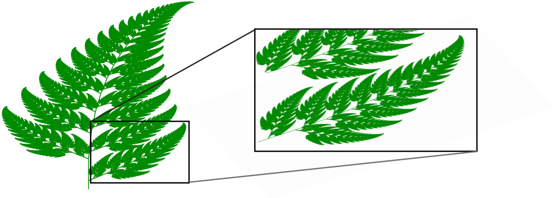
Autosimilarité multivariée : estimation des exposants d'autosimilarité, tests bootstrap d'égalité des exposants et applications

Charles-Gérard Lucas
Soutenance de thèse
19 octobre 2023

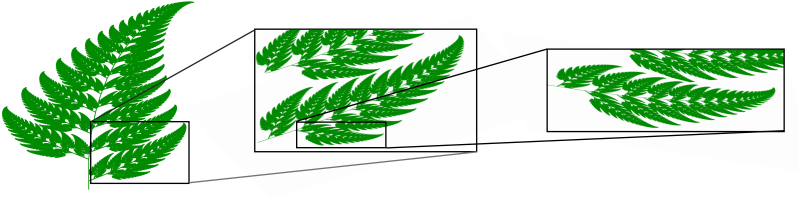
Patrice Abry CNRS, Laboratoire de Physique, ENS de Lyon
Herwig Wendt CNRS, IRIT, Université de Toulouse
Gustavo Didier Université Tulane, Nouvelle-Orléans, USA

Invariance d'échelle

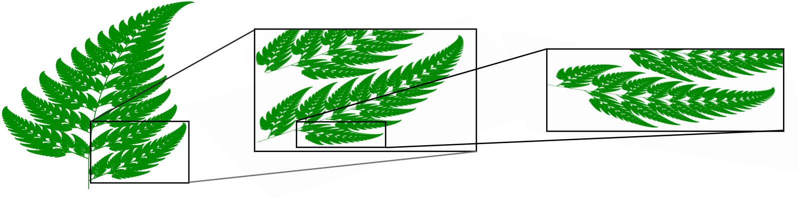
Invariance d'échelle

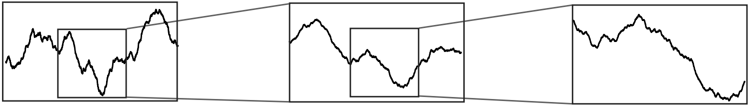


Invariance d'échelle



Invariance d'échelle



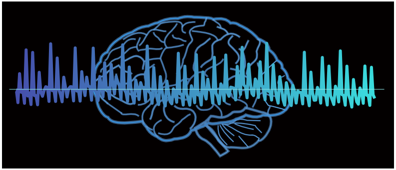


Invariance d'échelle dans le monde réel

Cours de la bourse

Source: Julien Eichinger - stock.adobe.com

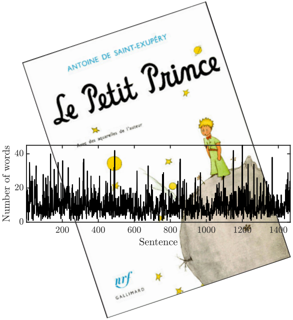
Activité cérébrale



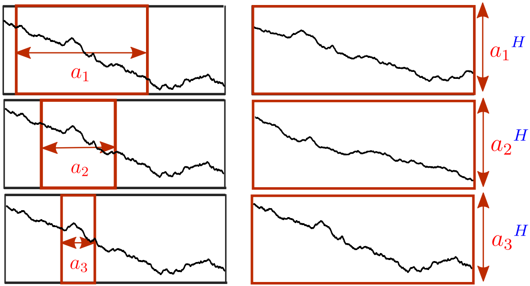
Traffic internet

Source: freepik.com

Stylométrie



Autosimilarité : définition



$$\{\mathcal{B}_H(t)\}_{t\in\mathbb{R}} \stackrel{fdd}{=} \left\{a^H \mathcal{B}_H(t/a)\right\}_{t\in\mathbb{R}}, \quad \forall a > 0$$

→ paramètre d'autosimilarité $0 < H < 1$

Objectif : Estimation de H à partir d'un signal de taille finie (N)

Autosimilarité : modèle

Mouvement brownien fractionnaire

Définition

Processus gaussien $\mathcal{B}_H$ autosimilaire de paramètre H

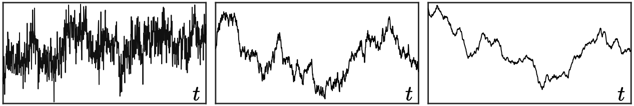
- Loi : $\mathcal{B}_H(t) \sim \mathcal{N}(0, \sigma_t^2), \quad \mathcal{B}_H(0) = 0$
- Variance : $\sigma_t^2 = \mathbb{E}\left[\mathcal{B}_H(t)\mathcal{B}_H(t)\right] = C_H |t|^{2H}$
- Accroissements stationnaires

→ Fonction de covariance :

$$\mathbb{E}\left[\mathcal{B}_H(t)\mathcal{B}_H(s)\right] = \frac{C_H}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)$$

→ Rugosité (Hölder) $h = H$

$\mathcal{B}_H(t)$

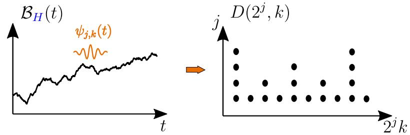


$H = 0.1$ $H = 0.5$ $H = 0.7$

Autosimilarité : analyse

Transformée en ondelettes

$$D(2^j, k) = \langle {\color{orange}\psi_{j,k}} | \mathcal{B}_{\color{blue}H} \rangle$$



$${\color{orange}\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)}$$

↳ ondelette mère ψ_0 translatée et dilatée

Spectre d'ondelettes

Variance à l'échelle 2^j

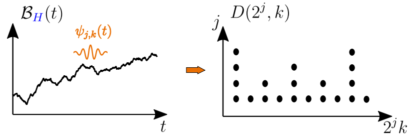
$$S(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D(2^j, k)^2$$

$n_j = N/2^j$, N : taille d'échantillon

Autosimilarité : analyse

Transformée en ondelettes

$$D(2^j, k) = \langle \psi_{j,k} | \mathcal{B}_H \rangle$$



$$\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$$

↳ ondelette mère ψ_0 translatée et dilatée

Spectre d'ondelettes

Variance à l'échelle 2^j

$$S(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D(2^j, k)^2$$

$n_j = N/2^j$, N : taille d'échantillon

$\mathcal{B}_H$ autosimilaire

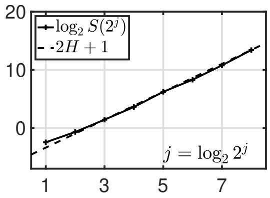
$\Rightarrow$ loi de puissance : $S(2^j) \propto (2^j)^{(2H+1)}$

[Flandrin, 1992]

$\rightarrow$ Régression linéaire en log-log

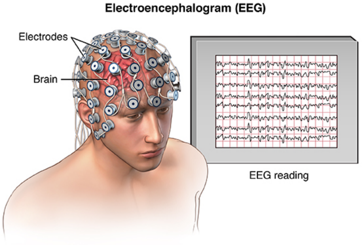
$$\hat{H} = \frac{1}{2}\sum_{j=j_1}^{j_2} \omega_j \log_2 S(2^j) - \frac{1}{2}$$

[Abry et Veitch, 1998]

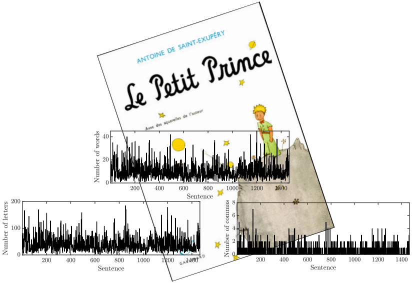


Données multivariées

Collection de signaux décrivant un même système



Source: speakingofresearch.com/tag/eeg/



→ M analyses univariées ?

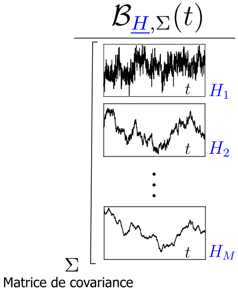
→ 1 analyse multivariée ?

Gliederung

Autosimilarité multivariée : extension naturelle

Collection de M mouvements browniens fractionnaires corrélés

[Mason et Xiao, 2002; Amblard et Cœurjolly, 2011]



- Modélisation : $\underline{H} = (H_1, \dots, H_M)$, Σ
- M relations d'autosimilarité univariées :

$$\{(\mathcal{B}_{\underline{H},\Sigma}(t))_m\}_{t\in\mathbb{R}} \overset{fdd}{=} \{a^{H_m}(\mathcal{B}_{\underline{H},\Sigma}(t/a))_m\}_{t\in\mathbb{R}}, \quad \forall a > 0$$

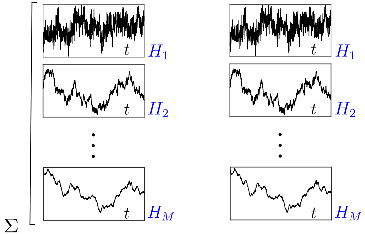
Autosimilarité multivariée : extension naturelle

Collection de M mouvements browniens fractionnaires corrélés

[Mason et Xiao, 2002; Amblard et Cœurjolly, 2011]

Matrice de mélange $W = \mathbb{I}$ $\times$ $\mathcal{B}_{\underline{H},\Sigma}(t)$ $=$ $\mathcal{B}_{\underline{H},\Sigma,W}(t)$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$



Matrice de covariance

- Modélisation : $\underline{H} = (H_1, \ldots, H_M)$, Σ, $W = \mathbb{I}$
- M relations d'autosimilarité univariées :

$$\{(\mathcal{B}_{\underline{H},\Sigma}(t))_m\}_{t\in\mathbb{R}} \overset{fdd}{=} \{a^{H_m}(\mathcal{B}_{\underline{H},\Sigma}(t/a))_m\}_{t\in\mathbb{R}}, \quad \forall a > 0$$

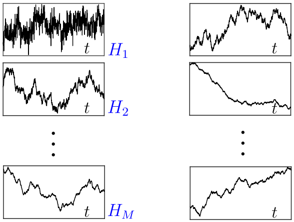
Autosimilarité multivariée : modèle

Mouvement brownien fractionnaire multivarié

[Abry et al., 2019]

Matrice de mélange W $\times$ $\mathcal{B}_{\underline{H},\Sigma}(t)$ $=$ $\mathcal{B}_{\underline{H},\Sigma,W}(t)$

$$\begin{bmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,M} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ W_{M,1} & W_{M,2} & \cdots & W_{M,M} \end{bmatrix}$$



Σ
Matrice de covariance

- Modélisation : $\underline{H} = (H_1, \ldots, H_M)$, Σ, W réelle et inversible
- Relation d'autosimilarité multivariée :

$$\{\mathcal{B}_{\underline{H},\Sigma,W}(t)\}_{t\in\mathbb{R}} \overset{fdd}{=} \left\{a^{\underline{\underline{H}}}\mathcal{B}_{\underline{H},\Sigma,W}(t/a)\right\}_{t\in\mathbb{R}}, \quad \forall a > 0$$

$$a^{\underline{\underline{H}}} = \sum_{k\geq 0} \frac{1}{k!} \log^k a \, \underline{\underline{H}}^k, \qquad \underline{\underline{H}} = W \,\mathrm{diag}(\underline{H})\, W^{-1}$$

Autosimilarité multivariée : modèle

Mouvement brownien opérateur-fractionnaire [Didier et Pipiras, 2011]

$$\mathcal{B}_{\underline{\underline{H}},A}(t) = \int_{\mathbb{R}} \frac{e^{itf}-1}{if} \left(f_+^{-(\underline{\underline{H}}-\frac{1}{2}\mathbb{I})} A + f_-^{-(\underline{\underline{H}}-\frac{1}{2}\mathbb{I})} \overline{A} \right) \tilde{B}(\mathrm{d}f)$$

- $f \in \mathbb{R}$, $\mathbb{E}[\tilde{B}(\mathrm{d}f)\tilde{B}(\mathrm{d}f)^\star] = \mathrm{d}f$
- $\underline{\underline{H}}$: matrice complexe de taille $M \times M$
- A : matrice complexe de taille $M \times M$, $\det AA^\star > 0$

Mouvement brownien fractionnaire multivarié

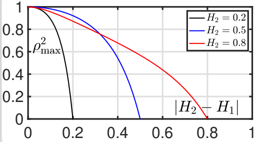
$$\mathcal{B}_{\underline{H},\Sigma,W} = \mathcal{B}_{\underline{\underline{H}},A} \quad \text{avec} \quad \begin{cases} \underline{\underline{H}} = W \,\mathrm{diag}(\underline{H})\, W^{-1} \\ AA^\star = W\,(G \odot \Sigma)\, W^\top \end{cases}$$

- $G_{m,m'} = \frac{1}{2\pi}\Gamma(H_m + H_{m'} + 1)\sin\left((H_m + H_{m'})\frac{\pi}{2}\right)$
- $\det AA^\star > 0 \Rightarrow \det G \odot \Sigma > 0$

[Lucas et al., 2023a]

$M = 2$

Corrélation maximale

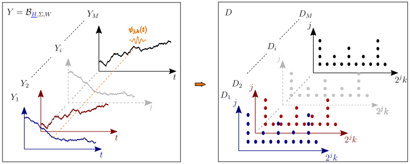


Gliederung

Autosimilarité multivariée : analyse

Transformée en ondelettes multivariée

$$D(2^j,k) = (D_1(2^j,k),\ldots,D_M(2^j,k))$$



Spectre d'ondelettes

Covariance à l'échelle 2^j

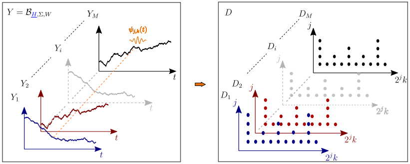
$$S(2^j) = \begin{bmatrix} S_{1,1}(2^j) & \cdots & S_{1,M}(2^j) \\ \vdots & \ddots & \vdots \\ S_{M,M}(2^j) & \cdots & S_{M,1}(2^j) \end{bmatrix}$$

$$S_{m,m'}(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D_m(2^j,k)D_{m'}(2^j,k)$$

Autosimilarité multivariée : analyse

Transformée en ondelettes multivariée

$$D(2^j, k) = (D_1(2^j, k), \ldots, D_M(2^j, k))$$



Spectre d'ondelettes

Covariance à l'échelle 2^j

$$S(2^j) = \begin{bmatrix} S_{1,1}(2^j) & \cdots & S_{1,M}(2^j) \\ \vdots & \ddots & \vdots \\ S_{M,M}(2^j) & \cdots & S_{M,1}(2^j) \end{bmatrix}$$

$$S_{m,m'}(2^j) = \frac{1}{n_j} \sum_{k=1}^{n_j} D_m(2^j, k) D_{m'}(2^j, k)$$

Cas $W = \mathbb{I}$ [Wendt et al., 2018]

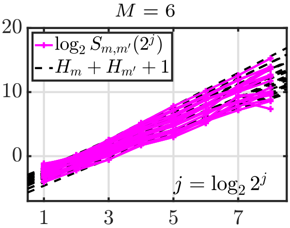
$\mathcal{B}_{\underline{H},\Sigma,W}$ autosimilaire

$\Rightarrow$ loi de puissance :

$$S_{m,m'}(2^j) \propto (2^j)^{(H_m + H_{m'} + 1)}$$

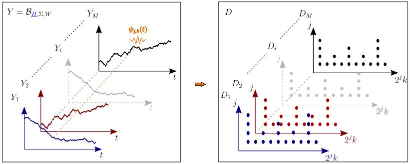
$\rightarrow$ Régression linéaire en log-log

$$\hat{H}_{m,m'} = \frac{1}{2} \sum_{j=j_1}^{j_2} \omega_j \log_2 S_{m,m'}(2^j) - \frac{1}{2}$$



Autosimilarité multivariée : analyse

Transformée en ondelettes multivariée
$$D(2^j, k) = (D_1(2^j, k), \ldots, D_M(2^j, k))$$



Spectre d'ondelettes

Covariance à l'échelle 2^j

$$S(2^j) = \begin{bmatrix} S_{1,1}(2^j) & \cdots & S_{1,M}(2^j) \\ \vdots & \ddots & \vdots \\ S_{M,M}(2^j) & \cdots & S_{M,1}(2^j) \end{bmatrix}$$

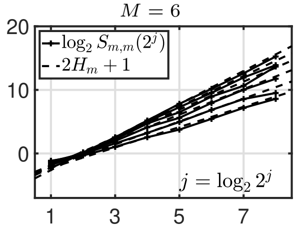
$$S_{m,m'}(2^j) = \frac{1}{n_j} \sum_{k=1}^{n_j} D_m(2^j, k) D_{m'}(2^j, k)$$

Cas $W = \mathbb{I}$ [Wendt et al., 2018]

$\mathcal{B}_{\underline{H},\Sigma,W}$ autosimilaire
$\Rightarrow$ loi de puissance :
$$S_{m,m}(2^j) \propto (2^j)^{(2H_m+1)}$$

$\rightarrow$ Régression linéaire en log-log

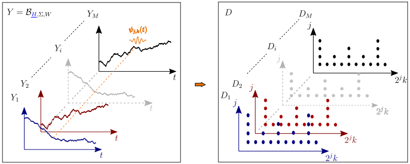
$$\hat{H}_m^{\mathrm{U}} = \frac{1}{2} \sum_{j=j_1}^{j_2} \omega_j \log_2 S_{m,m}(2^j) - \frac{1}{2}$$



