2.2 - Modèles autorégressifs à moyenne mobile (ARMA)

Durant les périodes sèches, les débits proviennent principalement des eaux souterraines. Cette contribution peut être représentée par un modèle autorégressif en ajoutant à cette valeur les apports aléatoires des pluies qui sont représentées par un modèle de moyenne mobile (MA) on arrive à présenter les débits de rivières.

Cette combinaison donne lieu au processus mixte autorégressif d'ordre p et de moyenne mobile d'ordre q ARMA (p,q), [autrefois appelés modèles de BOX-JENKINS] .

Les modèles ARMA sont construits avec un minimum de paramètres à estimer et ce, comparés aux modèles autorégressifs de grands ordres.

La formulation mathématique du modèle ARMA est la suivante

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + V_t - \theta_1 V_{t-1} - \theta_2 V_{t-2} - \dots - \theta_q V_{t-q}$$

Z_t sont des variables normalement distribuées de moyennes nulles

$\phi_{j/j=1,p}$: coefficients autorégressifs

$\theta_{j/j=1,q}$: coefficients de moyenne mobile

Les paramètres du modèle sont estimés à partir des données historiques . Ces paramètres sont :

La moyenne μ , la variance σ_{V_t} de la variable indépendante V_t et les coefficients ϕ_i et θ_j .
Nous avons donc au total $p+q+2$ paramètres à estimer.

I.2.3 - Modèles multidimensionnels

I.2.3.1 - Modèles multisites

Les modèles présentés dans les parties précédentes sont des modèles de génération des débits annuels à site unique .

Dans le cas d'une modélisation à plusieurs sites , la simulation devient très complexe , et cette complexité dépend des données (disponibilité et validité) et de la relation existant entre les débits des différents sites.

Le plus simple des modèles est obtenu en supposant que les débits Q_t^i au site i ($i = 2, N$) ne sont qu'une fraction du débit total Q_t^1 .

$$Q_t^i = \alpha_i Q_t^1 \quad i=2, N$$

où α_i paramètre qui dépend des caractéristiques du bassin fluvial i

Les débits Q_t^1 sont générés par un modèle représentatif qui satisfait à toutes les conditions citées dans les parties précédentes.

Une fois que les débits Q_t^1 sont générés , les débits Q_t^i sont synthétisés par l'équation ()

a - Modèle général

Il est parfois nécessaire de générer des débits synthétiques pour plusieurs sites. L'idéal est de pouvoir utiliser un modèle multidimensionnel .

Ces modèles ne sont qu'une extension des modèles autorégressifs et des modèles ARMA.

Leur forme générale est donnée par:

$$Z_{t+1} = A Z_t + B \epsilon_t$$

Z_t : Matrice colonne des débits annuels transformés , de moyenne nulle pour l'année y (de dim $N \times 1$)

ϵ_t : Matrice colonne des variables aléatoires normales standardisées

A,B: Paramètres du modèle (de dim $N \times N$)

Une autre forme des modèles multidimensionnels est présentée dans la seconde partie qui traite du modèle de désagrégation.

En multipliant l'équation (2) par la transposée de la matrice Z_t puis en prenant les valeurs espérées on obtient:

$$A = E[Z_{t+1} Z_t^T] / E[Z_t Z_t^T]$$

ou

$$A = S_1 / S_0$$

S_1 : Matrice des covariances et cross-covariances à pas -1

S_0 : Matrice des covariances et cross-covariances à pas 0

b - Modèles multisaïsons multisites

Il est parfois exigé dans l'étude d'un système de connaître la variation des débits pour chaque année. Cela nous permet par exemple , de détecter les périodes sèches et les périodes humides.

Il existe deux approches pour la génération des débits et la détermination de leurs variation annuelle .

La première est la désagrégation des débits annuels et ce afin de produire des débits saisonniers. Cette méthode permet la reproduction des caractéristiques annuelles et saisonnières des débits.

La seconde approche nous permet de générer des débits saisonniers en utilisant une des méthodes séquentielles citées précédemment .

I.3 - TRAITEMENT DES DONNEES

Les possibilités d'utilisation d'un modèle mathématique ainsi que les limites de sa validation sont dictées par la quantité et la qualité des données .

De plus les modèles stochastiques en hydrologie imposent des conditions d'utilisation qui se basent sur des hypothèses de la statistique . A cette fin les données doivent subir plusieurs tests pour qu'elles soient statistiquement acceptables à la modélisation.

Quatre tests sont utilisés.

- # Test d'homogénéité des données
- # Test d'indépendance
- # Test de normalité
- # Test de stationnarité

I.3.1 - Qualité des données

Le test d'homogénéité vérifie la cohérence interne des valeurs de l'échantillon. Deux propriétés indésirables dans les séries chronologiques en hydrologie, la non homogénéité des données, qui est un phénomène très courant en hydrologie, due à des effets humains ou à des facteurs perturbateurs naturels, et l'inconsistance qui est produite par des erreurs systématiques. Ces deux propriétés sont les principaux responsables des tendances (Trends) ou des changements soudains (Jumps, slippages).

Pour déterminer proprement les caractéristiques des séries, l'inconsistance et la non homogénéité des séries doivent être identifiées et supprimées. La tendance et les cycles doivent être testés pour qu'ils soient statistiquement significatifs et physiquement (ou historiquement) justifiés.

L'identification, la description et la suppression de la non homogénéité et de l'inconsistance sont donc des aspects importants dans l'analyse des séries chronologiques. Il est plus fiable si ces changements sont justifiés par des tests statistiques et des justifications physiques ou historiques.

Ces caractéristiques sont identifiées soit par :

La fonction de la tendance dont ses paramètres doivent être sensiblement différent de Zéro. La fonction de la tendance est de la forme :

$$x_t = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n$$

où b_0, b_1, \dots, b_n : paramètres à estimer

Soit par :

Un test effectué sur les caractéristiques statistiques de base des sous-séries des séries historiques qui doivent être statistiquement différentes entre elles. Il existe pour cela plusieurs techniques de comparaison des caractéristiques statistiques des sous-séries telle que celle de Student que nous rappelons.

On décompose la série en deux sous-séries, le test de Student sur la moyenne est décrit par :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s \sqrt{\frac{(N_1 + N_2)}{(N_1 N_2)}}}$$

avec

$$s = \sqrt{\frac{\sum_{i=1}^{N_1} (x_i - \bar{X}_1)^2 + \sum_{j=1}^{N_2} (x_j - \bar{X}_2)^2}{N_1 + N_2 - 2}}$$

La variable t suit la distribution de Student. t doit être inférieur à la variable critique t_c tirée à partir des tables de Student (95%).

I.3.2 - Test d'indépendance

La dépendance des séries chronologiques est représenté par la fonction r_k ou par la fonction spectre.

Pour des séries indépendantes la population des corrélogrammes est égale a zéro pour tout $k > 0$.

Le test d'indépendance repose donc sur le principe que la fonction r_k fluctue autour d'une valeur nulle. ANDERSON a pu déterminer les limites des séries indépendantes par des corrélogrammes pour des niveaux de confiance de 95% et de 99% suivant la loi Normale.

$$r_k(95\%) = \frac{-1 \pm 1.960\sqrt{(N-K-1)}}{N-K}$$

$$r_k(99\%) = \frac{-1 \pm 2.326\sqrt{(N-K-1)}}{N-K}$$

I.3.3 - Test de Normalité

Dans l'analyse des séries chronologiques en hydrologie, les variables sont supposées normalement distribuées. Il faudra donc normaliser les variables qui présentent une dissymétrie dans leur distribution.

La technique consiste simplement à passer d'une variable à une autre par une certaine fonction. Tout le problème réside dans le choix de cette fonction.

Si les variables de départ suivent par exemple la loi Log-Normale 2 on adopte la transformation logarithmique.

$$Y = \ln(X)$$

ou bien la transformation Log-translaté.

$$Y = \ln(X-c)$$

Si par contre la fonction de distribution des variables est la loi Gamma, la fonction adoptée dans ce cas sera:

$$y = \sqrt{x}$$

ou la transformation en puissance:

$$Y = a(X-c)^b$$

b étant de l'ordre de 1/2, 1/3, 1/4

Le principal avantage de l'utilisation de cette approche qu'est la normalisation est que les meilleurs techniques de la statistique et des processus stochastiques sont développées pour des processus qui suivent des lois normales. D'autre part la préservation des caractéristiques après transformation par le modèle n'est pas assurée pour les séries originales (avant transformation).

Un test de normalité est présenté au chapitre ()

- Test d'asymétrie

I.4 - ESTIMATION DES PARAMETRES

L'estimation des paramètres est l'un des principes de base et est une des techniques nécessaires pour la modélisation stochastique des séries chronologiques .

Considérons un échantillon de N variables X_1, \dots, X_N et un modèle stochastique de paramètres α et β . Les valeurs estimées sont notées par $\hat{\alpha}$ et $\hat{\beta}$.

I.4.1 - Propriétés des estimateurs

- On dit que α est un estimateur non biaisé si :

$$E[\hat{\alpha}] - \alpha = 0$$

α paramètre de la population

- La différence entre le paramètre α et $\hat{\alpha}$ est dite erreur de l'estimation. L'espérance du carré de cette erreur est appelée MSE (moyenne du carré de l'erreur).

$$E[(\alpha - \hat{\alpha})^2] = \text{Var}(\hat{\alpha}) + (\alpha - E[\hat{\alpha}])^2$$

Si $\hat{\alpha}$ est non biaisé $\text{MSE} = \text{Var}(\hat{\alpha})$ ce qui veut dire que le bon choix de $\hat{\alpha}$ doit être avec une MSE minimale. Si $\hat{\alpha}$ est non biaisé et est à MSE minimale, $\hat{\alpha}$ est dit un estimateur efficace.

La méthode d'estimation prend en considération ces propriétés, mais comme l'estimateur est souvent non biaisé et il n'a pas une MSE minimale, celui choisi doit être donc moins biaisé et à moindre MSE.

Si ceci n'est pas possible, l'analyste doit juger laquelle de ces deux propriétés est la plus désirable pour le cas étudié puis choisira l'estimateur.

- Soit $\hat{\alpha}_n$ un estimateur du paramètre α à partir d'un échantillon de taille N .

Si $\hat{\alpha}_n \rightarrow \alpha$ tout en augmentant N , alors $\hat{\alpha}_n$ est un estimateur consistant.

Si l'estimateur utilise le maximum d'informations contenues dans les données, l'estimateur est alors jugé suffisant .

I.4.2 - Méthodes d'estimation

Il existe trois méthodes d'estimation des paramètres, la plus simple est la méthode des moments.

a - Méthode des moments

La méthode consiste à évaluer les moments calculés à partir de l'échantillon et les moments du même ordre de la population.

$$E[X^k] = \frac{1}{N} \sum_{i=1}^N X_i^k$$

x : variable aléatoire de la population
 x_i : variable aléatoire de l'échantillon
 N : taille de l'échantillon
 k : ordre du moment

Si le modèle contient p paramètres on doit calculer les p premiers moments qui donneront p équations à résoudre et dont les solutions seront les paramètres à estimer .

Cette méthode est souvent utilisée comme une première approximation pour l'estimation des paramètres, à l'aide d'autres méthodes.

En général les moments estimés, excepté la moyenne sont souvent biaisés. Pour cela on applique un ajustement afin de les rendre non biaisés.

Notons que les moments estimés seront efficaces si la distribution est la loi Normale ($C_s = 0$).

b - Méthode des moindres carrés

Considérons une série chronologique $Y_1 \dots Y_N$ comme échantillon, représentée par le modèle suivant.

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, \alpha_1, \alpha_2, \dots, \alpha_p) + \epsilon_t$$

$\alpha_{1,2,\dots,p}$: paramètres du modèle

ϵ_t : résidu de moyenne nulle

La méthode des moindres carrés (MC) est basée sur le fait que la somme des carrés des différences entre les valeurs observées $Y_1 \dots Y_N$ et valeurs estimées $Y_t = f(Y_{t-1}, \dots, \alpha_1, \dots, \alpha_p)$ $t=1, N$ est minimale. En d'autres termes ceci veut dire que les $\alpha_1 \dots \alpha_p$ sont choisis de façon à ce que $\sum \epsilon_t^2$ soit minimale.

$$\sum_{t=1}^N e_t^2 = \sum_{t=1}^N (Y_t - \hat{Y}_t)^2$$

Cette somme est minimale si les dérivées partielles par rapport à $\alpha_1, \dots, \alpha_p$ respectivement, sont nulles.

$$\frac{\partial \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}{\partial \alpha_1} = 0, \dots, \frac{\partial \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}{\partial \alpha_p} = 0$$

On obtient donc un système de p équations à p inconnues. La résolution de ce système nous permet de déterminer les paramètres estimés $\hat{\alpha}_1, \dots, \hat{\alpha}_p$.

c - Méthode du Maximum de Vraisemblance

Considérons le modèle:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, \alpha_1, \dots, \alpha_p) + e_t$$

d'une série chronologique échantillonnée Y_1, \dots, Y_N .

Posons $\alpha = \{\alpha_1, \dots, \alpha_p\}$.

La loi de probabilité des e_1, \dots, e_N est appelée fonction de vraisemblance $L(\cdot)$ et s'écrit:

$$L(\cdot) = f(e_1; \alpha) \cdot f(e_2; \alpha) \cdot \dots \cdot f(e_N; \alpha) = \prod_{t=1}^N f(e_t; \alpha)$$

L'estimé du paramètre α par la méthode du Maximum de Vraisemblance (MDV) est obtenu quand la fonction $L(\cdot)$ est maximale. Le même estimé est obtenu si le log la fonction de vraisemblance est maximal.

La fonction log de vraisemblance s'écrit:

$$L.L(\cdot) = \ln \prod_{t=1}^N f(\epsilon_t; \alpha) = \sum_{t=1}^N \ln f(\epsilon_t, \alpha)$$

Cette fonction est maximale si:

$$\frac{\partial L.L(\cdot)}{\partial \alpha_1} = 0, \dots, \frac{\partial L.L(\cdot)}{\partial \alpha_p} = 0$$

La résolution de ce système d'équation donne les valeurs des $\alpha_1, \dots, \alpha_p$. Ces estimateurs sont efficaces, consistants et suffisants.

Si les ϵ_t sont normalement distribués de moyenne nulle et de variance σ_ϵ^2 , la fonction $L(\cdot)$ s'écrit

$$L(\cdot) = \frac{1}{(\sqrt{2\pi} \sigma_\epsilon)^N} \exp \left[-\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^N \epsilon_t^2 \right]$$

Le log de la fonction de vraisemblance s'écrit donc:

$$L.L(\cdot) = -N \ln(\sqrt{2\pi} \sigma_\epsilon) - \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^N \epsilon_t^2$$

Cette fonction est maximale si:

$$\frac{\partial \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}{\partial \alpha_1} = 0, \dots, \frac{\partial \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}{\partial \alpha_p} = 0$$

Qui est le résultat donné par la méthode des moindres carrés. Mais la méthode du maximum de vraisemblance donne une meilleure approximation, car par exemple:

$$\text{Si } Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

$$\text{La MMC nous donne: } \phi_1 = \frac{\sum Y_t Y_{t-1}}{\sum Y_{t-1}^2} \quad t = 2, \dots, N$$

$$\text{La MMV nous donne: } \phi_1 = \frac{\sum_{t=1}^N Y_t Y_{t-1}}{\sum_{t=1}^N Y_{t-1}^2}$$

Pour la deuxième méthode la valeur de Y_0 est incluse dans l'équation.

