

Rapport scientifique no 139

par

J.C. Auclair, M. Bisson et M. Lachance

Etude des variations de l'alcalinité  
des lacs de tête de la région de Québec

INRS-Eau  
Université du Québec  
C.P. 7500  
Sainte-Foy (Québec)  
G1V 4C7

Mars 1982

Etude des variations de l'alcalinité  
des lacs de tête de la région de Québec

Rapport scientifique no 139

par

J.C. Auclair, M. Bisson et M. Lachance

INRS-Eau  
Université du Québec  
C.P. 7500  
Sainte-Foy (Québec)  
G1V 4C7

Mars 1982



## TABLE DES MATIERES

	Page
TABLE DES MATIERES .....	i
LISTE DES FIGURES .....	iii
LISTE DES TABLEAUX .....	iv
RESUME .....	2
1. Position du problème et études antérieures .....	3
1.1 Introduction .....	3
1.2 Méthodes utilisées dans la littérature pour estimer des données manquantes .....	4
1.3 Méthode proposée .....	5
2. Développements théoriques .....	7
2.1 Hypothèses et notations .....	7
2.2 Estimation de $\rho$ à partir d'une valeur échantil- lonnée $r$ .....	9
2.3 Utilisation de l'estimateur "moyenne générale" pour simuler les valeurs absentes d'un échantillon long .....	10
2.4 Utilisation de l'estimateur "moyenne locale" pour simuler les valeurs absentes d'un échantillon long .....	12

2.5	Estimation de $r$ pour un échantillon long comportant des données manquantes occasionnelles .....	16
2.6	Etude de la suppression des valeurs absentes occasionnelles .....	19
2.7	Comparaison des biais théoriques .....	23
2.8	Cas de l'échantillon court .....	24
2.9	Calcul des biais théoriques dans le cas d'un échantillon court comprenant des valeurs absentes .....	25
3.	Etude des biais résiduels et des variances d'estimation par simulation .....	50
3.1	Technique de Monte-Carlo .....	50
3.2	Présentation des résultats .....	51
4.	Biais introduits par la transformation $Z$ de Fisher ..	70
4.1	Rappel sur la transformation de Fisher .....	70
4.2	Application aux autocorrélations .....	70
4.3	Simulations .....	73
5.	Utilisation des autocorrélations débiaisées comme un test d'indépendance .....	77
	REFERENCES BIBLIOGRAPHIQUES .....	83

## LISTE DES FIGURES

	Page
Figure 1 : Abaques reliant $r_0$ , $r_i$ et $r_c$ .....	27
Figure 2 : Comparaison des biais pour différents estimateurs des valeurs absentes .....	28
Figure 3 : Evolution des biais des estimateurs selon les cor- rections effectuées ( $N = 60$ , $n = 9$ ) .....	55
Figure 4 : Pourcentage des échantillons classés comme indépen- dants ( $\alpha = 0.05$ ) selon les estimateurs $r$ ( $N = 100$ , $n = 20$ ) .....	80
Figure 5 : Pourcentage des échantillons classés comme indé- pendants ( $\alpha = 0.05$ ) selon les estimateurs $r$ ( $N = 20$ , $n = 4$ ) .....	81

## LISTE DES TABLEAUX

	Page
Tableau 1 : Biais moyen $\rho - E(r)$ d'un échantillon de longueur $N$ provenant d'une série Markovienne d'ordre 1 ..	29
Tableau 2 : Variance d'estimation d'une série markovienne d'ordre 1 de paramètre $\rho$ et de longueur $N$ .....	30
Tableau 3 : Biais $r - E(r_0)$ liés à l'utilisation de l'estimateur "moyenne générale" pour un échantillon long comportant $\alpha\%$ valeurs absentes .....	31
Tableau 4 : Biais $r - E(r_i)$ liés à l'utilisation de l'estimateur "moyenne locale" pour un échantillon long comportant $\alpha\%$ valeurs absentes .....	32
Tableau 5 : Valeurs de $r$ en fonction de $r_0$ et $r_1$ pour un échantillon long avec des données manquantes occasionnelles .....	33
Tableau 6 : Biais $r - E(r_D)$ liés à l'utilisation de l'estimateur $r_D$ pour un échantillon long comportant $\alpha\%$ de valeurs absentes .....	34
Tableau 7 : Biais combiné $\rho - E(r_0)$ estimateur "moyenne générale" .....	35
Tableau 8 : Biais combiné $\rho - E(r_0)$ estimateur "moyenne générale" .....	36
Tableau 9 : Biais combiné $\rho - E(r_0)$ estimateur "moyenne générale" .....	37

Tableau 10: Biais combiné $\rho - E(r_0)$ estimateur "moyenne générale" .....	38
Tableau 11: Biais combiné $\rho - E(r_0)$ estimateur "moyenne générale" .....	39
Tableau 12: Biais combiné $\rho - E(r_i)$ estimateur "moyenne locale" .....	40
Tableau 13: Biais combiné $\rho - E(r_i)$ estimateur "moyenne locale" .....	41
Tableau 14: Biais combiné $\rho - E(r_i)$ estimateur "moyenne locale" .....	42
Tableau 15: Biais combiné $\rho - E(r_i)$ estimateur "moyenne locale" .....	43
Tableau 16: Biais combiné $\rho - E(r_i)$ estimateur "moyenne locale" .....	44
Tableau 17: Biais combiné $\rho - E(r_D)$ valeurs absentes omises	45
Tableau 18: Biais combiné $\rho - E(r_D)$ valeurs absentes omises	46
Tableau 19: Biais combiné $\rho - E(r_D)$ valeurs absentes omises	47
Tableau 20: Biais combiné $\rho - E(r_D)$ valeurs absentes omises	48
Tableau 21: Biais combiné $\rho - E(r_D)$ valeurs absentes omises	49



Tableau 22: Biais résiduels moyens $\rho - E(r)$ calculés sur 2000 échantillons complets de longueur $N$ et ayant subi la correction de longueur de KENDALL	56
Tableau 23: Augmentation de la variance d'estimation induite par la correction de longueur pour 2,000 échantil- lons complets .....	57
Tableau 24: Biais globaux induits par les estimateurs .....	58
Tableau 25: Biais résiduels induits par différents estimateurs non corrigés pour l'effet de longueur .....	59
Tableau 26: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur .....	60
Tableau 27: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur puis débiaisés ( $N = 100$ ) .....	61
Tableau 28: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur puis débiaisés ( $N = 100$ ) .....	62
Tableau 29: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur puis débiaisés ( $N = 100$ ) .....	63
Tableau 30: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur puis débiaisés ( $N = 100$ ) .....	64
Tableau 31: Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur puis débiaisés ( $N = 100$ ) .....	65

Tableau 32: Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur puis débiaisés (N = 100) .....	66
Tableau 33: Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur puis débiaisés (N = 80) .....	67
Tableau 34: Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur puis débiaisés (N = 60) .....	68
Tableau 35: Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur puis débiaisés (N = 40) .....	69
Tableau 36: Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur puis débiaisés (N = 20) .....	70
Tableau 37: Biais résiduels théoriques de la transformation Z de Fisher (N = 100) .....	76
Tableau 38: Biais résiduels moyens et variances d'estimation des autocorrélations - transformation de Fisher N = 100 .....	77
Tableau 39: Puissance comparée des estimateurs débiaisés pour tester l'indépendance des séries ( $\alpha = 0.05$ ) dans les cas les plus critiques (20% de valeurs absentes) .....	82
Tableau 40: Pourcentage des séries de Markov de longueur 100 classées indépendantes ( $\alpha = 0.05$ ) selon le nombre n de valeurs absentes et l'estimateur débiaisé utilisé .....	83



## ABSTRACT

### PERSISTENCE ESTIMATION FROM A TIME-SERIES CONTAINING OCCASIONAL MISSING DATA

In a real life situation, such as that encountered with weekly sampled water-quality parameters, measured time-series are often incomplete. The missing data complicate the estimation of the lag-one autocorrelation coefficient, a most important parameter in the evaluation of serial persistence necessary for efficient sampling strategies. In the situation of occasional missing data, we have investigated the biases and sampling variances resulting from the introduction of simple but non-optimal estimators (e.g. the "general mean" estimator, the "local mean" estimator and the "combined mean" estimator) or from compressing the time-series by simply neglecting the missing data. These studies were carried out in two ways:

- 1) via theoretical developments, in the case of large samples and
- 2) via simulation of Markovian series in the case of smaller samples where a classical short sample correction was combined with the theoretical biases derived from the large sample case.

The efficiencies (residual biases and sampling variances), which depend on the sample lengths and on the persistence parameter of the parent population of the estimators, have been evaluated when 5%, 10%, 15% and 20% of the data were missing. The proposed approach (combined mean estimator) leads, in many cases, to better results than would be obtained using classical techniques derived from Fisher's Z transformation.

Key-words: autocorrelation, missing data, estimator, bias, Fisher.

## RESUME

Dans la vie réelle, les séries mesurées comportent des données manquantes, ce qui rend difficile l'estimation du coefficient d'auto-corrélation d'ordre 1, un des paramètres les plus importants pour évaluer la persistance de la population-mère. Dans le cas de valeurs absentes occasionnelles, les biais et variances d'estimation résultant de l'utilisation d'estimateurs simples mais non optimaux des valeurs absentes ont été étudiés:

- 1) théoriquement, dans le cas d'échantillons longs,
- 2) par simulation de séries markoviennes dans le cas d'échantillons courts où une correction classique de longueur a été combinée aux biais théoriques calculés dans le cas d'échantillons longs.

Les performances (biais résiduels et variances d'estimation) selon la longueur de l'échantillon et la persistance de la population-mère des estimateurs: "moyenne générale", "moyenne locale", "moyenne combinée" et "valeurs absentes omises" ont été évaluées et donnent dans de nombreux cas de meilleurs résultats que la technique classique faisant appel à la transformation Z de Fisher.

Mots-clés: Autocorrélation, valeur absente, estimateur, biais, Fisher.

## 1. POSITION DU PROBLEME ET ETUDES ANTERIEURES

### 1.1 Introduction

L'estimation du coefficient d'autocorrélation d'ordre 1 d'une série de temps stationnaire constitue une étape importante dans l'évaluation de la persistance à court terme de cette série; cette connaissance est utile, à la fois pour la compréhension du phénomène représenté par la série, mais aussi pour déterminer le pas de temps optimal à l'acquisition de données ultérieures; il arrive que, pour évaluer ce coefficient d'autocorrélation, on dispose d'un échantillon de données prélevées de façon systématique à pas de temps égal, mais comprenant, soit des données occasionnellement absentes, soit des valeurs aberrantes qui pourraient fausser toute interprétation. Dans ce cas, deux techniques ont été décrites pour évaluer l'autocorrélation: la première consiste à calculer, sur chaque tronçon continu de données, un coefficient de corrélation qui peut être entâché d'un biais considérable, étant donné la longueur de chaque tronçon, à effectuer la transformation de FISHER qui normalise la distribution de ces coefficients, à combiner les valeurs transformées en fonction des longueurs partielles des tronçons et à effectuer la transformation inverse qui procure une évaluation de l'autocorrélation de l'ensemble. Une autre alternative consiste à évaluer les valeurs manquantes au moyen d'un modèle plus ou moins complexe, établi à partir de l'ensemble des données valides en donnant aux valeurs manquantes des valeurs qui conservent les caractéristiques essentielles de l'échantillon: moyenne, variance et persistance. Quand on dispose de suffisamment de données, on essaie alors de modéliser la série, par les méthodes d'interpolation directe et inverse de BOX et JENKINS pour obtenir un estimateur linéaire non biaisé (BLUE) de chaque valeur absente, puis calculer la corrélation sur la série ainsi complétée. Cette technique exige un effort relativement grand. Une telle procédure peut être itérative et même éventuellement diverger, si le nombre de valeurs manquantes est assez élevé et si la distance entre elles est trop petite pour éliminer les interactions.

## 1.2 Méthodes utilisées dans la littérature pour estimer des données manquantes

Il existe, dans la littérature récente, de nombreux textes traitant de l'estimation de données manquantes de séries uni ou multivariées; il s'agit essentiellement d'un problème d'interpolation de données en utilisant la tendance générale des données. WILKINSON (1958) et PREECE (1971) ont appliqué cette technique à des données expérimentales incomplètes. D'autres techniques graphiques similaires ont été proposées, mais comme elles n'utilisent pas la structure des autocorrélations, leur application à l'étude des séries de temps est très limitée.

Plus récemment, BRUBACHER et WILSON (1976) ont développé une technique basée sur le principe de moindres carrés et une approche prévisionnelle qui permet d'estimer l'effet d'un jour férié sur la consommation horaire d'électricité en utilisant les consommations normales antérieures et postérieures. Les valeurs interpolées sont obtenues par une méthode de projection directe et inverse ("forecasting", "backforecasting") en minimisant des séries résiduelles; on obtient alors un système de  $n$  équations à résoudre pour  $n$  valeurs interpolées. Le rapport de la consommation actuelle à la consommation historique durant les jours fériés passés permet d'évaluer la demande future pour des jours fériés similaires. La technique utilisée paraît bien adaptée au problème traité, quoique les calculs puissent devenir compliqués dès que plusieurs valeurs absentes sont à reconstituer simultanément. La raison du succès de la méthode et aussi la raison pour laquelle celle-ci ne peut se généraliser à d'autres séries de temps, tient à la nature même de la consommation d'électricité. Un modèle du type BOX et JENKINS (1970) peut être identifié pour l'ensemble de la série à partir d'une section courte de données: les fluctuations hebdomadaires ont une bien plus grande variabilité que les fluctuations annuelles; quelques semaines de données permettent donc une identification convenable

du modèle global qui n'est pas influencé significativement par les jours fériés.

D'ASTOUS et HIPEL (1979) ont présenté une technique basée sur les modèles de BOX et JENKINS et un modèle d'intervention qui permet de reconstituer, de façon plus générale, les données absentes. Les simulations effectuées, par ces auteurs, sur des séries homogènes et hétérogènes de débits annuels, montrent l'efficacité et la généralité de la méthode.

### 1.3 Méthode proposée

Dans ce rapport, nous nous proposons d'utiliser deux estimateurs non optimaux, mais très simples de données manquantes qui induisent des biais faciles à évaluer dans le calcul de l'autocorrélation de l'échantillon complété et ainsi de les corriger.

Le premier estimateur, constitué par la moyenne générale de l'échantillon, conduit à sous-estimer les valeurs calculées de  $r_1$ . Le deuxième estimateur, constitué par la moyenne locale, c'est-à-dire la moyenne arithmétique entre la valeur précédant la valeur manquante et celle la suivant, conduit à surestimer les valeurs calculées de  $r_1$ .

Les biais et les variances des deux estimateurs seront évalués et comparés théoriquement.

On note que les estimateurs utilisés ne font aucune hypothèse quant au processus générateur de chaque échantillon, ce qui est cohérent avec le fait qu'une même courte trace peut provenir de populations mères générées par des processus différents.

Des séries markoviennes d'ordre 1 générées selon l'algorithme de BOX et MULLER (1958) seront simulées pour tester, par la méthode de Monte-Carlo l'efficacité de la correction des biais selon les valeurs



de  $\rho$  de la population mère en combinaison avec la correction pour échantillon court déjà étudié par WALLIS et O'CONNELL (1972).

Un développement théorique simple permettant de calculer  $\rho$  sans corriger les biais sera présenté dans le cas d'un échantillon long.

Dans la même veine, le biais introduit dans  $r$  en raccourcissant la série des valeurs manquantes, technique simple utilisée quelquefois dans la pratique, sera évalué dans le cas où la série suit un processus Markovien d'ordre 1.

A titre de comparaison, les estimations de  $\rho_1$  évaluées à partir de la transformation  $Z$  de FISHER, corrigées pour la longueur des séries seront effectuées dans quelques cas pour vérifier l'efficacité des méthodes proposées.

## 2. DEVELOPPEMENTS THEORIQUES

### 2.1 Hypothèses et notations

On considère un échantillon  $X_i$  ( $i=1, \dots, N$ ) tiré d'une série stationnaire, de longueur  $N$  (connue et grande) dans laquelle  $n$  valeurs sont absentes. Les  $n$  valeurs absentes ne se suivent pas et sont peu nombreuses par rapport à  $N$  ( $n \ll N$ ) l'échantillon se présente donc sous la forme de  $(n + 1)$  tronçons  $Z_j$  ( $j=1 \dots n + 1$ ) séparés par  $n$  valeurs isolées absentes:

$$Z_1 ? Z_2 ? \dots Z_j \dots ? Z_n ? Z_{n+1}$$

On sait que le calcul des autocorrélations sur des échantillons courts sont sujets à de larges biais et il peut être intéressant d'estimer les valeurs manquantes et de calculer  $r_1$  sur une série longue plutôt que d'évaluer les coefficients largement biaisés sur chacun des tronçons et de les recombinaison en utilisant, par exemple, la transformation  $Z$  de FISHER.