Autosimilarité multivariée : analyse

Transformée en ondelettes multivariée

$$D(2^j,k) = (D_1(2^j,k),\dots,D_M(2^j,k))$$



Spectre d'ondelettes

Covariance à l'échelle 2^j

$$S(2^j) = \begin{bmatrix} S_{1,1}(2^j) & \cdots & S_{1,M}(2^j) \\ \vdots & \ddots & \vdots \\ S_{M,M}(2^j) & \cdots & S_{M,1}(2^j) \end{bmatrix}$$

$$S_{m,m'}(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D_m(2^j,k)D_{m'}(2^j,k)$$

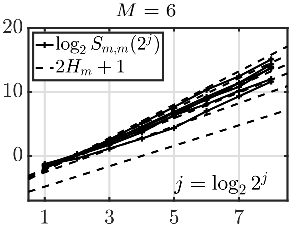
Cas $W \neq \mathbb{I}$ [Wendt et al., 2018]

$\mathcal{B}_{\underline{H},\Sigma,W}$ autosimilaire avec $W \neq \mathbb{I}$
$\Rightarrow$ mélange de M^2 lois de puissance :

$$S_{m,m'}(2^j) = \sum_{k,n} \alpha_{k,n}^{(m,m')}(2^j)^{(H_k+H_n+1)}$$

$\rightarrow$ Régression linéaire en log-log

$$\hat{H}_m^U = \frac{1}{2}\sum_{j=j_1}^{j_2} \omega_j \log_2 S_{m,m}(2^j) - \frac{1}{2}$$



Autosimilarité multivariée : analyse

Décomposition spectrale:

$$S(2^j) = U(2^j) \begin{bmatrix} \hat{\lambda}_1(2^j) & 0 & \cdots & \cdots & 0 \\ 0 & \hat{\lambda}_2(2^j) & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \hat{\lambda}_M(2^j) \end{bmatrix} U(2^j)^{-1}$$

Estimation multivariée [Abry et Didier, 2018a; 2018b]

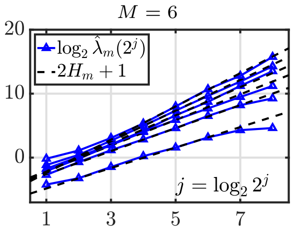
$\mathcal{B}_{\underline{H},\Sigma,W}$ autosimalaire

$\Rightarrow$ loi de puissance asymptotique :

$$\hat{\lambda}_m(2^j) \underset{j\to+\infty}{\propto} (2^j)^{(2H_m+1)}$$

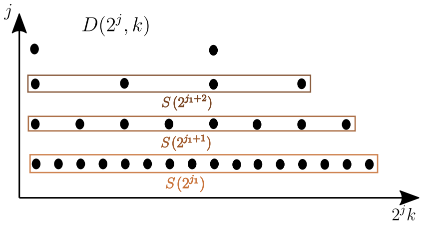
$\rightarrow$ Régression linéaire en log-log

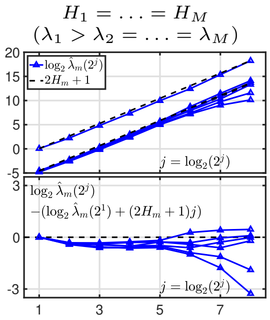
$$\hat{H}_m^{\mathrm{M}} = \frac{1}{2}\sum_{j=j_1}^{j_2} \omega_j \log_2 \hat{\lambda}_m(2^j) - \frac{1}{2}$$



Biais de répulsion

Nombre fini n_j de coefficients d'ondelettes disponibles
$\Rightarrow \hat{\lambda}_m(2^j) - \hat{\lambda}_{m'}(2^j) \gg \lambda_m(2^j) - \lambda_{m'}(2^j)$ pour $H_m = H_{m'}$ [Tao, 2012]





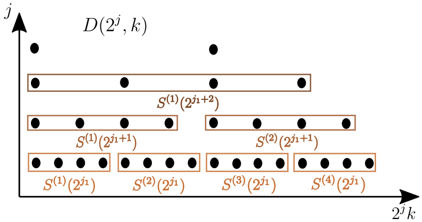
⚠ Nombre n_j de coefficients d'ondelettes disponibles varie avec 2^j
$\Rightarrow$ écart $\hat{\lambda}_{m'}(2^j) - \hat{\lambda}_m(2^j)$ croissant avec 2^j
$\Rightarrow$ biais dans les régressions linéaires

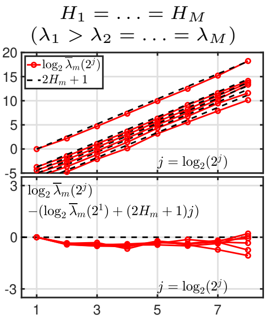
Correction de l'estimation

– Ré-échantillonnage par bloc bootstrap temps-échelle [Wendt et al., 2019]
– Reproduction du biais au travers des échelles :

$\rightarrow$ $S^{(w)}(2^j) = \frac{1}{n_{j_2}} \sum_{k \in F_w} D(2^j,k) D(2^j,k)^\top$

$\mathrm{Card}(F_w) = n_{j_2}, \; w = 1, \ldots, 2^{j_2 - j}$





$\rightarrow$ Valeurs propres $\hat{\lambda}_1^{(w)}(2^j), \ldots, \hat{\lambda}_M^{(w)}(2^j)$ de $S^{(w)}(2^j)$

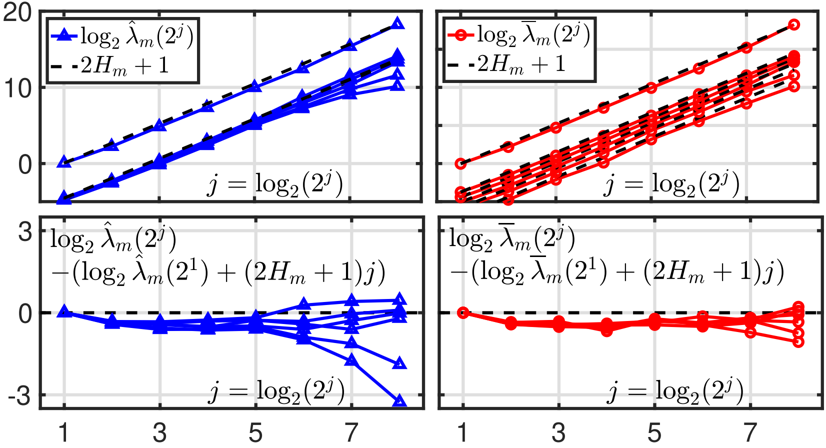
↳ $\hat{\lambda}_{m'}^{(w)}(2^j) - \hat{\lambda}_m^{(w)}(2^j)$ similaires à toutes les échelles 2^j

$\rightarrow$ Moyenne des logarithmes : $\log_2 \bar{\lambda}_m(2^j) = \langle \log_2 \hat{\lambda}_m^{(w)}(2^j) \rangle_w$

$\rightarrow$ Régression linéaire : $\hat{H}_m^{\mathrm{M,bc}} = \frac{1}{2} \sum_{j=j_1}^{j_2} \omega_j \log_2 \bar{\lambda}_m(2^j) - \frac{1}{2}$

Correction du biais de répulsion

Estimation non corrigée contre estimation corrigée



Étude des performances asymptotiques théoriques

Cadre asymptotique

Double limite :

- $N \to +\infty$
- $2^{j_1}, \ldots, 2^{j_2} \to +\infty$

Vitesse de convergence des $2^{j_1}, \ldots, 2^{j_2}$:

$$j_1 = j_1^0 + \log_2 a(N)$$
$$\vdots$$
$$j_2 = j_2^0 + \log_2 a(N)$$

avec

- $a(N) \sim N^\beta, \quad 1/(2\varpi + 1) < \beta < 1$
- $\varpi = \min\left\{\min_{\{2 \le i \le M |\ H_i - H_{i-1} > 0\}}(H_i - H_{i-1}), \frac{H_1}{2} + \frac{1}{4}\right\}$

[Abry et al., 2018a; 2018b]

Étude des performances asymptotiques théoriques

Consistance

Pour $\mathcal{B}_{\underline{H},\Sigma,W}$,

$\underline{\hat{H}}^{\mathrm{U}} \to \underline{H}$ si $W = \mathbb{I}$ [Wendt et al., 2018]

$\underline{\hat{H}}^{\mathrm{M}} \to \underline{H}$ [Abry et Didier, 2018a; 2018b]

$\underline{\hat{H}}^{\mathrm{M,bc}} \to \underline{H}$ [Lucas et al., 2023a]

Étude des performances asymptotiques théoriques

Consistance

Pour $\mathcal{B}_{\underline{H},\Sigma,W}$,

$\underline{\hat{H}}^{\mathrm{U}} \to \underline{H}$ si $W = \mathbb{I}$	[Wendt et al., 2018]
$\underline{\hat{H}}^{\mathrm{M}} \to \underline{H}$	[Abry et Didier, 2018a; 2018b]
$\underline{\hat{H}}^{\mathrm{M,bc}} \to \underline{H}$	[Lucas et al., 2023a]

Normalité asymptotique

Pour $\mathcal{B}_{\underline{H},\Sigma,W}$, si $\lambda_1(a(N)2^j) \neq \ldots \neq \lambda_M(a(N)2^j)$ lorsque $N \to +\infty$, alors

$$\sqrt{\frac{a(N)}{N}}\left(\underline{\hat{H}}^{\mathrm{M,bc}} - \underline{\hat{H}}\right) \to \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{B}})$$

[Lucas et al., 2023a]

Étude des performances asymptotiques théoriques

Consistance

Pour $\mathcal{B}_{\underline{H},\Sigma,W}$,

$$\underline{\hat{H}}^{\mathrm{U}} \to \underline{H} \text{ si } W = \mathbb{I}$$ [Wendt et al., 2018]

$$\underline{\hat{H}}^{\mathrm{M}} \to \underline{H}$$ [Abry et Didier, 2018a; 2018b]

$$\underline{\hat{H}}^{\mathrm{M,bc}} \to \underline{H}$$ [Lucas et al., 2023a]

Normalité asymptotique

Pour $\mathcal{B}_{\underline{H},\Sigma,W}$, si $\lambda_1(a(N)2^j) \neq \ldots \neq \lambda_M(a(N)2^j)$ lorsque $N \to +\infty$, alors

$$\sqrt{\frac{a(N)}{N}} \left(\underline{\hat{H}}^{\mathrm{M,bc}} - \underline{\hat{H}} \right) \to \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{B}})$$

[Lucas et al., 2023a]

- Effets de taille finie ?
- Impact des paramètres $\underline{H}$, Σ et W sur la corrélation $\Sigma_{\mathbf{B}}$?

Étude des performances empirique de taille finie

Simulations : 1000 réalisations, $M = 6$ composantes

- Tailles d'échantillon $N \in \{2^{13}, \ldots, 2^{18}\}$
- $2^{j_1} = a(N)2^6, \ldots, 2^{j_2} = a(N)2^9$ with $a(N) = 2^{\lfloor .9 \log_2 N/2^{13} \rfloor}$
- $\Sigma = \Sigma_0$
- 4 configurations :
  - $W \neq \mathbb{I}$ et $H_1 \neq \ldots \neq H_M$
  - $W \neq \mathbb{I}$ et $H_1 = \ldots = H_M$
  - $W = \mathbb{I}$ et $H_1 \neq \ldots \neq H_M$
  - $W = \mathbb{I}$ et $H_1 = \ldots = H_M$

Mesures de performance des estimateurs $\underline{\hat{H}}$:

- $s\mathrm{Bias} = (\mathbb{E}\underline{\hat{H}} - \underline{H})(\mathbb{E}\underline{\hat{H}} - \underline{H})^\top$
- $\mathrm{Cov} = \mathbb{E}[(\mathbb{E}\underline{\hat{H}} - \underline{\hat{H}})(\mathbb{E}\underline{\hat{H}} - \underline{\hat{H}})^\top]$
- $MSE = \mathbb{E}[(\underline{\hat{H}} - \underline{H})(\underline{\hat{H}} - \underline{H})^\top] = \mathrm{sBias} + \mathrm{Cov}$
  (erreur quadratique moyenne)

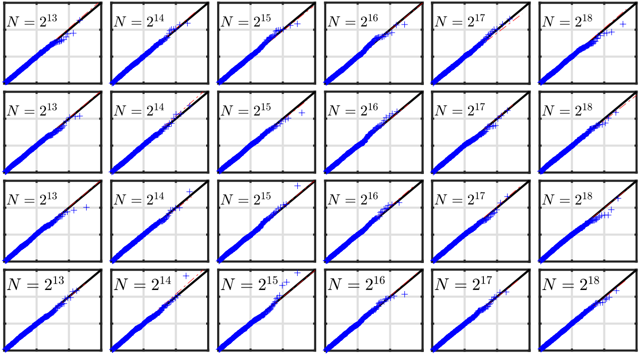
$\rightarrow$ norme spectrale $\|.\|_2$

Distribution normale multivariée de $\underline{\hat{H}}^{\mathrm{M,bc}}$

Diagramme quantile-quantile :

$$\left(\underline{\hat{H}}^{\mathrm{M,bc}} - \mathbb{E}\underline{\hat{H}}^{\mathrm{M,bc}}\right) \mathrm{Cov}\left(\underline{\hat{H}}^{\mathrm{M,bc}}\right)^{-1} \left(\underline{\hat{H}}^{\mathrm{M,bc}} - \mathbb{E}\underline{\hat{H}}^{\mathrm{M,bc}}\right)^{\top}$$

contre distribution du χ^2 à M degrés de liberté



$W \neq \mathbb{I}$
$H_1 \neq \ldots \neq H_M$

$W \neq \mathbb{I}$
$H_1 = \ldots = H_M$

$W = \mathbb{I}$
$H_1 \neq \ldots \neq H_M$

$W = \mathbb{I}$
$H_1 = \ldots = H_M$

$\rightarrow$ Normalité multivariée approximative de $\underline{\hat{H}}^{\mathrm{M,bc}}$, même à faible N

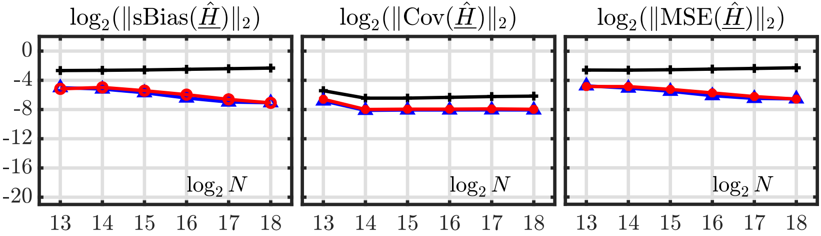
Performances d'estimation comparées

$$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



$\rightarrow$ $\underline{\hat{H}}^{\mathrm{U}}$ biaisé par le paramètre dominant H_M

$\rightarrow$ performances similaires entre $\underline{\hat{H}}^{\mathrm{M}}$ et $\underline{\hat{H}}^{\mathrm{M,bc}}$

$$\underline{\hat{H}}^{\mathrm{M,bc}} \sim \underline{\hat{H}}^{\mathrm{M}} \gg \underline{\hat{H}}^{\mathrm{U}}$$

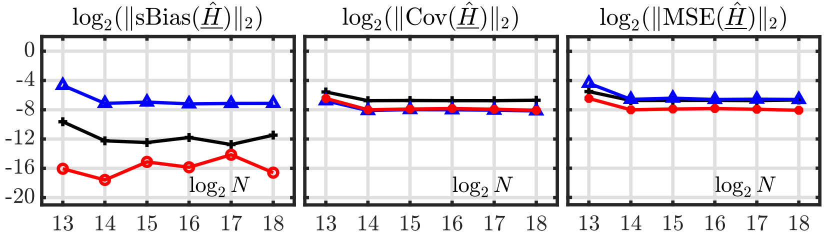
Performances d'estimation comparées

$$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 = \ldots = H_M \end{array}}$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



→ $\underline{\hat{H}}^{\mathrm{U}}$ moins biaisé que $\underline{\hat{H}}^{\mathrm{M}}$ car $H_1 = \ldots = H_M$

→ $\underline{\hat{H}}^{\mathrm{M,bc}}$ meilleur que $\underline{\hat{H}}^{\mathrm{M}}$: correction du biais de répulsion

$$\underline{\hat{H}}^{\mathrm{M,bc}} \geq \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$$

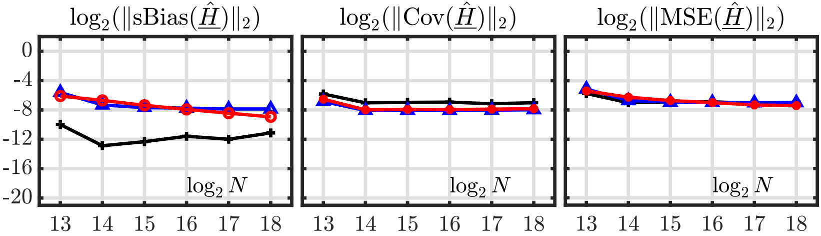
Performances d'estimation comparées

$$\boxed{\begin{array}{c} W = \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$$

