Estimation ondelette des paramètres de stabilité et d'autosimilarité des processus α-stables autosimilaires

Patrice ABRY[1], Béatrice PESQUET-POPESCU[2], Murad S. TAQQU[3]

[1]CNRS UMR 5672, Laboratoire de Physique, ENS Lyon, 46, Allée d'Italie, 69364 LYON Cedex 07
tél : (+33)4 72 72 84 93 - Fax : (+33)4 72 72 80 80
Ce travail a été réalisé avec le soutien du Programme Télécommunications TL97035 du CNRS
et de l'Opération Thématique *Fractales et Ondelettes* du GDR PRC ISIS.

[2]Image and Communication Group, *Laboratoires d'Electronique de Philips*, 94453 Liméil-Brévannes, FRANCE

[3]*Department of Mathematics , 111 Cummington Street, Boston University, Boston, MA 02215-2411, USA*
tel: (+1) (617) 353-3022 - Fax: (+1) (617) 353-8100
Ce travail a été réalisé avec le soutien partiel du grant ANI-9805623 de l'Université de Boston, USA.
pabry@physique.ens-lyon.fr, http://www.physique.ens-lyon.fr/ts, pesquet@lep-philips.fr,
murad@math.bu.edu, http://math.bu.edu/people/murad

Résumé – Les processus α-stable autosimilaires (tels que le mouvement linéaire fractionnaire stable) offrent un cadre intéressant pour modéliser deux grandes classes de phénomènes de diffusion anormaux : variance infinie des pas de la marche aléatoire et dépendance statistique à longue portée entre ces pas. Nous proposons, pour ces processus, des estimateurs ondelettes des paramètres de stabilité, α, et d'autosimilarité, H, caractérisant ces deux anomalies. Du fait de l'adéquation entre l'analyse multirésolution et les phénomènes d'autosimilarité, ces estimateurs présentent d'excellentes performances statistiques, étudiées ici, à la fois, théoriquement et par simulations numériques.

Abstract – Self-similar α-stable processes (such as the linear fractional stable motion) offer a relevant framework to model two major classes of anomalous diffusion phenomena: infinite variance of the steps of the random walks and long range statistical dependence among those steps. For such processes, we define, here, wavelet-based estimators for the stability parameter α and the self-similarity parameter H, that describe those two anomalies. Due to the intimate adequacy between the multiresolution analysis and the self-similarity phenomenon, these estimators are exhibiting excellent statistical performance, that are studied here both theoretically and from numerical simulations.

1 Motivation

Le caractère anormal de certains phénomènes de diffusion peut être modélisé essentiellement de deux façons. D'une part, les pas de la marche aléatoire, sous-jacente au phénomène de diffusion, peuvent être à forte dépendance statistique plutôt qu'indépendants. D'autre part, ils peuvent rester statistiquement indépendants mais ne plus être de variance (ou même de moyenne) finie. Les processus α-stables autosimilaires [17] constituent le paradigme idéal pour la modélisation et l'étude de ces situations où les deux difficultés (variance infinie et dépendance statistique à longue portée) apparaissent conjointement. Ces processus sont, en effet, caractérisés par deux paramètres, de stabilité $0 < \alpha \leq 2$, et d'autosimilarité $0 < H < 1$, et leurs moments d'ordre supérieurs à α sont infinis et ils sont caractérisés par de la dépendance à longue portée dès que $H > 1/\alpha$ [17]. Nous proposons ici des estimateurs des paramètres α et H, construits à partir des coefficients d'ondelettes du processus. Il a déjà été indiqué [1, 2, 3] que les analyses en ondelettes constituent un excellent outil pour l'étude de l'autosimilarité. Les estimateurs proposés ici bénéficient de cette adéquation : ils sont de grandes simplicités conceptuelle et pratique – ils consistent en régression linéaire – et présentent de très bonnes performances statistiques.

Nous rappelons brièvement les définitions des processus α-stables autosimilaires et les propriétés statistiques de leurs coefficients d'ondelettes. Nous donnons ensuite la définition des estimateurs des paramètres H et α et étudions leurs performances, à la fois théoriquement et au moyen de simulations numériques.

