

Record Number:
Author, Monographic: Cluis, D.//Lebrun, S.//Laberge, C.
Author Role:
Title, Monographic: Correction analytique du biais de variance d'échantillons markoviens
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1984
Original Publication Date: Septembre 1984
Volume Identification:
Extent of Work: iii, 11
Packaging Method: pages
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, Rapport de recherche
Series Volume ID: 171
Location/URL:
ISBN: 2-89146-169-X
Notes: Rapport annuel 1984-1985
Abstract: 10.00\$
Call Number: R000171
Keywords: rapport/ ok/ dl

Correction analytique du biais de
variance d'échantillons markoviens

par

D. CLUIS, S. LEBRUN et C. LABERGE

Rapport scientifique No 171

Septembre 1984

Code	[R000028]
Rapport	[73-74]
Date	[1973]
Auteur	[Couillard, D.]
Et_al	[Lacasse+, J.P., Slivitzky, M.]
Titre	«nt-Laurent. - Tome I: aspects juridiques des conséquences du proj»
References	[INRS-Eau, rapport de recherche no 28, 64 p. - Rapport rédigé pour»

Please respond. (F10) when complete. (Alt)H for help.

Table des matières

	<u>Page</u>
Table des matières	i
Liste des tableaux	ii
Résumé et abstract	iii
1. Position du problème et études antérieures	1
1.1 Problématique	1
1.2 Méthodes utilisées dans la littérature	4
1.3 Méthode proposée	5
2. Résultats	7
2.1 Résultats obtenus en appliquant les 2 correc- tions analytiques décrites	7
2.2 Comparaison avec le facteur de correction de Tasker et Gilroy	7
2.3 Résultats en termes de biais sur la variance estimée	8
2.4 Résultats en termes de probabilité de sous- estimer	9
2.5 Résultats en termes d'écart quadratique moyen	11
3. Zones d'application	11
Bibliographie	12

Liste des tableaux

- Tableau 1: Facteur de correction $W(\rho, n)$ de Tasker et Gilroy
- Tableau 2: Nombre d'échantillons utilisés pour le calcul
- Tableau 3: Facteur de correction analytique $W'(\rho, n)$ obtenu
- Tableau 4: Variance du facteur de correction $W'(\rho, n)$ obtenu
- Tableau 5: Rapport $\frac{W'(\rho, n) - W(\rho, n)}{W(\rho, n)} \cdot 100$
- Tableau 6: Biais résiduel absolu $\hat{\sigma} - \sigma$
- Tableau 7: Probabilité de sous-estimer la variance (α^-)
- Tableau 8: Biais positif moyen (E_e^+)
- Tableau 9: Biais négatif moyen (E_e^-)
- Tableau 10: Écart quadratique moyen

Liste des figures

- Figure 1: Influence d'une mauvaise estimation de la variance d'un phénomène hydraulique sur l'optimum économique. (Cas I: sous-estimation; cas II: surestimation)
- Figure 2: Variance du facteur de correction $W'(\rho, n)$
- Figure 3: Délimitation de zones de biais résiduels selon N et ρ

Résumé

Ce rapport propose un estimateur de la variance d'une population markovienne d'ordre un à partir des paramètres d'un échantillon court. L'estimateur proposé est dérivé de l'application successive de deux corrections théoriques classiques et ses performances sont évaluées en termes de biais pour des échantillons de longueur et de persistance variables. L'intérêt de la correction est mis en évidence sur un exemple d'application économique concernant le dimensionnement d'un barrage-réservoir de protection contre les inondations.

Mots-clé / Descripteurs:

Estimateur de variance / inondation / barrage-réservoir / dimensionnement / population markovienne d'ordre un / échantillon court /

Abstract

This report proposes an estimator for the variance of a first-order markovian population inferred from the parameters of a single short sample. The suggested estimator is derived from the successive application of two classical and theoretical corrections and its performances are evaluated in terms of residual biases with regards to the variable lengths and persistences of the studied samples.

The practical use of this correction is enhanced by an example of economic application dealing with the design of a flood protection dam.

Keywords / Descriptors:

Variance estimator / flood protection dam / design / first-order markovian population / single short sample /

1. Position du problème et études antérieures

Dans les développements qui suivent, nous utiliserons les notations suivantes:

- μ, σ, ρ : moyenne, écart-type et coefficient de corrélation de la population
 m, s, r : moyenne, écart-type et coefficient de corrélation de l'échantillon de longueur N
 $\hat{\mu}, \hat{\sigma}, \hat{\rho}$: estimation de la moyenne de l'écart-type et du coefficient de corrélation de la population.

1.1 Problématique

Considérons une série constituée d'éléments $X_1, X_2 \dots X_N$ provenant d'un processus stationnaire persistant.

Les processus Markoviens s'avèrent d'une très grande importance lorsqu'il s'agit de modéliser des séries chronologiques en géophysique: en effet, ces modèles permettent de tenir compte de la plus grande partie des phénomènes de persistance que l'on retrouve dans la nature et c'est la représentation la plus simple, compatible avec la faible longueur de l'échantillon.

Lorsqu'on utilise de tels modèles, il est alors possible d'estimer les valeurs de la moyenne μ et de l'écart-type σ par l'utilisation des formules usuelles:

$$m = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

$$s = \left[\frac{1}{N-1} \sum_{i=1}^N (x_i - m)^2 \right]^{\frac{1}{2}} \quad (2)$$

On sait que même pour des échantillons courts non indépendants, la moyenne m de l'échantillon ainsi calculée est un estimateur non biaisé de μ , alors que la quantité s calculée par (2) est un estimateur biaisé de σ .