$+$ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



$\rightarrow$ performances de $\underline{\hat{H}}^{\mathrm{M}}$ et $\underline{\hat{H}}^{\mathrm{M,bc}}$ similaires au cas $W \neq \mathbb{I}$

$\rightarrow$ compromis biais-covariance entre estimations univariée et multivariées

$$\underline{\hat{H}}^{\mathrm{M,bc}} \sim \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$$

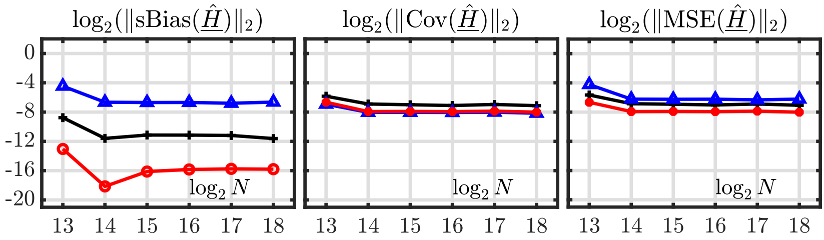
Performances d'estimation comparées

$$\boxed{\begin{matrix} W = \mathbb{I} \\ H_1 = \ldots = H_M \end{matrix}}$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



→ $\underline{\hat{H}}^{\mathrm{U}}$ moins biaisé que $\underline{\hat{H}}^{\mathrm{M}}$ car $H_1 = \ldots = H_M$

→ $\underline{\hat{H}}^{\mathrm{M,bc}}$ meilleur que $\underline{\hat{H}}^{\mathrm{M}}$: correction du biais de répulsion

$\Rightarrow$ similaire au cas $W \neq \mathbb{I}$ et $H_1 = \ldots = H_M$

$$\underline{\hat{H}}^{\mathrm{M,bc}} \geq \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$$

Synthèse des performances d'estimation

$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$	$\Rightarrow$	$\underline{\hat{H}}^{\mathrm{M,bc}} \sim \underline{\hat{H}}^{\mathrm{M}} \gg \underline{\hat{H}}^{\mathrm{U}}$
$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 = \ldots = H_M \end{array}}$	$\Rightarrow$	$\underline{\hat{H}}^{\mathrm{M,bc}} \geq \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$
$\boxed{\begin{array}{c} W = \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$	$\Rightarrow$	$\underline{\hat{H}}^{\mathrm{M,bc}} \sim \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$
$\boxed{\begin{array}{c} W = \mathbb{I} \\ H_1 = \ldots = H_M \end{array}}$	$\Rightarrow$	$\underline{\hat{H}}^{\mathrm{M,bc}} \geq \underline{\hat{H}}^{\mathrm{M}} \sim \underline{\hat{H}}^{\mathrm{U}}$

$\underline{\hat{H}}^{\mathrm{M,bc}}$ peut être utilisé dans toutes les configurations

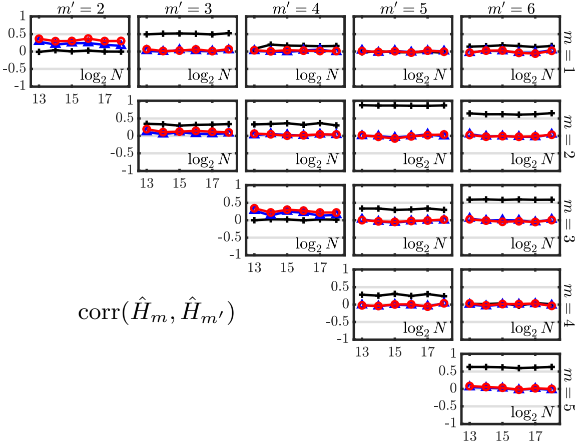
Corrélation des estimateurs

$$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$$

$+$ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



$\mathrm{corr}(\hat{H}_m, \hat{H}_{m'})$

$\rightarrow$ Quasi-décorrélation de $\underline{\hat{H}}^{\mathrm{M,bc}}$ et $\underline{\hat{H}}^{\mathrm{M}}$, même pour N faible

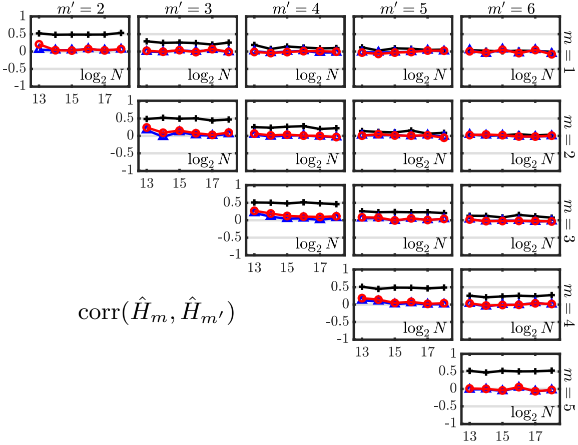
Corrélation des estimateurs

$$\boxed{\begin{array}{c} W = \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



$\mathrm{corr}(\hat{H}_m, \hat{H}_{m'})$

$\rightarrow$ Quasi-décorrélation de $\underline{\hat{H}}^{\mathrm{M,bc}}$ et $\underline{\hat{H}}^{\mathrm{M}}$, même pour N faible

Gliederung

Tests d'égalité

Fluctuation de l'estimation : $\hat{H}_m \neq \hat{H}_{m'}$ même si $H_m = H_{m'}$

Problème :

- Combien de paramètres distincts dans $\underline{H} = (H_1, \ldots, H_M)$?
- Combien de paramètres H_m prennent chacune de ces valeurs ?

Tests d'égalité

Fluctuation de l'estimation : $\hat{H}_m \neq \hat{H}_{m'}$ même si $H_m = H_{m'}$

Problème :

- Combien de paramètres distincts dans $\underline{H} = (H_1, \ldots, H_M)$?
- Combien de paramètres H_m prennent chacune de ces valeurs ?

Solutions :

- Tester $H_1 = \ldots = H_M$ [Lucas et al., EUSIPCO, 2021]

Tests d'égalité

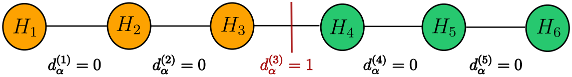
Fluctuation de l'estimation : $\hat{H}_m \neq \hat{H}_{m'}$ même si $H_m = H_{m'}$

Problème :

- Combien de paramètres distincts dans $\underline{H} = (H_1, \ldots, H_M)$?
- Combien de paramètres H_m prennent chacune de ces valeurs ?

Solutions :

- Tester $H_1 = \ldots = H_M$ [Lucas et al., EUSIPCO, 2021]
- Trier $\hat{H}_{\tau(1)} \leq \ldots \leq \hat{H}_{\tau(M)} \Rightarrow M - 1$ tests pour $H_m = H_{m+1}$ [Lucas et al., ICASSP, 2022; Lucas et al., GRETSI, 2022]



Tests d'égalité

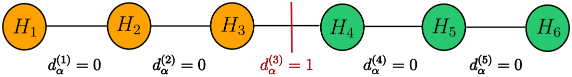
Fluctuation de l'estimation : $\hat{H}_m \neq \hat{H}_{m'}$ même si $H_m = H_{m'}$

Problème :

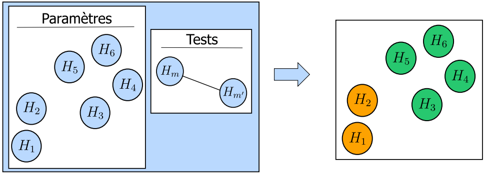
- Combien de paramètres distincts dans $\underline{H} = (H_1, \ldots, H_M)$?
- Combien de paramètres H_m prennent chacune de ces valeurs ?

Solutions :

- Tester $H_1 = \ldots = H_M$ [Lucas et al., EUSIPCO, 2021]
- Trier $\hat{H}_{\tau(1)} \leq \ldots \leq \hat{H}_{\tau(M)} \Rightarrow M - 1$ tests pour $H_m = H_{m+1}$ [Lucas et al., ICASSP, 2022; Lucas et al., GRETSI, 2022]



- $M(M-1)/2$ tests pour $H_m = H_{m'}$, $m \neq m'$ [Lucas et al., 2023]

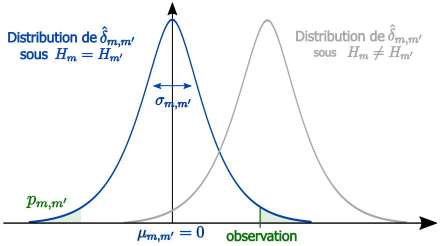


Test pour $H_m = H_{m'}$

- Formulation :
  - $M(M-1)/2$ hypothèses : $H_m = H_{m'}, \quad 1 \leq m < m' \leq M$
  - Estimées multivariées corrigées $\hat{H}_1^{\mathrm{M,bc}}, \ldots, \hat{H}_M^{\mathrm{M,bc}}$
  - Statistiques $\hat{\delta}_{m,m'} = \hat{H}_{m'}^{\mathrm{M,bc}} - \hat{H}_m^{\mathrm{M,bc}}$

Test pour $H_m = H_{m'}$

- Formulation :
  - $M(M-1)/2$ hypothèses : $H_m = H_{m'}, \quad 1 \leq m < m' \leq M$
  - Estimées multivariées corrigées $\hat{H}_1^{\mathrm{M,bc}}, \ldots, \hat{H}_M^{\mathrm{M,bc}}$
  - Statistiques $\hat{\delta}_{m,m'} = \hat{H}_{m'}^{\mathrm{M,bc}} - \hat{H}_m^{\mathrm{M,bc}}$
- Distribution de $\hat{\delta}_{m,m'}$:
  - approximativement gaussienne en taille finie
  - centrée sous $H_m = H_{m'}$

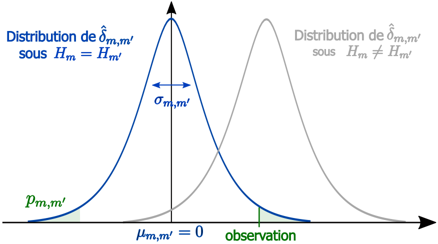


Test pour $H_m = H_{m'}$

- Formulation :
  - $M(M-1)/2$ hypothèses : $H_m = H_{m'}, \quad 1 \leq m < m' \leq M$
  - Estimées multivariées corrigées $\hat{H}_1^{\mathrm{M,bc}}, \ldots, \hat{H}_M^{\mathrm{M,bc}}$
  - Statistiques $\hat{\delta}_{m,m'} = \hat{H}_{m'}^{\mathrm{M,bc}} - \hat{H}_m^{\mathrm{M,bc}}$
- Distribution de $\hat{\delta}_{m,m'}$:
  - approximativement gaussienne en taille finie
  - centrée sous $H_m = H_{m'}$
- p-valeurs :

$$p_{m,m'} = 2\left(1 - F\left(\frac{|\hat{\delta}_{m,m'}|}{\sigma_{m,m'}}\right)\right)$$

F : fonction de répartition de la loi gaussienne standard

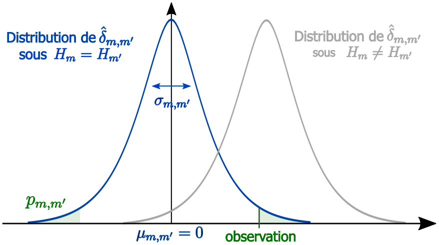


Test pour $H_m = H_{m'}$

- Formulation :
  - $M(M-1)/2$ hypothèses : $H_m = H_{m'}$, $\quad 1 \leq m < m' \leq M$
  - Estimées multivariées corrigées $\hat{H}_1^{\mathrm{M,bc}}, \ldots, \hat{H}_M^{\mathrm{M,bc}}$
  - Statistiques $\hat{\delta}_{m,m'} = \hat{H}_{m'}^{\mathrm{M,bc}} - \hat{H}_m^{\mathrm{M,bc}}$
- Distribution de $\hat{\delta}_{m,m'}$:
  - approximativement gaussienne en taille finie
  - centrée sous $H_m = H_{m'}$
- p-valeurs :

$$p_{m,m'} = 2\left(1 - F\left(\frac{|\hat{\delta}_{m,m'}|}{\sigma_{m,m'}}\right)\right)$$

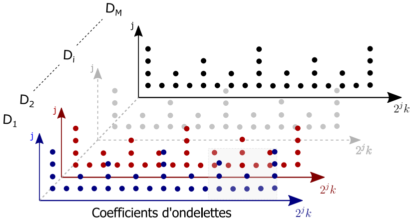
F : fonction de répartition de la loi gaussienne standard



⚠ Une seule observation $\Rightarrow \sigma_{m,m'}$ inconnu
$\rightarrow$ estimation de $\sigma_{m,m'}$ par bootstrap

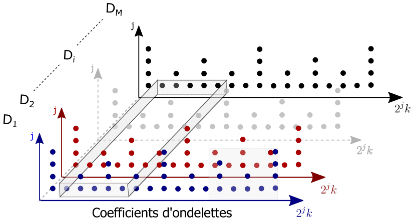
Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

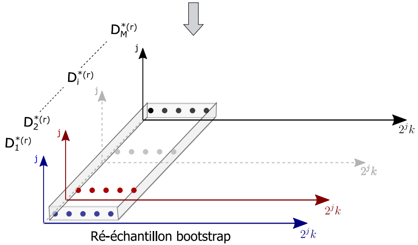
[Wendt et al., 2007; 2018]



Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

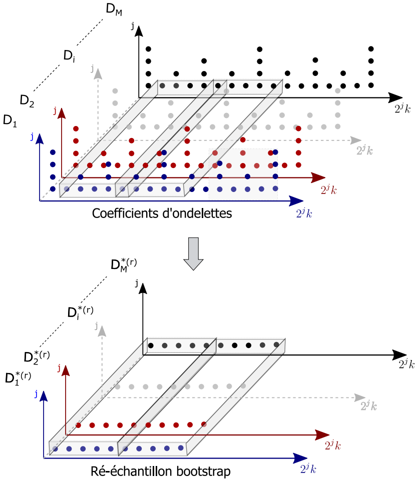
[Wendt et al., 2007; 2018]





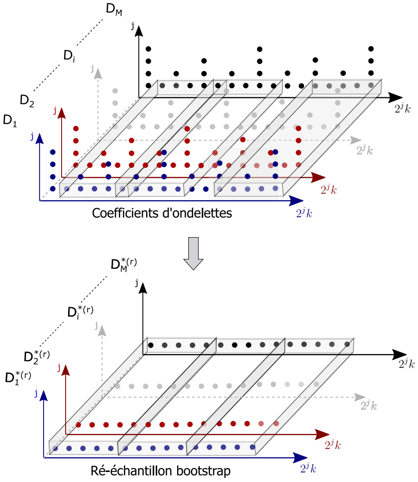
Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



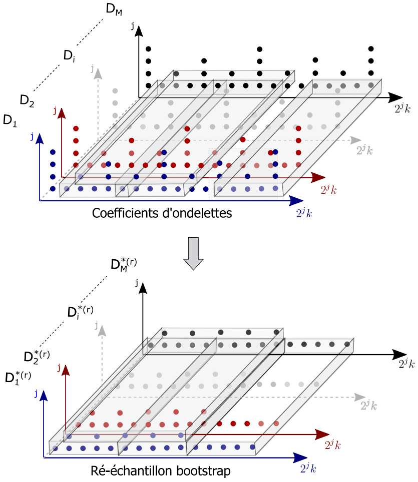
Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



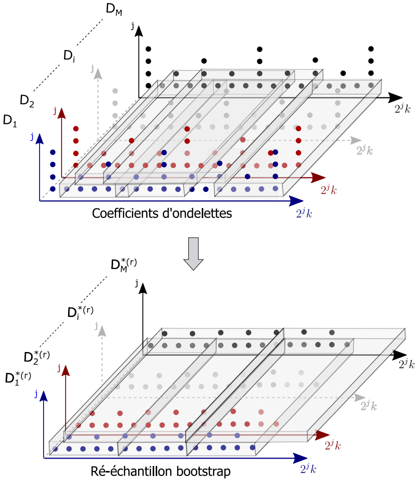
Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



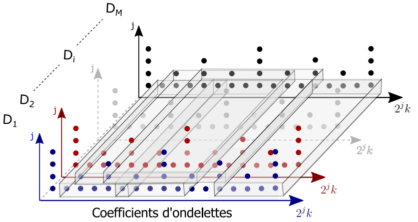
Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

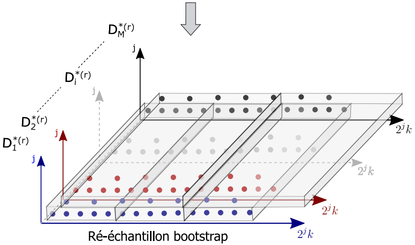
[Wendt et al., 2007; 2018]



Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]