Il y a eu plusieurs discussions par PARZEN (1964) et RODRIGUEZ-ITURBE (1971) concernant la forme appropriée à donner à la variance et à la covariance d'une série à partir des valeurs échantillonnées. Dans cette étude, nous avons choisi d'utiliser les expressions non biaisées qui peuvent s'appliquer indifféremment aux échantillons plus ou moins longs.

$$\text{var} = \frac{1}{N-1} \sum_{i=1}^N (X_i - m)^2$$

$$\text{cov} = \frac{1}{N-1} \sum_{i=1}^{N-1} (X_i - m) (X_{i+1} - m)$$

où 
$$m = \frac{1}{N} \sum_{i=1}^N X_i$$

l'expression correspondante de l'autocorrélation est celle de JENKINS et WATTS (1968) et de BOX et JENKINS (1970)

$$r = \frac{\text{cov}}{\text{var}} = \frac{\sum_{i=1}^{N-1} (X_i - m) (X_{i+1} - m)}{\sum_{i=1}^N (X_i - m)^2}$$

Si on considère maintenant l'échantillon recentré, ces expressions se réduisent, avec  $m = 0$ , à :

$$\text{var} = \frac{1}{N-1} \sum_{i=1}^N X_i^2$$

$$\text{cov} = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i X_{i+1}$$

et 
$$r = \frac{\text{cov}}{\text{var}} = \frac{\sum_{i=1}^{N-1} X_i X_{i+1}}{\sum_{i=1}^N X_i^2} \quad [1]$$

On évalue maintenant les espérances mathématiques des expressions utilisées:

$$E (X_i X_{i+1}) = \text{cov} \quad [2]$$

$$E (X_i^2) = \frac{N-1}{N} \text{ var} \quad [3]$$

## 2.2 Estimation de $\rho$ à partir d'une valeur échantillonnée $r$

Plusieurs chercheurs comme SOPER et al. (1918), ORCUTT (1948) et SASTRY (1951) se sont intéressés aux relations entre les valeurs de l'autocorrélation évaluée sur quelques traces des échantillons de longueur limitée et à la valeur de l'autocorrélation de la population-mère sous-jacente. Dans une revue des travaux antérieurs, MARIOTT et POPE (1954) ont montré que le biais dans le calcul des autocorrélations d'une série à partir de celui d'un échantillon peut provenir de 2 sources: si la moyenne de la population-mère est connue, les valeurs estimées des autocorrélations sont biaisées, en général vers zéro, à cause de la longueur courte de l'échantillon. Souvent, en pratique, la moyenne de la population-mère n'est pas connue et doit être estimée à partir de l'échantillon, ce qui introduit un nouveau biais qui est toujours négatif pour une série longue.

Les deux sources de biais peuvent se renforcer ou se compenser; elles ne sont pas indépendantes et ne peuvent donc s'étudier séparément. C'est pourquoi WALLIS et O'CONNELL (1972) ont étudié par simulation sur 20,000 échantillons, pour différentes populations-mères connues, les biais engendrés selon la longueur des échantillons et les valeurs  $\rho$  de la population-mère. Si on suppose celle-ci générée pour un processus Markovien d'ordre 1, ce qui est souvent le cas en hydrologie, ils ont montré que, dans ces circonstances, l'expression circulaire de l'autocorrélation  $r$  d'un échantillon permettait le calcul

d'un biais théorique très proche de celui qui affectait l'équation [1].

$$E(r) = \rho - \frac{1}{N} (1 + 4\rho) = \rho \frac{N-4}{N} - \frac{1}{N}$$

ce qui s'écrit aussi:

$$\rho = \frac{N}{N-4} \left[ E(r) + \frac{1}{N} \right] = \frac{E(r) + \frac{1}{N}}{1 - \frac{4}{N}} \quad [4]$$

Cette correction [4] due à KENDALL (1954) peut être très significative pour des échantillons courts ( $N \leq 100$ , par exemple):

$$\rho - E(r) = \frac{1}{N} (1 + 4\rho) = \frac{4E(r) + 1}{N - 4}$$

Toujours pour des séries markoviennes d'ordre 1, BOX et JENKINS (1970) ont démontré que la variance d'estimation de  $\rho$  pour des échantillons longs pouvait être significative et était donnée approximativement par:

$$\text{var}(r) = \frac{1 - \rho^2}{N} \quad [5]$$

Ces biais et ces variances d'estimation ont été calculés aux tableaux 1 et 2 pour des échantillons de longueur  $N = 100, 80, 60, 40$  et  $20$  et des autocorrélations de la population-mère  $\rho = 0, 0.1 \dots$  et  $0.9$ .

### 2.3 Utilisation de l'estimateur "moyenne générale" pour simuler les valeurs absentes d'un échantillon long

Notons  $r_0$  la valeur du coefficient d'autocorrélation de l'échantillon obtenue par l'équation [1] en remplaçant les  $n$  valeurs absentes  $X_0$  (abs) de l'échantillon par la moyenne générale soit zéro; on s'aperçoit que:

- au numérateur, chaque valeur absente fait disparaître deux produits.

- au dénominateur chaque valeur absente fait disparaître un carré.

Si on remplace les termes disparus par leurs espérances mathématiques (équations [2] et [3]) on obtient:

$$E(r_0) = \frac{N - 2n - 1}{N - n} \frac{\text{cov}}{\frac{N-1}{N} \text{var}} = \frac{1 - \frac{2n+1}{N}}{1 - \frac{n}{N}} \cdot \frac{N}{N-1} r$$

soit, en inversant:

$$r = \frac{(1 - \frac{1}{N})(1 - \frac{n}{N})}{1 - \frac{2n+1}{N}} E(r_0) \quad [6]$$

Cette expression peut se développer facilement jusqu'aux termes en  $N^{-2}$ .

$$r \approx (1 - \frac{1}{N})(1 - \frac{n}{N})(1 + \frac{2n+1}{N} + [\frac{2n+1}{N}]^2) E(r_0)$$

$$r \approx (1 - \frac{n+1}{N} + \frac{n}{N^2})(1 + \frac{2n+1}{N} + \frac{4n^2 + 4n + 1}{N^2}) E(r_0)$$

$$\approx [1 + \frac{n}{N} + \frac{2n(n+1)}{N^2}] E(r_0) \quad [7]$$

Ce qui peut s'exprimer sous forme de biais absolu:

$$BA = r - E(r_0) = \frac{n}{N} (1 + \frac{2(n+1)}{N}) E(r_0)$$

On note que la correction est du signe de  $E(r_0)$ , donc souvent positive, dans le cas des séries persistantes; dans ces cas, l'estimateur sous-estime la vraie valeur.

Cas particulier: Si l'on suppose de plus  $n \gg 1$  et si l'on pose

$\frac{n}{N} = \alpha$  , les expressions précédentes se simplifient:

$$r \approx [1 + \alpha + 2\alpha^2] E(r_0) \quad [8]$$

$$BA \approx \alpha [1 + 2\alpha] E(r_0)$$

Le tableau 3 présente les biais moyens liés à l'utilisation de l'estimateur "moyenne générale" pour un échantillon long comportant  $\alpha\%$  valeurs absentes.

#### 2.4 Utilisation de l'estimateur "moyenne locale" pour simuler les valeurs absentes d'un échantillon long

Notons  $r_i$  la valeur du coefficient d'autocorrélation obtenue par l'équation [1] en remplaçant les  $n$  valeurs absentes  $x_i$  (abs) par la moyenne arithmétique des éléments précédant et suivant:

$$x_i \text{ (abs)} = \frac{x_{i+1} + x_{i-1}}{2}$$

- chaque produit du numérateur contenant une valeur absente s'estime par:

$$x_i \text{ (abs)} x_{i+1} = \frac{x_{i+1}^2}{2} + \frac{x_{i-1} x_{i+1}}{2}$$

- chaque carré du dénominateur contenant une valeur absente s'estime par:

$$x_i \text{ (abs)}^2 = \frac{x_{i+1}^2}{4} + \frac{x_{i-1}^2}{4} + \frac{x_{i+1} x_{i-1}}{2}$$

Si on remplace les produits ainsi créés par leur espérance mathématique déduite des valeurs présentes dans l'échantillon (équations [2] et [3]) et si, autour des valeurs absentes, on fait l'approximation  $E[X_{i+1} X_{i-1}] \approx E[X_{i+1} X_i]$ , et si, autour des valeurs absentes, on fait l'approximation  $E[X_{i+1} X_{i-1}] \approx E[X_{i+1} - X_i]$ , on obtient:

$$E(r_i) = \frac{(N-2n-1) \text{cov} + 2n \left( \frac{N-1}{2N} \text{var} + \frac{\text{cov}}{2} \right)}{(N-n) \frac{N-1}{N} \text{var} + n \left( \frac{N-1}{2N} \text{var} + \frac{\text{cov}}{2} \right)}$$

ce qui s'écrit

$$E(r_i) = \frac{2(N-n-1) N \frac{\text{cov}}{\text{var}} + 2n (N-1)}{n N \frac{\text{cov}}{\text{var}} + (2N - n) (N-1)} = \frac{2N (N-n-1) r + 2n (N-1)}{n N r + (2N - n) (N - 1)} \quad [9]$$

soit, en inversant:

$$r = \frac{\frac{(2N-n)(N-1)}{2(N-n-1)N} E(r_i) - \frac{n(N-1)}{N(N-n-1)}}{1 - \frac{n}{2(N-n-1)} E(r_i)} \quad [10]$$

Cette expression peut se développer jusqu'aux termes en  $N^{-2}$ :

$$r = \frac{\frac{(N-\frac{n}{2})(N-1)}{N(N-n-1)} E(r_i) - \frac{n(N-1)}{N(N-n-1)}}{\frac{N(N-n-1) - \frac{n}{2} E(r_i)}{N(N-n-1)}} = \frac{(1 - \frac{n}{2N})(1 - \frac{1}{N}) E(r_i) - \frac{n}{N} (1 - \frac{1}{N})}{(1 - \frac{n+1}{N}) - \frac{n}{N} \frac{E(r_i)}{2}}$$

le numérateur s'écrit  $E(r_i) - \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) - \frac{1}{N} E(r_i) + \frac{n}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)$



le dénominateur s'écrit  $1 - \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) - \frac{1}{N}$ , ce qui se développe en utilisant l'identité:

$$\frac{1}{1 - \epsilon} \approx 1 + \epsilon + \frac{\epsilon^2}{2} + \dots \quad \text{soit}$$

$$1 + \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) + \frac{1}{N} + \frac{n^2}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)^2 + \frac{1}{N^2} + \frac{2n}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)$$

à l'ordre  $N^{-2}$ .

Si on exécute le produit des deux expressions, il vient:

$$\begin{aligned} r = E(r_i) + \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) (E(r_i) - 1) + \frac{n^2}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)^2 (E(r_i) - 1) \\ + \frac{n}{N^2} \left( \frac{E(r_i)}{2} + 1 \right) (E(r_i) - 1) \end{aligned}$$

ce qui peut s'écrire:

$$\begin{aligned} r = E(r_i) + \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) (E(r_i) - 1) + \\ + \frac{n^2}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)^2 (E(r_i) - 1) \left[ 1 + \frac{1}{\frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right)} \right] \quad [11] \end{aligned}$$

Cette expression s'exprime sous forme de biais absolu:

$$BA = r - E(r_i) = \frac{n}{N} \left( \frac{E(r_i)}{2} + 1 \right) (E(r_i) - 1) + \frac{n^2}{N^2} \left( \frac{E(r_i)}{2} + 1 \right)^2 (E(r_i) - 1) \left[ 1 + \frac{1}{\frac{E(r_i)}{n} \left( \frac{E(r_i)}{2} + 1 \right)} \right]$$

On note que ces corrections sont toujours négatives: l'estimateur surestime toujours la vraie valeur.

Cas particulier: Si on suppose de plus que  $n \gg 1$ , et si l'on pose  $\frac{n}{N} = \alpha$ , les expressions précédentes se simplifient:

$$r \approx E(r_i) - \alpha \left( \frac{E(r_i)}{2} + 1 \right) (1 - E(r_i)) - \alpha^2 \left( \frac{E(r_i)}{2} + 1 \right)^2 (1 - E(r_i))$$

ou

$$r \approx E(r_i) - \alpha \left( 1 - \frac{E(r_i)}{2} - \frac{E(r_i)^2}{2} \right) - \alpha^2 \left( 1 - \frac{3 E(r_i)^2}{4} - \frac{E(r_i)^3}{4} \right)$$

$$BA \approx r - E(r_i) = -\alpha \left( 1 - \frac{E(r_i)}{2} - \frac{E(r_i)^2}{2} \right)$$

$$- \alpha^2 \left( 1 - \frac{3 E(r_i)^2}{4} - \frac{E(r_i)^3}{4} \right) \quad [12]$$

Le tableau 4 présente les biais liés à l'utilisation de l'estimateur à moyenne locale pour un échantillon long comportant  $\alpha\%$  valeurs absentes.

## 2.5 Estimation de $r$ pour un échantillon long comportant des données manquantes occasionnelles: estimateur combiné

Grâce aux estimateurs précédents qui ne font aucune hypothèse sur le processus générateur de la série, on dispose de 2 évaluations de  $r$  en fonction de  $E(r_0)$  et  $E(r_i)$  et de  $\alpha = \frac{n}{N}$ ; quand on ne dispose que d'un seul échantillon long comportant des données manquantes,  $E(r_0)$  et  $E(r_i)$  peuvent être estimées par les valeurs échantillonnées  $r_0$  et  $r_i$ , ce qui revient à négliger la position des valeurs absentes et les caractéristiques numériques particulières. On écrit alors:

$$E(r_0) = r_0$$

$$E(r_i) = r_i$$

Avec cette approximation, les équations [7] et [11] se simplifient et s'écrivent:

$$r \approx r_0 + r_0 \alpha + 2 r_0 \alpha^2$$

et

$$r \approx r_i + \left[-1 + \frac{r_i}{2} + \frac{r_i^2}{2}\right]\alpha + \left[-1 + \frac{3 r_i^2}{4} + \frac{r_i^3}{4}\right]\alpha^2$$

On note que  $\alpha$  étant positif et petit devant 1, on a toujours:

$$1 > r_i > r > r_0 > 0 \quad \text{pour } r_i \text{ et } r_0 > 0$$

et

$$1 > r_i > 0 > r_0 > r > -1 \quad \text{pour } r_i > 0 \text{ et } r_0 < 0$$

et

$$0 > r_i > r_0 > r > -1 \quad \text{pour } r_i \text{ et } r_0 < 0$$

Si en première approximation, on ne considère que la partie linéaire des 2 équations, on peut facilement éliminer  $\alpha$  entre elles:

$$\alpha = \frac{r - r_0}{r_0} = \frac{r - r_i}{-1 + \frac{r_i}{2} + \frac{r_i^2}{2}} \quad \text{soit}$$

$$r \left[ \frac{1}{r_0} - \frac{2}{-2 + r_i + r_i^2} \right] = 1 - \frac{2 r_i}{-2 + r_i + r_i^2} = \frac{-2 - r_i + r_i^2}{-2 + r_i + r_i^2}$$

Notons  $r_c$  l'estimateur combiné défini par l'expression:

$$r_c = \frac{r_0 (2 + r_i - r_i^2)}{2 r_0 + 2 - r_i - r_i^2} = \frac{r_0 (1 + r_i) (2 - r_i)}{2 r_0 + (1 - r_i) (2 + r_i)} \quad [13]$$

Cette expression [13] permet d'évaluer  $r_c$  à partir de  $r_0$  et de  $r_i$  (tableau 5) et de tracer dans les coordonnées  $r_0$  et  $r_i$  en respectant la série d'inégalités précédentes, des abaques permettant d'évaluer très simplement  $r_c$  sans débiaiser formellement les évaluations  $r_0$  et  $r_i$  avec le pourcentage de valeurs manquantes  $\alpha = \frac{n}{N}$ .