Étant donné le biais de cet estimateur, il devient beaucoup moins intéressant de l'utiliser pour des applications hydrologiques. Ainsi, les intervalles de confiance et les tests statistiques ($m \mp k\sigma$), associés aux risques assumés, sont intimement reliés à la valeur de la variance. Il sera alors d'un grand intérêt d'avoir une valeur non-biaisée de cette variance du phénomène hydrologique.

Voyons un exemple illustrant la nécessité d'avoir cet estimateur non-biaisé. Supposons qu'un gouvernement décide de construire un barrage. Supposons également que les deux seuls coûts à considérer soient: premièrement, un coût de construction qui ne dépend pas d'une composante aléatoire mais de la grandeur du barrage, et deuxièmement, les coûts des dommages pouvant subvenir, qui, eux, dépendent d'une composante stochastique en plus de dépendre de la grandeur du barrage (Watt et Wilson, 1979).

Il est évident que l'on devra estimer ces coûts des dommages possibles. Généralement, pour de tels phénomènes, on utilise une estimation par intervalle de confiance plutôt qu'une estimation ponctuelle; on diminue ainsi le risque d'avoir trop souvent des débordements.

Supposons maintenant que le coût de construction soit une fonction linéaire de la grandeur du barrage et que le coût anticipé, avec l'estimation par intervalle de confiance, soit d'une forme exponentielle décroissante par rapport à la grandeur du barrage. La figure 1 illustre graphiquement de telles hypothèses.

Voyons d'abord sur cette figure 1, l'effet qu'aurait une surestimation des coûts des dommages. Dans un tel cas le gouvernement aurait en main les courbes II, alors que le phénomène (en réalité) serait expliqué par les courbes I. On s'aperçoit que le gouvernement choisira comme coût total minimum le point D, mais construira alors un barrage beaucoup trop gros, car au point E, le barrage coûterait beaucoup moins cher et aurait les mêmes risques de dommages futurs.

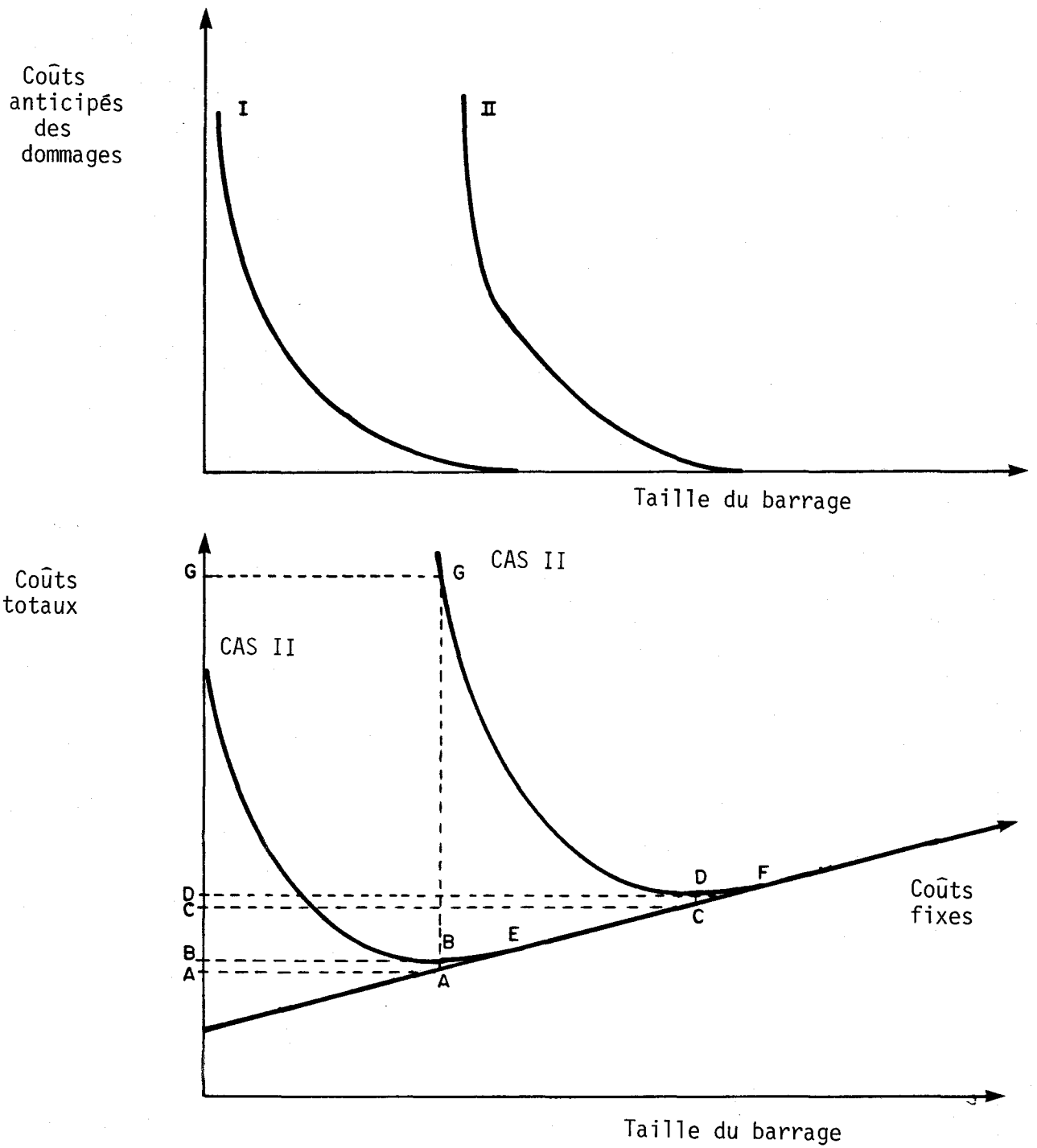


FIGURE 1 : Influence d'une mauvaise estimation de la variance d'un phénomène hydraulique sur l'optimum économique (cas I: sous-estimation; cas II: surestimation).