Dans tout les cas, et quelque soit la méthode utilisée, les paramètres doivent être estimés conjointement, car les variables de l'échantillon sont mutuellement dépendantes. Il va sans dire que cette estimation est très complexe dans le cas d'un modèle ayant un grand nombre de paramètres.

I.5 - TESTS DU MODELE

Après l'estimation des paramètres du modèle choisi, il faudrait effectuer des tests pour vérifier la capacité du modèle à reproduire toutes les caractéristiques statistiques des séries historiques, cette étape s'effectue au moyen des tests suivants:

- Test d'adéquation (ou de représentativité) qui comporte:
 - # Test d'indépendances des résidus .
 - # Test de normalité des résidus .
 - # Test d'indépendances des résidus vis-à-vis Z_t .
- Test de parcimonie.
- Test de ressemblance des caractéristiques statistiques.
- Test de validation.

1.5.1 - TEST D'INDEPENDANCES DES RESIDUS

Il existe plusieurs tests d'indépendances tel le test de portemanteau, test du periodégramme cumulatifs, test d'Anderson et d' autres ...

Présentons celui de portemanteau:

Considérons une série chronologique x_i de taille N représentée par un modèle AR(p), p est l'ordre du modèle, le test utilise le nombre Q donné par la relation suivante :

$$Q = N \sum_{k=1}^L r_k^2(\epsilon)$$

avec

$r_k(\epsilon)$: correlogramme des résidus
 L : le maximum de décalage considéré

qui est approximativement distribué comme une $\chi^2_{\alpha}(L-p)$ où α est le seuil de confiance, le test nous permet de comparer Q à la valeur théorique $\chi^2_{\alpha}(L-p)$.

Si Q est inférieur à $\chi^2_{\alpha}(L-p)$ les ϵ_t sont indépendants.

1.5.2 - TEST DE NORMALITE DES RESIDUS

On peut citer comme test de normalité le test d'asymétrie . En se basant sur le fait que la loi normale représente une asymétrie dans sa distribution c'est à dire que son coefficient d'asymétrie g est nulle.

Si la série suit une lois normale g a une distribution asymptotiquement normale de moyenne nulle et de variance égale à $6/N$ (SNEDECOR-COCHRAN, 1967). Les intervalles de confiance, et ce pour une probabilité $(1-\alpha)$, de g sont:

$$[U_{1-\alpha/2} \sqrt{(\sigma/N)} ; U_{1-\alpha/2} \sqrt{(\sigma/N)}]$$

Si la valeur de g calculée est à l' intérieur de l'intervalle de confiance l' hypothèse de normalité est acceptée.

En effet ce test n' est valable que pour des séries de taille supérieur à 150; SNEDECOR et COCHRAN 1967 [] ont proposé un tableau du test d' asymétrie pour des séries de taille inférieur à 150,

I.5.3 - TEST D' INDEPENDANCE DES ε_t VIS-à-VIS Z_t

On dit que deux variables sont indépendantes si leurs espérance est nulle c' est à dire que :

$$\text{Cov}(\varepsilon_t , Z_{t-k}) = 0 \quad k \geq 1$$

on calcule donc

$$C_k = \frac{\sum_{t=1}^N \epsilon_t (Z_{t-k} - Z)}{\sum_{t=1}^N \epsilon_t^2 \left(\sum_{t=1}^N (Z_t - Z)^2 \right)^{1/2}}$$

C_k est significatif si elle satisfait la condition suivante pour un intervalle de confiance 95% .

$$C_k > 2 / \sqrt{N-k}$$

I.5.4 - TEST DE PARCIMONIE

Le principe de la modélisation est de trouver un modèle avec un nombre minimum de paramètres qui reproduit adéquatement les caractéristiques de la série historique, c' est le principe de parcimonie.

La parcimonie se traduit par le rapport

$$S = N / k$$

où N: nombre d' observations
k: nombre de paramètres

Ce nombre doit être supérieur à une valeur égale à 15. Pour les modèles ARMA la parcimonie est mesurée par le nombre AIC (le critère d'information d'AKAIKE, AKAIKE 1979) qui doit être minimale.

$$AIC = N \ln(\sigma^2_{\epsilon}) + 2(p+q)$$

I.5.5 - TEST DE RESSEMBLANCE DES CARACTERISTIQUES STATISTIQUES

Ce test consiste à comparer les résultats des caractéristiques des séries chronologiques et ceux des séries générées.

Les caractéristiques principales sont la moyenne, la variance et le coefficient d'asymétrie. La comparaison des caractéristiques des séries historiques et des séries générées est faite sur la base des statistique qui dérivent des erreurs relatives (SRIKANTAN ET MAHON 1982 []).

$$ER_t = \frac{CG_t - CH_t}{CH_t}$$

$$VER_n = \sum_{t=1}^{12} ER_t^2 / 12 \quad n= 1, \dots, N_c$$

Où:

ER_{τ} : Erreur relative périodique correspondant à la période τ

CG_{τ} : caractéristique statistique de la série générée

CH_{τ} : caractéristique statistique de la série historique

VER_n : moyenne des carrés des erreurs relatives correspondent à la caractéristique n .

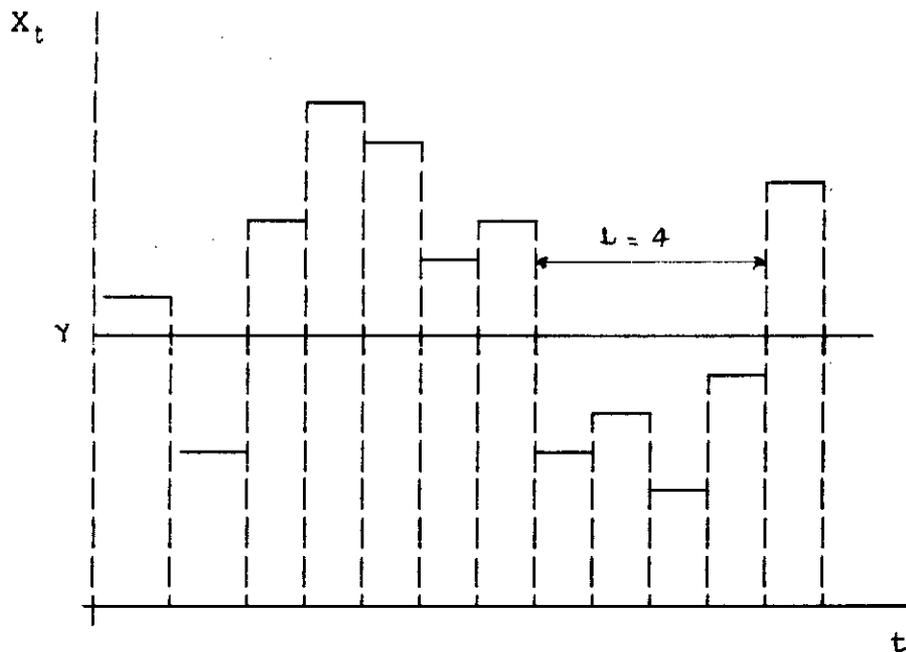
N_c : nombre de caractéristiques considérées

Le même principe pour les séries globales.

I.5.6 - TEST DE VALIDATION

Le but de ce test est de faire la comparaison des caractéristiques des RUNS des séries historiques et des séries générées.

Le RUN est une caractéristique des suites homogènes c'est à dire, si on considère une série de débits X_t de taille N représentée sur la figure ci dessus, la ligne droite (Y) représente la demande en eau :



- Définition du RUN. -

Lorsque $X_t < Y$ durant un ou plusieurs intervalles consécutifs ceci correspond à "un RUN négatif" le contraire ça sera un "RUN positif". ces RUNS sont présentées par leur longueurs (nombre d' intervalle) ou leurs volumes.

Les expressions de caractéristique des suites homogènes (RUN) des longueurs sont données par :

-La moyenne:

$$L_N = \frac{1}{M} \sum_{j=1}^M L(j)$$

-La variance:

$$S_N^2(L) = \frac{1}{M-1} \sum_{j=1}^M (L(j) - \bar{L}_M)^2$$

-La longueur maximale:

$$L_N^* = \max_M (L(1), L(2), \dots, L(M))$$

CHAPITRE DEUX

LE MODELE DE DESAGREGATION

Qu'est-ce que la Désagregation?

L'analyse stochastique est basée sur la supposition d'un processus stationnaire, c'est à dire que les propriétés statistiques restent stationnaires dans le temps. Ceci implique que les données historiques peuvent faire l'objet d'une grande extension (série synthétique).

La synthèse des séquences hydrologiques a été présentée comme étant un outil utile en ce qui concerne la gestion des ressources hydrauliques.

Dans le but de réaliser un tel objectif, les séquences synthétiques doivent être conformes aux données de base (données historiques) et ce sur le plan des paramètres statistiques bien sûr.

Avant, la synthèse des valeurs hydrologiques utilisaient le modèle MARKOVIEN. Ce modèle concerne la moyenne, la variance et les coefficients de corrélation, l'inconvénient était que la statistique des données n'était plus conservative lorsqu'on aggregait d'un niveau à un autre. Ce problème a été surmonté par le modèle de VALENCIA et SCHAAKE qui ont apporté une importante contribution pour la désagregation des valeurs annuelles, de façon à ce que la statistique soit préservée à plus d'un niveau.

Ce modèle présente une capacité de désagregation de séquences multivariantes annuelles qui peuvent provenir de n'importe quel modèle de génération annuelle. Mais ces niveaux d'agrégation sont liés avec le passé par la statistique des niveaux annuels. Ce problème a été surmonté à l'aide du modèle de MEJIA et ROUSSELLE qui ont construit un modèle plus réaliste.

Quelques améliorations ont été faites par LANE pour rendre ce modèle plus simple et donc plus facile à utiliser.

II.1 - Historique

Le modèle de désagregation est l'une des techniques de modélisation des séries chronologiques en hydrologie.

La première approche a été proposée en 1967 par HANNS et CAMPBELL mais un modèle plus acceptable fut présenté par VALENCIA et SCHAAKE en 1973.

En 1976 MEJIA et ROUSSELLE ont porté une modification au modèle en le rendant plus réaliste.

LANE en 1979 a proposé quelques améliorations en développant un calcul simple pour une application praticable tant sur le plan temporel que sur le plan spatial.

Nous pouvons énumérer quatre (04) types de modèles:

- # Le modèle de base
- # Le modèle étendu
- # le modèle condensé
- # le modèle spatial

Le but primordial de n'importe quel modèle de désagregation est la conservation des propriétés statistiques à plus d'un niveau.

Si nous prenons par exemple le cas d'une désagregation annuelle en mensuelle, le but est donc de préserver les propriétés ou les caractéristiques statistiques des données historiques du niveau mensuel, certaines propriétés statistiques du niveau annuel et les propriétés entre les deux niveaux, le mensuel et l'annuel.

Les propriétés considérées comme importantes [5]
sont les suivantes :

- La moyenne
- La variance
- La loi de distribution
- La covariance

II.2 - Description générale du modèle

La désagregation est un procédé avec lequel une série chronologique générée dépend d'une série déjà disponible. La série originale est donc désagregée en sous-séries; leur somme donne ce qu'on appelle une série mère.

La génération des sous-séries est réalisée par un modèle linéaire désigné pour préserver les propriétés statistiques considérées comme importantes entre les éléments de la sous-série, les éléments de la subsérie et les éléments de la série mère. De cette manière, les propriétés statistiques seront préservées aux niveaux de la série mère et de la sous-série ; De plus les relations entre les deux niveaux seront maintenues

On peut à cet effet citer comme exemple de désagregation temporelle , la désagregation des séries annuelles en mensuelles. Pour ce qui est de la désagregation spatiale, elle concerne la désagregation du débit total d'une station en débits aux différentes stations (sites) d'un bassin fluvial.

II.2.1 - Formulation mathématique du modèle

Les modèles de désagregation ont une forme de dépendance linéaire qui est donnée par:

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

Y: série à générer (sous-série)
X: série originale (série mère)
 ε : série des variables aléatoires
A, B: paramètres du modèle

En général le modèle est multivariable ce qui implique que Y, X et ε soient des vecteurs et A et B des matrices. Notons que X et Y suivent la loi Normale de moyenne nulle et de leur propre variance. Les termes stochastiques de la matrice colonne ε suivent la loi Normale centrée réduite.

Si les différents termes du modèle ne suivent pas une loi de distribution de moyenne nulle, le modèle aura la forme suivante:

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

II.2.2 - Modèle de désagregation à site unique

Ce modèle se présente sous trois formes:

- Le modèle de base
- Le modèle étendu
- Le modèle condensé

a) Modèle de base (Modèle de VALENCIA-SCHAAKE)

Le modèle de base est donné sous la forme suivante:

$$Y = A X + B \epsilon$$

Ce modèle n'est valable que pour des variables distribuées normalement; Les termes de la série X et de la série Y doivent donc être transformées en valeurs normales centrées.

D'après plusieurs auteurs, ce modèles est considéré comme le plus claire d' où le plus facile des trois, mais il à l'inconvénient de prendre en compte un nombre important de paramètres.

Ce modèle tel qu'il présenté preserve la covariance entre les valeurs annuelles - saisonnières et entre les valeurs saisonnières - saisonnières.

Si w est le nombre de saisons on aura w^2+w moments préservés.

b) Modèle étendu

Developpé par MEJIA et ROUSSELLE, il n'est qu'une simple extension du modèle de base. Un terme est ajouté pour préserver la covariance saisonnière entre les saisons de l'année courante et les saisons de l'année antécédante.

Ce modèle est de la forme:

$$Y = A X + B \epsilon + C Z$$

où Z: Matrice des valeurs saisonnières des années antécédantes.

Malgré que ce modèle soit plus compliqué que le premier, il est aussi considéré comme simple et clair par certains auteurs.

L'ajout du troisième terme rend l'estimation des paramètres plus compliquée et leur nombre devient par là même très excessif.

c) Modèle condensé

LANE a développé une approche qui remplace les valeurs des paramètres non importants du modèle étendu en valeurs nulles. Ainsi, le nombre de paramètres estimés et donc le nombre de moments préservés sera réduit.

L'approche utilise la forme du modèle étendu mais chaque saison a son propre modèle:

$$Y_t = A_t X + B_t \epsilon + C_t Y_{t-1}$$

Remarquons que ce modèle est d'ordre 1 (à pas -1). Ainsi si w est le nombre de saisons nous aurons w équations individuelles et trois paramètres pour chaque équation, d'où un total de $3w$ paramètres à estimer. Cette réduction du nombre de paramètres rend le modèle plus praticable.

Remarque

Notons que les deux premiers modèles ont le problème de structure inconsistence c'est à dire que le début correspondant à la dernière saison est generé en préservant la covariance entre cette valeur et les (w-1) valeurs précédantes, Tandis que la première valeur est generée sans présevation de la covariance entre cette valeur et n'importe quelle valeur saisonnière.