2 Processus α-stables autosimilaires

2.1 Définition

On définit les processus α-stables à l'aide de la représentation intégrale introduite dans [17] :

$$x(t) = \int_{\mathbb{R}} f(t,u) M(du) \,, \tag{1}$$

où $M(du)$ est une mesure α-stable que nous supposerons symétrique (SαS) [17, 13] et $f(t,u)$ un noyau d'intégration qui contrôle la dépendance statistique du processus. Pour certaines formes du noyau, le processus est autosimilaire [17, 9, 15, 14, 5]. Rappelons rapidement qu'un processus est dit statistiquement autosimilaire, avec paramètre d'autosimilarité H, si [17]:

$$\forall c > 0, \{c^{-H}x(ct), t \in \mathbb{R}\} \stackrel{d}{=} \{x(t), t \in \mathbb{R}\} \qquad (2)$$

où $\stackrel{d}{=}$ indique l'égalité de toutes les distributions finies des processus. Rappelons qu'un processus est dit à accroissements stationnaires si les distributions finies des processus $\{y_h(t) = x(t+h) - x(t), t \in \mathbb{R}\}$ ne dépendent pas de t. Deux exemples particulièrement intéressants pour l'étude qui suit sont les vols de Lévy et le mouvement linéaire fractionnaire stable (LFSM) [17].

Le premier est défini par

$$f(t,u) = \begin{cases} 1 & \text{si } 0 \le u \le t, \\ -1 & \text{si } t \le u \le 0, \\ 0 & \text{sinon.} \end{cases}$$

Ses accroissements $x(t+1) - x(t)$ sont stationnaires et indépendants. Il est autosimilaire de paramètre $H = 1/\alpha$.

Le second est défini par un paramètre $-\infty < d < 1 - 1/\alpha$ et son noyau est donné (dans le cas d'un LFSM bien équilibré) par

$$f(t,u) = (t-u)_+^d - (-u)_+^d,$$

où $(t)_+ = t$ si $t \ge 0$ et 0 sinon. Ce processus est autosimilaire de paramètre $H = d + 1/\alpha$. Ses accroissements sont stationnaires mais statistiquement dépendants. On parlera de dépendance à longue portée [7] lorsque $d > 0$ [17]. La codifférence de ces accroissements $y_h(t)$ décroît alors comme :

$$|\text{Cod } y_h(t), y_h(t+\tau)| \sim |\tau|^{\alpha(H-1)}, |\tau| \to +\infty,$$

pour toutes les paires (α, H) auxquelles nous nous intéressons ici [17]. Cette décroissance algébrique (en loi de puissance de τ) et donc lente, de la dépendance statistique indique que n'importe quelle paire d'échantillons, aussi éloignés soient-ils l'un de l'autre, présente une forte liaison statistique, impossible à négliger sans manquer quelque chose d'essentiel dans l'analyse des données : c'est la dépendance à longue portée, qui vient significativement compliquer l'analyse des données [7, 3]. Pour la définition de la codifférence, on pourra consulter [17], rappelons simplement que pour $\alpha = 2$, codifférence et covariance coïncident.

2.2 Coefficients d'ondelettes

Notons $d_x(j,k) = \langle x, \psi_{j,k} \rangle$ les coefficients d'ondelettes [8, 12] du processus x étudié. Les fonctions $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ forment la base d'ondelettes obtenues par dilatation et translation de l'ondelette-mère ψ_0. Cette dernière est caractérisée par son nombre de moments nuls N, tel que

$$\forall m \in \{0, \ldots, N-1\}, \quad \int_{\mathbb{R}} t^m \, \psi_0(t) dt = 0.$$

Les coefficients d'ondelettes d'un processus $\alpha-$stable autosimilaire à accroissements stationnaires vérifient les propriétés suivantes [4, 6, 9, 10, 14, 15, 5] :

P0 : Sous réserve de conditions peu restrictives de décroissance temporelle de l'ondelette, que l'on supposera toujours satisfaites et sous certaines hypothèses sur le processus étudié, les $d_x(j,k)$ sont des variables aléatoires α-stables, de représentation intégrale :

$$d_x(j,k) = \int_{\mathbb{R}} h_{j,k}(u) M(du),$$

avec $h_{j,k}(u) = \int_{\mathbb{R}} f(t,u)\psi_{j,k}(t)dt$.

