



HAL
open science

Statistique des réseaux : inférence, comparaison et classification. Application à la connectivité cérébrale

Sophie Achard

► **To cite this version:**

Sophie Achard. Statistique des réseaux : inférence, comparaison et classification. Application à la connectivité cérébrale. Méthodologie [stat.ME]. Univ. Grenoble Alpes, 2016. tel-02064855

HAL Id: tel-02064855

<https://theses.hal.science/tel-02064855>

Submitted on 12 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Mémoire présenté pour l'obtention du diplôme
d'**HABILITATION À**
DIRIGER DES RECHERCHES

École Doctorale Electronique, Electrotechnique, Automa-
tique, Traitement du Signal (EEATS)
Spécialité : **Signal, Image, Parole, Télécoms (SIPT)**

Présentée par
Sophie ACHARD

Statistique des réseaux :
inférence, comparaison et classification
Application à la connectivité cérébrale

Diplôme soutenu publiquement le **17 mai 2016**,
devant le jury composé de :

Anne GUÉRIN-DUGUÉ

Professeur, Université Grenoble Alpes, Présidente du jury

Maureen CLERC

DR, INRIA, Rapporteur

Jérôme IDIER

DR, CNRS, Université de Nantes, Rapporteur

Dimitri VAN de VILLE

Professeur, École polytechnique fédérale de Lausanne, Université de
Genève, Rapporteur

Emmanuel BARBIER

DR, INSERM, Université Grenoble Alpes, Examineur

Florence FORBES

DR, INRIA - Université Grenoble Alpes, Examinatrice

Stéphane KREMER

PU-PH, Université de Strasbourg, Examineur

Question everything. Learn something. Answer nothing.
Euripides.

A ma grand-mère, à mes parents,

Résumé :

Dans ce mémoire, je m'intéresse à l'analyse statistique de données multivariées qui sont structurées en réseaux (ou graphes). Chaque nœud du réseau correspond à une série temporelle et les connexions ou arêtes relient deux nœuds. Mes travaux de recherche sur l'analyse statistique des réseaux se déclinent en trois parties : l'inférence, la comparaison et la classification. L'inférence des réseaux consiste à estimer les connexions à partir des observations multivariées. Les séries temporelles observées présentent souvent de la longue mémoire et dans le cas de séries temporelles multivariées, l'estimation des corrélations par ondelettes est biaisée, nous avons donc introduit la notion de connectivité fractale qui permet de prendre en compte la longue mémoire dans les estimateurs. Le réseau ou graphe est obtenu après un seuillage des connexions pour extraire seulement celles qui sont significatives, soit par exploitation de la notion de tests multiples soit par optimisation d'un critère de vraisemblance. Enfin, dans le cadre de mon travail pour la compréhension de la connectivité cérébrale, l'objectif est de quantifier des différences entre populations de sujets sains ou malades. L'imagerie par résonance magnétique fonctionnelle (IRMf) permet d'enregistrer le fonctionnement du cerveau de manière non invasive et avec une résolution spatiale de l'ordre du millimètre. Les données recueillies sont donc des séries temporelles multivariées correspondant à différentes régions spatiales du cerveau. Je me suis attachée à développer des méthodes de comparaison ou de classification des réseaux basées soit sur l'extraction de métriques topologiques comme l'efficacité globale ou le clustering, soit sur la définition d'un score dérivé d'un critère de vraisemblance. Mes recherches sur le développement de nouvelles méthodes théoriques se sont toujours beaucoup enrichies du contexte interdisciplinaire à l'interface du traitement du signal et des images, des statistiques et des applications médicales. Les applications en neurosciences ont une portée clinique comme l'étude de patients en état de conscience altérée.

Summary :

This manuscript presents my recent work on the statistical analysis of multivariate data structured in a network (or graph). A node in the network corresponds to a time series, and the edges or connections are linking two nodes. My research results in the statistical analysis of networks are divided in three parts : the inference, the comparison, and the classification of networks. Network inference consists in estimating the connections from the multivariate observations of a given system. The observed time series are often long memory and the usage of wavelets in order to compute the correlations introduces a bias in the estimation. We define the fractal connectivity in order to design estimators of correlations that are not affected by the presence of long memory. The network is then obtained by thresholding the correlations computed for pairs of nodes. The significant connections are extracted by using multiple testing procedures or maximum likelihood criterion optimisations. Finally, I present applications of this research on the understanding of brain connectivity. My objective is to quantify differences between healthy volunteers and patients. We are using functional magnetic resonance imaging (fMRI) to acquire noninvasive data on the brain functional process with a spatial resolution around 1 millimeter. The acquired data are multivariate time series where each time series corresponds to a spatial region of the brain. Networks for each subject are inferred and then compared using topological metrics like global efficiency or clustering, or using a score derived on a likelihood criterion. My investigations on new theoretical methods have been fertilised by their interdisciplinary context, at the intersection of signal and image processing, statistics and medical applications with genuine clinical outreach.

