

Record Number: 11960
Author, Monographic: Abi-Zeid, I./Mathier, L./Bobée, B.
Author Role:
Title, Monographic: La génération de données synthétiques: comparaison des approches directes et indirectes
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1996
Original Publication Date: Juin 1996
Volume Identification:
Extent of Work: 92
Packaging Method: pages incluant un annexe
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, rapport de recherche
Series Volume ID: 467
Location/URL:
ISBN: 2-89146-441-9
Notes: Rapport annuel 1996-1997
Abstract: 15.00\$
Call Number: R000467
Keywords: rapport/ ok/ dl

**LA GÉNÉRATION DE DONNÉES
SYNTHÉTIQUES: COMPARAISON DES
APPROCHES DIRECTES ET INDIRECTES**

**LA GÉNÉRATION DE DONNÉES SYNTHÉTIQUES:
COMPARAISON DES APPROCHES DIRECTES ET INDIRECTES**

par

**Irène Abi-Zeid
Louis Mathier
Bernard Bobée**

**Institut national de la recherche scientifique, INRS-Eau
2800, rue Einstein, Case postale 7500, SAINTE-FOY (Québec), G1V 4C7**

Rapport de recherche No R-467

Juin 1996

RÉSUMÉ

Ce travail a pour but d'effectuer une comparaison des approches directes et indirectes de la génération de données synthétiques hydrologiques à un site. Les méthodes directes génèrent directement des données saisonnières (mensuelles, hebdomadaires). Les méthodes indirectes génèrent en un premier lieu des données annuelles, et elles utilisent ensuite des méthodes de désagrégation pour en arriver aux données saisonnières. La méthodologie de génération de données synthétiques par une approche directe (modèles ARMA) est présentée en détail. Cinq modèles mensuels sont étudiés: des modèles ARMA à paramètres constants avec et sans pondération annuelle, des modèles ARMA à paramètres variables avec et sans pondération annuelle, et un modèle de désagrégation (SPIGOT). Les données utilisées sont les apports nets mensuels sur le lac Supérieur entre 1900 et 1989.

Afin d'évaluer les performances des diverses méthodes, des mesures d'erreurs quantifiant l'écart entre les moyennes des caractéristiques statistiques mensuelles et annuelles des données générées, et les caractéristiques statistiques mensuelles et annuelles des données historiques sont calculées. Les cinq modèles sont comparés par rapport à ces mesures d'erreur.

Il est difficile de trancher d'une façon générale laquelle des approches, directe ou indirecte est meilleure pour générer des données synthétiques. Dans ce travail, l'approche directe sans pondération s'est avérée globalement meilleure pour préserver les caractéristiques mensuelles et l'approche indirecte s'est avérée meilleure pour préserver les caractéristiques annuelles. Nos conclusions sont basées strictement sur des caractéristiques statistiques.

TABLE DES MATIÈRES

1. INTRODUCTION.....	1
2. LES MÉTHODES DE GÉNÉRATION DES DONNÉES SYNTHÉTIQUES.....	3
2.1 Le choix des caractéristiques statistiques de validation.....	5
3. LA MODÉLISATION DE LA SÉRIE HYDROLOGIQUE.....	9
3.1 Les tests de normalité et les tests d'indépendance.....	9
3.2 La normalisation et la désaisonnalisation des données.....	12
4. MÉTHODOLOGIE.....	15
4.1 L'analyse préliminaire et l'identification du modèle.....	15
4.2 La construction du modèle.....	16
4.2.1 La modélisation ARMA.....	16
4.2.1.1 L'identification.....	17
4.2.1.2 L'estimation des paramètres du modèle.....	17
4.2.2 La modélisation de séries périodiques à paramètres variables.....	17
4.3 La vérification du modèle et l'analyse des résidus.....	18
5. LE MODÈLE DE DÉSAGRÉGATION.....	21
6. APPLICATION.....	23
6.1 La série annuelle.....	23
6.2 La série mensuelle.....	32
6.2.1 L'approche Directe.....	34
6.2.1.1 ARMA à paramètres constants.....	35
6.2.1.2 ARMA à paramètres variables.....	43
6.2.1.3 La pondération des données mensuelles générées.....	46
6.2.2 L'approche indirecte.....	46
7. RÉSULTATS.....	49
7.1 Les caractéristiques statistiques de cette étude.....	49
7.2 Comparaison des performances des modèles.....	63
8. CONCLUSION.....	69
9. BIBLIOGRAPHIE.....	71
10. ANNEXE 1.....	75

LISTE DES TABLEAUX

Tableau 1: Les statistiques de la série annuelle historique x_t	23
Tableau 2: Les statistiques annuelles de la série z_t	25
Tableau 3: Les moyennes mensuelles périodiques de $x_{v,\tau}$	33
Tableau 4: Les écarts types mensuels périodiques de $x_{v,\tau}$	33
Tableau 5: Les moyennes mensuelles périodiques de $y_{v,\tau}$	33
Tableau 6: Les écarts types mensuels périodiques de $y_{v,\tau}$	33
Tableau 7: Les coefficients d'asymétrie périodiques de $y_{v,\tau}$	34
Tableau 8: Les paramètres du modèle AR(2).....	39
Tableau 9: Les paramètres du modèle AR(1)V périodique à paramètres variables.....	43
Tableau 10: Les autocorrélations périodiques $r_{1,\tau}(\hat{\xi}_\tau)$ des résidus du modèle AR(1)V.....	44
Tableau 11: Les modèles de désagrégation et les seuils de transformation.....	47
Tableau 12: Les statistiques annuelles historiques et générées.....	50
Tableau 13: Les statistiques de surplus historiques et générées.....	50
Tableau 14: Les statistiques mensuelles générées, AR(2)np.....	51
Tableau 15: Les statistiques mensuelles générées, AR(2)p.....	51
Tableau 16: Les statistiques mensuelles générées, AR(1)Vnp.....	52
Tableau 17: Les statistiques mensuelles générées, AR(1)Vp.....	52
Tableau 18: Les statistiques mensuelles générées, SPIGOT.....	53
Tableau 19: Les valeurs absolues des erreurs relatives des statistiques mensuelles.....	65
Tableau 20: Les valeurs absolues des erreurs absolues des coefficients de corrélation mensuels.....	66
Tableau 21: Les erreurs relatives des statistiques annuelles.....	67
Tableau 22: La comparaison des erreurs moyennes des statistiques mensuelles.....	67
Tableau 23: La comparaison des erreurs moyennes globales des statistiques annuelles.....	67
Tableau 24: La comparaison des erreurs moyennes globales des statistiques mensuelles.....	68

LISTE DES FIGURES

Figure 1: La droite de Henri des données annuelles historiques, x_t	24
Figure 2: Les apports nets annuels historiques du lac Supérieur.....	25
Figure 3: Le corrélogramme des données annuelles transformées, z_t	26
Figure 4: Le corrélogramme partiel des données annuelles transformées, z_t	27
Figure 5: Le périodogramme cumulatif des données annuelles transformées, z_t	28
Figure 6: Le corrélogramme des résidus du modèle AR(0) pour z_t	30
Figure 7: La droite de Henri résidus du modèle AR(0) pour z_t	31
Figure 8: Les données historiques mensuelles $x_{v,\tau}$	32
Figure 9: Les autocorrélations périodiques d'ordre 1 $r_{1,\tau}$ de $z_{v,\tau}$	34
Figure 10: Les autocorrélations périodiques d'ordre 2 $r_{2,\tau}$ de $z_{v,\tau}$	35
Figure 11: Le corrélogramme des données mensuelles transformées, z_t	36
Figure 12: Le corrélogramme partiel des données mensuelles transformées, z_t	37
Figure 13: Le périodogramme cumulatif des données mensuelles transformées, z_t	38
Figure 14: Le corrélogramme des résidus du modèle mensuel AR(2).....	40
Figure 15: Le périodogramme cumulatif des résidus du modèle mensuel AR(2).....	41
Figure 16: La droite de Henri des résidus du modèle mensuel AR(2).....	42
Figure 17: La droite de Henri des résidus du modèle AR(1)V.....	45
Figure 18: Les autocorrélations annuelles, AR(2)np.....	53
Figure 19: Les autocorrélations annuelles, AR(0).....	54
Figure 20: Les autocorrélations annuelles, AR(1)Vnp.....	54
Figure 21: Les autocorrélations annuelles, AR(0) SPIGOT.....	55
Figure 22: Le graphique des fréquences des données annuelles AR(2)np.....	56
Figure 23: Le graphique des fréquences des données annuelles AR(0).....	56
Figure 24: Le graphique des fréquences des données annuelles AR(1)Vnp.....	57
Figure 25: Le graphique des fréquences des données annuelles SPIGOT.....	57
Figure 26: Le graphique de la période de retour des données annuelles AR(2)np.....	58
Figure 27: Le graphique de la période de retour des données annuelles AR(0).....	58
Figure 28: Le graphique de la période de retour des données annuelles AR(1)Vnp.....	59
Figure 29: Le graphique de la période de retour des données annuelles SPIGOT.....	59
Figure 30: Les autocorrélations mensuelles d'ordre 1, AR(1)Vp et AR(1)Vnp.....	60
Figure 31: Les autocorrélations mensuelles d'ordre 1, AR(2)p et AR(2)np.....	61
Figure 32: Les autocorrélations mensuelles d'ordre 1, SPIGOT.....	61
Figure 33: Les autocorrélations mensuelles d'ordre 2, AR(1)Vp et AR(1)Vnp.....	62
Figure 34: Les autocorrélations mensuelles d'ordre 2, AR(2)p et AR(2)np.....	62
Figure 35: Les autocorrélations mensuelles d'ordre 2, SPIGOT.....	63

1. INTRODUCTION

La seule utilisation de données historiques lors de la conception d'ouvrages hydrauliques ne permet pas de tester les divers plans de conception et les différentes politiques de gestion par rapport à une gamme d'événements futurs possibles. La raison est que les données observées ne sont qu'une des réalisations possibles du processus stochastique. La génération de données synthétiques permet d'obtenir des séries hydrologiques qui seront vraisemblablement rencontrées durant la période de planification. Il est alors possible de déterminer les paramètres de conception des ouvrages pour chaque série de données, et d'obtenir une distribution des valeurs de ces paramètres. On peut ainsi mieux comprendre la variabilité et la plage de variabilité des performances futures des systèmes conçus.

La génération de données synthétiques est une technique qui permet de simuler des données hydrologiques qui ressemblent aux données historiques observées et qui en possèdent les caractéristiques statistiques. Les données synthétiques n'améliorent pas la qualité des données. Elles ne créent pas d'information additionnelle mais elles permettent une exploitation complète de l'information disponible en générant différents scénarios de données.

Plusieurs méthodes de génération existent, mais les deux approches principales consistent en l'approche directe et l'approche indirecte. Les méthodes directes génèrent directement des données saisonnières (mensuelles, hebdomadaires). Les méthodes indirectes génèrent en un premier lieu des données annuelles, et elles utilisent ensuite des méthodes de désagrégation pour en arriver aux données saisonnières. Un des buts de ce travail est d'effectuer une comparaison entre ces deux approches, et ce pour des modèles univariés. L'autre but est de développer une méthodologie formelle à suivre; une description détaillée des étapes est donc présentée.

Les données utilisées sont les apports nets mensuels sur le lac Supérieur (*Net Basin Supply*, NBS) qui représentent la quantité d'eau qu'un lac reçoit des précipitations sur sa surface ainsi que sur son bassin versant moins l'évaporation nette de la surface du lac pendant un mois. Deux méthodes de calculs peuvent être utilisées pour calculer les apports nets: la méthode de calcul par bilan hydrique et la méthode de calcul par composante (Mathier *et al.*, 1992).

Les méthodes directes de génération de données mensuelles sont représentées dans ce travail par quatre modèles univariés de la famille ARMA qui s'ajustent le mieux aux données historiques, selon la méthodologie proposée par Box et Jenkins (1976): AR(2) à paramètres constants avec et sans pondération, et AR(1) à paramètres variables avec et sans pondération. La méthode indirecte est représentée par le modèle de désagrégation du logiciel SPIGOT (Grygier et Stedinger, 1991), dans sa version univariée. Trente séries de 90 ans sont générées par chacune des méthodes et les caractéristiques statistiques jugées importantes et devant être reproduites sont comparées avec les caractéristiques des données historiques.

Afin d'évaluer les performances des diverses méthodes, des mesures d'erreurs quantifiant l'écart entre les moyennes des caractéristiques statistiques mensuelles et annuelles des données générées, et les caractéristiques statistiques mensuelles et annuelles des données historiques sont calculées. Les cinq modèles sont comparés par rapport à ces mesures d'erreur.

2. LES MÉTHODES DE GÉNÉRATION DES DONNÉES SYNTHÉTIQUES

D'après McMahon et Mein (1986), Hazen (1914) a été le premier à reconnaître l'importance de simuler des données hydrologiques. Il combina les séries de débits annuels observés de 14 rivières pour obtenir un ensemble synthétique de 300 années. À partir d'une série historique de débits Sudler (1927) choisit 50 valeurs représentatives du débit annuel qu'il inscrivit sur 50 morceaux de carton. Ces cartons furent brassés et tirés au hasard sans remise pour obtenir une suite de 50 années de débits. Cette expérience fut répétée 20 fois. Malgré les lacunes de cette approche, celle-ci fut probablement la première tentative d'élaboration d'un modèle de génération de données stochastiques. Barnes (1954) trouva que les débits annuels étaient distribués selon une loi normale et utilisa une approche Monte Carlo pour générer 1000 années de données. Il était le premier à utiliser une procédure qui tient compte d'une distribution statistique des débits. Entre 1936 et le milieu des années 1960, Hurst (1965) étudia le Nil et développa des techniques d'échantillonnage pour générer des débits annuels. Toutefois, la modélisation stochastique prit réellement son envol dans les années 60 et plusieurs méthodes virent le jour (Salas *et al.*, 1980). Les modèles stochastiques se présentent sous plusieurs formes: modèle autorégressif (Matalas, 1967), modèle *Fractional Gaussian Noise* (FGN) (Mandelbrot et Wallis, 1968), modèle ARMA (Carlson *et al.*, 1970), modèle *Broken Line* (BL) (Mejia *et al.*, 1972), modèle *Shot-Noise* (Weiss, 1973), modèle de Processus Intermittents (Yakowitz, 1973), modèle *Markov Mixture* (Jackson, 1975), modèle *ARMA-Markov Mixture* (Lettenmaier et Burges, 1977), modèle *Pattern Recognition* (Panu, 1978), et modèle *General Mixture* (Boes et Salas, 1978). Bien que chaque modèle ait ses avantages et ses désavantages, ils ont tous été critiqués pour une ou plusieurs des raisons suivantes: incapacité de reproduire la dépendance à long terme, incapacité de reproduire la dépendance à court terme, difficulté d'estimation des paramètres, manque de justification physique, trop de paramètres.

La plupart des modèles ont des versions univariées (série chronologique unique à un seul site) et multivariées (séries chronologiques multiples, à plusieurs sites, ou séries multiples à un seul site). Les modèles d'intérêt dans cette étude sont univariés, la génération de données synthétiques étant effectuée pour un seul site, le lac Supérieur. Les séries hydrologiques se distinguent par le pas de temps utilisé; les deux types traités dans cette étude sont les séries périodiques dont le pas de temps est une fraction de l'année

(mensuelles) exhibant ainsi une cyclicité annuelle, et celles dont le pas de temps est d'un an (annuelles), supposées stationnaires.

Les méthodes étudiées peuvent être classées en deux grandes catégories: les méthodes directes et les processus de désagrégation (méthode indirecte). Les méthodes directes génèrent directement des données saisonnières, et ont souvent des difficultés à préserver les caractéristiques annuelles des séries de données. Les méthodes indirectes génèrent des données annuelles utilisées par des méthodes de désagrégation pour en arriver aux données saisonnières. La modélisation par désagrégation est un processus par lequel des séries chronologiques saisonnières sont générées à partir d'une série annuelle déjà disponible. Typiquement, la série annuelle est générée dans un premier temps par un modèle stochastique, et elle est ensuite désagrégée en sous-séries. La désagrégation peut être temporelle: une série annuelle est désagrégée en sous-séries mensuelles, ou spatiale: une série de débit d'une rivière est désagrégée en sous-séries de débits de ses tributaires. La désagrégation a le mérite de préserver les caractéristiques statistiques à plusieurs niveaux. Elle permet souvent la réduction du nombre des paramètres par rapport à l'approche directe et augmente la flexibilité des méthodes de génération de données. Harms et Campbell (1967) ont proposé la première approche de désagrégation mais le premier modèle de désagrégation complet est celui de Valencia et Schaake (1973), dont découle le modèle de Mejia et Rousselle (1976). Lane (1979) a apporté des améliorations à cette méthode dans son logiciel LAST, et Grygier et Stedinger (1988) en ont fait de même dans le logiciel SPIGOT.

Une troisième catégorie de méthodes est celle de l'agrégation. Dans ces modèles, les données annuelles sont calculées à partir de données saisonnières. Vecchia *et al.* (1983) ont démontré que cette approche permet un gain significatif dans l'efficacité d'estimation des paramètres. Toutefois, cette approche n'est pas considérée dans ce travail.

Plusieurs études ont été effectuées afin de comparer les divers modèles de génération de débits (McMahon *et al.*, 1972; Lettenmaier et Burges, 1977; McLeod et Hipel, 1978; Hipel *et al.*, 1979; Lettenmaier et Burges, 1982; Srikanthan et McMahon, 1982; Pereira *et al.*, 1984; McMahon et Mein, 1986; C.E.A., 1990; Rassam *et al.*, 1992; Mathier *et al.*, 1992;). L'expérience acquise a démontré que le choix judicieux de modèles AR et ARMA a produit de bons résultats pour la plupart des applications en hydrologie opérationnelle (Salas *et al.* 1980); c'est pour cette raison que les modèles ARMA sont utilisés comme représentants des méthodes directes dans cette étude. Les travaux de C.E.A. (1990) ainsi

que ceux de Rassam *et al.* (1992), et Mathier *et al.* (1992) nous ont amenés à considérer SPIGOT comme représentant les méthodes de désagrégation.

Le développement et l'utilisation de modèles stochastiques hydrologiques incluent les étapes suivantes (Stedinger et Taylor, 1982):

1. l'obtention des données historiques et autre information;
2. le choix de modèles décrivant les distributions des données hydrologiques et l'estimation de paramètres;
3. le choix d'un modèle de dépendance temporelle et spatiale approprié;
4. la vérification de ce modèle;
5. la validation de ce modèle, et
6. l'utilisation de ce modèle.

Une distinction doit être faite entre les étapes 4 et 5. L'étape 4 consiste à vérifier si le modèle conserve bien les caractéristiques statistiques explicites qu'il est censé reproduire. Cette étape permet la détection des erreurs dans l'élaboration du modèle. La validation du modèle permet de voir si les caractéristiques des données générées sont consistantes avec celles des données historiques. La validation permet donc d'évaluer la pertinence du choix du modèle par rapport au problème étudié.

2.1 Le choix des caractéristiques statistiques de validation

Le choix des caractéristiques fondamentales qui devront être reproduites par les séries générées est très important. Lors de ce choix, on doit tenir compte de l'application ultime des données générées, (par exemple, des niveaux de lac simulés à partir des apports nets générés, pour l'opération d'un système hydrologique), donc des caractéristiques qui sont importantes pour l'opération du système étudié. Il est important de voir si les résultats de l'application sont sensibles aux changements dans le modèle de génération de données.

Alors que doit on reproduire? Malheureusement, il n'y a pas de réponse unique et définitive à cette question. Les caractéristiques statistiques réelles de la population ne sont pas

connues, les valeurs observées ne sont qu'une réalisation parmi les possibilités infinies de réalisation du processus stochastique, et les statistiques estimées de l'échantillon sont des variables aléatoires. Il y a aussi le problème de définition et d'interprétation des caractéristiques statistiques dont les principales sont la moyenne, l'écart type, le coefficient d'asymétrie et l'autocorrélation. La moyenne et l'écart type sont les caractéristiques les moins incertaines. Le coefficient d'asymétrie est par contre très incertain, et la capacité d'un modèle à le reproduire dépend largement de la taille de l'échantillon (Klemes, 1972). Les plages de variation (*range*) dont découle le phénomène de Hurst, et les séries de surplus ou de déficit (*runs*) (série de mesures consécutives inférieures ou supérieures à un seuil donné), sont aussi importantes bien que controversées.

Lettenmaier et Burges (1982) ont exprimé plusieurs critiques face à l'approche traditionnelle de validation, qui consiste à comparer des moyennes de caractéristiques générées avec celles des données historiques, car elle ne reflète pas l'usage ultime des données. Ils proposent des mesures de validation propres à chaque application. Les difficultés principales découlent du fait que les moments synthétiques et historiques ne sont pas indépendants, car les moments historiques et les paramètres des modèles sont estimés à partir du même échantillon historique, introduisant ainsi un biais dans l'estimation des moments. Ils ont souligné l'avantage de présenter les distributions complètes des mesures de validation, et non seulement leur moyenne et leur variance.

En général, la procédure de validation d'un modèle annuel ou mensuel inclut la comparaison de certaines caractéristiques des données générées avec celles des données historiques. Quelques unes de ces caractéristiques sont (McMahon et Mein, 1986):

- la moyenne et la variabilité des statistiques telles que la moyenne annuelle ou mensuelle, l'écart type, le coefficient d'asymétrie, le coefficient de variation, l'autocorrélation, et le coefficient de Hurst;
- les courbes de fréquence-durée basées sur les données générées;
- la moyenne des événements extrêmes;
- les corrélogrammes des événements;
- la moyenne et la variabilité des corrélations spatiales d'ordre 0 et 1, pour une analyse multisite, et

- les caractéristiques des variables reliées à l'usage, comme la capacité d'emmagasinage calculée à partir de séries de débits, ou les niveaux de lacs simulés à partir des apports nets. Par exemple, il s'avère important de conserver les caractéristiques annuelles pour un système d'emmagasinage ayant une longue période critique, tandis qu'il est plus important de préserver les caractéristiques mensuelles pour un site de production d'énergie au fil de l'eau.

3. LA MODÉLISATION DE LA SÉRIE HYDROLOGIQUE

Le processus de modélisation peut être simple ou complexe, selon les caractéristiques de la série historique disponible, et le type du modèle utilisé. En général, ce processus est composé de quatre phases principales:

1. l'identification de la composition du modèle (modèle univarié ou multivarié);
2. le choix du type de modèle (ARMA, FGN ou autre);
3. l'analyse préliminaire et l'identification de la forme du modèle (paramètres périodiques ou constants);
4. la construction du modèle et l'estimation des paramètres.

Les séries peuvent être annuelles $\{x_t\}$ (une mesure par an) où t est l'indice des années variant entre 1 et N , le nombre des années; elles peuvent être périodiques $\{x_{v,\tau}\}$ (mensuelles, journalières, etc.), où τ est l'indice des saisons variant entre 1 et ω , (12 ou 365, par exemple), et v est l'indice des années variant entre 1 et N .

Afin d'utiliser les modèles stochastiques de type ARMA, les données hydrologiques doivent être stationnaires et normalement distribuées. La plupart des données hydrologiques ne sont pas distribuées selon une loi normale, elles sont souvent limitées par zéro (non négatives) ou leurs courbes de fréquence sont asymétriques. De plus, les séries saisonnières sont non stationnaires, elles doivent donc être normalisées et désaisonnalisées.

3.1 Les tests de normalité et les tests d'indépendance

Plusieurs tests existent pour vérifier si les données suivent une loi normale. Ces tests varient en puissance et sont plus ou moins appropriés selon les circonstances (Shapiro *et al.*, 1968). Dans cette étude, le test de la droite de Henri, le test de corrélation de Filliben, le test du Khi-deux, et le test de Kolmogorov-Smirnov sont utilisés.