Pour ce qui est du troisième modèle, la somme des données saisonnières ne donne pas exactement la série annuelle, car les données saisonnières sont générées séparément. Mais il faut noter que ce même problème surgit dans les trois approches si les données subissent n'importe quelle transformation.

II.2.3 - Modèle temporel multisite

Ce modèle a la même forme que le modèle à site unique, seulement les matrices sont plus larges, donc plus compliquées à manipuler.

$$Y = A X + B \epsilon + C Z$$

X: Matrice colonne contenant les valeurs annuelles de chaque site.

Y: Matrice colonne des valeurs saisonnières pour chaque site.

Si w est le nombre de saisons, n le nombre de sites et w' le nombre de saisons de la matrice Z , alors les dimensions des matrices seront:

$$\dim(Y) = n \omega * 1$$

$$\dim(X) = n * 1$$

$$\dim(e) = n \omega * 1$$

$$\dim(Z) = n \omega' * 1$$

d'où

$$\dim(A) = n \omega * n$$

$$\dim(B) = n \omega * n \omega$$

$$\dim(C) = n \omega * n \omega'$$

II.3 - Propriétés du modèle

1) Préservation des valeurs générées

Puisque toutes les données (série mère, termes stochastiques) ont une moyenne nulle, et comme les coefficients du modèle sont constants, les sous-séries auront elles aussi des moyennes nulles et suivront comme la série mère, une distribution Normale.

2) Préservation de la somme

La somme des données générées donne la série mère mais cela est vrai seulement si les données originales n'ont pas subies de transformations. Pour cela, un ajustement de ces données s'impose.

3) Préservation de la variance et de la covariance

Afin d'éviter tout effet dû aux transformations, on doit examiner les données générées pour que les moments soient préservés.

II.4 - DIFFERENTES PHASES DE LA DESAGREGATION

A - Analyse préliminaire et identification du modèle

Avant toute opération (Génération) sur les données, l'échantillon doit satisfaire certaines hypothèses imposées par la statistique.

1) Ajustement des données

La validité d'un modèle mathématique est limitée par la quantité et la qualité des données. Les données doivent donc subir un test de qualité qui consiste à les homogénéiser.

2) Normalité

Les séries chronologiques utilisées dans de tels modèles (modèles stochastiques) doivent suivre une loi statistique bien déterminée pour pouvoir reproduire des séries de même distribution. De plus les techniques de base de la statistique (estimation de paramètres, tests du modèle ..) sont améliorées pour la loi normale. A cette fin les données subissent un test de normalité.

3) Stationnarité des séries

Le but des études étant la génération des données, les séries historiques doivent être stationnaires dans le temps.

4) Indépendance des données

La statistique impose le fait que les données utilisées doivent être indépendantes sinon l'échantillon ne sera plus représentatif et la série ne pourra plus faire l'objet d'une extension dans le temps .

5) Après passage par ces tests, on soustrait les moyennes des données transformées pour pouvoir les injecter dans le modèle

6) Identification du modèle

Le choix de la forme du modèle consiste à identifier le nombre de séries mère et de sous-séries et le nombre de désagregations demandé.

B - ESTIMATION DES PARAMETRES

La méthode la plus utilisée pour l'estimation des paramètres est la méthode des moments. Cette méthode consiste à évaluer les moments calculés à partir de l'échantillon et ceux de la population puis ressortir les valeurs à estimer.

Prenons la forme générale du modèle de désagrégation suivante:

$$Y = A X + B \varepsilon \quad (1)$$

Multiplions l'équation (1) par la transposée de X puis prenons les valeurs estimées :

$$E(YX^T) = A E(XX^T) + B E(\varepsilon X^T) \quad (2)$$

Comme ε et X^T sont des valeurs indépendantes :

$$E(\varepsilon X^T) = E(\varepsilon) \cdot E(X^T) = 0 \quad (3)$$

Car $\varepsilon \sim N(0,1)$

L'équation (2) devient :

$$E(YX^T) = A E(XX^T) \quad (4)$$

D'où:

$$A = [E(YX^T)] / [E(XX^T)] \quad (5)$$

Pour estimer B on multiplie l'équation (1) par Y^T d'où l'équation:

$$E(YY^T) = A E(XY^T) \quad (6)$$

Qui peut s'écrire:

$$E(Y\bar{Y}^T) = [A E(X\bar{X}^T)]^T \quad (7)$$

Notons que $(AB)^T = \bar{B} A^T$
L'équation (7) devient :

$$E(Y\bar{Y}^T) = E(Y\bar{X}^T) A^T \quad (8)$$

D'autre part multipliant l'équation (1) par $(AY+B\epsilon)^T$:

$$E(Y\bar{X}^T) A^T = A E(X\bar{X}^T) A^T + BB^T \quad (9)$$

En combinant l'équation (8) et (9) on obtient:

$$E(Y\bar{Y}^T) = A E(X\bar{X}^T) A^T + BB^T \quad (10)$$

Ce qui implique:

$$BB^T = E(Y\bar{Y}^T) - A E(X\bar{X}^T) A^T \quad (11)$$

D'où:

$$BB^T = E(Y\bar{Y}^T) - [E(Y\bar{X}^T)] [E(X\bar{X}^T)]^{-1} [E(X\bar{Y}^T)]$$

Posant $E(i\bar{j}^T) = S_{ij}$ qui représente la covariance des éléments i et j . L'équation (5) et (12) s'écrivent:

$$A = S_{YX} \cdot S^{-1}_{XX}$$

$$BB^T = S_{YY} - S_{YX} \cdot S^{-1}_{XX} \cdot S_{XY}$$

Notons que S_{ij} est la matrice transposée de la matrice S_{ji} .

Les paramètres estimés du modèle de base seront donc :

$$A = S_{YX} S_{XX}^{-1}$$

$$BB^T = S_{YY} - S_{YX} S_{XX}^{-1} S_{XY}$$

Les paramètres estimés du modèle étendu, en appliquant le même principe :

$$A = (S_{YX} - S_{YZ} S_{ZZ}^{-1} S_{ZX}) (S_{XX} - S_{XZ} S_{ZZ}^{-1} S_{ZX})^{-1}$$

$$C = (S_{YZ} - A S_{XZ}) S_{ZZ}^{-1}$$

$$BB^T = S_{YY} - A S_{XY} - C S_{ZY}$$

Pour le modèle condensé :
Rappelons que chaque saison a son propre modèle .

$$A_t = \begin{bmatrix} S_{YX}(t, t) - S_{YY}(t, t-1) S_{YY}^{-1}(t-1, t-1) S_{YX}(t-1, t) \\ S_{XX}(t, t) - S_{XY}(t, t-1) S_{YY}^{-1}(t-1, t-1) S_{YX}(t-1, t) \end{bmatrix}$$

$$C_t = [S_{YZ}(t, t-1) - A_t S_{XZ}(t, t-1)] S_{ZZ}^{-1}(t-1, t-1)$$

$$B_t B_t^T = S_{YY}(t, t) - A_t S_{XY}(t, t) - C_t S_{ZY}(t-1, t)$$

Notons que la matrice B est obtenue par la décomposition de cholesky.

C - TESTS DU MODELE

C.1 - L'ADEQUATION DES MODELES DE DESAGREGATION

Les séquences générées sont analysées afin de vérifier les hypothèses du modèle .

* A cause de la transformation des données la distribution est espérée préservée adéquatement, il est recommandé donc de

- Tracer les données historiques transformées sur papier de probabilité de la loi normale pour vérifier la linéarité du tracé.

- Tracer les données historiques non transformées et les données générées sur papier de probabilité de la loi normale pour les comparées afin de s'assurer que la distribution générée est raisonnable vu qu'elle a été historiquement observée.

* La linéarité des corrélations des données doit être vérifiée car les données historiques tendent généralement à une linéaire corrélation.

* La troisième recommandation est d'examiner des corrélation préservées indirectement.

Il faut signaler que les tests statistiques sont évitables , à cause du contexte de la désagrégation , ces tests sans compté les difficultés d' application, sont fragiles.

C.2 - TEST DE PARCIMONIE

La parcimonie représente un grand problème pour la désagrégation à cause du nombre excessif des paramètres. Pour pouvoir désagréger il faut que le nombre d'observation soit supérieur au moins dix fois le nombre de paramètres [], voir annexe pour que ce test soit vérifié on joue sur le nombre des paramètres en réduisant la taille de la matrice.

CHAPITRE TROIS

RESULTATS

PRESENTATION DES RESULTATS

Nous exposons dans ce paragraphe les résultats obtenus lors de la mise en exécution des programmes TR., PARD., DESG. et TST., et ce pour les données traitées du site de benibehdel.

Une condition d' utilisation du modèle de désagrégation est la normalité de la série . Comme les données présentent une dissymétrie dans leur distribution (d' après le test) une transformation s' est avérée nécessaire .

Les données ainsi transformées sont utilisées pour estimer les paramètres du modèle .

Nous procédons donc a la génération de dix séries synthétiques afin de tester la fiabilité de ces paramètres, en calculant les caractéristiques statistiques globales et périodiques

(moyenne et ecart-type) pour chacune des séries générées, le test consiste donc a comparer ces valeurs en utilisant les intervalles de confiance correspondant a un seuil de 5% .

Nous rappelons que trois modèles de désagrégation sont utilisés et qu' a chacun d' eux sont appliqués deux transformations différentes; une transformation Log. , et une transformation Log-translatee.

Les graphes présentes en pages (64 a 79) illustrent les résultats obtenus des tests de ressemblance des caractéristiques statistiques (moyenne , ecart type et distribution) globales et périodiques correspondant a chaque couple modele-transformation.

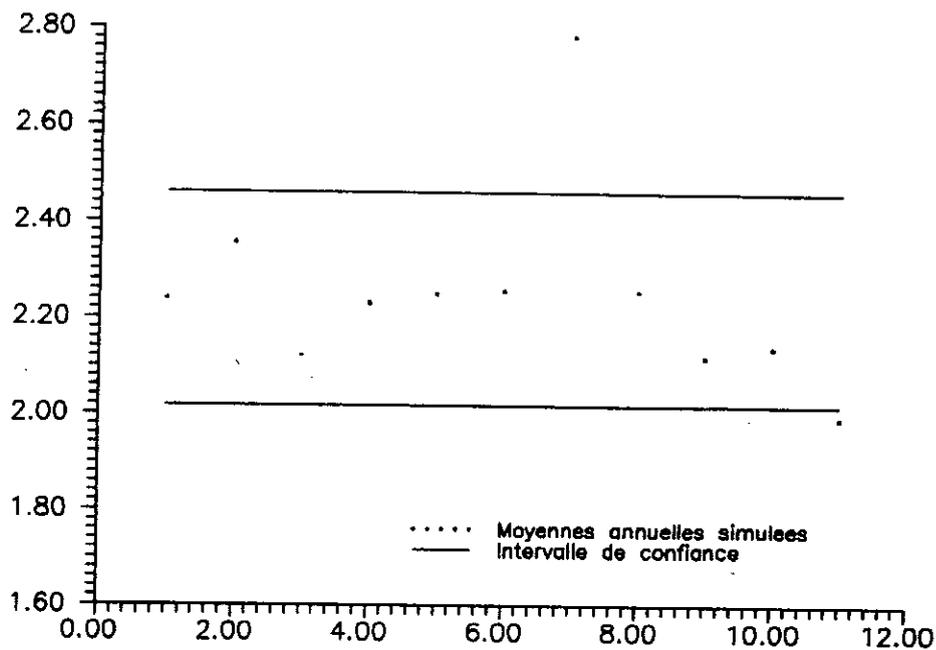
Les résultats correspondants à la transformation 1 (la log) indique la préservation de la somme seulement , même le caractère mensuel n' apparaît pas. (Le résultat était le même en désagrégeant par deux étapes : désagrégation des données

annuelles en saisonnières , puis la désagrégation de ces dernières en données mensuelles).

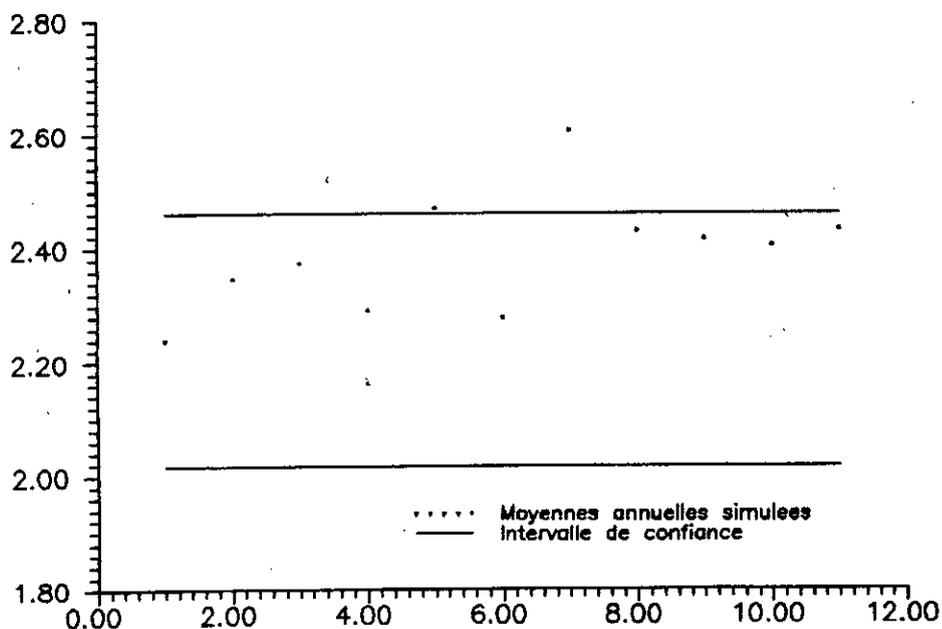
Pour cela nous avons procédé à la simulation des débits en utilisant une autre transformation. nous avons appliqué donc la log translatée en espérant que le problème provient de

la transformation utilisée.

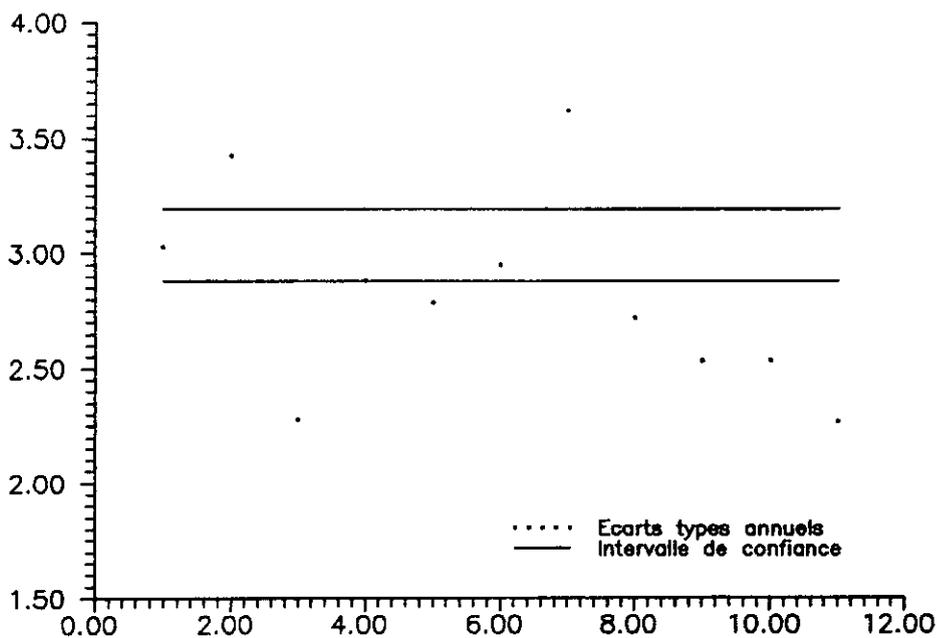
En fin nous remarquons que toutes les caractéristiques statistiques , appart l' écart type annuel , testées sont préservées .et ce pour certaines séries générées.