Table des matières

1	Introduction	10
1.1	Modèles à longue mémoire	10
1.2	Représentations graphiques	11
1.3	Connectivité cérébrale	12
1.4	Plan du manuscrit	12
2	Longue dépendance, ondelettes et corrélation	13
2.1	Modèles à longue mémoire : de l'univarié au multivarié	14
2.1.1	La longue mémoire en statistiques : définitions et estimations	14
2.1.2	Le cas multivarié	16
2.2	Ondelettes et corrélation	17
2.2.1	Définition ondelettes et corrélation	17
2.2.2	Estimation et propriétés asymptotiques	20
3	Connectivité fractale	21
3.1	Estimation des paramètres du modèle, et le retard de phase	22
3.1.1	Approximation par Whittle et ondelettes	23
3.1.2	Résultats de convergence	24
3.2	Intérêt du modèle	25
3.2.1	Comparaison approche univariée et multivariée	25
3.2.2	Convergence asymptotique de la corrélation	25
3.3	Perspectives	28
4	Estimation de structure de graphes	28
4.1	Corrélation ou corrélation partielle?	29
4.2	La corrélation comme outil pour extraire les graphes	31
4.3	Méthode gloutonne avec la corrélation partielle	35
4.4	Graphes différentiels	39
4.5	Détection de rupture pour des signaux multivariés	40
4.6	Perspectives	40
5	Méthodes statistiques de comparaisons des graphes	41
5.1	Quels outils pour comparer des graphes?	42
5.2	Introduction d'un nouvel index	44
5.3	Validation du coefficient de perturbation des hubs	44
5.4	Comparaison par profil de probabilités	45
5.5	Perspectives	50
6	Logiciels	52
6.1	brainwaver	52
6.2	multiwave	53
6.3	modèles graphiques	53
7	Conclusion et projets futurs	53
7.1	Axe 1 : propriétés statistiques du signal BOLD	54
7.2	Axe 2 : Développement d'outils d'exploration des réseaux	54
7.3	Axe 3 : Développement en lien avec les données réelles	55
	Références	56

A	Annexe : texte complet des publications mentionnées dans le manuscrit	59
A.1	(ACHARD et al. 2006)	59
A.2	(ACHARD et al. 2008)	60
A.3	(ACHARD et al. 2012)	60
A.4	(RICHIARDI et al. 2013)	61
A.5	(FALLANI et al. 2014)	61
A.6	(COSTARD et al. 2014)	61
A.7	(HARLÉ et al. 2014), (HARLÉ et al. 2015)	62
A.8	(ACHARD et GANNAZ 2016)	62

1 Introduction

Mes travaux de recherche se situent à l'interface du traitement du signal et de l'image, des statistiques et de plusieurs applications, notamment en sciences de la vie : analyse de systèmes cognitifs.

Depuis la fin de ma thèse, mes recherches se sont toujours beaucoup enrichies du contexte interdisciplinaire et de la motivation des applications tout en conservant la rigueur nécessaire pour développer des méthodes statistiques efficaces, ce qui s'avère souvent contraignant et nécessite généralement de nouveaux développements. En effet, dans le cadre d'applications avec des données réelles, il est à la fois important de pouvoir vérifier les hypothèses exigées par la théorie, mais aussi savoir les adapter, les modifier et les exploiter afin de mettre en place des méthodes adéquates aux différentes applications.

Actuellement, les données réelles d'observations de systèmes cognitifs par exemple se composent d'observations acquises simultanément sur des échelles de temps de plus en plus grandes. Les observations d'un système sont dites multivariées si elles s'écrivent sous la forme d'une ou plusieurs matrices.

Notons X_1 la variable aléatoire réelle correspondante à l'individu ou l'objet 1 et $X_1(1), \dots, X_1(N)$ le vecteur d'observations. Cette variable peut représenter par exemple la taille ou le poids d'un individu, la longueur ou la densité d'un objet.