Le test de la droite de Henri consiste à tracer la distribution empirique de la série sur du papier de probabilité normale, ce qui correspond à un graphique des observations ordonnées en fonction de leurs quantiles normaux (Perreault et Bobée, 1994). L'hypothèse

de la normalité n'est pas rejetée lorsque le graphique correspond à peu près à une droite (vérification visuelle). La corrélation r entre ces deux variables est la statistique de corrélation de Filliben. Pour normaliser des données, on choisit la transformation qui correspond à une distribution ayant la valeur de r la plus élevée (Grygier et Stedinger, 1991).

Le test du Kolmogorov-Smirnov calcule une statistique D_n basée sur le degré de ressemblance des distributions empiriques et hypothétiques (Normale) où n est la taille de l'échantillon. L'hypothèse nulle (normalité) n'est pas rejetée lorsque D_n est plus petite que la valeur critique correspondant à un niveau de signification donné (Hogg et Tanis, 1977).

Le test du Khi-deux consiste à arranger les données en ordre de grandeur croissant et à choisir k classes C_i définies à partir des tables de probabilité normale, telles que la probabilité de chaque classe est $1/k$. N_i est le nombre d'observations qui appartiennent à la classe C_i . Une statistique $\hat{\chi}^2$ est calculée en fonction de la différence entre les fréquences N_i/n et les probabilités correspondantes des classes C_i . L'hypothèse de normalité n'est pas rejetée au niveau α lorsque $\hat{\chi}^2 < \chi^2_{k-2}(1-\alpha)$ (Salas *et al.* 1980).

Plusieurs méthodes existent également pour vérifier l'indépendance d'une série chronologique (Salas *et al.* 1980). Les méthodes utilisées ici sont le test du corrélogramme de Anderson, le test du périodogramme cumulatif, le test du Porte Manteau, et le test de corrélation critique pour les résidus des modèles périodiques à paramètres variables. Pour les deux premiers tests, on doit calculer le corrélogramme, le corrélogramme partiel et le périodogramme cumulatif. Un corrélogramme d'une série stationnaire z_t est le graphique de r_k (l'estimateur du maximum de vraisemblance de ρ_k qui exprime la corrélation entre z_t et z_{t-k}) en fonction de k (l'ordre du coefficient de corrélation); le corrélogramme partiel d'une série est une autre façon de représenter la structure de dépendance temporelle de la série; un corrélogramme partiel d'une série stationnaire z_t est le graphique de $\hat{\phi}_{kk}$ en fonction de k où $\hat{\phi}_{kk}$ est l'estimateur de ϕ_{kk} , le $k^{\text{ième}}$ coefficient autorégressif du modèle AR(k); ϕ_{kk} exprime la corrélation entre z_t et z_{t-k} étant donné $z_{t-1}, \dots, z_{t-k-1}$, c'est à dire lorsqu'on enlève les effets de $z_{t-1}, \dots, z_{t-k-1}$. L'annexe 1 contient les définitions des statistiques utilisées dans cette étude.

Le périodogramme cumulatif d'une série z_t est le graphique de P_i en fonction de h_i qui, lorsque la série est indépendante, est dispersé autour d'une ligne joignant (0,0) et (0.5,1). Une périodicité dans la fréquence apparaît comme une déviation de la ligne droite:

$$P_i = \frac{\sum_{j=1}^i \text{MSD}(h_j)}{\sigma_z^2}, \quad i=1 \dots N$$

où σ_z^2 est la variance de z_t , N est le nombre de données, $h_j=j/N$ est la fréquence, et

$$\text{MSD}(h_j) = \frac{2}{N^2} \left[\left(\sum_{t=1}^N z_t \cos(2\pi h_j t) \right)^2 + \left(\sum_{t=1}^N z_t \sin(2\pi h_j t) \right)^2 \right]$$

$\text{MSD}(h_j)$ est interprétée comme la partie de la variance σ_z^2 qui est expliquée par l'harmonique j . Les périodogrammes sont souvent tracés avec les limites de Kolmogorov-Smirnov qui sont à des distances:

$$\pm \frac{K_\alpha}{N'}, \quad N' = \frac{N-2}{2} \text{ si } N \text{ est pair, ou } N' = \frac{N-1}{2} \text{ si } N \text{ est impair.}$$

K_α est la statistique de Kolmogorov. Lorsqu'un nombre de Points P_i supérieur à αN tombe à l'extérieur des limites, la série n'est pas indépendante.

Le test de Anderson consiste à calculer l'intervalle de confiance du corrélogramme d'une série tel que:

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k}$$

L'hypothèse d'indépendance de la série n'est pas rejetée lorsque le corrélogramme est compris à l'intérieur de l'intervalle.

Le test du Porte Manteau peut être utilisé pour tester si les résidus $\{\varepsilon_t\}$ d'un modèle ARMA(p,q) sont indépendants en calculant Q , tel que $Q = N \sum_{k=1}^L \hat{r}_k^2(\varepsilon)$, où L est entre 10% et 30% de N , et $\hat{r}_k^2(\varepsilon)$ est l'estimateur du maximum de vraisemblance de la fonction d'autocorrélation de la série ε_t . L'hypothèse d'indépendance des résidus d'un modèle ARMA(p,q) n'est pas rejetée lorsque $Q < \hat{\chi}_{(L-p-q)}^2(1-\alpha)$.

Le test de corrélation critique pour les résidus des modèles périodiques à paramètres variables consiste à calculer la valeur critique $r_{1-\alpha/2}$ où t est le coefficient de Student et:

$$r_{1-\alpha/2} = \frac{t(N-2)_{1-\alpha/2}}{\left[N-2 + t^2(N-2)_{1-\alpha/2} \right]^{1/2}}$$

Lorsque les coefficients de corrélation d'ordre 1 des résidus $r_{1,\tau}(\hat{\xi})$ sont inférieurs à $r_{1-\alpha/2}$, l'hypothèse d'indépendance des résidus n'est pas rejetée.

3.2 La normalisation et la désaisonnalisation des données

En général, trois approches sont possibles pour la modélisation des données hydrologiques asymétriques:

1. transformer les données historiques en données normales avant la modélisation;
2. modéliser les données asymétriques et trouver la distribution appropriée des résidus non normaux;
3. trouver une relation entre les deux premiers moments de la série originale asymétrique et ceux de la série normalisée afin de préserver les moments de la série originale.

L'avantage principal de l'approche (1) réside dans le fait que la majorité des modèles statistiques et stochastiques sont développées pour des processus normaux. Il est donc plus simple de normaliser les données que de développer de nouveaux outils. Par contre, la transformation des données introduit un biais dans les propriétés statistiques de la série générée qui ressembleront aux statistiques des données transformées mais pas nécessairement à celles des données originales. La méthode de transformation des moments de Matalas, 1967 (approche 3) est utilisée dans ce travail.

Plusieurs transformations peuvent être utilisées pour obtenir une série normalisée y_t à partir des données originales x_t :

La transformation de Box-Cox:

$$y_i^\lambda = \frac{[(x_i + \text{const})^\lambda - 1]}{\lambda} \quad \lambda \neq 0$$

$$y_i^\lambda = \ln(x_i + \text{const}) \quad \lambda = 0$$

La transformation de racine carrée est souvent appliquée lorsque x_t suit une loi Gamma à deux paramètres: $y_t = \sqrt{x_t}$

La transformation logarithmique

La transformation logarithmique $y_t = \ln(x_t - c)$ est la plus utilisée en hydrologie car elle s'applique lorsque les données x_t suivent une loi Lognormale à 3 paramètres (2 paramètres, lorsque $c = 0$). Pour les données périodiques, le paramètre c peut dépendre de la période, la transformation a alors la forme : $y_{v,\tau} = \ln(x_{v,\tau} - c_\tau)$.

Le seuil c_τ peut être calculé (Stedinger et Taylor, 1982) par:

$$\hat{c}_\tau = \frac{x \max_\tau x \min_\tau - xmed_\tau^2}{x \max_\tau x \min_\tau - 2xmed_\tau}$$

où $x \max_\tau$, $x \min_\tau$, $xmed_\tau$, représentent respectivement le maximum, le minimum et la médiane des données hydrologiques pour la saison τ .

La transformation des moments de Matalas (1967) qui correspond à une transformation logarithmique est utilisée pour normaliser les données. On doit résoudre des équations implicites où \bar{x}_τ , $s_{x\tau}$, $g_{x\tau}$ correspondent respectivement à la moyenne, à l'écart type et au coefficient d'asymétrie de la saison τ de la série originale $x_{v,\tau}$ et \bar{y}_τ , $s_{y\tau}$, $g_{y\tau}$ correspondent à la moyenne, à l'écart type et au coefficient d'asymétrie de la saison τ de la série normalisée $y_{v,\tau}$ telle que:

$$y_{v,\tau} = \ln(x_{v,\tau} - c_\tau), \quad c_\tau = g_{y\tau}$$

$$g_{x\tau} = \frac{\exp[3s_{y\tau}^2] - 3\exp[s_{y\tau}^2] + 2}{\{\exp[s_{y\tau}^2] - 1\}^2}$$

$$s_{x\tau}^2 = \exp\left[2\left(s_{y\tau}^2 + \bar{y}_\tau\right)\right] - \exp\left[s_{y\tau}^2 + 2\bar{y}_\tau\right]$$

$$\bar{x}_\tau = g_{y\tau} + \exp\left(.5s_{y\tau}^2 + \bar{y}_\tau\right)$$

Une question additionnelle (Salas *et al.* 1980) se pose lorsque les données sont périodiques ou lorsque les données annuelles présentent une saisonnalité. Ces composantes périodiques doivent être éliminées. Bien que la différenciation saisonnière élimine complètement la contribution périodique, elle cause une distorsion du spectre et n'est pas recommandée pour la simulation. Il est préférable (Delleur *et al.* 1976) de désaisonnaliser les données en soustrayant la moyenne et en divisant le tout par l'écart type telle que $y'_{v,\tau} = \frac{x_{v,\tau} - \hat{\mu}_{x,\tau}}{\hat{\sigma}_{x,\tau}}$.

À quel moment la normalisation doit-elle être effectuée? Par exemple, la normalisation pourrait être effectuée (a) directement sur la série originale $x_{v,\tau}$ ou (b) sur la série désaisonnalisée $y'_{v,\tau}$. Lorsque l'approche (a) est suivie, les périodicités de la moyenne et de l'écart type de la série $y'_{v,\tau}$ seront déformées par rapport à ceux de la série originale $x_{v,\tau}$; par contre, lorsque l'approche (b) est adoptée, $y'_{v,\tau}$ aura des valeurs positives et négatives, ce qui risque de nécessiter des transformations plus complexes. Dans ce travail, l'approche (a) est adoptée.

4. MÉTHODOLOGIE

La procédure à suivre se résume en trois étapes principales:

1. l'analyse préliminaire et l'identification du modèle;
2. la construction du modèle incluant l'estimation des paramètres et les tests d'ajustement du modèle;
3. la vérification du modèle et l'analyse des résidus incluant les tests optionnels du modèle.

L'annexe 1 contient les formules utilisées pour calculer les diverses statistiques.

4.1 L'analyse préliminaire et l'identification du modèle

Le but de cette analyse est de vérifier la normalité et la stationnarité des données originales, d'effectuer les transformations nécessaires pour normaliser et désaisonnaliser les séries, d'identifier le type et l'ordre du modèle, et d'identifier, pour les données périodiques, s'il s'agit d'un modèle à paramètres constants ou variables. Les sous-étapes sont les suivantes:

- a. vérifier si la série hydrologique originale est normale. Si la série est normale, poser $y_{v,\tau} = x_{v,\tau}$ pour les données périodiques, $y_t = x_t$ pour les données annuelles, et passer à la sous-étape c. Si la série n'est pas normale, aller à b.
- b. transformer la série originale périodique $x_{v,\tau}$ (annuelle x_t) en une série normalisée appelée $y_{v,\tau}$ (annuelle y_t) par une des transformations décrites dans la section 3.2.
- c. faire le graphique de la série périodique $y_{v,\tau}$ ou de la série annuelle y_t ; observer ses caractéristiques principales (déterminer si la série annuelle est stationnaire en vérifiant par inspection visuelle si la moyenne et la variance sont constantes), calculer la moyenne et l'écart type de l'échantillon, \bar{y} et s pour la série annuelle et \bar{y}_τ et s_τ pour la série périodique.
- d. lorsque la série annuelle y_t semble stationnaire, poser $z_t = y_t$. Pour la série périodique $y_{v,\tau}$, obtenir $z_{v,\tau}$ sa version désaisonnalisée par la standardisation $z_{v,\tau} = \frac{y_{v,\tau} - \bar{y}_\tau}{s_\tau}$. La

technique de différenciation pour désaisonnaliser les données ne doit pas être utilisée car elle crée une série différenciée à variance infinie, et un spectre déformé qui cause des problèmes pour la simulation (Kavvas et Delleur, 1975).

- e. dans le cas de la série périodique, faire le graphique de la fonction de corrélation $r_{k,\tau}$ de la série $y_{v,\tau}$ en fonction de τ . Généralement, on se limite à $k=2$.

$$r_{k,\tau} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)(x_{v,\tau-k} - \bar{x}_{\tau-k})}{s_\tau s_{\tau-k}}$$

Si suite à une inspection visuelle, les coefficients d'autocorrélation ne varient pas en fonction de la saison, un modèle ARMA à paramètres constants peut être ajusté.

4.2 La construction du modèle

La section 4.2.1 décrit la procédure pour ajuster un modèle ARMA à une série annuelle z_t ou à une série mensuelle à paramètres constants $z_t = z_{v,\tau}$ où $t = 12(v-1) + \tau$. La section 4.2.2 décrit l'ajustement d'un modèle ARMA à paramètres variables pour une série mensuelle.

4.2.1 La modélisation ARMA

Une série hydrologique z_t , normale et approximativement stationnaire qui suit un modèle ARMA(p,q) à paramètres constants est décrite par le processus suivant:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

où ε_t sont des variables normales indépendantes de moyenne 0 et variance 1. On dénote ARMA(p,0) par AR(p), ARMA(0,q) par MA(q), et ARMA(0,0) par AR(0).

Dans le cas de la série annuelle, t est l'indice des années, il varie de 1 à N ; dans le cas de la série périodique à paramètres constants, t varie de 1 à $\omega * N$; par exemple pour une série mensuelle, t varie de 1 à $12 * N$, on a alors $z_t = z_{v,\tau}$ où $t = 12(v-1) + \tau$.

Plusieurs logiciels informatiques commerciaux tels que SAS, STATGRAPH, SPSS, et MINITAB peuvent être utilisés pour exécuter les étapes suivantes de modélisation Box-Jenkins d'un processus ARMA.

4.2.1.1 L'identification

À partir des graphiques suivants, il est possible de déterminer p et q du modèle ARMA:

- le graphique de la série historique;
- le corrélogramme de la série historique;
- le corrélogramme partiel de la série historique, et
- le périodogramme de la série historique.

Lorsque le périodogramme indique que les données ne sont pas indépendantes, p et/ou q sont différents de zéro. Un modèle AR(p) a un corrélogramme qui s'atténue vers l'infini, et un corrélogramme partiel qui devient 0 après l'ordre p; un modèle MA(q) a un corrélogramme fini qui devient 0 après l'ordre q et un corrélogramme partiel qui s'atténue vers l'infini; et un modèle ARMA(p,q) a un corrélogramme et un corrélogramme partiel qui s'étendent vers l'infini (Box et Jenkins, 1976).

4.2.1.2 L'estimation des paramètres du modèle

Les sous-étapes à suivre sont les suivantes:

- a. estimer les paramètres ϕ_i et θ_i par la méthode du maximum de vraisemblance ou la méthode de minimisation des écarts au carré;
- b. vérifier que les paramètres estimés sont significatifs, l'estimateur du paramètre, divisé par son écart type, suit approximativement une loi de student;
- c. vérifier que les paramètres estimés rencontrent les conditions de stationnarité et d'inversibilité assurant ainsi une variance finie; c'est à dire, les racines r et s des polynômes $1 - \sum_{i=1}^p \phi_i r^i = 0$ et $1 - \sum_{j=1}^q \theta_j s^j = 0$ doivent se situer à l'extérieur du cercle unitaire. Pour AR(1) celles-ci correspondent à $-1 < \phi_1 < 1$ et pour AR(2) elles correspondent à $-1 < \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $\phi_1 + \phi_2 < 1$.

4.2.2 La modélisation de séries périodiques à paramètres variables

La modélisation des données périodiques est plus complexe que celle des données annuelles à cause de l'influence du cycle interannuel qui produit des variations dans les

caractéristiques statistiques de la série. Toutefois, la plupart des séries hydrologiques périodiques à paramètres variables peuvent être modélisées par un AR(1)V ou un AR(2)V. L'expression de AR(p)V pour une série normalisée et désaisonnalisée est (Salas *et al.*, 1980):

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

où $\xi_{v,\tau}$ sont des variables normales indépendantes de moyenne 0 et variance 1; $\sigma_{\varepsilon,\tau}^2$ est la variance des résidus $\xi_{v,\tau}$.

Dans le cas d'un modèle AR(1)V:

$$\begin{aligned} \hat{\phi}_{1,\tau} &= \hat{\rho}_{1,\tau}, \quad \tau = 1.. \omega \\ \hat{\sigma}_{\varepsilon,\tau}^2 &= 1 - \hat{\phi}_{1,\tau} \hat{\rho}_{1,\tau}, \quad \tau = 1.. \omega \end{aligned}$$

et dans le cas d'un modèle AR(2)V:

$$\begin{aligned} \hat{\phi}_{1,\tau} &= \frac{\hat{\rho}_{1,\tau} - \hat{\rho}_{1,\tau-1} \hat{\rho}_{2,\tau}}{1 - \hat{\rho}_{1,\tau-1}^2}, \quad \tau = 1.. \omega \\ \hat{\phi}_{2,\tau} &= \frac{\hat{\rho}_{2,\tau} - \hat{\rho}_{1,\tau-1} \hat{\rho}_{2,\tau}}{1 - \hat{\rho}_{1,\tau-1}^2}, \quad \tau = 1.. \omega \\ \hat{\sigma}_{\varepsilon,\tau}^2 &= 1 - \hat{\phi}_{1,\tau} \hat{\rho}_{1,\tau} - \hat{\phi}_{2,\tau} \hat{\rho}_{2,\tau}, \quad \tau = 1.. \omega \end{aligned}$$

En pratique, on doit vérifier que les estimateurs $\hat{\sigma}_{\varepsilon,\tau}^2$ sont en effet non négatifs.

4.3 La vérification du modèle et l'analyse des résidus

Les sous-étapes de vérification et de choix du modèle sont les suivantes:

- les résidus du modèle doivent constituer un bruit blanc, (données indépendantes distribuées selon une loi $N(0,1)$);
- les paramètres inclus dans le modèle doivent être significativement différents de 0, c'est-à-dire égaux à deux à trois fois leur écart type; lorsqu'ils ne le sont pas, on les laisse tomber;

- c. le principe de parcimonie des paramètres étant important, lorsque plusieurs modèles ARMA(p,q) (à paramètres constants) rencontrent toutes les exigences requises, on discrimine parmi eux, en choisissant le modèle qui minimise le critère d'Akaike AIC(p,q), où:

$$\text{AIC}(p,q) = N \ln(\hat{\sigma}_e^2) + (p+q)$$

N est la taille de l'échantillon et $\hat{\sigma}_e^2$ est l'estimateur de maximum de vraisemblance de la variance des résidus.

5. LE MODÈLE DE DÉSAGRÉGATION

L'approche indirecte de génération de données saisonnières consiste en la génération de données annuelles en un premier temps, et la désagrégation de ces données annuelles en données saisonnières par une méthode de désagrégation. La méthode de désagrégation présentée dans Grygier et Stedinger (1990), lorsque Z_n sont des données annuelles normales et $Z_{n,t}$ des données mensuelles normales, est la suivante:

$$Z_{n,1} = \alpha_1 + \gamma_1 Z_n + V_{n,1}$$

$$Z_{n,2} = \alpha_2 + \gamma_2 Z_n + \delta_2 \Lambda_{n,t} + V_{n,2}$$

$$Z_{n,t} = \alpha_t + \beta_t Z_{n,t-1} + \gamma_t Z_n + \delta_t \Lambda_{n,t} + V_{n,t}, \quad 3 \leq t \leq 12$$

où $\Lambda_{n,t} = \sum_{s=1}^{t-1} w_s Z_{n,s}$. $V_{n,t}$ sont des variables aléatoires normales indépendantes de moyenne 0, α_t , β_t , γ_t , et δ_t sont les paramètres de régression, et w_s sont des poids qui dépendent de la transformation effectuée sur les données originales Q_n et $Q_{n,t}$ afin de les normaliser et d'obtenir Z_n et $Z_{n,t}$.

SPIGOT (Grygier et Stedinger, 1990) est un logiciel d'approche indirecte qui génère des séries de débits mensuels à plusieurs sites (multivariées), à partir de données annuelles par une technique de désagrégation temporelle. Il contient six modules qui peuvent être combinés différemment selon les besoins de l'application. Ces six modules sont:

1. un modèle univarié AR(1) ou AR(0) générant des données annuelles agrégées;
2. un modèle AR(1) multivarié générant des vecteurs de données annuelles;
3. un modèle de désagrégation annuel-mensuel générant des données mensuelles à partir de données annuelles univariées générées,
4. un modèle de désagrégation de données annuelles univariées en données mensuelles multivariées à plusieurs sites;
5. un modèle annuel multivarié de désagrégation en données mensuelles à plusieurs sites;
et

6. un modèle de désagrégation spatiale de données mensuelles en un site à des données mensuelles à plusieurs sites.

Dans ce travail, la combinaison des modules (1) et (3) est appliquée. Le module (1) est utilisé pour générer des données annuelles à partir d'un AR(0), et le module (3) utilise la méthode de désagrégation décrite précédemment pour désagréger les données générées par le module (1) en des données mensuelles.

6. APPLICATION

Des données mensuelles et annuelles d'apports nets (1000 pcs) sur le lac Supérieur, entre 1900 et 1989, sont utilisées pour comparer les approches directes et indirectes de simulation de données synthétiques mensuelles. L'approche directe est représentée par quatre modèles de la famille ARMA à paramètres constants et à paramètres variables avec ou sans pondération annuelle. L'approche indirecte est représentée par le modèle de désagrégation implanté dans le logiciel SPIGOT. L'objectif est de préserver le mieux possible les caractéristiques mensuelles et annuelles. Toutefois aucun modèle ne peut conserver toutes ces caractéristiques. En général, les méthodes directes préservent bien les caractéristiques mensuelles et les méthodes indirectes préservent bien les caractéristiques annuelles.

L'année hydrologique utilisée commence au mois d'octobre et se termine au mois de septembre. Les données des quatre modèles directs sont générées à l'aide d'un programme écrit en C sur MACINTOSH. Les données annuelles sont x_t , $t=1..90$. Les données périodiques sont $x_{v,\tau}$, $\tau=1..12$, et $v=1..90$. $\tau=1$ correspond au mois de janvier et $\tau=12$ au mois de décembre. Le logiciel STATGRAPH a été utilisé pour la construction des modèles ARMA pour les séries annuelles et les séries périodiques à paramètres constants.

6.1 La série annuelle

Un modèle univarié de la famille ARMA est ajusté aux données annuelles d'apports nets du lac Supérieur. Le tableau 1 présente les statistiques annuelles de la série x_t .

Moyenne	870.14
Écart type	204.03
Coefficient d'asymétrie	.03
Coefficient de Hurst	.55

Tableau 1: Les statistiques de la série annuelle historique x_t

La normalité des données

La figure 1 présente la droite de Henri (STATGRAPH) de x_t . Le coefficient de corrélation de Filliben est de 0.997 (SPIGOT) ce qui est supérieur à .986, la valeur critique à $\alpha=0.05$.