#### Calcul du biais de $r_c$

Cas particulier: Si on suppose de plus que  $n \gg 1$  et si l'on pose  $\frac{n}{N} = \alpha$ , on peut développer le biais sur  $r_c$  dans ce cas asymptotique:

$$r = E(r_0) \cdot 1 + \alpha + 2\alpha^2$$

$$\text{et } r = \left[ -1 \frac{E(r_i)}{2} + \frac{E(r_i)^2}{2} \right] \alpha + \left[ -1 + \frac{3}{4} E(r_i)^2 + \frac{1}{4} E(r_i)^3 \right] \alpha^2$$

d'où:

$$E(r_i) - E(r_0) = 2\alpha [-2 + E(r_i) + E(r_i)^2 - 2E(r_0)] + \\ \alpha^2 [-1 + \frac{3}{2} E(r_i)^2 + \frac{3}{2} E(r_i)^3 - 2 E(r_0)]$$

L'équation précédente développée jusqu'à l'ordre  $\alpha^2$  s'écrit:

$$\alpha = \frac{r - E(r_0)}{E(r_0) (1 + 2\alpha)} = \frac{r - E(r)}{E(r_i) E(r_i)^2 (-1 + \frac{E(r_i)}{2} + \frac{E(r_i)^2}{2}) + \alpha (-1 + \frac{3}{2} E(r_i)^2 + \frac{3}{2} E(r_i)^3)}$$

d'où l'on extrait la valeur exacte de  $r$ :

$$r = \frac{E(r_0) [(-1 - \frac{E(r_i)}{2} + \frac{E(r_i)^2}{2} + \alpha (-1 - 2 E(r_i) + \frac{3}{2} E(r_i)^2 + \frac{3}{2} E(r_i)^3))]}{(-E(r_0) - 1 + \frac{E(r_i)}{2} + \frac{E(r_i)^2}{2}) + \alpha (-1 - 2 E(r_0) + \frac{3}{2} E(r_i)^2 + \frac{3}{2} E(r_i)^3)}$$

$$r = \frac{E(r_0) [2 + E(r_i) - E(r_i)^2]}{2 E(r_0) + 2 - E(r_i) - E(r_i)^2} \frac{1 - \frac{\alpha}{2} \frac{-4 - 8 E(r_i) + 3 E(r_i)^2 + E(r_i)^3}{2 + E(r_i) - E(r_i)^2}}{1 - \frac{\alpha}{2} \frac{-8 E(r_0) - 4 + 3 E(r_i)^2 + E(r_i)^3}{2 E(r_0) + 2 - E(r_i) - E(r_i)^2}}$$

Ce qui, à l'ordre  $\alpha$  s'écrit après développement:

$$r = E(r_c) [1 - \alpha \frac{[E(r_i) - E(r_0)][E(r_i)^2 + E(r_i) - 2]}{[E(r_i) + 1][2 E(r_0) + 2 - E(r_i) - E(r_i)^2]}]$$

Si on remplace  $E(r_i) - E(r_0)$  au même degré d'approximation, il vient:

$$r = E(r_c) \left[ 1 - 2 \alpha^2 \frac{2 - E(r_i) - E(r_i)^2}{1 + E(r_i)} \right]$$

Si on estime  $E(r_c)$  par  $E(r_i)$  dans la correction, on trouve:

$$BA = r - E(r_c) = -2 \alpha^2 \left[ \frac{2 E(r_c)}{1 + E(r_c)} - E(r_c)^2 \right]$$

expression négative et très faible, qui réduit bien le biais à l'ordre  $\alpha^2$ .

Les lignes d'iso-autocorrélation  $r_c = k$  ( $|k| < 1$ ) sont représentées par une famille de courbes:

$$r_0 = \frac{k (2 - r_i - r_i^2)}{2 - 2k + r_i - r_i^2} \quad [14]$$

Ces abaques ont été tracées à la figure 1 en tenant compte des mêmes inégalités. On note un point singulier autour de  $r_i = 1$  où  $r_c$  est mal défini.

## 2.6 Etude de la suppression des valeurs absentes occasionnelles

Il arrive quelquefois dans la pratique que, quand on dispose d'un échantillon assez long comprenant quelques valeurs absentes, on calcule l'autocorrélation sur l'échantillon raccourci des valeurs manquantes. Une telle expression est biaisée et c'est ce biais que nous allons évaluer.

Si on note par  $r_D$  la valeur de l'autocorrélation calculée pour un échantillon de longueur  $(N - n)$ , chaque valeur absente fait disparaître au numérateur de l'équation [1] deux produits  $X_{i-1} X_i + X_i X_{i+1}$  et fait apparaître un produit  $X_{i-1} X_{i+1}$ . Au dénominateur, chaque valeur absente fait disparaître un carré  $X_i^2$ .

$$\text{Notons } \text{cov}2 = \frac{1}{N-2} \sum_{i=1}^{N-2} X_i X_{i+2} \text{ et } E(X_{i-1} X_{i+1}) = \text{cov}2$$

Si l'on estime les produits ainsi créés par leurs espérances mathématiques déduites des valeurs présentes dans l'échantillon, on obtient:

$$E(r_D) = \frac{(N-2n-1) \text{cov} + n \text{cov}2}{(N-n) \frac{N-1}{N} \text{var}} = \frac{N (N-2n-1)}{(N-n)(N-1)} r + \frac{nN}{(N-n)(N-1)} r_2 \quad [15]$$

Si  $r_2$  est négligeable ( $r_2 \neq 0$ ), on trouve une équation identique à l'équation [5].

Dans le cas général, on ne peut préciser le biais que si l'on définit le processus générateur de la série; si l'on suppose que le phénomène suit un processus Markovien d'ordre 1, ce qui est fréquent en hydrologie; on a alors  $r_2 = r^2$ ; l'équation précédente devient:

$$(N-n) (N-1) E(r_D) = N (N-2n-1) r + n N r^2$$

$$\text{soit } \frac{n}{N} r^2 + (1 - \frac{2n+1}{N}) r - (1 - \frac{n}{N}) (1 - \frac{1}{N}) E(r_D) = 0$$

$$r = \frac{-(1 - \frac{2n+1}{N}) \pm \sqrt{[1 - \frac{2n+1}{N}]^2 + 4 \frac{n}{N} (1 - \frac{n}{N}) (1 - \frac{1}{N}) E(r_D)}}{2 \frac{n}{N}}$$

La seule solution acceptable est la solution comportant le signe + d'où:

$$r = \frac{-(1 - \frac{2n+1}{N}) + (1 - \frac{2n+1}{N}) \sqrt{1 + \frac{\frac{4n}{N} (1 - \frac{n}{N}) (1 - \frac{1}{N}) E(r_D)}{[1 - \frac{2n+1}{N}]^2}}}{2 \frac{n}{N}} \quad [16]$$

Développons la quantité sous le radical en puissance de N jusqu'à l'ordre  $N^{-2}$ :

$$\begin{aligned} Z &= \frac{4 \frac{n}{N} (1 - \frac{n}{N}) (1 - \frac{1}{N}) E(r_D)}{[1 - \frac{2n+1}{N}]^2} = \frac{4 \frac{n}{N} E(r_D) (1 - \frac{n+1}{N} + \frac{n}{N^2})}{1 - \frac{4n+2}{N} + \frac{4n^2+4n+1}{N^2}} \\ &= 4 \frac{n}{N} E(r_D) (1 - \frac{n+1}{N} + \frac{n}{N^2}) (1 + \frac{4n+2}{N} + \frac{12n^2 + 12n + 3}{N^2}) \\ &= 4 \frac{n}{N} E(r_D) [1 + \frac{3n+1}{N} + \frac{8n^2 + 7n + 1}{N^2}] \end{aligned}$$

Utilisant l'identité  $(1 + \epsilon)^{\frac{1}{2}} \equiv 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{8} + \dots$ , on tire à l'ordre  $N^{-3}$ :

$$\begin{aligned} \sqrt{1+Z} &= 1 + \frac{2n}{N} E(r_D) [1 + \frac{3n+1}{N} + \frac{8n^2 + 7n + 1}{N^2}] \\ &\quad - \frac{16n^2}{8N^2} E(r_D)^2 [1 + \frac{6n+2}{N}] \\ &= 1 + \frac{2n}{N} E(r_D) [1 + \frac{(3 - E(r_D)) n+1}{N}] \end{aligned}$$



Le tableau 6 présente les biais liés à l'utilisation de ce mode d'estimation pour un échantillon long comportant  $\alpha\%$  de valeurs absentes.

## 2.7 Comparaison des biais théoriques

Les tableaux 3, 4 et 6 présentent les biais pour des échantillons longs, comportant  $\alpha\%$  valeurs absentes; les équations correspondantes sont:

$$r - E(r_0) = \alpha (1 + 2\alpha) E(r_0) \quad [8]$$

$$r - E(r_i) = -\alpha \left[ 1 - \frac{1}{2} E(r_i) - \frac{1}{2} E(r_i)^2 \right] - \alpha^2 \left[ 1 - \frac{3}{4} E(r_i)^2 - \frac{1}{4} E(r_i)^3 \right] \quad [12]$$

$$r - E(r_D) = \alpha [(1 - E(r_D)) + \alpha (2 - E(r_D))] E(r_D) \quad [18]$$

Ces biais ont été comparés à la figure 2 pour  $\alpha = 5\%$  et  $\alpha = 20\%$ . On y note que, dans tous les cas, le biais introduit par la suppression des valeurs absentes est le plus faible, que les biais introduits par les calculs de  $r_0$  et de  $r_i$  sont des signes opposés et du même ordre de grandeur,  $r_0$  étant nul pour  $E(r) = 0$  et croissant linéairement, et  $r_i$  étant nul pour  $E(r) = 1$  et croissant en valeur absolue non linéairement pour des valeurs de  $E(r)$  plus faibles. On note aussi que ces deux biais ont à peu près la même valeur pour  $E(r) = 0.5$ ; Quand on ne dispose que d'un échantillon pour évaluer  $E(r)$ , donc les corrections de biais, l'effet de cette erreur d'échantillonnage est minimisé si on utilise l'estimateur  $r_0$  pour une valeur échantillonnée petite ( $\hat{r} < 0.5$ ) et l'estimateur  $r_i$  pour une valeur échantillonnée grande ( $\hat{r} > 0.5$ ) en appliquant bien entendu, les corrections de biais pertinentes. Cette recommandation est très générale, car aucune hypothèse n'a été faite sur le processus générateur de la population-mère; le biais induit par l'estimateur combiné  $r_c$  étant de l'ordre de  $\alpha^2$ , il n'a pas été représenté à la figure 2.

## 2.8 Cas d'un échantillon court

Les corrections précédentes [8], [12] et [18] ont été calculées pour des échantillons longs et centrés; avec ces hypothèses, la valeur vraie de l'échantillon complet  $r$ , que l'on cherche à évaluer par le débiaisage des estimateurs  $r_0$ ,  $r_i$  et  $r_D$ , est identique à celle de la population-mère  $\rho$ .

Si l'échantillon évalué est court, sa moyenne vraie peut être modifiée par les estimateurs ce qui introduit un nouveau biais; KENDALL (1954) a présenté une correction applicable à un échantillon court, mais complet pour évaluer l'autocorrélation de la population-mère; WALLIS et O'CONNELL (1972) ont montré que cette correction bien que dérivée théoriquement de la définition circulaire de  $\rho$  peut s'appliquer aussi bien à la définition classique d'une série "ouverte".

$$\rho = \frac{E(r) + \frac{1}{N}}{1 - \frac{4}{N}} \quad [4]$$

Cette correction, présentée au paragraphe 2.2 contient deux sources de biais liées qui vont se combiner à l'effet des valeurs absentes estimées, proprement dites (leur nombre et, éventuellement leurs positions dans l'échantillon) tel que représenté par les équations [6], [10] et [16]:

$$r = \frac{(1 - \frac{1}{N}) (1 - \frac{n}{N})}{1 - \frac{2n+1}{N}} E(r_0) \quad [6]$$

$$r = \frac{\frac{(2N-n)(N-1)}{2(N-n-1)} \frac{1}{N} E(r_i) - \frac{n}{N} \frac{(N-1)}{(N-n-1)}}{1 - \frac{n}{2(N-n-1)} E(r_i)} \quad [10]$$

Les résultats sont présentés aux tableaux numérotés comme suit:

cor- rection \ lon- gueur					
	100	80	60	40	20
$E(r_0)$	T-7	T-8	T-9	T-10	T-11
$E(r_i)$	T-12	T-13	T-14	T-15	T-16
$E(r_D)$	T-17	T-18	T-19	T-20	T-21

Dans le chapitre suivant, nous allons tester par simulation si, dans le cas de la population-mère Markovienne, l'application successive des deux corrections peut fournir des estimations non-biaisées de  $r$ .

Les biais résiduels moyens seront calculés et surtout les variances d'estimation seront calculées, ce qui permettra d'évaluer le mode d'estimation des valeurs absentes donnant un biais résiduel moyen et une variance d'estimation minimaux.

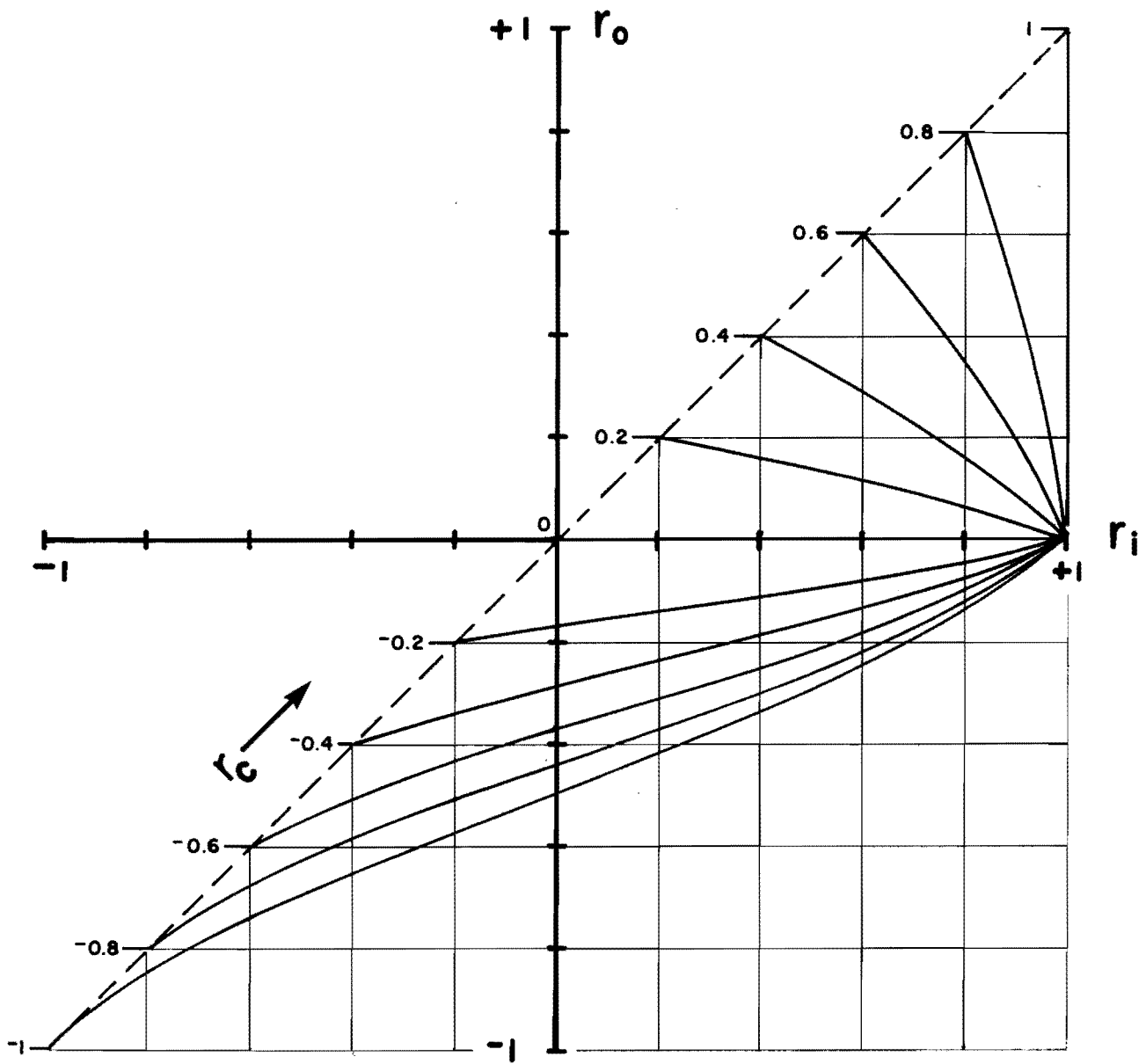


Figure 1 : Abaques reliant  $r_0$  ,  $r_i$  et  $r_c$ .