Cependant, une sous-estimation n'est guère plus intéressante. Si le gouvernement a en main les courbes I et que le phénomène hydraulique réagit plutôt selon les courbes II, il est facile de voir que dans ce cas, le coût en capital initial sera plus petit car le gouvernement construira un barrage au point B, au lieu du point D qu'il aurait dû atteindre en réalité, mais les coûts des dommages futurs seront beaucoup plus gros que prévus, soit (G-A) plutôt que (B-A).

Dans les deux cas, on s'aperçoit de l'importance économique d'une valeur non-biaisée de la variance du phénomène.

1.2 Méthodes utilisées dans la littérature

En toute généralité, un estimateur non biaisé de σ peut être écrit sous la forme:

$$\hat{\sigma} = W(\rho, N) \cdot s$$

où $W(\rho, N)$ est un facteur de correction, fonction de ρ et de N .

Voici les méthodes employées jusqu'ici pour débiaiser σ :

- Dixon et Massey (1957) donnent un estimateur non biaisé de σ pour des valeurs indépendantes ($\rho = 0$) sous l'hypothèse de normalité:

$$\hat{\sigma} = \left[1 + \frac{1}{4(N-1)} \right] \cdot s \quad (3)$$

- Pour leur part, Dawdy et Matalas (1964) donnent pour une population markovienne le facteur par lequel on doit multiplier s^2 pour obtenir une estimation non biaisée de σ^2 lorsque ρ est connu:

$$\hat{\sigma}^2 = \left[1 - \frac{1-\rho^2}{N(1-\rho)^2} + \frac{2\rho(1-\rho^N)}{N^2(1-\rho)^2} \right]^{-1} \cdot s^2 \quad (4)$$

- Enfin, Tasker et Gilroy (1982) nous donnent une table de correction numérique représentant les valeurs de $W(\rho, N)$ pour chaque combinaison de ρ et N .

Ce facteur de correction $W(\rho, N)$ a été obtenu par la méthode de Monte-Carlo appliquée à 1500 échantillons. Pour chaque échantillon tiré d'une population markovienne de paramètres ρ et N , l'écart-type s a été calculé par la formule (2) et la valeur de l'écart-type moyen \bar{s} a été déterminée. La valeur du facteur de correction de biais $W(\rho, N)$ pour chaque combinaison de ρ et N a été évaluée comme:

$$W(\rho, n) = \frac{\sigma}{E(s)} = \frac{1}{\bar{s}}$$

Cette correction numérique de Tasker et Gilroy (1982) est présentée au tableau 1.

1.3 Méthode proposée

Ce qui a motivé ce travail, c'est qu'il s'avère peu pratique d'avoir à se référer à une table à chaque fois que l'on doit estimer la variance d'une série de persistance markovienne, ce qui est une opération usuelle en hydrologie.

Nous proposons de remplacer la table de correction numérique donnée par Tasker et Gilroy (1982) par un estimateur analytique de σ dont nous évaluons par la suite les performances et la gamme d'applicabilité.

Dans le cas de séries à persistance markovienne, on connaît 2 corrections théoriques dans des cas particuliers:

D'une part, l'estimateur de variance de Dawdy et Matalas (1964) qui s'applique à un échantillon provenant d'une population markovienne dont l'autocorrelation ρ est connue (équation 4).

Tableau 1: Facteur de correction $W(\rho, n)$ de Tasker et Gilroy.

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	1.028	1.018	1.013	1.010	1.009	1.006	1.005	1.004	1.003	1.003
.10	1.040	1.026	1.019	1.015	1.013	1.010	1.007	1.005	1.004	1.004
.20	1.061	1.037	1.025	1.023	1.017	1.015	1.010	1.010	1.005	1.004
.30	1.079	1.051	1.036	1.029	1.021	1.019	1.015	1.012	1.010	1.008
.40	1.107	1.068	1.050	1.041	1.036	1.026	1.022	1.015	1.013	1.012
.50	1.147	1.096	1.069	1.059	1.045	1.038	1.028	1.026	1.018	1.013
.60	1.204	1.140	1.101	1.085	1.068	1.049	1.040	1.036	1.026	1.021
.70	1.304	1.200	1.158	1.123	1.103	1.077	1.063	1.049	1.038	1.032
.80	1.492	1.335	1.252	1.207	1.175	1.134	1.100	1.088	1.065	1.053
.90	1.945	1.688	1.537	1.434	1.375	1.283	1.229	1.191	1.144	1.117
.95	2.630	2.244	2.006	1.832	1.732	1.567	1.463	1.397	1.301	1.242

Tableau 2: Nombre d'échantillons utilisés pour le calcul.