L'hypothèse de la normalité des données annuelles n'est donc pas rejetée, aucune transformation n'est nécessaire, et $y_t = x_t$.

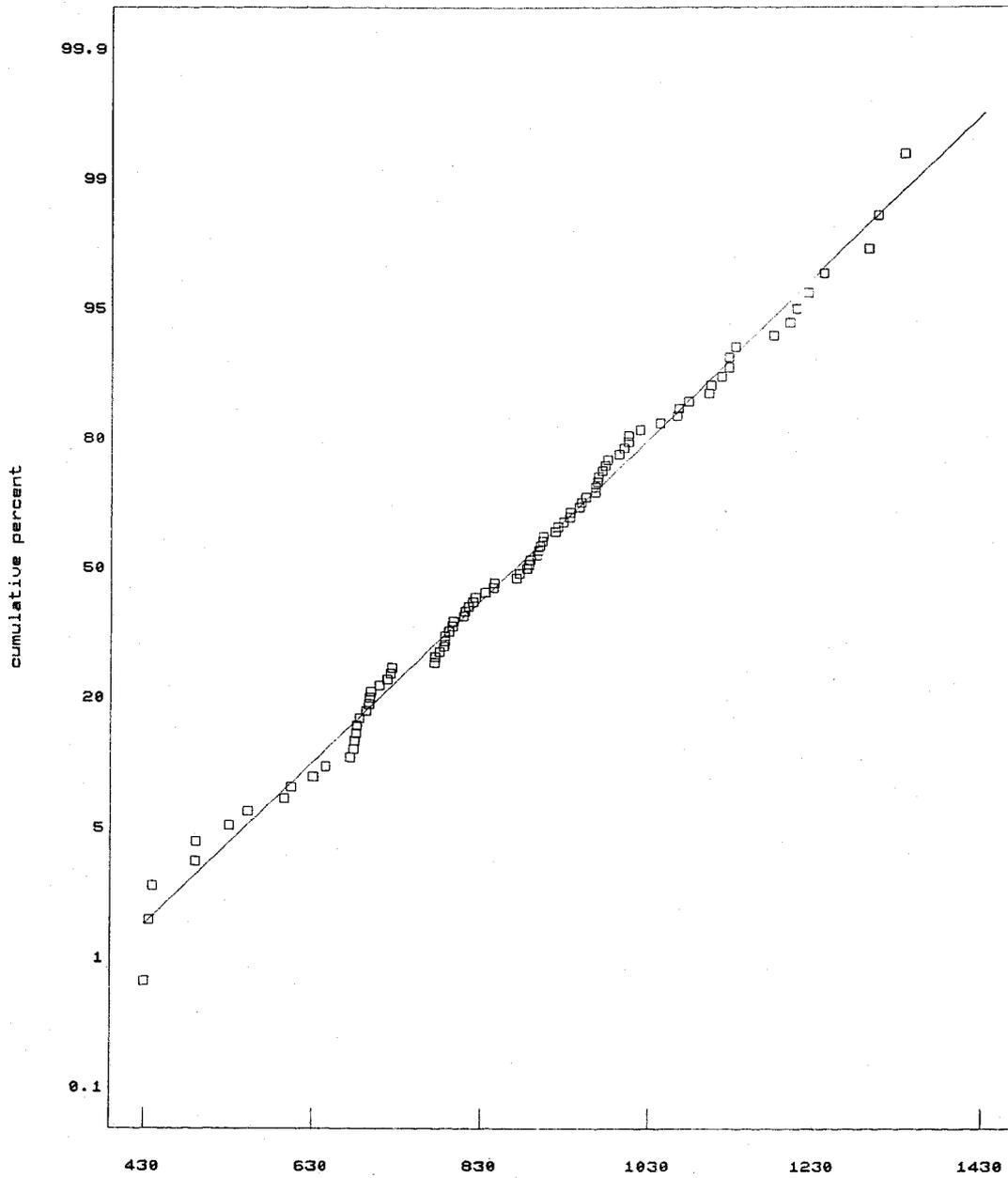


Figure 1: La droite de Henri des données annuelles historiques, x_t

La figure 2 présente y_t en fonction du temps. Les données semblent stationnaires, il n'y a pas de tendance ni de cycle. La série y_t est pondérée afin d'obtenir la série $z_t = y_t/87.01$ dont la moyenne est 10. Le dénominateur correspond à la moyenne de la série historique divisée par 10. Cette pondération sert à garder les variables transformées du même ordre de grandeur afin de réduire les problèmes de mauvais conditionnement lors de l'ajustement du modèle ARMA.. Le tableau 2 présente la moyenne et l'écart type de la série z_t .

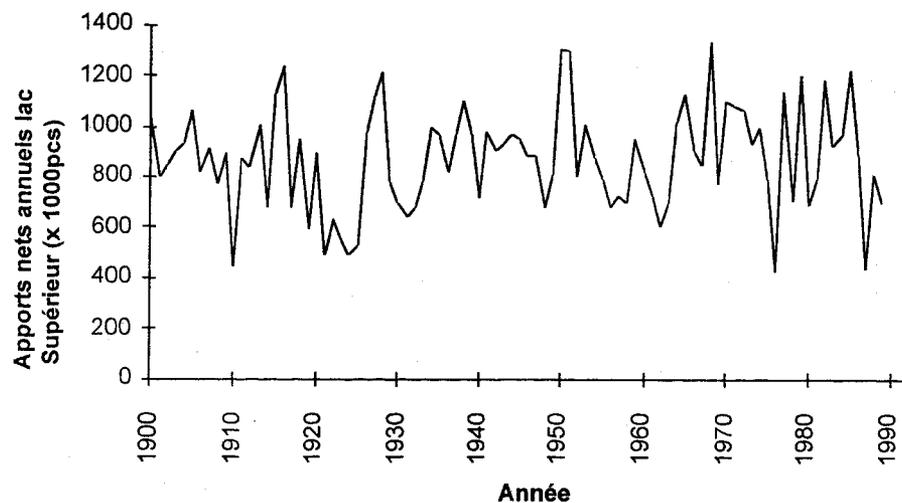


Figure 2: Les apports nets annuels historiques du lac Supérieur

Moyenne	10.0
Écart type	30.25

Tableau 2: Les statistiques annuelles de la série z_t

La construction du modèle

Les figures 3 à 5 présentent le corrélogramme, le corrélogramme partiel, et le périodogramme cumulatif de z_t qui indiquent que les données annuelles sont indépendantes. Un modèle AR(0) est donc approprié. Les données générées suivent une normale (10,30.25). Le modèle est:

$$z_t = 10.0 + 5.5 \varepsilon_t, \text{ où } \varepsilon_t \text{ sont des variables normales } N(0,1) \text{ indépendantes.}$$

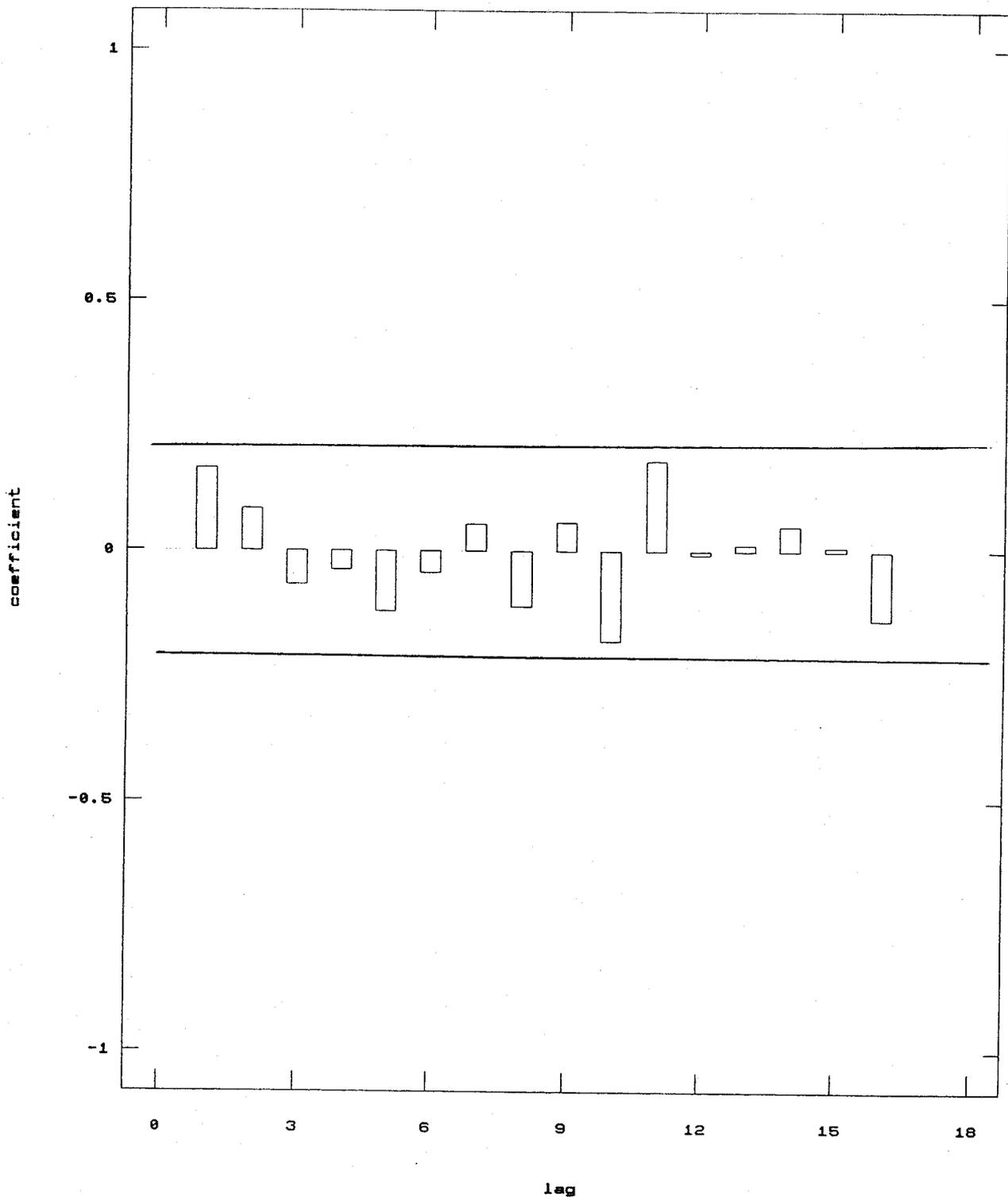


Figure 3: Le corrélogramme des données annuelles transformées, z_t

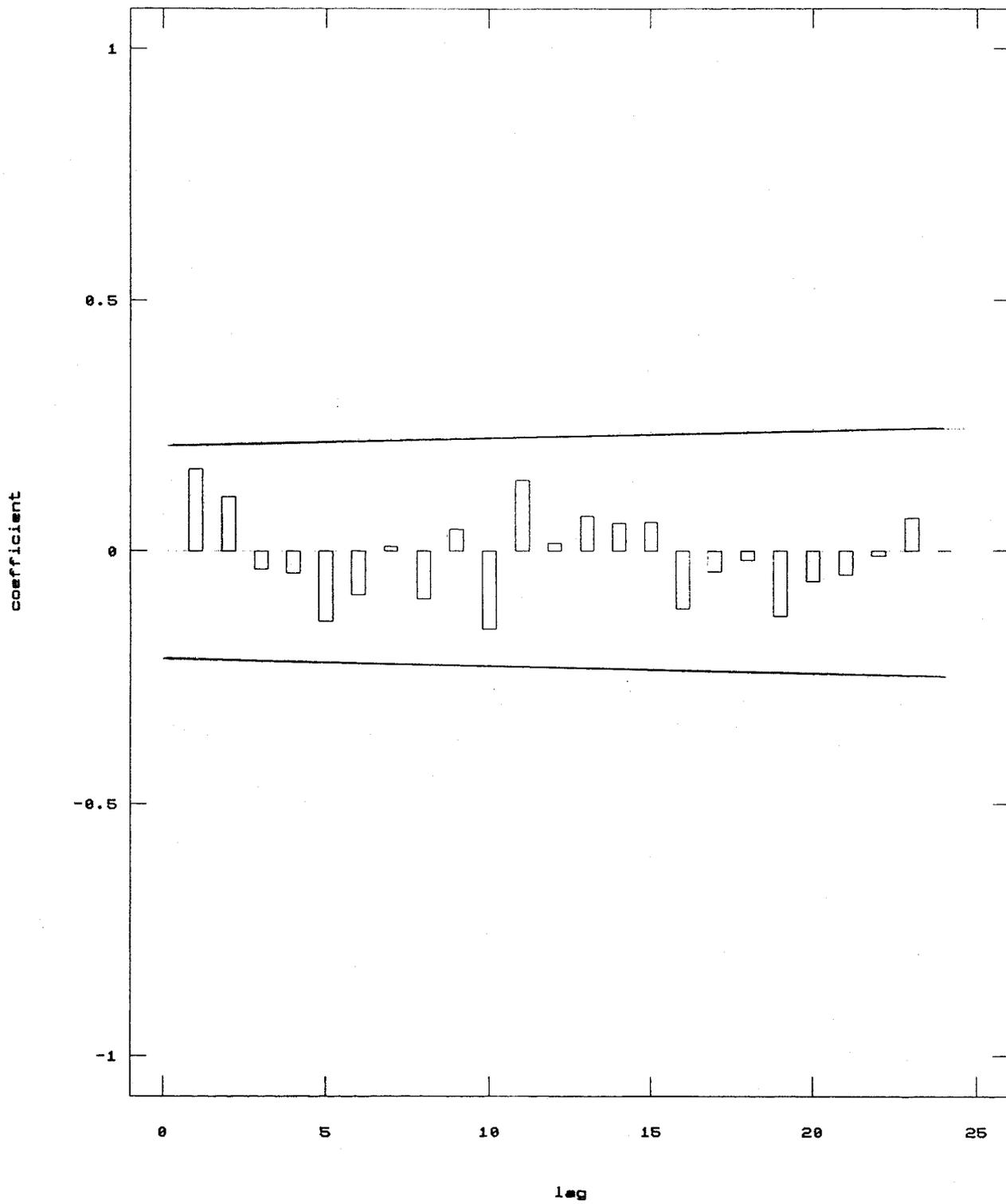


Figure 4: Le corrélogramme partiel des données annuelles transformées, z_t

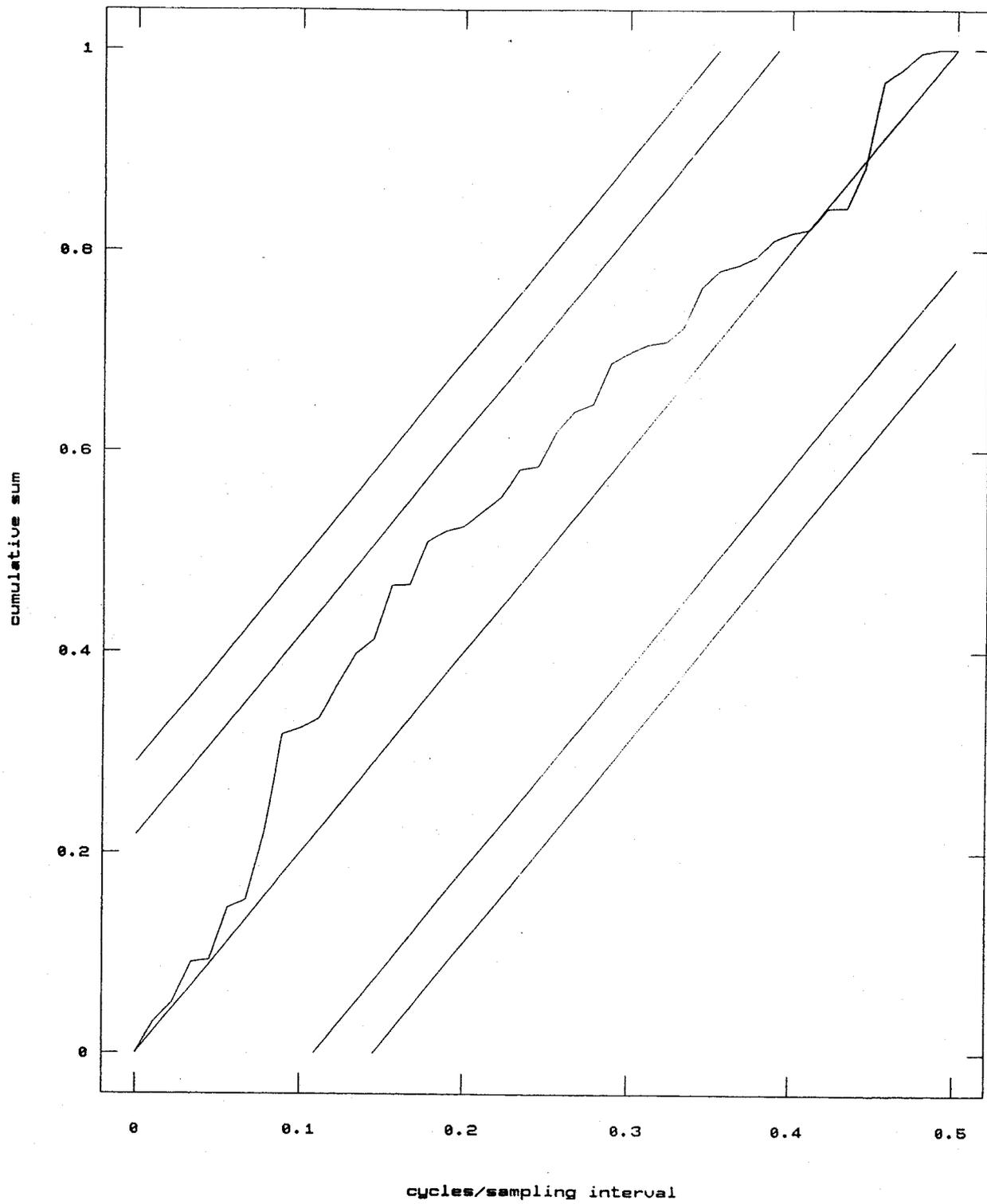


Figure 5: Le périodogramme cumulatif des données annuelles transformées, Z_t

Les tests d'ajustement du modèle et l'analyse des résidus

Les figures 6 et 7 présentent le corrélogramme des résidus du modèle AR(0) et la droite de Henri respectivement. Le corrélogramme indique l'indépendance des résidus car aucun coefficient d'autocorrélation n'est significativement différent de 0 au niveau 5%. La statistique de Kolmogorov-Sminrnov D_{90} est .04 ($< 0.17 = C = \frac{1.63}{\sqrt{n}}$), et la droite de Henri ne rejettent pas l'hypothèse de normalité des résidus. On en déduit que le modèle AR(0) s'ajuste bien aux données annuelles z_t .

Trente séries de 90 ans de données annuelles sont générées selon le modèle: $z_t = 10.0 + 5.5 \varepsilon_t$, où ε_t sont des variables normales $N(0,1)$ indépendantes. Donc, $y_t = z_t * 87.01$ et $x_t = y_t$.

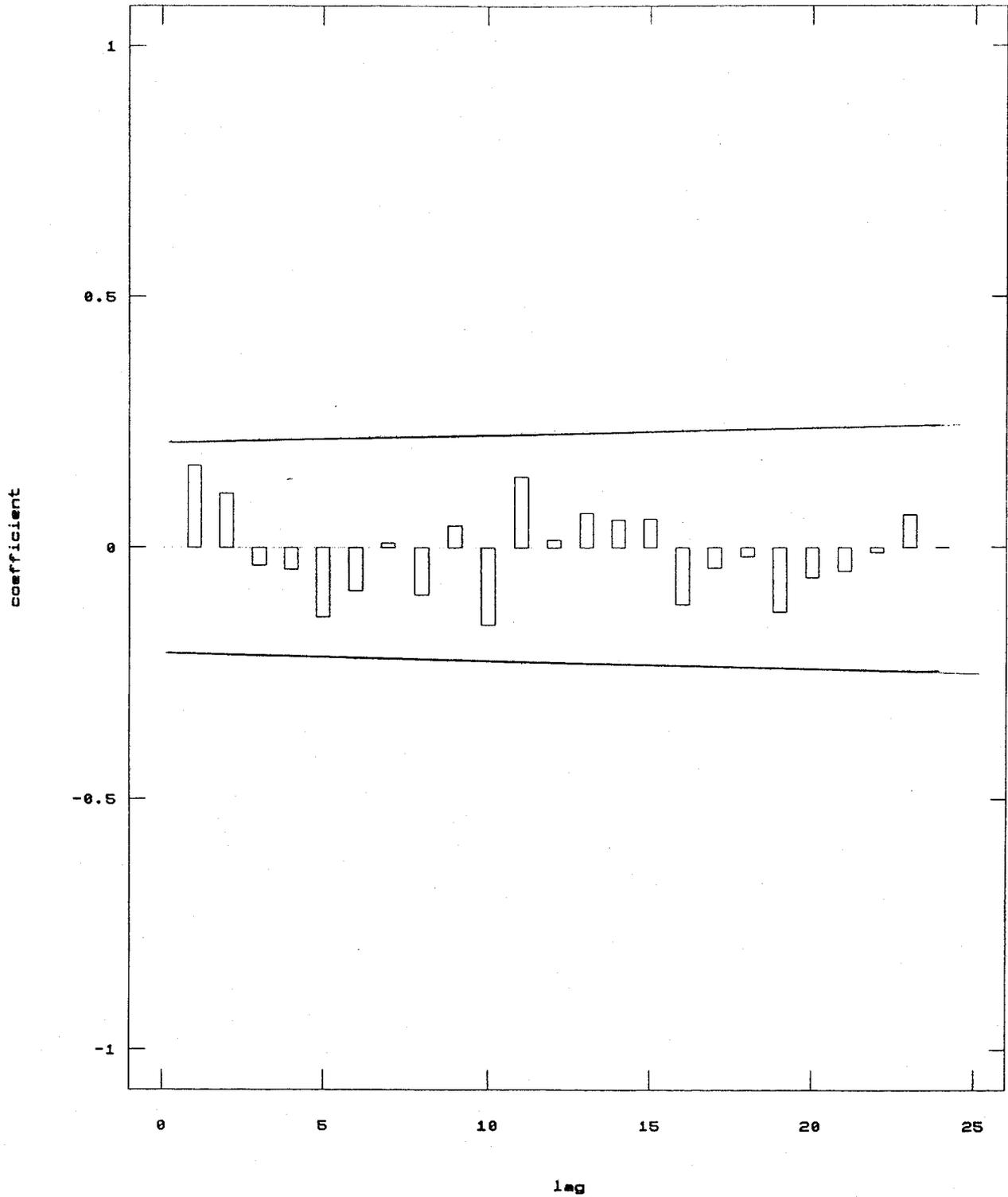


Figure 6: Le corrélogramme des résidus du modèle AR(0) pour z_t

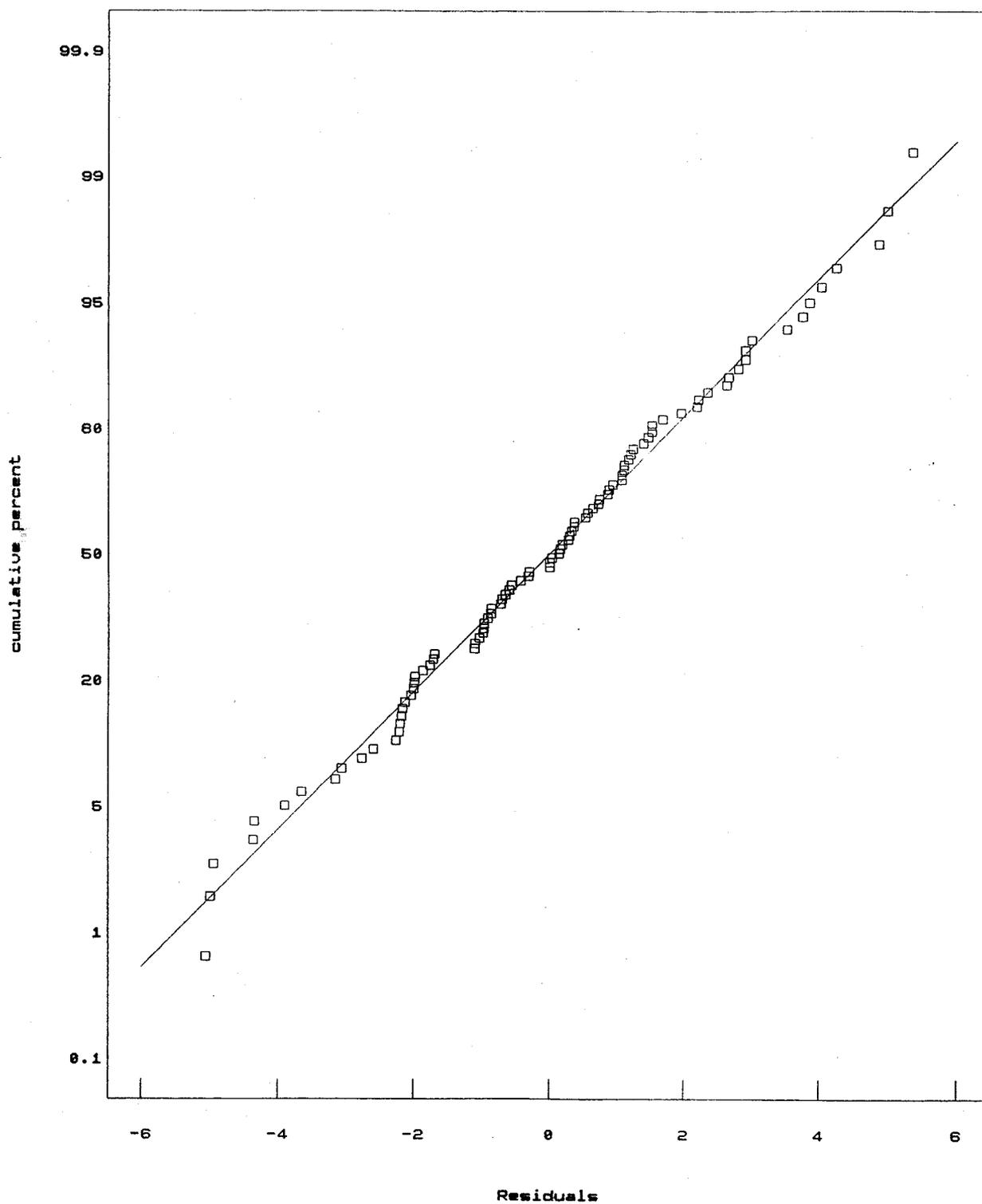


Figure 7: La droite de Henri résidus du modèle AR(0) pour z_t

6.2 La série mensuelle

Un modèle univarié à paramètres constants de la famille ARMA, et un modèle univarié à paramètres variables de la famille ARMA sont ajustés aux données mensuelles $x_{v,\tau}$ d'apports nets au lac Supérieur. La figure 8 présente la série chronologique des données historiques mensuelles $x_{v,\tau}$.