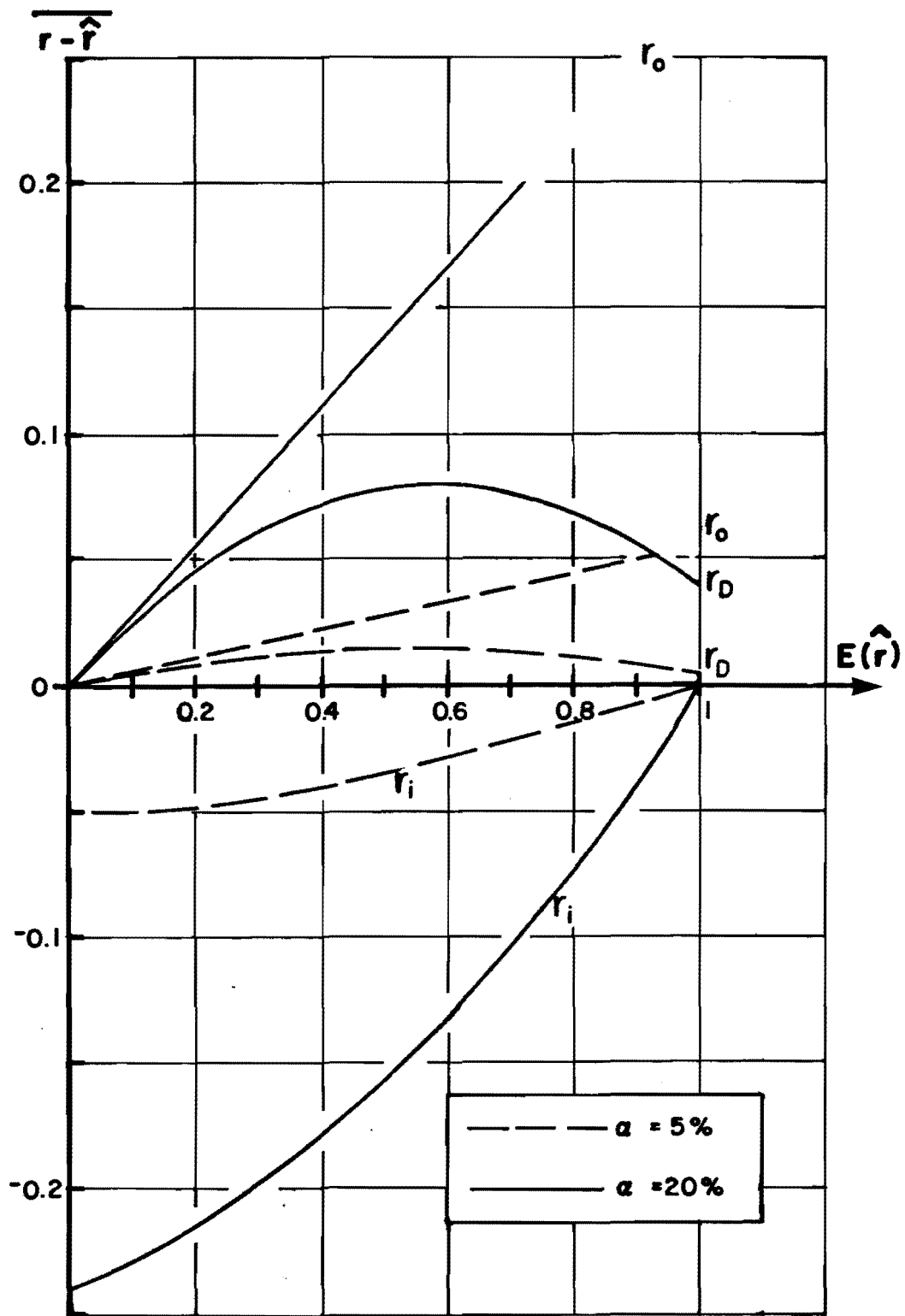


Figure 2 : Comparaison des biais pour différents estimateurs des valeurs absentes.

Tableau 1 : Biais moyen  $\rho - E(r)$  d'un échantillon de longueur N provenant d'une série Markovienne d'ordre 1.

N E(r)					
	100	80	60	40	20
0	0.010	0.013	0.018	0.028	0.063
0.1	0.014	0.018	0.025	0.039	0.088
0.2	0.019	0.024	0.032	0.050	0.112
0.3	0.023	0.029	0.039	0.061	0.138
0.4	0.027	0.034	0.046	0.072	0.163
0.5	0.031	0.039	0.054	0.083	0.188
0.6	0.035	0.045	0.061	0.094	0.212
0.7	0.040	0.050	0.068	0.106	0.238
0.8	0.044	0.055	0.075	0.117	*
0.9	0.048	0.060	0.082	*	*

\* Valeurs impossibles

Tableau 2 : Variance d'estimation d'une série markovienne d'ordre 1 de paramètre  $\rho$  et de longueur  $N$ .

$\rho \backslash N$	N				
	20	40	60	80	100
0	0.050	0.025	0.017	0.013	0.010
0.1	0.049	0.025	0.017	0.012	0.010
0.2	0.048	0.024	0.016	0.012	0.010
0.3	0.045	0.023	0.015	0.011	0.009
0.4	0.042	0.021	0.014	0.010	0.008
0.5	0.038	0.019	0.013	0.009	0.007
0.6	0.032	0.016	0.011	0.008	0.006
0.7	0.026	0.013	0.008	0.006	0.005
0.8	0.018	0.009	0.006	0.004	0.004
0.9	0.010	0.005	0.003	0.002	0.002
1	0	0	0	0	0

\* Valeurs impossibles

Tableau 3 : Biais  $r - E(r_0)$  liés à l'utilisation de l'estimateur "moyenne générale" pour un échantillon long comportant  $\alpha\%$  valeurs absentes.

$\alpha$ $E(r_0)$	5%	10%	15%	20%
0	0	0	0	0
0.05	0.003	0.006	0.010	0.014
0.1	0.006	0.012	0.020	0.028
0.2	0.011	0.024	0.039	0.056
0.3	0.017	0.036	0.059	0.084
0.4	0.022	0.048	0.078	0.112
0.5	0.028	0.060	0.098	0.140
0.6	0.033	0.072	0.117	0.168
0.7	0.039	0.084	0.136	0.196
0.8	0.044	0.096	0.156	*
0.9	0.050	*	*	*
0.95	*	*	*	*
1	*	*	*	*

\* Valeurs impossibles



Tableau 4 : Biais  $r - E(r_i)$  liés à l'utilisation de l'estimateur  
"moyenne locale" pour un échantillon long comportant  $\alpha\%$   
valeurs absentes.

$\alpha$ $E(r_i)$	5%	10%	15%	20%
0	-0.053	-0.111	-0.172	-0.240
0.05	-0.051	-0.107	-0.169	-0.235
0.1	-0.050	-0.104	-0.164	-0.229
0.2	-0.047	-0.098	-0.154	-0.215
0.3	-0.043	-0.090	-0.142	-0.198
0.4	-0.038	-0.081	-0.127	-0.179
0.5	-0.033	-0.070	-0.111	-0.156
0.6	-0.028	-0.059	-0.093	-0.131
0.7	-0.022	-0.046	-0.073	-0.103
0.8	-0.015	-0.032	-0.051	-0.072
0.9	-0.008	-0.017	-0.027	-0.037
0.95	-0.004	-0.008	-0.013	-0.019
1	0	0	0	0

Tableau 5 : Valeurs de  $r$  en fonction de  $r_0$  et  $r_1$  pour un échantillon long avec des données manquantes occasionnelles

$r_i \backslash r_0$	-1	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
-1	-1.00																				
-0.9		-0.90																			
-0.8			-0.80																		
-0.7				-0.70																	
-0.6					-0.60																
-0.5						-0.50															
-0.4							-0.40														
-0.3								-0.30													
-0.2									-0.20												
-0.1										-0.10											
0											0.00										
0.1												0.10									
0.2													0.20								
0.3														0.30							
0.4															0.40						
0.5																0.50					
0.6																	0.60				
0.7																		0.70			
0.8																			0.80		
0.9																				0.90	
1																					1.00

Tableau 6 : Biais  $r - E(r_D)$  liés à l'utilisation de l'estimateur  $r_D$  pour un échantillon long comportant  $\alpha\%$  de valeurs absentes.

$\alpha$ $E(r_D)$	5%	10%	15%	20%
0	0	0	0	0
0.05	0.003	0.006	0.009	0.013
0.1	0.005	0.011	0.018	0.026
0.2	0.009	0.020	0.032	0.046
0.3	0.012	0.026	0.043	0.062
0.4	0.014	0.030	0.050	0.074
0.5	0.014	0.033	0.054	0.080
0.6	0.014	0.032	0.055	0.082
0.7	0.013	0.030	0.052	0.078
0.8	0.010	0.026	0.046	0.070
0.9	0.007	0.019	0.036	0.058
0.95	0.005	0.015	0.030	0.049
1	*	*	*	*

\* Valeur impossible

Tableau 7 : Biais combiné  $\rho - E(r_0)$ , estimateur "moyenne générale".

$E(r_0)$	N = 100							
	n = 5		n = 10		n = 15		n = 20	
	A	P	A	P	A	P	A	P
0	0.011	0.010	0.012	0.010	0.013	0.010	0.014	0.010
0.1	0.021	0.020	0.029	0.028	0.040	0.037	0.054	0.050
0.2	0.031	0.031	0.047	0.045	0.067	0.064	0.094	0.090
0.3	0.041	0.041	0.064	0.063	0.094	0.092	0.133	0.130
0.4	0.051	0.051	0.082	0.080	0.121	0.119	0.173	0.170
0.5	0.061	0.061	0.099	0.098	0.148	0.146	0.213	0.210
0.6	0.071	0.071	0.117	0.115	0.175	0.173	0.253	0.249
0.7	0.082	0.081	0.134	0.133	0.202	0.200	0.293	0.289
0.8	0.092	0.091	0.152	0.150	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* Valeur impossible

Tableau 8 : Biais combiné  $\rho - E(r_0)$ , estimateur "moyenne générale".

$E(r_0)$	N = 80							
	n = 4		n = 8		n = 12		n = 16	
	A	P	A	P	A	P	A	P
0	0.014	0.013	0.015	0.013	0.016	0.013	0.018	0.013
0.1	0.025	0.024	0.034	0.032	0.045	0.042	0.059	0.055
0.2	0.036	0.036	0.052	0.051	0.073	0.070	0.100	0.096
0.3	0.048	0.047	0.071	0.070	0.102	0.099	0.142	0.138
0.4	0.059	0.058	0.090	0.088	0.130	0.127	0.184	0.179
0.5	0.070	0.070	0.109	0.107	0.159	0.156	0.225	0.221
0.6	0.082	0.081	0.128	0.126	0.187	0.184	0.267	0.262
0.7	0.093	0. 92	0.146	0.144	0.216	0.213	*	*
0.8	0.104	0.103	0.165	0.163	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

Tableau 9 : Biais combiné  $\rho - E(r_0)$ , estimateur "moyenne générale".

$E(r_0)$	N = 60							
	n = 3		n = 6		n = 9		n = 12	
	A	P	A	P	A	P	A	P
0	0.019	0.018	0.020	0.018	0.022	0.018	0.024	0.019
0.1	0.032	0.031	0.041	0.039	0.053	0.049	0.069	0.062
0.2	0.046	0.044	0.062	0.060	0.084	0.080	0.113	0.107
0.3	0.059	0.058	0.083	0.081	0.115	0.111	0.158	0.151
0.4	0.072	0.071	0.104	0.102	0.146	0.142	0.202	0.196
0.5	0.085	0.084	0.125	0.123	0.177	0.173	0.247	0.240
0.6	0.099	0.098	0.147	0.144	0.208	0.204	0.291	0.285
0.7	0.112	0.111	0.168	0.165	0.239	0.235	*	*
0.8	0.125	0.124	0.189	0.186	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

Tableau 10 : Biais combiné  $\rho - E(r_0)$ , estimateur "moyenne générale".

$E(r_0)$	N = 40							
	n = 2		n = 4		n = 6		n = 8	
	A	P	A	P	A	P	A	P
0	0.029	0.028	0.031	0.028	0.034	0.028	0.038	0.028
0.1	0.047	0.045	0.057	0.054	0.071	0.064	0.088	0.079
0.2	0.065	0.063	0.083	0.079	0.107	0.101	0.139	0.129
0.3	0.082	0.081	0.109	0.105	0.143	0.137	0.190	0.180
0.4	0.100	0.098	0.135	0.131	0.180	0.174	0.241	0.231
0.5	0.118	0.116	0.160	0.157	0.216	0.210	0.291	0.281
0.6	0.135	0.134	0.186	0.183	0.253	0.246	0.342	0.332
0.7	0.153	0.151	0.212	0.208	0.289	0.283	*	*
0.8	0.170	0.169	*	*	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

Tableau 11 : Biais combiné  $\rho - E(r_0)$ , estimateur "moyenne générale".

$E(r_0)$	N = 20							
	n = 1		n = 2		n = 3		n = 4	
	A	P	A	P	A	P	A	P
0	0.066	0.063	0.071	0.062	0.077	0.062	0.086	0.062
0.1	0.099	0.095	0.134	0.105	0.133	0.118	0.159	0.135
0.2	0.132	0.128	0.156	0.147	0.188	0.173	0.232	0.208
0.3	0.164	0.161	0.199	0.190	0.243	0.228	0.305	0.281
0.4	0.197	0.193	0.241	0.233	0.299	0.284	0.377	0.353
0.5	0.230	0.226	0.284	0.275	0.354	0.339	0.450	0.426
0.6	0.263	0.259	0.326	0.318	*	0.394	*	*
0.7	0.295	0.292	*	*	*	*	*	*
0.8	*	*	*	*	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible



Tableau 12 : Biais combiné  $\rho - E(r_i)$ , estimateur "moyenne locale".

$E(r_i)$	N = 100							
	n = 5		n = 10		n = 15		n = 20	
	A	P	A	P	A	P	A	P
0	-0.042	-0.044	-0.100	-0.105	-0.166	-0.174	-0.239	-0.251
0.1	-0.034	-0.037	-0.090	-0.095	-0.152	-0.161	-0.223	-0.235
0.2	-0.027	-0.030	-0.079	-0.084	-0.137	-0.146	-0.204	-0.216
0.3	-0.019	-0.021	-0.066	-0.072	-0.120	-0.129	-0.182	-0.195
0.4	-0.010	-0.013	-0.052	-0.058	-0.100	-0.109	-0.156	-0.170
0.5	0.000	-0.003	-0.036	-0.043	-0.078	-0.088	-0.126	-0.141
0.6	0.010	0.007	-0.019	-0.026	-0.053	-0.064	-0.093	-0.109
0.7	0.021	0.017	0.000	-0.008	-0.026	-0.038	-0.056	-0.074
0.8	0.032	0.029	0.079	0.011	0.004	-0.010	-0.015	-0.034
0.9	0.044	0.041	0.041	0.032	0.036	0.021	0.030	0.009

Tableau 13 : Biais combiné  $\rho - E(r_i)$ , estimateur "moyenne locale".

$E(r_i)$	N = 80							
	n = 4		n = 8		n = 12		n = 16	
	A	P	A	P	A	P	A	P
0	-0.039	-0.042	-0.097	-0.104	-0.163	-0.173	-0.236	-0.251
0.1	-0.031	-0.034	-0.080	-0.093	-0.148	-0.159	-0.219	-0.234
0.2	-0.022	-0.025	-0.073	-0.080	-0.132	-0.143	-0.198	-0.214
0.3	-0.012	-0.016	-0.059	-0.066	-0.113	-0.124	-0.174	-0.191
0.4	-0.002	-0.006	-0.043	-0.051	-0.091	-0.104	-0.147	-0.164
0.5	0.009	0.005	-0.027	-0.035	-0.068	-0.080	-0.116	-0.134
0.6	0.020	0.016	-0.008	-0.017	-0.041	-0.056	-0.081	-0.101
0.7	0.032	0.028	0.012	0.002	-0.013	-0.028	-0.042	-0.064
0.8	0.045	0.040	0.033	0.023	0.019	-0.002	0.001	-0.023
0.9	0.058	0.053	0.056	0.045	0.053	0.034	0.049	0.022

Tableau 14 : Biais combiné  $\rho - E(r_i)$ , estimateur "moyenne locale".

$E(r_i)$	N = 60							
	n = 3		n = 6		n = 9		n = 12	
	A	P	A	P	A	P	A	P
0	-0.034	-0.039	-0.093	-0.101	-0.158	-0.172	-0.231	-0.251
0.1	-0.024	-0.028	-0.079	-0.088	-0.141	-0.155	-0.212	-0.232
0.2	-0.013	-0.018	-0.064	-0.074	-0.122	-0.137	-0.188	-0.210
0.3	-0.001	-0.006	-0.048	-0.058	-0.101	-0.116	-0.162	-0.184
0.4	0.011	0.006	-0.030	-0.040	-0.077	-0.094	-0.133	-0.155
0.5	0.024	0.018	-0.010	-0.022	-0.050	-0.068	-0.097	-0.123
0.6	0.037	0.032	0.010	-0.002	-0.021	-0.041	-0.059	-0.087
0.7	0.052	0.046	0.033	0.019	0.010	-0.011	-0.017	-0.047
0.8	0.066	0.060	0.057	0.042	0.045	0.022	0.030	-0.003
0.9	0.082	0.075	0.082	0.067	0.082	0.057	0.082	0.045

Tableau 15 : Biais combiné  $\rho - E(r_i)$ , estimateur "moyenne locale".

$E(r_i)$	N = 40							
	n = 2		n = 4		n = 6		n = 8	
	A	P	A	P	A	P	A	P
0	-0.024	-0.031	-0.082	-0.096	-0.147	-0.169	-0.221	-0.252
0.1	-0.010	-0.016	-0.064	-0.078	-0.126	-0.148	-0.196	-0.228
0.2	0.006	-0.001	-0.045	-0.060	-0.102	-0.125	-0.167	-0.201
0.3	0.022	0.014	-0.024	-0.039	-0.075	-0.100	-0.135	-0.170
0.4	0.038	0.030	-0.001	-0.017	-0.045	-0.072	-0.099	-0.136
0.5	0.056	0.047	0.024	0.006	-0.014	-0.042	-0.058	-0.099
0.6	0.074	0.065	0.050	0.030	0.021	-0.010	-0.013	-0.057
0.7	0.092	0.083	0.077	0.056	0.059	-0.026	0.037	-0.012
0.8	0.112	0.102	0.106	0.084	0.100	0.063	0.092	0.038
0.9	*	*	*	*	*	*	*	0.093

\* valeur impossible

Tableau 16 : Biais combiné  $\rho - E(r_i)$ , estimateur "moyenne locale".