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	1485	1500	1500	1500	1500	1500	1500	1500	1500	1500
.10	1468	1498	1500	1500	1500	1500	1500	1500	1500	1500
.20	1436	1500	1500	1500	1500	1500	1500	1500	1500	1500
.30	1419	1492	1498	1500	1500	1500	1500	1500	1500	1500
.40	1354	1484	1496	1500	1500	1500	1500	1500	1500	1500
.50	1299	1445	1490	1499	1500	1500	1500	1500	1500	1500
.60	1210	1390	1467	1489	1498	1500	1500	1500	1500	1500
.70	1108	1292	1411	1453	1483	1499	1500	1500	1500	1500
.80	1009	1156	1255	1360	1416	1471	1491	1499	1500	1500
.90	885	1012	1056	1137	1184	1304	1373	1421	1474	1495
.95	843	946	1009	1044	1073	1114	1164	1218	1307	1384

D'autre part, Wallis et O'Connell (1972) donnent un estimateur non biaisé de ρ en fonction de r et de N :

$$\hat{\rho} = \frac{r + 1/N}{1 - 4/N} \quad (5)$$

Dans cette formule, r est estimé par l'algorithme de Box et Jenkins (1970) et la correction de biais circulaire de Kendall (1954).

L'estimateur analytique proposé s'obtient en estimant pour chaque échantillon markovien le paramètre ρ de la population-mère dont il provient puis en introduisant cette valeur dans l'estimateur de variance de Dawdy et Matalas (1964).

On obtient:

$$\hat{\sigma}^2 = \left[1 - \frac{1 - \hat{\rho}^2}{N(1-\hat{\rho})^2} + \frac{2 \hat{\rho} (1-\hat{\rho}^N)}{N^2 (1-\hat{\rho})^2} \right]^{-1} \cdot s^2 \quad (6)$$

Notons $W'(\rho, N)$ le facteur par lequel on multiplie s pour obtenir $\hat{\sigma}$:

$$\hat{\sigma} = W'(\rho, N) \cdot s$$

Pour évaluer l'efficacité de cette estimation analytique de σ , nous avons utilisé la technique de Monte-Carlo en générant 1500 échantillons synthétiques de longueur 100, 80, 60, 50, 40, 30, 25, 20, 15 et 10 provenant de populations markoviennes, dont le paramètre ρ connu, vaut 0, 0.1, 0.2, ..., 0.8, 0.9 et 0.95. Ces séries sont générées en utilisant la formule récurrente:

$$x_i = \rho x_{i-1} + (1 - \rho^2)^{\frac{1}{2}} \varepsilon_i.$$

Dans cette formule, les ε_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).

Tableau 3: Facteur de correction analytique $W'(\rho, n)$ obtenu.

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	1.053	1.035	1.026	1.020	1.017	1.013	1.010	1.008	1.006	1.005
.10	1.059	1.041	1.031	1.024	1.020	1.015	1.012	1.010	1.008	1.006
.20	1.069	1.048	1.036	1.029	1.025	1.019	1.015	1.013	1.009	1.008
.30	1.076	1.055	1.043	1.035	1.029	1.023	1.018	1.015	1.011	1.009
.40	1.088	1.066	1.052	1.042	1.036	1.027	1.022	1.019	1.014	1.011
.50	1.099	1.077	1.062	1.052	1.044	1.035	1.028	1.024	1.018	1.015
.60	1.112	1.093	1.077	1.065	1.056	1.044	1.036	1.031	1.023	1.019
.70	1.128	1.111	1.094	1.083	1.073	1.058	1.049	1.042	1.032	1.026
.80	1.142	1.133	1.121	1.109	1.098	1.082	1.070	1.061	1.048	1.040
.90	1.158	1.159	1.154	1.146	1.140	1.124	1.113	1.102	1.086	1.074
.95	1.165	1.167	1.166	1.164	1.160	1.158	1.150	1.144	1.130	1.117

Tableau 4: Variance du facteur de correction $W'(\rho, n)$ obtenu.

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
.10	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000
.20	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
.30	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
.40	.003	.001	.001	.000	.000	.000	.000	.000	.000	.000
.50	.004	.002	.001	.001	.000	.000	.000	.000	.000	.000
.60	.004	.003	.002	.001	.001	.000	.000	.000	.000	.000
.70	.006	.004	.003	.002	.001	.001	.000	.000	.000	.000
.80	.006	.006	.005	.004	.003	.002	.001	.001	.000	.000
.90	.008	.009	.008	.007	.007	.005	.004	.003	.002	.001
.95	.009	.010	.009	.009	.009	.009	.007	.007	.005	.004