Dans un contexte multivarié, le système observé comporte plusieurs objets ou individus, ce nombre sera noté p . Dans ce cas simple, la matrice des observations $p \times N$ sera construite en juxtaposant les observations pour chaque objet ou individu,

$$\mathbf{X} = (\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(N)) = \begin{bmatrix} X_1(1) & X_1(2) & \dots & X_1(N) \\ X_2(1) & X_2(2) & \dots & X_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ X_p(1) & X_p(2) & \dots & X_p(N) \end{bmatrix}$$

Le vecteur aléatoire à étudier est noté $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ où T définit l'opérateur transposé. Chacune des composantes de ce vecteur aléatoire est une variable aléatoire associée à un individu ou objet d'étude.

Par analogie au cas univarié, la loi gaussienne multivariée permet de caractériser un jeu de données avec seulement deux paramètres : le vecteur des moyennes et la matrice de variance-covariance.

Par définition, le vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ est gaussien, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, si sa densité de probabilité vérifie,

$$f(\mathbf{x}) = (2\pi)^{-\frac{1}{2}p} |\Sigma|^{-\frac{1}{2}} \exp^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X}-\boldsymbol{\mu})}.$$

Nous étudions dans ce manuscrit les propriétés temporelles des séries multivariées observées (modèles à longue mémoire ou longue dépendance) et leur représentation sous forme de graphes. Ces approches sont illustrées dans le cadre de la connectivité cérébrale observée par Imagerie par Résonance Magnétique fonctionnelle (IRMf).

1.1 Modèles à longue mémoire

Des exemples classiques d'analyse statistique de séries multivariées comprennent entre autres, les tests d'indépendance, l'estimation de matrice de covariance, les modèles linéaires généralisés et analyse de variance (voir par exemple les livres de référence (ANDERSON

2003 ; MUIRHEAD 1982)). Dans une grande majorité d'études, les observations sont supposées indépendantes ($\mathbf{X}(i)$ indépendant de $\mathbf{X}(j)$, pour tous $i, j, 1 \leq i \neq j \leq N$). Quand les observations ne sont pas indépendantes, on parlera de séries chronologiques, temporelles ou de processus, c'est souvent le cas en économie, finance, géoscience où les observations sont des mesures dans le temps (BOX et al. 2011 ; PERCIVAL et WALDEN 2000). Des exemples sont les modèles autorégressifs (AR), les modèles à moyenne mobile (MA), et bien d'autres. Ce sont aussi les modèles à longue mémoire qui seront étudiés dans la suite de ce manuscrit.

En effet, l'observation d'un système sur une longue période permet la modélisation de la fonction d'auto-covariance. Dans les cas classiques, par exemple ARMA (Autorégressifs à moyenne mobile), cette fonction d'auto-covariance décroît rapidement en l'infini, i.e. en termes mathématiques, la fonction d'autocovariance est sommable. Les modèles à longue mémoire permettent une autre modélisation de la fonction d'auto-covariance avec une décroissance lente en l'infini,

$$\lim_{\tau \rightarrow \infty} \frac{\text{Cov}(X_i(k), X_i(k + \tau))}{C|\tau|^{-\alpha}} = 1$$

où C est une constante. Quand $0 < \alpha < 1$, la fonction d'auto-covariance décroît très lentement quand τ tend vers l'infini, et elle n'est pas sommable. Cette modélisation engendre des difficultés dans l'estimation de la moyenne ou de la variance (BERAN 1994). Un exemple classique est le mouvement brownien fractionnaire, $B^H(t)$ est l'unique processus aléatoire gaussien centré et continu dont la variance est donnée par $E(B^H(s)B^H(t)) = C(|s|^{2H} + |t|^{2H} - |s - t|^{2H})/2$. Lorsque $H > 1/2$, le mouvement brownien fractionnaire présente de la longue dépendance.

1.2 Représentations graphiques

En outre, dans le cadre d'observations multivariées, il est souvent utile et nécessaire d'étudier la structure des données multivariées. La plus ancienne de ces techniques est l'analyse en composantes principales, une analyse linéaire des données qui permet de regrouper des composantes fortement corrélées entre elles et d'extraire des composantes le plus décorréélées possibles. Ceci permet souvent la réduction du nombre de composantes à étudier (voir par exemple (JOLLIFFE 2002)). Ce champ de recherche est très vaste et regroupe de nombreux développements comme l'analyse de facteur, l'analyse en composantes indépendantes, des approches non linéaires d'extraction de composantes et bien d'autres. Une alternative à ces approches consiste à étudier le lien entre la distribution conjointe des données et la structure des données sous la forme d'un graphe. Ce graphe aura pour nœud les composantes du vecteur d'observations, X_1, X_2, \dots, X_p , et pour arêtes le couplage entre les variables aléatoires.