P1 : Les $d_x(j,k)$ reproduisent, de façon exacte, l'autosimilarité du processus, à travers la relation essentielle :

$$\{d_x(j,k), k \in \mathbb{Z}\}) \stackrel{d}{=} \{2^{j(H+\frac{1}{2})} d_x(0,k), k \in \mathbb{Z}\}.$$

Cette propriété découle fondamentalement du fait que l'outil d'analyse, la base d'ondelette, est construit à partir de l'opérateur de changement d'échelle, intimement lié à la propriété d'autosimilarité.

P2 : Les $d_x(j,k)$ forment, à chaque octave j, des suites stationnaires. Cette propriété découle du fait que, par définition, $N \ge 1$.

P3 : La dépendance statistique des $d_x(j,k)$ peut être étudiée par le biais de leur codifférence. Dans le cas du LFSM, il est montré dans [9, 11] que, lorsque $|2^j k - 2^{j'} k'| \to +\infty$,

$$|\text{Cod } d_x(j,k), d_x(j',k')| \le C|2^j k - 2^{j'} k'|^{-(\alpha/2)(N-H)}. \qquad (3)$$

Cette relation met en évidence le rôle fondamental du nombre de moments nuls de l'ondelette-mère. En augmentant N, on peut réduire, autant que l'on veut la portée de la dépendance statistique entre des paires de coefficients d'ondelettes suffisamment éloignés et situés à la même octave j.

P3LOG : De plus, grâce à **P0** et **P3**, on peut étudier la structure de covariance de la variable $\log_2 |d_x(j,k)|$ et montrer que [9, 11]:

$$|\text{Cov } \log_2 |d_x(j,k)|, \log_2 |d_x(j,k')|| \le C|k - k'|^{-(\alpha/4)(N-H)}. \qquad (4)$$

De nouveau, N permet de réduire la portée de la corrélation. Cette propriété est idéalisée dans la suite en :
ID : Les variables aléatoires $\{\log_2 |d_x(j,k)|, k \in \mathbb{Z}, j \in \mathbb{Z}^+\}$ sont décorrélées les unes des autres.

3 Estimation de H et α

3.1 Estimation de H

Nous allons maintenant estimer les paramètres H et α à partir des propriétés des coefficients d'ondelettes décrites ci-dessus. Commençons par H. De la propriété fondamentale **P1**, on déduit

$$\mathbb{E}\log_2 |d_x(j,k)| = j(H + 1/2) + \mathbb{E}\log_2 |d_x(0,k)|$$

qui suggère de procéder à l'estimation de H par régression linéaire effectuée sur un graphe $\mathbb{E}\log_2|d_x(j,k)|$ versus $\log_2(2^j) = j$. Pour réaliser pratiquement l'estimation de H, à partir d'une seule réalisation de longueur finie du processus, il est nécessaire d'estimer la quantité $\mathbb{E}\log_2|d_x(j,k)|$. Les propriétés **P2** (stationnarité) et **P3** (faible dépendance) conduisent naturellement à proposer l'estimateur :

$$Y_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\log_2|d_x(j,k)|,$$

où n_j est le nombre de coefficients d'ondelettes disponibles à l'octave j. On définit alors l'estimateur du paramètre H par la régression linéaire suivante :

$$\hat{H} = \sum_j w_j Y_j - 1/2, \tag{5}$$

où $\sum_j$ désigne la somme sur la gamme d'octaves $\{j_1,\ldots,j_2\}$ où se fait la régression linéaire et les poids w_j satisfont les relations usuelles, $\sum_j j w_j = 1$ et $\sum_j w_j = 0$, et sont définis par : $w_j = (1/a_j)(S_0 j - S_1)/(S_0 S_2 - S_1^2)$ avec $S_m = \sum_{j=j_1}^{j_2} a_j^{-1} j^m$ $(m = 0,1,2)$ et les a_j sont des nombres arbitraires. On sait que l'estimateur est de variance minimale si les a_j sont proportionnels aux variances des Y_j.

