

Record Number: 16800
Author, Monographic: Asselin, J.//Ouarda, T. B. M. J.//Fortin, V.//Bobée, B.
Author Role:
Title, Monographic: Une procédure bayésienne bivariée pour détecter un décalage de la moyenne
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1999
Original Publication Date: Mars 1999
Volume Identification:
Extent of Work: 44
Packaging Method: pages incluant 3 annexes
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, rapport de recherche
Series Volume ID: 528
Location/URL:
ISBN: 2-89146-305-6
Notes: Rapport annuel 1998-1999
Abstract: Chaire en hydrologie statistique Hydro-Québec / CRSNG
Call Number: R000528
Keywords: rapport/ dl/ ok

**UNE PROCÉDURE BAYÉSIENNE
BIVARIÉE POUR DÉTECTER UN
DÉCALAGE DE LA MOYENNE**

**UNE PROCÉDURE BAYÉSIENNE BIVARIÉE POUR
DÉTECTER UN DÉCALAGE DE LA MOYENNE**

par:

Jérôme Asselin

Taha B.M.J. Ouarda

Vincent Fortin

Bernard Bobée

Chaire en Hydrologie Statistique

Institut National de la Recherche Scientifique, INRS-Eau

2800 Einstein, C.P. 7500, Sainte-Foy (Québec) G1V 4C7

Rapport de recherche N° R-528

Mars 1999

ÉQUIPE DE RECHERCHE

**Chaire en Hydrologie Statistique
INRS-Eau**

Jérôme Asselin
Taha B.M.J. Ouarda
Vincent Fortin
Bernard Bobée

Table des matières

1	Introduction	1
2	Revue de littérature.....	3
3	Distributions a posteriori de τ et δ	5
	3.1 Distribution conjointe a posteriori de τ et δ	6
	3.2 Distributions marginales a posteriori.....	8
	3.3 Généralisation des distributions a priori.....	9
	3.4 Test de simulation	10
4	Application.....	15
5	Conclusion.....	21
6	Bibliographie.....	23
	Appendice A : Calcul de la densité conjointe de X et Y	25
	Appendice B : Résolution de l'intégrale (5).....	27
	Appendice C : Calcul de la fonction de distribution conjointe.....	35

1 Introduction

On a X_i et Y_i , $i = 1, 2, \dots, n$, une série de variables aléatoires observables à valeurs dans \mathbf{R} . On veut tester si la moyenne des Y_i a pu subir un changement systématique $\delta \in \mathbf{R}$ à un certain moment $\tau \in \{1, 2, \dots, n-1\}$ alors que les X_i sont uniformes dans le temps. On souhaite estimer les paramètres τ et δ en tenant compte d'une éventuelle corrélation entre X_i et Y_i pour $i = 1, 2, \dots, n$. Un tel scénario peut se retrouver en météorologie. Par exemple, on pourrait être intéressé à étudier les précipitations enregistrées par un pluviomètre en profitant des données de précipitations d'une station voisine, dans le but d'identifier la date de début d'un mauvais fonctionnement de l'appareillage. De même, en hydrologie, on peut être intéressé à étudier l'impact des changements climatiques sur les régimes hydrologiques. On pourrait alors étudier les débits moyens annuels d'une rivière en tenant compte des températures moyennes annuelles de la région.

La section 2 contient une revue de littérature exposant différents modèles qui ont été développés dans le passé pour la détection de changements systématiques. Dans la section 3, on décrit d'abord la structure de changement étudiée dans ce travail de recherche et on en énonce les hypothèses de départ et les distributions *a priori* des paramètres du modèle. Les sections 3.1 et 3.2 donnent les fonctions de distribution conjointe et marginales *a posteriori* pour les paramètres τ et δ . La section 3.3 présente une méthode pour généraliser les hypothèses sur les distributions *a priori*. La section 3.4 donne les résultats d'un test de simulation impliquant deux changements systématiques sur la moyenne. On retrouve une application des résultats théoriques de la section 3 dans la section 4. Les données utilisées sont les débits moyens annuels des rivières Moisie et Romaine dans la province de Québec, Canada.

2 Revue de littérature

Plusieurs chercheurs ont déjà étudié le problème de la détection de changements systématiques pour différentes structures de changement. Maronna et Yohai (1978) ont développé un test bivarié pour détecter de tels changements sous l'hypothèse que les (X_i, Y_i) suivent tous la même

loi normale bivariée. Le modèle qu'ils ont utilisé est $Y_i = \begin{cases} b + cX_i + U_i, & 1 \leq i \leq \tau \\ b + \delta + cX_i + U_i, & \tau < i \leq n \end{cases}$ où $b, \delta,$

c et τ sont des termes constants et les U_i sont les termes d'erreur aléatoire. Ce modèle sera considéré dans ce rapport. Lettenmaier et al. (1994) ont présenté les résultats d'une application du test de Maronna et Yohai (1978) en combinant des données météorologiques et hydrologiques. Dans le cas univarié du problème, une procédure bayésienne a été proposée par Lee et Heghinian

(1977). En prenant la même notation, leur modèle était $Y_i = \begin{cases} b + U_i, & 1 \leq i \leq \tau \\ b + \delta + U_i, & \tau < i \leq n \end{cases}$. Ouarda et

al. (1996, 1998) ont appliqué cette procédure sur les débits moyens annuels de plusieurs stations de jaugeage de la province de Québec, Canada, dans le but de détecter la date d'un éventuel changement climatique. Perreault et al. (1998) ont appliqué la méthode de Lee et Heghinian (1977) pour la détection d'un accroissement de précipitations annuelles. Bernier (1994) a généralisé la procédure de Lee et Heghinian (1977) en considérant différentes structures de changement de la moyenne dans le cas univarié. Solow (1988) a aussi développé un test bayésien

univarié sur le modèle de régression linéaire $Y_i = \begin{cases} a_0 + b_0 i + U_i, & 1 \leq i \leq \tau \\ a_1 + b_1 i + U_i, & \tau < i \leq n \end{cases}$ où a_0, a_1, b_0 et b_1

sont des termes constants *a priori* inconnus. Il a appliqué sa méthode sur un problème de détection des changements climatiques. Chin Choy et Broemeling (1980) ont présenté un test

bayésien multivarié avec pour modèle $Y_i = \begin{cases} \mathbf{X}_i \beta_1 + U_i, & 1 \leq i \leq \tau \\ \mathbf{X}_i \beta_2 + U_i, & \tau < i \leq n \end{cases}$ où les \mathbf{X}_i sont des vecteurs

$1 \times p$ de p variables indépendantes et β_1 et β_2 sont les vecteurs $p \times 1$ des coefficients

correspondants. Ce modèle fait varier tous les paramètres du modèle de régression entre les deux phases, contrairement à celui de Maronna et Yohai (1978) — utilisé dans ce rapport — dont la structure du changement entre les phases n'implique que l'ordonnée à l'origine. Les travaux de Ouarda et *al.* (1998) et de Perreault et *al.* (1998) contiennent plusieurs autres références sur l'ensemble des techniques qui ont été développées sur le sujet. Dans le présent rapport, on propose une approche similaire à celle de Lee et Heghinian dans le sens que les mêmes distributions *a priori* seront introduites dans le modèle et que les distributions *a posteriori* de τ et δ sont fournies. La différence majeure entre notre approche et celle de Lee et Heghinian (1977) consiste en l'ajout dans leur modèle d'une variable aléatoire observable X_i qui pourrait être corrélée avec la variable réponse Y_i . Le modèle ainsi obtenu avait déjà été introduit par Maronna et Yohai (1978).

3 Distributions a posteriori de τ et δ

On considère le modèle

$$Y_i = b_i + cX_i + U_i, \quad i = 1, 2, \dots, n, \quad (1)$$

où les X_i sont des variables aléatoires observables, indépendantes et identiquement distribuées (iid) selon une loi normale d'espérance μ et de variance σ^2 ($N(\mu, \sigma^2)$) et où les U_i sont des termes d'erreurs aléatoires iid suivant une loi $N(0, \sigma_u^2)$. Les X_i sont indépendantes des U_i . Les

b_i sont des paramètres tels que $b_i = b + a_i\delta$, $i = 1, 2, \dots, n$, où $a_i = \begin{cases} 0, & 1 \leq i \leq \tau \\ 1, & \tau < i \leq n \end{cases}$. Les paramètres

τ , δ , b , c , μ , σ^2 et σ_u^2 sont des constantes inconnues. On peut en déduire que les variables Y_i sont indépendantes avec loi $N(b + a_i\delta + c\mu, c^2\sigma^2 + \sigma_u^2)$. On peut aussi noter que le modèle (1)

implique une corrélation entre les X_i et les Y_i . En effet, on a $Cov(X_i, Y_i) = c\sigma^2$, $i = 1, 2, \dots, n$, ce qui implique que la matrice de variance-covariance du vecteur (X_i, Y_i) est

$\Sigma = \begin{pmatrix} \sigma^2 & c\sigma^2 \\ c\sigma^2 & c^2\sigma^2 + \sigma_u^2 \end{pmatrix}$ pour $i = 1, 2, \dots, n$. Ainsi, sous les hypothèses du modèle (1), on peut

montrer (voir Appendice A) que la fonction de densité conjointe de X_i et Y_i est celle d'une normale bivariée

$$f_{X,Y}(x_i, y_i) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \begin{pmatrix} x_i - \mu \\ y_i - b - a_i\delta - c\mu \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x_i - \mu \\ y_i - b - a_i\delta - c\mu \end{pmatrix}\right) \quad (2)$$

avec $|\Sigma| = \sigma^2\sigma_u^2$ et $\Sigma^{-1} = \frac{1}{\sigma^2\sigma_u^2} \begin{pmatrix} c^2\sigma^2 + \sigma_u^2 & -c\sigma^2 \\ -c\sigma^2 & \sigma^2 \end{pmatrix}$. En substituant $|\Sigma|$ et Σ^{-1} dans (2), on a

$$f_{X,Y}(x_i, y_i) = \frac{1}{2\pi\sigma\sigma_u} \exp\left(-\frac{1}{2} \left(\frac{(x_i - \mu)^2}{\sigma^2} + \frac{(y_i - b - a_i\delta - cx_i)^2}{\sigma_u^2} \right)\right). \quad (3)$$

On introduit maintenant le contexte bayésien du problème en proposant des distributions *a priori* pour les paramètres inconnus. Les distributions *a priori* des paramètres τ , δ , b et μ suivent Lee et Heghinian (1977). On a convenu que la distribution *a priori* de c est normale.