Par exemple pour des variables gaussiennes, le couplage pour une paire de variables peut se formaliser comme une corrélation non nulle ou *suffisamment* grande. DEMPSTER (1972) a proposé une méthode de sélection de covariance qui permet, pour des variables gaussiennes, d'identifier les zéros de l'inverse de la matrice de covariance et ainsi de définir les paires de nœuds du graphe où il n'y a pas d'arêtes. C'est le cadre des modèles graphiques où le graphe représente la structure de dépendance conditionnelle des données. Un modèle graphique est ainsi défini par une famille de probabilités reposant sur une structure de graphe dirigé ou non dirigé, voir figure 1 (WHITTAKER 1990 ; LAURITZEN 1996). Dans ce manuscrit, seuls les graphes non dirigés sont étudiés, pour les graphes dirigés voir par exemple (PEARL 2000).

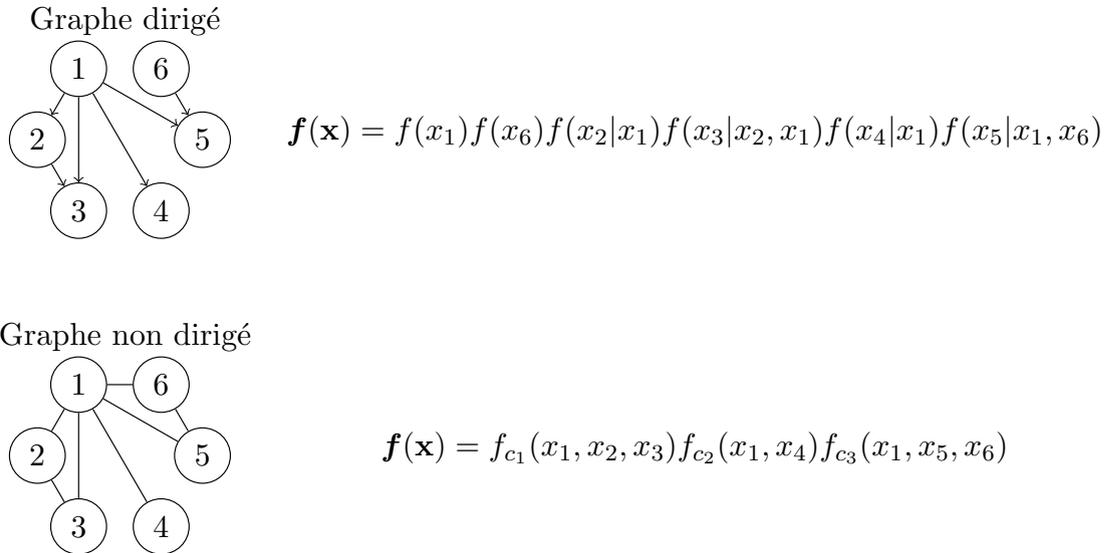


FIGURE 1 – Exemple de représentations de loi avec un graphe dirigé et un graphe non dirigé. Dans le cadre de graphe non dirigé, et avec des hypothèses sur la densité, $f(x_i|x_j, x_k)$ est représentée par la densité conditionnelle de x_i sachant x_j et x_k . Pour le graphe non dirigé, cette analogie n'est plus possible et f_{c_i} désigne une fonction positive associée à une clique c_i du graphe (JORDAN 2004).

1.3 Connectivité cérébrale

L'imagerie par résonance magnétique fonctionnelle (IRMf) est un moyen de mesurer l'activité cérébrale in vivo, de manière indirecte et non invasive par la mesure de l'effet BOLD (Blood Oxygen Level Dependent). Les acquisitions d'IRMf correspondent donc à l'enregistrement de séries temporelles pour chaque voxel (un voxel est un cube de quelques millimètres de côté) couvrant ensemble le volume entier du cerveau. Avec ces données, les neuroscientifiques sont intéressés par le fait de quantifier et caractériser les liens qui se mettent en place entre les parties anatomiques du cerveau, c'est ce qu'on appelle le réseau de connectivité cérébrale. La figure 2 illustre le processus implémenté pour l'extraction des réseaux de connectivité. De nombreux résultats ont été obtenus en clinique pour la caractérisation de perte de connectivité dans différentes pathologies : schizophrénie (LYNALL et al. 2010), Alzheimer (BUCKNER et al. 2009), hyperactivité (WANG et al. 2009), désordres de la conscience (CAUDA et al. 2009 ; ACHARD et al. 2012) ...