$E(r_i)$	N = 20							
	n = 1		n = 2		n = 3		n = 4	
	A	P	A	P	A	P	A	P
0	0.012	-0.003	-0.046	-0.077	-0.110	-0.160	-0.184	-0.254
0.1	0.041	0.025	-0.012	-0.044	-0.072	-0.124	-0.141	-0.214
0.2	0.071	0.055	0.024	-0.010	-0.030	-0.085	-0.093	-0.170
0.3	0.102	0.085	0.062	0.026	0.015	-0.043	-0.039	-0.122
0.4	0.134	0.116	0.102	0.063	0.064	0.002	0.019	-0.070
0.5	0.168	0.748	0.144	0.102	0.117	0.049	0.084	-0.073
0.6	0.202	0.780	0.189	0.143	0.173	0.099	0.155	0.048
0.7	0.236	0.214	0.235	0.186	0.234	0.153	0.232	0.113
0.8	*	*	*	*	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

Tableau 17 : Biais combiné  $\rho - E(r_D)$ , valeurs absentes omises.

$E(r_D)$	N = 100							
	n = 5		n = 10		n = 15		n = 20	
	A	P	A	P	A	P	A	P
0	0.011	0.010	0.011	0.010	0.012	0.010	0.012	0.010
0.1	0.020	0.020	0.025	0.025	0.029	0.028	0.030	0.029
0.2	0.028	0.028	0.038	0.037	0.044	0.043	0.045	0.044
0.3	0.035	0.035	0.048	0.048	0.057	0.057	0.058	0.058
0.4	0.042	0.042	0.057	0.057	0.067	0.068	0.068	0.069
0.5	0.047	0.047	0.064	0.065	0.076	0.077	0.076	0.077
0.6	0.051	0.052	0.069	0.071	0.082	0.083	0.081	0.083
0.7	0.054	0.056	0.073	0.075	0.086	0.088	0.084	0.086
0.8	0.057	0.059	0.075	0.077	0.087	0.090	0.084	0.088
0.9	0.058	0.060	0.075	0.078	0.087	0.091	0.082	0.086

Tableau 18 : Biais combiné  $\rho - E(r_D)$ , valeurs absentes omises.

$E(r_D)$	N = 80							
	n = 4		n = 8		n = 12		n = 16	
	A	P	A	P	A	P	A	P
0	0.011	0.013	0.011	0.013	0.012	0.013	0.012	0.013
0.1	0.020	0.024	0.025	0.028	0.029	0.032	0.029	0.033
0.2	0.028	0.033	0.038	0.042	0.044	0.048	0.045	0.049
0.3	0.035	0.042	0.048	0.054	0.057	0.063	0.058	0.064
0.4	0.042	0.049	0.057	0.065	0.068	0.075	0.068	0.076
0.5	0.047	0.056	0.064	0.073	0.076	0.085	0.076	0.085
0.6	0.051	0.062	0.069	0.080	0.082	0.093	0.081	0.092
0.7	0.055	0.066	0.073	0.086	0.086	0.093	0.084	0.097
0.8	0.057	0.070	0.075	0.089	0.087	0.102	0.084	0.099
0.9	0.058	0.073	0.075	0.091	0.087	*	0.082	0.098

\* valeur impossible

Tableau 19 : Biais combiné  $\rho - E(r_D)$ , valeurs absentes omises.

$E(r_D)$	N = 60							
	n = 3		n = 6		n = 9		n = 12	
	A	P	A	P	A	P	A	P
0	0.019	0.018	0.020	0.018	0.020	0.018	0.020	0.018
0.1	0.031	0.030	0.037	0.035	0.040	0.039	0.041	0.039
0.2	0.042	0.042	0.052	0.051	0.058	0.057	0.059	0.058
0.3	0.052	0.052	0.065	0.065	0.074	0.073	0.074	0.074
0.4	0.061	0.062	0.077	0.077	0.087	0.088	0.087	0.088
0.5	0.070	0.070	0.087	0.088	0.098	0.100	0.097	0.099
0.6	0.077	0.078	0.095	0.097	0.107	0.109	0.104	0.108
0.7	0.083	0.085	0.101	0.104	0.113	0.117	0.109	0.114
0.8	0.088	0.090	0.105	0.110	0.117	0.122	0.111	0.118
0.9	0.092	0.095	*	*	*	*	*	*

\* valeur impossible



Tableau 20 : Biais combiné  $\rho - E(r_D)$ , valeurs absentes omises.

$E(r_D)$	N = 40							
	n = 2		n = 4		n = 6		n = 8	
	A	P	A	P	A	P	A	P
0	0.029	0.028	0.031	0.028	0.031	0.028	0.031	0.028
0.1	0.045	0.044	0.052	0.049	0.056	0.053	0.056	0.053
0.2	0.061	0.060	0.071	0.069	0.077	0.075	0.077	0.076
0.3	0.075	0.074	0.088	0.088	0.097	0.096	0.096	0.096
0.4	0.087	0.088	0.103	0.104	0.113	0.114	0.112	0.113
0.5	0.100	0.101	0.117	0.119	0.128	0.130	0.125	0.128
0.6	0.110	0.113	0.128	0.132	0.139	0.144	0.135	0.141
0.7	0.120	0.123	0.138	0.143	0.148	0.155	0.142	0.151
0.8	0.129	0.133	0.145	0.156	0.155	0.165	0.147	0.158
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

Tableau 21 : Biais combiné  $\rho - E(r_D)$ , valeurs absentes omises.

$E(r_D)$	N = 20							
	n = 1		n = 2		n = 3		n = 4	
	A	P	A	P	A	P	A	P
0	0.066	0.063	0.068	0.063	0.070	0.063	0.070	0.063
0.1	0.096	0.093	0.103	0.099	0.108	0.102	0.107	0.101
0.2	0.125	0.123	0.136	0.134	0.142	0.139	0.140	0.138
0.3	0.152	0.152	0.166	0.166	0.174	0.174	0.170	0.171
0.4	0.178	0.180	0.194	0.197	0.203	0.206	0.196	0.202
0.5	0.203	0.207	0.220	0.226	0.228	0.236	0.219	0.229
0.6	0.227	0.232	0.243	0.253	0.251	0.263	0.239	0.255
0.7	0.249	0.257	0.264	0.278	0.270	0.288	0.255	0.277
0.8	*	*	*	*	*	*	*	*
0.9	*	*	*	*	*	*	*	*

\* valeur impossible

### 3. Etude des biais résiduels et des variances d'estimation par simulation.

#### 3.1 Technique de Monte-Carlo

Pour évaluer l'efficacité des corrections de biais théoriques précédentes liées aux différents modes d'estimation des valeurs absentes, nous avons utilisé la technique de Monte-Carlo en générant 2000 échantillons synthétiques de longueur 100, 80, 60, 40 et 20 provenant de populations markoviennes d'ordre 1, dont le paramètre  $\rho$  est connu et vaut 0, 0.1, 0.2 ... 0.8, 0.9 et 0.95. Ces séries sont générées en utilisant la formule récurrente.

$$x_i = \rho x_{i-1} + (1 - \rho^2)^{\frac{1}{2}} \varepsilon_i \quad [19]$$

Dans cette formule, les  $\varepsilon_i$  sont des éléments tirés d'une population normale et indépendante de moyenne 0 et d'écart-type 1, générés par l'algorithme de BOX et MULLER (1958). Cette technique consiste à tirer au hasard, en utilisant en double précision le générateur pseudo-aléatoire contenu dans les bibliothèques de la plupart des ordinateurs, deux nombres  $U_1$  et  $U_2$  distribués uniformément dans l'intervalle 0 ... 1; BOX et MULLER ont démontré que les quantités

$$x_1 = \sqrt{-2 \ln U_1} \cos 2\pi U_2 \quad \text{et}$$

$$x_2 = \sqrt{-2 \ln U_1} \sin 2\pi U_1$$

sont des éléments indépendants tirés d'une population normale de moyenne 0 et d'écart-type 1, NIP (0, 1), et peuvent donc être utilisés comme  $\varepsilon_i$ .

A partir des 2000 séries ainsi générées, des valeurs absentes non consécutives ont été artificiellement créées pour tester la capacité des estimateurs précédents et des corrections de biais associées à régénérer le coefficient d'autocorrélation  $\rho$  de la population originale. Dans une première étape, la position des valeurs manquantes était aléatoire, mais comme les résultats se sont avérés stables à  $10^{-3}$  près pour 2000 échantillons (moyennes et variances), elle a ensuite été fixée, pour l'ensemble des 2000 séries de façon à rendre les valeurs absentes à peu près équidistantes.

### 3.2 Présentation des résultats

Pour chacun des estimateurs étudiés, nous avons calculé sur les 2000 échantillons le biais moyen résiduel et la variance d'estimation du coefficient d'autocorrélation calculé pour différentes valeurs du  $\rho$  de la population-mère. Pour chiffrer la puissance de nos corrections, nous avons évalué les mêmes statistiques sur 4 séries d'échantillons:

Séries 1 : Sur des échantillons complets (sans valeur manquante) pour servir de référence à la puissance des estimateurs.

Séries 2 : Sur des échantillons pour lesquels aucune correction n'a été effectuée.

Séries 3 : Sur des échantillons pour lesquels la correction de longueur n'a pas été effectuée.

Séries 4 : Sur des échantillons corrigés à la fois pour les biais des estimateurs et pour la longueur courte des séries par la formule de KENDALL (1954) telle qu'étudiée par WALLIS et O'CONNELL (1972) et représentée par notre équation [4].

En ce qui concerne les séries 1, nous présentons d'abord au tableau 22 le biais résiduel moyen; on note, par rapport au tableau 1 une diminution très importante du biais résiduel. Nous présentons ensuite

au tableau 23 l'augmentation de variance induite par la correction de longueur sur des échantillons complets pour différentes longueurs de l'échantillon et différentes valeurs du paramètre  $\rho$  de la population-mère; sur ce tableau, les colonnes 1 représentent les variances d'estimation pour des séries générées selon l'algorithme de l'équation [19] et les colonnes 2 représentent ces mêmes variances quand l'estimation de l'autocorrélation est complétée par la correction de longueur de KENDALL; on note une augmentation très importante de la variance pour des séries courtes ( $N=20$ ); on note que ces variances d'estimation sont similaires à celles développées théoriquement pour un échantillon long (équation [5]) dans le cas markovien (tableau 2) pour  $N = 100$  et  $80$ , mais sont nettement plus importantes pour  $N = 40$  et  $20$  surtout si  $\rho$  est grand, (proche de 1).

En ce qui concerne les séries 2, 3 et 4 des résultats partiels,  $N = 100$   $n = 10$ ,  $N = 60$   $n = 9$  et  $N = 20$  et  $n = 4$  sont présentés aux tableaux 24, 25 et 26.

Le tableau 24 montre l'étendue des biais obtenus quand on applique les estimateurs sous correction (séries 2); on note pour les séries assez longues  $r_c$  et  $r_D$  sont les plus efficaces alors que pour les séries très courtes tous les estimateurs sont très mauvais.

Le tableau 25 présente les résultats pour nos quatre estimateurs de la série 3; on note que les biais résiduels  $y$  sont dans tous les cas très importants et positifs, ce qui était prévisible (correction de KENDALL, non effectuée) et nous n'avons pas poussé plus avant cette analyse.

En ce qui concerne les séries 4, l'analyse complète a été effectuée avec les deux corrections. Pour les estimateurs  $r_0$ ,  $r_i$  et  $r_D$  il existe deux permutations possibles de calcul en effectuant la correction de longueur soit antérieurement, soit postérieurement; en ce qui concerne l'estimateur  $r_c$ , il existe 5 combinaisons possibles dans l'ordre des corrections; les résultats des simulations montrent que si

la correction de longueur est effectuée sur  $r_0$ ,  $r_i$  et  $r_D$  antérieurement à la correction pour les valeurs absentes, les résultats sont les meilleurs à la fois pour les biais résiduels et la variance d'échantillonnage. Ce sont les seuls résultats que nous présenterons ici. Pour montrer l'importance de la correction de longueur, nous présentons au tableau 26, les résultats comparables aux tableaux 24 et 25; la figure 3, les réductions de biais résiduels sont comparés dans le cas intermédiaire  $N = 60$   $n = 9$ ; on y note qu'il y a une réduction d'un facteur 3 environ du biais résiduel et que les estimateurs  $r_D$  et  $r_C$  sont les plus efficaces.

Les résultats complets pour les 2,000 simulations sont présentés aux tableaux 27 à 31 pour les biais résiduels et 32 à 36 pour les variances résiduels dans les 4 estimateurs selon la grille suivante:

Longueur N	100	80	60	40	20
Biais résiduel	T.27	T.28	T.29	T.30	T.31
Variance résiduelle	T.32	T.33	T.34	T.35	T.36

L'examen de ces tableaux et leur comparaison avec les tableaux 22 et 23 montrent que les biais et variances résiduels après les 2 corrections, sont du même ordre de grandeur (à un facteur 2 à 3) dans la plupart des cas que les biais et variances résiduels des échantillons complets corrigés pour l'effet de longueur; l'estimateur  $r_0$  présente un fort biais si  $\rho$  est élevé, mais si  $\rho$  est inférieur à 0.5, c'est le meilleur estimateur. En ce qui concerne  $r_i$ , c'est le meilleur estimateur dans le cas où  $\rho$  est supérieur à 0.5;  $r_C$  et  $r_D$  sont tous deux très efficaces sur l'ensemble de la gamme de  $\rho$ , avec

toutefois un avantage à  $r_c$ , car  $r_D$  présente autour de  $\rho = 1$  une zone d'instabilité due au point singulier de sa définition. En hydrologie, où  $\rho$  reste presque toujours inférieur à 0.7, l'emploi de  $r_c$  est suggéré, car c'est un estimateur asymptotiquement sans biais (à l'ordre  $\alpha^2 = \frac{n^2}{N^2}$ ) et que, étant construit à partir de  $r_0$  et de  $r_i$  qui ne font pas d'hypothèse sur le processus générateur, il est plus susceptible de donner de bons résultats que  $r_D$ , dont le débiaisage théorique suppose un processus markovien d'ordre 1.

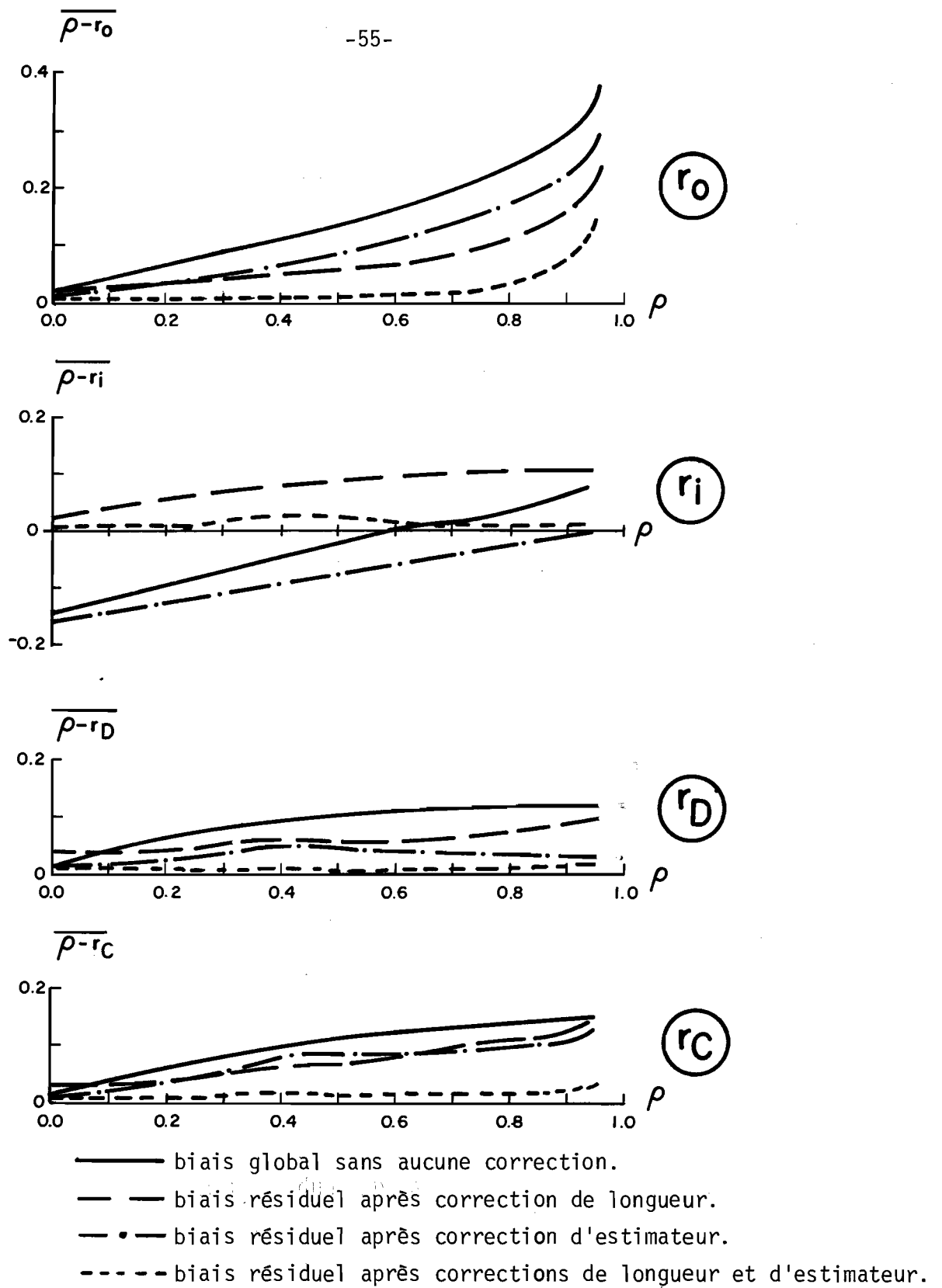


Figure 3 : Evolution des biais des estimateurs selon les corrections effectuées ( $N = 60$ ,  $n = 9$ ).