$\pi(\tau) = 1/(n-1)$, $\tau = 1, 2, \dots, n-1$, la distribution uniforme discrète ;

$\delta \sim N(0, \sigma_\delta^2)$ avec σ_δ^2 connue ;

$b \sim N(0, \sigma_b^2)$ avec σ_b^2 connue ;

$c \sim N(0, \sigma_c^2)$ avec σ_c^2 connue ;

$\pi(\mu)$ est arbitraire pour $\mu \in \mathbf{R}$;

$\pi(\sigma)$ est arbitraire pour $\sigma \in [\varepsilon, \infty[$, $\varepsilon > 0$;

$\pi(\sigma_u) \propto \sigma_u^{-1}$ avec $\sigma_u \in \mathbf{S} \subset [\varepsilon_u, \infty[$, $\varepsilon_u > 0$.

On considère les paramètres τ , δ , b , c , μ , σ et σ_u comme indépendants. On remarque que les distributions *a priori* pour les paramètres μ et σ sont plus générales donc moins précises que les autres. On pourra constater dans la prochaine section que leurs densités *a priori* n'ont aucune influence sur l'estimation *a posteriori* des autres paramètres.

3.1 Distribution conjointe *a posteriori* de τ et δ

On présentera ici la densité conjointe *a posteriori* de τ et δ , soit $\pi(\tau, \delta | \underline{x}, \underline{y})$ où \underline{x} et \underline{y} sont les vecteurs des observations x_i et y_i , $i = 1, 2, \dots, n$. La notation « \propto » indique la proportionnalité entre deux valeurs. Selon le théorème de Bayes, on a

$$\pi(\tau, \delta | \underline{x}, \underline{y}) \propto \pi(\underline{x}, \underline{y} | \tau, \delta) \pi(\tau, \delta) = \int \int \int \int \int \int \int \int \int \int \pi(\underline{x}, \underline{y}, b, c, \mu, \sigma, \sigma_u | \tau, \delta) \pi(\tau, \delta) db dc d\mu d\sigma d\sigma_u .$$

En raison de l'indépendance des paramètres,

$$\pi(\tau, \delta | \underline{x}, \underline{y}) \propto \int \int \int \int \int \int \int \int \int \int \pi(\underline{x}, \underline{y} | b, c, \mu, \sigma, \sigma_u, \tau, \delta) \pi(b) \pi(c) \pi(\mu) \pi(\sigma) \pi(\sigma_u) \pi(\tau) \pi(\delta) db dc d\mu d\sigma d\sigma_u \quad (4)$$

On note ici que

$$\begin{aligned} \pi(\underline{x}, \underline{y} | b, c, \mu, \sigma, \sigma_u, \tau, \delta) &= \prod_{i=1}^n f_{X,Y}(x_i, y_i) \\ &= \frac{1}{(2\pi)^n \sigma^n \sigma_u^n} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} + \frac{\sum_{i=1}^n (y_i - b - a_i \delta - c x_i)^2}{\sigma_u^2} \right)\right). \end{aligned}$$

En substituant les densités *a priori* dans (4) et en supprimant quelques termes constants en τ et δ de l'intégrale, on obtient

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \int_S \int_E \int_R \int_R \int_R \frac{\pi(\mu)\pi(\sigma)}{\sigma^n \sigma_u^{n-1}} \\ &\times \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} + \frac{\sum_{i=1}^n (y_i - b - a_i \delta - c x_i)^2}{\sigma_u^2} + \frac{b^2}{\sigma_b^2} + \frac{c^2}{\sigma_c^2} \right)\right) db da d\mu d\sigma d\sigma_u. \end{aligned} \quad (5)$$

Les détails de la résolution de cette intégrale sont présentés dans l'appendice B. On propose de calculer numériquement les intégrales dont la résolution analytique est trop compliquée. L'expression finale pour la densité conjointe *a posteriori* de τ et δ est définie pour $n \geq 2$.

$$\pi(\tau, \delta | \underline{x}, \underline{y}) = \frac{1}{\Omega} \int_S \frac{1}{\sigma_u^{n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} (T(\tau, \sigma_u)\delta^2 - 2B(\tau, \sigma_u)\delta + C(\tau, \sigma_u))\right) d\sigma_u \quad (6)$$

pour $\tau \in \{1, 2, \dots, n-1\}$ et $\delta \in \mathbf{R}$ où

$$\Omega = \sqrt{2\pi} \sum_{\tau=1}^{n-1} \int_S \frac{1}{\sigma_u^{n-2}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)T(\tau, \sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} \left(C(\tau, \sigma_u) - \frac{B^2(\tau, \sigma_u)}{T(\tau, \sigma_u)}\right)\right) d\sigma_u,$$

$$S(\sigma_u) = \frac{n\sigma_b^2}{\sigma_u^2 + n\sigma_b^2},$$

$$R(\sigma_u) = \frac{\sigma_u^2}{\sigma_c^2} + \sum_{i=1}^n x_i^2 - S(\sigma_u)n\bar{x}^2,$$

$$B(\tau, \sigma_u) = \left(\sum_{i=\tau+1}^n y_i - S(\sigma_u)(n-\tau)\bar{y} \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=1}^n x_i y_i - S(\sigma_u)n\bar{x}\bar{y} \right) \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u)(n-\tau)\bar{x} \right),$$

$$C(\tau, \sigma_u) = \left(\sum_{i=1}^n y_i^2 - S(\sigma_u) n \bar{y}^2 \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=1}^n x_i y_i - S(\sigma_u) n \bar{x} \bar{y} \right)^2$$

et

$$I(\tau, \sigma_u) = \frac{\sigma_u^2}{\sigma_\delta^2} + (n - \tau) \left(1 - S(\sigma_u) \left(\frac{n - \tau}{n} \right) \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n - \tau) \bar{x} \right)^2.$$

De même, pour $\tau_0 \in \{1, 2, \dots, n-1\}$ et $\delta_0 \in \mathbf{R}$, la fonction de distribution cumulative conjointe *a posteriori* est donnée par

$$P(\tau \leq \tau_0, \delta \leq \delta_0 | \underline{x}, \underline{y}) = \frac{\sqrt{2\pi}}{\Omega} \sum_{\tau=1}^{\tau_0} \int_s \frac{1}{\sigma_u^{n-2}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u) I(\tau, \sigma_u)}} \times \exp\left(-\frac{1}{2\sigma_u^2} \left(C(\tau, \sigma_u) - \frac{B^2(\tau, \sigma_u)}{I(\tau, \sigma_u)} \right)\right) \Phi\left(2\sqrt{2} \frac{\sqrt{I(\tau, \sigma_u)}}{\sigma_u} \left(\delta_0 - \frac{B(\tau, \sigma_u)}{I(\tau, \sigma_u)} \right)\right) d\sigma_u \quad (7)$$

où Φ est la fonction de distribution cumulative d'une loi normale standardisée $N(0,1)$, c'est-à-dire

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

3.2 Distributions marginales *a posteriori*

On peut déduire la fonction de distribution marginale *a posteriori* de τ à partir de la valeur de Ω . Le développement détaillé du facteur de réduction Ω est présenté dans l'appendice C.

$$\pi(\tau | \underline{x}, \underline{y}) = \frac{\sqrt{2\pi}}{\Omega} \int_s \frac{1}{\sigma_u^{n-2}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u) I(\tau, \sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} \left(C(\tau, \sigma_u) - \frac{B^2(\tau, \sigma_u)}{I(\tau, \sigma_u)} \right)\right) d\sigma_u \quad (8)$$

pour $\tau \in \{1, 2, \dots, n-1\}$.

De même, la densité marginale *a posteriori* de δ est obtenue en sommant sur τ la densité conjointe *a posteriori* $\pi(\tau, \delta | \underline{x}, \underline{y})$.

$$\pi(\delta | \underline{x}, \underline{y}) = \frac{1}{\Omega} \sum_{\tau=1}^{n-1} \int_{\mathcal{S}} \frac{1}{\sigma_u^{n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} (T(\tau, \sigma_u)\delta^2 - 2B(\tau, \sigma_u)\delta + C(\tau, \sigma_u))\right) d\sigma_u \quad (9)$$

On pourra visualiser dans l'exemple de simulation de la prochaine section que $\tau | \underline{x}, \underline{y}$ et $\delta | \underline{x}, \underline{y}$ ne sont pas indépendantes, c'est-à-dire qu'en général $\pi(\tau, \delta | \underline{x}, \underline{y}) \neq \pi(\tau | \underline{x}, \underline{y})\pi(\delta | \underline{x}, \underline{y})$.

3.3 Généralisation des distributions *a priori*

On a introduit précédemment le modèle (1) et les distributions *a priori* des paramètres. Toutefois, les distributions choisies ne permettent pas d'inclure l'information *a priori* sur les espérances des paramètres δ , b et c . C'est pourquoi on propose ici un modèle équivalent qui contourne ce problème. Soit $\tilde{\delta} = \delta + \mu_\delta$, $\tilde{b} = b + \mu_b$ et $\tilde{c} = c + \mu_c$ avec δ , b et c qui suivent les distributions *a priori* données précédemment et avec μ_δ , μ_b et μ_c qui sont des constantes réelles connues. On pose le modèle

$$Y_i = \tilde{b} + a_i \tilde{\delta} + \tilde{c} X_i + U_i, \quad i = 1, 2, \dots, n, \quad (10)$$

ce qui est équivalent à

$$Y_i - \mu_b - a_i \mu_\delta - \mu_c X_i = b + a_i \delta + c X_i + U_i, \quad i = 1, 2, \dots, n.$$

En posant $W_i := Y_i - \mu_b - a_i \mu_\delta - \mu_c X_i$, on obtient à nouveau le modèle initial (1)

$$W_i = b + a_i \delta + c X_i + U_i, \quad i = 1, 2, \dots, n. \quad (11)$$

On note toutefois que la variable W_i doit être mise à jour à chaque fois qu'on modifie la variable τ dans le cas où $\mu_\delta \neq 0$, car la fonction a_i dépend de τ . Cependant, ce modèle permet l'utilisation de tous les résultats des sections précédentes et les distributions *a posteriori* des paramètres de ce modèle sont à une translation près de celles des paramètres du modèle (10).