1.4 Plan du manuscrit

Ce manuscrit présente l'étude d'observations multivariées avec présence de longue mémoire ou longue dépendance. L'estimation de la structure des données est proposée sous la forme d'un graphe. L'objectif est d'extraire des caractéristiques les plus robustes et reproductibles possibles afin d'aider à l'interprétation de données cérébrales.

Ce manuscrit se compose de quatre parties qui s'enrichissent des problématiques en neurosciences :

- Le modèle longue mémoire multivarié et la corrélation par ondelettes
- La connectivité fractale
- L'estimation de la structure de graphe
- Méthodes statistiques de comparaisons des graphes

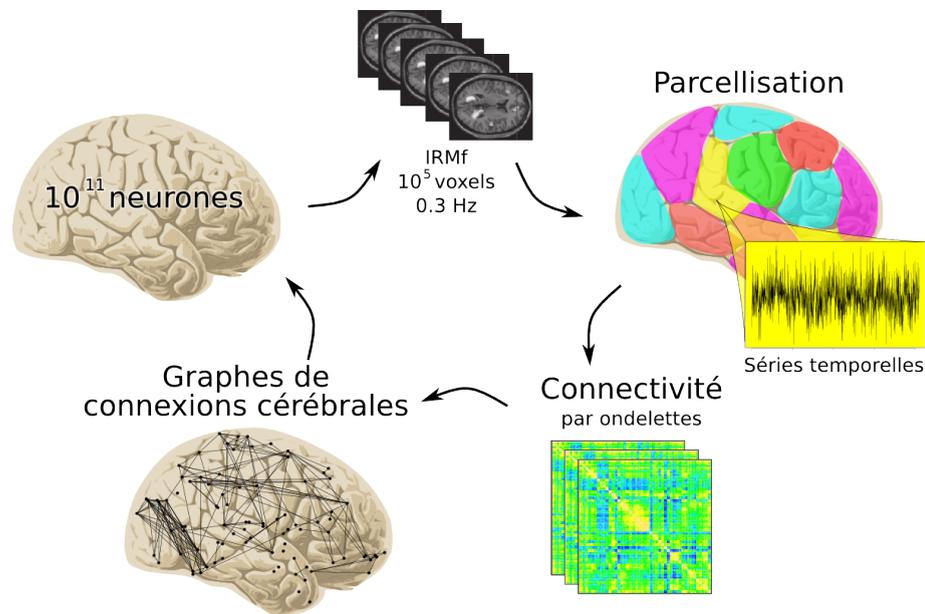


FIGURE 2 – Méthodes de construction des graphes de connectivité cérébrale

Dans la suite du document je vais détailler ces quatre parties en explicitant l’apport méthodologique et les applications sur données réelles. J’évoquerai pour chaque partie en dernier lieu les perspectives. Chaque partie a donné lieu à des publications qui sont résumées dans le tableau suivant, le texte publié est reproduit en annexe (les publications indiquées par * désignent celles effectuées dans le cadre d’un encadrement de thèse).

corrélation par ondelettes	(ACHARD et al. 2006) Annexe A.1 (COEURJOLLY et al. 2013)
connectivité fractale	(ACHARD et al. 2008) Annexe A.2 (ACHARD et GANNAZ 2016) Annexe A.8
structure de graphe	(AMINI et al. 2011 ; AMINI et al. 2010)* (COSTARD et al. 2014)* Annexe A.6 (HARLÉ et al. 2014 ; HARLÉ et al. 2015)* Annexe A.7
comparaisons des graphes	(ACHARD et al. 2012) Annexe A.3 (RICHIARDI et al. 2013) Annexe A.4
application médicale	(FALLANI et al. 2014) Annexe A.5 (ACHARD et BULLMORE 2007) (RIDLEY et al. 2015)

2 Longue dépendance, ondelettes et corrélation

Nous détaillons dans cette partie la définition du modèle à longue mémoire avec l’extension au cas multivarié. Les ondelettes sont particulièrement adaptées pour l’analyse des propriétés de séries temporelles à longue mémoire. Nous présentons les résultats existants pour l’étude de la corrélation de paire de séries temporelles.