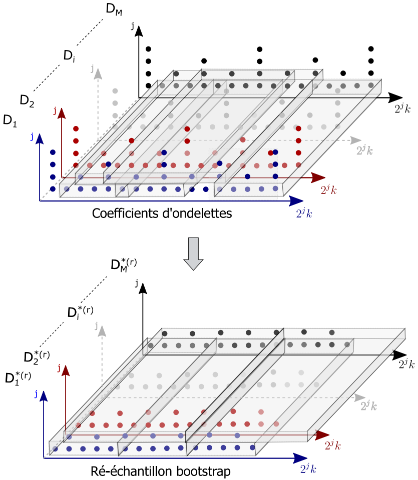




R ré-échantillons d'ondelettes
$D^{*(r)} = (D_1^{*(r)}, \dots, D_M^{*(r)})$

Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



R ré-échantillons d'ondelettes

$$D^{*(r)} = (D_1^{*(r)}, \ldots, D_M^{*(r)})$$

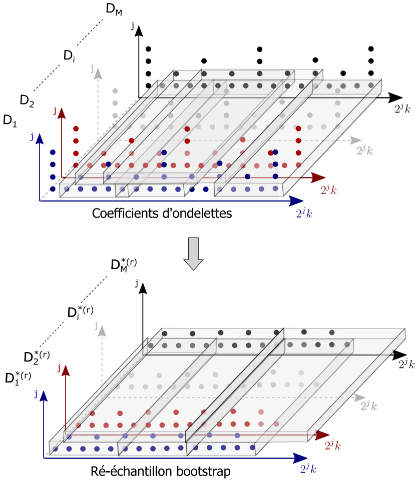
$$\Downarrow$$

R estimées bootstrap

$$\underline{\hat{H}}^{*(r)} = (\hat{H}_1^{*(r)}, \ldots, \hat{H}_M^{*(r)})$$

Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

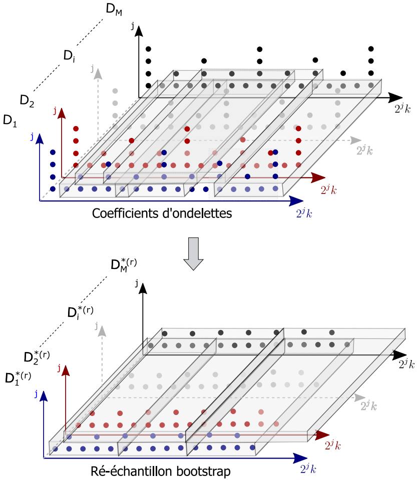
[Wendt et al., 2007; 2018]



R ré-échantillons d'ondelettes
$$D^{*(r)} = (D_1^{*(r)}, \dots, D_M^{*(r)})$$
$$\Downarrow$$
R estimées bootstrap
$$\underline{\hat{H}}^{*(r)} = (\hat{H}_1^{*(r)}, \dots, \hat{H}_M^{*(r)})$$
$$\Downarrow$$
R statistiques de test bootstrap
$$\hat{\delta}_{m,m'}^{*(r)} = \hat{H}_{m'}^{*(r)} - \hat{H}_m^{*(r)}$$

Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



R ré-échantillons d'ondelettes

$$D^{*(r)} = (D_1^{*(r)}, \dots, D_M^{*(r)})$$

$$\Downarrow$$

R estimées bootstrap

$$\underline{\hat{H}}^{*(r)} = (\hat{H}_1^{*(r)}, \dots, \hat{H}_M^{*(r)})$$

$$\Downarrow$$

R statistiques de test bootstrap

$$\hat{\delta}_{m,m'}^{*(r)} = \hat{H}_{m'}^{*(r)} - \hat{H}_m^{*(r)}$$

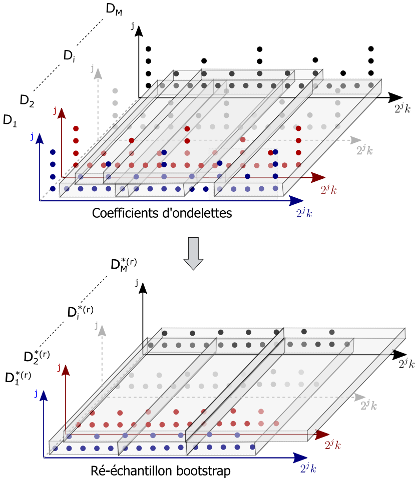
$$\Downarrow$$

Estimées de paramètre de test

$$\hat{\sigma}_{m,m'}^{*} = \sqrt{\mathrm{Var}^*(\hat{\delta}_{m,m'}^{*})}$$

Ré-échantillonnage par bloc bootstrap temps-échelle multivarié

[Wendt et al., 2007; 2018]



R ré-échantillons d'ondelettes

$$D^{*(r)} = (D_1^{*(r)}, \ldots, D_M^{*(r)})$$

$$\Downarrow$$

R estimées bootstrap

$$\underline{\hat{H}}^{*(r)} = (\hat{H}_1^{*(r)}, \ldots, \hat{H}_M^{*(r)})$$

$$\Downarrow$$

R statistiques de test bootstrap

$$\hat{\delta}_{m,m'}^{*(r)} = \hat{H}_{m'}^{*(r)} - \hat{H}_m^{*(r)}$$

$$\Downarrow$$

Estimées de paramètre de test

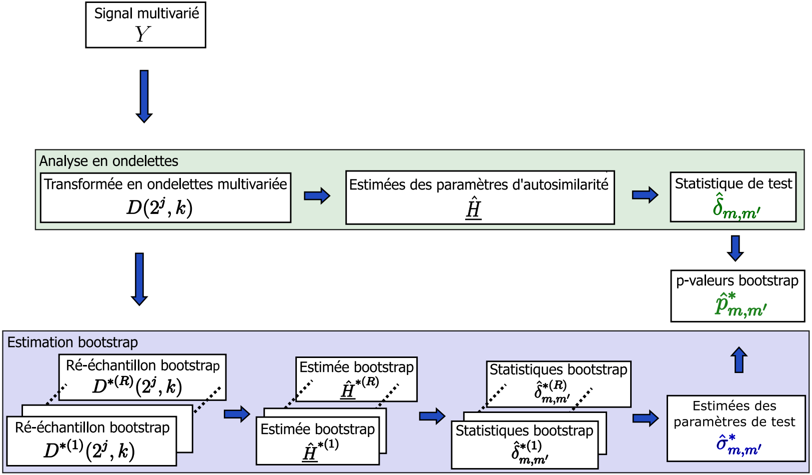
$$\hat{\sigma}_{m,m'}^{*} = \sqrt{\mathrm{Var}^*(\hat{\delta}_{m,m'}^{*})}$$

$$\Downarrow$$

p-valeurs bootstrap

$$\hat{p}_{m,m'}^{*}$$

Organigramme de la procédure de test



Validation de la procédure bootstrap

Simulations : 1000 réalisations, $M = 6$, $N = 2^{16}$

$$\underline{H} = (0.8, 0.8, 0.8, 0.8, 0.8, 0.8)$$

Estimation de $\sigma_{m,m'}$ par boostrap
(Moyenne de Monte Carlo $\pm$ intervalle de confiance à 95%)

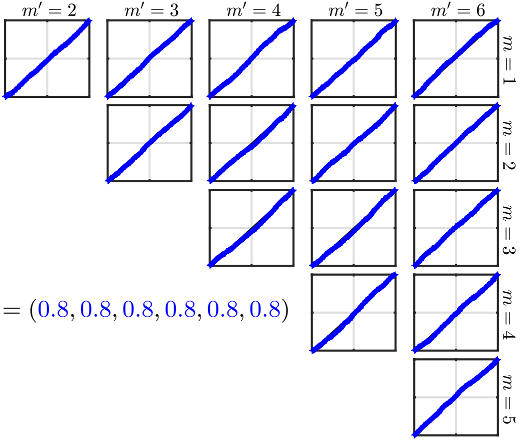
$\times 10^2$	2	3	4	5	6	m' / m
$\sigma_{m,m'}$	7.81	7.63	7.59	7.35	7.48	1
$\hat{\sigma}^*_{m,m'}$	7.59 $\pm$0.74	7.51 $\pm$0.76	7.42 $\pm$0.77	7.50 $\pm$0.77	7.65 $\pm$0.80	
$\sigma_{m,m'}$		6.83	6.84	6.86	7.00	2
$\hat{\sigma}^*_{m,m'}$		6.54 $\pm$0.57	6.70 $\pm$0.62	6.82 $\pm$0.65	6.98 $\pm$0.71	
$\sigma_{m,m'}$			5.82	6.36	6.43	3
$\hat{\sigma}^*_{m,m'}$			5.63 $\pm$0.53	6.19 $\pm$0.55	6.38 $\pm$0.62	
$\sigma_{m,m'}$				5.73	6.18	4
$\hat{\sigma}^*_{m,m'}$				5.47 $\pm$0.49	5.89 $\pm$0.58	
$\sigma_{m,m'}$					5.45	5
$\hat{\sigma}^*_{m,m'}$					5.39 $\pm$0.56	

$\rightarrow$ Reproduction de la distribution $\hat{\delta}_{m,m'}$ sous $H_m = H'_m$

Validation de la procédure de test

Simulations : 1000 réalisations, $M = 6$, $N = 2^{16}$

Diagramme quantile-quantile
$\hat{p}^*_{m,m'}$ contre distribution uniforme



$\hat{p}^*_{m,m'}$ uniforme
sous $H_m = H_{m'}$

$\underline{H} = (0.8, 0.8, 0.8, 0.8, 0.8, 0.8)$

Validation de la procédure bootstrap

Simulations : 1000 réalisations, $M = 6$, $N = 2^{16}$

$$\underline{H} = (0.4, 0.4, 0.6, 0.6, 0.8, 0.8)$$

Estimation de $\sigma_{m,m'}$ par boostrap
(Moyenne de Monte Carlo $\pm$ intervalle de confiance à 95%)

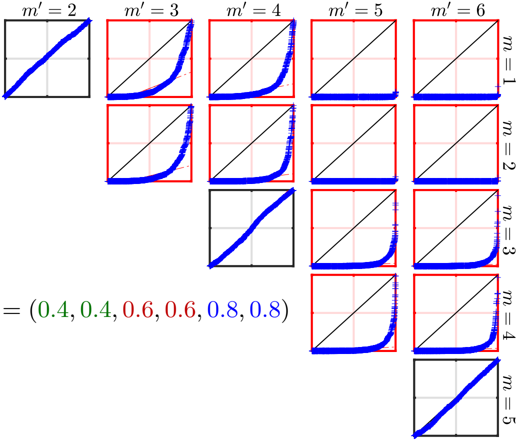
$\times 10^2$	2	3	4	5	6	m' / m
$\sigma_{m,m'}$	6.32	7.95	7.56	7.39	7.27	1
$\hat{\sigma}^*_{m,m'}$	6.83 $\pm$0.77	7.82 $\pm$0.86	7.73 $\pm$0.83	7.65 $\pm$0.82	7.69 $\pm$0.82	1
$\sigma_{m,m'}$		7.36	7.20	7.08	6.87	2
$\hat{\sigma}^*_{m,m'}$		7.11 $\pm$0.79	6.85 $\pm$0.73	6.79 $\pm$0.77	6.65 $\pm$0.74	2
$\sigma_{m,m'}$			7.12	7.38	7.23	3
$\hat{\sigma}^*_{m,m'}$			6.99 $\pm$0.75	7.18 $\pm$0.84	7.11 $\pm$0.83	3
$\sigma_{m,m'}$				7.20	7.03	4
$\hat{\sigma}^*_{m,m'}$				6.78 $\pm$0.75	6.67 $\pm$0.79	4
$\sigma_{m,m'}$					6.94	5
$\hat{\sigma}^*_{m,m'}$					6.39 $\pm$0.70	5

$\rightarrow$ Sous $H_m \neq H'_m$, reproduction de la distribution $\hat{\delta}_{m,m'}$ sous $H_m = H'_m$

Puissance du test

Simulations : 1000 réalisations, $M = 6$, $N = 2^{16}$

Diagramme quantile-quantile
$\hat{p}^*_{m,m'}$ contre distribution uniforme



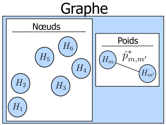
$\underline{H} = (0.4, 0.4, 0.6, 0.6, 0.8, 0.8)$

$\hat{p}^*_{m,m'}$ uniforme sous $H_m = H_{m'}$

$\hat{p}^*_{m,m'}$ s'écarte de la distribution uniforme avec $H_{m'} - H_m \neq 0$

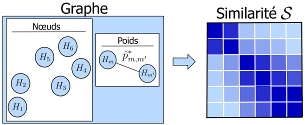
Stratégie de partitionnement par graphe

– Graphe : nœuds H_m, matrice de similarité $\mathcal{S} = (\hat{p}^*_{m,m'})_{m,m'}$



Stratégie de partitionnement par graphe

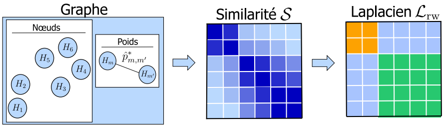
– Graphe : nœuds H_m, matrice de similarité $\mathcal{S} = (\hat{p}^*_{m,m'})_{m,m'}$



Stratégie de partitionnement par graphe

- Graphe : nœuds H_m, matrice de similarité $\mathcal{S} = (\hat{p}^*_{m,m'})_{m,m'}$
- Partitionnement spectral [Filippone et al., 2008]
  - Laplacien $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$

    Matrice des degrées (diagonale) D : $D_{m,m} = \sum_{m'=1}^{M} \mathcal{S}_{m,m'}$



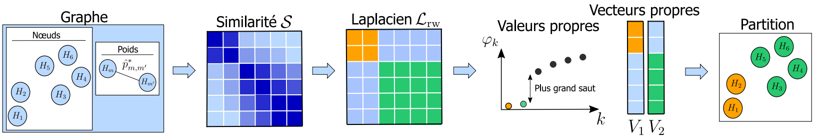
Stratégie de partitionnement par graphe

- Graphe : nœuds H_m, matrice de similarité $\mathcal{S} = (\hat{p}^*_{m,m'})_{m,m'}$
- Partitionnement spectral [Filippone et al., 2008]
  - Laplacien $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$

    Matrice des degrées (diagonale) D : $D_{m,m} = \sum_{m'=1}^{M} \mathcal{S}_{m,m'}$
  - Décomposition spectrale : $\mathcal{L}_{rw} = V \begin{bmatrix} \varphi_1 & 0 & \cdots & \cdots & 0 \\ 0 & \varphi_2 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \varphi_M \end{bmatrix} V^{-1}$

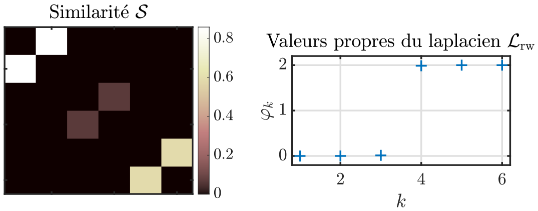
  → Estimation du nombre de groupes : $\hat{N}_{\mathrm{C}} = \underset{k}{\operatorname{argmax}}(\varphi_{k+1} - \varphi_k)$ [Azran et Ghahramani, 2006]

  → Partitionnement : k-moyenne sur $V_1, \ldots, V_{\hat{N}_{\mathrm{C}}}$

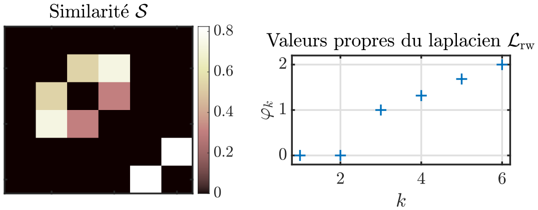


Statégie de partitionnement par graphe

$$\underline{H} = (0.4, 0.4, 0.6, 0.6, 0.8, 0.8)$$



$$\underline{H} = (0.4, 0.6, 0.6, 0.6, 0.8, 0.8)$$



Normalisation de $\mathcal{L}_{\mathrm{rw}} \rightarrow H_m$ uniques non détectés

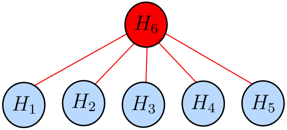
Détection de nœuds isolés

Suppression des H_m uniques avant partitionnement

→ Décisions de rejet de Benjamini-Hochberg :

rejeter $H_m = H_{m'}$ si $\hat{p}^*_{m,m'} < d_\alpha^{(m,m')}$

taux de fausses découvertes : $\alpha = 0.05$



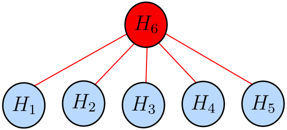
Détection de nœuds isolés

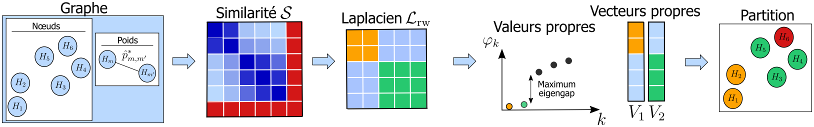
Suppression des H_m uniques avant partitionnement

$\rightarrow$ Décisions de rejet de Benjamini-Hochberg :

rejeter $H_m = H_{m'}$ si $\hat{p}^*_{m,m'} < d_\alpha^{(m,m')}$

taux de fausses découvertes : $\alpha = 0.05$





Performances d'estimation du nombre de groupes

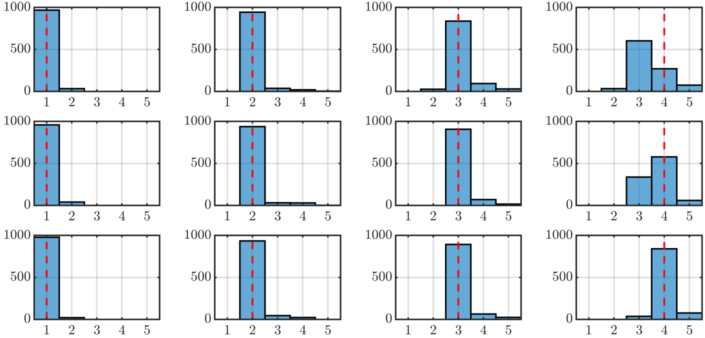
Simulations : 1000 réalisations, $M = 20$ composantes