3.4 Test de simulation

On présente dans cette section un test de simulation pour étudier le comportement des distributions *a posteriori* lorsque deux changements systématiques ont eu lieu à deux moments différents. Ce test servira à souligner les avantages ainsi que les faiblesses de la procédure bayésienne bivariée proposée dans le présent rapport. Les données ont été simulées avec la fonction génératrice *normrnd* de la version 5.1 du logiciel *MATLAB* (*The Math Works Inc.*, 1997) sur une plate-forme compatible *IBM Pentium* 133 MHz avec 256 K de mémoire Cache et 32 Mega-octets de mémoire vive. Les programmes développés pour effectuer les tracés des distributions *a posteriori* sont disponibles auprès des auteurs. L'algorithme utilisé pour effectuer l'intégration numérique est la méthode récursive "Newton Cotes 8 panel" — fonction *quad8* de *MATLAB*.

Le modèle de simulation est donné par

$$Y_i = b' + a_{1,i}\delta'_1 + a_{2,i}\delta'_2 + c'X_i + U_i, \quad i = 1, 2, \dots, n$$

avec $a_{1,i} = \begin{cases} 0, & 1 \leq i \leq \tau_1 \\ 1, & \tau_1 < i \leq n \end{cases}$ et $a_{2,i} = \begin{cases} 0, & 1 \leq i \leq \tau_2 \\ 1, & \tau_2 < i \leq n \end{cases}$. Les valeurs des paramètres de simulation

sont

$$n = 60$$

$$\delta'_1 = 40$$

$$X_i \sim N(\mu = 100, \sigma^2 = 15^2)$$

$$\delta'_2 = -40$$

$$U_i \sim N(0, 20^2)$$

$$\tau_1 = 20$$

$$b' = -50$$

$$\tau_2 = 40$$

$$c' = 2$$

Les distributions *a priori* des paramètres du modèle (1)

$$Y_i = b_i + cX_i + U_i, \quad i = 1, 2, \dots, n$$

sont

$$\pi(\tau) = 1/(n-1), \quad \tau = 1, 2, \dots, n-1;$$

$$\delta \sim N(\mu_\delta = 0, \sigma_\delta^2 = 100^2) ;$$

$$b \sim N(\mu_b = -40, \sigma_b^2 = 40^2) ;$$

$$c \sim N(\mu_c = 3, \sigma_c^2 = 3^2) ;$$

$$\pi(\sigma_u) \propto \sigma_u^{-1} \text{ avec } \sigma_u \in \mathcal{S} = [5, 50].$$

Les distributions *a priori* de μ et σ sont arbitraires. Comme $\mu_b = -40$ et $\mu_c = 3$, on utilisera la généralisation décrite dans la section précédente. On pose $W_i := Y_i - \mu_b - \mu_c X_i$. Les figures 1 et 2 montrent les diagrammes de dispersion des couples de variables (X_i, Y_i) et (X_i, W_i) . Les données simulées sont présentées dans le tableau 1.

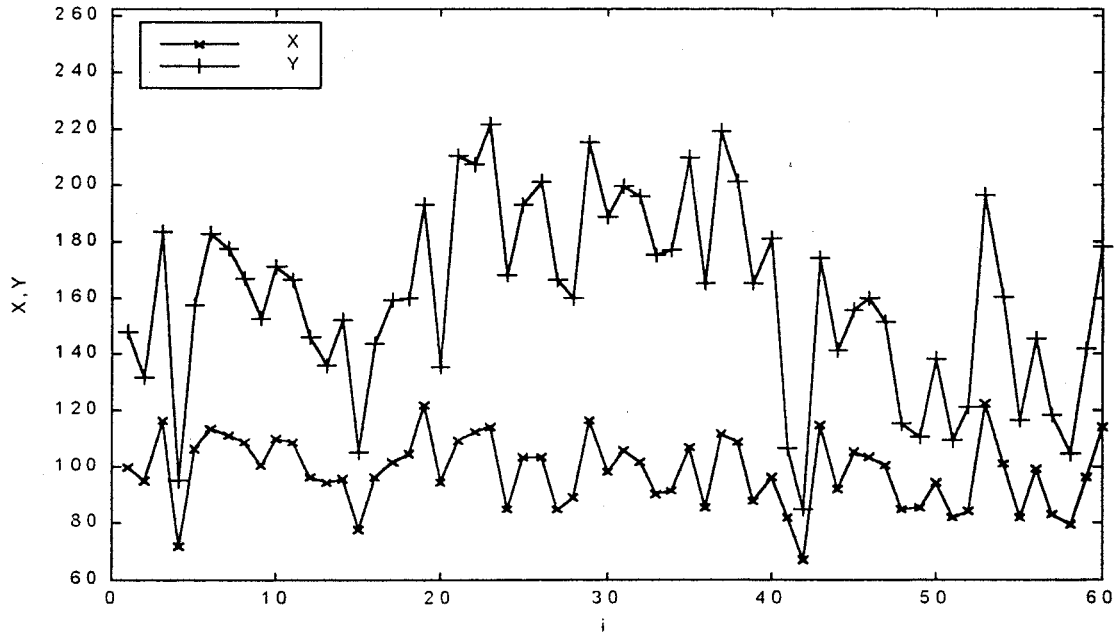


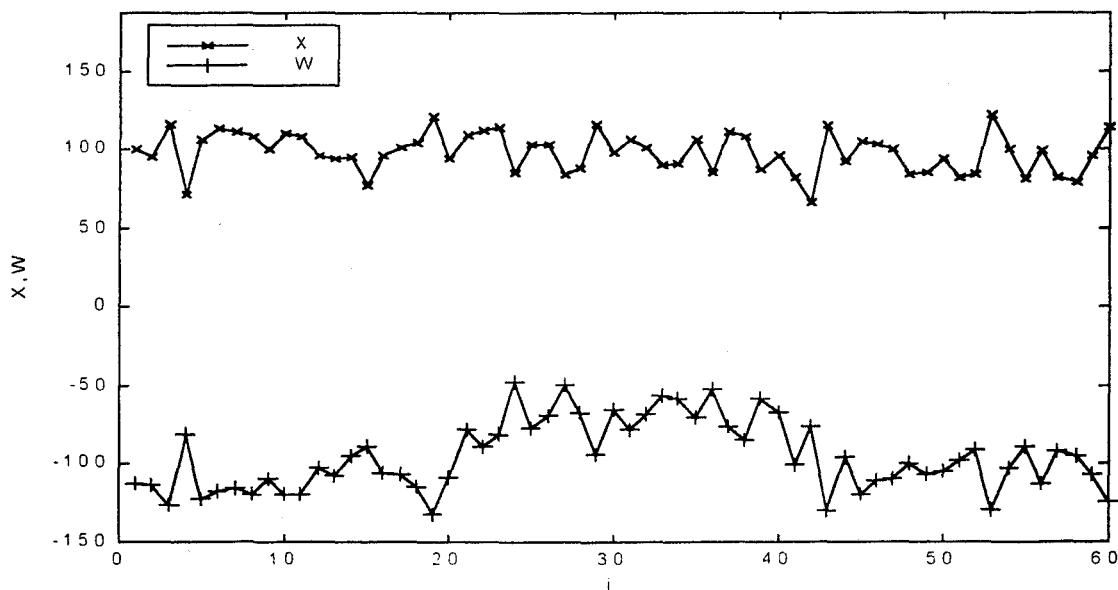
Figure 1 : Série chronologique des couples (X_i, Y_i) en fonction de i .

Tableau 1 : Données générées pour le test de simulation.**Variable X**

100.0006	95.2321	116.4251	71.8901	106.4227	113.4346
110.9644	108.6679	100.6047	110.1563	108.5335	96.1653
94.3380	95.5617	77.8730	96.4899	101.7767	104.7221
121.6526	94.7354	109.3485	111.9857	114.1133	85.1186
103.1805	103.5682	84.8835	88.8693	116.2344	98.0275
105.8482	101.3198	90.4680	91.6064	106.6548	85.7514
111.7177	108.5344	87.6743	96.0159	82.1833	66.9652
114.7951	92.2205	104.9105	103.5109	100.3220	84.9408
85.7928	94.3836	82.2117	84.1615	122.0872	100.8362
81.7402	99.3816	83.0748	79.7608	96.0835	114.3020

Variable Y

147.8385	132.1363	183.4768	95.2187	157.1131	182.8237
177.8745	167.1476	152.8459	171.1859	166.1335	145.9596
135.7343	152.0393	105.0640	143.5495	158.8872	159.7407
192.8270	135.3090	210.1691	207.2905	221.7983	168.3551
192.9022	201.4264	166.0371	159.7700	215.2640	188.9107
199.6970	196.0896	175.0142	176.7723	209.7609	164.8459
219.3896	201.0565	165.2496	181.2482	106.3463	85.2169
174.3078	141.5167	155.7956	159.6654	151.7406	115.2722
110.7322	138.4712	109.3702	121.3952	196.7131	160.1345
116.4369	145.5452	118.0514	104.4761	142.0694	178.3629

**Figure 2 : Série chronologique des couples (X_i, W_i) en fonction de i .**

Les figures 3 et 4 donnent les distributions conjointes *a posteriori* de τ et δ , et les distributions marginales de τ et δ respectivement. On y remarque que les deux décalages ont bien été détecté. Toutefois, les amplitudes des décalages sont sous-estimées par rapport à la réalité. Ceci peut être expliqué par le fait que le modèle (1) n'est pas conçu pour détecter deux décalages. Ainsi, chaque amplitude du décalage tient compte du reste de l'ensemble de données. La présence des deux cloches séparées sur la figure 3 donne un contre-exemple qui permet de rejeter l'hypothèse d'indépendance *a posteriori* des paramètres τ et δ . On souligne cependant que l'amplitude du décalage a été estimée correctement par la procédure dans le cadre d'une première simulation dans laquelle un seul décalage a été généré.