2.1 Modèles à longue mémoire : de l'univarié au multivarié

2.1.1 La longue mémoire en statistiques : définitions et estimations

Les phénomènes longues dépendances (aussi appelés longues mémoires) permettent de décrire des structures d'auto-covariance dans lesquelles la dépendance décroît lentement quand le "retard" augmente. Pour une série temporelle, ceci se traduit empiriquement par le fait que les corrélations prises entre les observations décroissent vers zéro plus lentement que le taux attendu pour par exemple des observations indépendantes ou pour des modèles ARMA (AutoRegressive Moving Average) ou Markov (BERAN 1994; BERAN 2010). Ces phénomènes ont été décrits du point de vue statistique depuis les années 50 par les travaux de MANDELBROT et NESS (1968) avec l'introduction et l'étude du mouvement brownien fractionnaire (fBm). Puis les travaux de GRANGER et JOYEUX (1980) et HOSKING (1981) ont proposé les modèles ARIMA (autoregressive integrated moving average) qui sont une extension des modèles ARMA en ajoutant de la longue dépendance. Les différents modèles à longue mémoire sont décrits précisément dans le livre de PERCIVAL et WALDEN (2000). Enfin le livre de BERAN (1994) sur les statistiques pour les séries à longue mémoire rend compte des difficultés de définir des estimateurs robustes en présence de longue dépendance.

Une définition classique d'une série à longue mémoire s'exprime à l'aide de la densité spectrale.

Définition 1 (Longue dépendance ou longue mémoire). Soit $X = \{X(k), k \in \mathbb{Z}\}$ un processus à valeurs réelles, stationnaire au second-ordre. Notons f sa densité spectrale telle que,

$$\gamma(t) = E[(X(k) - E(X(k)))(X(k+t) - E(X(k+t)))] = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda.$$

Le processus X est dit à longue dépendance (au sens des corrélations) si il existe d , $0 < d < 1$ tel que la densité spectrale f vérifie

$$\lim_{\lambda \rightarrow 0} \frac{f(\lambda)}{C|\lambda|^{-2d}} = 1,$$

où C est une constante, $C > 0$.

Dans certains cas, cette définition est équivalente à la caractérisation d'une décroissance lente de la fonction d'auto-covariance (BERAN 1994, p. 43), ce qui justifie les termes de longue dépendance ou longue mémoire. Pour comparaison, on peut citer quelques processus à courte mémoire comme les processus ARMA (AutoRegressive Moving Average), et bien sûr, le bruit blanc gaussien.

Dans la suite de ce travail, nous nous sommes concentrés sur un modèle à différences finies qui a l'avantage de ne pas forcer la stationnarité des processus. Ce modèle dénoté $M(d)$, pour *memory of parameter d* (appelé dans la suite processus à mémoire de paramètre d) a été introduit par HURVICH. et RAY (1995) puis étudié en particulier par MOULINES et al. (2007) et MOULINES et al. (2008).

Tout d'abord, définissons l'opérateur à différences finies.

Définition 2 (opérateurs à différences finies). Notons $\Delta^D X$ le processus obtenu par la différence finie d'ordre D du processus X .

Pour $D = 1$, le processus à différences finies est $(\Delta X)(k) = X(k) - X(k-1)$, pour $k \in \mathbb{Z}$.

Puis pour les autres valeurs de D , le processus est obtenu par récurrence.

Pour $D = 2$, le processus à différences finies est $(\Delta^2 X)(k) = (\Delta(\Delta X))(k) = X(k) - 2X(k-1) + X(k-2)$, pour $k \in \mathbb{Z}$.

Finalement, pour D entier quelconque, $(\Delta^D X)(k) = (\Delta^{D-1}(\Delta X))(k)$

Nous définissons aussi l'espace fonctionnel suivant,

Définition 3. Soit deux réels β et L tels que, $0 < \beta \leq 2$ et $L > 0$, notons $\mathcal{H}(\beta, L)$ l'ensemble des fonctions suivant

$$\mathcal{H}(\beta, L) = \{g \text{ non négative, symétrique définie sur }]-\pi, \pi[\text{ telle que} \\ \text{pour tout } \lambda \in (-\pi, \pi), |g(\lambda) - g(0)| \leq L|g(0)||\lambda|^\beta\}. \quad (1)$$

Cet espace $\mathcal{H}(\beta, L)$ permet de définir des fonctions non négatives symétriques, bornées sur $]-\pi, \pi[$ et bornées par des valeurs non nulles sur un voisinage de l'origine.

Le processus $M(d)$ est maintenant défini en imposant la stationnarité pour un ordre fini de différences et en modélisant les phénomènes courtes mémoires par les fonctions appartenant à $\mathcal{H}(\beta, L)$.