Histogrammes des estimées $\hat{N}_C$ du nombre de partitions N_C

$N = 2^{17}$

$N = 2^{18}$

$N = 2^{19}$



$\underline{H} = ({\color{red}0.7, \ldots, 0.7})$ $\qquad M = {\color{red}20}$

$\underline{H} = ({\color{red}0.6, \ldots, 0.6}, {\color{blue}0.8, \ldots, 0.8})$ $\qquad M = {\color{red}10} + {\color{blue}10}$

$\underline{H} = ({\color{green}0.5, \ldots, 0.5}, {\color{red}0.6, \ldots, 0.6}, {\color{blue}0.7, \ldots, 0.7})$ $\qquad M = {\color{green}7} + {\color{red}7} + {\color{blue}6}$

$\underline{H} = ({\color{orange}0.4, \ldots, 0.4}, {\color{green}0.5, \ldots, 0.5}, {\color{red}0.6, \ldots, 0.6}, {\color{blue}0.7})$ $\qquad M = {\color{orange}7} + {\color{green}6} + {\color{red}6} + {\color{blue}1}$

Performances du partitionnement

NMI: information mutuelle normalisée
(*entropie conjointe des partitions exactes et estimées*)
ARI: indice de Rand ajusté
(*paires d'éléments correctement séparées ou rassemblées*)

Performances (entre 0 et 1)
(Moyenne de Monte Carlo $\pm$ intervalle de confiance à 95%)

		2 partitions	3 partitions	4 partitions
$N = 2^{17}$	NMI	0.99 ± 0.00	0.85 ± 0.01	0.83 ± 0.01
	ARI	0.98 ± 0.01	0.79 ± 0.01	0.75 ± 0.01
$N = 2^{18}$	NMI	0.99 ± 0.00	0.97 ± 0.00	0.95 ± 0.00
	ARI	0.98 ± 0.01	0.96 ± 0.01	0.92 ± 0.01
$N = 2^{19}$	NMI	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
	ARI	0.98 ± 0.01	0.97 ± 0.01	0.97 ± 0.01

$\underline{H} = (0.6, \dots, 0.6, 0.8, \dots, 0.8)$ $\quad M = 10 + 10$

$\underline{H} = (0.5, \dots, 0.5, 0.6, \dots, 0.6, 0.7, \dots, 0.7)$ $\quad M = 7 + 7 + 6$

$\underline{H} = (0.4, \dots, 0.4, 0.5, \dots, 0.5, 0.6, \dots, 0.6, 0.7)$ $\quad M = 7 + 6 + 6 + 1$

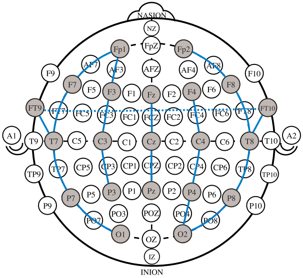
$\rightarrow$ efficacité du partitionnement spectral, même à faible N

Gliederung

Prédiction des crises d'épilepsie : données

Outils prêts pour des données du monde réel:

- Stylométrie : 6 indicateurs $\rightarrow$ roman original ou traduit ?
- Données physiologiques 4-variées $\rightarrow$ détection de la somnolence [Lucas et al., EMBC, 2022]
- Prédiction de crises d'épilepsie [Lucas et al., EUSIPCO, 2023]



19 paires d'électrodes

Base de données

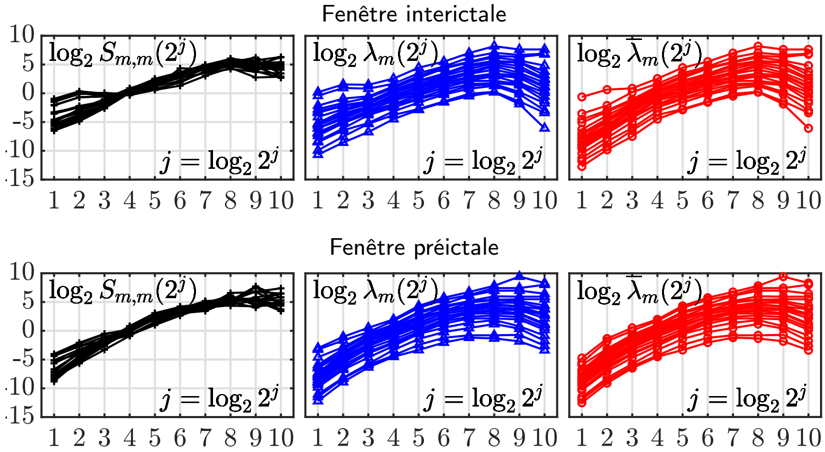
- CHB-MIT Scalp EEG
  physionet.org/content/chbmit/
- 8 sujets pédiatriques
- 19 canaux échantillonés à 256Hz

Objectif : classification binaire de fenêtres 19-variées de 2 minutes

- état préictal : période quelques minutes avant une crise d'épilepsie
- état interictal : période loin d'une crise d'épilepsie

Prédiction de crises d'épilepsie : invariance d'échelle

Analyse par patient

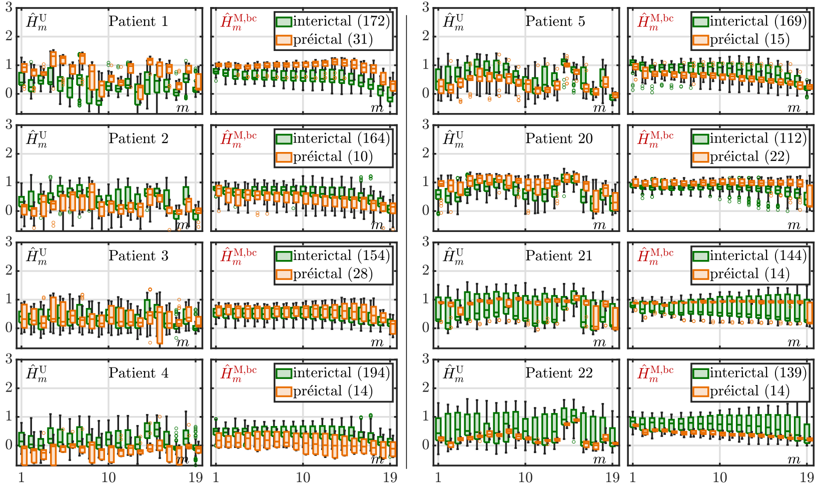


$\Rightarrow$ Comportement invariant d'échelle multivarié sur $2^1 - 2^8 = 0.5\text{Hz} - 85\text{Hz}$

Comparaison des distributions des paramètres d'autosimilarité

Fréquences d'analyse $10\text{Hz} - 85\text{Hz}$ [Gadhoumi et al., 2015]

$\Rightarrow$ Échelles d'analyse $2^{j_1} = 2^1$ – $2^{j_2} = 2^4$

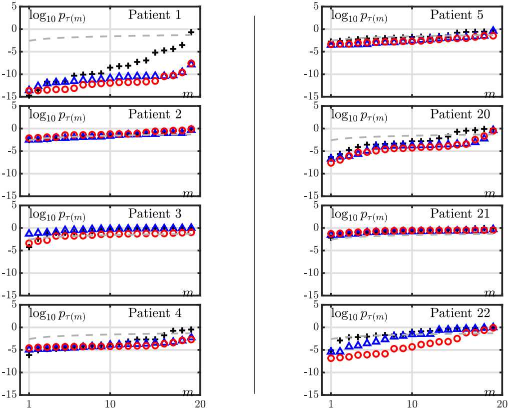


→ Différences entre distributions interictales et preictales par patient

→ Différences entre les différents patients

Prédiction des crises d'épilepsie : invariance d'échelle

Test de Wilcoxon :
p-valeurs (interictal contre preictal) pour $\underline{\hat{H}}^{\mathrm{U}}$, $\underline{\hat{H}}^{\mathrm{M}}$ et $\underline{\hat{H}}^{\mathrm{M,bc}}$
Seuils de Benjamini-Hochberg (comparaison multiple)



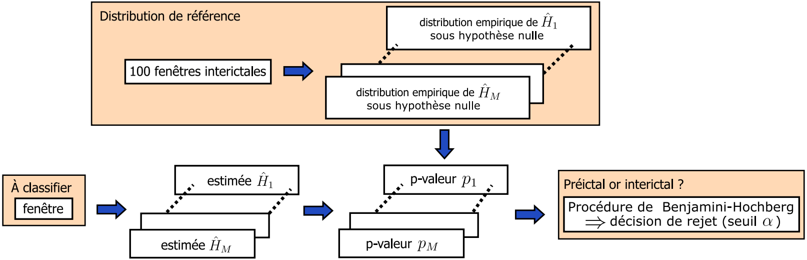
→ Différences entre états interictaux et preictaux par patient

- significatives pour tous les patients
- plus significatives avec $\underline{\hat{H}}^{\mathrm{M,bc}}$

Prédiction des crises d'épilepsie : classification

– Classification par patient [Lucas et al., EUSIPCO, 2023]

– Procédure pour chaque patient:



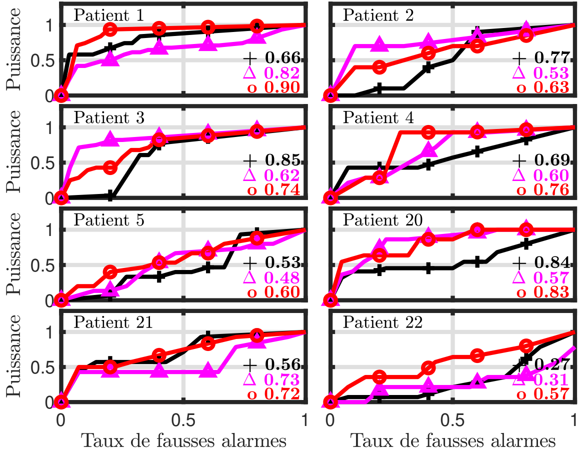
où les $\hat{H}_1, \ldots, \hat{H}_M$ sont

- soit les estimées univariées $\hat{H}_1^U, \ldots, \hat{H}_M^U$,
- soit les estimées multivariées corrigées $\hat{H}_1^{M,bc}, \ldots, \hat{H}_M^{M,bc}$.

– Même procédure avec $M(M+1)/2$ estimées multivariées classiques $\hat{H}_{m,m'}$

Prédiction des crises d'épilepsie : performances de classification

Courbes ROC avec AUC (aire sous la courbe)



+ M estimées univariées $\hat{H}_m^U$

△ $M(M+1)/2$ estimées multivariées classiques $\hat{H}_{m,m'}$

o M estimées multivariées corrigées $\hat{H}_m^{M,bc}$

$$\hat{H}_m^{M,bc} \sim \hat{H}_{m,m'} \geq \hat{H}_m^U$$

Résultats et perspectives

Résultats

- Estimation de $\underline{H}$ robuste au biais de répulsion
- Tests d'égalité : partitionnement de $\underline{H}$ en paramètres égaux
- Boîte à outils MatLab en ligne : github.com/charlesglucas/ofbm_tools
- Applications biomédicales : procédure de classification

Travaux en cours

- Grande dimension : M grand avec N limité
  - → Triple limite : $M \to +\infty$, $N \to +\infty$, $2^{j_1}, \dots, 2^{j_2} \to +\infty$
  - → Perte de la normalité multivariée de l'estimateur
  - → Nouvelle approche du partitionnement de $\underline{H}$ [Lucas et al., GRETSI, 2023]

Perspectives

- Prise en compte de l'irréversibilité en temps :

$$\{\mathcal{B}_{\underline{H},\Sigma,W}(t)\}_{t\in\mathbb{R}} \overset{fdd}{\neq} \{\mathcal{B}_{\underline{H},\Sigma,W}(-t)\}_{t\in\mathbb{R}}$$

- Extension à deux dimensions : texture d'images et anisotropie (e.g. imagerie hyperspectrale)

$$\{\mathcal{B}_{\underline{H},\Sigma,W}(a^{E}x)\}_{x\in\mathbb{R}^2} \overset{fdd}{=} \{a^{\underline{H}}\mathcal{B}_{\underline{H},\Sigma,W}(x)\}_{x\in\mathbb{R}^2}, \quad \forall a \in \mathbb{R}^2$$

E : matrice 2×2 caractérisant l'anisotropie

Communications

– **Estimation**

[Lucas et al., 2023a] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Multivariate self-similarity: Multiscale eigen structures for the estimation of Hurst exponents. Soumis.

– **Applications**

[Lucas et al., EMBC, 2022] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Drowsiness detection from polysomnographic data using multivariate selfsimilarity-based eigen-wavelet analyses, EMBC, Glasgow, Scotland, 2022.

[Lucas et al., EUSIPCO, 2023] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Epileptic seizure prediction from eigen-wavelet multivariate selfsimilarity analysis of multi-channel EEG signals, EUSIPCO, Helsinki, Finland, 2023.

– **Dénombrement et regroupement de paramètres égaux**

[Lucas et al., EUSIPCO, 2021] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Bootstrap for testing the equality of selfsimilarity exponents across multivariate time series, EUSIPCO, Dublin, Ireland, 2021.

[Lucas et al., ICASSP, 2022] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Counting the number of dif- ferent scaling exponents in multivariate scale-free dynamics: clustering by bootstrap in the wavelet domain, ICASSP, Singapore, 2022.

[Lucas et al., GRETSI, 2022] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Multivariate time-scale bootstrap for testing the equality of selfsimilarity parameters, GRETSI, Nancy, France, 2022.

[Lucas et al., 2023b] C-G. Lucas, P. Abry, H. Wendt and G. Didier, Multivariate self-similarity parameter counting: Spectral clustering using wavelet-domain bootstrap. En cours d'écriture.

– **Grande dimension**

[Lucas et al., GRETSI, 2023] C-G. Lucas, P. Abry, H. Wendt, G. Didier and O. Orejola, Bootstrap based test for the unimodality of estimated Hurst exponents. Performance assessment in a high-dimensional analysis setting, GRETSI, Grenoble, France, 2023.

MERCI

POUR VOTRE ATTENTION !

Lois de puissance des valeurs propres

Idée de la démonstration [Abry et al., 2018a; 2018b]

$\mathcal{B}_{\underline{\underline{H}},\Sigma,W}$ autosimalaire $\Rightarrow S(a(N)2^j) \overset{\mathbb{P}}{\sim} a(N)^{(\underline{\underline{H}}+\frac{1}{2}\mathbb{I})} S(2^j) a(N)^{(\underline{\underline{H}}^\top+\frac{1}{2}\mathbb{I})}$

$$\text{avec } a(N)^{\underline{\underline{H}}} = a(N)^{W\text{diag}(\underline{H})W^{-1}}$$

$$= W \begin{bmatrix} a(N)^{H_1} & 0 & \cdots & \cdots & 0 \\ 0 & a(N)^{H_2} & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & a(N)^{H_M} \end{bmatrix} W^{-1}$$

$\rightarrow$ Valeurs propres $\hat{\lambda}_m(2^j)$ de $S(2^j)$ asymptotiquement bornées :

$$\hat{\lambda}_m(2^j) \overset{\mathbb{P}}{\rightarrow} l_m \text{ avec } 0 < l_m \leq B_m$$

$\rightarrow$ Encadrement des valeurs propres $\hat{\lambda}_m(a(N)2^j)$ de $S(a(N)2^j)$ par les valeurs propres de $a(N)^{(\underline{\underline{H}}+\frac{1}{2}\mathbb{I})} a(N)^{(\underline{\underline{H}}^\top+\frac{1}{2}\mathbb{I})}$, i.e. par les $a(N)^{2H_m+1}$

$\rightarrow \hat{\lambda}_m(a(N)2^j) \propto a(N)^{2H_m+1}$

Étude des performances d'estimation empiriques

Simulations de Monte Carlo :

- 1000 réalisations
- $M = 6$ composantes
- Tailles d'échantillon $N \in \{2^{13}, \ldots, 2^{18}\}$
- 2 différent vecteurs $\underline{H}$:
  - $H_1 = \ldots = H_M = 0.7$
  - $H_1 \neq \ldots \neq H_M$: $\underline{H} = (0.4, 0.5, 0.6, 0.65, 0.7, 0.8)$
- Matrice de mélange aléatoire non orthonormale $W \neq \mathbb{I}$ ou $W = \mathbb{I}$
- Corrélation $\Sigma \neq \mathbb{I}$ avec $\Sigma_{m,m'} = 0.7^{|m'-m|}$, $1 \leq m \leq m' \leq M$

Analyse en ondelettes :

- Ondelette mère : Daubechies2
- Régressions linéaires de $2^{j_1} = a(N)2^6$ à $2^{j_2} = a(N)2^9$
  avec $\log_2 a(N) = \lfloor \beta \log_2 N/N_0 \rfloor$, $\beta = 0.9$, $N_0 = 2^{13}$