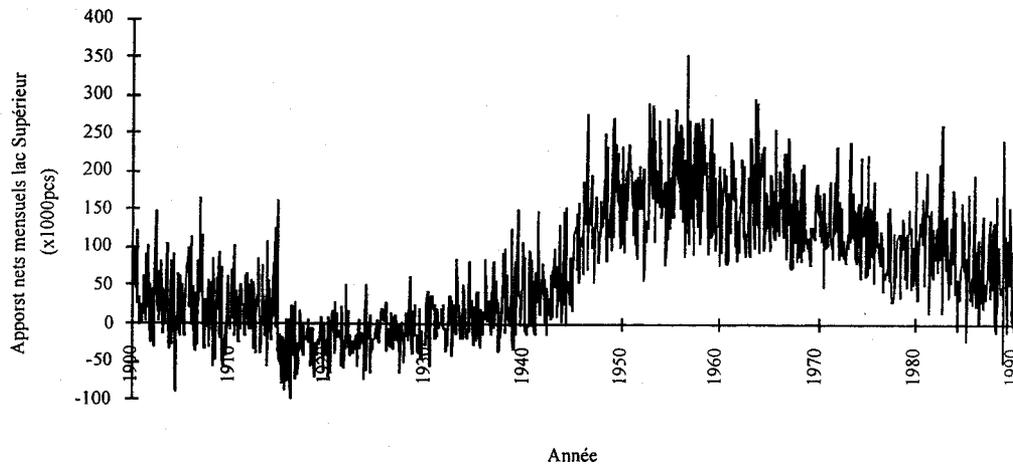


Figure 8: Les données historiques mensuelles $x_{v,\tau}$

La normalité des données

Dans un premier temps, il s'agit de déterminer si les données mensuelles sont normales. Les coefficients de corrélation de Filliben (r) ont été calculés pour tous les mois, pour les transformations basées sur les lois Normale, Lognormale, Lognormale-3, et Gamma à 3 paramètres (dans SPIGOT). Les valeurs de r les plus élevées sont celles de la loi Lognormale-3 pour tous les mois sauf les mois d'octobre et décembre. Les valeurs de r les plus élevées pour ces deux mois sont celles de la loi Normale. L'hypothèse de normalité est alors rejetée pour tous les mois sauf les mois d'octobre et décembre. La Lognormale-3 est la plus appropriée pour les 10 autres mois dont les données, $x_{v,\tau}$ sont transformées pour obtenir des données normales $y_{v,\tau}$:

$$y_{v,\tau} = \ln(x_{v,\tau} - c_\tau), \quad c_\tau = g_{y,\tau}$$

Les moments de $y_{v,\tau}$ sont obtenus en résolvant les équations des moments de Matalas (1967) de la section 3.2. Les tableaux 3 et 4 présentent les moyennes et les écarts types périodiques de la série historique $x_{v,\tau}$ tandis que les tableaux 5, 6, et 7 présentent les moyennes, les écarts types, et les coefficients d'asymétrie périodiques de la série transformée $y_{v,\tau}$ calculés selon la méthode des moments.

Moyenne de l'échantillon $x_{v,\tau}$, \bar{x}_τ											
Jan	Fév	Mar	Avr	Mai	Jun	Juil	Aou	Sep	Oct	Nov	Déc
-13.6	10.4	45.3	148.7	185.6	157.7	129.6	100.6	73.3	37.88	18.2	-23.6

Tableau 3: Les moyennes mensuelles périodiques de $x_{v,\tau}$

Écart type de l'échantillon $x_{v,\tau}$, $s_{x\tau}$											
Jan	Fév	Mar	Avr	Mai	Jun	Juil	Aou	Sep	Oct	Nov	Déc
24.3	27.9	42.8	49.8	59.2	52.3	40.1	43.1	55.9	46.5	44.6	28.9

Tableau 4: Les écarts types mensuels périodiques de $x_{v,\tau}$

Méthode des moments, Moyenne de $y_{v,\tau}$, \bar{y}_τ											
Jan	Fév	Mar	Avr	Mai	Jun	Juil	Aou	Sep	Oct	Nov	Déc
5.82	5.16	5.4	5.97	6.65	5.95	5.33	5.64	5.33	-	5.45	-

Tableau 5: Les moyennes mensuelles périodiques de $y_{v,\tau}$

Méthode des moments, Écart type de $y_{v,\tau}$, $s_{y\tau}$											
Jan	Fév	Mar	Avr	Mai	Jun	Juil	Aou	Sep	Oct	Nov	Déc
.072	.157	.19	.125	.076	.134	.189	.15	.26	-	.186	-

Tableau 6: Les écarts types mensuels périodiques de $y_{v,\tau}$

Méthode des moments, coefficient d'asymétrie de $y_{v,\tau}$, $g_{y,\tau} = c_\tau$											
Jan	Fév	Mar	Avr	Mai	Jun	Juil	Aou	Sep	Oct	Nov	Déc
-350.8	-166.3	-175.6	-247.8	-592.9	-230.6	-80.66	-184.8	-139.8	-	-219.3	-

Tableau 7: Les coefficients d'asymétrie périodiques de $y_{v,\tau}$

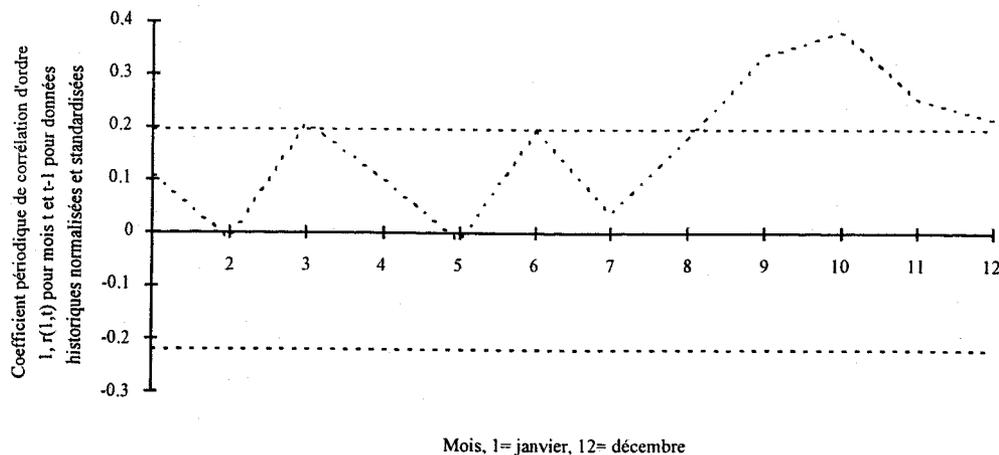
Afin d'éliminer les saisonnalités des données périodiques, la série $z_{v,\tau}$ est standardisée par

$$z_{v,\tau} = \frac{y_{v,\tau} - \bar{y}_\tau}{s_{y,\tau}} \text{ pour tous les mois sauf les mois d'octobre et de décembre pour lesquels}$$

$$z_{v,\tau} = \frac{x_{v,\tau} - \bar{x}_\tau}{s_{x,\tau}}. \text{ La série } z_{v,\tau} \text{ ainsi obtenue est une série de variables normales } N(0,1).$$

6.2.1 L'approche Directe

La série $z_{v,\tau}$ est modélisée par une approche directe ARMA, dans le but de générer des données synthétiques. Les graphiques de $r_{1,\tau}$ et $r_{2,\tau}$ en fonction de τ aux figures 9 et 10 suggèrent qu'un modèle ARMA à paramètres constants est adéquat car les coefficients d'autocorrélation d'ordre 1 et d'ordre 2 ne varient pas grandement. La corrélation d'ordre 1 est significative à 5 % pour tous les mois sauf les mois de janvier, février, avril, mai et juillet. La corrélation d'ordre deux est généralement non significative sauf pour les mois de septembre, novembre et décembre. Toutefois, par souci de rigueur, les deux approches à paramètres constants et à paramètres variables sont utilisées et comparées dans cette étude.

Figure 9: Les autocorrélations périodiques d'ordre 1 $r_{1,\tau}$ de $z_{v,\tau}$

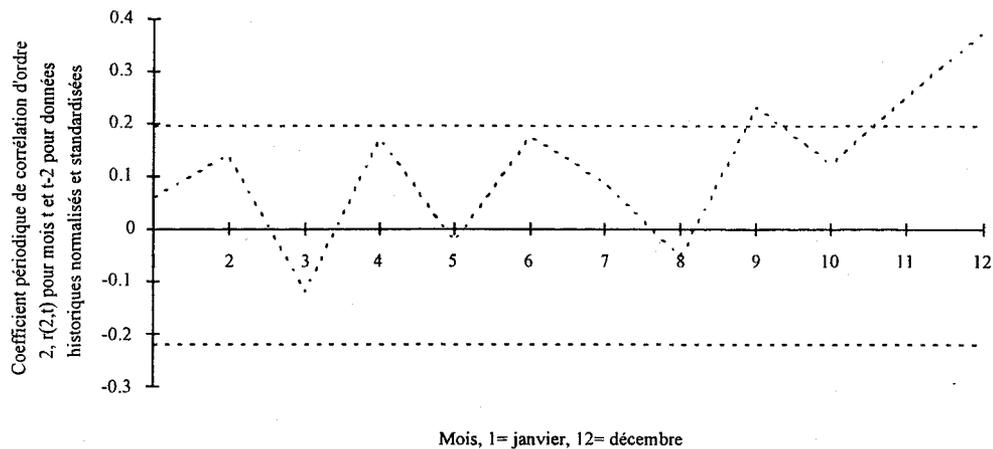


Figure 10: Les autocorrélations périodiques d'ordre 2 $r_{2,\tau}$ de $z_{v,\tau}$

6.2.1.1 ARMA à paramètres constants

Dans cette approche, la série $z_t = z_{v,\tau}$ où $t = 12(v-1) + \tau$ est utilisée.

La construction du modèle

Les figures 11, 12, et 13 présentent le corrélogramme, le corrélogramme partiel, et le périodogramme cumulatif de z_t qui indiquent que les données mensuelles ne sont pas indépendantes et qui suggèrent même une dépendance d'ordre 2 car l'autocorrélation d'ordre 2 est significative au niveau 5%. Il est important de noter que lorsque les données mensuelles sont considérées en bloc, l'autocorrélation d'ordre 2 est significative et l'autocorrélation d'ordre 1 ne l'est pas. Toutefois, les corrélations mensuelles d'ordre 2 ne sont pas significatives et les corrélations mensuelles d'ordre 1 sont significatives telles qu'indiquée par les figures 9 et 10.

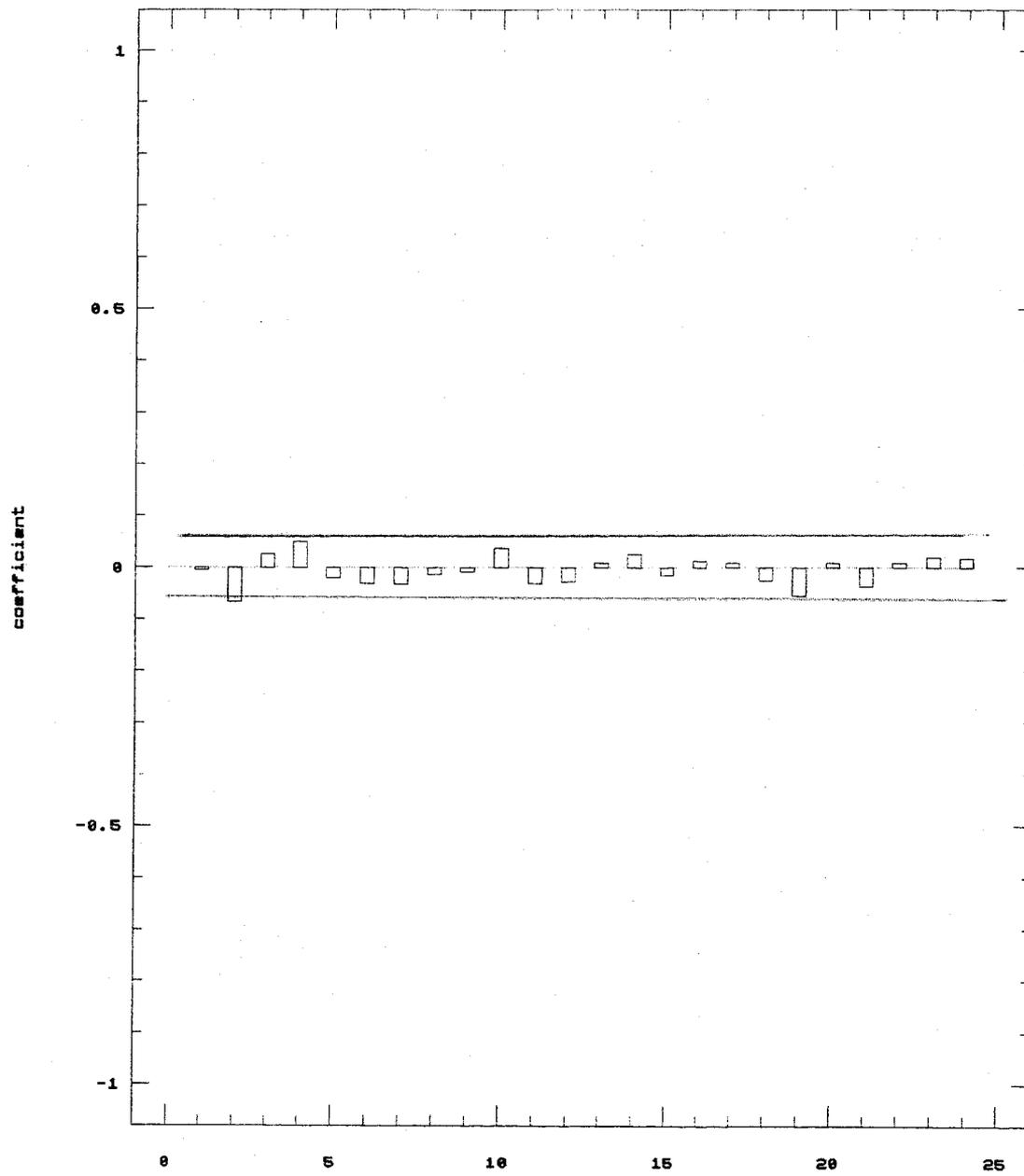


Figure 11: Le corrélogramme des données mensuelles transformées, z_t

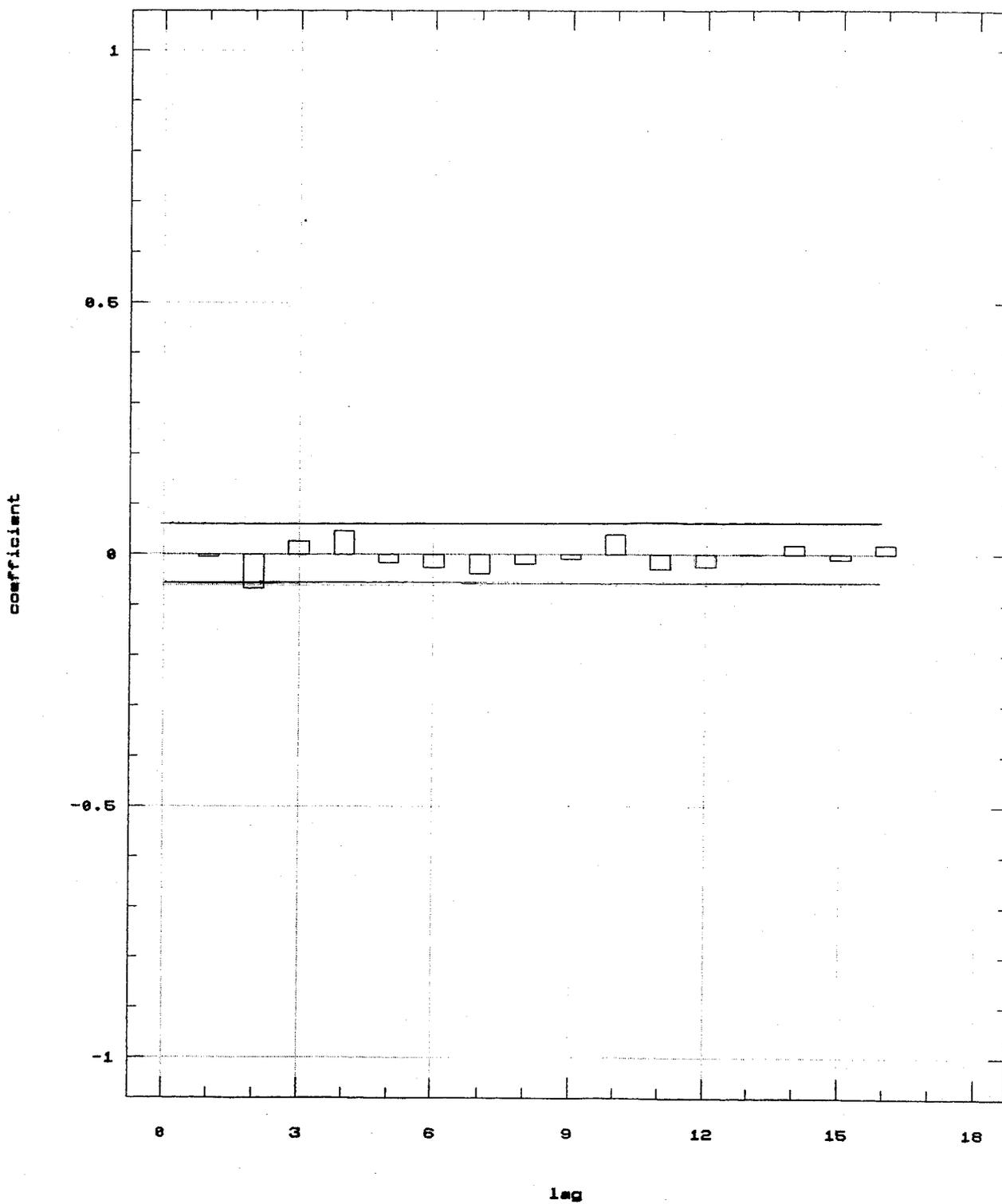


Figure 12: Le corrélogramme partiel des données mensuelles transformées, z_t

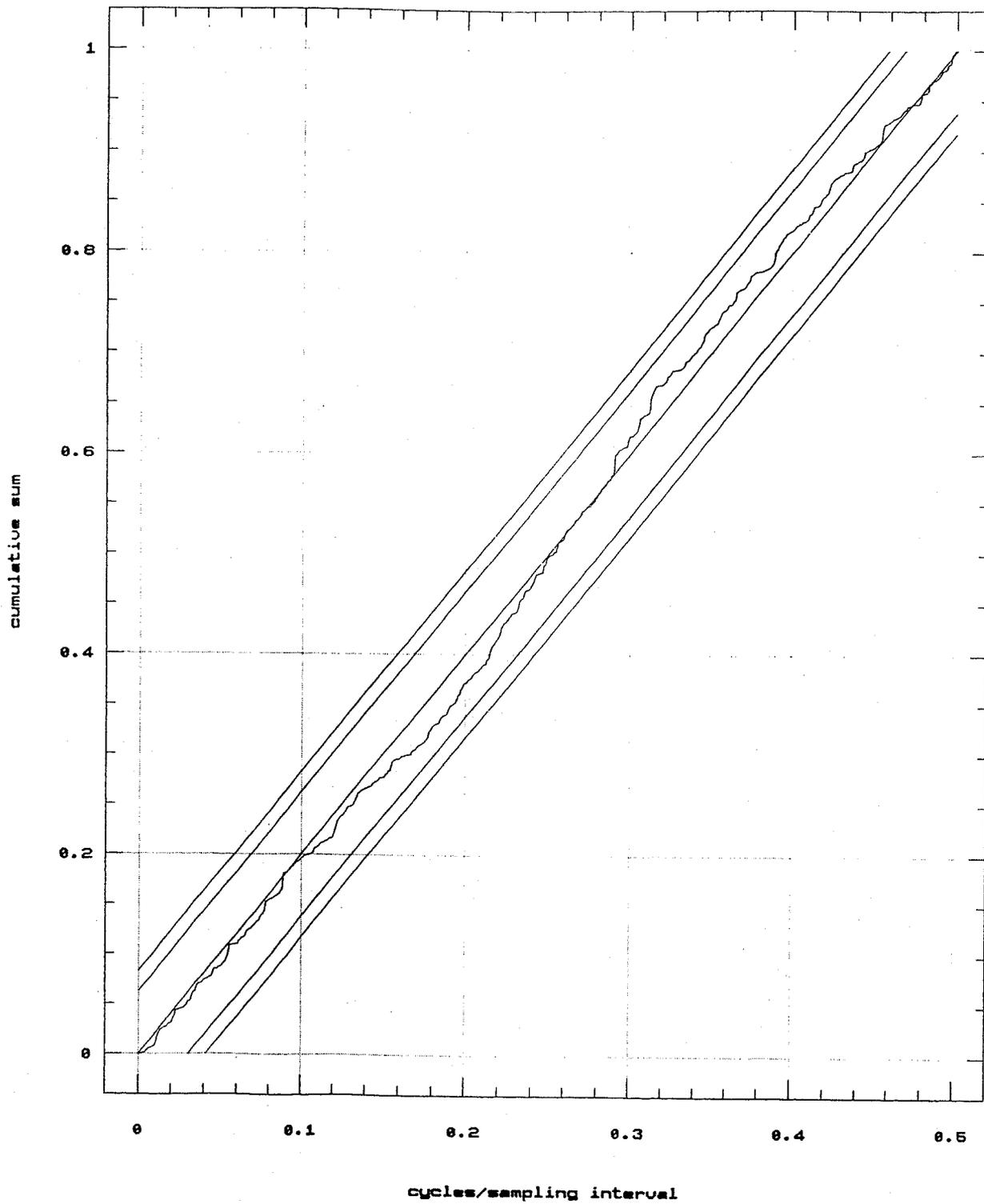


Figure 13: Le périodogramme cumulatif des données mensuelles transformées, z_t

Après avoir essayé plusieurs modèles dont AR(1), AR(2), ARMA(1,1), MA(2), le modèle AR(2) est jugé le plus approprié, car il est le seul à avoir des résidus normaux et indépendants. Les paramètres du modèle AR(2) sont présentés au tableau 8.