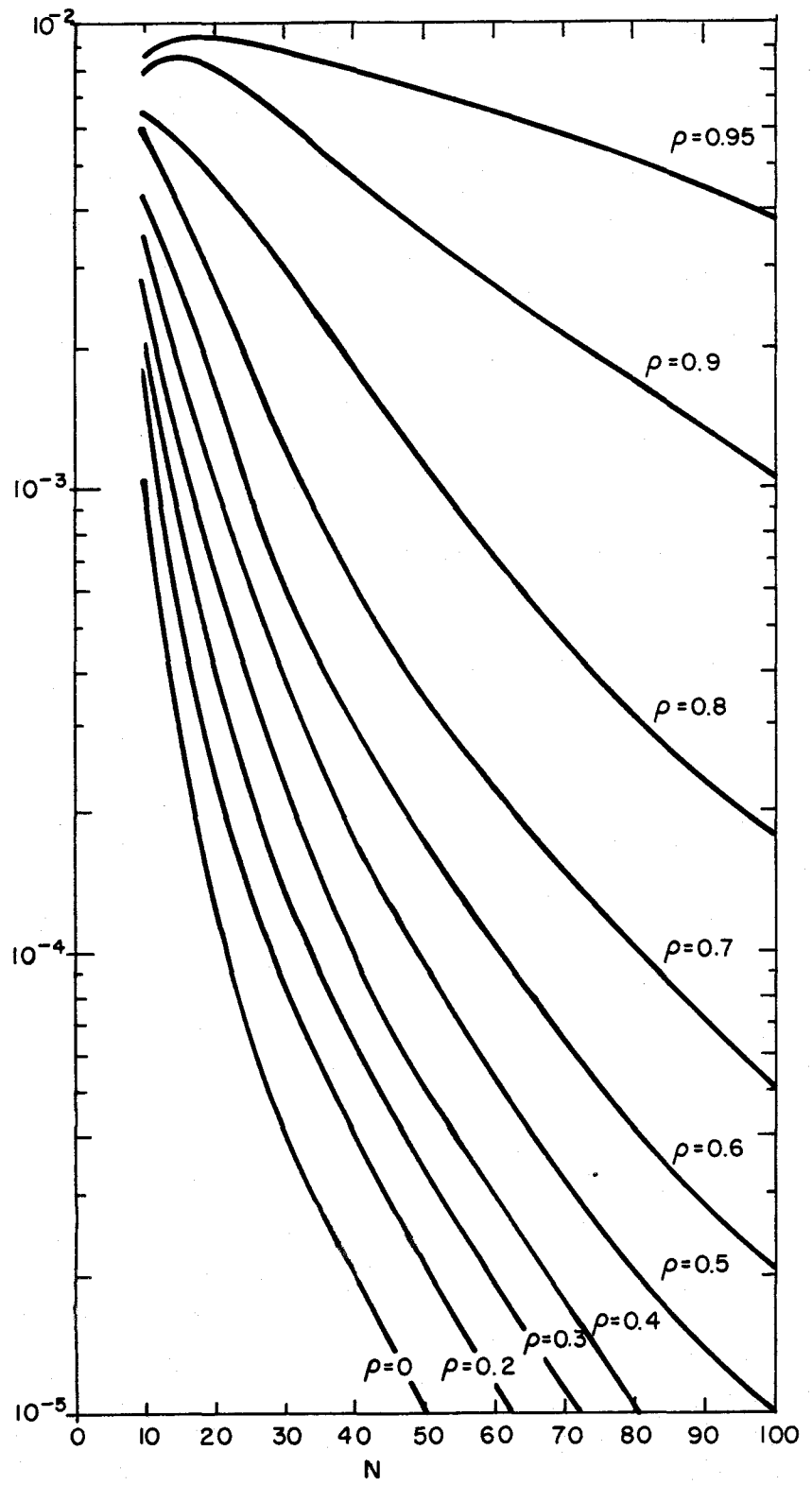


FIGURE 2: Variance du facteur de correction $W'(\rho, n)$.

2. Résultats

2.1 Résultats obtenus en appliquant les 2 corrections analytiques décrites

Le tableau 2 présente le nombre réel d'échantillons utilisés pour les calculs de l'estimateur $\hat{\sigma}$ de σ . Nous avons éliminé les échantillons qui conduisaient à un $\hat{\rho}$ supérieur à 1, car le premier terme de la formule (6) devenait négatif.

De ce fait, nous ne discuterons pas les résultats obtenus pour les ρ et N où il a fallu éliminer des échantillons car ils sont plus ou moins fortement faussés.

Sur chacun des tableaux suivants, un trait sera fait pour délimiter la zone que nous analyserons, i.e., celle où les 1500 échantillons sont considérés.

Au tableau 3, nous avons la table des facteurs de correction obtenus en appliquant successivement les 2 corrections analytiques. Nous comparerons ces résultats à la méthode de Tasker et Gilroy (1982) en 2.2.

Le tableau 4 représente la variance de ce facteur de correction $W'(\rho, N)$. La figure 2, quant à elle, nous donne l'allure graphique de cette variance, en fonction de N , et ce, pour chaque ρ étudié.

Puisque cette variance est faible dans la zone étudiée, nous jugerons donc pertinents les résultats obtenus.

2.2 Comparaison avec le facteur de correction de Tasker et Gilroy (1982)

Pour étudier l'efficacité de notre méthode, nous avons comparé les résultats obtenus avec le facteur de correction numérique de Tasker et Gilroy (1982).

Pour ce faire, nous avons établi l'erreur relative sur le facteur de correction obtenu par:

$$\frac{W'(\rho, N) - W(\rho, N)}{W(\rho, N)} \quad (7)$$

Les résultats de cette opération sont présentés au tableau 5.

On note d'abord que pour ρ fixe, le rapport décroît avec N . On y remarque que dans la zone étudiée, l'efficacité de notre estimateur est supérieur à 96% d'une correction qui reste proche de 1. Dans cette zone, il s'avèrera avantageux d'utiliser les 2 corrections analytiques successivement puisqu'elles évitent l'emploi de la table.

2.3 Résultats en termes de biais sur la variance estimée

Un critère important pour l'efficacité d'une méthode d'estimation de la variance est l'étude du biais résiduel.

Nous avons donc calculé le biais $(\hat{\sigma} - \sigma)$ absolu, ce qui équivaut à $\hat{\sigma} - 1$ car $\sigma = 1$ pour notre population générée et où $\hat{\sigma}$ a été obtenu par l'application de la méthode proposée.

Au tableau 6 nous présentons ces résultats.

Dans la zone étudiée, le biais résiduel est toujours inférieur à 4%, donc la méthode s'avère efficace.

Quant à la figure 3, elle nous permet de voir rapidement les points forts et les points faibles de notre correction analytique. Ainsi, on s'aperçoit qu'elle est très efficace pour des échantillons de 30 valeurs et plus ayant une persistance "moyenne", soit entre 0,2 et 0,6.

On voit, également, que plus la persistance est forte et plus il faut un échantillon de grande taille pour avoir un biais acceptable.