Il est immédiat de vérifier que $\hat{H}$ est un estimateur non biaisé de H et ce même pour une observation de durée finie du processus [4]. Sa variance peut être majorée analytiquement par une borne impliquant α, H et N [11, 4, 9] :

$$\mathrm{Var}\ \hat{H} \leq C n^{-1/(1+1/(\alpha(N-H)))}.$$

Cependant, en supposant l'indépendance (du log) des coefficients d'ondelettes (idéalisation **ID** ci-dessus), on peut déduire que la variance de $\hat{H}$ se comporte comme :

$$\mathrm{Var}\ \hat{H} \simeq (\log_2(e))^2\pi^2(1+2/\alpha^2)(\sum_j w_j^2/n_j)/12. \tag{6}$$

Ce résultat montre que Var $\hat{H}$ ne dépend pas de la valeur H du paramètre estimé, mais dépend de α, d'où l'intérêt de l'estimer. Cette expression de la variance est de plus minimale si $(\sum_j w_j^2/n_j)$ est minimale, ce qui est obtenu si $a_j \propto \mathrm{Var}\ Y_j$. Or, sous **ID**, $VarY_j \propto n_j^{-1}$. C'est ce choix qui est systématiquement retenu dans l'implantation pratique de l'estimateur $\hat{H}$. Avec ce choix, et comme $n_j \simeq 2^{-j}n$, où n est la longueur du processus étudié, on obtient :

$$\mathrm{Var}\ \hat{H} \simeq \left((\log_2(e))^2\pi^2(1+2/\alpha^2)(\sum_j w_j^2 2^j)/12\right)/n.$$

L'estimateur $\hat{H}$ exhibe donc une variance qui décroît comme $1/n$, en dépit de dépendance à longue portée dans le processus, un résultat non trivial [7]. Dans les simulations numériques décrites ci-dessous et rapportées sur les figures, la courbe de variance en pointillé a été tracé à partir de l'expression approchée (6). Notons, de plus, que cette variance ne dépend pas du paramètre à estimer H, elle dépend, par contre de α.

Simulations numériques. Une étude numérique a été effectuée sur *nbreal* réalisations de n échantillons du LFSM, pour divers couples de paramètres (α, H). Elles sont obtenues par évaluation numérique de l'intégrale (1), conformément à l'algorithme de synthèse proposé dans [17], chapitre 7. Dans les essais réalisés, $\alpha = 0.6, 0.8, 1, 1.25, 1.50, 1.75$ et $d = \pm 0.22$, pour les LFSM, de façon à étudier des situations à courte $(d < 0)$ ou longue $(d > 0)$ dépendance avec un même $|H - 1/\alpha|$, $nbreal = 80$, $2^8 \leq n \leq 2^{15}$, $j_1 = 3$, $j_2 = \log_2(n) - 1$. Les ondelettes utilisées sont des Daubechies, avec différents nombres de moments nuls. Les figures 1 et 2 comparent biais et variance prédits théoriquement à ceux observés dans les simulations numériques.

Les simulations numériques confirment le non biais de $\hat{H}$, cf. figure 1, pour tous les couples (α, H) envisagés. L'accord entre variance observée dans les simulations numériques et prédite en supposant l'idéalisation **ID** ci-dessus est peu influencé par la présence ou non de dépendance à longue portée (d positif ou négatif) mais se dégrade significativement lorsque α décroît, cf. figure 2.

3.2 Estimation de α

Pour estimer α, l'idée repose sur le constat que l'estimateur de H présente d'excellentes performances pour l'estimation du paramètre H des Lévy-stables. Dans ce cas particulier, $H = 1/\alpha$ et l'estimateur de H mesure donc également α. Pour des LFSM ou d'autres processus autosimilaires α-stable, ceci n'est plus vrai $(H \neq 1/\alpha)$ du fait de la non indépendance des accroissements du processus. L'idée est alors de *détruire* la structure de dépendance des accroissements en effectuant une permutation aléatoire de ceux-ci, et donc heuristiquement, de transformer un processus à accroissements fortement dépendants en un processus à accroissements indépendants. Soient $y_1(k) = x(k+1) - x(k)$ les accroissements du processus x, soit π une permutation des n échantillons dont on dispose, tirée suivant une loi uniforme parmi les $n!$ permutations possibles. On obtient ainsi $y_1'(k) = y_1(\pi_k)$ et on calcule ensuite la somme cumulée de ce processus : $x'(k) = \sum_{\ell=0}^{k} y_1'(\ell)$. On estime alors α en déterminant H comme précédemment, à partir des coefficients en ondelettes $d_x'(j,k)$ du processus x' et en utilisant $\widehat{1/\alpha} = \hat{H}$.