Paramètre	Estimateur	Écart type (É.t.)	Estimateur/É.t.
ϕ_1	-.0048	.030	-.1576
ϕ_2	-.06708	.030	-2.20

Tableau 8: Les paramètres du modèle AR(2)

Le paramètre ϕ_1 n'est pas significativement différent de 0 car la valeur-t qui suit une distribution de student est plus faible que la valeur critique 1.64 ($=t_{.95}$). Ce paramètre n'est donc pas inclus dans le modèle. Le modèle devient:

$$z_t = -.068 z_{t-2} + \varepsilon_t, \text{ où } \varepsilon_t \text{ sont des variables normales } N(0,1) \text{ indépendantes.}$$

On vérifie que le modèle rencontre les conditions de stationnarité et d'inversibilité de la section 4.2.1.b par:

$$-1 < \phi_2 = -.06708 < 1; \quad -.0048 + (-.06708) < 1; \quad \text{et } -.06708 - (-.0048) < 1.$$

L'analyse des résidus

Les figures 14, 15, et 16 présentent l'information concernant les résidus du modèle AR(2). Le corrélogramme ainsi que le périodogramme cumulatif indiquent l'indépendance des résidus. La droite de Henri et le test du Khi-deux ($\hat{\chi}^2 = 14.26 < \chi^2 (.95) = 28.8$ à 18 degrés de liberté) indiquent la normalité des résidus. On en déduit que le modèle AR(2) s'ajuste bien aux données mensuelles z_t . C'est un modèle parcimonieux, à un paramètre significatif.

Trente séries de 90 ans de données mensuelles sont générées selon le modèle:

$$z_t = .068 z_{t-2} + \varepsilon_t, \text{ où } \varepsilon_t \text{ sont des variables normales } N(0,1) \text{ indépendantes;}$$

$$y_{v,\tau} = z_t * s_{y,\tau} + \bar{y}_\tau, \text{ sauf pour les mois d'octobre et décembre;}$$

$$x_{v,\tau} = \exp(y_{v,\tau}) + c_\tau, \text{ sauf pour les mois d'octobre et décembre;}$$

$$x_{v,\tau} = z_t * s_{x,\tau} + \bar{x}_\tau, \text{ pour les mois d'octobre et décembre.}$$

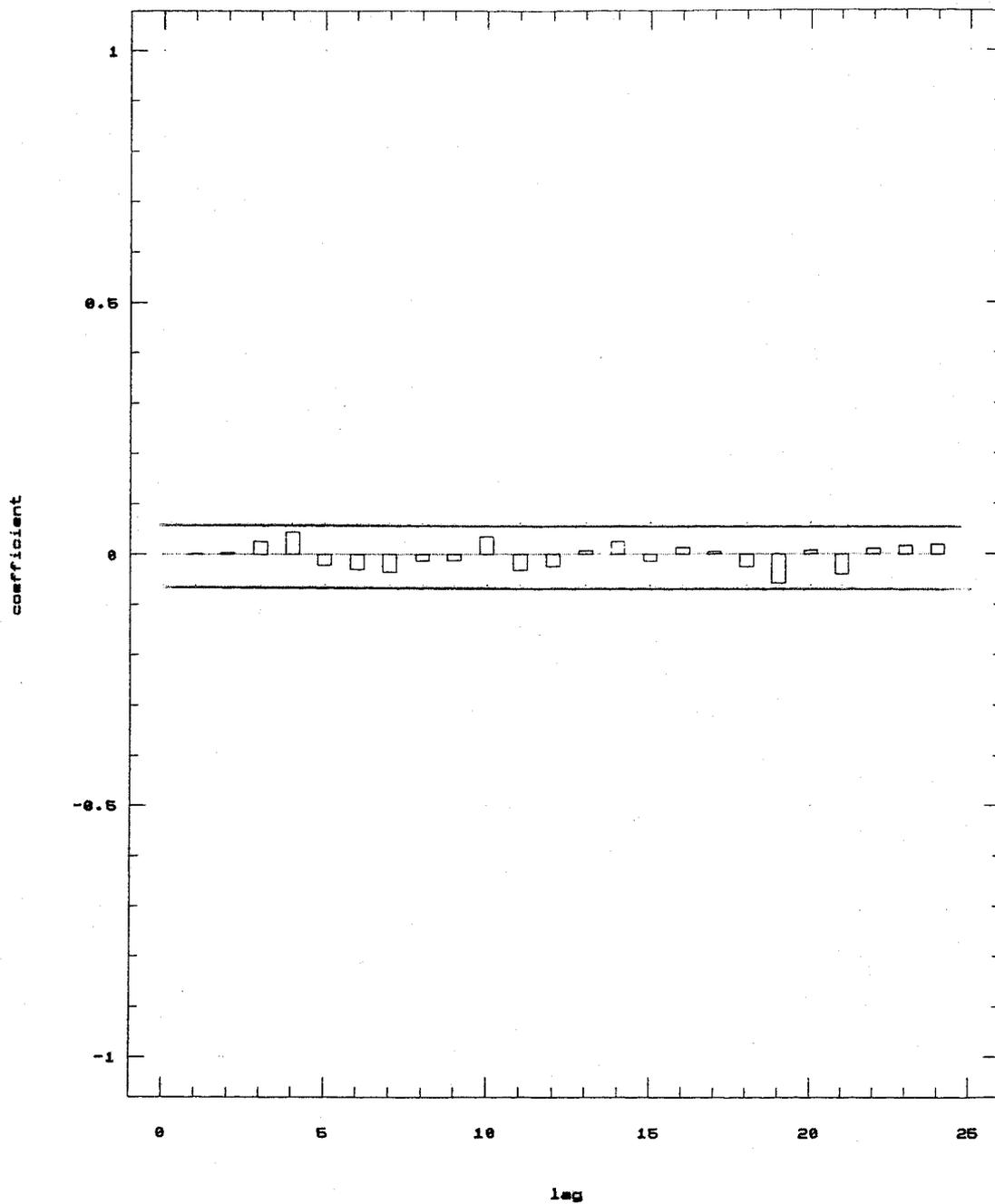


Figure 14: Le corrélogramme des résidus du modèle mensuel AR(2)

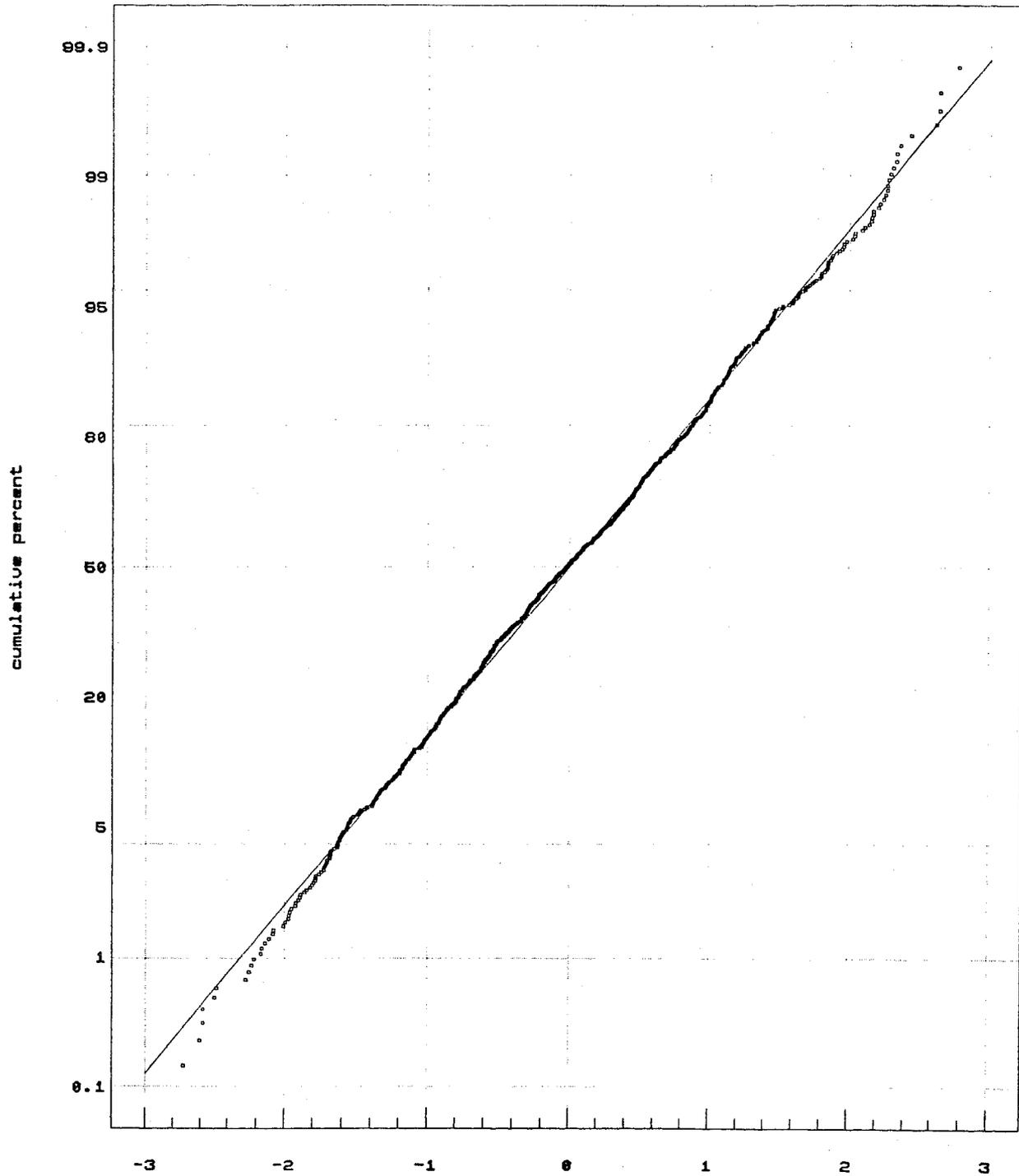


Figure 15: Le périodogramme cumulatif des résidus du modèle mensuel AR(2)

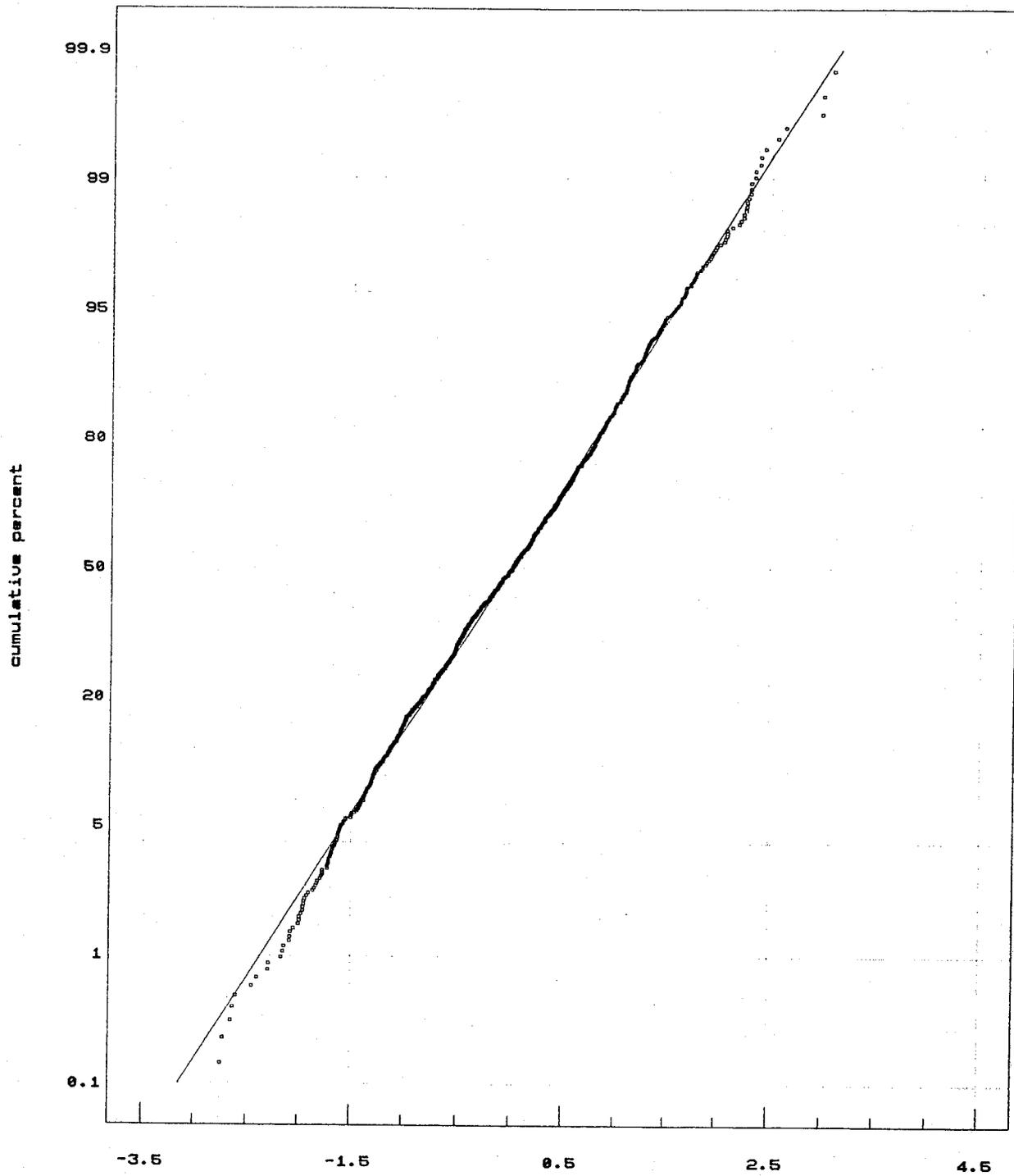


Figure 16: La droite de Henri des résidus du modèle mensuel AR(2)

6.2.1.2 ARMA à paramètres variables

Douze modèles différents doivent être ajustés aux données normalisées et désaisonnalisées $z_{v,\tau}$, chaque modèle correspondant à chacun des mois.

La construction des modèles

Le modèle AR(1)V (V pour indiquer que c'est un modèle périodique à paramètres variables) est ajusté aux données $z_{v,\tau}$. Ce modèle d'ordre 1 est retenu parce que la corrélation d'ordre 2 n'est pas significative pour la plupart des mois (figure 10). Le tableau 9 présente les estimateurs de $\phi_{1,\tau}$ et les estimateurs de la variance des résidus $\sigma_{\varepsilon,\tau}^2$.

Mois	$\hat{\phi}_{1,\tau}$	variance des résidus $\hat{\sigma}_{\varepsilon,\tau}^2$
janvier-décembre, $\tau = 1$.108	.994
février-janvier	-.01	.999
mars-février	.209	.978
avril-mars	.099	.995
mai-avril	-.016	.999
juin-mai	.196	.981
juillet-juin	.035	.999
août-juillet	.179	.984
septembre-juillet	.335	.942
octobre-septembre	.377	.926
novembre-octobre	.253	.967
décembre-novembre $\tau = 12$.213	.977

Tableau 9: Les paramètres du modèle AR(1)V périodique à paramètres variables

Les modèles sont alors:

$$z_{v,\tau} = \hat{\phi}_{1,\tau} z_{v,\tau-1} + \hat{\sigma}_{\varepsilon,\tau} \xi_{v,\tau} \text{ où } \xi_{v,\tau} \text{ sont des variables indépendantes } N(0,1);$$

$$y_{v,\tau} = z_{v,\tau} * s_{y\tau} + \bar{y}_{\tau}, \text{ sauf pour les mois d'octobre et décembre;}$$

$$x_{v,\tau} = \exp(y_{v,\tau}) + c_{\tau}, \text{ sauf pour les mois d'octobre et décembre;}$$

$$x_{v,\tau} = z_{\tau} * s_{x\tau} + \bar{x}_{\tau}, \text{ pour les mois d'octobre et décembre.}$$

La vérification des modèles et l'analyse des résidus

Les résidus pour chacun des douze modèles sont calculés par $\hat{\xi}_{v,\tau} = \frac{z_{v,\tau} - \hat{\phi}_{1,\tau} z_{v,\tau-1}}{\hat{\sigma}_{\varepsilon,\tau}}$. Pour vérifier leur normalité, une série $\hat{\xi}_t, t=1, \dots, (N\omega-1)$ est formée à partir de $\hat{\xi}_{v,\tau}$. La figure 17 présente la droite de Henri, et l'hypothèse de la normalité des résidus n'est donc pas rejetée.

Le test de corrélation critique est utilisé pour vérifier l'indépendance des résidus. Le tableau 10 contient les coefficients de corrélation d'ordre 1 des résidus $r_{1,\tau}(\hat{\xi}_t)$ qui sont tous inférieurs à la valeur critique 0.2 ($=r_{1-\alpha/2}$, où $\alpha = .05$); l'hypothèse d'indépendance des résidus n'est donc pas rejetée.

Coefficient d'autocorrélation mensuelles d'ordre 1 des résidus du modèle AR(1)V, $r_{1,\tau}(\hat{\xi}_t)$	
janvier-décembre	-.008
février-janvier	-.013
mars-février	.0213
avril-mars	-.038
mai-avril	-.001
juin-mai	-.012
juillet-juin	.0126
août-juillet	-.038
septembre-août	-.074
octobre-septembre	.013
novembre-octobre	.024
décembre-novembre	-.091

Tableau 10: Les autocorrélations périodiques $r_{1,\tau}(\hat{\xi}_t)$ des résidus du modèle AR(1)V

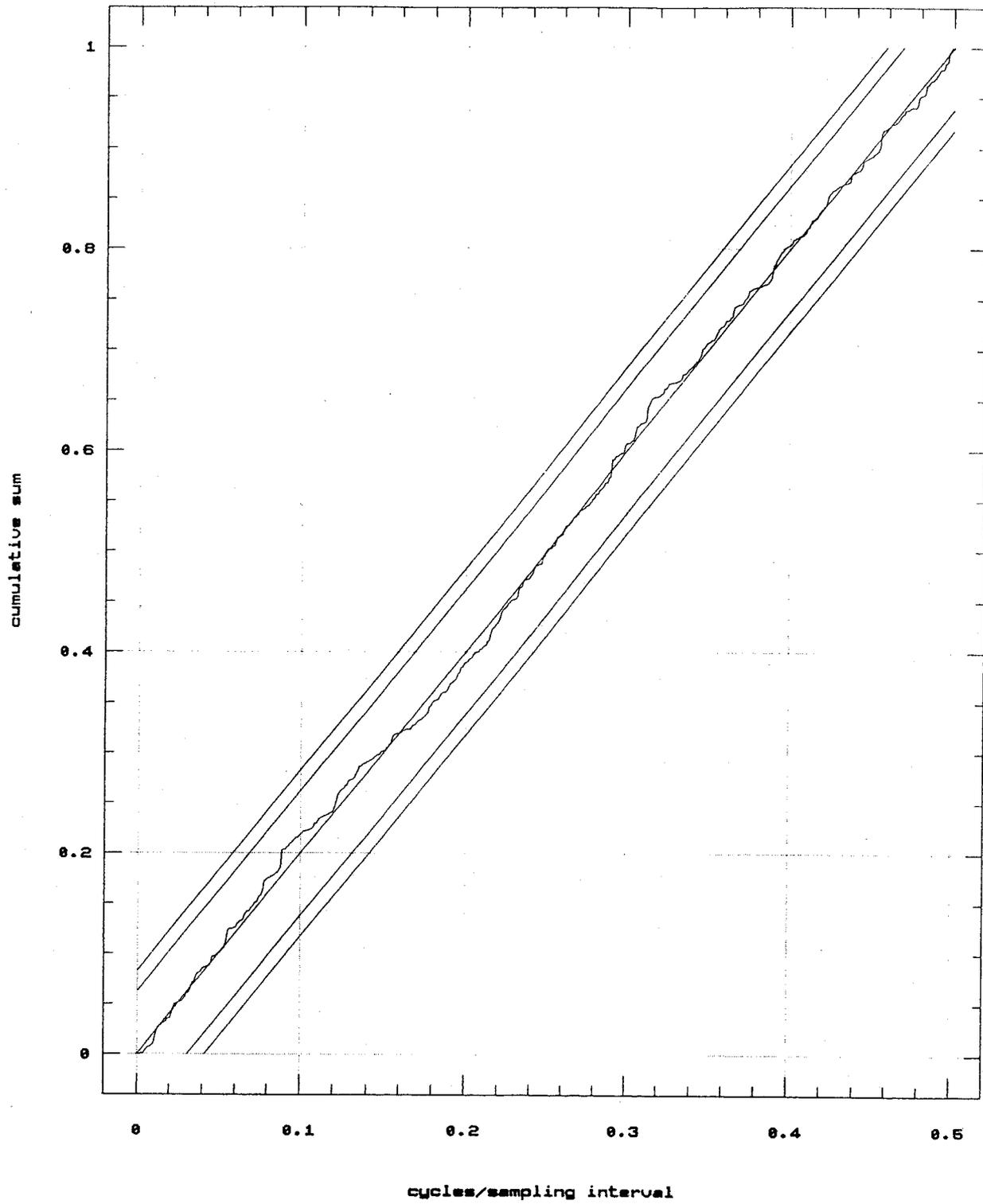


Figure 17: La droite de Henri des résidus du modèle AR(1)V

6.2.1.3 La pondération des données mensuelles générées

Afin de tenter de préserver les caractéristiques annuelles, les données mensuelles générées $x_{v,\tau}$ sont pondérées pour produire $x_{p_{v,\tau}}$ dont la somme sur un an est égale à x_v , l'apport net annuel correspondant généré par un modèle AR(0). Cette approche est appelée génération avec pondération, et elle est comparée à l'approche sans pondération. La série pondérée est $x_{p_{v,\tau}}$ obtenue par Harms et Cambell, (1967):

$$x_{p_{v,\tau}} = \left(\frac{x_v}{\sum_{\tau=1}^{12} x_{v,\tau}} \right) x_{v,\tau}, \quad v=1,..90, \text{ et } \tau=1,..12.$$

6.2.2 L'approche indirecte

L'approche indirecte utilisée est la méthode de désagrégation du logiciel SPIGOT. Les données annuelles originales Q_n , $n=1$ à 90, sont pondérées par la division par 87.01 telles que dans la section 6.1: $Z_n=Q_n/87.01$. Les données mensuelles originales $Q_{n,t}$, $t=1$ à 12, sont transformées selon une Lognormale-3 sauf pour les mois d'octobre et de décembre telles que: $Z_{n,t} = \ln (Q_{n,t} - \tau_t)$ (tableau 11). Les données annuelles normales pondérées Z_n sont générées à l'aide d'un modèle AR(0), $Z_n = 10 + V_n$, et sont ensuite désagrégées pour obtenir des données mensuelles normales $Z_{n,t}$ selon le modèle de la section 5.0 dont les équations sont présentées au tableau 11. Seuls les paramètres significatifs sont retenus. $V_{n,t}$ et V_n sont des variables aléatoires normales de moyenne 0, $\Lambda_{n,t} = \sum_{s=1}^{t-1} w_s Z_{n,s}$, et $w_t = \exp(\mu_t + \sigma_t^2)$ où μ_t et σ_t sont respectivement la moyenne et l'écart type des données mensuelles $Z_{n,t}$.