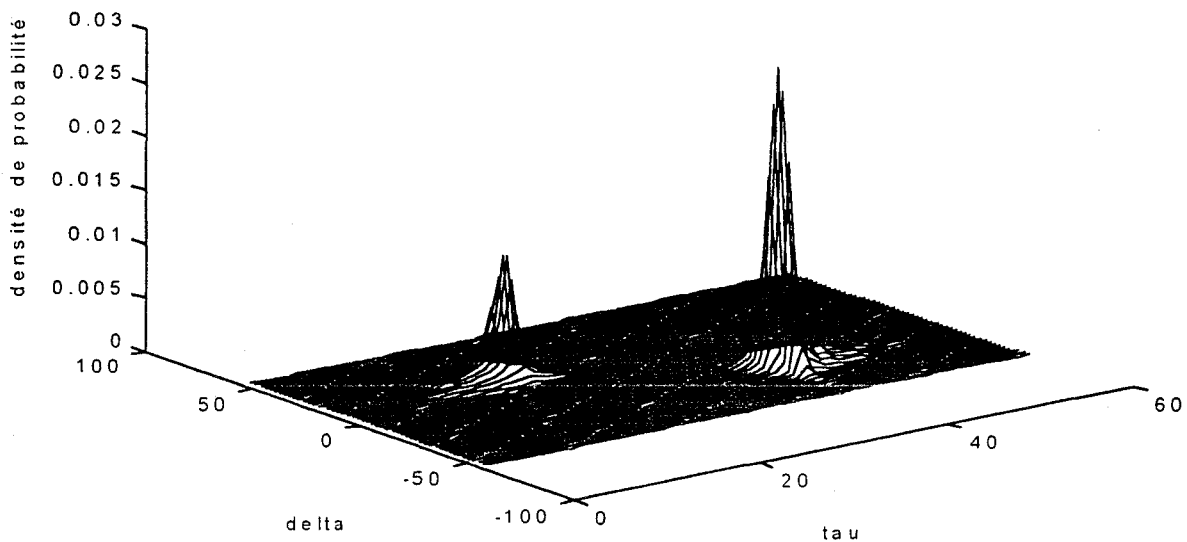


Figure 3 : Densité de probabilité conjointe *a posteriori* de τ (tau) et δ (delta).

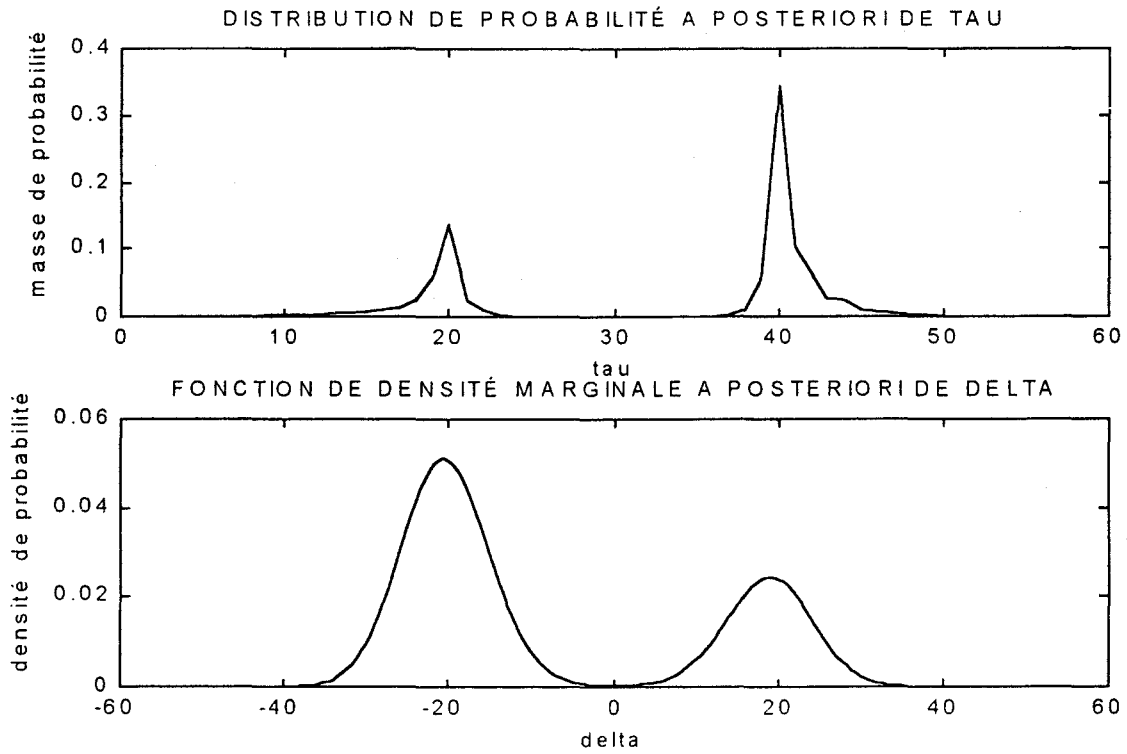


Figure 4 : Distribution marginales *a posteriori* de τ (tau) et δ (delta).

4 Application

Les rivières Moisie et Romaine sont situées sur la Basse Côte-Nord dans la province de Québec. Au milieu des années 80, on a observé une brusque diminution des débits de crues printanières sur ces rivières. La figure 5 montre la série chronologique des débits moyens de mai et de juin pour ces rivières. Les données numériques sont dans le tableau 2.

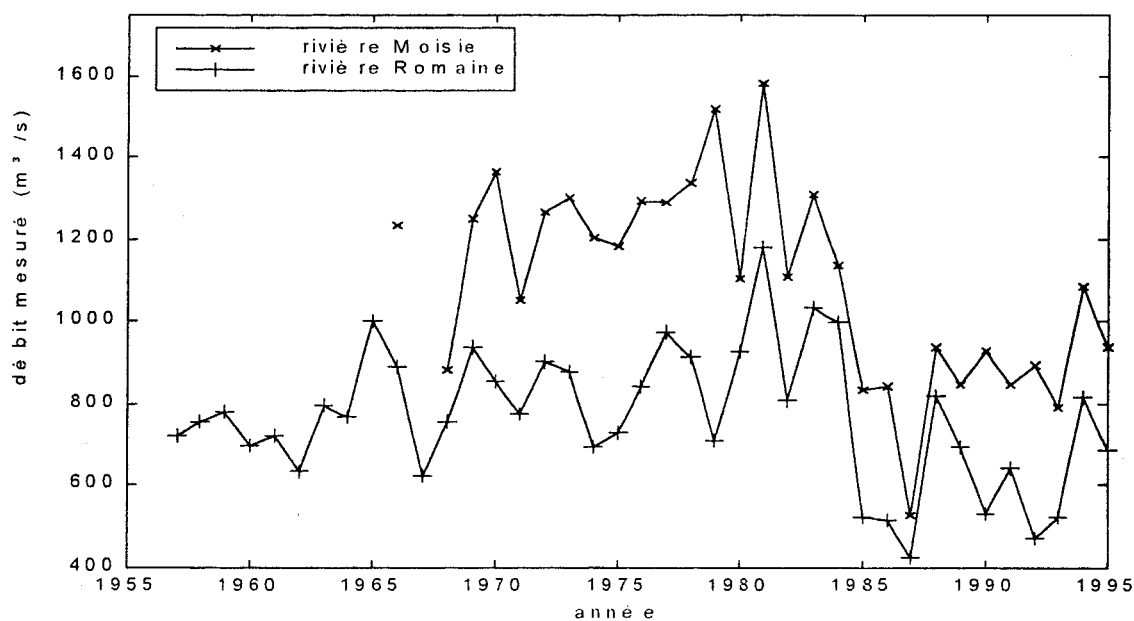


Figure 5 : Série chronologique des débits des rivières Moisie et Romaine.

Tableau 2 : Données des rivières Moisie et Romaine.

Les débits sont en m³/s.

Rivière Moisie

Année	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
Débit	—	1234.0	—	884.0	1250.0	1365.0	1053.5	1265.5	1303.5	1204.0	1185.0
Année	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
Débit	1293.5	1290.0	1340.0	1520.0	1104.0	1585.0	1110.0	1309.0	1134.5	834.0	843.0
Année	1987	1988	1989	1990	1991	1992	1993	1994	1995		
Débit	528.5	939.5	845.5	930.0	846.5	896.5	790.0	1085.0	938.5		

Rivière Romaine

Année	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Débit	NaN	722.5	755.0	781.5	698.0	719.0	633.0	794.0	769.0	1001.0
Année	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
Débit	891.0	623.5	757.0	939.0	854.5	775.5	901.5	878.0	692.0	727.5
Année	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Débit	842.5	973.5	913.5	709.0	928.0	1180.0	806.0	1034.0	999.5	523.5
Année	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Débit	515.0	423.5	819.0	692.5	529.0	640.0	470.5	522.5	817.0	686.5

On tentera de vérifier à l'aide du modèle développé dans le présent rapport si l'une des rivières a subi une baisse plus importante que l'autre. En effet, sous l'hypothèse $H_0: \delta = 0$, on a une relation uniforme entre les deux rivières sur toute la période de données et la baisse du régime de la rivière Romaine est explicable par la baisse de régime de la rivière Moisie. Ce genre de test ne peut être effectué avec la procédure de Lee et Heghinian, qui n'inclut pas une deuxième variable aléatoire dans le modèle.

Pour les calculs, on n'a conservé que les années pour lesquelles on dispose de mesures concomitantes pour ces deux rivières. On a donc une donnée manquante (année 1967) séparant la série de débits. Étant donné que le modèle utilisé présuppose l'indépendance des observations, on a simplement lié les années 1966 et 1968 et on a alors considéré qu'il n'y avait aucune donnée manquante.

On a d'abord déterminé des distributions *a priori* peu informatives sur les paramètres.

$$\pi(\tau) = 1/(n-1), \quad \tau = 1, 2, \dots, n-1;$$

$$\delta \sim N(\mu_\delta = 0, \sigma_\delta^2 = 1000^2);$$

$$b \sim N(\mu_b = 0, \sigma_b^2 = 500^2);$$

$$c \sim N(\mu_c = 1, \sigma_c^2 = 2^2);$$

$$\pi(\sigma_u) \propto \sigma_u^{-1} \text{ avec } \sigma_u \in S = [50, 500].$$

Puisque $\mu_c \neq 0$, on a dû effectuer une transformation des données avant de calculer les distributions *a posteriori*. On note W la variable transformée. La validation des hypothèses de normalité ne peut pas être effectuée de manière automatique pour la variable transformée, car la moyenne de cette variable dépend de l'année du changement de régime. Bien que le graphique de

normalité des débits de la rivière Moisie (figure 6) semble montrer une anomalie dans les données, il a été vérifié par un test de Shapiro-Wilk (Shapiro et Wilk, 1965) que l'hypothèse de normalité est respectée pour cette rivière.