Tableau 5: Rapport $\frac{W'(\rho, n) - W(\rho, n)}{W(\rho, n)} \cdot 100$

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.024	.017	.012	.010	.008	.006	.005	.004	.003	.003
.10	.018	.014	.012	.009	.007	.005	.005	.006	.004	.003
.20	.007	-.010	.011	.006	.008	.004	.005	.002	.005	.003
.30	-.002	.004	.007	.006	.008	.004	.003	.003	.002	.002
.40	-.018	-.002	.002	.002	.000	.001	.000	.003	.002	-.000
.50	-.042	-.017	-.006	-.007	-.000	-.003	-.000	-.002	-.000	.001
.60	-.076	-.041	-.022	-.018	-.011	-.005	-.004	-.005	-.003	-.002
.70	-.135	-.074	-.055	-.035	-.027	-.017	-.013	-.007	-.005	-.005
.80	-.235	-.152	-.105	-.082	-.065	-.046	-.028	-.025	-.017	-.013
.90	-.405	-.313	-.249	-.201	-.171	-.124	-.094	-.075	-.051	-.039
.95	-.557	-.480	-.419	-.361	-.330	-.261	-.214	-.182	-.131	-.101

Tableau 6: Biais résiduel absolu $\hat{\sigma} - \sigma$

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.027	.023	.019	.013	.012	.006	.005	.003	.002	.001
.10	.027	.020	.018	.012	.009	.007	.008	.008	.008	.006
.20	.019	.012	.008	.004	.004	.003	.004	.001	.002	.001
.30	.007	.013	.011	.008	.004	.004	.003	-.000	-.001	.001
.40	-.011	.003	.006	.003	.004	.001	-.001	.000	-.002	-.001
.50	-.034	-.019	-.011	-.006	-.001	.002	.003	.003	.001	.001
.60	-.069	-.029	-.016	-.011	-.011	-.004	-.004	-.003	-.002	-.002
.70	-.116	-.064	-.049	-.029	-.024	-.016	-.009	-.008	-.007	-.005
.80	-.223	-.140	-.099	-.076	-.060	-.047	-.032	-.027	-.022	-.015
.90	-.391	-.289	-.225	-.182	-.152	-.112	-.085	-.072	-.047	-.034
.95	-.553	-.473	-.413	-.363	-.320	-.248	-.201	-.165	-.118	-.091

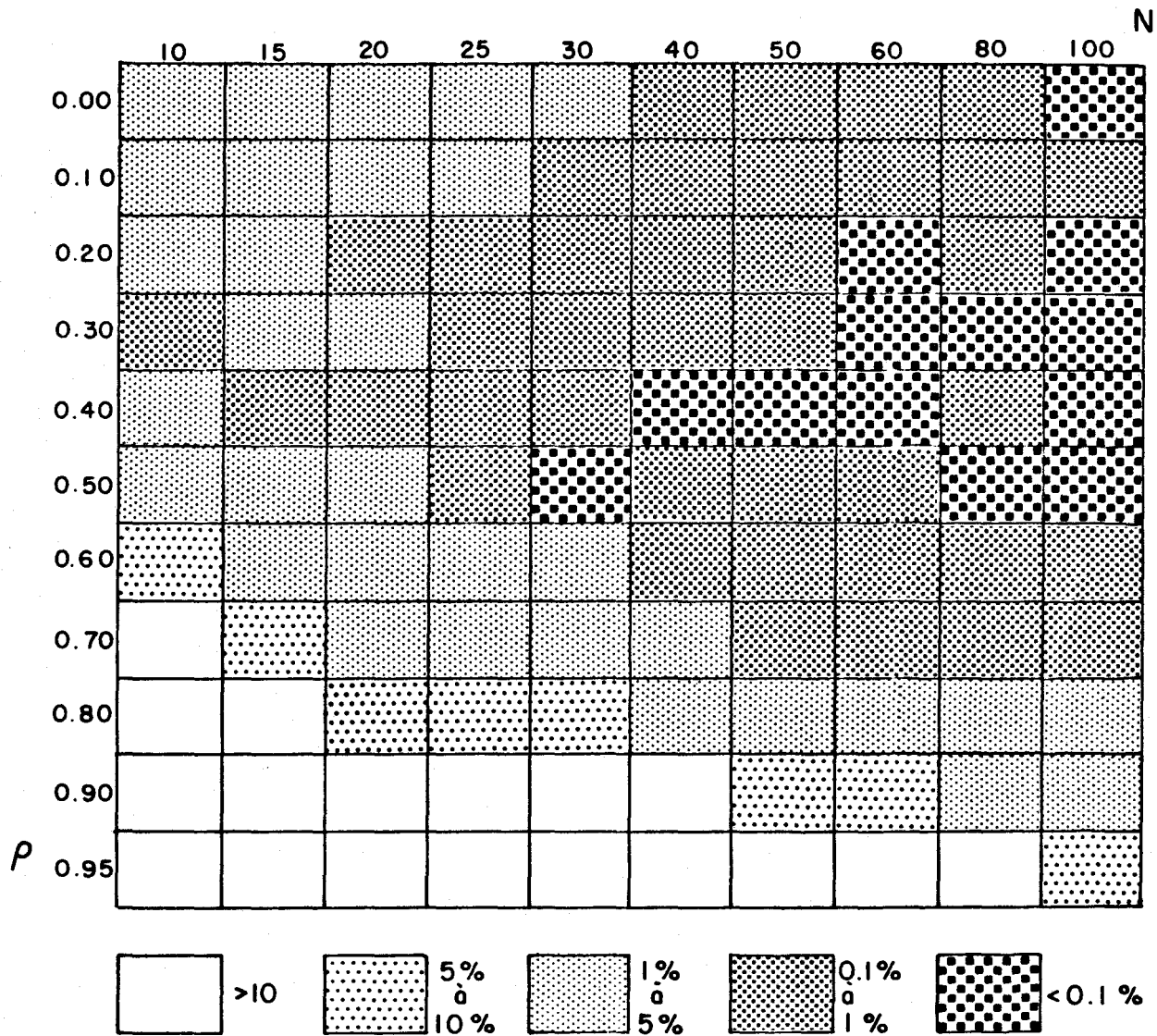


FIGURE 3: Délimitation des zones de biais résiduels selon N et ρ.