Mois	Seuil τ_t	Équation
octobre	-	$Z_{n,1} = -21.7 + 3.2 Z_n + V_{n,1}$
novembre	-107.8	$Z_{n,2} = 4.2 + .05 Z_n + V_{n,2}$
décembre	-	$Z_{n,3} = -44.4 + 1.9 Z_n + V_{n,3}$
janvier	-211.8	$Z_{n,4} = 5.2 + .01 Z_n + V_{n,4}$
février	-563.3	$Z_{n,5} = 6.5 + .007 Z_n + V_{n,5}$
mars	-287.3	$Z_{n,6} = 7.2 + .02 Z_n + V_{n,6}$
avril	-158.6	$Z_{n,7} = 6.4 + .02 Z_n + V_{n,7}$
mai	-122	$Z_{n,8} = 10.5 + .05 Z_n - .03 \Lambda_{n,8} + V_{n,8}$
juin	-60.7	$Z_{n,9} = 9.1 + .06 Z_n + V_{n,9}$
juillet	-54.9	$Z_{n,10} = 16.6 + .09 Z_n - .08 \Lambda_{n,10} + V_{n,10}$
août	-251	$Z_{n,11} = 14.4 + .06 Z_n - .04 \Lambda_{n,11} + V_{n,11}$
septembre	-194.2	$Z_{n,12} = 25.9 + .14 Z_n - .08 \Lambda_{n,12} + V_{n,12}$

Tableau 11: Les modèles de désagrégation et les seuils de transformation

Trente séries de 90 ans de données mensuelles $Q_{n,t}$ sont générées où:

$$Q_{n,t} = \exp(Z_{n,t}) + \tau_t.$$

7. RÉSULTATS

7.1 Les caractéristiques statistiques de cette étude

Dans cette étude, les critères de validation sont de nature statistique et ne dépendent pas de l'application. Ils sont basés sur les critères utilisés dans le travail de Rassam *et al.* (1992). Les caractéristiques mensuelles sont la moyenne périodique \bar{x}_τ , l'écart type périodique s_τ , le coefficient de corrélation d'ordre 1 $r_{1,\tau}$ entre les mois. Les caractéristiques annuelles sont la moyenne \bar{x} , l'écart type s , le coefficient d'asymétrie g , le coefficient de Hurst h , les coefficients d'autocorrélation d'ordre 1 jusqu'à 14, les courbes de fréquence de période de retour, la durée maximale de séquence de surplus (*surplus run length* (RL)) et la somme maximale de l'apport net en surplus (*surplus run sum* (RS)). Dans les deux derniers cas les seuils utilisés sont la moyenne traditionnelle ainsi que la moyenne mobile sur quatre ans. Les expressions mathématiques de ces caractéristiques telles qu'utilisées dans cette étude sont définies dans l'annexe 1. Des graphiques des données historiques en ordre décroissant, des 25^{ème} et 75^{ème} quantiles, et de la médiane des moyennes ordonnées des séries générées sont aussi inclus.

Les données mensuelles sont générées par les cinq méthodes: la méthode indirecte (SPIGOT), la méthode AR(2) à paramètres constants avec et sans pondération, la méthode AR(1)V à paramètres variables avec et sans pondération. Dans ce qui suit, AR(2)np représente le modèle AR(2) à paramètres constants sans pondération, AR(2)p est le modèle AR(2) à paramètres constants avec pondération, AR(1)Vnp est le modèle AR(1) à paramètres variables sans pondération, AR(1)Vp est le modèle AR(1) à paramètres variables avec pondération, et SPIGOT est le modèle de désagrégation. Quatre modèles génèrent les données annuelles: les données annuelles des modèles AR(1)Vp et AR(2)p proviennent d'un modèle AR(0) qui est utilisé pour la pondération (x_v de l'équation de la section 6.2.1.3); les données annuelles de SPIGOT sont aussi générées par un AR(0) mais par un logiciel différent; et les données annuelles de AR(1)Vnp et AR(2)np correspondent aux sommes des données mensuelles générées par les modèles respectifs. Toutes les statistiques sont calculées à l'aide d'un programme C sur MACINTOSH.

Les tableaux 12 et 13 présentent une comparaison entre les caractéristiques des données annuelles historiques et simulées. La moyenne, l'écart type, les maxima et minima des caractéristiques statistiques annuelles et mensuelles des trente séries générées sont calculés.

On note qu'en général, les quatre modèles annuels reproduisent assez bien les caractéristiques annuelles.

Statistiques annuelles générées						
Historique			SPIGOT	AR (2) np	AR(0)	AR(1)Vnp
Moyenne	870.14	moyenne	878.72	874.51	868.64	874.89
		écart type	31.46	16.12	24.46	20.09
		max	983.42	911.46	910.76	922.01
		min	821.81	851.40	816.02	846.37
Écart type	204.03	moyenne	208.74	146.78	202.24	174.94
		écart type	25.38	12.11	21.92	13.46
		max	258.79	175.54	271.10	207.63
		min	163.01	126.26	169.95	155.63
Asym	.03	moyenne	.03	0.07	0.02	0.07
		écart type	.25	0.30	0.22	0.26
		max	.46	0.69	0.47	0.52
		min	-.5	-0.47	-0.45	-0.47
Hurst	.55	moyenne	.53	0.53	0.53	0.54
		écart type	.05	0.04	0.05	0.04
		max	.63	0.59	0.61	0.61
		min	.43	0.44	0.43	0.44

Tableau 12: Les statistiques annuelles historiques et générées

Maximum Historique			Statistiques annuelles générées								
			SPIGOT		AR(2)np		AR(0)		AR(1)Vnp		
1 an	4 ans		1 an	4 ans	1 an	4 ans	1 an	4 ans	1 an	4 ans	
RL (an)	7	12	moy.	6.3	12.97	6.03	12.90	5.43	10.97	6.53	12.43
			é.t.	2.82	6.71	1.88	4.96	1.36	3.35	2.18	4.95
			max	17.0	40.0	13.00	25.00	8.00	19.00	13.0	24.0
			min	3.0	6.0	4.00	5.00	3.00	6.00	4.0	5.0
RS (pcs)	858.76	1303.9	moy.	1227.2	1450.4	833.47	992.60	1056.6	1158.0	1097.6	1248.9
			é.t.	583.13	896.7	291.71	444.95	267.52	489.38	358.7	590.3
			max	2866.2	5244.9	1668.8	2445.6	1573.3	2501.5	2183.7	3129.5
			min	411.95	395.66	453.49	319.97	663.31	499.58	582.9	361.4

Tableau 13: Les statistiques de surplus historiques et générées

Les tableaux 14 à 18 présentent la comparaison entre les moyennes et les écarts types mensuels générés et historiques pour chacune des cinq méthodes mensuelles. Toutes les méthodes semblent reproduire adéquatement les statistiques historiques.

Mois	Moyennes mensuelles, AR(2)np					Écarts types mensuels, AR(2)np				
	moy	é.t.	max	min	hist	moy	é.t.	max	min	hist
jan=1	-13.1	2.06	-8.8	-16.99	-13.5	24.7	1.86	29.03	22.17	24.3
2	10.17	2.34	14.98	5.36	10.43	27.6	2.35	33.36	23.5	27.9
3	47.4	4.85	56.38	36.27	45.29	43.6	2.86	48.67	38.69	42.8
4	150.2	6.35	164.86	139.04	148.7	50.0	4.92	57	38	49.7
5	186.8	5.06	197.51	176.78	185.6	57.9	4.92	66.42	48.7	59.2
6	157.8	5.05	169.93	146.28	157.7	53.0	4.46	63.17	44.15	52.2
7	130.3	3.45	136.86	123.38	129.6	41.3	2.88	49.53	34.02	40.0
8	100.3	5.3	111.81	90.05	100.6	43.7	3.6	51.21	35.95	43.0
9	73.21	6.96	86.27	60.97	73.31	56.0	4.83	68.28	46.58	55.9
10	37.41	6.01	46.93	25.94	37.88	46.6	3.83	55.65	40.08	46.5
11	18.65	4.66	30.33	8.52	18.2	44.2	4.67	54.4	35.46	44.5
déc=12	-24.8	2.53	-20.48	-30.47	-23.6	29.1	2.3	34.15	25.11	28.9

Tableau 14: Les statistiques mensuelles générées, AR(2)np

Mois	Moyennes mensuelles, AR(2)p					Écarts types mensuels, AR(2)p				
	moy	é.t.	max	min	hist	moy	é.t.	max	min	hist
jan=1	-14.3	2.72	-8.37	-22.68	-13.57	27.1	3.02	34.34	21.83	24.31
2	9.91	2.62	15.85	4.58	10.43	29.2	3.16	34.7	23.61	27.93
3	46.39	5.28	56.47	30.11	45.29	45.0	4.15	55.23	37.32	42.8
4	150.9	6.32	166.01	134.13	148.7	61.4	5.62	75.68	48.75	49.76
5	188.0	8.92	216.6	166.6	185.64	73.5	6.88	86.38	59.01	59.25
6	158.5	6.71	173.57	146.92	157.7	66.7	7.17	86.86	57.05	52.27
7	131.1	5.76	147.6	119.8	129.57	53.5	5.69	66.47	42.16	40.09
8	101.0	4.87	111.77	88.78	100.62	49.7	5.47	64.98	39.33	43.05
9	70.95	5.51	79.31	55.92	73.31	56.7	4.86	65.27	47.84	55.9
10	35.19	6.02	46.78	23.04	37.88	49.8	5.87	62.83	38.22	46.5
11	17.24	5.58	29.28	6.4	18.2	45.9	4.82	54.2	37.46	44.56
déc=12	-26.5	3.31	-19.4	-31.23	-23.63	33.4	3.16	39.58	25.29	28.91

Tableau 15: Les statistiques mensuelles générées, AR(2)p

Mois	Moyennes mensuelles, AR(1)Vnp					Écarts types mensuels, AR(1)Vnp				
	moy	é.t.	max	min	hist	moy	é.t.	max	min	hist
jan=1	-13.1	2.07	-9.18	-16.58	-13.5	24.7	1.94	29.56	22.14	24.3
2	10.08	2.3	14.75	5.26	10.43	27.5	2.39	33.14	23.55	27.9
3	47.33	5.04	56.77	36.59	45.29	43.8	2.98	49.35	37.51	42.8
4	150.4	6.3	165.12	140	148.7	50.0	5.01	57.2	37.34	49.7
5	187.0	5.08	197.37	176.9	185.6	57.9	4.88	66.19	49.19	59.2
6	158.2	4.88	170.18	148.3	157.7	53.0	4.6	63.66	43.74	52.2
7	130.3	3.4	136.83	123.6	129.5	41.3	2.82	49.54	34.71	40.0
8	100.4	5.17	111.5	91.16	100.6	43.9	3.84	51.58	34.77	43.0
9	73.5	6.78	85.79	58.97	73.31	55.9	5.13	68.39	46.5	55.9
10	37.34	5.3	46.78	28.34	37.88	46.6	4.05	54.94	37.7	46.5
11	18.48	4.3	27.94	9.23	18.2	43.9	4.13	52.55	35.27	44.5
déc=12	-24.8	2.6	-20.23	-30.39	-23.6	29.1	2.44	34.19	24.99	28.9

Tableau 16: Les statistiques mensuelles générées, AR(1)Vnp

Mois	Moyennes mensuelles, AR(1)Vp					Écarts types mensuels, AR(1)Vp				
	moy	é.t.	max	min	hist	moy	é.t.	max	min	hist
jan=1	-15.0	2.96	-8.65	-23.91	-13.5	28.7	4.55	44.04	22.49	24.3
2	9.32	2.85	15.42	2.26	10.4	29.8	4.49	45.76	23.44	27.9
3	46.06	5.28	56.94	29.58	45.29	45.3	3.91	55.32	38.24	42.8
4	152.6	6.43	166.52	134.77	148.7	63.6	6.07	77.66	49.11	49.7
5	189.6	9.07	220.2	167.78	185.6	75.5	7.65	92.23	58.1	59.2
6	160.1	6.93	176.58	148.22	157.7	68.1	8.05	90.23	57.07	52.2
7	132.0	5.83	149.55	122.15	129.5	54.6	6.77	73.19	41.87	40.0
8	100.8	4.51	110.12	89.55	100.6	48.4	5.47	62.21	37.99	43.0
9	70.62	5.45	78.78	56.97	73.31	55.8	6.04	69.47	44.14	55.9
10	34.21	5.84	46.79	25.7	37.88	51.5	10.0	90.81	36.35	46.5
11	15.8	5.36	28.59	4.38	18.2	46.7	6.05	65.24	37.57	44.5
déc=12	-27.6	3.62	-19.56	-32.8	-23.6	35.6	4.52	47.52	23.43	28.9

Tableau 17: Les statistiques mensuelles générées, AR(1)Vp

Mois	Moyennes mensuelles, SPIGOT					Écarts types mensuels, SPIGOT				
	moy	é.t.	max	min	hist	moy	é.t.	max	min	hist
jan=1	-13.4	3.2	-6.45	-18.58	-13.5	25.8	3.04	32.42	19.67	24.3
2	9.2	4.38	17.33	-3.22	10.43	28.8	2.36	33.15	23.78	27.9
3	46.12	6.77	58.68	31.48	45.29	42.8	5.4	57.89	35.03	42.8
4	148.7	9.31	172.39	128.84	148.7	50.5	5.9	68.94	42.78	49.7
5	186.2	8.7	203.83	171.34	185.6	60.3	6.34	78.44	48.76	59.2
6	157.0	7.79	173.99	138.35	157.7	53.0	6.73	70.53	41.38	52.2
7	131.5	5.35	140.69	121.15	129.5	41.5	5.54	56.4	32.5	40.0
8	103.7	8.56	125.27	88.32	100.6	44.1	4.03	53.61	36.35	43.0
9	73.31	7.66	89.56	55.72	73.31	56.5	7.83	72.56	46.45	55.9
10	38.51	7.4	54.3	24.61	37.88	47.1	4.97	61.81	37.32	46.5
11	20.42	6.34	36.05	7.99	18.2	47.3	6.56	62.99	36.63	44.5
déc=12	-22.7	4.19	-17.17	-33.8	-23.6	27.4	2.73	35.54	21.58	28.9

Tableau 18: Les statistiques mensuelles générées, SPIGOT

Les figures 18 à 21 présentent les corrélogrammes annuels d'ordres 1 à 14 de la moyenne des trente séries générées.

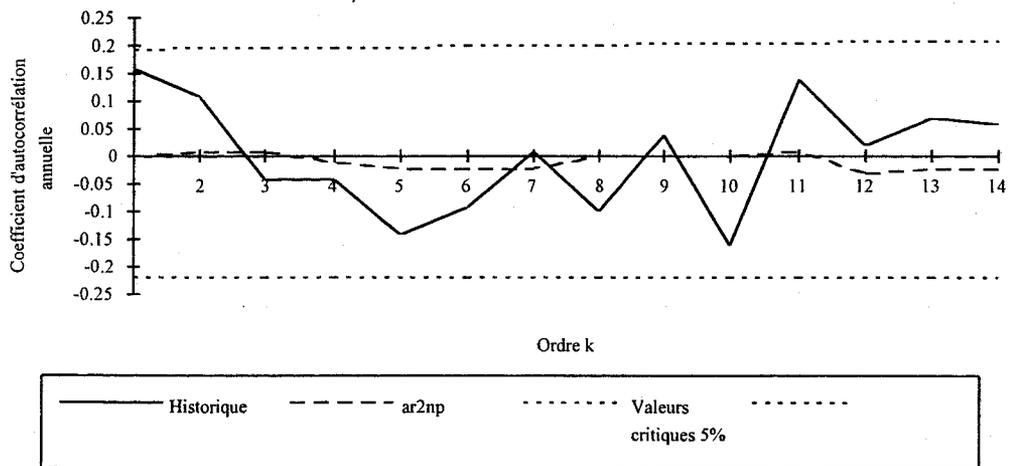


Figure 18: Les autocorrélations annuelles, AR(2)np

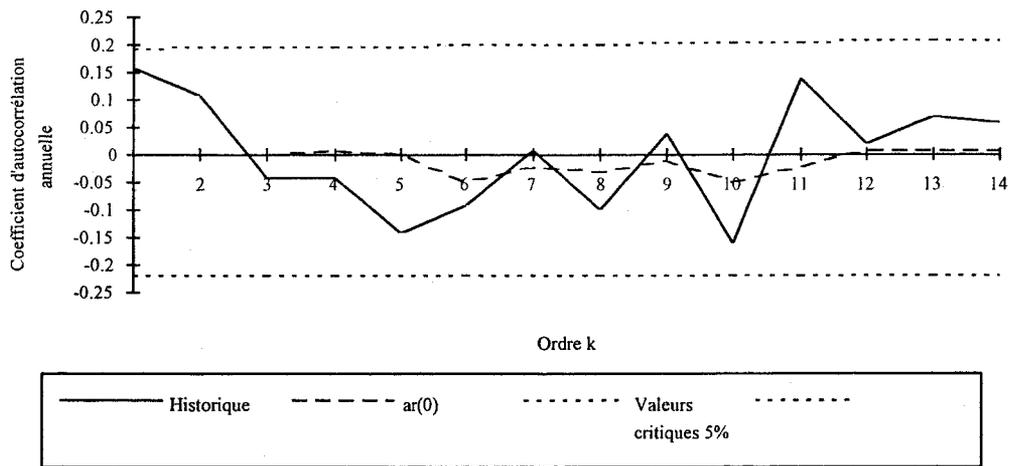


Figure 19: Les autocorrélations annuelles, AR(0)

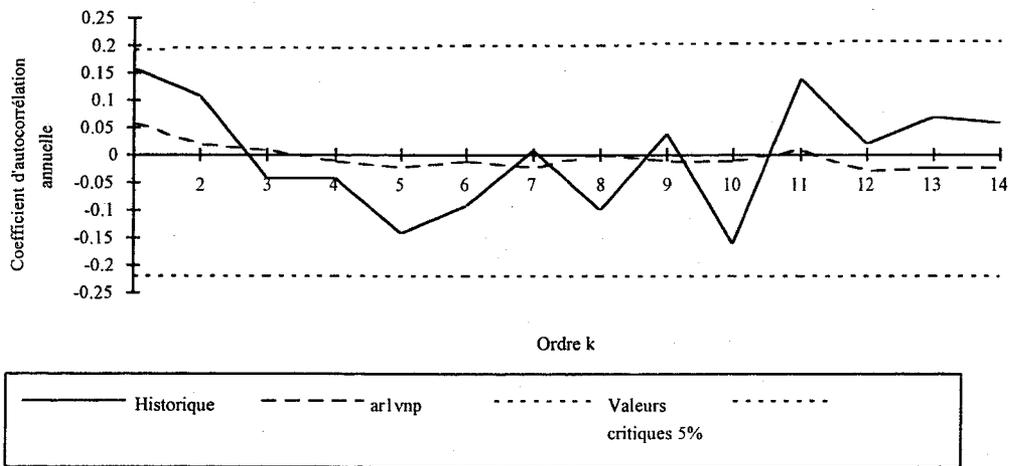


Figure 20: Les autocorrélations annuelles, AR(1)Vnp

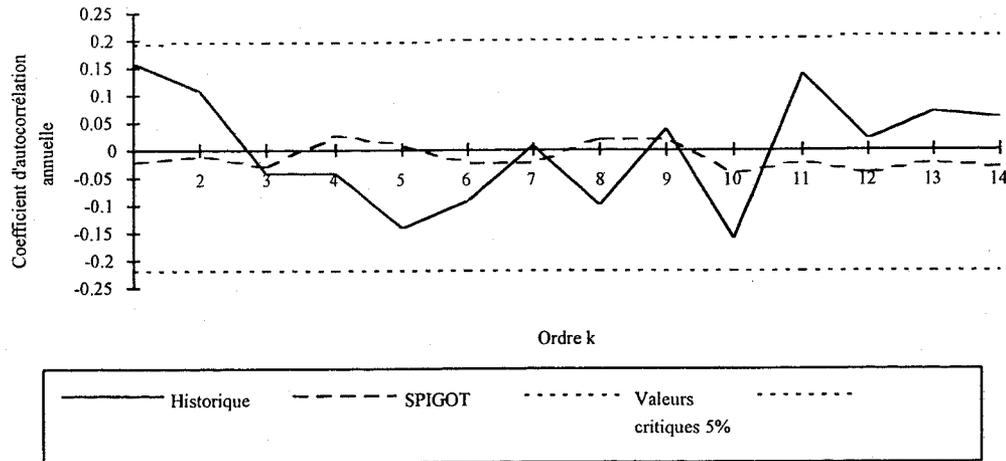


Figure 21: Les autocorrélations annuelles, AR(0) SPIGOT

En examinant les graphiques des autocorrélations annuelles aux figures 18 à 21, il est difficile de déceler au premier coup d'oeil une différence significative entre les quatre méthodes annuelles. De toute façon, à l'instar des données historiques, les données annuelles générées par les quatre méthodes annuelles n'exhibent pas d'autocorrélation significative.

Les figures 22 à 29 présentent les graphiques des fréquences et les graphiques des périodes de retour. Chaque série générée $\{x_{ij}\}$ est ordonnée en ordre décroissant par rapport aux 90 années pour donner $\{x_{ord_a_{ij}}\}$ ($i=1,..90$ est l'indice des années, $j = 1,..30$ est l'indice des séries); ensuite les moyennes $\{x_{moy_a_i}\}$ des trente séries ordonnées $x_{ord_a_{ij}}$ sont calculées pour chacune des 90 années et sont mises en ordre décroissant. C'est cette série finale qui est utilisée pour tracer le graphique de période de retour des données générées. Pour les graphiques de fréquence, la série $\{x_{ord_a_{ij}}\}$ est mise en ordre décroissant par rapport aux trente séries (indice j) pour former la série $x_{ord_ord_a_{ij}}$; les médianes, les 25^{ième} et 75^{ième} quantiles sont ensuite extraits des trente séries $\{x_{ord_ord_a_{ij}}\}$ pour chaque année i , et tracés en fonction de l'année.

Les graphiques des fréquences (figures 22 à 25) indiquent que les méthodes AR(2)_{np} et AR(1)_{Vnp} surestiment les faibles extrêmes des apports nets annuels et sous-estiment les extrêmes élevés; ces deux méthodes n'utilisent pas la pondération annuelle. Les graphiques des périodes de retour (figures 26 à 29) suggèrent que AR(2)_{np} et AR(1)_{Vnp} génèrent des données plus faibles que les données historiques dans le cas d'une période de retour

supérieure à dix ans. Il est difficile de discriminer sur ce point entre les deux autres approches par l'inspection visuelle.