Performances d'estimation

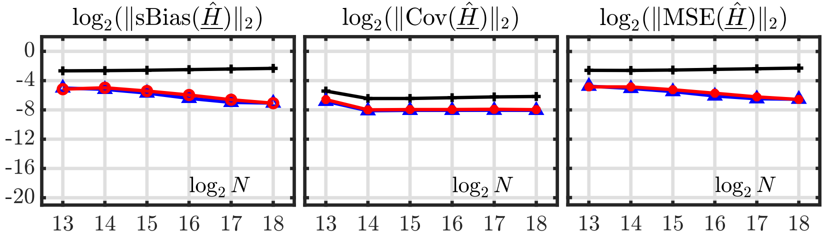
$$W \neq \mathbb{I}$$
$$H_1 \neq \ldots \neq H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

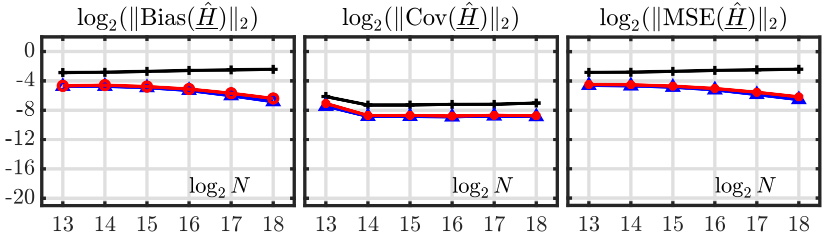
Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Ondelette mère : Daubechies2



Ondelette mère : Daubechies10



Performances d'estimation

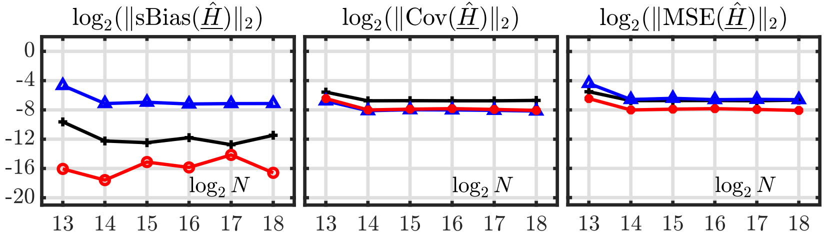
$$W \neq \mathbb{I}$$
$$H_1 = \ldots = H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

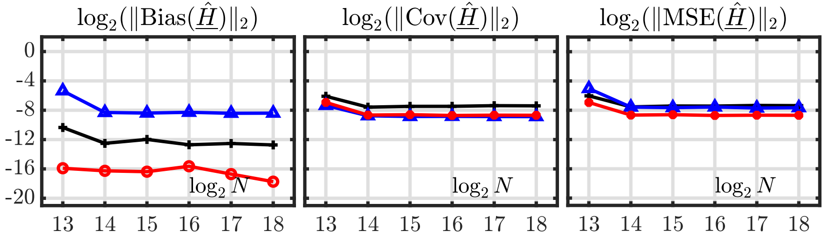
△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Ondelette mère : Daubechies2



Ondelette mère : Daubechies10



Performances d'estimation

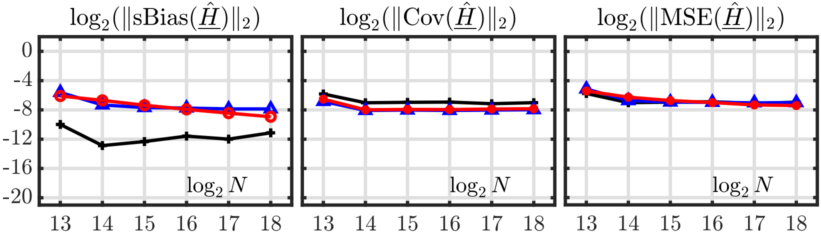
$$W = \mathbb{I}$$
$$H_1 \neq \ldots \neq H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

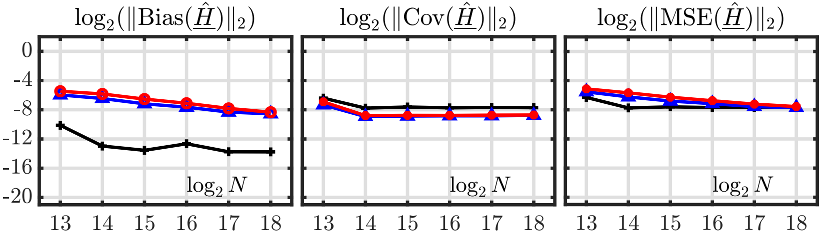
△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Ondelette mère : Daubechies2



Ondelette mère : Daubechies10



Performances d'estimation

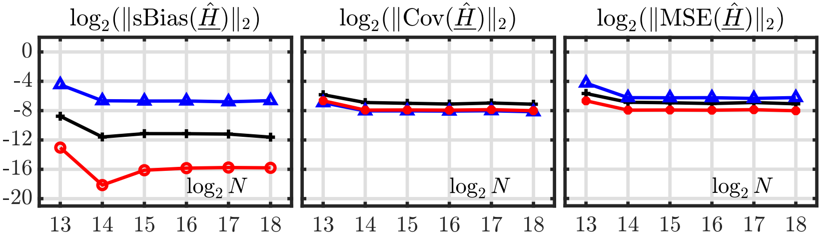
$$W = \mathbb{I}$$
$$H_1 = \ldots = H_M$$

$+$ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

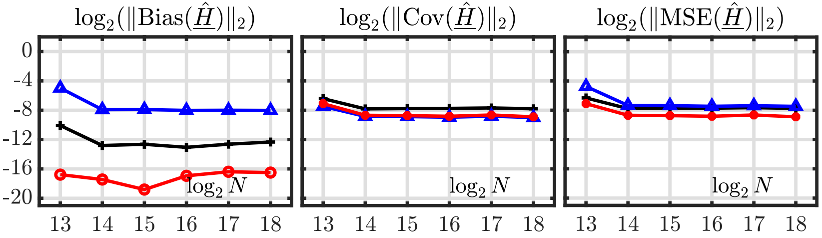
Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Ondelette mère : Daubechies2



Ondelette mère : Daubechies10



Performances d'estimation

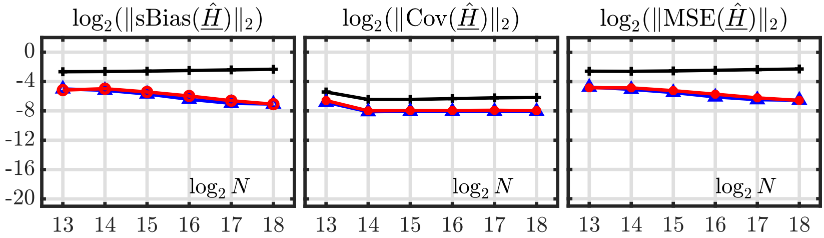
$$W \neq \mathbb{I}$$
$$H_1 \neq \ldots \neq H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

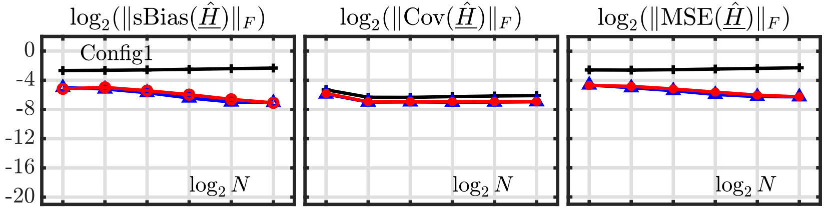
Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Norme spectrale :



Norme de Frobenius :



Performances d'estimation

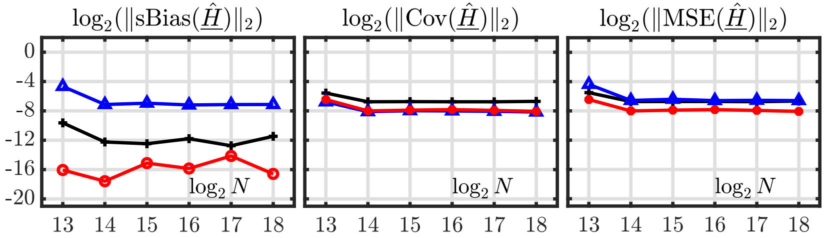
$$\boxed{\begin{array}{c} W \neq \mathbb{I} \\ H_1 = \ldots = H_M \end{array}}$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

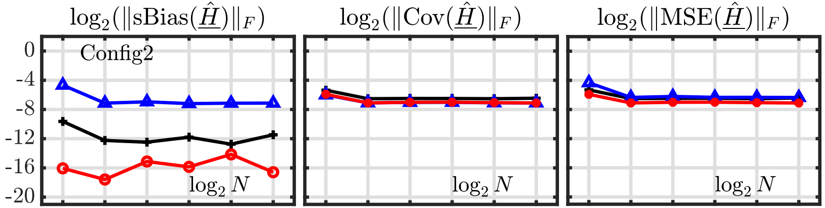
Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Norme spectrale :



Norme de Frobenius :



Performances d'estimation

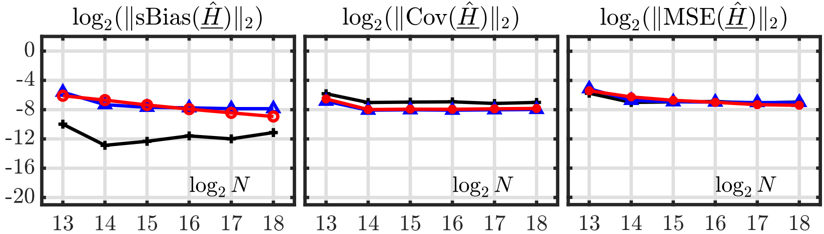
$$\boxed{\begin{array}{c} W = \mathbb{I} \\ H_1 \neq \ldots \neq H_M \end{array}}$$

$+$ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

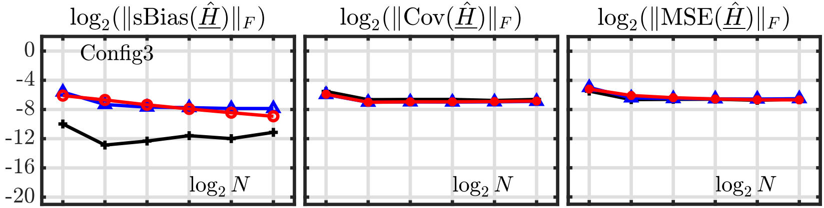
Δ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Norme spectrale :



Norme de Frobenius :



Performances d'estimation

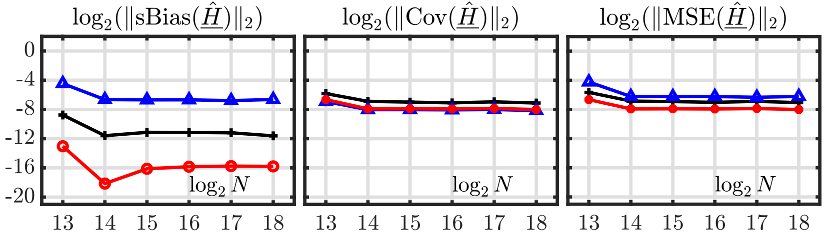
$$W = \mathbb{I}$$
$$H_1 = \ldots = H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

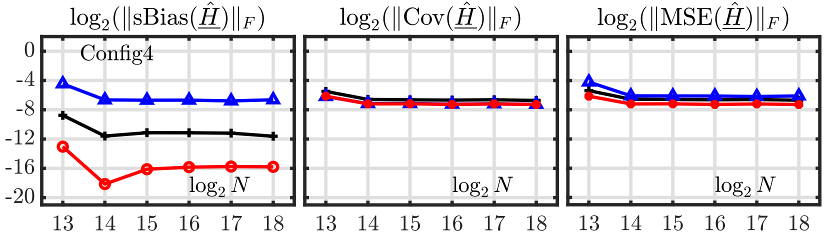
△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$

Norme spectrale :



Norme de Frobenius :



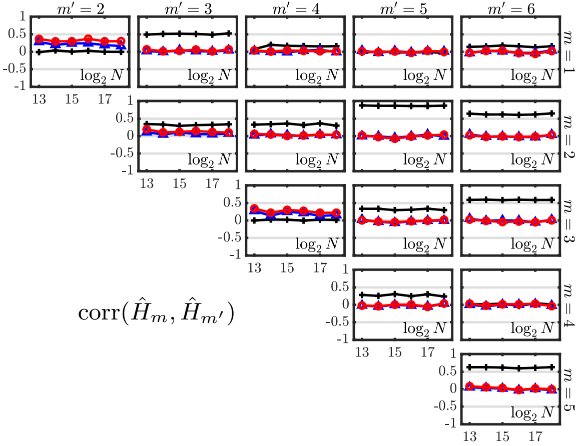
Corrélation des estimateurs

$$W \neq \mathbb{I} \quad H_1 = \ldots = H_M$$

+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



→ Quasi-décorrélation de $\underline{\hat{H}}^{\mathrm{M,bc}}$ et $\underline{\hat{H}}^{\mathrm{M}}$, même pour N faible

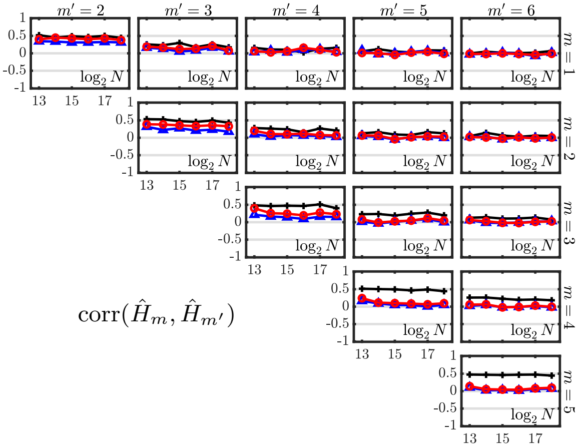
Corrélation des estimateurs

$$W = \mathbb{I}$$
$$H_1 = \ldots = H_M$$

\+ estimation univariée $\underline{\hat{H}}^{\mathrm{U}}$

△ estimation multivariée $\underline{\hat{H}}^{\mathrm{M}}$

o estimation multivariée corrigée $\underline{\hat{H}}^{\mathrm{M,bc}}$



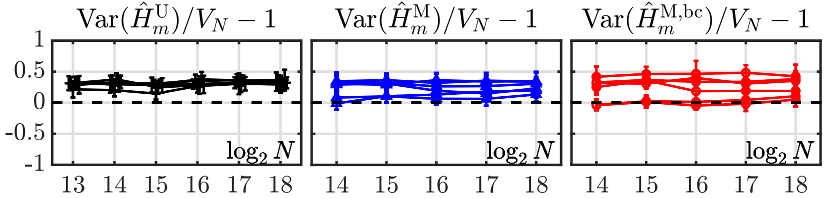
$\mathrm{corr}(\hat{H}_m, \hat{H}_{m'})$

→ Quasi-décorrélation de $\underline{\hat{H}}^{\mathrm{M,bc}}$ et $\underline{\hat{H}}^{\mathrm{M}}$, même pour N faible

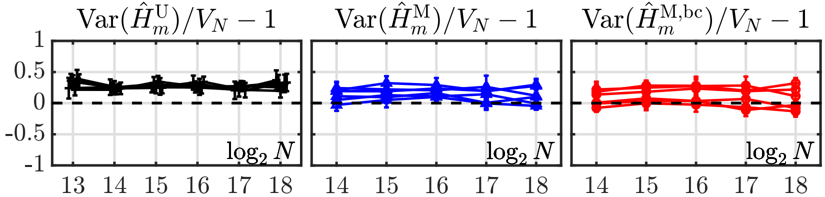
Variance des estimateurs

$$V_N = 2(\log_2 e)^2 \sum_{j=j_1}^{j_2} \frac{w_j^2}{n_{a,j}}$$

$W \neq \mathbb{I}$
$H_1 \neq \ldots \neq H_M$



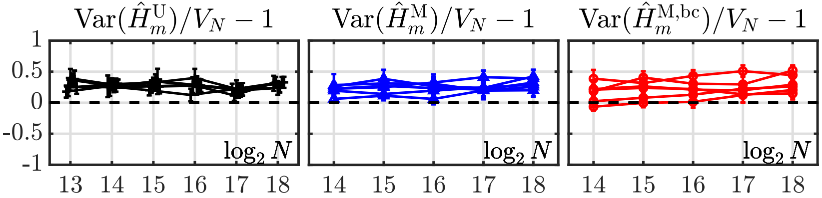
$W \neq \mathbb{I}$
$H_1 = \ldots = H_M$



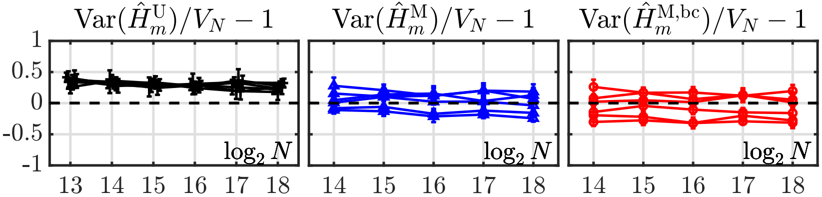
Variance des estimateurs

$$V_N = 2(\log_2 e)^2 \sum_{j=j_1}^{j_2} \frac{w_j^2}{n_{a,j}}$$

$W = \mathbb{I}$
$H_1 \neq \ldots \neq H_M$



$W = \mathbb{I}$
$H_1 = \ldots = H_M$



Étude des performances de partitionnement empiriques

Analyse en ondelettes :