Tableau 22 : Biais résiduel moyens  $\rho - E(r)$  calculés sur 2,000 échantillons complets de longueur N et ayant subi la correction de longueur de KENDALL.

$\rho \backslash N$					
	100	80	60	40	20
0	0.003	0.003	0.003	0.005	0.000
0.1	-0.001	0.001	0.003	0.005	-0.006
0.2	-0.003	-0.004	-0.003	-0.003	-0.003
0.3	-0.000	0.002	0.001	-0.002	-0.003
0.4	0.003	0.005	0.005	0.006	0.003
0.5	-0.000	0.001	-0.001	-0.001	0.005
0.6	0.005	0.003	0.004	0.005	0.009
0.7	0.000	0.000	0.002	0.007	0.020
0.8	0.004	0.007	0.008	0.017	0.027
0.9	0.006	0.009	0.016	0.024	0.044
0.99	0.008	0.011	0.018	0.032	0.066

\* valeur impossible

Tableau 23 : Augmentation de la variance d'estimation induite par la correction de longueur pour 2,000 échantillons complets.

N ρ	100		80		60		40		20	
0	0.010	0.010	0.012	0.013	0.016	0.018	0.023	0.028	0.044	0.068
0.1	0.010	0.011	0.012	0.014	0.016	0.018	0.023	0.028	0.044	0.069
0.2	0.010	0.010	0.012	0.013	0.016	0.018	0.023	0.028	0.041	0.064
0.3	0.009	0.010	0.011	0.012	0.015	0.017	0.023	0.028	0.042	0.066
0.4	0.009	0.009	0.011	0.012	0.015	0.017	0.022	0.027	0.043	0.067
0.5	0.008	0.008	0.010	0.011	0.013	0.015	0.020	0.025	0.043	0.065
0.6	0.007	0.008	0.009	0.010	0.012	0.014	0.018	0.022	0.038	0.060
0.7	0.006	0.006	0.007	0.008	0.011	0.012	0.017	0.021	0.037	0.057
0.8	0.005	0.005	0.006	0.007	0.009	0.010	0.015	0.018	0.035	0.054
0.9	0.003	0.004	0.005	0.005	0.007	0.008	0.012	0.015	0.029	0.046
0.95	0.002	0.003	0.003	0.004	0.005	0.006	0.011	0.013	0.028	0.044
	1	2	1	2	1	2	1	2	1	2

Tableau 24 : Biais globaux induits par les estimateurs.

	N = 100				n = 10				N = 60				n = 9				N = 20				n = 4			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.011	-0.092	0.013	0.010	0.021	-0.130	0.025	0.020	0.051	-0.157	0.065	0.050	0.051	-0.157	0.065	0.050	0.051	-0.157	0.065	0.050	0.051	-0.157	0.065	0.050
0.1	0.023	-0.077	0.025	0.026	0.043	-0.110	0.046	0.044	0.084	-0.119	0.097	0.082	0.084	-0.119	0.097	0.082	0.084	-0.119	0.097	0.082	0.084	-0.119	0.097	0.082
0.2	0.038	-0.061	0.035	0.044	0.061	-0.091	0.057	0.063	0.135	-0.070	0.141	0.133	0.135	-0.070	0.141	0.133	0.135	-0.070	0.141	0.133	0.135	-0.070	0.141	0.133
0.3	0.057	-0.042	0.048	0.063	0.092	-0.062	0.080	0.092	0.178	-0.028	0.174	0.174	0.178	-0.028	0.174	0.174	0.178	-0.028	0.174	0.174	0.178	-0.028	0.174	0.174
0.4	0.072	-0.028	0.056	0.077	0.119	-0.036	0.096	0.116	0.233	0.019	0.213	0.225	0.233	0.019	0.213	0.225	0.233	0.019	0.213	0.225	0.233	0.019	0.213	0.225
0.5	0.087	-0.016	0.061	0.086	0.136	-0.022	0.099	0.125	0.279	0.054	0.242	0.263	0.279	0.054	0.242	0.263	0.279	0.054	0.242	0.263	0.279	0.054	0.242	0.263
0.6	0.106	0.001	0.068	0.098	0.167	0.002	0.110	0.144	0.330	0.092	0.265	0.304	0.330	0.092	0.265	0.304	0.330	0.092	0.265	0.304	0.330	0.092	0.265	0.304
0.7	0.120	0.010	0.065	0.097	0.196	0.020	0.111	0.152	0.388	0.133	0.290	0.347	0.388	0.133	0.290	0.347	0.388	0.133	0.290	0.347	0.388	0.133	0.290	0.347
0.8	0.140	0.025	0.068	0.099	0.231	0.044	0.117	0.162	0.463	0.174	0.314	0.401	0.463	0.174	0.314	0.401	0.463	0.174	0.314	0.401	0.463	0.174	0.314	0.401
0.9	0.170	0.040	0.068	0.096	0.287	0.069	0.121	0.174	0.569	0.218	0.341	0.486	0.569	0.218	0.341	0.486	0.569	0.218	0.341	0.486	0.569	0.218	0.341	0.486
0.95	0.205	0.047	0.067	0.097	0.356	0.080	0.121	0.193	0.697	0.250	0.366	0.629	0.697	0.250	0.366	0.629	0.697	0.250	0.366	0.629	0.697	0.250	0.366	0.629

Tableau 25 : Biais résiduels induits par différents estimateurs non corrigés pour l'effet de longueur.

	N = 100      n = 10				N = 60      n = 9				N = 20      n = 4			
$\rho$	$r_0$	$r_i$	$r_D$	$r_C$	$r_0$	$r_i$	$r_D$	$r_C$	$r_0$	$r_i$	$r_D$	$r_C$
0	0.013	0.014	0.016	0.014	0.026	0.027	0.037	0.030	0.070	0.069	0.134	0.094
0.1	0.013	0.023	0.017	0.015	0.031	0.046	0.040	0.035	0.078	0.096	0.137	0.098
0.2	0.018	0.032	0.020	0.021	0.030	0.054	0.036	0.037	0.110	0.136	0.154	0.130
0.3	0.026	0.044	0.027	0.031	0.045	0.073	0.050	0.054	0.131	0.167	0.164	0.154
0.4	0.031	0.050	0.031	0.038	0.057	0.087	0.059	0.069	0.170	0.203	0.190	0.194
0.5	0.033	0.052	0.034	0.042	0.054	0.087	0.056	0.069	0.194	0.223	0.203	0.218
0.6	0.043	0.059	0.041	0.052	0.071	0.097	0.067	0.085	0.227	0.245	0.215	0.249
0.7	0.046	0.056	0.040	0.052	0.083	0.097	0.070	0.092	0.269	0.270	0.234	0.287
0.8	0.055	0.060	0.047	0.059	0.105	0.103	0.081	0.106	0.334	0.293	0.255	0.339
0.9	0.076	0.061	0.052	0.065	0.151	0.110	0.092	0.125	0.442	0.319	0.282	0.430
0.95	0.109	0.062	0.056	0.072	0.223	0.112	0.097	0.151	0.600	0.345	0.307	0.608

Tableau 26 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur.

	N = 100      n = 10				N = 60      n = 9				N = 20      n = 4			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.002	-0.001	0.004	0.003	0.006	-0.004	0.014	0.011	0.001	-0.054	0.056	0.044
0.1	-0.002	0.004	0.001	0.000	-0.004	0.009	0.010	0.008	-0.014	-0.048	0.037	0.013
0.2	-0.002	0.009	-0.001	0.002	-0.004	0.010	-0.001	0.002	0.001	-0.026	0.038	0.021
0.3	0.002	0.017	0.003	0.007	0.005	0.022	0.007	0.013	0.002	-0.016	0.028	0.022
0.4	0.003	0.019	0.002	0.010	0.011	0.030	0.010	0.020	0.026	-0.000	0.036	0.039
0.5	0.003	0.016	0.002	0.009	0.001	0.021	0.000	0.012	0.031	-0.004	0.030	0.032
0.6	0.008	0.019	0.005	0.015	0.011	0.023	0.005	0.019	0.047	-0.007	0.023	0.033
0.7	0.007	0.012	0.000	0.010	0.017	0.015	0.003	0.017	0.075	-0.006	0.025	0.039
0.8	0.013	0.010	0.004	0.011	0.033	0.014	0.008	0.020	0.131	-0.009	0.031	0.047
0.9	0.030	0.007	0.006	0.012	0.075	0.012	0.014	0.026	0.241	-0.007	0.044	0.155
0.95	0.062	0.006	0.008	0.013	0.145	0.009	0.017	0.033	0.426	-0.011	0.064	0.303

Tableau 27 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N=100).

N = 100	n = 5				n = 10				n = 15				n = 20			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.003	0.001	0.004	0.003	0.002	-0.001	0.004	0.003	0.004	-0.002	0.009	0.006	0.002	-0.006	0.012	0.006
0.1	-0.002	0.002	0.000	-0.001	-0.002	0.004	0.001	0.000	-0.001	0.009	0.003	0.002	-0.004	0.009	0.005	0.001
0.2	-0.002	0.003	-0.002	-0.001	-0.002	0.009	-0.001	0.002	-0.001	0.016	-0.000	0.004	-0.000	0.023	0.003	0.008
0.3	0.002	0.009	0.002	0.004	0.002	0.017	0.003	0.007	0.002	0.026	0.004	0.010	0.001	0.035	0.004	0.013
0.4	0.003	0.011	0.002	0.006	0.003	0.019	0.002	0.010	0.007	0.030	0.004	0.016	0.003	0.039	0.002	0.017
0.5	0.000	0.008	0.001	0.004	0.003	0.016	0.002	0.009	0.000	0.028	0.001	0.013	0.005	0.037	0.002	0.020
0.6	0.006	0.011	0.004	0.009	0.008	0.019	0.005	0.015	0.010	0.029	0.006	0.021	0.012	0.038	0.005	0.026
0.7	0.005	0.006	0.001	0.005	0.007	0.012	0.000	0.010	0.010	0.018	0.001	0.016	0.010	0.025	-0.000	0.020
0.8	0.008	0.007	0.004	0.008	0.013	0.010	0.004	0.011	0.017	0.014	0.003	0.015	0.024	0.018	0.003	0.021
0.9	0.019	0.007	0.006	0.009	0.030	0.007	0.006	0.012	0.042	0.008	0.006	0.014	0.056	0.008	0.005	0.017
0.95	0.038	0.008	0.008	0.012	0.062	0.006	0.008	0.013	0.091	0.005	0.008	0.017	0.122	0.003	0.007	0.022

Tableau 28 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N=80).

N = 80	n = 4				n = 8				n = 12				n = 16			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.003	0.001	0.005	0.004	0.002	-0.002	0.005	0.004	-0.003	0.004	0.010	0.007	0.004	-0.007	0.016	0.010
0.1	0.001	0.004	0.003	0.002	0.001	0.006	0.005	0.003	0.010	0.002	0.007	0.005	-0.002	0.009	0.010	0.003
0.2	-0.004	0.002	-0.003	-0.002	-0.004	0.006	-0.003	-0.001	0.014	-0.002	-0.000	0.004	-0.001	0.019	0.003	0.007
0.3	0.004	0.011	0.004	0.007	0.005	0.018	0.005	0.010	0.026	0.005	0.006	0.012	0.004	0.035	0.007	0.016
0.4	0.005	0.012	0.004	0.008	0.005	0.020	0.004	0.011	0.031	0.010	0.007	0.019	0.006	0.038	0.006	0.020
0.5	0.002	0.009	0.002	0.005	0.004	0.017	0.003	0.010	0.026	0.012	0.002	0.014	0.007	0.036	0.004	0.022
0.6	0.005	0.009	0.003	0.008	0.007	0.016	0.003	0.013	0.025	0.010	0.004	0.019	0.011	0.033	0.003	0.024
0.7	0.005	0.005	0.001	0.006	0.008	0.010	0.001	0.010	0.016	0.013	0.001	0.016	0.013	0.022	-0.001	0.019
0.8	0.012	0.009	0.007	0.070	0.018	0.012	0.007	0.014	0.015	0.024	0.006	0.018	0.033	0.018	0.006	0.024
0.9	0.025	0.009	0.010	0.013	0.039	0.009	0.009	0.015	0.009	0.053	0.009	0.018	0.069	0.008	0.008	0.020
0.95	0.048	0.010	0.012	0.015	0.078	0.008	0.010	0.018	0.006	0.112	0.010	0.022	0.148	0.003	0.009	0.029

Tableau 29 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N=60).

N = 60	n = 3				n = 6				n = 9				n = 12			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.004	0.002	0.007	0.005	0.003	-0.003	0.007	0.005	0.006	-0.004	0.014	0.011	0.004	-0.011	0.022	0.012
0.1	0.002	0.005	0.004	0.003	0.001	0.003	0.005	0.003	0.004	0.009	0.010	0.008	-0.004	0.003	0.010	0.003
0.2	-0.004	0.001	-0.003	-0.002	-0.004	0.005	-0.002	-0.000	-0.004	0.010	-0.001	0.002	-0.001	0.015	0.005	0.008
0.3	0.004	0.010	0.005	0.007	0.004	0.016	0.005	0.009	0.005	0.022	0.007	0.013	0.003	0.029	0.007	0.015
0.4	0.006	0.013	0.006	0.009	0.007	0.020	0.007	0.013	0.011	0.030	0.010	0.020	0.009	0.035	0.008	0.022
0.5	0.001	0.006	0.001	0.004	0.004	0.014	0.001	0.009	0.001	0.021	0.000	0.012	0.008	0.031	0.004	0.021
0.6	0.008	0.009	0.004	0.009	0.011	0.014	0.004	0.014	0.011	0.023	0.005	0.019	0.015	0.029	0.004	0.024
0.7	0.009	0.006	0.003	0.008	0.013	0.011	0.003	0.013	0.017	0.015	0.003	0.017	0.022	0.019	0.002	0.022
0.8	0.017	0.010	0.009	0.013	0.024	0.012	0.009	0.017	0.033	0.014	0.008	0.020	0.045	0.016	0.008	0.026
0.9	0.038	0.015	0.017	0.020	0.055	0.014	0.015	0.023	0.075	0.012	0.014	0.026	0.096	0.010	0.013	0.029
0.95	0.068	0.016	0.019	0.023	0.108	0.013	0.017	0.027	0.145	0.009	0.017	0.033	0.195	0.005	0.015	0.046



Tableau 30 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N=40).

N = 40	n = 2				n = 4				n = 6				n = 8			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.006	0.003	0.010	0.008	0.005	-0.005	0.012	0.009	0.007	-0.008	0.022	0.015	0.010	-0.014	0.040	0.024
0.1	0.004	0.004	0.007	0.005	0.003	0.000	0.009	0.007	0.006	0.006	0.018	0.013	-0.002	-0.007	0.019	0.009
0.2	-0.002	0.000	-0.001	-0.001	-0.001	0.002	0.002	0.003	-0.001	0.005	0.002	0.005	0.002	0.008	0.015	0.013
0.3	0.003	0.007	0.004	0.001	0.002	0.011	0.005	0.007	0.005	0.015	0.008	0.012	0.002	0.019	0.011	0.015
0.4	0.008	0.012	0.007	0.010	0.009	0.017	0.009	0.014	0.013	0.026	0.014	0.021	0.014	0.028	0.013	0.025
0.5	0.003	0.006	0.003	0.005	0.007	0.011	0.005	0.011	0.005	0.017	0.004	0.014	0.013	0.023	0.009	0.023
0.6	0.011	0.008	0.006	0.010	0.015	0.011	0.005	0.015	0.018	0.017	0.007	0.021	0.024	0.020	0.007	0.026
0.7	0.017	0.009	0.009	0.013	0.024	0.011	0.009	0.018	0.029	0.015	0.009	0.022	0.036	0.014	0.006	0.026
0.8	0.030	0.016	0.018	0.022	0.042	0.017	0.018	0.026	0.057	0.015	0.016	0.030	0.070	0.014	0.016	0.034
0.9	0.057	0.022	0.026	0.031	0.082	0.019	0.025	0.033	0.114	0.014	0.022	0.038	0.139	0.008	0.022	0.040
0.95	0.106	0.029	0.035	0.044	0.160	0.024	0.033	0.052	0.213	0.019	0.032	0.063	0.271	0.010	0.029	0.074

Tableau 31 : Biais résiduels induits par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N=20).