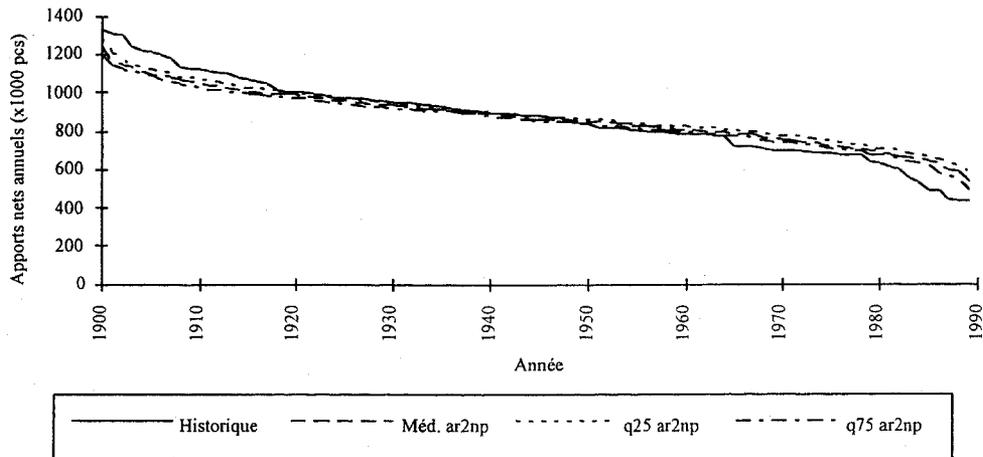


Figure 22: Le graphique des fréquences des données annuelles AR(2)np

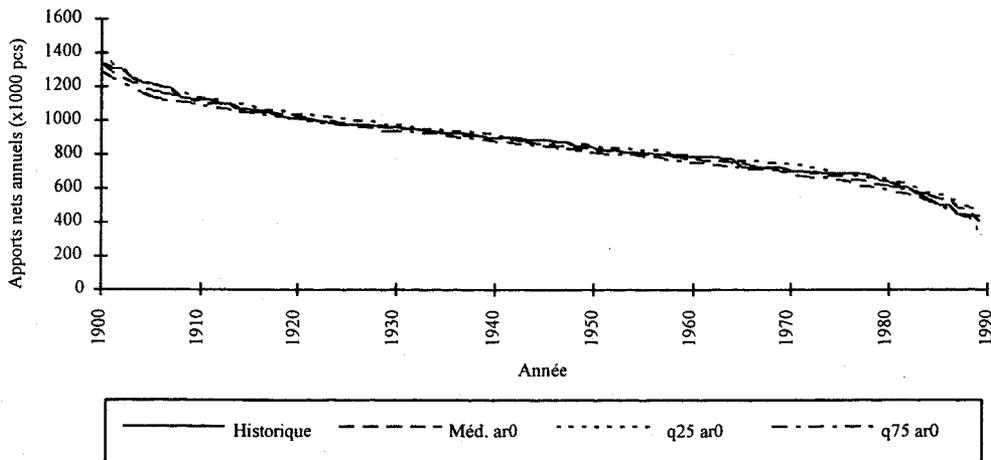


Figure 23: Le graphique des fréquences des données annuelles AR(0)

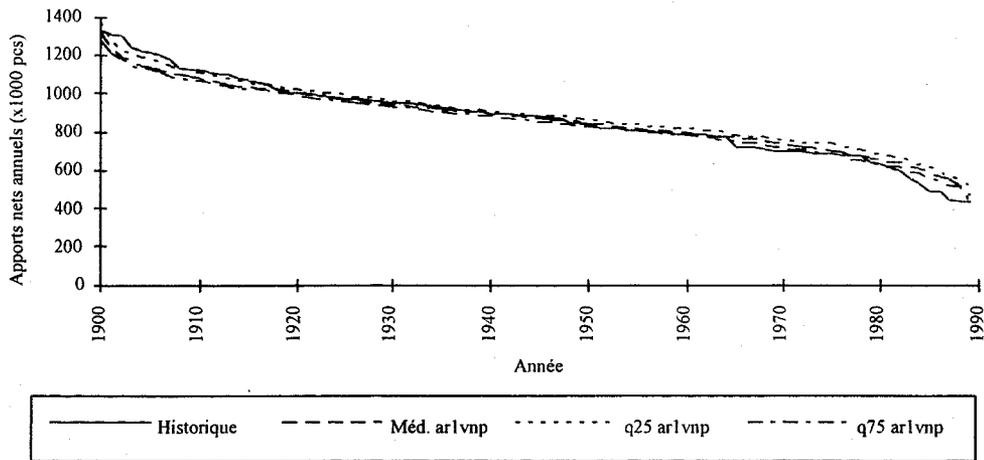


Figure 24: Le graphique des fréquences des données annuelles AR(1)Vnp

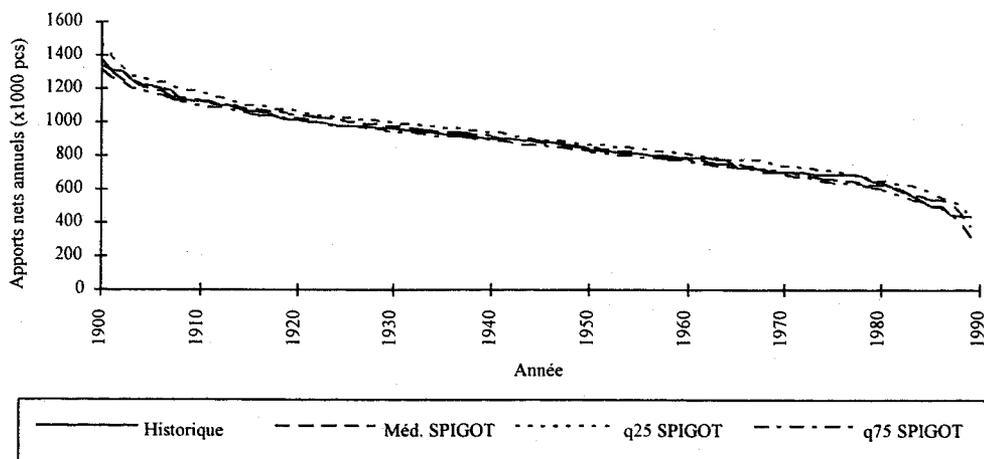


Figure 25: Le graphique des fréquences des données annuelles SPIGOT

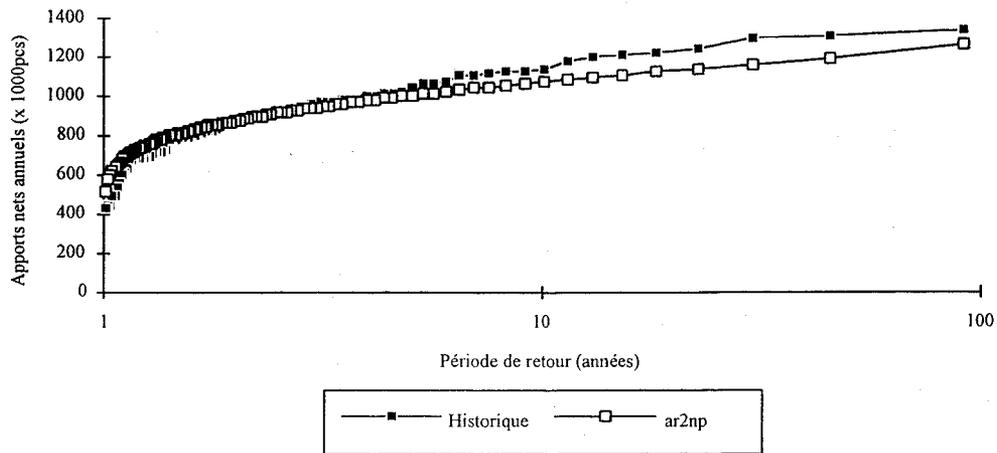


Figure 26: Le graphique de la période de retour des données annuelles AR(2)np

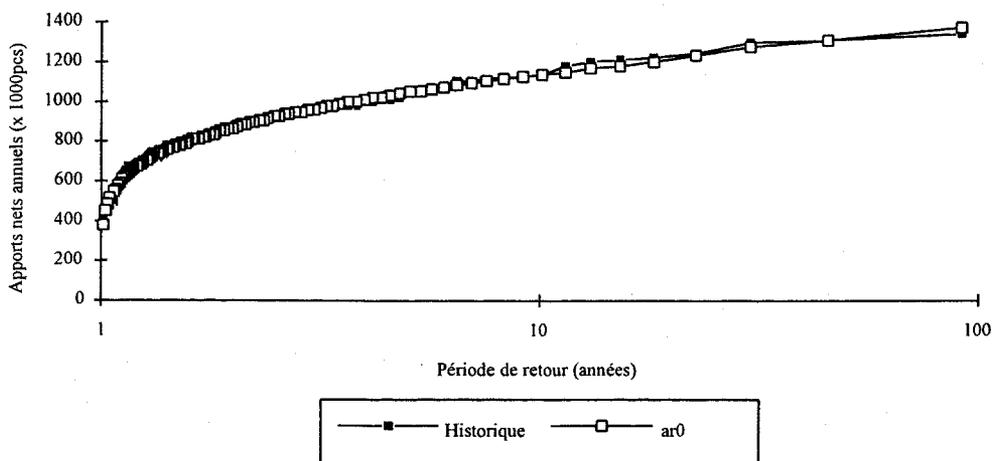


Figure 27: Le graphique de la période de retour des données annuelles AR(0)

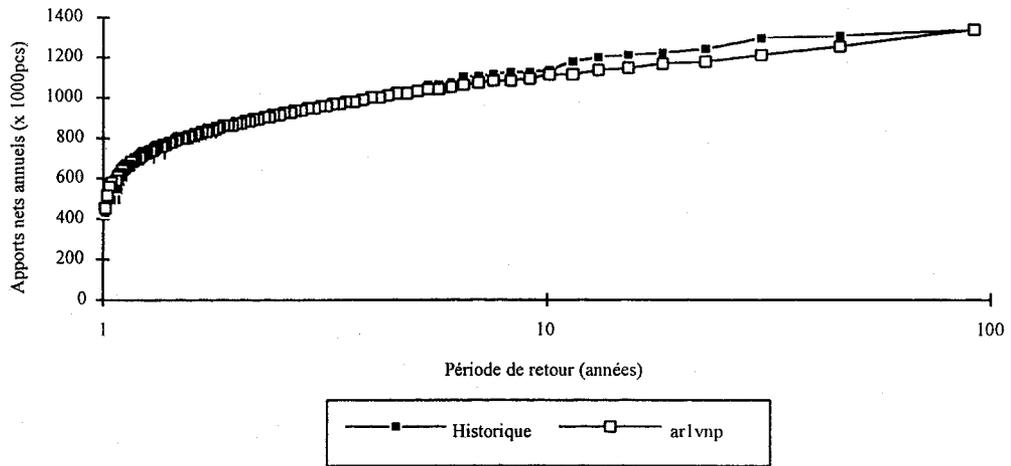


Figure 28: Le graphique de la période de retour des données annuelles AR(1)Vnp

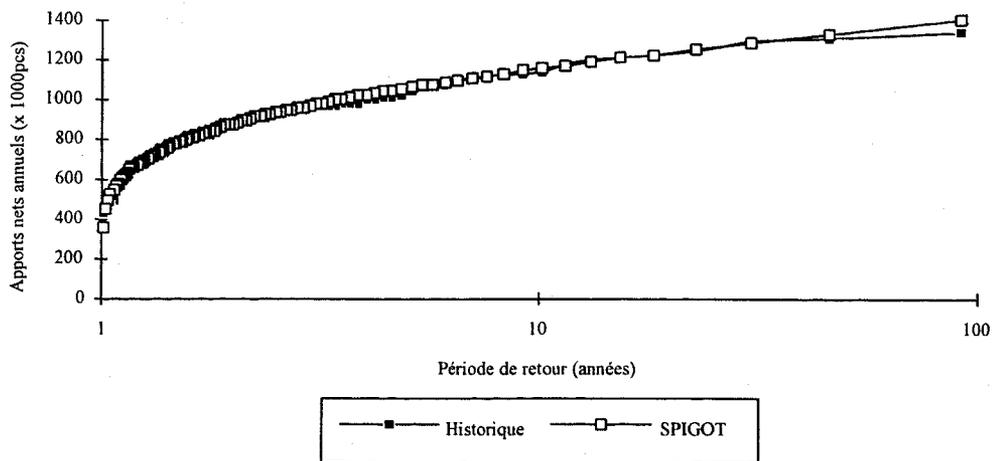


Figure 29: Le graphique de la période de retour des données annuelles SPIGOT

Les figures 30 à 32 présentent les autocorrélations mensuelles d'ordre 1, $r_{1,t}$ pour les cinq méthodes. On note que AR(1)Vnp reproduit très bien ces corrélations, ce qui est normal puisque c'est un modèle d'ordre 1. AR(1)Vp et AR(2)p créent des corrélations significatives là où il n'y en a pas (les mois 5 à 7), et SPIGOT suit assez bien le corrélogramme historique mensuel d'ordre 1 mais le surestime pour les mois 5 à 8 et le sous-estime pour les mois 9 et 10. Les figures 33 à 35 présentent les autocorrélations mensuelles d'ordre 2, $r_{2,t}$, qui sont en général non significatives pour les données historiques. Les méthodes sans pondération AR(1)Vp et AR(2)p donnent des corrélations autour de zéro pour tous les mois, donc non significatives. Les méthodes avec pondération AR(1)Vnp et AR(2)np ne reproduisent pas bien les corrélations d'ordre 2, et les surestiment pour les mois 6 à 8. En général, seul SPIGOT reproduit bien les corrélations mensuelles d'ordre 2.

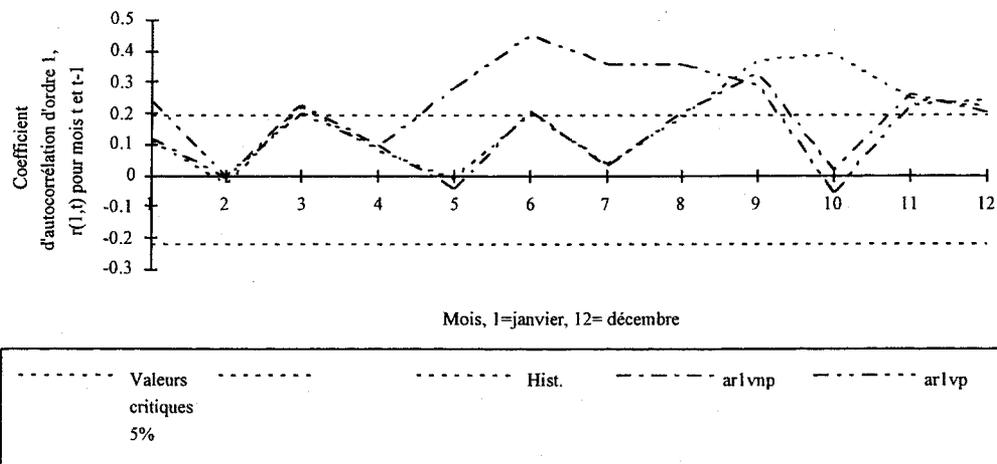


Figure 30: Les autocorrélations mensuelles d'ordre 1, AR(1)Vp et AR(1)Vnp

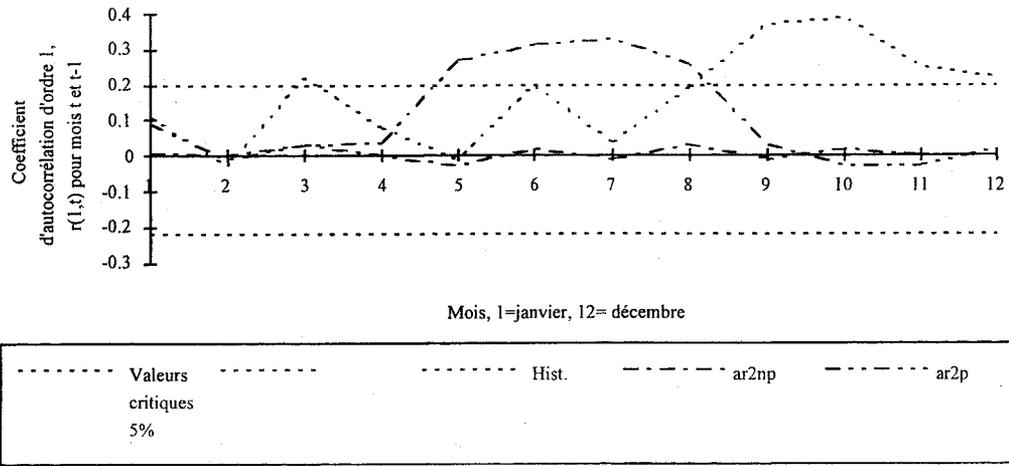


Figure 31: Les autocorrélations mensuelles d'ordre 1, AR(2)p et AR(2)np

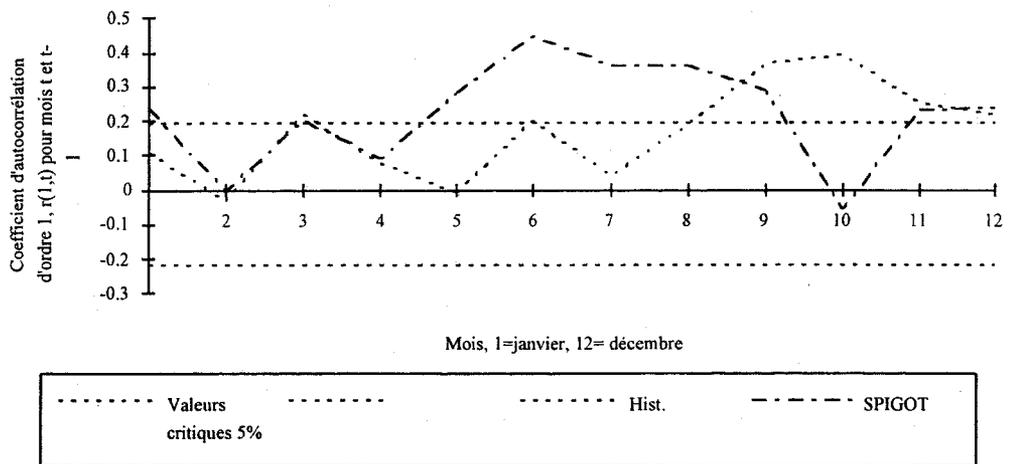


Figure 32: Les autocorrélations mensuelles d'ordre 1, SPIGOT

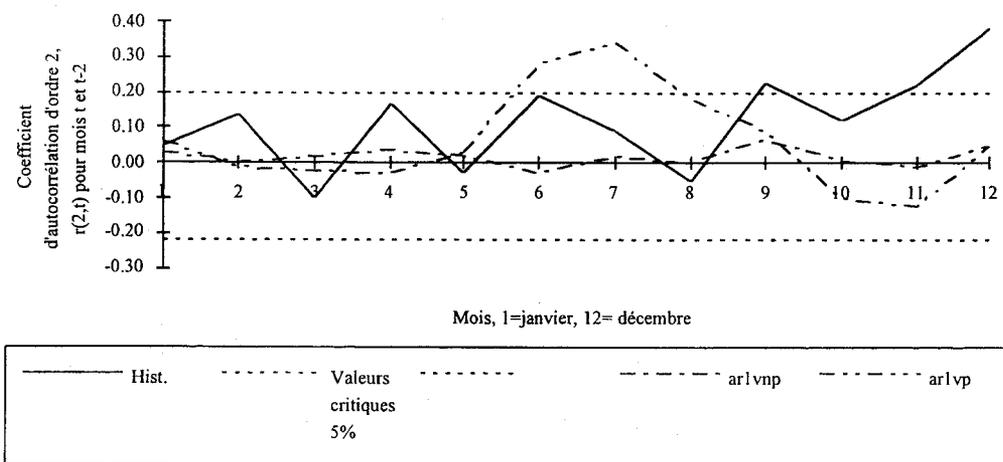


Figure 33: Les autocorrélations mensuelles d'ordre 2, AR(1)Vp et AR(1)Vnp

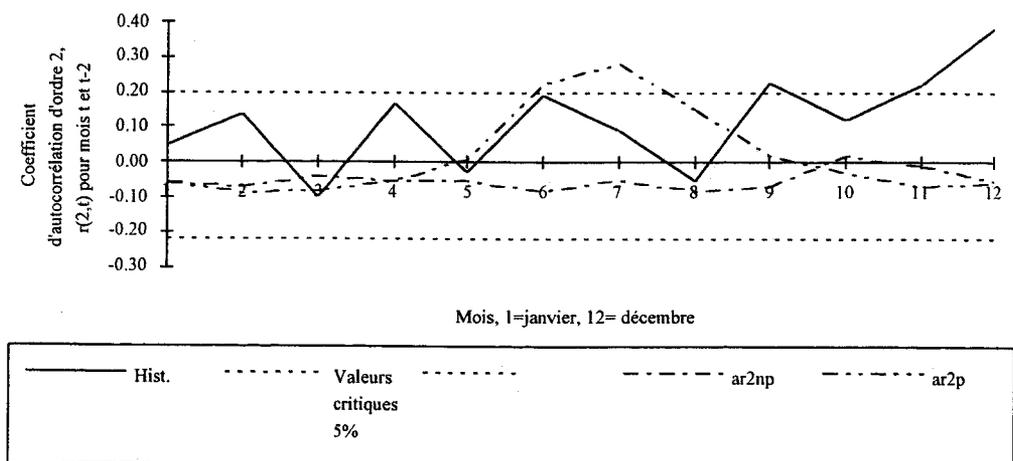


Figure 34: Les autocorrélations mensuelles d'ordre 2, AR(2)p et AR(2)np

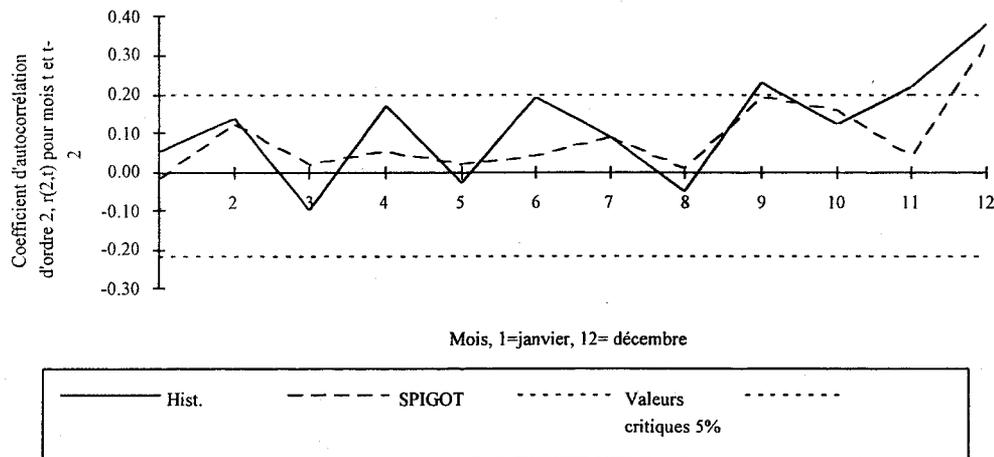


Figure 35: Les autocorrélations mensuelles d'ordre 2, SPIGOT

7.2 Comparaison des performances des modèles

Tous les modèles stochastiques ont leurs limitations. On suppose en général que les modèles $AR(p)$ préservent la moyenne historique, l'écart type et les premiers p coefficients d'autocorrélation. En ce qui concerne la structure de corrélation, il est préférable d'utiliser le terme ressemblance dans le sens que le corrélogramme des données générées par le modèle et celui des données historiques auront des formes et des ordres de grandeur similaires.