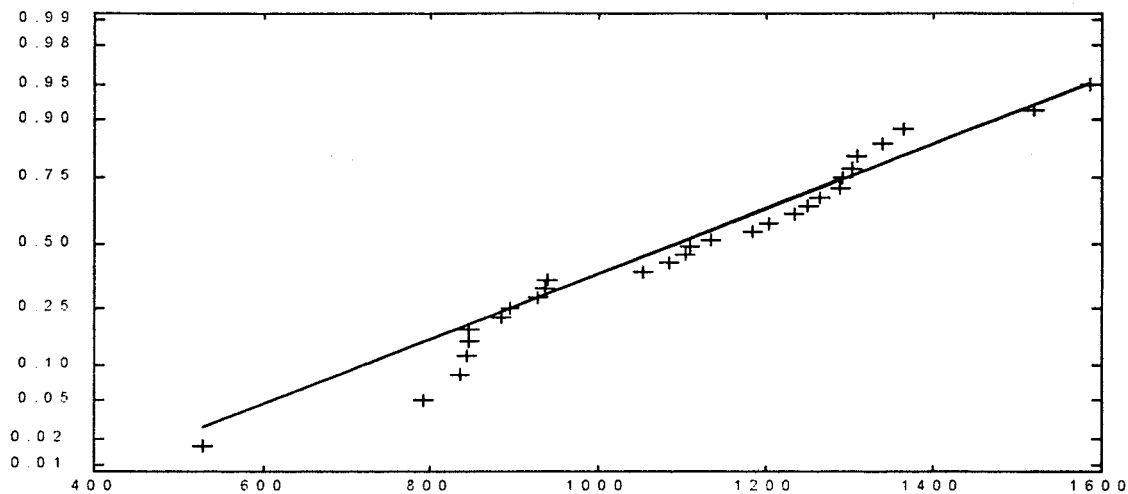


Figure 6 : Diagramme de normalité des débits de la rivière Moisie.

Pour vérifier l'hypothèse de normalité de la rivière Romaine, il faudrait connaître à l'avance le moment d'un éventuel changement par rapport à la rivière Moisie. On convient donc d'effectuer d'abord l'analyse bayésienne pour ensuite valider la normalité des données de la rivière Romaine. Les tracés des distributions conjointes *a posteriori* de τ et δ (figure 7) et des distributions marginales *a posteriori* de τ et δ (figure 8) ne montrent pas une diminution significativement plus importante de la moyenne des débits de la rivière Romaine par rapport à celle de la rivière Moisie. On détecte cependant un possible changement en 1984. Ainsi, en visualisant le graphique de la densité *a posteriori* de $\delta | (\tau = 1984)$, c'est-à-dire la densité *a posteriori* de δ sous l'hypothèse qu'un décalage a eu lieu en 1984 (figure 9), on constate que la probabilité que $\delta \geq 0$ est plus faible avec un seuil de 8%. Il semble donc que la grande diminution du régime de la rivière Romaine depuis 1984 ne soit pas entièrement explicable par la baisse des débits de la rivière Moisie. Cette affirmation sera validée si la normalité des données transformées de la rivière Romaine est respectée. Sous l'hypothèse que le changement a bien eu lieu en 1984, le calcul numérique de l'espérance du paramètre δ donne

$$E(\delta | (\tau = 1984), \underline{x}, \underline{y}) = \int_{\mathbf{R}} \delta \pi(\delta | (\tau = 1984), \underline{x}, \underline{y}) d\delta \approx -96.77.$$

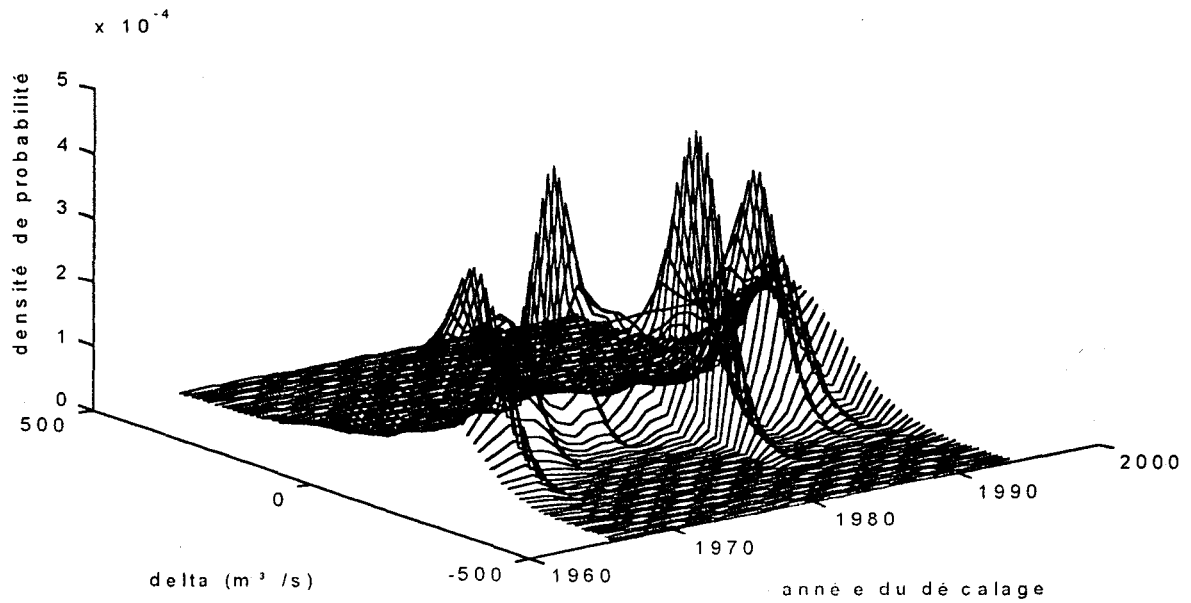


Figure 7 : Densité de probabilité conjointe *a posteriori* de l'année de décalage et de l'amplitude du décalage.

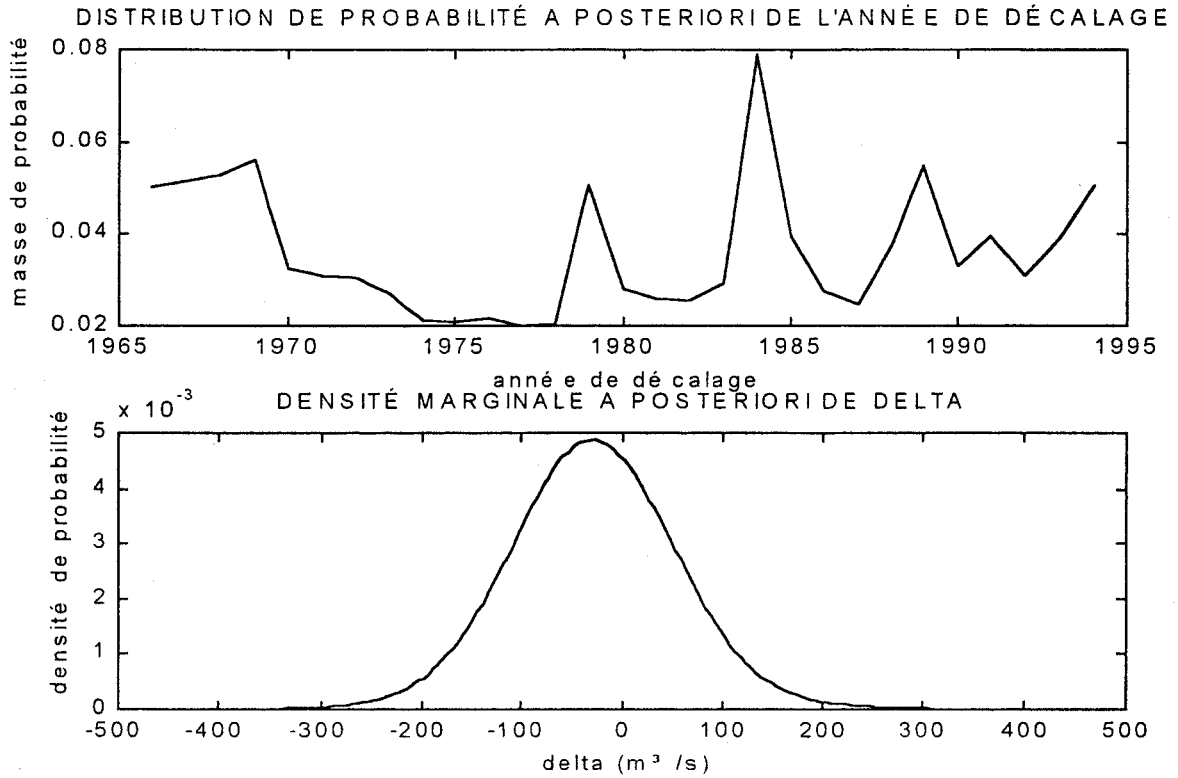


Figure 8 : Distributions marginales *a posteriori* de l'année de décalage et de l'amplitude du décalage.

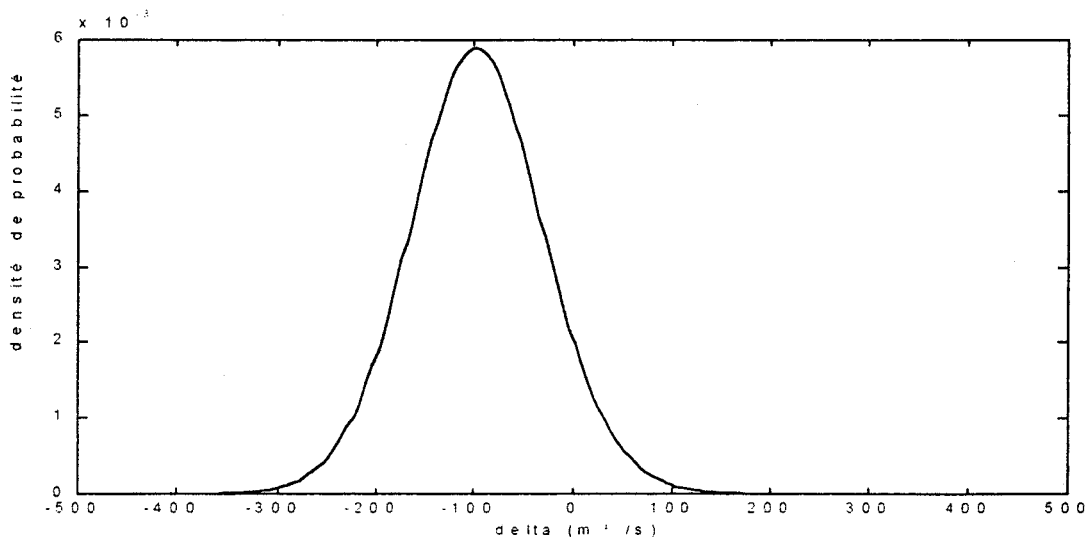


Figure 9 : Distribution de l'amplitude du décalage *a posteriori* étant donné que l'année du décalage est 1984.