N = 20	n = 1				n = 2				n = 3				n = 4			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.001	-0.032	0.008	0.006	-0.001	-0.022	0.014	0.011	0.001	-0.032	0.039	0.023	0.001	-0.054	0.056	0.044
0.1	-0.010	-0.023	-0.002	-0.007	-0.010	-0.025	0.004	-0.003	-0.006	-0.023	0.027	0.010	-0.014	-0.048	0.037	0.013
0.2	-0.004	-0.016	-0.000	-0.001	-0.001	-0.010	0.008	0.005	-0.002	-0.016	0.015	0.010	0.001	-0.026	0.038	0.021
0.3	0.003	0.001	0.006	0.005	0.004	-0.002	0.010	0.009	0.010	0.001	0.025	0.019	0.002	-0.016	0.028	0.022
0.4	0.009	0.015	0.009	0.010	0.014	0.002	0.013	0.016	0.013	0.005	0.022	0.021	0.026	-0.000	0.036	0.039
0.5	0.013	0.002	0.014	0.012	0.019	0.004	0.019	0.018	0.019	0.002	0.019	0.021	0.031	-0.004	0.030	0.032
0.6	0.023	-0.001	0.014	0.017	0.034	0.001	0.015	0.023	0.034	-0.001	0.016	0.027	0.047	-0.007	0.023	0.033
0.7	0.036	0.005	0.027	0.028	0.050	0.011	0.027	0.032	0.065	0.005	0.027	0.036	0.075	-0.006	0.025	0.039
0.8	0.060	0.004	0.037	0.040	0.083	0.014	0.034	0.042	0.112	0.004	0.033	0.045	0.131	-0.009	0.031	0.047
0.9	0.106	0.012	0.056	0.060	0.146	0.024	0.050	*	0.203	0.012	0.047	*	0.241	-0.007	0.044	*
0.95	0.195	0.031	0.078	*	0.275	0.045	0.072	*	0.346	0.031	0.067	*	0.426	-0.011	0.064	*

\* Valeurs instables

Tableau 32 : Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N = 100).

N = 100	n = 5				n = 10				n = 15				n = 20			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.011	0.011	0.012	0.011	0.013	0.013	0.015	0.013	0.015	0.014	0.019	0.015	0.017	0.014	0.024	0.018
0.1	0.012	0.011	0.012	0.012	0.013	0.012	0.014	0.013	0.015	0.013	0.017	0.014	0.018	0.015	0.023	0.018
0.2	0.012	0.011	0.012	0.011	0.013	0.012	0.013	0.012	0.015	0.013	0.016	0.014	0.018	0.014	0.019	0.016
0.3	0.011	0.010	0.011	0.010	0.012	0.011	0.012	0.011	0.014	0.012	0.014	0.012	0.016	0.013	0.016	0.014
0.4	0.011	0.010	0.010	0.010	0.012	0.011	0.011	0.011	0.014	0.012	0.013	0.012	0.016	0.012	0.014	0.013
0.5	0.010	0.009	0.009	0.009	0.011	0.010	0.010	0.010	0.013	0.010	0.011	0.010	0.014	0.012	0.012	0.012
0.6	0.009	0.008	0.008	0.008	0.010	0.009	0.009	0.009	0.011	0.009	0.009	0.009	0.013	0.010	0.010	0.010
0.7	0.008	0.007	0.007	0.007	0.009	0.007	0.007	0.008	0.010	0.008	0.007	0.008	0.011	0.009	0.008	0.009
0.8	0.007	0.006	0.006	0.006	0.008	0.006	0.006	0.007	0.009	0.007	0.006	0.007	0.010	0.007	0.006	0.008
0.9	0.006	0.004	0.004	0.004	0.007	0.004	0.004	0.005	0.009	0.005	0.004	0.005	0.011	0.005	0.004	0.006
0.95	0.007	0.003	0.003	0.004	0.011	0.003	0.003	0.005	0.019	0.003	0.003	0.007	0.028	0.004	0.003	0.012

\* Valeurs instables

Tableau 33 : Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N = 80).

N = 80	n = 4				n = 8				n = 12				n = 16			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.014	0.014	0.015	0.015	0.016	0.016	0.019	0.017	0.019	0.017	0.024	0.020	0.022	0.018	0.031	0.023
0.1	0.015	0.014	0.015	0.014	0.017	0.015	0.018	0.016	0.018	0.017	0.021	0.018	0.023	0.019	0.030	0.023
0.2	0.015	0.014	0.015	0.014	0.016	0.015	0.017	0.015	0.019	0.016	0.020	0.018	0.022	0.018	0.025	0.020
0.3	0.014	0.013	0.014	0.013	0.016	0.014	0.016	0.015	0.018	0.015	0.018	0.016	0.021	0.017	0.021	0.018
0.4	0.014	0.013	0.013	0.013	0.015	0.014	0.014	0.014	0.018	0.015	0.017	0.016	0.021	0.016	0.018	0.017
0.5	0.012	0.012	0.012	0.012	0.014	0.013	0.013	0.013	0.017	0.014	0.014	0.014	0.019	0.015	0.016	0.015
0.6	0.011	0.010	0.010	0.010	0.013	0.011	0.011	0.011	0.015	0.012	0.012	0.012	0.017	0.013	0.012	0.013
0.7	0.010	0.009	0.009	0.009	0.012	0.010	0.009	0.010	0.014	0.011	0.010	0.011	0.015	0.011	0.010	0.012
0.8	0.009	0.007	0.007	0.008	0.010	0.008	0.007	0.008	0.012	0.009	0.008	0.009	0.014	0.010	0.008	0.011
0.9	0.008	0.005	0.005	0.006	0.010	0.006	0.005	0.007	0.013	0.006	0.005	0.008	0.016	0.007	0.005	0.010
0.95	0.011	0.004	0.004	0.005	0.017	0.004	0.004	0.007	0.026	0.005	0.004	0.011	0.038	0.006	0.004	0.021

\* Valeurs instables

Tableau 34 : Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N = 60).

N = 60	n = 3				n = 6				n = 9				n = 12			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.020	0.020	0.021	0.020	0.022	0.019	0.026	0.023	0.026	0.024	0.034	0.027	0.030	0.026	0.046	0.032
0.1	0.020	0.020	0.021	0.020	0.023	0.018	0.025	0.022	0.026	0.023	0.030	0.025	0.031	0.025	0.040	0.030
0.2	0.020	0.020	0.021	0.020	0.023	0.018	0.024	0.022	0.026	0.023	0.028	0.025	0.031	0.024	0.035	0.028
0.3	0.019	0.018	0.020	0.018	0.022	0.016	0.022	0.020	0.025	0.021	0.025	0.022	0.029	0.022	0.029	0.025
0.4	0.019	0.018	0.019	0.018	0.021	0.016	0.020	0.019	0.025	0.021	0.024	0.022	0.029	0.023	0.026	0.024
0.5	0.017	0.016	0.017	0.016	0.019	0.014	0.018	0.017	0.023	0.018	0.020	0.019	0.025	0.020	0.021	0.021
0.6	0.016	0.015	0.015	0.015	0.019	0.012	0.016	0.016	0.021	0.017	0.017	0.018	0.024	0.018	0.018	0.019
0.7	0.015	0.013	0.013	0.014	0.018	0.011	0.014	0.015	0.020	0.016	0.014	0.016	0.022	0.017	0.015	0.018
0.8	0.014	0.011	0.011	0.011	0.016	0.009	0.011	0.012	0.018	0.013	0.011	0.014	0.021	0.014	0.011	0.016
0.9	0.013	0.009	0.008	0.010	0.016	0.007	0.008	0.012	0.021	0.010	0.008	0.013	0.027	0.011	0.009	0.018
0.95	0.016	0.016	0.007	0.008	0.025	0.005	0.007	0.012	0.037	0.008	0.007	0.018	0.053	0.009	0.007	0.063

\* Valeurs instables

Tableau 35 : Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N = 40).

N = 40	n = 2				n = 4				n = 6				n = 8			
$\rho$	$r_0$	$r_1$	$r_D$	$r_c$	$r_0$	$r_1$	$r_D$	$r_c$	$r_0$	$r_1$	$r_D$	$r_c$	$r_0$	$r_1$	$r_D$	$r_c$
0	0.032	0.031	0.034	0.032	0.035	0.034	0.041	0.037	0.040	0.037	0.054	0.043	0.048	0.041	0.080	0.056
0.1	0.032	0.031	0.033	0.032	0.036	0.032	0.039	0.036	0.041	0.036	0.048	0.041	0.049	0.040	0.065	0.050
0.2	0.031	0.030	0.032	0.030	0.034	0.032	0.036	0.033	0.040	0.036	0.044	0.039	0.048	0.038	0.056	0.046
0.3	0.032	0.030	0.032	0.031	0.035	0.033	0.037	0.033	0.040	0.035	0.042	0.037	0.048	0.038	0.052	0.044
0.4	0.030	0.029	0.030	0.029	0.033	0.031	0.032	0.031	0.039	0.034	0.038	0.035	0.045	0.037	0.043	0.040
0.5	0.028	0.026	0.027	0.027	0.032	0.029	0.030	0.029	0.038	0.031	0.033	0.033	0.043	0.034	0.037	0.037
0.6	0.026	0.023	0.024	0.024	0.029	0.025	0.026	0.026	0.035	0.028	0.028	0.029	0.039	0.030	0.030	0.033
0.7	0.027	0.023	0.023	0.024	0.031	0.025	0.024	0.027	0.033	0.027	0.025	0.029	0.040	0.030	0.027	0.034
0.8	0.025	0.019	0.019	0.021	0.030	0.021	0.020	0.024	0.034	0.023	0.021	0.029	0.039	0.026	0.021	0.033
0.9	0.025	0.016	0.016	0.019	0.031	0.017	0.016	0.023	0.041	0.019	0.016	0.029	0.050	0.021	0.017	0.041
0.95	0.033	0.014	0.014	0.021	0.048	0.016	0.015	0.032	0.065	0.018	0.015	0.061	0.087	0.019	0.014	0.228

\* Valeurs instables

Tableau 36 : Variances résiduelles induites par différents estimateurs corrigés pour l'effet de longueur, puis débiaisés (N = 20).

N = 20	n = 1				n = 2				n = 3				n = 4			
$\rho$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$	$r_0$	$r_i$	$r_D$	$r_c$
0	0.076	0.074	0.081	0.078	0.084	0.082	0.102	0.094	0.091	0.087	0.141	0.173	0.112	0.096	0.181	0.229
0.1	0.077	0.077	0.084	0.079	0.085	0.083	0.102	0.092	0.099	0.089	0.135	0.111	0.117	0.100	0.177	0.152
0.2	0.072	0.070	0.077	0.072	0.080	0.077	0.091	0.083	0.095	0.085	0.116	0.100	0.110	0.094	0.156	0.133
0.3	0.074	0.072	0.077	0.074	0.085	0.079	0.091	0.086	0.094	0.085	0.108	0.099	0.114	0.094	0.137	0.139
0.4	0.075	0.073	0.077	0.075	0.084	0.077	0.087	0.085	0.101	0.089	0.107	0.102	0.119	0.096	0.134	0.136
0.5	0.072	0.068	0.072	0.071	0.082	0.075	0.082	0.081	0.098	0.083	0.094	0.095	0.115	0.093	0.113	0.123
0.6	0.069	0.064	0.066	0.070	0.081	0.069	0.072	0.083	0.094	0.076	0.080	0.107	0.108	0.086	0.094	0.136
0.7	0.071	0.061	0.063	0.068	0.083	0.067	0.069	0.080	0.096	0.074	0.074	0.097	0.109	0.084	0.082	0.167
0.8	0.073	0.059	0.061	0.097	0.088	0.064	0.064	0.073	0.104	0.073	0.070	0.242	0.117	0.081	0.074	*
0.9	0.077	0.050	0.052	0.109	0.097	0.056	0.055	*	0.112	0.063	0.058	*	0.134	0.073	0.063	*
0.95	0.100	0.047	0.048	*	0.122	0.052	0.051	*	0.138	0.059	0.052	*	0.172	0.065	0.054	*

\* Valeurs instables

#### 4. Biais introduits par la transformation Z de Fisher

##### 4.1 Rappel sur la transformation de Fisher

Quand la corrélation entre deux populations n'est pas nulle, la distribution des coefficients de corrélation n'est ni normale, ni même symétrique, ce qui rend impossible les tests d'hypothèse ( $\rho = \alpha$ ) ou l'établissement des intervalles de confiance de  $\rho$ .

FISHER [1921] a donné une solution à ce problème en transformant les valeurs échantillonnées  $r$  en une quantité  $Z$ , distribuée presque normalement avec un écart-type  $\sigma_Z$  pratiquement indépendant de la valeur de la corrélation de la population dont l'échantillon est tiré; la relation entre  $Z$  et  $r_f$  est donnée par:

$$Z = \frac{1}{2} [\ln (1 + r_f) - \ln (1 - r_f)] = \tanh^{-1} r_f$$

$$\text{ou } r_f = \tanh Z$$

##### 4.2 Application aux autocorrélations

Dans le cas d'une autocorrélation, on a affaire à une corrélation "intra-classe" ou "fraternelle", c'est-à-dire que la corrélation s'applique sur deux séries de mêmes caractéristiques; dans ce cas, Fisher a montré que l'écart-type de la distribution normale des  $Z$  vaut  $\frac{1}{\sqrt{m - 3/2}}$  où  $m$  est la longueur de l'échantillon ayant servi au calcul de  $Z$ . Cette propriété peut être utilisée à plusieurs fins. Pour notre problème nous l'utilisons pour estimer la valeur du  $r_f$  de la population.

Sur chaque segment  $k$  de  $m_k$  données ininterrompues, on calcule un coefficient d'autocorrélation  $r_k$ , éventuellement on corrige ce



coefficient pour l'effet de longueur de l'échantillon, puis on calcule les valeurs de  $z_k$  correspondantes qui sont combinées en un  $Z$  global pondéré conduisant à une valeur  $\rho$  de la population-mère selon la séquence d'opérations suivantes:

calcul de  $r_k$

$$z_k = \tanh^{-1} r_k$$

$$Z = \frac{\sum_{k=1}^k z_k \cdot (m_k - \frac{3}{2})}{(\sum_{k=1}^k m_k) - \frac{3}{2}}$$

$$\hat{r}_f = \tanh Z$$

Remarque 1 : une telle procédure n'utilise pas toute l'information qui nous est disponible à savoir que les échantillons ne sont séparés que par une valeur absente.

Remarque 2 : nous n'avons pas trouvé dans la littérature d'exemple où la correction de longueur de l'échantillon était utilisée en combinaison avec la transformation  $Z$  de Fisher.

• biais résiduel dans le cas de valeurs absentes équidistantes:

Dans nos notations, les équations précédentes s'écrivent:

$$N = (n + 1) p + n$$

$$p = \frac{N - n}{n + 1}$$

$$Z = \frac{\sum_{i=1}^{n+1} z \cdot \left( \frac{N-n}{n+1} - \frac{3}{2} \right)}{\left( \sum_{i=1}^{n+1} \frac{N-n}{n+1} \right) - \frac{3}{2}} = \frac{\left( \frac{N-n}{n+1} - \frac{3}{2} \right) \cdot \sum_{i=1}^{n+1} z}{N - n - \frac{3}{2}}$$

Fisher (1921) a montré que chaque évaluation de  $z$  comportait un biais par excès  $\Delta z = - \frac{r_f}{2 \left( \frac{N-n}{n+1} - 1 \right)}$  ce qui conduit à un biais par excès  $\Delta Z$ :

$$\Delta Z \approx - \frac{r_f}{2(N+1)} \frac{N - \frac{5}{2}n - \frac{3}{2}}{(N-2n-1) \left( N-n-\frac{3}{2} \right)}$$

$$\Delta Z = - \frac{r_f}{2N(n+1)} \left( 1 + \frac{n+2}{2N} \right)$$

$$\text{d'où } r = \tanh (Z + \Delta Z) = \frac{\tanh Z + \tanh \Delta Z}{1 + \tanh Z \tanh \Delta Z}$$

$$r \approx (\tanh Z + \tanh \Delta Z) (1 - \tanh Z \tanh \Delta Z)$$

$$r \approx \tanh Z + \tanh \Delta Z - \tanh Z^2 \tanh \Delta Z$$

$$\approx r_f + \tanh \Delta Z (1 - r_f^2)$$

Si on développe  $\tanh \Delta Z$  autour de 0:

$$\tanh x = x - \frac{x^3}{3} + \epsilon (x^3)$$

$$r \approx r_f = \frac{r_f}{2N(n+1)} \left( 1 + \frac{n+2}{2N} \right) (1 - r_f^2)$$

d'où le biais absolu:

$$BA = r - r_f = - \frac{r_f (1 - r_f^2)}{2N (n+1)} \left(1 + \frac{n+2}{2N}\right)$$

Ce biais a été calculé au tableau 37; on y note qu'il est très faible et qu'il peut être négligé devant les corrections de longueur.

#### 4.3 Simulations

Pour tester la puissance de la transformation de Fisher, nous utilisons nos 2000 échantillons de longueur 100 obtenus par la génération d'un processus Markovien d'ordre 1 et de coefficient d'auto-corrélation  $\rho = 0, 0.1 \dots 0.8, 0.9$  et  $0.95$ .

Dans cette série, on impose 5% et 10% de valeurs absentes équidistantes, ce qui crée des tronçons de longueurs approximatives 16 et 8 éléments.