Définition 4 (processus $M(d)$). Soit $X = \{X(k), k \in \mathbb{Z}\}$ un processus à valeurs réelles, non nécessairement stationnaire. Le processus X est dit processus mémoire de paramètre d , noté $M(d)$, si pour tout entier $D > d - 1/2$, le processus à différences finies d'ordre D , $\Delta^D X$ est stationnaire au second ordre, et sa densité spectrale est définie par,

$$f_{\Delta^D X}(\lambda) = |1 - e^{-i\lambda}|^{2(D-d)} f^*(\lambda), \text{ pour } \lambda \in]-\pi, \pi[,$$

où $f^* \in \mathcal{H}(\beta, L)$ est une fonction non négative symétrique, bornée sur $]-\pi, \pi[$ et bornée non nulle sur un voisinage de l'origine. Cette fonction modélise la courte mémoire du processus.

Quelques remarques sont à faire pour cette définition.

Si $d < 1/2$ Le processus X est stationnaire au second ordre et sa densité spectrale est donnée par

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \text{ pour } \lambda \in]-\pi, \pi[. \quad (2)$$

Si $0 < d < 1/2$, le processus X est dit à longue dépendance. Si $d = 0$, le processus est dit à courte mémoire. Si $d < 0$, le processus X est dit à mémoire négative. Enfin, si $d < -1/2$ Le processus X est non inversible.

Si $d > 1/2$ Le processus X est non stationnaire. Dans ce cas, la fonction (2) n'est pas intégrable sur $]-\pi, \pi[$ et n'est donc pas une densité spectrale. Cependant, en utilisant la notion de densité spectrale généralisée (YAGLOM 1958), on peut utiliser la forme (2) pour des processus non stationnaires (MOULINES et al. 2007).

Comme décrit dans MOULINES et al. (2007), selon le choix de la fonction f^* , cette définition 4 de processus $M(d)$ contient la classe des processus classiquement utilisés comme ARIMA et ARFIMA.

Ces modèles ont suscité l'intérêt du point de vue statistique pour la définition d'estimateur pour le paramètre d . On peut citer des approches paramétriques YAJIMA (1985), FOX et TAQQU (1986), DAHLHAUS (1989) et GIRAITIS et al. (1997). En considérant la forme de

la densité spectrale dans la définition 1, on remarque qu'en prenant le logarithme, la décroissance de la densité spectrale au voisinage de zéro est linéaire par rapport au paramètre d . Cette constatation est à l'origine de nombreux estimateurs dits semi-paramétriques basés sur le périodogramme (GEWEKE et PORTER-HUDAK 1983; ROBINSON 1994; ROBINSON 1995a). Puis, d'autres estimateurs ont été proposés en utilisant une décomposition en ondelettes au lieu de la transformée de Fourier (FLANDRIN 1992; ABRY et VEITCH 1998; PERCIVAL et WALDEN 2000; MOULINES et al. 2008).

2.1.2 Le cas multivarié

Dans le cadre du traitement de données réelles, il est de plus en plus fréquent d'être confronté à des données multivariées présentant par exemple les enregistrements de capteurs dans plusieurs points de l'espace. On peut citer les données météorologiques avec des capteurs de températures dans différentes parties du globe terrestre (PERCIVAL et WALDEN 2000), les données cérébrales avec l'enregistrement de différentes parties du cerveau simultanément (BUZSÁKI 2006). Ces séries temporelles présentent souvent de la longue mémoire, mais aussi des interactions qui restent à identifier. Il est donc nécessaire de s'interroger sur la définition d'un processus multivarié à longue dépendance. Dans ce travail, nous nous sommes concentrés sur l'extension au cas multivarié des processus $M(d)$ (CHAMBERS 1995; MOULINES et al. 2007; ACHARD et al. 2008). La définition mathématique est donnée dans la définition suivante 5, on peut résumer cette définition par le fait que chaque composante du processus est $M(d)$, suivant la définition 4, et la matrice de densité spectrale est le produit d'une matrice diagonale et d'une matrice de covariance (formule (5)).

Définition 5 (Processus $M(d)$ multivarié). Soit $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)^T$ un processus à valeurs réelles. Soit $\mathbf{d} = (d_1, d_2, \dots, d_p)$ un vecteur, $\mathbf{d} \in \mathbb{R}^p$.