On peut ainsi ajuster les valeurs de la variable transformée W pour créer une variable dont la moyenne resterait inchangée dans le temps. La formule de transformation est

$$\tilde{W}_i = W_i - \alpha_i E(\delta | (\tau = 1984), \underline{x}, \underline{y}).$$

Le diagramme de normalité de cette nouvelle variable (figure 10) montre que la normalité des données transformées de la rivière Romaine est effectivement acceptable sous l'hypothèse que le changement a eu lieu en 1984.

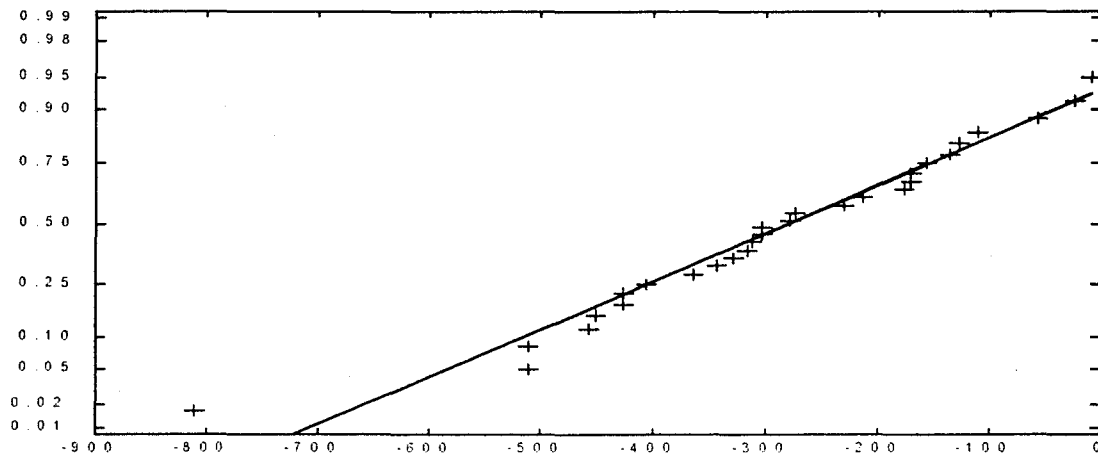
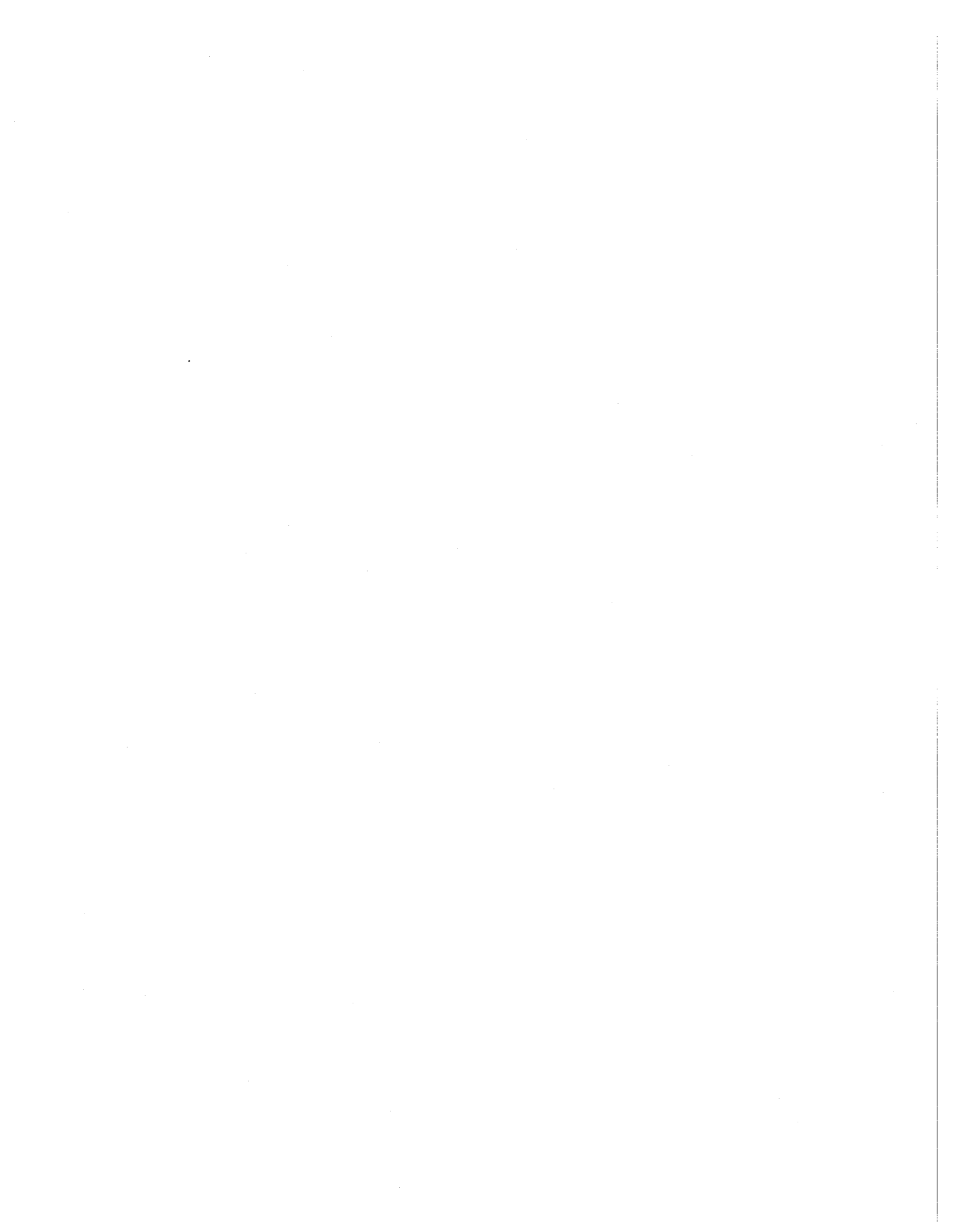


Figure 10 : Diagramme de normalité des données ajustées de la rivière Romaine sous l'hypothèse qu'un décalage a eu lieu en 1984.

5 Conclusion

Le modèle de détection d'un décalage systématique de la moyenne d'une variable présenté dans ce rapport de recherche permet de tenir compte d'une autre variable corrélée avec la première. Toutefois, les distributions *a posteriori* contiennent une intégrale qui est résolue numériquement, ce qui pourrait avoir pour effet d'augmenter légèrement le temps de calcul par ordinateur.

La procédure étant basée sur des couples de données concomitantes, on a dû dans l'exemple d'application éliminer les premières données de la rivière Romaine (1956-1964) puisque les débits de la rivière Moisie sont seulement disponibles à partir de 1965. Comme les données manquantes sont fréquentes dans la réalité, il serait intéressant de trouver un moyen de les inclure dans la structure de changement exploitée dans ce rapport.



6 Bibliographie

- Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis*. Second Edition. Springer Series in Statistics. Springer-Verlag New York, Inc., New York, NY.
- Bernier, J. (1994). Statistical detection of changes in geophysical series *in* *Engineering Risk in Natural Resources Management*. Duckstein L. And E. Parent (eds.), NATO Advanced Studies Institute Series, Series E - Vol. 275, Kluwer : 159-176.
- Chin Choy, J. H. et L. D. Broemeling (1980). Some Bayesian Inferences for a Changing Linear Model. *Technometrics*, 22(1): 71-78.
- Gradshteyn, I. S. et I. M. Ryzhik (1994). *Table of Integrals, Series, and Products*, Fifth Edition. Alan Jeffrey, Editor, University of Newcastle upon Tyne, England.
- Lee, A. S.F. et S. M. Heghinian (1977). A Shift of the Mean Level in a Sequence of Independent Normal Random Variables : A Bayesian Approach. *Technometrics*, 19(4), 503-506.
- Lettenmaier, D. P., E. F. Wood et J. R. Wallis (1994). Hydro-Climatological Trends in the Continental United States, 1948-88. *Journal of Climate*, Vol. 7 : 586-607.
- Maronna, R. et V. J. Yohai (1978). A Bivariate Test for the Detection of a Systematic Change in Mean. *J. Am. Statist. Assoc.*, 73 :640-645.
- Ouarda, T. B.M.J., P. F. Rasmussen, B. Bobée (1996). Rationalisation du réseau hydrométrique de la Province de Québec pour le suivi des changements climatiques. Institut National de la Recherche Scientifique, INRS-Eau, Sainte-Foy, Québec. No R-476, 71 pp.
- Ouarda, T. B.M.J., P. F. Rasmussen, J.-F. Cantin, B. Bobée, R. Laurence, V. D. Hoang et G. Barabé (1998). Identification d'un réseau hydrométrique pour le suivi des modifications climatiques dans la province de Québec. *Revue Des Sciences De l'Eau* (sous presse).
- Perreault, L., M. Haché, M. Slivitzky et B. Bobée (1998). Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Hydrology and Hydraulics* (accepté).

Shapiro, S.S., et M.B. Wilk (1965). An Analysis of Variance test for Normality, Complete Samples, *Biometrika*, 52 : 591-611.

Solow, A. R. (1988). A Bayesian Approach to Statistical Inference about Climate Change. *Journal of Climate*, Vol. 1 : 512-521.

The Math Works, Inc. (1997). MATLAB, The Language of Technical Computing, Version 5.1, Natick, MA.