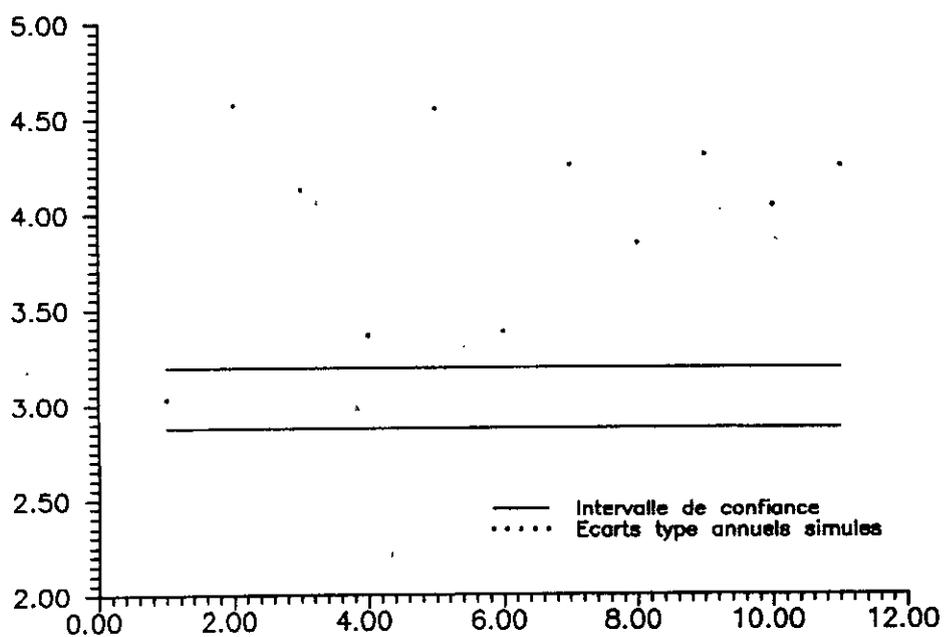
Modele 1 - Transformation 1



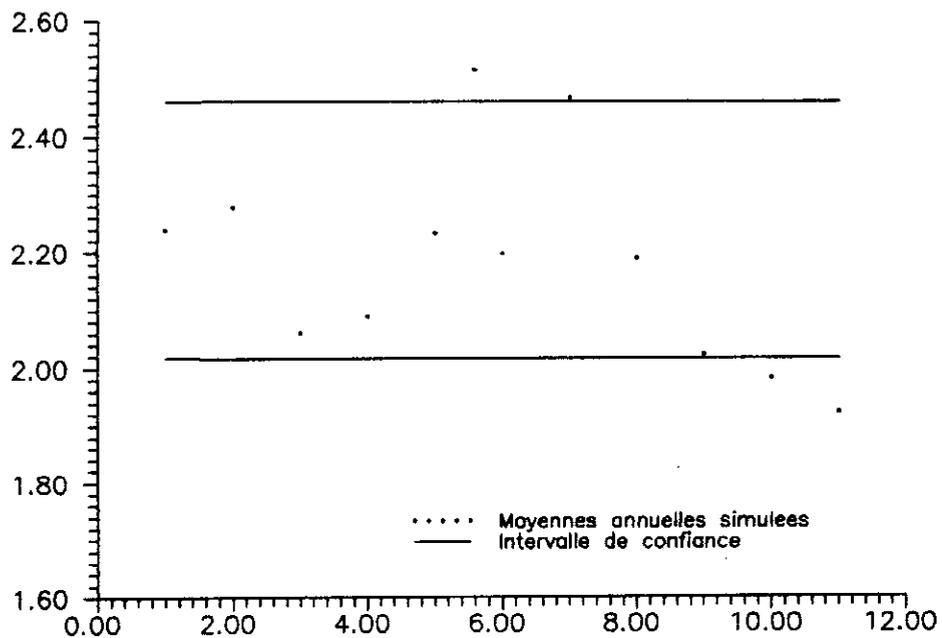
Modele 1 - transformation 2



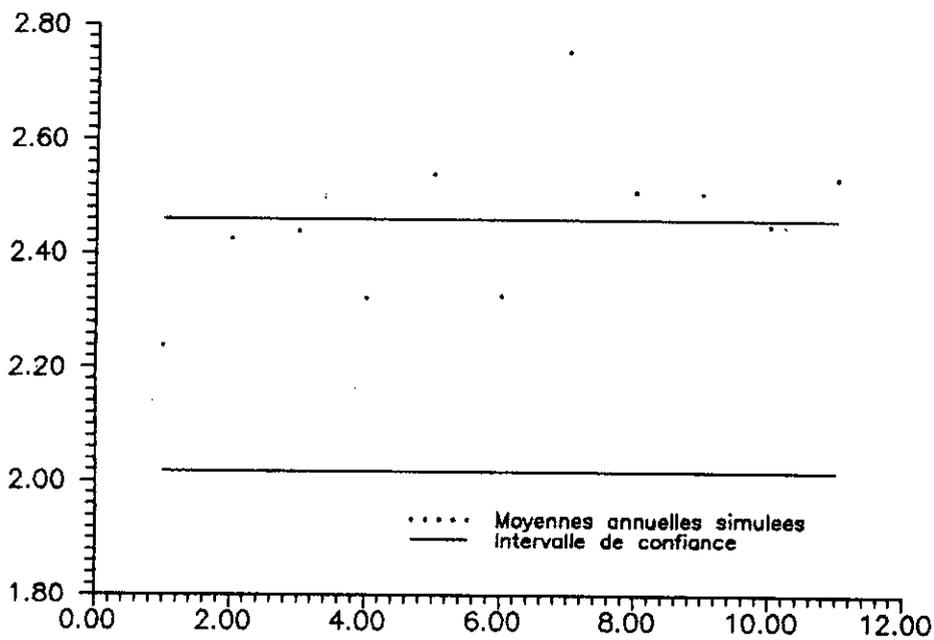
Modele 1 - Transformation 1



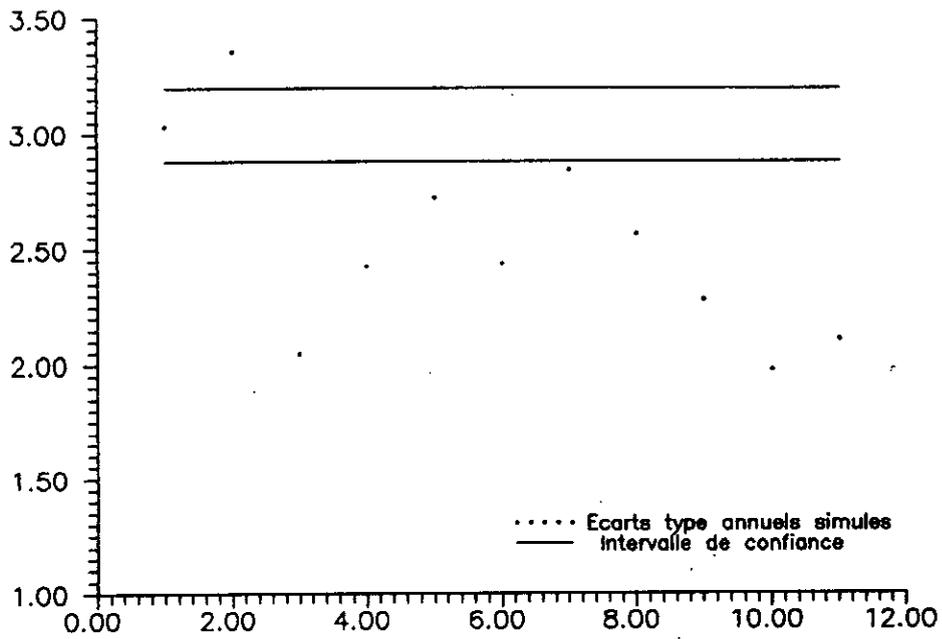
Modele 1 - Transformation 2



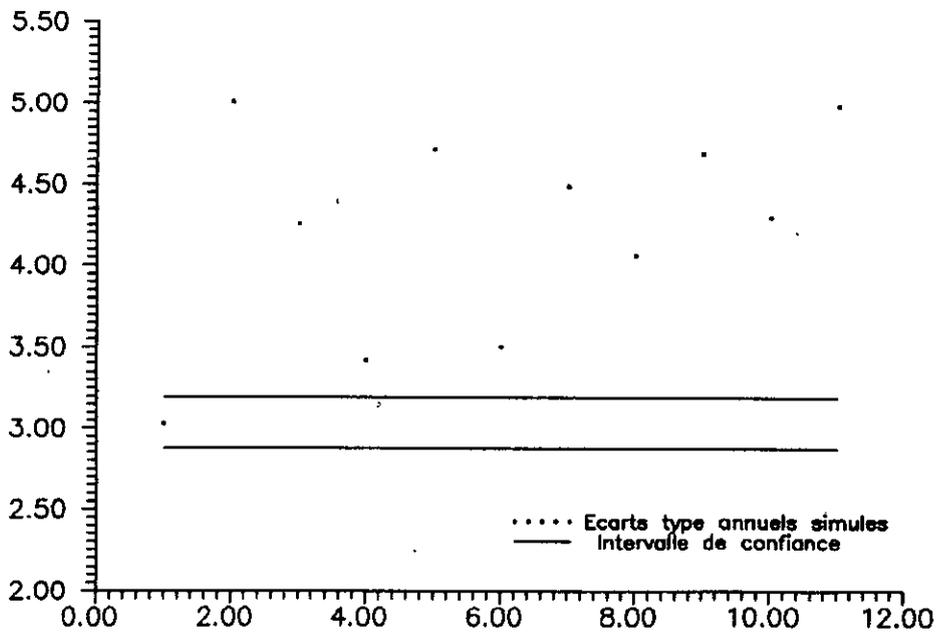
Modele 2 - Transformation 1



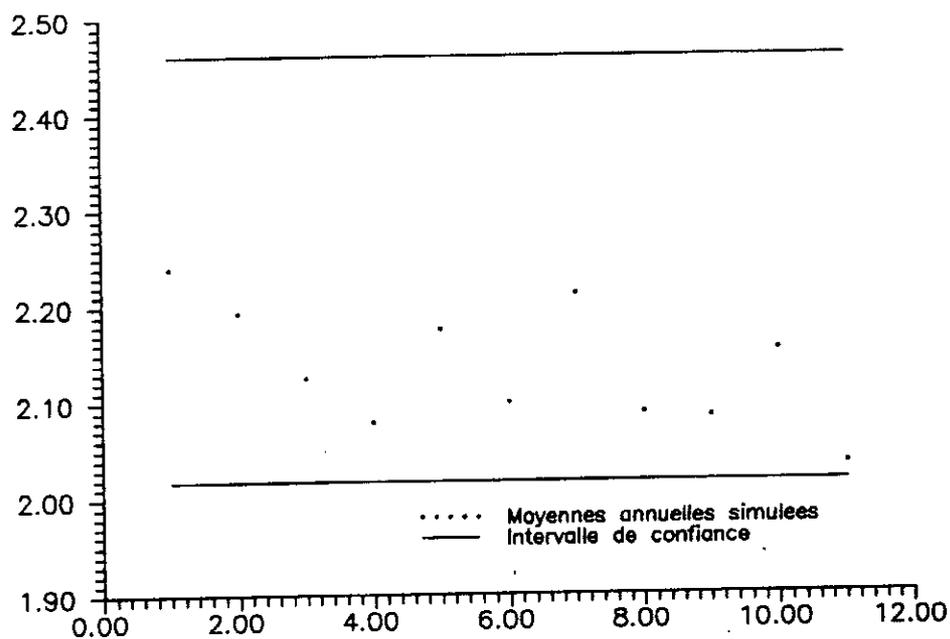
Modele 2 - Transformation 2