- Ondelette mère ψ_0 : Daubechies3
- Régression linéaire de :
  - 2^8 à 2^{11} pour $M = 6$
  - 2^7 à 2^{10} pour $M = 20$
- Mélange W inversible et non orthonormal aléatoire
- $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_M)\rho\mathrm{diag}(\sigma_1, \ldots, \sigma_M)^T$
  - $\rho = 0.5$ et $\sigma_m = 1$ pour $M = 6$
  - $\rho = 0.5$ et $\sigma_m = 2^m$ pour $M = 20$

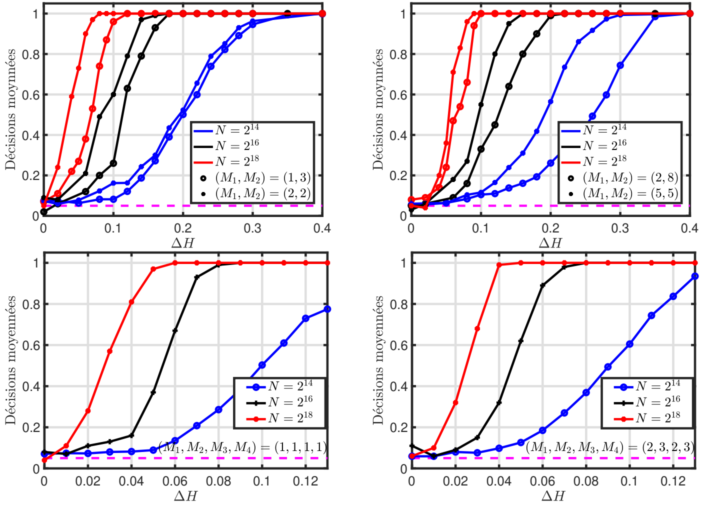
Bootstrap :

- $R = 500$ ré-échantillons
- Taille des blocs $N_B = \text{taille du support de } \psi_0 = 6$

Test pour $H_1 = \ldots = H_M$: statistique et puissance

Statistique : $T^* = (\underline{\hat{H}}^{\mathrm{M,bc}} - \langle \hat{H}_m^{\mathrm{M,bc}} \rangle_m)^\top \left(\hat{\Sigma}^*_{\underline{\hat{H}}} \right)^{-1} (\underline{\hat{H}}^{\mathrm{M,bc}} - \langle \hat{H}_m^{\mathrm{M,bc}} \rangle_m)$

$\rightarrow$ sous $H_1 = \ldots = H_M$, χ^2 à $M-1$ degrés de liberté pour N grand



$$\underline{H} = 0.5 + (0, \ldots, 0, \Delta H, \ldots, \Delta H, 2\Delta H, \ldots, 2\Delta H, 3\Delta H, \ldots, 3\Delta H)$$
$$M = M_1 + M_2 + M_3 + M_4$$

Test pour $H_m = H_{m+1}$: procédure

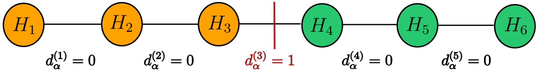
- Formulation :
  - $M-1$ hypothèse : $H_m = H_{m+1}$, $1 \leq m \leq M$
  - Tri des estimées corrigées : $\hat{H}_{\tau(1)} \leq \ldots \leq \hat{H}_{\tau(M)}$
  - Statistiques : $\tilde{\delta}_m = \hat{H}_{\tau(m+1)} - \hat{H}_\tau(m)$

- Distribution approximative :
  - $\tilde{\delta}_m$ normale repliée pour N grand et M petit
  - $\tilde{\delta}_m$ demi-normale $\mathcal{FN}(0, \tilde{\sigma}_m)$ pour N grand et M petit sous $H_m = H_{m+1}$

- p-valeurs : $\tilde{p}_m = 1 - F_{\mathcal{FN}(0,\tilde{\sigma}_m)}(\tilde{\delta}_m)$

- Décisions $d_\alpha^{(m)}$: procédure de Benjamini-Hochberg sur les p-valeurs $\tilde{p}_m$

- Partitionnement :



$\rightarrow$ Estimation de $\tilde{\sigma}_m$ par bootstrap

Test pour $H_m = H_{m+1}$: estimation des paramètres

Estimation du paramètre $\tilde{\sigma}_m$ de la loi demi-normale par bootstrap :

- Reproduction de $H_1 = \ldots = H_M$ [Lucas et al., ICASSP, 2022] :

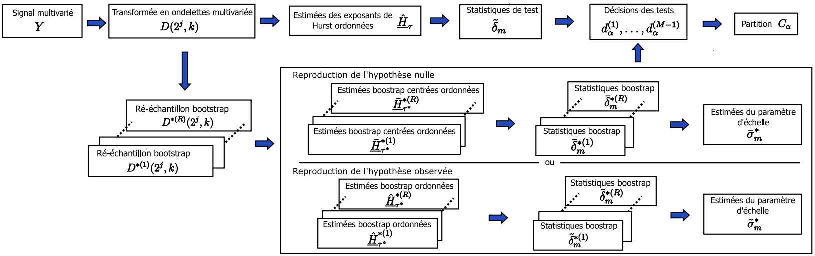
$$\bar{\delta}_m^{*(r)} = \bar{H}^{*(r)}_{\tau^{*(r)}(m+1)} - \bar{H}^{*(r)}_{\tau^{*(r)}(m)}, \quad \underline{\bar{H}}^{*(r)} = \underline{\hat{H}}^{*(r)} - \langle \underline{\hat{H}}^{*(r)} \rangle_r$$

$\Rightarrow$ estimer le paramètre $\bar{\sigma}_m^*$ de loi demi-normale

- Reproduction de l'hypothèse observée [Lucas et al., GRETSI, 2022] :

$$\tilde{\delta}_m^{*(r)} = \hat{H}^{*(r)}_{\tau^{*(r)}(m+1)} - \hat{H}^{*(r)}_{\tau^{*(r)}(m)}$$

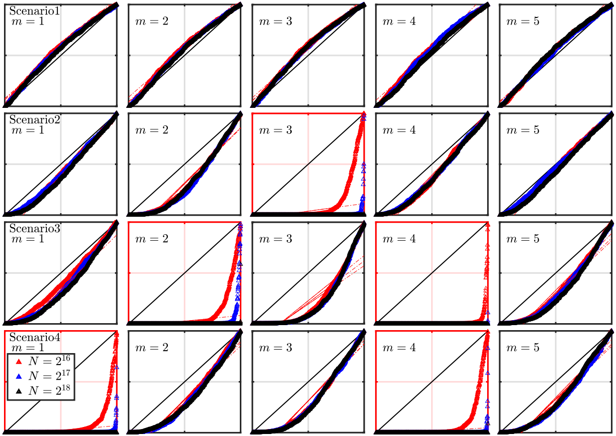
$\Rightarrow$ estimer le paramètre $\tilde{\sigma}_m^*$ de loi normale repliée



Test pour $H_m = H_{m+1}$: performances

Reproduction de $H_1 = \ldots = H_M$

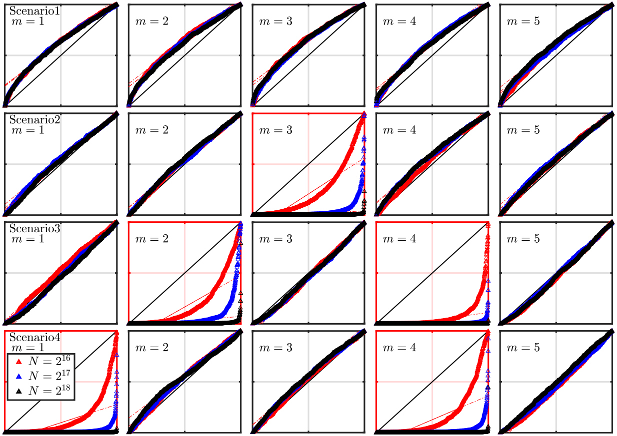
Diagrammes quantile-quantile
$\bar{p}_m^*$ contre distribution uniforme



Test pour $H_m = H_{m+1}$: performances

Reproduction de l'hypothèse observée

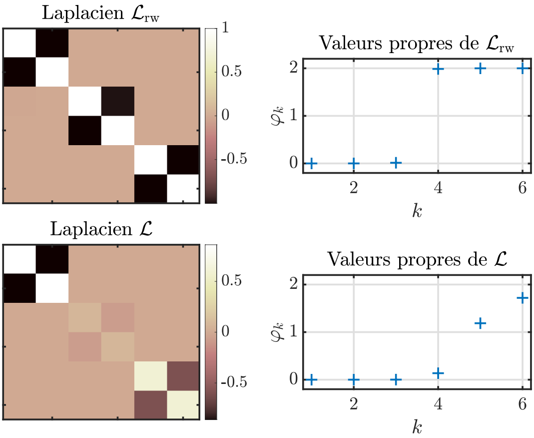
Diagrammes quantile-quantile
$\tilde{p}_m^*$ contre distribution uniforme



Partitionnement : comparaison de laplaciens

Partitionnement spectral sur

- Laplacien normalisé $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$
  Désavantage : ne permet pas d'identifier des nœuds isolés
- Laplacien combinatoire $\mathcal{L} = D - W$
  Désavantage : moins adapté que $\mathcal{L}_{rw}$ pour la convergence
  [Von Luxburg, 2008; Sarkar, 2015]

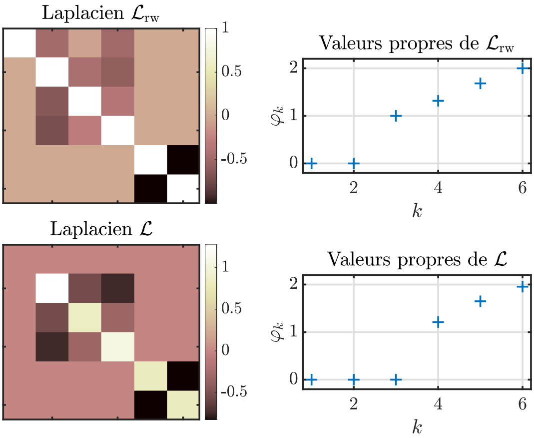


$$\underline{H} = (0.4, 0.4, 0.6, 0.6, 0.8, 0.8)$$

Partitionnement : comparaison de laplaciens

Partitionnement spectral sur

- Laplacien normalisé $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$
  Désavantage : ne permet pas d'identifier des nœuds isolés
- Laplacien combinatoire $\mathcal{L} = D - W$
  Désavantage : moins adapté que $\mathcal{L}_{rw}$ pour la convergence
  [Von Luxburg, 2008; Sarkar, 2015]

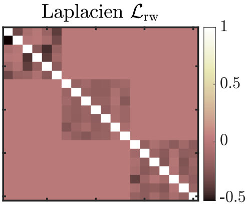


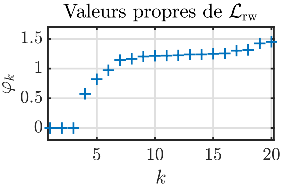
$$\underline{H} = (0.4, 0.6, 0.6, 0.6, 0.8, 0.8)$$

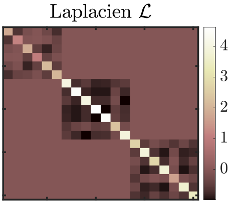
Partitionnement : comparaison de laplaciens

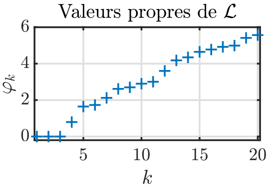
Partitionnement spectral sur

- Laplacien normalisé $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$
  Désavantage : ne permet pas d'identifier des nœuds isolés
- Laplacien combinatoire $\mathcal{L} = D - W$
  Désavantage : moins adapté que $\mathcal{L}_{rw}$ pour la convergence
  [Von Luxburg, 2008; Sarkar, 2015]







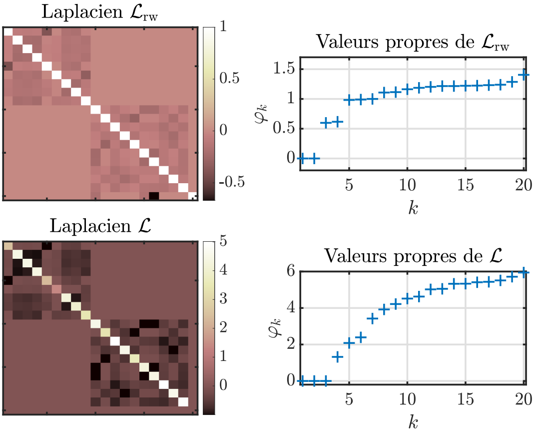


$$\underline{H} = (0.4, \ldots, 0.4, 0.6, \ldots, 0.6, 0.8, \ldots, 0.8), \quad M = 6 + 7 + 7$$

Partitionnement : comparaison de laplaciens

Partitionnement spectral sur

- Laplacien normalisé $\mathcal{L}_{rw} = D^{-1}(D - \mathcal{S})$
  Désavantage : ne permet pas d'identifier des nœuds isolés
- Laplacien combinatoire $\mathcal{L} = D - W$
  Désavantage : moins adapté que $\mathcal{L}_{rw}$ pour la convergence
  [Von Luxburg, 2008; Sarkar, 2015]



$\underline{H} = (0.4, \ldots, 0.4, 0.6, \ldots, 0.6, 0.8), \quad M = 9 + 10 + 1$

Partitionnement : Indice de Rand ajusté

ARI, Adjusted Rand Index

2 partitions de $\mathcal{V} = \{1, \ldots, M\}$: $U = \{U_1, U_2, \ldots, U_R\}$, $V = \{V_1, V_2, \ldots, V_C\}$.

$$\mathrm{RI} = (a+b)/\binom{M}{2}$$

- a : nombre de paires d'éléments de $\mathcal{V}$ dans le même sous-sensemble à la fois dans V and dans U
- b : nombre de paires d'éléments de $\mathcal{V}$ dans différents sous-sensembles à la fois dans V and dans U

$$\mathrm{ARI} = \frac{\mathrm{RI} - \mathbb{E}\mathrm{RI}}{\max(\mathrm{RI}) - \mathbb{E}\mathrm{RI}}$$

Partitionnement : Information mutuelle normalisée

NMI, Normalized Mutual Information

2 partitions de $\mathcal{V} = \{1, \ldots, M\}$: $U = \{U_1, U_2, \ldots, U_R\}$, $V = \{V_1, V_2, \ldots, V_C\}$.

$$NMI = \frac{H(U) + H(V) - H(U,V)}{\sqrt{H(U)H(V)}}$$

avec

$$H(U) = -\sum_i q_{i,.} \log_2(q_{i,.})$$

$$H(V) = -\sum_j q_{.,j} \log_2(q_{.,j})$$

$$H(U,V) = -\sum_{i,j} q_{i,j} \log_2(q_{i,j})$$

with:

- $q_{i,j} = P(U_i \cap V_j)$: proportion d'éléments à la fois dans U_i et V_j
- $q_{i,.} = P(U_i)$: proportion d'éléments dans U_i
- $q_{.,j} = P(V_j)$: proportion d'éléments dans V_j

Prédiction de crises d'épilepsie : pré-traitement et analyse des données

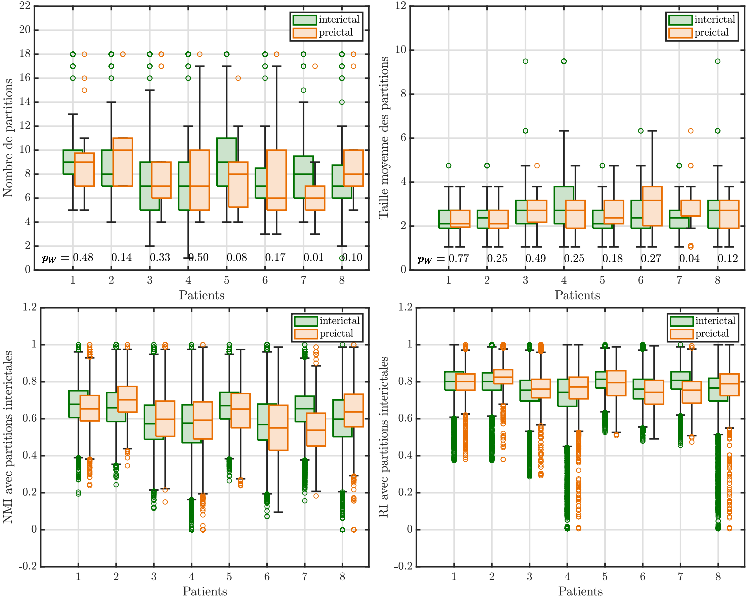
- Univarié : $N/2^j = n_j \gg 8$
- Multivarié : $N/2^j = n_j \gg 4M$ ($\Rightarrow S(2^j)$ de rang plein)

$\rightarrow M = 19 \Rightarrow \simeq 2^6 \Rightarrow N \gg 2^{j+6}$
$\rightarrow$ Lois de puissance de 2^1 à $2^8 \Rightarrow N > 2^{8+6} = 2^{14}$
$\rightarrow$ Fréquence d'échantillonnage $f_0 = 256\mathrm{Hz} \Rightarrow$ fenêtres de $2^{15}/256 \simeq 2\mathrm{mn}$