Appendice A : Calcul de la densité conjointe de X et Y

Soit $f_U(u)$, la fonction de densité de la variable aléatoire U du modèle (1), $F_U(u)$, sa fonction de distribution cumulative, et $F_X(x)$, la fonction de distribution cumulative de la variable aléatoire X . On a

$$\begin{aligned} F_{X,Y}(x,y) &:= P(X \leq x, Y \leq y) \\ &= \int_{\mathbf{R}} P(X \leq x, cX \leq y - b_i - u) f_U(u) du. \end{aligned}$$

Si $c = 0$, alors

$$F_{X,Y}(x,y) = \int_{-\infty}^{y-b_i} F_X(x) f_U(u) du = F_X(x) F_U(y - b_i),$$

ce qui implique que

$$f_{X,Y}(x,y) = f_X(x) f_U(y - b_i).$$

Si $c > 0$, alors

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{y-b_i-cx}^{\infty} F_X\left(\frac{y-b_i-u}{c}\right) f_U(u) du + F_X(x) F_U(y - b_i - cx) \\ &= \int_{y-b_i-cx}^v F_X\left(\frac{y-b_i-u}{c}\right) f_U(u) du + \int_v^{\infty} F_X\left(\frac{y-b_i-u}{c}\right) f_U(u) du + F_X(x) F_U(y - b_i - cx) \end{aligned}$$

où $v > y - b_i - cx$.

En utilisant la formule de différentiation d'une integrale définie contenant un paramètre (Gradshteyn et Ryzhik, 1994), on obtient

$$\frac{\partial}{\partial x} F_{X,Y}(x,y) = c F_X(x) f_U(y - b_i - cx) + f_X(x) F_U(y - b_i - cx) - c F_X(x) f_U(y - b_i - cx)$$

et donc

$$f_{X,Y}(x,y) = f_X(x) f_U(y - b_i - cx).$$

Par symétrie, ce résultat est aussi vrai dans le cas où $c < 0$. On a donc en général que

$$f_{X,Y}(x,y) = f_X(x)f_Y(y - b_i - cx).$$

Cette formule peut se ramener à la densité d'une loi normale bivariée. On suggère pour montrer cette dernière affirmation de partir de la densité de loi normale bivariée pour arriver au dernier résultat.

□

Appendice B : Résolution de l'intégrale (5)

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \int_{\mathbb{S}} \int_{\varepsilon} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\sigma^n \sigma_u^{n-1}} \\ &\times \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} + \frac{\sum_{i=1}^n (y_i - b - a_i \delta - cx_i)^2}{\sigma_u^2} + \frac{h^2}{\sigma_b^2} + \frac{c^2}{\sigma_c^2} \right)\right) dhdc d\mu d\sigma d\sigma_u \end{aligned} \quad (5)$$

On remarque tout d'abord que les termes en μ et σ peuvent être isolés dans l'intégration. C'est-à-dire

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \int_{\varepsilon} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\sigma^n} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right)\right) d\mu d\sigma \\ &\times \int_{\mathbb{S}} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sigma_u^{n-1}} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - b - a_i \delta - cx_i)^2}{\sigma_u^2} + \frac{h^2}{\sigma_b^2} + \frac{c^2}{\sigma_c^2} \right)\right) dhdc d\sigma_u . \end{aligned}$$

Or, $\int_{\varepsilon} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\sigma^n} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right)\right) d\mu d\sigma$ converge. En effet, puisque $\sigma \geq \varepsilon > 0$,

$$0 \leq \int_{\varepsilon} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\sigma^n} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right)\right) d\mu d\sigma$$

$$\leq \int_{\varepsilon}^{\infty} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\sigma^n} d\mu d\sigma \leq \int_{\varepsilon}^{\infty} \int_{\mathbb{R}} \frac{\pi(\mu)\pi(\sigma)}{\varepsilon^n} d\mu d\sigma = \frac{1}{\varepsilon^n}.$$

On réduit ainsi l'intégrale (5) à

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \\ &\times \int_{\mathbb{S}} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sigma_u^{n+1}} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - b - a_i \delta - cx_i)^2}{\sigma_u^2} + \frac{b^2}{\sigma_b^2} + \frac{c^2}{\sigma_c^2} \right)\right) db dc d\sigma_u. \end{aligned} \quad (12)$$

Les intégrales sur b et c peuvent se ramener à une forme connue (Gradshteyn et Ryzhik, 1994)

$$\int_{\mathbb{R}} \exp\left(-\frac{1}{p}(t^2 - rt)\right) dt = \sqrt{\pi p} \exp\left(\frac{r^2}{4p}\right) \text{ lorsque } p > 0. \quad (13)$$

On a

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right) \\ &\times \int_{\mathbb{S}} \frac{1}{\sigma_u^{n+1}} \int_{\mathbb{R}} \exp\left(-\frac{c^2}{2\sigma_c^2}\right) \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - b - a_i \delta - cx_i)^2}{\sigma_u^2} + \frac{b^2}{\sigma_b^2} \right)\right) db dc d\sigma_u. \end{aligned} \quad (14)$$

Soit $S(\sigma_u) := \frac{n\sigma_b^2}{\sigma_u^2 + n\sigma_b^2}$, en utilisant l'équation (13), on obtient la solution de l'intégrale sur b .

$$\begin{aligned} &\int_{\mathbb{R}} \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - b - a_i \delta - cx_i)^2}{\sigma_u^2} + \frac{b^2}{\sigma_b^2} \right)\right) db \\ &= \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \left(\frac{h^2 - 2hS(\sigma_u) \left(\bar{y} - \frac{n-\tau}{n} \delta - c\bar{x} \right) + S(\sigma_u) \frac{1}{n} \sum_{i=1}^n (y_i - a_i \delta - cx_i)^2}{\sigma_u^2 \frac{S(\sigma_u)}{n}} \right)\right) db \end{aligned}$$

$$= \sqrt{2\pi \frac{S(\sigma_u)}{n}} \sigma_u \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta - cx_i)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta - c\bar{x} \right)^2}{\sigma_u^2} \right) \right)$$

En insérant ce résultat dans l'équation (12) et en laissant tomber les facteurs constants en τ , δ , c et σ_u , il reste à résoudre

$$\begin{aligned} \pi(\tau, \delta | \underline{x}, \underline{y}) &\propto \exp \left(-\frac{\delta^2}{2\sigma_\delta^2} \right) \int_S \frac{1}{\sigma_u^n} \sqrt{S(\sigma_u)} \\ &\times \int_{\mathbb{R}} \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta - cx_i)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta - c\bar{x} \right)^2}{\sigma_u^2} + \frac{c^2}{\sigma_c^2} \right) \right) dc d\sigma_u. \end{aligned} \quad (15)$$

On a

$$\begin{aligned} &\frac{\sum_{i=1}^n (y_i - a_i \delta - cx_i)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta - c\bar{x} \right)^2}{\sigma_u^2} + \frac{c^2}{\sigma_c^2} \\ &= \frac{1}{\left(\frac{\sigma_u^2}{R(\sigma_u)} \right)} \left(c^2 - 2c \frac{\sum_{i=1}^n x_i (y_i - a_i \delta) - S(\sigma_u) n \bar{x} \left(\bar{y} - \frac{n-\tau}{n} \delta \right)}{R(\sigma_u)} \right. \\ &\quad \left. + \frac{\sum_{i=1}^n (y_i - a_i \delta)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta \right)^2}{R(\sigma_u)} \right) \end{aligned}$$

$$\text{où } R(\sigma_u) = \frac{\sigma_u^2}{\sigma_c^2} + \sum_{i=1}^n x_i^2 - S(\sigma_u) n \bar{x}^2.$$

On note que $R(\sigma_u) > \sum_{i=1}^n (x_i - \bar{x})^2 \geq 0$. On peut donc utiliser l'équation (13) et on obtient

$$\int_{\mathbb{R}} \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta - cx_i)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta - c\bar{x} \right)^2}{\sigma_u^2} + \frac{c^2}{\sigma_c^2} \right) \right) dc$$

$$= \frac{\sqrt{2\pi}\sigma_u}{\sqrt{R(\sigma_u)}} \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta \right)^2}{\sigma_u^2} \right. \right. \\ \left. \left. - \frac{\left(\sum_{i=1}^n x_i (y_i - a_i \delta) - S(\sigma_u) n \bar{x} \left(\bar{y} - \frac{n-\tau}{n} \delta \right) \right)^2}{\sigma_u^2 R(\sigma_u)} \right) \right).$$

On a donc

$$\pi(\tau, \delta | \underline{x}, \underline{y}) \propto \exp \left(-\frac{\delta^2}{2\sigma_\delta^2} \right) \int_S \frac{1}{\sigma_u^{n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta \right)^2}{\sigma_u^2} \right. \right. \\ \left. \left. - \frac{\left(\sum_{i=1}^n x_i (y_i - a_i \delta) - S(\sigma_u) n \bar{x} \left(\bar{y} - \frac{n-\tau}{n} \delta \right) \right)^2}{\sigma_u^2 R(\sigma_u)} \right) \right) d\sigma_u.$$

(16)

Plusieurs méthodes ont été utilisées pour tenter de résoudre la dernière intégrale. L'intégration directe, l'intégration par parties, le développement en série de Taylor du terme exponentiel, le changement de l'ordre d'intégration, aucune n'a mené à une solution analytique. On s'est souvent référé à Gradshteyn et Ryzhik (1994). On proposera donc d'intégrer numériquement la dernière intégrale. Il reste maintenant à calculer le facteur réducteur Ω de la densité $\pi(\tau, \delta | \underline{x}, \underline{y})$

$$\Omega = \sum_{\tau=1}^{n-1} \int_S \frac{1}{\sigma_u^{n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \int_R \exp \left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n (y_i - a_i \delta)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta \right)^2}{\sigma_u^2} \right. \right. \\ \left. \left. - \frac{\left(\sum_{i=1}^n x_i (y_i - a_i \delta) - S(\sigma_u) n \bar{x} \left(\bar{y} - \frac{n-\tau}{n} \delta \right) \right)^2}{\sigma_u^2 R(\sigma_u)} + \frac{\delta^2}{\sigma_\delta^2} \right) \right) d\delta d\sigma_u.$$

(17)

On va encore se ramener sous la forme de l'équation (13).