Dans le cas où la correction de longueur a été appliquée, il se peut que certaines valeurs de  $r_k$  corrigées par l'équation [4] donnent des valeurs supérieures à 1 en valeur absolue. Pour éviter ce problème, nous avons imposé une contrainte aux valeurs de  $z$  des tronçons de façon à ce que les valeurs de  $r_k$  corrigées restent comprises entre  $-0.98$  et  $+0.98$ . Pour les grandes valeurs de  $\rho$ , le résultat des simulations implique donc moins de 2000 séries.

Au tableau 38, nous présentons les biais résiduels moyens  $\overline{p - E(r_f)}$  et les variances d'estimation obtenus pour des séries de longueur 100 comportant 5 et 10 valeurs absentes sur lesquelles on a effectué la transformation de Fisher simple (colonnes 1, 2, 5 et 6) et la transformation de Fisher avec correction de longueur sur les tronçons (colonnes 3, 4, 7 et 8). On y note que la correction de longueur effectuée sur les tronçons améliore de façon drastique l'efficacité de la transformation de Fisher, tant en restant toutefois nettement infé-

rieur à celle des estimateurs testés au chapitre 3; on doit noter aussi que pour les fortes valeurs de  $\rho$ , les biais résiduels et les variances d'estimation calculés sont sans doute trop optimistes étant des conséquences des contraintes imposées aux valeurs  $z$  des  $k$  tronçons.

Tableau 38 : Biases résiduels moyens et variances d'estimation des autocorrélations - transformation de Fisher, N = 100.

$\rho$	n = 5				n = 10			
	$\overline{\rho - E(r_f)} \text{ var } r_f$		$\overline{\rho - E(r_f)} \text{ var } r_f$		$\overline{\rho - E(r_f)} \text{ var } r_f$		$\overline{\rho - E(r_f)} \text{ var } r_f$	
0	0.067	0.009	0.001	0.019	0.135	0.009	0.000	0.064
0.1	0.088	0.009	-0.011	0.019	0.171	0.008	-0.070	0.058
0.2	0.106	0.010	-0.028	0.021	0.210	0.009	-0.119	0.058
0.3	0.130	0.009	-0.039	0.020	0.255	0.009	-0.143	0.048
0.4	0.156	0.009	-0.048	0.021	0.303	0.009	-0.159	0.037
0.5	0.175	0.009	-0.069	0.019	0.347	0.008	-0.165	0.026
0.6	0.207	0.008	-0.068	0.017	0.401	0.009	-0.135	0.021
0.7	0.231	0.008	-0.068	0.012	0.448	0.008	-0.102	0.012
0.8	0.271	0.008	-0.035	0.009	0.510	0.008	-0.040	0.009
0.9	0.311	0.006	-0.013	0.005	0.571	0.007	0.027	0.006
0.95	0.336	0.006	-0.045	0.003	0.607	0.007	0.005	0.005
#	1	2	3	4	5	6	7	8

### 5. Utilisation des autocorrélations débiaisées comme un test d'indépendance

Plusieurs tests existent dans la littérature pour établir l'indépendance stochastique d'une série; la plupart de ces tests manquent de puissance quand on les applique à des séries courtes et surtout quand la valeur de  $\rho_1$  est faible. WALLIS et MATALAS (1971) ont effectué la revue des erreurs de type 2 produites par différents tests sur des échantillons générés par un processus Markovien d'ordre 1; le test de ANDERSON (1942) appliqué sur les valeurs corrigées s'avère plus puissant pour des échantillons courts que la plupart des tests classiques; ce test s'écrit:

$$IC(\hat{r}) = -1 \mp Z_{\alpha} (N-2)^{\frac{1}{2}} / (N-1)$$

où IC représente l'intervalle de confiance;  
 $\hat{r}$  est la valeur corrigée de l'autocorrélation d'ordre 1;  
 N est la longueur de l'échantillon  
 et  $Z_{\alpha}$  est la déviation standard normale correspondant à la probabilité  $\alpha$ .

si  $\alpha = 0.05$   $Z_{\alpha} = 1.645$

$\alpha = 0.01$   $Z_{\alpha} = 2.326$

Ce test a été appliqué aux valeurs calculées par nos estimateurs  $r_0$ ,  $r_i$  et  $r_D$ ,  $r_f$  et  $r_c$ , débiaisées et corrigées postérieurement pour l'effet de longueur (équation [4]). Les 2,000 échantillons ont été générés par un processus markovien d'ordre 1 et nous avons dénombré par le test précédant le pourcentage des échantillons soumis à une erreur de type 2, c'est-à-dire classés comme indépendants alors qu'ils proviennent d'une population-mère de paramètre  $\rho_1$ .

La perte de puissance du test due aux valeurs absentes estimées est mise en évidence au tableau 39, pour des échantillons de longueur  $N = 100$  et  $20$ , comprenant respectivement  $n = 20$  et  $4$  valeurs absentes; le cas classique  $n = 0$  est donné comme base de référence du test, on note que l'estimateur présentant la meilleure puissance face au test d'indépendance est l'estimateur  $r_i$ , ce qui était prévisible.

Ces résultats apparaissent aussi aux figures 4 et 5; on y note que les erreurs de classification peuvent être très importantes pour  $N$  petit et  $\rho_1$  grand; le meilleur estimateur, dans ce contexte est  $r_i$  si  $\rho_1$  est petit et  $r_D$  si  $\rho_1$  est grand.

Le tableau 40 montre que, pour des séries de longueur  $100$  et comportant  $5$  ou  $10$  valeurs absentes, les estimateurs  $r_C$  et  $r_i$  ont des performances à peu près équivalentes, mais nettement supérieures à celles obtenues en utilisant  $r_f$ , l'estimateur déduit de la transformation  $Z$  de FISHER.

Ces considérations recourent les résultats théoriques présentés à la figure 2.

En conclusion, sur l'ensemble des résultats, on peut dire que la perte de puissance induite par les différents estimateurs permet de les classer dans l'ordre décroissant d'efficacité suivant:

$r_i$ , suivi de  $r_C$ ,  $r_0$ ,  $r_D$  et enfin  $r_f$ .

A ce niveau, tous nos estimateurs sont plus puissants que la transformation  $Z$  de FISHER.

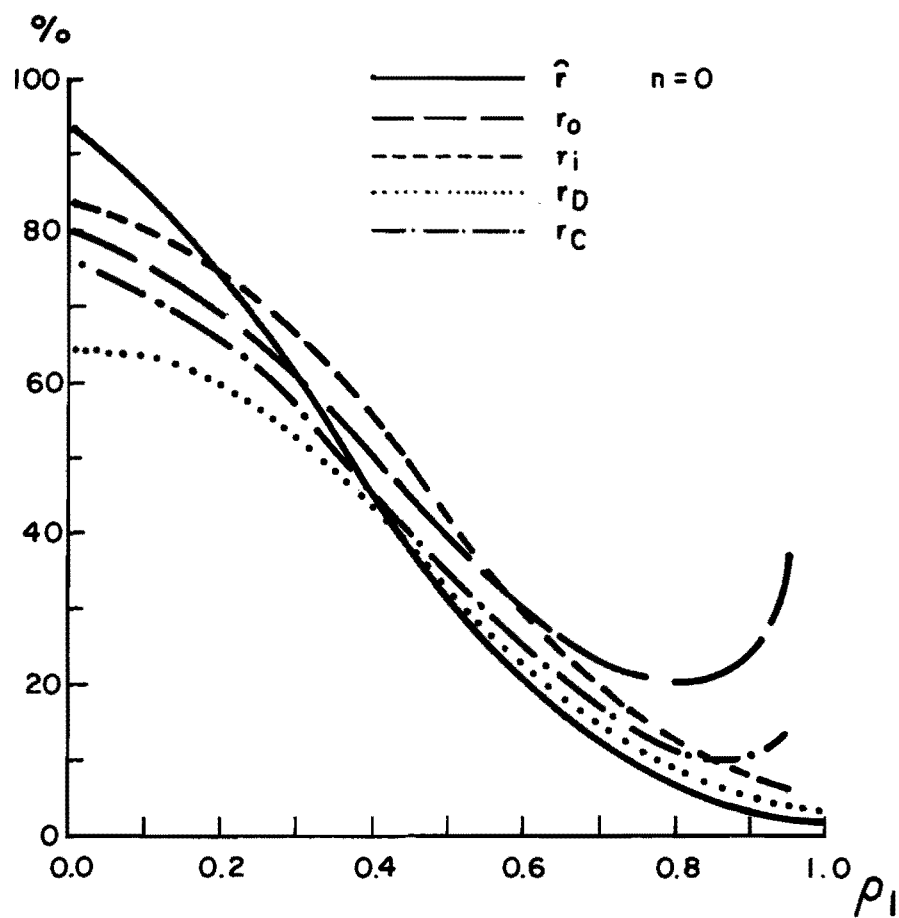


Figure 4 : Pourcentage des échantillons classés comme indépendants ( $\alpha = 0.05$ ) selon les estimateurs  $r$  ( $N = 20, n = 4$ ).



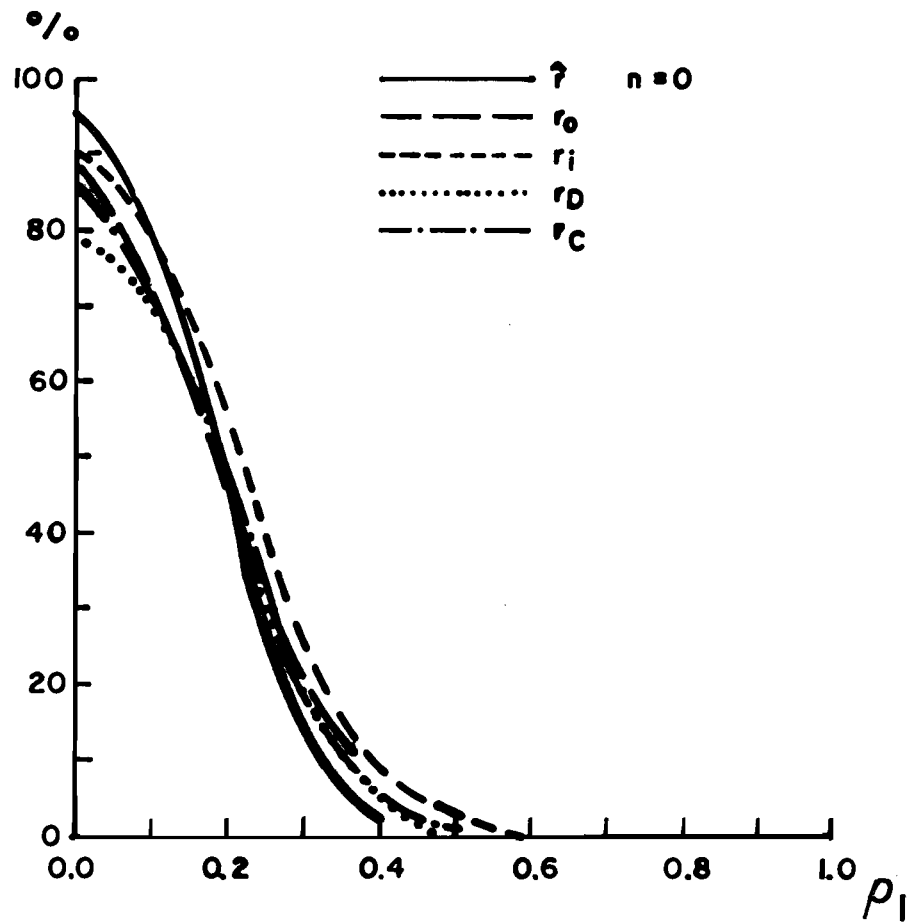


Figure 5 : Pourcentage des échantillons classés comme indépendants ( $\alpha = 0.05$ ) selon les estimateurs  $r$  ( $N = 100, n = 20$ ).

Tableau 39 : Puissance comparée des estimateurs débiaisés pour tester l'indépendance des séries ( $\alpha = 0.05$ ) dans les cas les plus critiques (20% de valeurs absentes).

$\rho_1$	N = 100					N = 20				
	n = 0	n = 20				n = 0	n = 4			
	r	$r_0$	$r_i$	$r_D$	$r_c$	r	$r_0$	$r_i$	$r_D$	$r_c$
0	95.1	87.6	90.2	79.3	86.7	90.7	80.3	84.9	64.8	79.3
0.1	79.0	71.7	79.9	69.7	72.2	84.8	75.5	80.3	63.8	75.5
0.2	43.3	45.5	57.1	44.9	45.8	75.3	70.4	76.6	60.2	71.6
0.3	12.7	19.8	27.5	19.2	19.1	61.5	60.4	67.8	53.2	62.1
0.4	1.9	5.1	7.8	5.1	5.0	45.0	51.0	56.1	42.1	51.7
0.5	0.1	1.0	1.4	0.8	0.9	33.3	40.7	42.8	32.4	41.0
0.6	0.0	0.2	0.2	0.1	0.2	19.4	30.6	31.2	22.7	30.0
0.7	0.0	0.0	0.0	0.0	0.0	12.5	23.8	19.7	14.6	21.7
0.8	0.0	0.0	0.0	0.0	0.0	6.8	22.1	13.0	9.3	17.1
0.9	0.0	0.0	0.0	0.0	0.0	3.4	23.5	8.0	5.3	14.3
0.95	0.0	0.0	0.0	0.0	0.2	3.0	37.7	6.0	4.4	17.9

t

Tableau 40 : Pourcentage des séries de Markov de longueur 100 classées indépendantes ( $\alpha = 0.05$ ) selon le nombre  $n$  de valeurs absentes et l'estimateur débiaisé utilisé.

$\rho_1$	$n = 0$	$n = 5$					$n = 10$				
	$r$	$r_0$	$r_i$	$r_D$	$r_f$	$r_c$	$r_0$	$r_i$	$r_D$	$r_f$	$r_c$
0	95.1	98.8	93.7	92.7	85.8	93.5	91.6	91.1	89.2	56.5	91.3
0.1	79.0	77.6	79.5	77.5	71.1	78.2	80.7	76.1	75.1	48.2	77.2
0.2	43.3	43.2	47.1	43.8	40.2	43.7	50.1	44.4	43.8	27.0	45.1
0.3	12.7	14.1	15.3	14.1	14.7	14.4	19.4	16.4	15.4	12.7	16.6
0.4	1.9	2.7	3.0	2.5	4.1	2.7	3.9	3.2	2.9	4.2	3.0
0.5	0.1	0.3	0.3	0.2	0.5	0.2	0.5	0.4	0.3	0.9	0.5
0.6	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.6	0.2
0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.95	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0

REFERENCES BIBLIOGRAPHIQUES

- ANDERSON, R.L. (1942).  
Distribution of the serial correlation coefficient. Ann. Math. Stat., B: 1-13.
- BOX, J.P. et G.M. JENKINS (1970).  
Time series analysis, forecasting and control, 553 p., Holden-Day, San Francisco, Calif.
- BOX, G.E.P. et M.E. MULLER (1958).  
A note on the generation of random normal deviates. Ann. Math. Statist., 29: 610-611.
- BRUBACHER, S.R. et G.T. WILSON (1976).  
Interpolating time series with applications to the estimation of holiday effects on electricity demand, Journal of the Royal Statistical Society, London, England, Series C (Applied Statistics), vol. 25, No. 2: 107-116.
- D'ASTOUS, F. et K.W. HIPEL (1979).  
Analysing environmental time series. ASCE, Env. Eng. Div., 105 (EES): 979-992.
- FISHER, R.A. (1921).  
On the probable error of a coefficient of correlation deduced from a small sample. Metron. 1(4): 3-32.
- JENKINS, M.G. et D.G. WATTS (1968).  
Spectral analysis and its applications, 525 p., Holden-Day. San Francisco, Calif.
- KENDALL, M.G. (1954).  
Note on bias in the estimation of autocorrelation. Biometrika, 41: 403-404.
- MARRIOTT, F.H.C. et J.A. POPE (1954).  
Bias in the estimation of autocorrelations, Biometrika, 42: 390-402.

ORCUTT, G.H. (1948).

A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919-1932. J.R. Statist. Soc. B.10: 1-54.

PARZEN, E. (1964).

An approach to empirical time series analysis, Radio-Science, 68(9): 937-951.

PREECE, D.A. (1971).

Iterative procedures for missing values in experiments. Technometrics, vol. 13, No. 4: 743-753.

RODRIGUEZ-ITURBE, I. (1971).

Structural analysis of hydrological sequences. Proceeding Warsaw Symposium. In mathematical models in hydrology, vol. 3: 1157-1165.

SASTRY, A.S.R. (1951).

Bias in estimation of serial coefficients. Sankhya, 11: 281-296.

SOPER, H.E., A.W. YOUNG, B.M. CAVE, A. LEE et K. PEARSON (1916).

On the distribution of the correlation coefficient in small samples - a cooperative study, Biometrika, 11: 328-413.

WALLIS, J.R. et MATALAS, N.C. (1971).

Correlogram analysis revisited. Wat. Res. Res., 7(6): 1448-1459.

WALLIS, J.R. et P.E. O'CONNELL (1972).

Small sample estimation of  $\rho_1$ . Water Resour. Res., 8(3): 707-712.

WILKINSON, G.N. (1958).

Estimation of missing values for the analysis of incomplete data. Biometrika, vol. 14, No 2: 257-286.