Pour étudier théoriquement les performances statistiques de cet estimateur, il faudrait justifier que les processus produits par la procédure de permutation des accroissements convergent vers des Lévy-stables ou tout au moins vers des processus à accroissements indépendants. Il est aisé de vérifier que la marginale de y' est α-stable. Il est également possible de franchir une première étape en montrant que n'importe quelle paire (y_l', y_k') (même pour k et l voisins) tend à être statistiquement indépendante lorsqu'on fait tendre le nombre d'échantillons sur lequel est réalisée la permutation (donc le nombre déchantillons de la réalisation analysée) vers l'infini (i.e., $n \to \infty$) [16]. Cette étude de performance est complétée par les simulations numériques décrites ci-dessus. En terme de biais, les simulations numériques montrent la présence d'un biais systématique. Celui-ci diminue lentement quand $n \to \infty$, i.e., asymptotiquement. Elles indiquent également que le biais augmente significativement, à n fixé, lorsque α décroît. Enfin, le biais, à n et α fixés, est nul pour les Lévy-stables (équivalent à $d = 0$), visible mais faible pour $d < 0$ et significativement plus important pour $d > 0$, i.e., en situation de dépendance à longue portée.

Références

[1] P. Abry, P. Gonçalvès and P. Flandrin, Wavelets, spectrum estimation and $1/f$ processes, in A. Antoniadis and G. Oppenheim, eds, *Wavelets and Statistics, Lectures Note in Statistics* **103**, pp. 15–30. Springer-Verlag, New York, 1995.

[2] P. Abry, D. Veitch, Wavelet analysis of long-range dependent traffic, *IEEE Trans. on Info. Theory*, 44(1):2–15, 1998.

[3] P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch, Wavelets for the analysis, estimation and synthesis of scaling data. To appear in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, eds., Wiley Interscience, 1999.

[4] P. Abry, L. Delbeke and P. Flandrin, Wavelet-based estimator for the self-similarity parameter of α-stable processes. *IEEE-ICASSP-99*, Phoenix (AZ), 1999.

[5] P. Abry, B. Pesquet-Popescu, P. Flandrin, L. Delbeke Wavelet Analysis of α-Stable Self-Similar Processes. Soumis à *Statistical Inference for Stochastic Processes*, mai 1999.

[6] R. Averkamp, C. Houdré, Some distributional properties of the continuous wavelet transform of random processes. *IEEE Trans. on Info. Theory*, Vol. 44, no. 3, pp. 1111–1124, May 1998.

[7] J.Beran, *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.

[8] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia (PA), 1992.

[9] L. Delbeke, *Wavelet based estimators for the scaling index of a self-similar with stationary increments*, PhD Thesis, KU Leuven, Belgium, 1998.

[10] L. Delbeke, P. Abry, Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion, submitted to *Stochastic Processes and their Applications*, preprint, 1997.

[11] L. Delbeke, J. Segers, The covariance of the logarithm of jointly symmetric stable random variables, preprint, 1998.

[12] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Boston, 1997.

[13] C.L. Nikias, M. Shao, *Signal processing with Alpha-Stable distributions and applications*. John Wiley and Sons, Inc., New York, 1995.

[14] B. Pesquet-Popescu. *Modélisation bidimensionnelle de processus non stationnaires et application à l'étude du fond sous-marin*, thèse de l'Ecole Normale Supérieure de Cachan, 1998.

[15] B. Pesquet-Popescu. Statistical properties of the wavelet decomposition of some non-gaussian self-similar processes. *Signal Processing* 75(3):, 1999.

[16] B. Pesquet-Popescu, M.S. Taqqu, P. Abry Permutations aléatoires des accroissements d'un mouvement linéaire fractionnaire stable. En préparation.