Étant donné que les calculs en cause ne sont pas complexes mais excessivement lourds, on ne donne ici que le résultat de la simplification du terme à l'exposant.

$$\begin{aligned} & \frac{\sum_{i=1}^n (y_i - \alpha_i \delta)^2 - S(\sigma_u) n \left(\bar{y} - \frac{n-\tau}{n} \delta \right)^2}{\sigma_u^2} - \frac{\left(\sum_{i=1}^n x_i (y_i - \alpha_i \delta) - S(\sigma_u) n \bar{x} \left(\bar{y} - \frac{n-\tau}{n} \delta \right) \right)^2}{\sigma_u^2 R(\sigma_u)} + \frac{\delta^2}{\sigma_\delta^2} \\ &= \frac{1}{\left(\frac{\sigma_u^2}{T(\tau, \sigma_u)} \right)} \left(\delta^2 - 2\delta \frac{B(\tau, \sigma_u)}{T(\tau, \sigma_u)} + \frac{C(\tau, \sigma_u)}{T(\tau, \sigma_u)} \right) \end{aligned}$$

où

$$\begin{aligned} B(\tau, \sigma_u) &= \left(\sum_{i=\tau+1}^n y_i - S(\sigma_u) (n-\tau) \bar{y} \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=1}^n x_i y_i - S(\sigma_u) n \bar{x} \bar{y} \right) \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x} \right), \\ C(\tau, \sigma_u) &= \left(\sum_{i=1}^n y_i^2 - S(\sigma_u) n \bar{y}^2 \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=1}^n x_i y_i - S(\sigma_u) n \bar{x} \bar{y} \right)^2 \end{aligned}$$

et

$$T(\tau, \sigma_u) = \frac{\sigma_u^2}{\sigma_\delta^2} + (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n} \right) \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x} \right)^2.$$

On note que l'expression (16) peut alors s'écrire

$$\pi(\tau, \delta | \underline{x}, \underline{y}) \propto \int_S \frac{1}{\sigma_u^{n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} (T(\tau, \sigma_u) \delta^2 - 2B(\tau, \sigma_u) \delta + C(\tau, \sigma_u))\right) d\sigma_u. \quad (18)$$

Pour pouvoir utiliser l'équation (13) dans le but de calculer l'intégrale sur δ , il faut que $T(\tau, \sigma_u) > 0$ pour $\tau = 1, 2, \dots, n-1$ et $\sigma_u > 0$. C'est ce qu'on montre ici. Soit $\tau \in \{1, 2, \dots, n-1\}$ et $\sigma_u > 0$, on a

$$\begin{aligned} T(\tau, \sigma_u) &= \frac{\sigma_u^2}{\sigma_\delta^2} + (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n} \right) \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x} \right)^2 \\ &> (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n} \right) \right) - \frac{1}{R(\sigma_u)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x} \right)^2 \end{aligned}$$

$$\geq (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) - \frac{1}{\left(\sum_{i=1}^n x_i^2 - S(\sigma_u) n \bar{x}^2\right)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x}\right)^2.$$

On va vérifier que la dernière expression est bornée inférieurement par 0.

$$\begin{aligned} & (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) - \frac{1}{\left(\sum_{i=1}^n x_i^2 - S(\sigma_u) n \bar{x}^2\right)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x}\right)^2 \geq 0 \\ \Leftrightarrow & \frac{1}{\left(\sum_{i=1}^n x_i^2 - S(\sigma_u) n \bar{x}^2\right)} \left(\sum_{i=\tau+1}^n x_i - S(\sigma_u) (n-\tau) \bar{x}\right)^2 \leq (n-\tau) \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \\ \Leftrightarrow & \left(\sum_{i=\tau+1}^n x_i\right)^2 - 2S(\sigma_u) (n-\tau) \bar{x} \sum_{i=\tau+1}^n x_i + S^2(\sigma_u) (n-\tau)^2 \bar{x}^2 \\ & \leq (n-\tau) \sum_{i=1}^n x_i^2 \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) - S(\sigma_u) (n-\tau) n \bar{x}^2 + S^2(\sigma_u) (n-\tau)^2 \bar{x}^2 \\ \Leftrightarrow & \sum_{i=1}^n x_i^2 \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) + S(\sigma_u) \left(2\bar{x} \sum_{i=\tau+1}^n x_i - n \bar{x}^2\right) - \frac{1}{(n-\tau)} \left(\sum_{i=\tau+1}^n x_i\right)^2 \geq 0 \\ \Leftrightarrow & \sum_{i=1}^n x_i^2 \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) + \frac{S(\sigma_u)}{n} \left(\left(\sum_{i=\tau+1}^n x_i\right)^2 - \left(\sum_{i=1}^{\tau} x_i\right)^2\right) - \frac{1}{(n-\tau)} \left(\sum_{i=\tau+1}^n x_i\right)^2 \geq 0 \\ \Leftrightarrow & \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \left[\left(\sum_{i=1}^{\tau} x_i^2 - \frac{1}{\tau} \left(\sum_{i=1}^{\tau} x_i\right)^2\right) + \left(\sum_{i=\tau+1}^n x_i^2 - \frac{1}{n-\tau} \left(\sum_{i=\tau+1}^n x_i\right)^2\right)\right] \\ & + \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \left(\frac{1}{\tau} \left(\sum_{i=1}^{\tau} x_i\right)^2 + \frac{1}{n-\tau} \left(\sum_{i=\tau+1}^n x_i\right)^2\right) \\ & + \frac{S(\sigma_u)}{n} \left(\left(\sum_{i=\tau+1}^n x_i\right)^2 - \left(\sum_{i=1}^{\tau} x_i\right)^2\right) - \frac{1}{(n-\tau)} \left(\sum_{i=\tau+1}^n x_i\right)^2 \geq 0 \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \left[\sum_{i=1}^{\tau} \left(x_i - \frac{1}{\tau} \left(\sum_{j=1}^{\tau} x_j\right)\right)^2 + \sum_{i=\tau+1}^n \left(x_i - \frac{1}{n-\tau} \left(\sum_{j=\tau+1}^n x_j\right)\right)^2 \right] \\
&\quad + \left(\sum_{i=1}^{\tau} x_i\right)^2 \left(\frac{1}{\tau} - S(\sigma_u) \left(\frac{n-\tau}{n\tau}\right) - \frac{S(\sigma_u)}{n}\right) \\
&\quad + \left(\sum_{i=\tau+1}^n x_i\right)^2 \left(\frac{1}{n-\tau} - \frac{S(\sigma_u)}{n} + \frac{S(\sigma_u)}{n} - \frac{1}{n-\tau}\right) \geq 0 \\
&\Leftrightarrow \left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \left[\sum_{i=1}^{\tau} \left(x_i - \frac{1}{\tau} \left(\sum_{j=1}^{\tau} x_j\right)\right)^2 + \sum_{i=\tau+1}^n \left(x_i - \frac{1}{n-\tau} \left(\sum_{j=\tau+1}^n x_j\right)\right)^2 \right] \\
&\quad + \frac{(1 - S(\sigma_u))}{\tau} \left(\sum_{i=1}^{\tau} x_i\right)^2 \geq 0
\end{aligned}$$

Comme $0 < S(\sigma_u) < 1$, on a bien

$$\begin{aligned}
&\left(1 - S(\sigma_u) \left(\frac{n-\tau}{n}\right)\right) \left[\sum_{i=1}^{\tau} \left(x_i - \frac{1}{\tau} \left(\sum_{j=1}^{\tau} x_j\right)\right)^2 + \sum_{i=\tau+1}^n \left(x_i - \frac{1}{n-\tau} \left(\sum_{j=\tau+1}^n x_j\right)\right)^2 \right] \\
&\quad + \frac{(1 - S(\sigma_u))}{\tau} \left(\sum_{i=1}^{\tau} x_i\right)^2 \\
&\geq \frac{\tau}{n} \left[\sum_{i=1}^{\tau} \left(x_i - \frac{1}{\tau} \left(\sum_{j=1}^{\tau} x_j\right)\right)^2 + \sum_{i=\tau+1}^n \left(x_i - \frac{1}{n-\tau} \left(\sum_{j=\tau+1}^n x_j\right)\right)^2 \right] \\
&\geq 0,
\end{aligned}$$

ce qui implique que $T(\tau, \sigma_u) > 0$ pour $\tau = 1, 2, \dots, n-1$ et $\sigma_u > 0$.

On peut maintenant utiliser l'équation (13) et on a

$$\Omega = \sqrt{2\pi} \sum_{\tau=1}^{n-1} \int_S \frac{1}{\sigma_u^{n-2}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)T(\tau, \sigma_u)}} \exp\left(-\frac{1}{2\sigma_u^2} \left(C(\tau, \sigma_u) - \frac{B^2(\tau, \sigma_u)}{T(\tau, \sigma_u)}\right)\right) d\sigma_u. \quad (19)$$

□

Appendice C : Calcul de la fonction de distribution conjointe

$$P\left(\tau \leq \tau_0, \delta \leq \delta_0 \mid \underline{x}, \underline{y}\right)$$

On énonce d'abord une généralisation de l'équation (13),

$$\int_{-r}^u \exp\left(-\frac{1}{p}(t^2 - rt)\right) dt = \sqrt{\pi p} \exp\left(\frac{r^2}{4p}\right) \Phi\left(\frac{2u - r}{\sqrt{2p}}\right) \text{ avec } p > 0 \quad (20)$$

où Φ est la fonction de distribution cumulative d'une loi normale standardisée $N(0,1)$. Soit $\tau_u \in \{1, 2, \dots, u-1\}$ et $\delta_u \in \mathfrak{R}$, en se basant sur le calcul de Ω effectué précédemment et en utilisant l'équation (20), on arrive à

$$\begin{aligned} P\left(\tau \leq \tau_0, \delta \leq \delta_0 \mid \underline{x}, \underline{y}\right) &= \frac{1}{\Omega} \sum_{\tau=1}^{\tau_0} \int_S \frac{1}{\sigma_u^{2n-1}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u)}} \\ &\quad \times \int_{-\delta_u}^{\delta_0} \exp\left(-\frac{1}{2\sigma_u^2} \left(T(\tau, \sigma_u) \delta^2 - 2B(\tau, \sigma_u) \delta + C(\tau, \sigma_u) \right)\right) d\delta d\sigma_u \\ &= \frac{\sqrt{2\pi}}{\Omega} \sum_{\tau=1}^{\tau_0} \int_S \frac{1}{\sigma_u^{2n-2}} \sqrt{\frac{S(\sigma_u)}{R(\sigma_u) T(\tau, \sigma_u)}} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_u^2} \left(C(\tau, \sigma_u) - \frac{B^2(\tau, \sigma_u)}{T(\tau, \sigma_u)} \right)\right) \Phi\left(2\sqrt{2} \frac{\sqrt{T(\tau, \sigma_u)}}{\sigma_u} \left(\delta_0 - \frac{B(\tau, \sigma_u)}{T(\tau, \sigma_u)} \right)\right) d\sigma_u \end{aligned}$$

□