La plupart des travaux de comparaison des méthodes par rapport à un ou plusieurs critères utilisent une approche ad hoc de comparaison des statistiques générées et des statistiques historiques. À la connaissance des auteurs, seules les études de Rassam *et al.* (1992) et de Mathier *et al.* (1992), ont utilisé une approche formelle de comparaison par la méthode multicritère du *compromise programming*. Un score de 1, 0, -1 est alloué par des experts à chacune des méthodes par rapport à chacune des caractéristiques retenues, et un score global est ensuite calculé pour chacune des méthodes. Ce score global est une somme pondérée des écarts entre les scores obtenus et les scores idéaux. L'approche qui obtient le plus bas score global, c'est à dire qui est en moyenne la plus proche de l'idéal, est considérée la meilleure. À noter que cette approche est quand même subjective dans l'allocation des scores -1,0, et 1.

Il est clair qu'une approche formelle est nécessaire pour choisir une "meilleure" méthode si celle-ci existe. Dans ce travail, un score reflétant la performance de chaque méthode est calculé en utilisant les écarts pondérés entre les statistiques générées et les statistiques historiques mensuelles et annuelles (erreurs). Des erreurs relatives sont calculées pour chacune des statistiques annuelles ainsi que pour les moyennes et les écarts types mensuels, et des erreurs absolues sont calculées pour les autocorrélations mensuelles d'ordre 1 et 2:

$$\text{Erreur relative} = \frac{(\text{valeur historique} - \text{moyenne des statistiques générées})}{(\text{valeur historique})}$$

$$\text{Erreur absolue} = (\text{valeur historique} - \text{moyenne des statistiques générées}).$$

Les valeurs suivies de deux astérisques aux tableaux 19 à 24 indiquent les modèles qui ont obtenu le meilleur score (la plus faible erreur) par rapport au critère correspondant.

Le tableau 19 présente les valeurs absolues des erreurs relatives pour les statistiques mensuelles (moyenne, écart type), le tableau 20 contient les valeurs absolues des erreurs absolues des coefficients de corrélation d'ordre 1 et d'ordre 2, et le tableau 21 contient les valeurs absolues des erreurs relatives des statistiques annuelles (moyenne, écart type, coefficient d'asymétrie, coefficient de Hurst, durée maximale des séquences de surplus, et somme maximale de l'apport net en surplus).

On constate au tableau 19 que le modèle SPIGOT produit des moyennes mensuelles proches des moyennes historiques pour le plus grand nombre de mois (cinq mois) et que AR(1)Vnp et AR(2)np produisent des écarts types mensuels proches des écarts types historiques pour le plus grand nombre de mois (six mois).

Valeurs absolues des erreurs relatives des statistiques mensuelles					
Statistiques	AR(2)np	AR(2)p	AR(1)Vp	AR(1)Vnp	SPIGOT
Moy. jan.	0.035	0.057	0.108	0.028	0.010**
Moy. fév.	0.025**	0.050	0.106	0.034	0.118
Moy. mar	0.047	0.024	0.017**	0.045	0.018
Moy. avr.	0.010	0.015	0.026	0.012	0.000**
Moy. mai	0.007	0.013	0.022	0.007	0.003**
Moy. juin	0.001**	0.006	0.015	0.003	0.004
Moy. juil	0.006**	0.012	0.019	0.006**	0.016
Moy. aou	0.003	0.004	0.002	0.001**	0.031
Moy. sep	0.001	0.032	0.037	0.003	0.000**
Moy. oct	0.012**	0.071	0.097	0.014	0.017
Moy. nov	0.025	0.053	0.132	0.015**	0.122
Moy. déc	0.052	0.124	0.168	0.050	0.039**
É.t. jan	0.018	0.118	0.183	0.017**	0.061
É.t. fév	0.012**	0.046	0.068	0.015	0.031
É.t. mar	0.019	0.053	0.059	0.024	0.002**
É.t. avr.	0.005**	0.235	0.280	0.005**	0.017
É.t. mai	0.022	0.241	0.275	0.022	0.019**
É.t. juin	0.015**	0.278	0.304	0.016	0.015**
É.t. juil	0.032	0.335	0.363	0.030**	0.031
É.t. aou	0.017**	0.155	0.126	0.021	0.026
É.t. sep	0.003	0.015	0.001**	0.001**	0.012
É.t. oct	0.003**	0.071	0.108	0.003**	0.015
É.t. nov	0.007**	0.031	0.049	0.013	0.063
É.t. déc	0.010	0.156	0.233	0.007**	0.051

Tableau 19: Les valeurs absolues des erreurs relatives des statistiques mensuelles

On constate au tableau 20 que le modèle AR(1)Vnp produit des corrélations mensuelles d'ordre 1 proches des corrélations historiques pour le plus grand nombre de mois (9 mois) et que SPIGOT produit des corrélations mensuelles d'ordre 1 proches des corrélations historiques pour le plus grand nombre de mois (7 mois).

Valeurs absolues des erreurs absolues des coefficients de corrélation mensuels d'ordre 1 et 2					
Statistiques	AR(2)np	AR(2)p	AR(1)Vp	AR(1)Vnp	SPIGOT
Corr. 1 jan-déc	.1	.02	.13	.01**	.02
Corr. 1 fév-jan	.03	.02**	.03	.02**	.04
Corr. 1 mar-fév	0.190	0.190	0.020	0.010**	0.180
Corr. 1 avr-mar	0.080	0.040	0.010**	0.020	0.070
Corr. 1 mai-avr	0.020**	0.280	0.290	0.030	0.030
Corr. 1 jun-mai	0.180	0.110	0.250	0.010**	0.110
Corr. 1 jui-jun	0.050	0.290	0.320	0.010**	0.050
Corr. 1 aou-jui	0.160	0.070	0.170	0.010**	0.010**
Corr. 1 sep-aou	0.380	0.340	0.080	0.040**	0.180
Corr. 1 oct-sep	0.370	0.420	0.450	0.370	0.180**
Corr. 1 nov-oct	0.250	0.280	0.020	0.010**	0.020
Corr. 1 déc-nov	0.220	0.200	0.020**	0.020**	0.020**
Corr. 2 jan-nov	0.110	0.110	0.010**	0.020	0.070
Corr. 2 fév-déc	0.210	0.230	0.150	0.140	0.020**
Corr. 2 mar-jan	0.060	0.020**	0.080	0.120	0.120
Corr. 2 avr-fév	0.220	0.220	0.200	0.130	0.120**
Corr. 2 mai-mar	0.020**	0.040	0.060	0.050	0.050
Corr. 2 jun-avr	0.270	0.030**	0.090	0.220	0.150
Corr. 2 jui-mai	0.140	0.190	0.250	0.070	0.000**
Corr. 2 aou-jun	0.030**	0.200	0.230	0.050	0.060
Corr. 2 sep-jui	0.300	0.210	0.140	0.160	0.040**
Corr. 2 oct-aou	0.100	0.150	0.220	0.110	0.040**
Corr. 2 nov-sep	0.230	0.290	0.340	0.230	0.180**
Corr. 2 déc-oct	0.430	0.440	0.330	0.330	0.050**

Tableau 20: Les valeurs absolues des erreurs absolues des coefficients de corrélation mensuels

Valeurs absolues des erreurs relatives des statistiques annuelles				
	AR(2)np	AR(0)	AR(1)Vnp	SPIGOT
Moyenne annuelle	.005	.001**	.005	.009
Écart type annuel	.28	.008**	.14	.023
Coefficient d'asym.	1.33	.33	1.3	0**
Coefficient de Hurst	.03	.036	.02**	.03
RL (1)	.14	.22	.21	.1**
RS(1)	.029**	.22	.27	.43
RL(4)	.07	.08	.04**	.08
RS(4)	.23	.11	.04**	.11

Tableau 21: Les erreurs relatives des statistiques annuelles

Afin de résumer toutes les mesures de performance (erreurs), une mesure globale est calculée pour chacune des méthodes pour les statistiques mensuelles, les coefficients de corrélation mensuels et les statistiques annuelles. La mesure globale est la moyenne des valeurs absolues des erreurs relatives (ou absolues). Les erreurs moyennes des statistiques mensuelles individuelles, des statistiques annuelles globales et des statistiques mensuelles globales sont calculées et présentées aux tableaux 22 à 24. La méthode qui a la plus faible erreur moyenne pour un critère donné reproduit le mieux les statistiques correspondantes.

Erreurs moyennes des statistiques mensuelles					
Statistiques	AR(2)np	AR(2)p	AR(1)Vp	AR(1)Vnp	SPIGOT
Moyenne	.019	.039	.063	.018**	.031
écart type	.014**	.145	.171	.015	.029
Corrélation ordre 1	.169	.188	.149	.047**	.076
Corrélation ordre 2	.177	.178	.175	.136	.075**

Tableau 22: La comparaison des erreurs moyennes des statistiques mensuelles

Erreurs moyennes globales des statistiques annuelles				
Statistiques	AR(2)np	AR(0)	AR(1)Vnp	SPIGOT
Annuelles	.26	.12**	.25	.12**

Tableau 23: La comparaison des erreurs moyennes globales des statistiques annuelles

Erreurs moyennes globales des statistiques mensuelles					
Statistiques	AR(2)np	AR(2)p	AR(1)Vp	AR(1)Vnp	SPIGOT
Moyenne et écart type	.016**	.092	.117	.016**	.03
Corrélation ordre 1 et ordre 2	.17	.18	.16	.09	.076**

Tableau 24: La comparaison des erreurs moyennes globales des statistiques mensuelles

En examinant les tableaux 22 à 24 on peut conclure qu'aucune des méthodes ne se dégage comme étant la championne, car il n'y en a aucune qui est la meilleure pour tous les critères. Toutefois, on note que SPIGOT est le meilleur en moyenne, dans le cas des corrélations mensuelles (surtout les corrélations d'ordre 1). Dans le cas des statistiques annuelles, SPIGOT et AR(0) sont comparables et les meilleurs, ce qui est normal car les données annuelles sont générées directement par un AR(0) dans les deux cas (les données annuelles ne sont pas les sommes de données mensuelles générées comme dans le cas de AR(1)Vnp et AR(2)np). Ce sont les méthodes directes sans pondération qui sont meilleures pour reproduire la plupart des statistiques mensuelles (AR(1)Vnp). On en conclut que les méthodes directes sans pondération reproduisent le mieux les statistiques mensuelles et que les méthodes indirectes (SPIGOT) reproduisent le mieux les statistiques annuelles. Ce résultat n'est pas une surprise. De plus, on peut conclure qu'une approche ARMA à paramètres variables n'est pas absolument nécessaire pour modéliser les données d'apports nets au lac Supérieur, un modèle à paramètres constants serait satisfaisant car les différences des scores des diverses méthodes ne sont pas très grandes. Le modèle à paramètres constants a l'avantage d'être parcimonieux (moins de paramètres à estimer).

8. CONCLUSION

Toutes les étapes d'une méthodologie de génération de données synthétiques par une approche directe (modèles ARMA) ont été présentées en détail. Cinq modèles mensuels ont été étudiés pour fin de comparaison des approches directes et indirectes: des modèles ARMA à paramètres constants avec et sans pondération annuelle, des modèles ARMA à paramètres variables avec et sans pondération annuelle, et un modèle de désagrégation.

Il est difficile de trancher d'une façon générale laquelle des approches, directe ou indirecte est meilleure pour générer des données synthétiques. La réponse dépend du problème, elle dépend des données, des caractéristiques statistiques que l'on veut préserver, et de l'application en simulation des données générées. Dans ce travail, l'approche directe sans pondération s'est avérée globalement meilleure pour préserver les caractéristiques mensuelles des apports nets mensuels au lac Supérieur et l'approche indirecte s'est avérée meilleure pour préserver les caractéristiques annuelles. Nos conclusions sont basées strictement sur des caractéristiques statistiques, qui ne sont pas nécessairement toujours les critères les plus importants. De plus, la méthode de comparaison des méthodes utilisée (erreur moyenne) est très subjective, et n'accorde pas de poids plus important à un critère par rapport à un autre. Qui plus est, les différences entre les scores de la méthode indirecte (SPIGOT) et des méthodes directes sont assez faibles et leur importance est sujette à interprétation.

9. BIBLIOGRAPHIE

- Barnes, F.B. (1954), Storage required for a city water supply, Journal of the Institution of Engineers, Australia, Vol. 26, pp. 198-
- Boes D.C. and J.D. Salas (1978), Nonstationarity in the mean and the Hurst phenomenon, Water Resources Research, Vol. 14, No. 1, pp. 135-143.
- Box, G.E.P. and G.M. Jenkins (1976), Time Series Analysis - Forecasting and Control, Holden-Day, 575 pages.
- Carlson R.F., MacCormick A.J.A., and D.G. Watts (1970), Application of linear models to four annual streamflow series, Water Resources Research, Vol. 6, No. 4, pp. 1070-1078.
- C.E.A. (Canadian Electrical Association) (1990), Evaluation of monthly flow generation models DRAFT, Acres International Limited, Niagara Falls, Ontario, 134 pages.
- Grygier J.C. and J.R. Stedinger (1988), Condensed disaggregation procedures and conservation corrections for stochastic hydrology, Water Resources Research, Vol. 24, No. 10, pp. 1574-1584.
- Grygier J.C. and J.R. Stedinger (1990), SPIGOT, a Synthetic Streamflow Generation Software Package, Technical Description, Version 2.6, Cornell University, Ithaca, New York 14853-3501.
- Grygier J.C. and J.R. Stedinger (1991), SPIGOT, a Synthetic Streamflow Generation Software Package, User's Manual, Version 2.6, Cornell University, Ithaca, New York 14853-3501.
- Harms, A.G. and Campbell, T.H. (1967), An extension to the Thomas-Fiering model for the sequential generation of streamflows, Water Resources Research, Vol. 3, pp. 653-661.
- Hazen, A. (1914), Storage to be provided in impounding reservoirs for municipal water supply, Transactions ASCE, Vol. 77, pp. 1539-1669
- Hipel K.W., McBean E.A. and A.I. McLeod (1979), Hydrologic generating model selection, Journal of the Water Resources Planning and Management Division, Vol. 105, No. WR2, pp. 223-242.
- Hogg R.V. and E.A. Tanis (1977), Probability and Statistical Inference, Macmillan Publishing Co., 533 pages.
- Hurst, H.E., Black R.P. and Y.M. Simaika (1965), Long Term Storage, Constable, London.
- Jackson, B.B. (1975), Markov mixture model for drought lengths, Water Resources Research, Vol. 11, No. 1, pp. 64-74.
- Kavvas M.L. and J.W. Delleur (1975), Removal of periodicities by differencing and monthly mean subtraction, Journal of Hydrology, Vol. 26, pp. 335-353.
- Klemes V. (1972), Comments on "Adequacy of Markovian models with cyclic components for stochastic streamflow simulation", by I.Rodriguez-Iturbe, D.R. Dawdy and L.E. Garcia, Water Resources Research, Vol. 8, No. 6, pp. 1613-1615.

Lane W.L. (1979), Applied stochastic techniques (LAST computer package), user manual. Division of Planning Technical Services, Bureau of Reclamation, Denver, Colorado.

Lettenmaier D.P. and S.J. Burges (1977), Operational assessment of hydrologic models of long-term persistence, Water Resources Research, Vol. 13, No. 1, pp. 113-124.

Lettenmaier D.P. and S.J. Burges (1982), Validation of synthetic streamflow models, Times Series Methods in Hydroscience, eds. A.H. El-Shaarawi and S.R. Esterby, Elsevier Scientific Publishing Company, pp. 424-444.

Loucks D.P., Stedinger J., and D.A. Haith (1981), Water Resource Systems Planning and Analysis, Prentice-Hall, Inc., 559 pages.

Mandelbrot B.B. and J.R. Wallis, Noah, Joseph and operational hydrology, Water Resources Research, Vol. 4, No. 5, pp. 909-918.

Matalas N.C. (1967), Mathematical assessment of synthetic hydrology, Water Resources Research, Vol. 3, No. 4, pp. 937-945.

Mathier L., Fagherazzi L., Rassam J.C. and B. Bobée (1992), Great Lakes Net Basin Simulation by a Stochastic Approach, Rapport Scientifique no 362, INRS-Eau, Université du Québec, 95 pages.

McLeod A.I. and K.W. Hipel (1978), Simulation procedures for Box-Jenkins models, Water Resources Research, Vol. 14, No. 5, pp. 969-975.

McMahon T.A., Codner G.P., and C. Philips (1972), Single and Multi-site operational hydrology, Nordic Hydrology, Vol. 3, pp. 214-238.

McMahon T.A. and R.G. Mein (1986), River and Reservoir Yield, Water Resources Publications, Littleton, Colorado, 368 pages.

Mejia J.M., Rodriguez-Iturbe I., and D.R. Dawdy (1972), Streamflow simulation 2, the Broken Line process as a potential model for hydrologic simulation, Water Resources Research, Vol. 8, No. 4, pp. 931-941.

Mejia J.M. and J. Rousselle (1976), Disaggregation models in hydrology revisited, Water Resources Research, Vol. 12, No. 2, pp. 185-186.

Panu U.S. (1978), Stochastic Synthesis of Monthly Streamflows Based on Pattern Recognition, Ph.d. dissertation, University of Waterloo, Waterloo, Ontario.

Pereira M.V.F., Oliveira G.C., Costa C.C.G. and J. Kelman (1984), Stochastic streamflow models for hydroelectric systems, Water Resources Research, Vol. 20, no. 3, pp. 379-390.

Perreault L. and B. Bobée, (In preparation 1994), Tests d'adéquation pour le logiciel AJUSTE-II, Rapport Scientifique, INRS-Eau.

Rassam J.-C., Mathier L., Fagherazzi L.D., Roy R., Bobée B., and L. Carballada (1992), Beauharnois - Les Cedres Spillway, Design Flood With A Stochastic Approach, Final Report To The Experts Committee, Hydro-Québec, Groupe Production, Transport Et Télécommunications, Vice Présidence Réseaux - Direction Production Et Échange D'énergie.

Salas J.D., and D.C. Boes (1980), Shifting level modelling of hydrologic series, Advances in Water Resources, Vol. 3, pp. 59-63.

Salas J.D., Delleur J.W., Yevjevich V. and W.L. Lane (1980), Applied Modeling of Hydrologic Time Series, Water Resources Publications, 481 pages.

Shapiro S.S., Wilk M.B., and H.J. Chen (1968), A comparative study of various tests of normality, American Statistical Association Journal, December, pp. 1343-1372.

Stedinger J.R. and M.R. Taylor (1982), Synthetic streamflow generation, 1. Model verification and validation, Water Resources Research, Vol. 18, no. 4, pp. 909-918.

Srikanthan R. and T.A. McMahon, Stochastic generation of monthly streamflows, (1982) Journal of the Water Resources Planning and Management Division, Vol. 108, No. HY3, pp. 419-441.

Sudler C.H. (1927), Storage required for regulation of streamflow, Transactions ASCE, Vol. 61, pp. 622-660.

Valencia D. and J.C. Schaake Jr. (1973), Disaggregation processes in stochastic hydrology, Water Resources Research, Vol. 9, No. 3, pp. 580-585.

Vecchia A.V., Obeysekera J.T., Salas J.D. and D.C. Boes (1983), Aggregation and estimation for low-order periodic ARMA models, Water Resources Research, Vol. 19, no. 5, pp. 1297-1306.

Weiss G. (1977), Shot Noise models for the generation of synthetic streamflow data, Water Resources Research, Vol. 13, no. 1, pp. 101-108.

Yakowitz S.J. (1973), A stochastic model for daily riverflow in an arid region, Water Resources Research, Vol. 9, No. 5, pp. 1271-1285.

10. ANNEXE 1

Les statistiques périodiques

La moyenne périodique de l'échantillon

$$\bar{x}_\tau = \frac{1}{N} \sum_{v=1}^N x_{v,\tau}$$

$\tau = 1, \dots, \omega$, N est le nombre d'années dans l'échantillon et ω est le nombre d'intervalles (saisons de l'année), c'est l'estimateur de la moyenne de la population, μ_τ .

La variance périodique de l'échantillon

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

s_τ^2 est l'estimateur de la variance de la population σ_τ^2 .

Le coefficient de corrélation de l'échantillon

$$r_{k,\tau} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)(x_{v,\tau-k} - \bar{x}_{\tau-k})}{s_\tau s_{\tau-k}}$$

c'est l'estimateur du coefficient de corrélation de la population $\rho_{k,\tau}$. Lorsque $\tau-k < 0$, on remplace $\tau-k$ par $\tau-k+12$.

Les statistiques annuelles

La moyenne de l'échantillon

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$$

N est le nombre d'années.

L'écart type de l'échantillon

$$s = \sqrt{\frac{1}{N-1} \sum_{t=1}^N (x_t - \bar{x})^2}$$

Le coefficient d'autocorrélation ρ_k d'ordre k d'une série stationnaire x_t

L'estimateur de $\rho_k = \text{Corr}(x_t, x_{t+k})$ est:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad \gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)] \quad \mu = E(x_t) = E(x_{t+k}), \quad k = 0, \pm 1, \pm 2, \dots$$

Cet estimateur est utilisé pour calculer les autocorrélations annuelles d'ordre 1 à 14 aux figures 18 à 22.

Un autre estimateur de ρ_k est celui du maximum de vraisemblance:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}_t)(x_{t+k} - \bar{x}_{t+k})}{\left[\sum_{t=1}^{N-k} (x_t - \bar{x}_t)^2 \sum_{t=1}^{N-k} (x_{t+k} - \bar{x}_{t+k})^2 \right]^{1/2}}$$

où \bar{x}_t est la moyenne des premières $N-k$ valeurs x_1, \dots, x_{N-k} ,

et \bar{x}_{t+k} est la moyenne des $N-k$ dernières valeurs x_{k+1}, \dots, x_N ;

le graphique de r_k en fonction de k est le corrélogramme de la série x_t .

Le coefficient d'autocorrélation partielle ϕ_{kk} d'ordre k d'un série stationnaire x_t (Box et Jenkins, 1976)

Si on note par ϕ_{kj} le $j^{\text{ième}}$ coefficient dans un processus autorégressive d'ordre k , alors ϕ_{kk} est le dernier coefficient. On obtient $\hat{\phi}_{kk}$ l'estimateur de ϕ_{kk} en résolvant les équations $r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{kk}r_{j-k}, j = 1, 2, \dots, k$.

Le coefficient d'asymétrie

$$g = \frac{N \sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)(N-2)s^3}$$

est un estimateur non biaisé.

Le coefficient de Hurst, h (Loucks *et al.*, 1981)

Ce coefficient mesure la persistance d'une série annuelle $\{q_i\}$ de N années. On calcule en premier, les écart cumulatifs de la moyenne S_n , où:

$$\bar{q}_n = \frac{1}{n} \sum_{y=1}^n q_y \quad S_y^n = \sum_{y^*=1}^y (q_{y^*} - \bar{q}_n), y=1, \dots, n, S_n^n = 0$$

$$R_n = \max_y (S_y^n) - \min_y (S_y^n) = S_M^n - S_m^n$$

$$h = \frac{\log_{10} \left(\frac{R}{s} \right)}{\log_{10} (n)}$$

s est l'écart type de la série $\{q_i\}$.

La courbe de période de retour

Pour obtenir la courbe de période de retour d'une série, les données annuelles sont ordonnées en ordre décroissant, et tracées en fonction de $(N+1)/T$ où T correspond à leur position dans la série, et N à la longueur de la série (formule Weibull).

La longueur d'une suite de surplus (*surplus run length* (RL) et l'apport net maximal en surplus d'une suite (*surplus run*) (RS) (Salas and Boes, 1980)

Un surplus est une valeur positive de $x_i - c$ où c est un seuil fixe. Une suite consécutive de L surplus est une suite de surplus d'une durée L, et la somme des surplus $x_i - c$ pour cette durée est le volume de surplus d'une séquence. RL est la plus longue séquence de surplus contenue dans l'échantillon $x_1 \dots x_n$ et RS est le maximum des apports nets en surplus. Le seuil c peut être la moyenne de l'échantillon, ou une moyenne mobile sur n années.