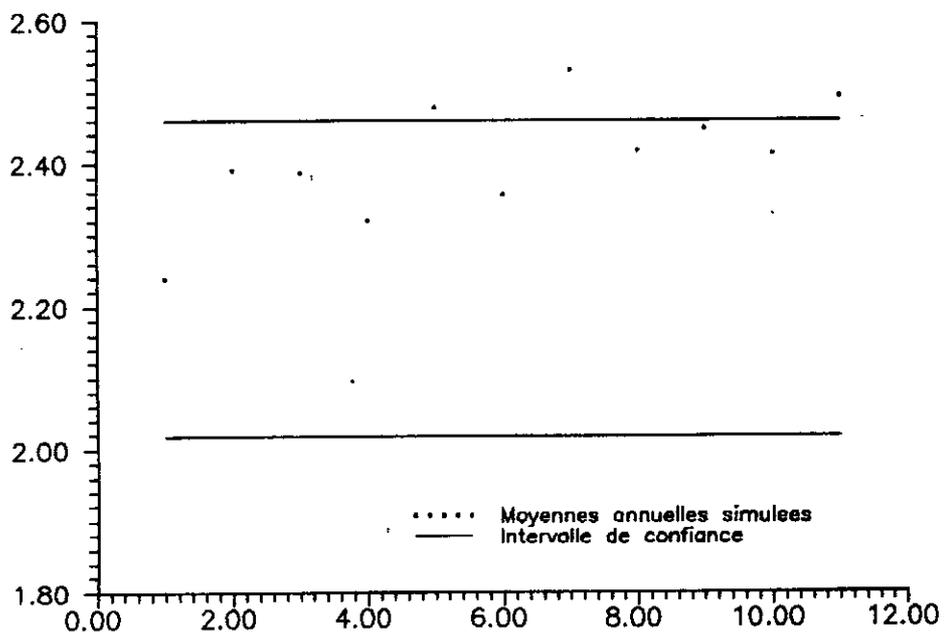
Modele 2 - Transformation 1



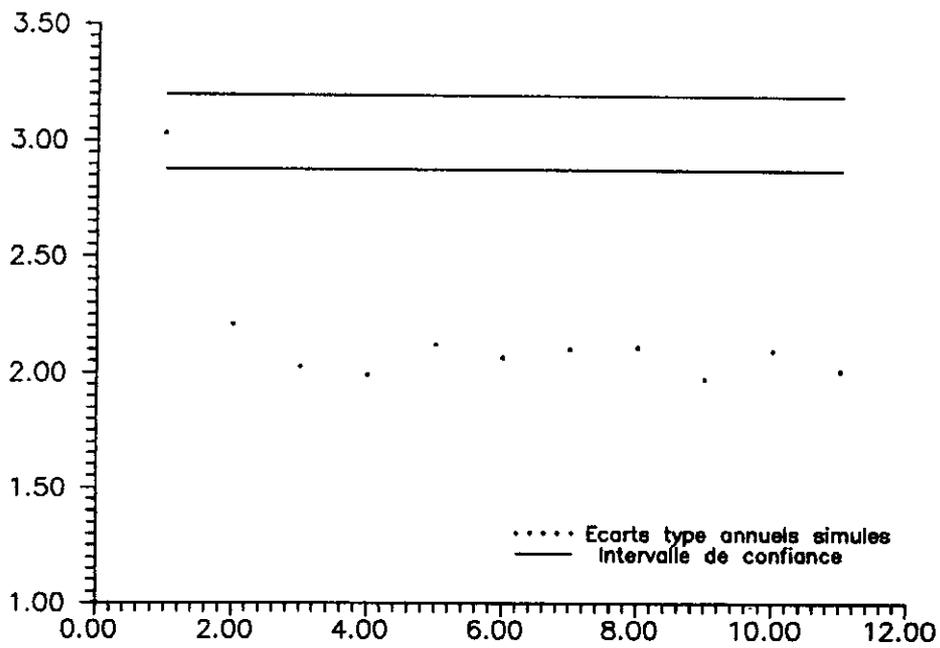
Modele 2 - Transformation 2



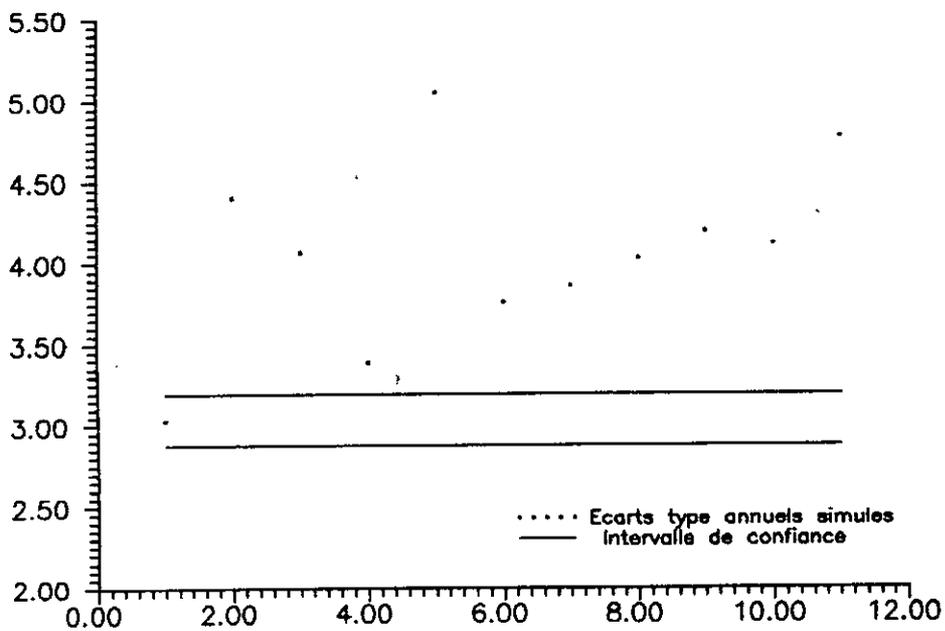
Modele 3 - Transformation 1



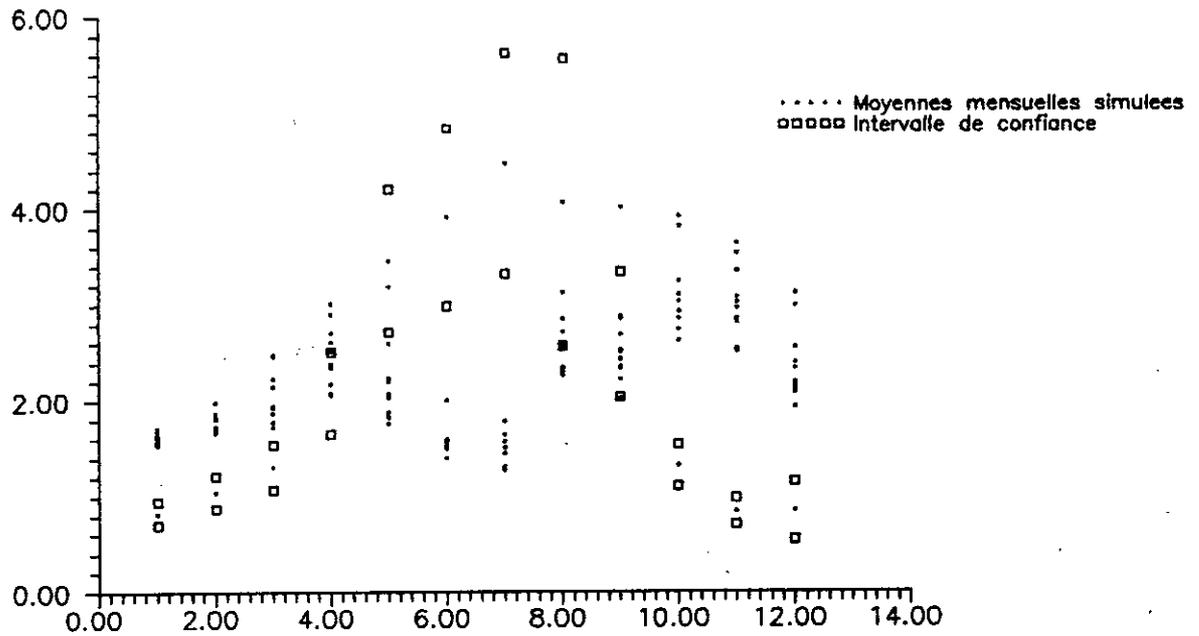
Modele 3 - Transformation 2



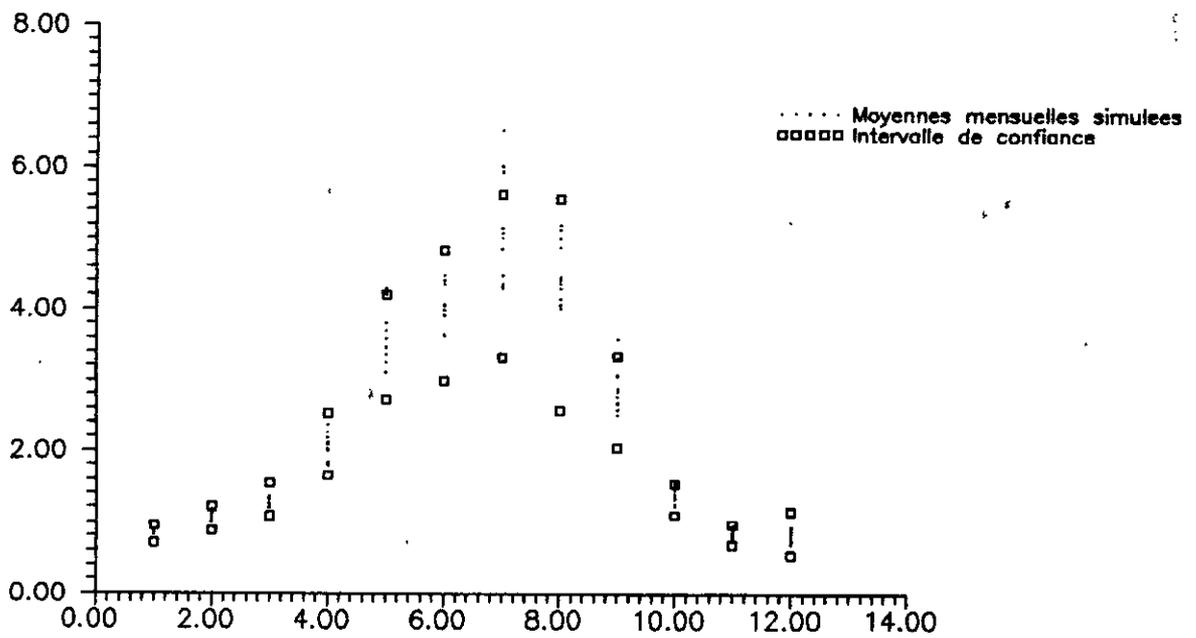
Modele 3 - Transformation 1



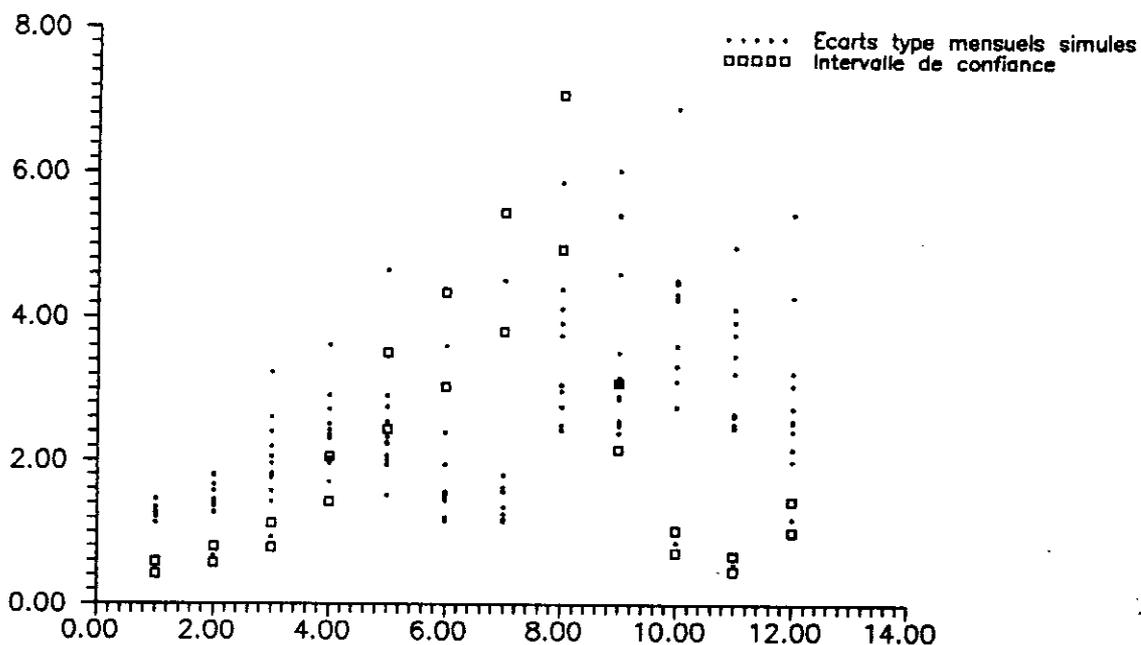
Modele 3 - Transformation 2



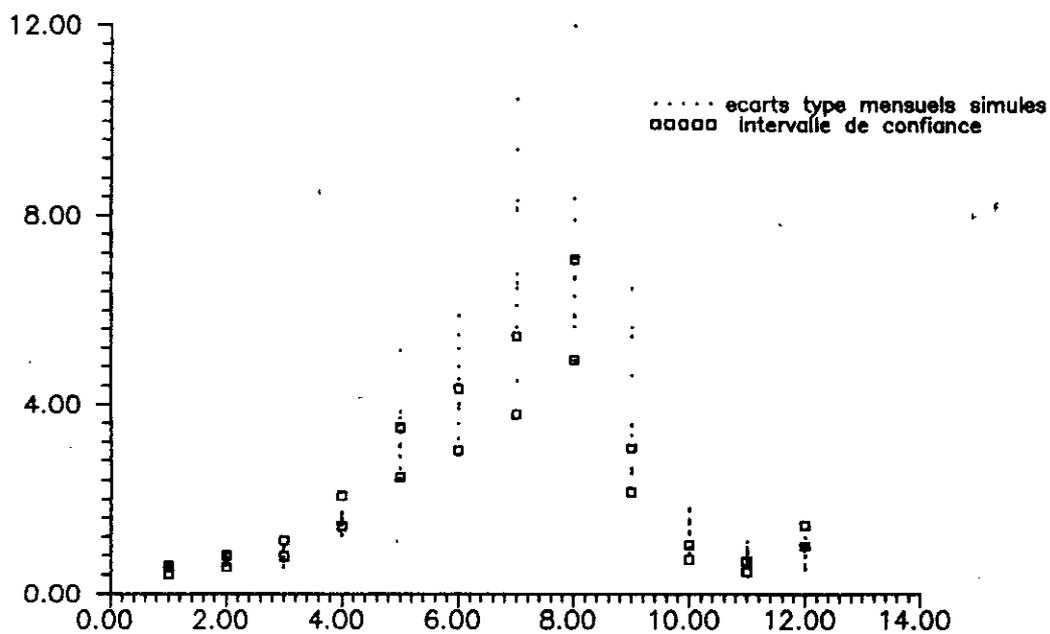