[17] G. Samorodnitsky, M. S. Taqqu, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, London, 1994.

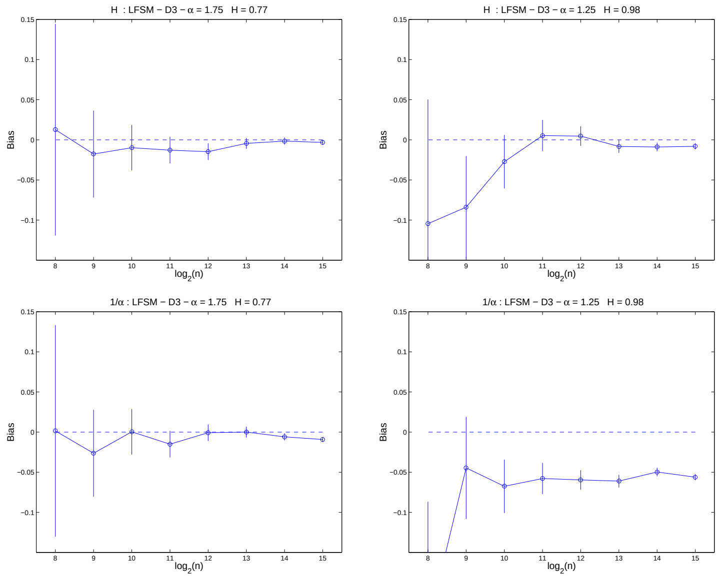

FIG. 1: **Biais d'estimation fonction de n.** *Valeurs estimées de H (haut) et $1/\alpha$ (bas) pour un LFSM $(\alpha, H) = (1.75, 0.77)$ (gauche) et $(\alpha, H) = (1.25, 0.98)$ (droite) pour différentes valeurs de n. Les barres verticales indiquent des intervalles de confiance (95%) de ces estimations. L'estimation de H est sans biais quelque soit H et α, au contraire, celle de α présente un biais léger qui croît avec d et lorsque α décroît.*

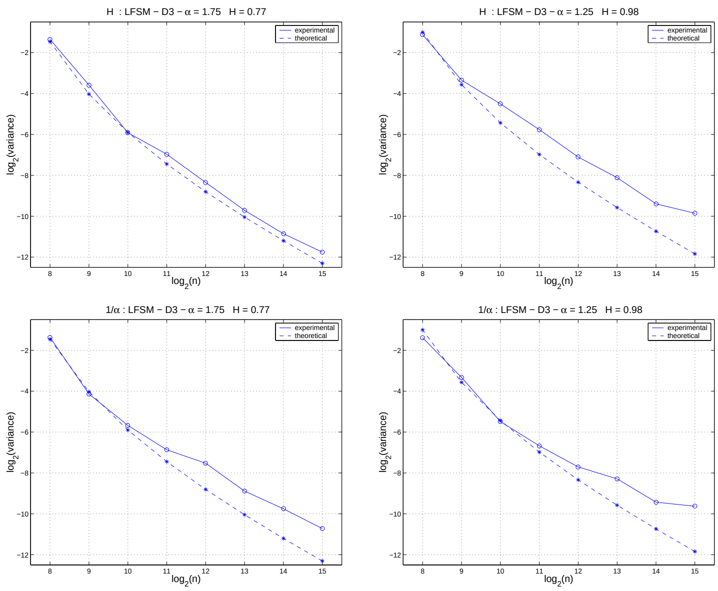

FIG. 2: **Variances d'estimation fonction de n.** *Variance des estimées de H (haut) et $1/\alpha$ (bas) pour un LFSM $(\alpha, H) = (1.75, 0.77)$ (gauche) et $(\alpha, H) = (1.25, 0.98)$ (droite) pour différentes valeurs de n. Le trait pointillé correspond au calcul analytique obtenu en supposant la décorrélation du log des coefficients d'ondelettes* **ID**, *le trait plein aux simulations numériques. L'accord est assez satisfaisant pour justifier l'usage de cette approximation, mais se détériore lorsque α diminue. On observe bien une décroissance en $1/n$ de la variance, qui est non triviale pour des processus auto-similaires avec dépendance à longue portée.*