2.4 Résultats en termes de probabilité de sous-estimer

Dans les applications en hydrologie, lorsque l'on veut estimer les risques, les primes, les périodes de retour, la validité d'un test de Student, etc..., il est important de savoir si l'on surestime ou si l'on sous-estime la variance.

D'où l'intérêt de calculer la quantité:

$$\alpha^- = \frac{\text{nombre de fois } (\hat{\sigma} - \sigma) < 0}{1500}$$

qui représente la probabilité de sous-estimer la variance, lors des 1500 tirages.

Le tableau 7 présente les résultats obtenus par ce calcul.

Les tableaux 8 et 9 présentent:

$$E_e^+ = \frac{\sum_{K=1}^{1500} \max(0, e_k)}{1500 (1 - \alpha^-)}$$

et

$$E_e^- = \frac{- \sum_{K=1}^{1500} \min(0, e_k)}{1500 \alpha^-}$$

où $e = \hat{\sigma} - 1$.

C'est-à-dire que nous avons la moyenne des biais positifs et celle des biais négatifs.

Tableau 7: Probabilité de sous-estimer la variance (α^-).

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.469	.473	.470	.474	.471	.480	.499	.497	.505	.498
.10	.470	.471	.467	.471	.477	.483	.453	.466	.457	.476
.20	.498	.497	.491	.499	.495	.481	.485	.495	.501	.510
.30	.515	.501	.501	.509	.515	.510	.513	.518	.512	.521
.40	.561	.529	.519	.519	.495	.518	.508	.518	.521	.512
.50	.609	.572	.561	.555	.525	.525	.511	.503	.509	.505
.60	.643	.603	.584	.573	.571	.543	.537	.544	.531	.545
.70	.705	.665	.632	.603	.592	.575	.561	.559	.549	.540
.80	.795	.726	.683	.661	.652	.630	.609	.597	.594	.578
.90	.913	.845	.798	.764	.741	.709	.678	.681	.645	.623
.95	.978	.949	.919	.896	.865	.820	.787	.760	.716	.695

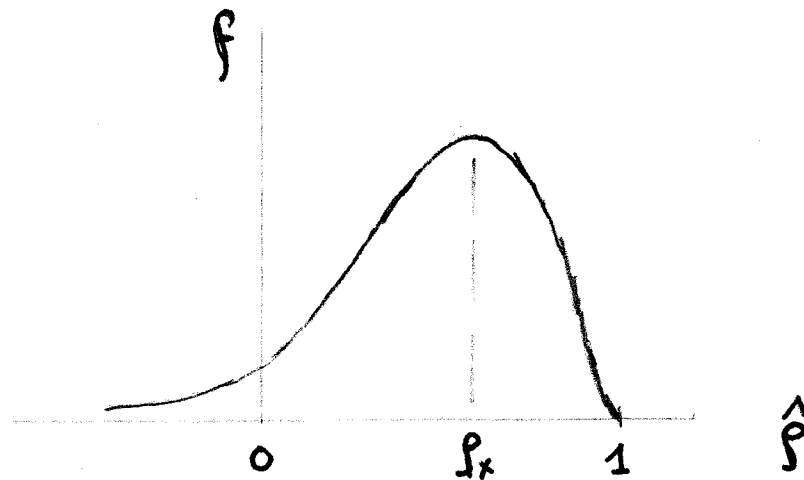
Tableau 8: Biais positif moyen (E_{e+}).

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.215	.176	.188	.126	.114	.096	.088	.079	.066	.057
.10	.216	.172	.144	.125	.113	.096	.081	.076	.066	.061
.20	.229	.176	.146	.128	.117	.096	.085	.078	.069	.061
.30	.227	.186	.157	.138	.126	.110	.097	.085	.072	.067
.40	.251	.203	.176	.151	.131	.111	.095	.089	.076	.066
.50	.262	.209	.184	.162	.145	.129	.114	.102	.084	.074
.60	.267	.239	.215	.191	.172	.143	.128	.119	.098	.090
.70	.284	.283	.238	.215	.196	.166	.149	.136	.115	.101
.80	.262	.271	.265	.248	.239	.204	.186	.167	.148	.131
.90	.214	.287	.288	.285	.284	.272	.258	.255	.221	.203
.95	.134	.203	.228	.256	.261	.283	.275	.276	.264	.254

Ce qui ressort de ces trois tableaux, c'est que l'on sous-estime plus que l'on surestime. Il y a donc une plus forte distribution des biais du côté négatif. Par contre, si l'on regarde la moyenne des biais qui sont positifs (tableau 8) et la moyenne des biais négatifs (tableau 9), on remarque que ces moyennes se ressemblent passablement. Nous avons donc une distribution presque symétrique au centre, mais beaucoup plus étendue vers la gauche.

Cette distribution spéciale est due au fait que le coefficient de corrélation est borné à 1. S'il n'y avait pas cette contrainte on pourrait s'attendre à ce que les coefficients estimés des échantillons aient une distribution proche de la distribution normale.