Modele 1 - Transformation 1



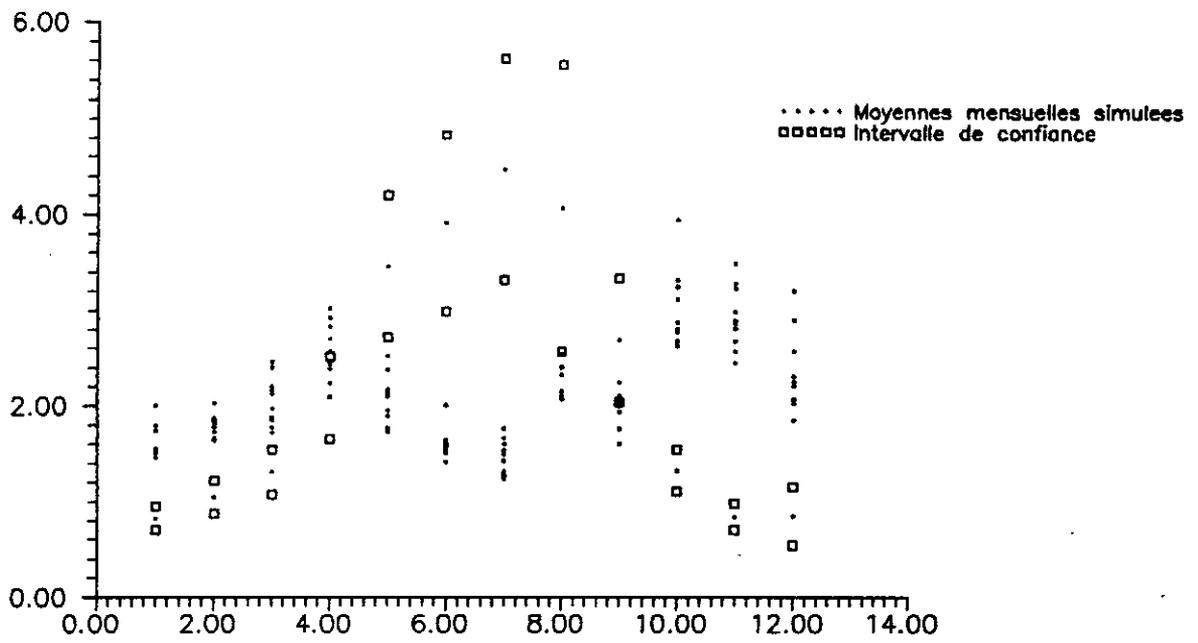
Modele 1 - Transformation 2



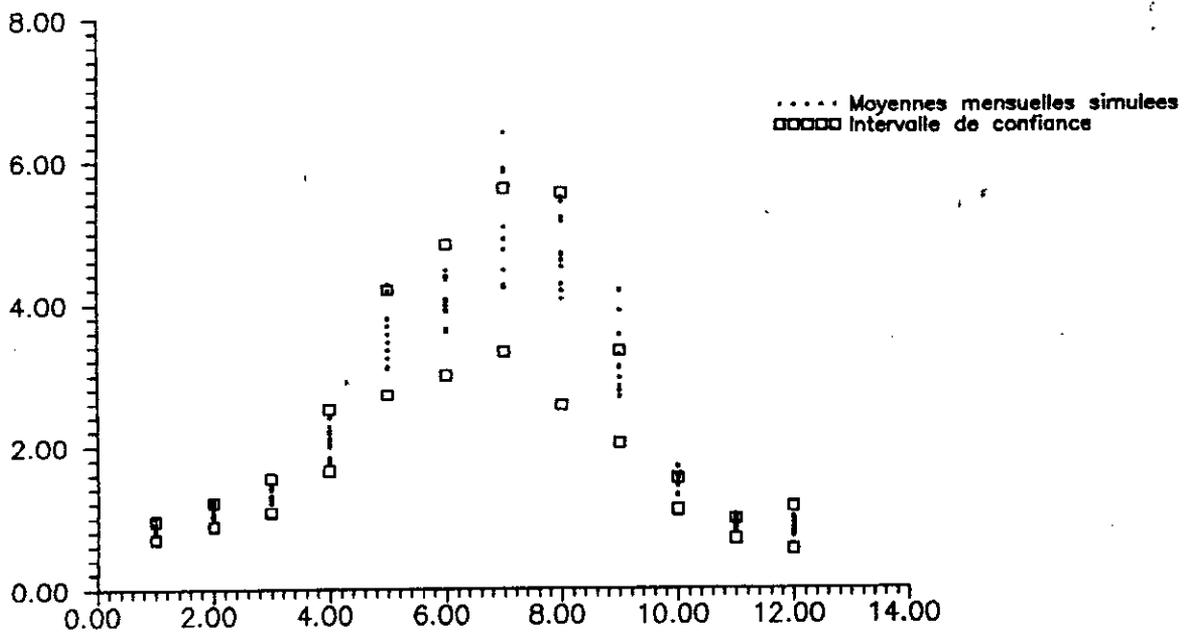
Modele 1 - Transformation 1



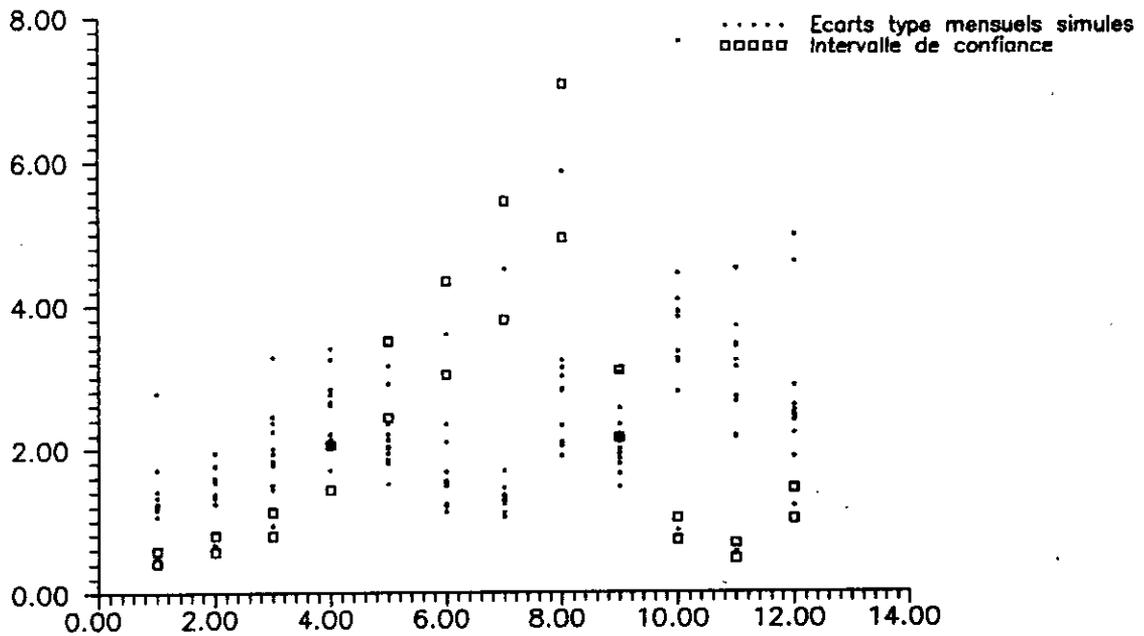
Modele 1 - Transformation 2



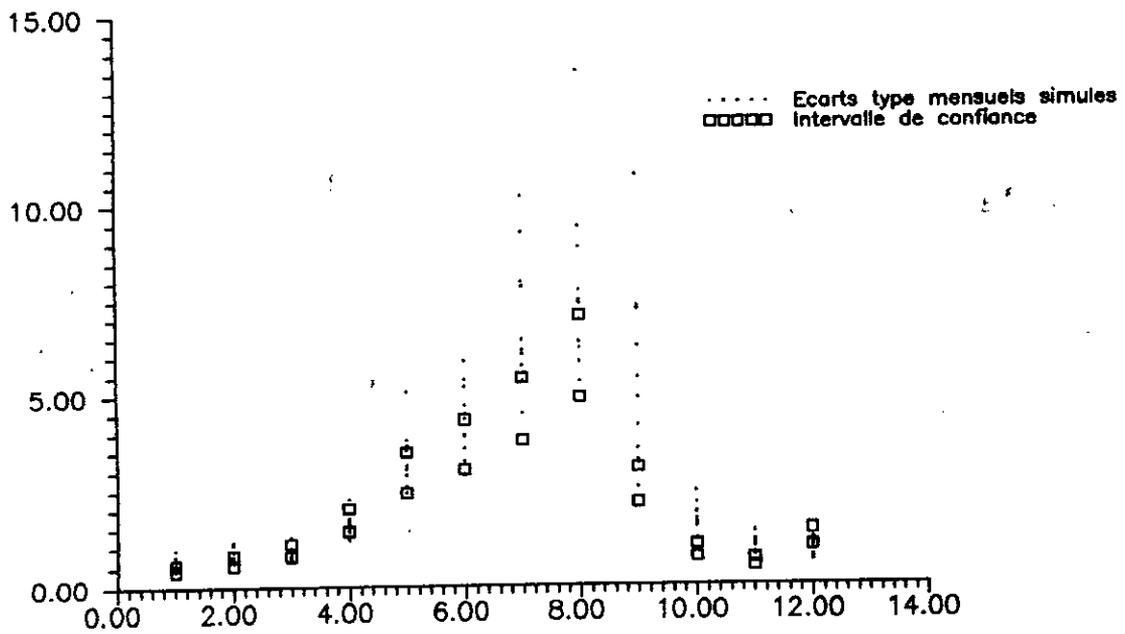
Modele 2 - Transformation 1



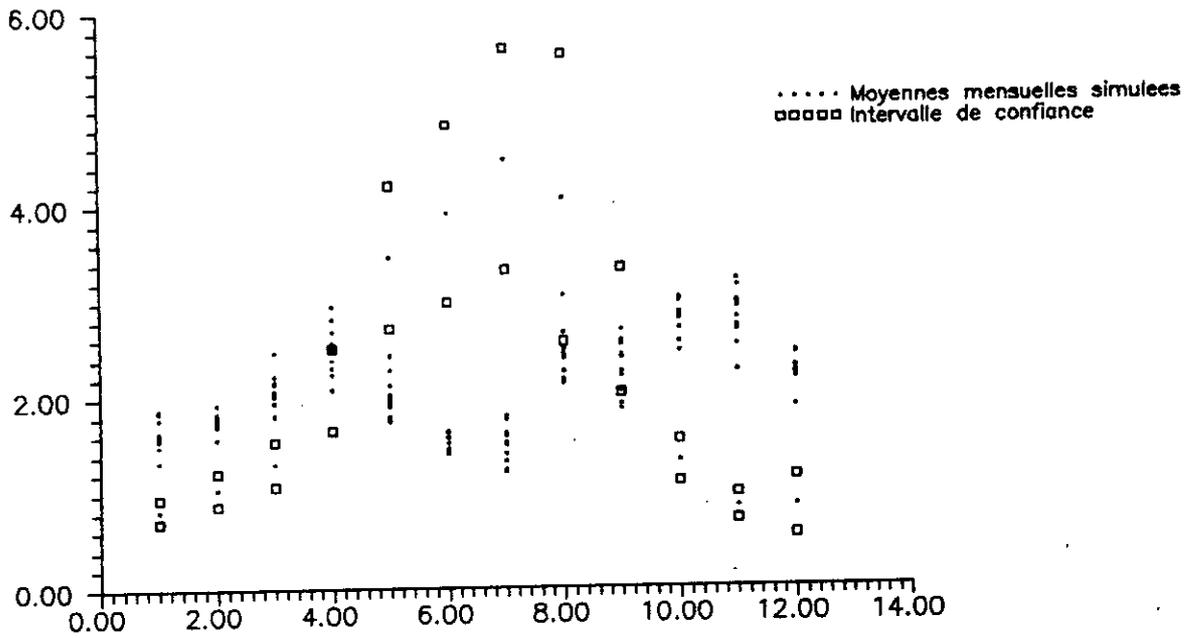
Modele 2 - Transformation 2



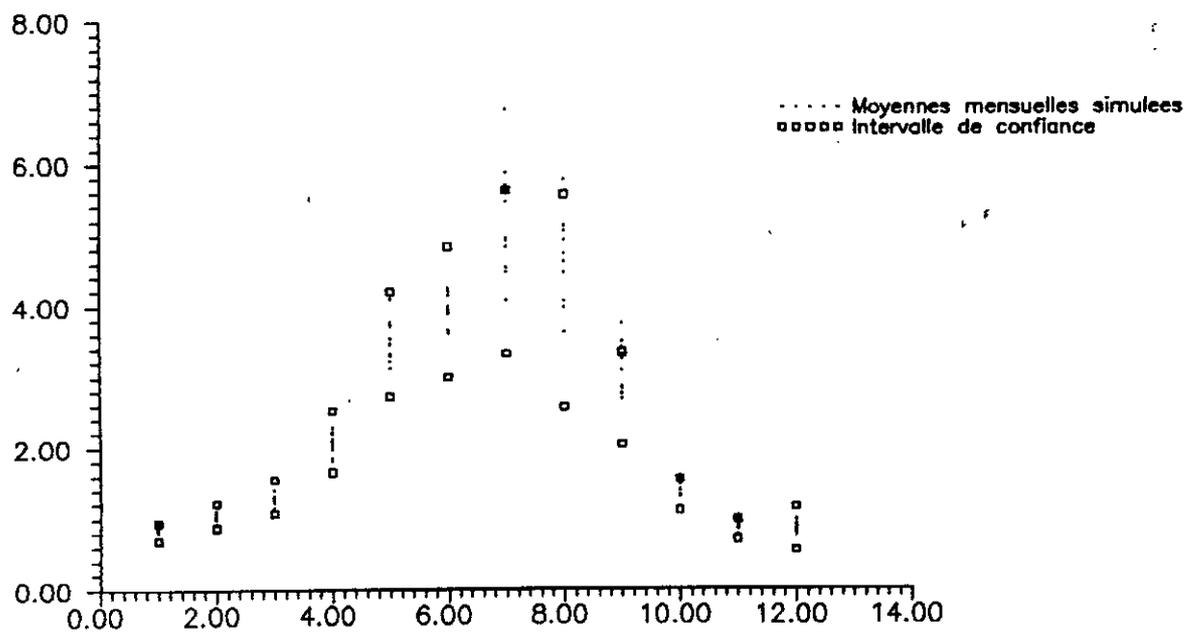