Ondelette mère : Daubechies2

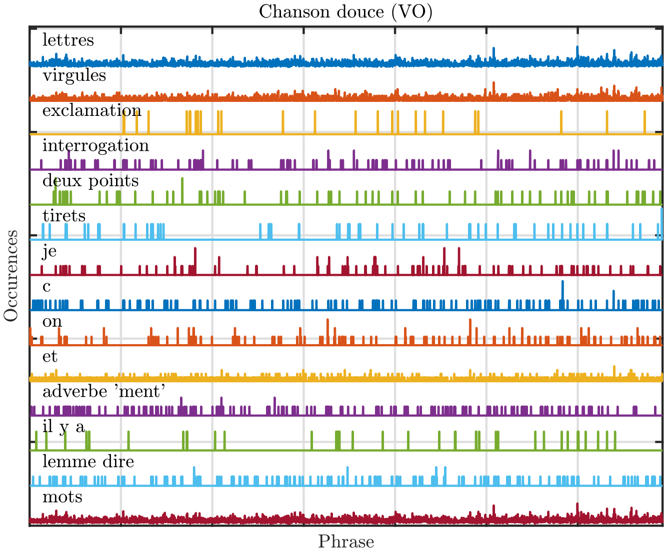
Bootstrap : $R = 500$ ré-échantillons, taille des blocs $N_B = 4$

Prédiction de crises d'épilepsie : partitionnement

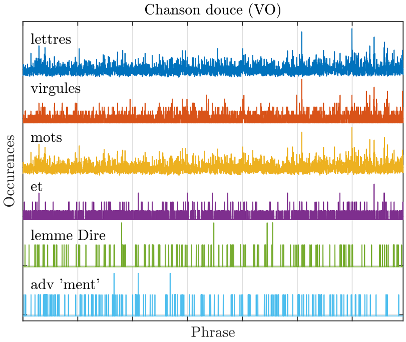


Stylométrie : romans VO/VT

Romans français : 20 en version originale (VO), 25 en version traduite (VT)
Taille des romans (nombre de phrases) : de 971 à $14\,587$, en moyenne $6\,157$



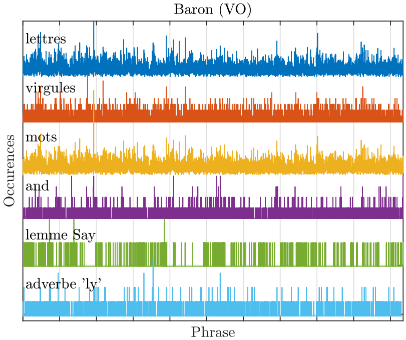
Stylométrie : analyse 6-variée



Corrélations	lettres	virgules	mots	et	lemme Dire	adv "ment"
lettres	1	0.67	0.98	0.45	0.11	0.14
virgules	0.67	1	0.66	0.20	0.11	0.05
mots	0.98	0.66	1	0.46	0.13	0.10
et	0.45	0.20	0.46	1	0.02	0.07
lemme Dire	0.11	0.11	0.13	0.02	1	0.04
adv 'ment'	0.14	0.05	0.10	0.07	0.04	1

Objectif : comparaison entre VO et VT par autosimilarité multivariée

Stylométrie : analyse 6-variée



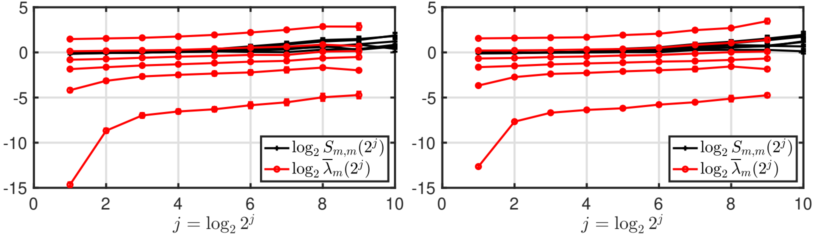
Corrélations	lettres	virgules	mots	et	lemme Say	adv 'ly'
lettres	1	0.51	0.97	0.39	-0.01	0.21
virgules	0.51	1	0.49	0.14	0.23	0.16
mots	0.97	0.49	1	0.39	0.00	0.17
and	0.39	0.14	0.39	1	-0.06	0.04
lemme Say	-0.01	0.23	0.00	-0.06	1	0.02
adv 'ly'	0.21	0.16	0.17	0.04	0.02	1

Objectif : comparaison entre VO et VT par autosimilarité multivariée

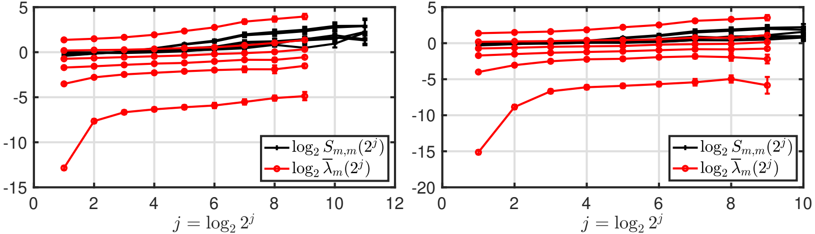
Stylométrie : analyse 6-variée

Analyse en ondelettes sur les échelles 2^3 à $2^{\log_2(N)-4}$

Romans francophones : lettres, virgules, mots, ‘et’, lemme Dire, adverbe ‘ment’



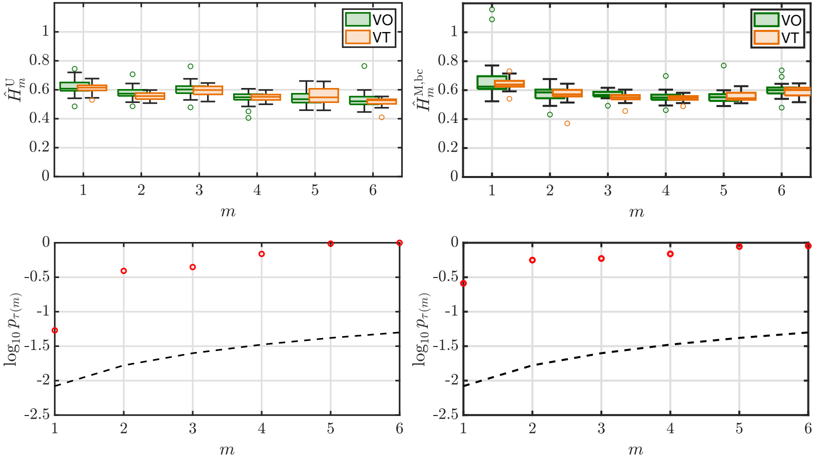
Romans anglophones : lettres, virgules, mots, ‘and’, lemme Say, adverbe ‘ly’



Stylométrie : analyse 6-variée

Analyse en ondelettes sur les échelles 2^3 à $2^{\log_2(N)-4}$

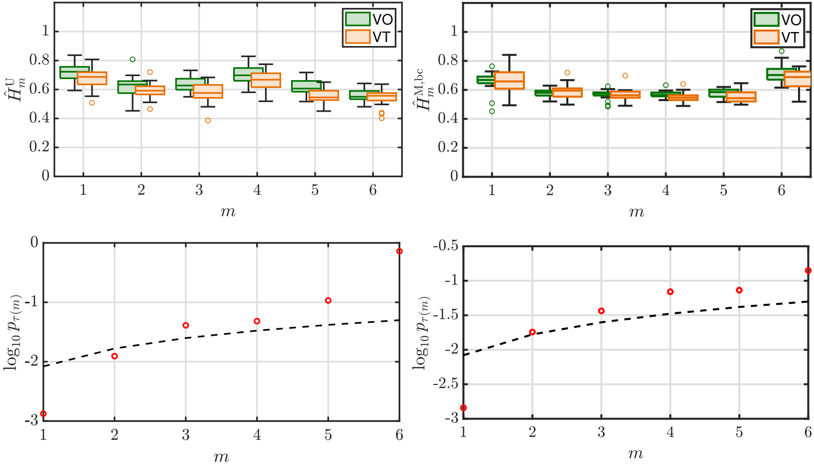
Romans francophones : lettres, virgules, mots, 'et', lemme Dire, adverbe 'ment'



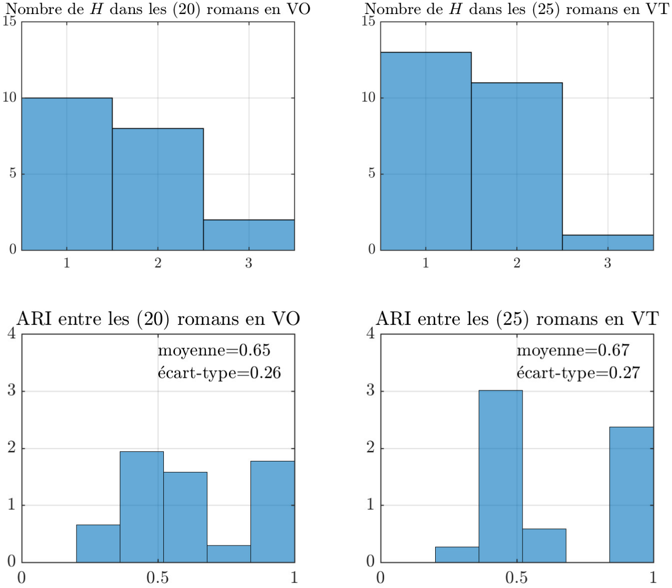
Stylométrie : analyse 6-variée

Analyse en ondelettes sur les échelles 2^3 à $2^{\log_2(N)-4}$

Romans anglophones : lettres, virgules, mots, 'and', lemme Say, adverbe 'ly'



Partitionnement par graphe des romans



Régime asymptotique de grande dimension

Dans certaines applications : M grand avec N limité

- Magnetoencephalographie (MEG): $M \sim 100$
- Imagerie par résonance magnétique fonctionnelle (IRMs): $M \sim 10000$

Cadre asymptotique

Triple limite :

- $N \to +\infty$
- $2^{j_1}, \ldots, 2^{j_2} \to +\infty$
- $M \to +\infty$

  vitesse contrôlée par

$$\frac{M}{N/2^{j_1}} \to c_1 \in [0,1]$$

$$\vdots$$

$$\frac{M}{N/2^{j_2}} \to c_J \in [0,1]$$

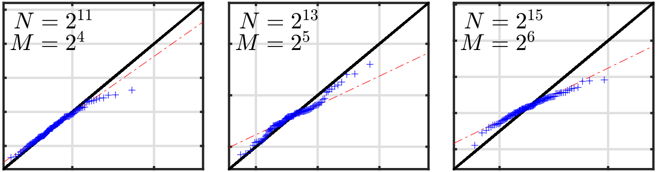
Étude numérique : 100 réalisations, $\Sigma = \mathbb{I}$, W orthonormale aléatoire

Enjeux de la grande dimension

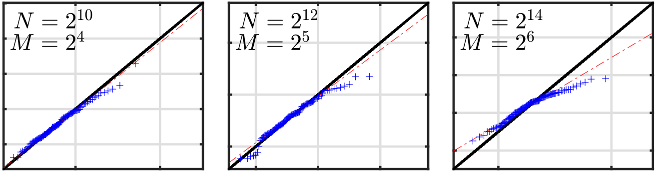
$$\underline{H} = (0.6, \ldots, 0.6, 0.8, \ldots, 0.8), \qquad M = M/2 + M/2$$

Diagramme quantile-quantile de $\underline{\hat{H}}$ contre χ^2 à M degrés de liberté

$$\frac{M}{N/2^{j_2}} = \frac{1}{8}$$



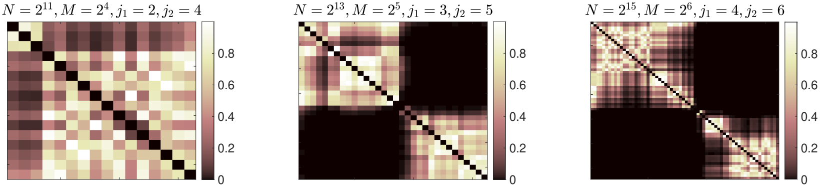
$$-\frac{M}{N/2^{j_2}} = \frac{1}{4}$$



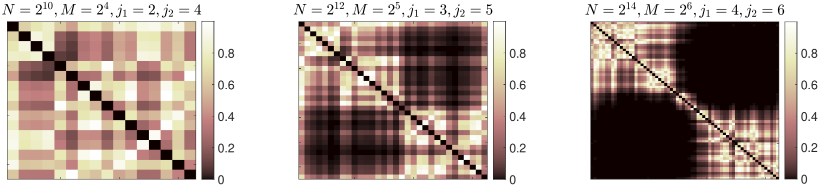
Enjeux de la grande dimension

$$\underline{H} = ({\color{red}0.6, \dots, 0.6}, {\color{blue}0.8, \dots, 0.8}), \qquad M = {\color{red}M/2} + {\color{blue}M/2}$$

Matrices de similarité $\mathcal{S}$ pour $\frac{M}{N/2^{j_2}} = \frac{1}{8}$



Matrices de similarité $\mathcal{S}$ pour $\frac{M}{N/2^{j_2}} = \frac{1}{4}$

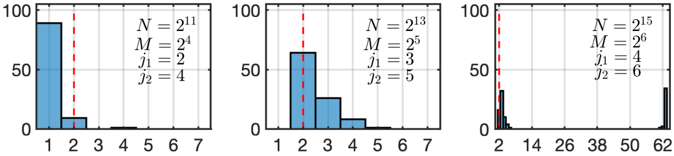


Enjeux de la grande dimension

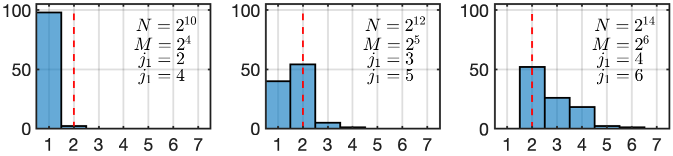
$$\underline{H} = ({\color{red}0.6, \ldots, 0.6}, {\color{blue}0.8, \ldots, 0.8}), \qquad M = {\color{red}M/2} + {\color{blue}M/2}$$

Histogramme des nombres de partitions estimés par partitionnement spectral

$$\frac{M}{N/2^{j_2}} = \frac{1}{8}$$



$$\frac{M}{N/2^{j_2}} = \frac{1}{4}$$

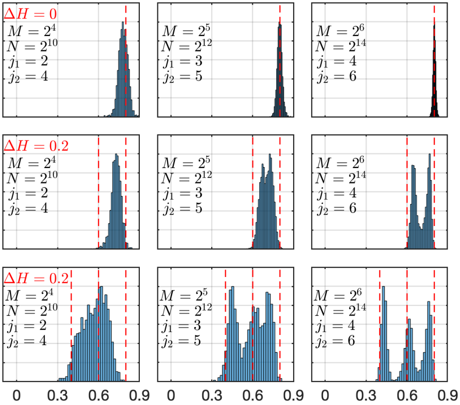


$\Rightarrow$ Surestimation

Comportement de l'estimateur en grande dimension

$$\underline{H} = (\underbrace{H_1, \ldots, H_1}_{p_1}, \underbrace{H_2, \ldots, H_2}_{p_2}, \ldots\ldots, \underbrace{H_{N_C}, \ldots, H_{N_C}}_{p_{N_C}})$$

Histogrammes des $\hat{H}_m$
au travers de $m = 1.\ldots.M$ et 100 réalisations



Proportions p_i de valeurs H_i dans $\underline{H}$ égales
avec $\Delta H = H_{i+1} - H_i$

Conjecture
Nombre N_C de valeurs
différentes H_i dans $\underline{H}$
$\Leftrightarrow$
Nombre de modes dans $\underline{\hat{H}}$
[Orejola et al., 2022]

$\downarrow$

Tests de multimodalité sur $\underline{\hat{H}}$

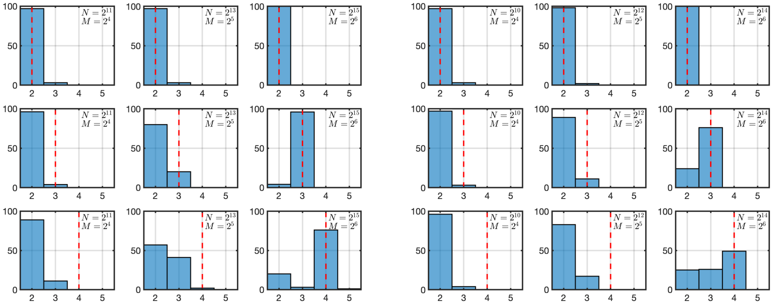
Dénombrement en grande dimension

→ $H_1 = \ldots = H_M$? ⇒ Test d'unimodalité sur $\underline{\hat{H}}$ [Lucas et al., GRETSI, 2023]

→ Estimation du nombre N_{C} de paramètres différents dans $\underline{H}$:

- Tester $\mathcal{H}_0^{(k)}$: "au plus $k \geq 2$ modes dans $\underline{\hat{H}}$" [Silverman, 1981]
- Estimée de $\hat{N}_{\mathrm{C}}$: plus petit k telle $\mathcal{H}_0^{(k)}$ non rejetée
- Histogrammes de $\hat{N}_{\mathrm{C}}$ sur 100 réalisations :

$\frac{M}{N/2^{j_2}} = \frac{1}{8}$ $\qquad\qquad$ $\frac{M}{N/2^{j_2}} = \frac{1}{4}$



N_{C} valeurs en mêmes proportions dans $\underline{H}$