Cependant, la borne supérieure a pour effet de compresser cette distribution du côté droit, amenant un résultat de la forme suivante:



où ρ_x = coefficient de corrélation de la population

Ce phénomène de compression amène les résultats cités quant à la probabilité de sous-estimer, et il faudra en tenir compte au moment de la décision d'utiliser ou non la correction analytique, car on a vu les problèmes pouvant subvenir suite à une sous-estimation de la variance dans les applications hydrauliques.

Tableau 9: Biais négatif moyen (E_{e^-}).

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.185	.149	.125	.112	.103	.091	.078	.074	.060	.055
.10	.186	.151	.126	.116	.104	.089	.081	.070	.060	.055
.20	.192	.155	.137	.120	.110	.097	.083	.076	.065	.058
.30	.199	.158	.135	.119	.112	.097	.086	.080	.069	.060
.40	.216	.176	.152	.134	.126	.102	.094	.082	.073	.065
.50	.224	.190	.163	.141	.133	.113	.103	.095	.079	.071
.60	.256	.205	.181	.161	.149	.128	.118	.106	.091	.080
.70	.283	.239	.216	.190	.175	.151	.133	.122	.108	.095
.80	.349	.294	.268	.242	.220	.194	.171	.158	.138	.122
.90	.448	.395	.355	.326	.305	.269	.248	.225	.195	.177
.95	.568	.510	.470	.434	.411	.365	.330	.304	.269	.242

Tableau 10: Écart quadratique moyen.

$\rho \backslash N$	10	15	20	25	30	40	50	60	80	100
.00	.254	.204	.172	.151	.136	.117	.104	.095	.079	.071
.10	.252	.202	.168	.150	.135	.116	.103	.093	.081	.073
.20	.261	.208	.177	.156	.142	.122	.106	.097	.084	.074
.30	.269	.218	.186	.163	.150	.132	.116	.105	.090	.080
.40	.292	.236	.203	.178	.160	.135	.120	.108	.093	.082
.50	.301	.249	.218	.192	.176	.153	.137	.122	.101	.090
.60	.324	.273	.248	.220	.200	.168	.153	.139	.119	.106
.70	.344	.318	.277	.251	.232	.198	.176	.162	.139	.123
.80	.384	.347	.328	.301	.279	.249	.222	.203	.177	.157
.90	.473	.428	.394	.375	.361	.331	.310	.291	.255	.233
.95	.592	.531	.491	.459	.435	.403	.374	.358	.327	.306

2.5 Résultats en termes d'écart quadratique moyen

Le tableau 10 présente l'écart quadratique moyen (RMSE).

Ce que nous remarquons, c'est que l'écart quadratique moyen décroît lorsque N augmente, et croît lorsque ρ augmente. Nous retrouvons la plus petite erreur quadratique moyenne à $\rho = 0$ et $N = 100$ et le plus grand écart quadratique moyen à $\rho = 0.95$ et $N = 10$.

De plus, dans la zone jugée acceptable, l'écart quadratique moyen ne dépasse pas 0.26.

3. Zones d'application

L'utilisation de cette méthode a comme grand avantage d'éliminer l'emploi de tables pour débiaiser l'estimation d'une variance.

Nous avons vu qu'elle n'est pas utilisable pour toutes les valeurs de ρ et de N . Comme guide, nous savons que si l'expression $\rho^N < 0.013$ est vérifiée, alors le biais obtenu sera inférieur à 10%, et si $\rho^N < .001$, alors le biais obtenu sera inférieur à 5%.

Examinons le cas particulier des débits moyens annuels:

a) Pour le fleuve Saint-Laurent à Ogdenburg (N.Y) $\rho = 0.7$ et $N = 97$, on a $\rho^N = 0.7^{97} = 9,43 \times 10^{-16}$ ce qui est très largement inférieur à 0.001. Donc, le biais devrait être inférieur à 5%. Si on consulte le tableau 5, le biais pour $\rho = 0,7$ et $N = 100$ est effectivement de -0,005; donc très acceptable. La probabilité de sous-estimer (tableau 6) se situe aux environs de 54%.

b) Quant aux tributaires du fleuve Saint-Laurent, qui ont en général une persistance de l'ordre de 0,4, on aura besoin de 5 valeurs pour avoir un biais inférieur à 10% et 8 valeurs pour garantir un biais inférieur à 5%.

Bibliographie

- BOX, J.P. et G.M. JENKINS. (1970).
Time series analysis forecasting and control. 553 pp., 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.
- DAWDY, D.R. et MATALAS, N.C. (1964).
Analysis of variance, covariance, and time series. In: Handbook of Applied Hydrology, Chow, V.T., ed., McGraw-Hill Book Co., p. 8-68 - 8-90.
- DIXON, W.J. et F.J. MASSEY. (1957).
Introduction to Statistical Analysis. McGraw-Hill, New York, 638 p.
- KENDALL, M.G. (1954).
Note on the bias in the estimation of autocorrelation. Biometrika, 42: 403-404.
- TASKER, G.D. et E.J. GILROY. (1982).
Comparison of Estimators of Standard Deviation for Hydrologic Time Series. Water Res. Research, 18(5): 1503-1508.
- WALLIS, J.R. et P.E. O'CONNELL. (1972).
Small sample estimation of ρ_1 . Water Res. Research, 8(3): 707-712.
- WATT, W.E. et K.C. WILSON. (1979).
An approach to optimal design of hydraulic structures. In: Reliability in Water Resources Management, p. 75-90, Water Resources Publications, Fort Collins. Colorado.