Modele 2 - Transformation 1



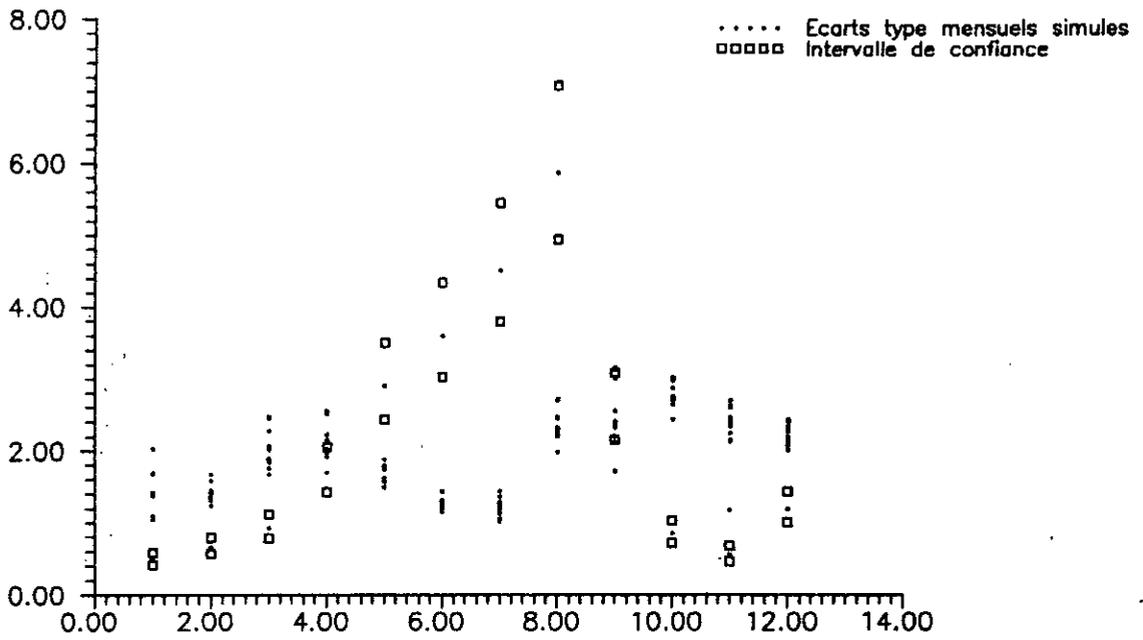
Modele 2 - Transformation 2



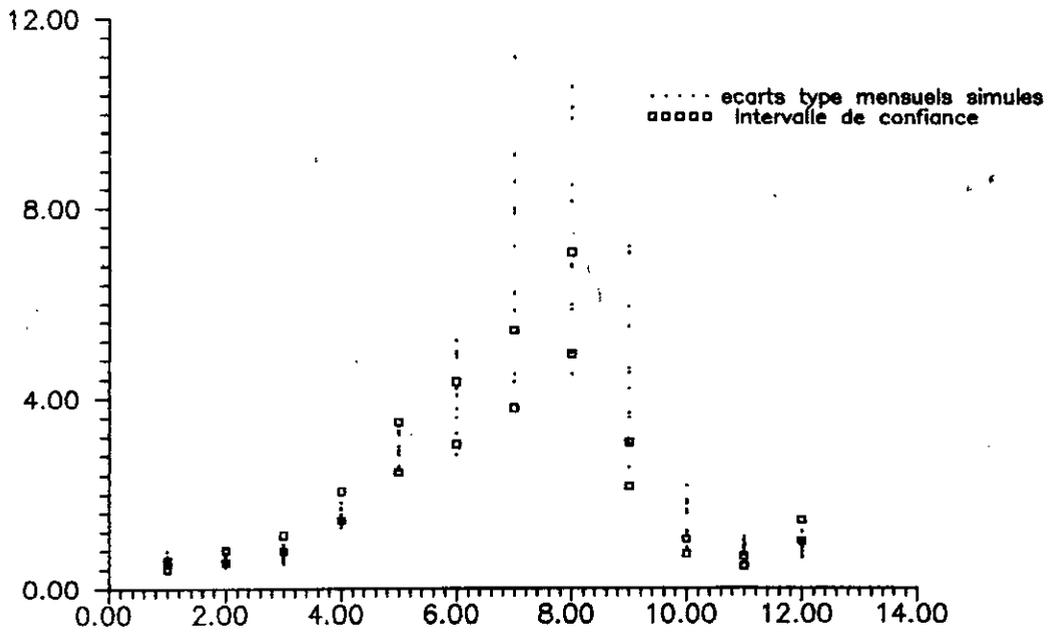
Modele 3 - Transformation 1



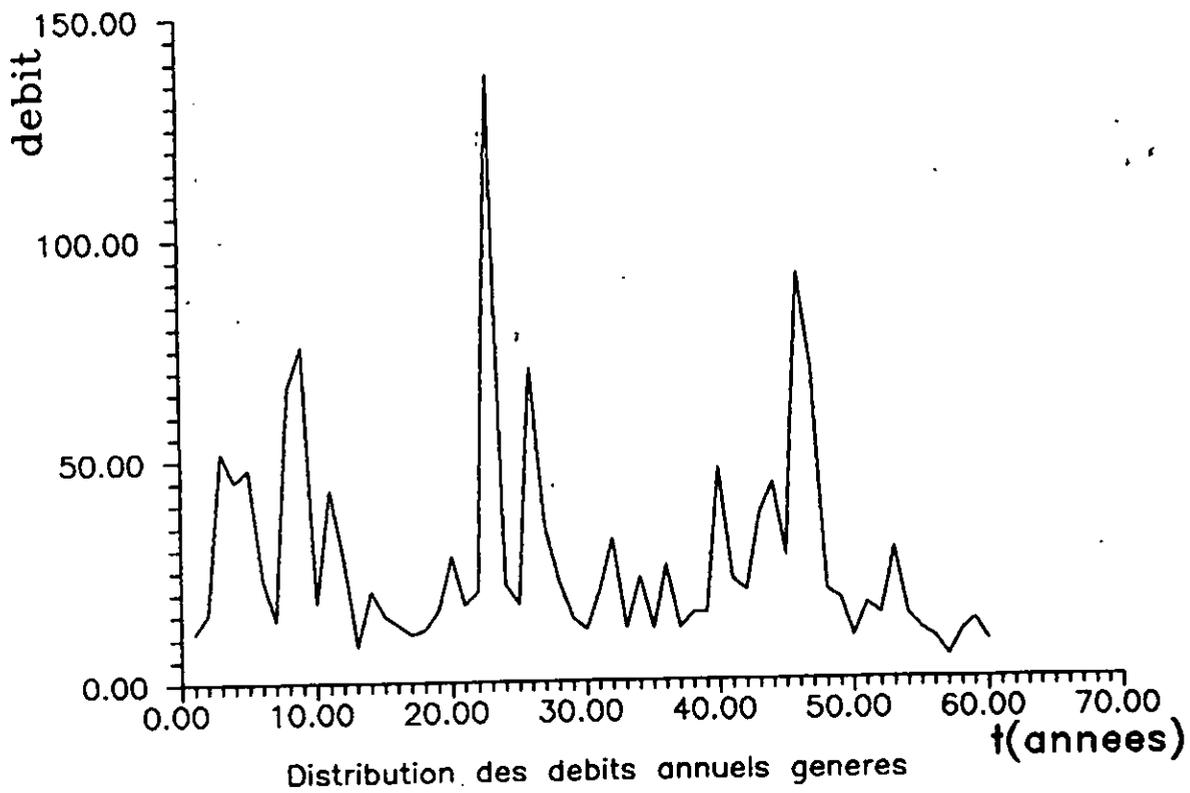
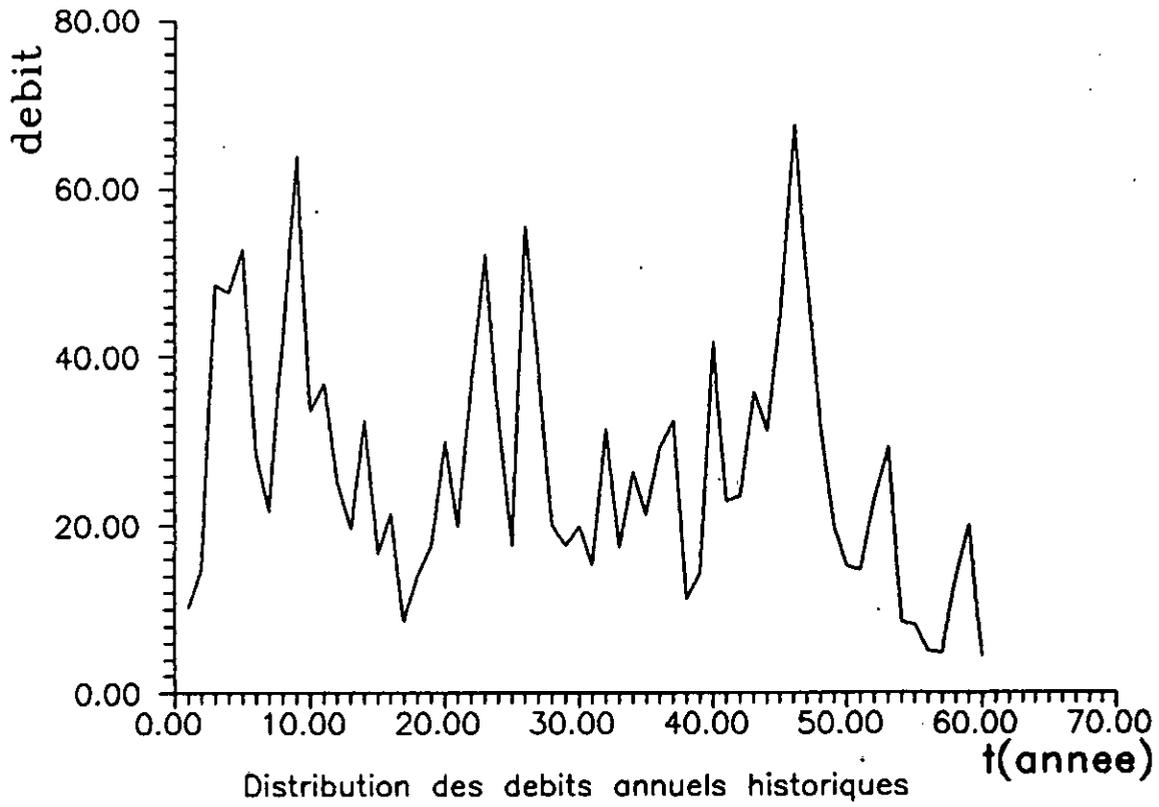
Modele 3 - Transformation 2

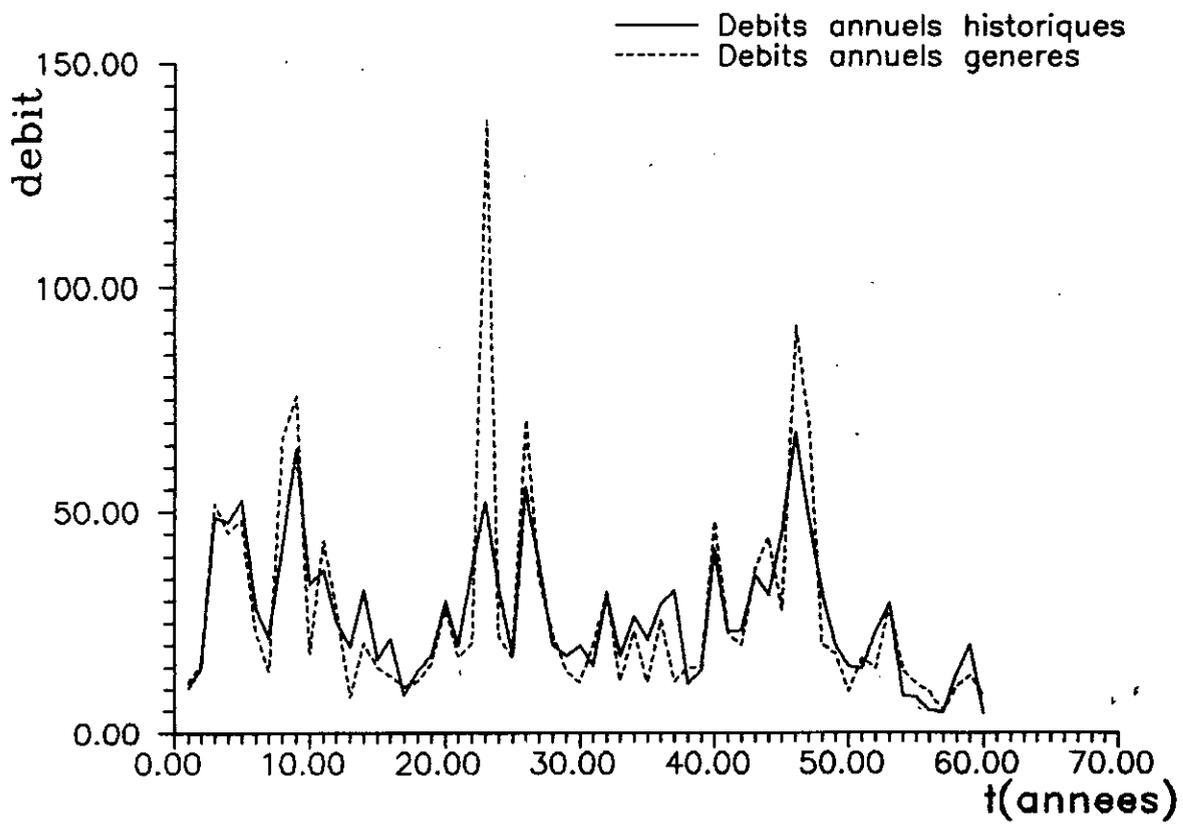


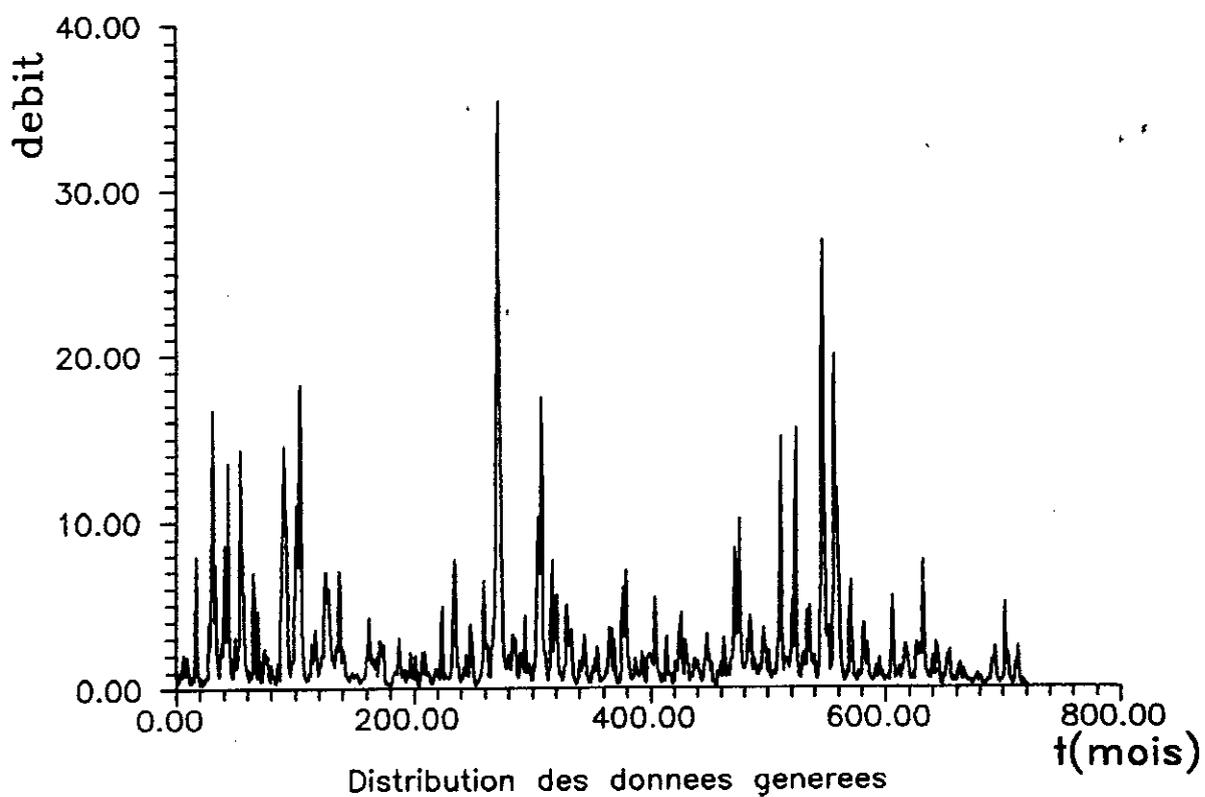
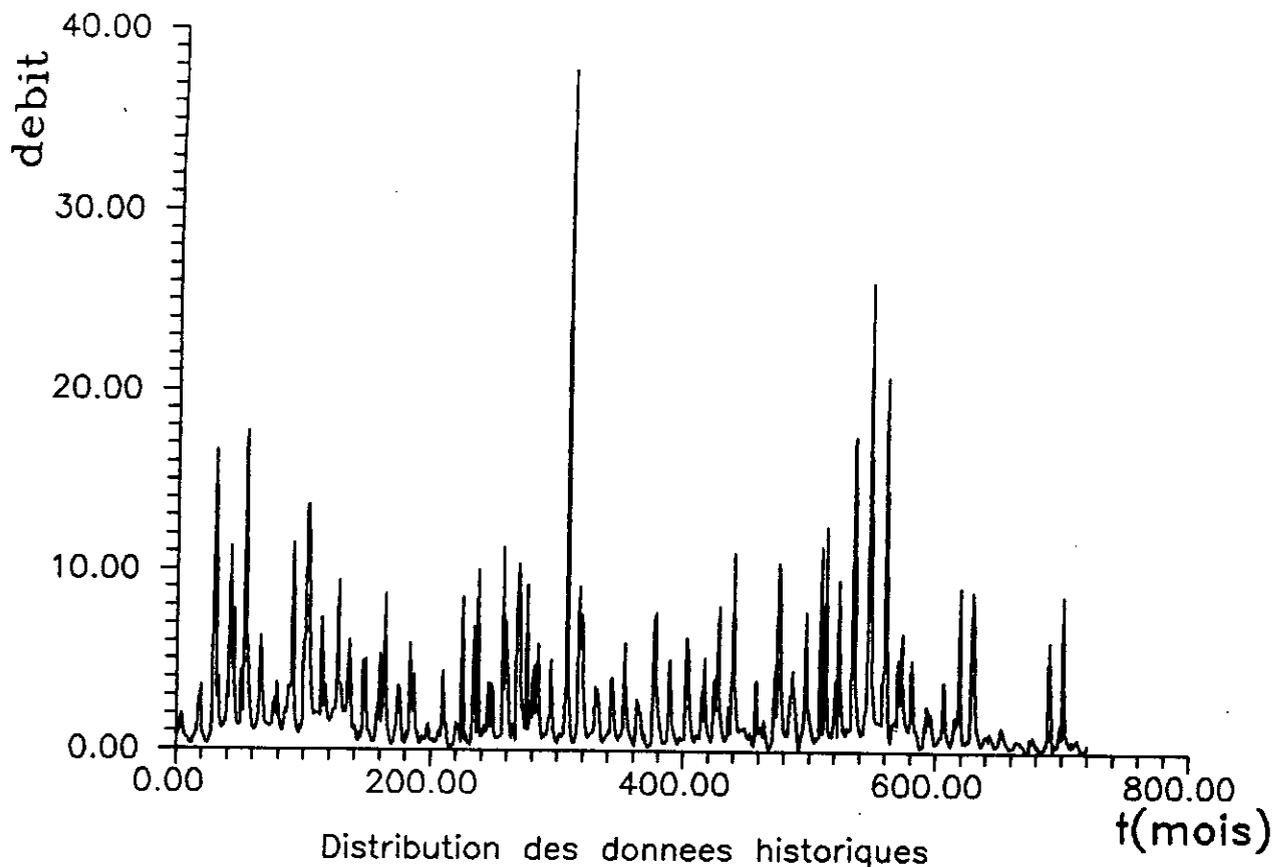
Modele 3 - Transformation 1



Modele 3 - Transformation 2







CONCLUSION

Pour pouvoir confirmer la représentativité du modèle de la série suscitée, il est indispensable de faire d'autres tests sur d'autres caractéristiques statistiques qui sont les corrélations préservées directement et les corrélations préservées indirectement.

Les trois formes du modèle ont donné plus au mois les mêmes résultats.

Un résultat important obtenu est celui de la transformation utilisée, elle a une influence directe sur la fiabilité du modèle, d'où l'importance de cette étape lors de la simulation des débits par un modèle de désagrégation.

L'ensemble des résultats obtenus était satisfaisant, pour surmonter le problème de non préservation d'une caractéristique statistique il suffit de faire quelques ajustements sur les données générées.

cette étude n'est qu'une première vue sur la simulation des données par un modèle de désagrégation et la capacité de ce dernier à reproduire les caractéristiques observées, le domaine de recherche reste donc ouvert surtout que ce type de technique fait appel à l'expérience.

ANNEXES

ANNEXE A: NOTIONS STATISTIQUES

ANNEXE B: NOTATIONS

ANNEXE C: CALCUL DES COVARIANCES

ANNEXE D: DETERMINATION DE B (TEL QUE $BBT=D$)

NOTIONS DE STATISTIQUES

L'étude des séries chronologiques et la génération des données nécessite au préalable la connaissance des éléments de base de la statistique que nous présentons ici.

Considérons un échantillon d'une série chronologique de N variables (X_t , $t=1,N$).

CARACTERISTIQUES GLOBALES

LA MOYENNE

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$$

\bar{X} est l'estimé de la moyenne de la population μ , \bar{X} mesure la tendance centrale de la série.

LA VARIANCE

$$S^2 = \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2$$

L'estimé de la variance de la population σ^2 , noté s^2 mesure la dispersion de la série. On applique un ajustement pour rendre cet estimateur non biaisé en remplaçant le dénominateur par N-1 le rapport \bar{X}/s^2 est dit le coefficient de variation.

LE COEFFICIENT D'ASYMETRIE

$$g = \frac{\frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^3}{s^3}$$

Pour avoir un estimateur non biaisé du coefficient d'asymétrie de la population ρ , g devient:

$$g = \frac{N \sum_{t=1}^N (X_t - \bar{X})^3}{(N-1)(N-2) s^3}$$

Le coefficient g mesure la dissymétrie de la série.

LA FONCTION D'AUTO-COVARIANCE

La fonction d'autocovariance mesure le degré de l'autodépendance linéaire de la série chronologique. L'autocovariance entre les deux valeurs X_t et X_{t+k} est donnée par la relation suivante:

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})$$

pour $0 \leq k \leq N$

C_k est appelée l'autocovariance à pas k , k étant le décalage entre l'ordre des deux valeurs. Pour le cas particulier $k=0$, C_0 devient la variance de la population s^2 .

L'estimé non biaisé de l'autocovariance de la population est donné en remplaçant le dénominateur par $(N-k)$.

LA FONCTION D'AUTOCORRELATION

Elle est définie comme étant le rapport entre l'Autocovariance d'ordre k et l'Autocovariance d'ordre 0. Ce rapport mesure la dépendance linéaire (ACF):

$$r_k = \frac{C_k}{C_0}$$

Le tracé de r_k en fonction du pas k donne ce qu'on appelle le correlogramme.

La correction non biaisée de la fonction d'autocorrelation r_k est donnée par la relation de KENDAL-STUART 1968 :

$$r'_k = 2r_k - 0.5 [r_k(1) + r_k(2)]$$

r_k et r'_k étant les coefficient d'autocorrelation de la première et la seconde moitié de la série chronologique.

En plus de ceci, la série chronologique est caractérisée par la fonction de distribution de ses variables. Les fonction les plus utilisées sont la loi normale, la loi log-normale, et la loi gama.

QUELQUES CARACTERISTIQUES PERIODIQUES

Ce sont des caractéristiques déterminées pour chaque intervalle de l'année .

considérons une séries chronologique périodique $X_{v,\tau}$ où v représente l'année et τ représente la saison (l'intervalle de l'année).

#La moyenne :

$$\bar{X}_\tau = \frac{1}{N} \sum_{v=1}^N X_{v,\tau} \quad \tau = 1, \dots, \omega$$

#La variance :

$$s_\tau^2 = \frac{1}{N} \sum_{v=1}^N (X_{v,\tau} - \bar{X}_\tau)^2$$

LES NOTATIONS

Soit un modèle de désagrégation de la forme suivante:

$$Y = AX + Be + CZ$$

Les notations utilisées sont suivantes :

$$Y_{v,\tau}^{(i)}$$

$$X_v^{(i)}$$

tel que :

$$i = 1, \dots, n$$

$$v = 1, \dots, N$$

$$\tau = 1, \dots, w$$

Où:

n est le nombre de sites (ou de séries chronologiques)

v est le nombre d'années

w est le nombre de saisons

CALCUL DES COVARIANCES

Considérons le cas d'une désagrégation trimestrielle pour deux sites ($n=2$ et $w=3$) afin d'illustrer le calcul d'estimation des covariances S_{XX} , S_{YY} , S_{YZ} et S_{ZZ}

LA COVARIANCE S_{XX}

$$S_{XX} = E(XX^T) = \frac{1}{N} \sum_{v=1}^N (XX^T)$$

Pour un nombre de séries chronologiques $n = 2$:

$$S_{XX} = E \begin{bmatrix} X_v^{(1)} \\ X_v^{(2)} \end{bmatrix} \begin{bmatrix} X_v^{(1)} & X_v^{(2)} \end{bmatrix}$$

$$S_{XX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} X_v^{(1)} \\ X_v^{(2)} \end{bmatrix} \begin{bmatrix} X_v^{(1)} & X_v^{(2)} \end{bmatrix}$$

$$S_{XX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} X_v^{(1)} & X_v^{(1)} & X_v^{(1)} & X_v^{(2)} \\ X_v^{(2)} & X_v^{(1)} & X_v^{(2)} & X_v^{(2)} \end{bmatrix}$$

le dénominateur est remplacé par $N-1$ pour avoir un estimateur non biaisé.

LA COVARIANCE S_{YX}

$$S_{YX} = E(YX^T)$$

$$S_{YX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} Y_{v,1}^{(1)} \\ Y_{v,2}^{(1)} \\ Y_{v,3}^{(1)} \\ \dots \\ Y_{v,1}^{(2)} \\ Y_{v,2}^{(2)} \\ Y_{v,3}^{(2)} \end{bmatrix} \begin{bmatrix} X_v^{(1)} & X_v^{(2)} \end{bmatrix}$$

LA COVARIANCE S_{YY}

$$S_{YY} = E(YY^T)$$

$$S_{YY} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} Y_{v,1} \\ Y_{v,2} \\ Y_{v,3} \\ \dots \\ Y_{v,1} \\ Y_{v,2} \\ Y_{v,3} \end{bmatrix} \begin{bmatrix} Y_{v,1} & Y_{v,2} & Y_{v,3} & \dots & Y_{v,1} & Y_{v,2} & Y_{v,3} \end{bmatrix}$$

LA MATRICE Z

Les éléments de la matrice Z sont les valeurs saisonnières correspondants aux années précédant l'année courante. Si on considère le cas de deux saisons, Z s'écrira sous la forme suivante :

$$Z_v = \begin{bmatrix} Y_{v-1,2}^1 \\ Y_{v-1,3}^1 \\ \dots \\ Y_{v-1,2}^2 \\ Y_{v-1,3}^2 \end{bmatrix}$$

v=1,N

LA COVARIANCE S_{ZZ}

$$S_{ZZ} = \frac{1}{N} \sum_{v=1}^N (Z_v Z_v^T)$$

LA COVARIANCE S_{ZX}

$$S_{ZX} = \frac{1}{N} \sum_{v=1}^N (Z_v X_v^T)$$

$$S_{ZX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} Y_{v-1,2}^1 \\ Y_{v-1,3}^1 \\ \dots \\ Y_{v-1,2}^2 \\ Y_{v-1,3}^2 \end{bmatrix} \begin{bmatrix} X_v^{(1)} & X_v^{(2)} \end{bmatrix}$$

LA COVARIANCE S_{ZY}

$$S_{ZY} = \frac{1}{N} \sum_{v=1}^N (Z_v Y_v^T)$$

$$S_{ZY} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} Y_{v-1,2}^1 \\ Y_{v-1,3}^1 \\ \dots \\ Y_{v-1,2}^2 \\ Y_{v-1,3}^2 \end{bmatrix} [Y_v^{(1)} \quad Y_v^{(2)}]$$

PROCEDE DE CALCUL POUR LA DETERMINATION DE B

Posons $BB^T = D$, B étant une matrice triangulaire inférieure.

Si D est définie positive

Les éléments non nuls de la matrice sont:

$$B(i, j) = D(i, j) / B(i, j) \quad \text{si } j=1, i=1, \dots, n$$

$$B(i, j) = \left[D(i, j) - \sum_{k=1}^{j-1} B(j, k) B(i, k) \right]^{1/2} \quad \text{si } i=j, j=2, \dots, n$$

$$B(i, j) = D(i, j) - \sum_{k=1}^{j-1} B(j, k) B(i, k) / B(j, j) \quad \text{si } j=2, \dots, n-1 \\ i=j+1, \dots, n$$

Si D est semi définie positive

Pour tout $k < 0$: $B(k, i) = 0$

Pour tout $k \geq 0$: $B(k, i) = 0$ si

$$D(i, i) - \sum_{j=1}^i B(i, j)^2 \leq 0$$

$$B(k, i) = \frac{D(k, i) - \sum_{j \neq i} B(i, j) B(k, j)}{\sqrt{D(i, i) - \sum_{j \neq i} B(i, j)^2}}$$

si

$$D(i, i) - \sum_{j \neq i} B(i, j)^2 > 0$$

-
- [10] . Tao, P. C., Delleur, J. W., 1976. Multistation, multiyear synthesis of hydrologic time series by disaggregation. Water Resource. 12, 6, pp. 1303-1312
- [11] . Valencia, D. R., Schaake, J. C., 1973. Disaggregation process in stochastic hydrology. Water Resources. 9, 3, pp. 580-585.
- [12] . Curry, K., Bras, R. L., 1978. Theory and applications of the multivariate broken line, disaggregation, and monthly AR streamflow generators to the Nile River. Technology adaptation program report 78-5, Massachusetts, 416 p., September.

-
- [1] . Lane, W. L., 1979. Applied stochastic techniques (Last computer package), User manual.
- [2] . Llamas, Hydrologie générale, Application
- [3] . Matalas, N. C., J. R. Wallis. Generation of synthetic flow sequences, in system approach to water Management, A. K. Biswas (ed.) Mc Graw-Hill Book Company, New York, 1976.
- [4] . Mejia, J. M., Rouselle, J., 1976. Disaggregation models in hydrology revisited. Water resource. 12, 2, pp. 185-186.
- [5] . Salas, J. D., Delleur, J. W., Yevjevich, V., Lane, W. L. Applied modeling of hydrologic time series. Water resources publication, 1980.
- [6] . Salas, J. D., Smith, R. A., 1980. Uncertainties in hydrologic time series analysis. Paper presented at the ASCE Spring meeting, Portland, Oregon, Preprint. 80-158.
- [7] . Snedecor, G. W., Cochran, W. G., 1967. Statistical methods, the Iowa State University Press, Iowa.
- [8] . Souag, D., Contribution à la gestion d'un réservoirs par les modèles de règles de décision linéaires (LDR et modélisation des débits, 1993. Thèse de magisters
- [9] . Stedinger, J. R., Vogel, R. M., 1984. Disaggregation procedure for generating serially correlated flows vectors. Water Resources. 20, 1, pp 47-56.

INTRODUCTION	01
I.LA SIMULATION	
I.1. Généralités	03
I.1.1. Définition	03
I.1.2. Génération des données	03
I.1.3. Modèles stochastiques de génération des données ..	04
I.1.4. Application simple de la simulation	04
I.1.5. La modélisation en hydrologie	07
Organigramme générale	
I.2. Modèles de simulation	11
I.2.1. Modèle autorégressif	11
I.2.2. Modèle ARMA	13
I.2.3. Modèles multidimensionnels	14
I.3. Traitement des données	18
I.4. Estimation des paramètres	23
I.5. Tests des modèles	29
II.MODELES DE DESAGREGATION	37
II.1. Historique	38

II.2. Description générale des modèles	39
II.2.1. Formulation mathématique du modèle	40
II.2.2. Modèle de désagrégation à site unique	40
II.2.2.1. Modèle de base	41
II.2.2.2. Modèle étendu	41
II.2.2.3. Modèle condensé	42
II.2.3. Modèle temporel multisite	43
II.3. Propriétés du modèle	44
II.4. Les phases de la désagrégation	45
II.5. Estimation des paramètres	47
II.6. Tests du modèle	50
RESULTATS	52
CONCLUSION	69
ANNEXES	70
REFERENCES	81
TABLES DES MATIERES	83