Le processus multivarié \mathbf{X} est dit $M(\mathbf{d})$, avec pour paramètres de mémoire le vecteur \mathbf{d} , si pour tout $\mathbf{D} > \mathbf{d} - 1/2$, $\mathbf{Z} = \text{diag}(\Delta^{D_\ell}, \ell = 1, \dots, p)\mathbf{X}$ est stationnaire au second ordre et sa matrice spectrale est définie par, pour tout (ℓ, m) , $1 \leq \ell, m \leq p$,

$$f_{\ell,m}^{(D_\ell, D_m)}(\lambda) = \frac{1}{2\pi} \Omega_{\ell,m} (1 - e^{-i\lambda})^{-d_\ell} (1 - e^{i\lambda})^{-d_m} f_{\ell,m}^S(\lambda), \quad \lambda \in [-\pi, \pi], \quad (3)$$

où les paramètres de mémoire sont donnés par $d_m^S = d_m - D_m$ pour tout m . Les fonctions $f_{\ell,m}^S(\cdot)$ correspondent à la modélisation des propriétés courtes mémoires du processus. Dans le cas multivarié, les fonctions $f_{\ell,m}^S(\cdot)$, vérifient la propriété analogue au cas multivarié avec en plus $f_{\ell,m}^S(0) = 1$ pour tout $(\ell, m) \in \{1, \dots, p\}^2$ et pour tout $\lambda \in (-\pi, \pi)$, $\max\{|\mathbf{f}^S(\lambda) - 1|, (\ell, m) \in \{1, \dots, p\}^2\}_\infty \leq L|\lambda|^\beta$, où \mathbf{f}^S est la version matricielle de $f_{\ell,m}^S$ avec $1 \leq \ell, m \leq p$. On notera de la même façon que dans le cas univarié, $\mathbf{f}^S(\cdot) \in \mathcal{H}(\beta, L)$ avec $0 < \beta \leq 2$ et $0 < L$

Cette définition est aussi valable pour des processus non stationnaires, et la densité spectrale généralisée des processus X_ℓ et X_m est définie par, pour tout (ℓ, m) , $1 \leq \ell, m \leq p$,

$$f_{\ell,m}(\lambda) = \frac{1}{2\pi} \Omega_{\ell,m} (1 - e^{-i\lambda})^{-d_\ell} (1 - e^{i\lambda})^{-d_m} f_{\ell,m}^S(\lambda), \quad \lambda \in [-\pi, \pi]. \quad (4)$$

Remarque. Nous pouvons donc exprimer la matrice de densité spectrale du processus multivarié \mathbf{X} par,

$$\mathbf{f}(\lambda) = \mathbf{\Omega} \circ (\mathbf{\Lambda}^0(\mathbf{d}) \mathbf{f}^S(\lambda) \mathbf{\Lambda}^0(\mathbf{d})^*), \quad \lambda \in [-\pi, \pi], \quad \text{avec } \mathbf{\Lambda}^0(\mathbf{d}) = \text{diag}((1 - e^{-i\lambda})^{-\mathbf{d}}) \quad (5)$$

où $\mathbf{d} = \mathbf{D} + \mathbf{d}^s$. Le symbole $*$ désigne l'opérateur conjugué et \circ désigne le produit de Hadamard.

Remarque (connectivité fractale). Dans le contexte semi-paramétrique dans lequel nous nous plaçons, la matrice $\mathbf{\Omega}$ est appelée *connectivité fractale* (ACHARD et al. 2008) ou bien *matrice de covariance basse-fréquence* (ROBINSON 2005).

2.2 Ondelettes et corrélation

L'estimation en présence de longue dépendance est complexe car les auto-corrélations à longues portées augmentent la variance des estimateurs. Ce phénomène s'illustre particulièrement simplement sur l'estimation de la moyenne (BERAN 1994).

Soient X_1, \dots, X_N , des variables aléatoires réelles suivant la même loi que la variable aléatoire X . Supposons que les moments d'ordre 1 et 2 de X sont finis et notés $\mu = \mathbb{E}(X)$ et $\sigma^2 = \mathbb{V}(X)$. Alors, la variance de l'estimateur de la moyenne empirique, noté $\widehat{X} := N^{-1} \sum_{i=1}^N X_i$ s'écrit :

$$\mathbb{V}(\widehat{X}) = \frac{\sigma^2}{N^2} \sum_{i,j=1}^N \rho_{i,j}$$

où $\rho_{i,j}$ désigne la corrélation de X_i et X_j . Dans le cas classique d'un échantillon de variables aléatoires, les corrélations sont nulles et on retrouve le résultat classique $\mathbb{V}(\widehat{X}) = \frac{\sigma^2}{N}$. Lorsque les variables aléatoires X_1, \dots, X_N ne sont plus indépendantes, la variance de la moyenne empirique peut-être fortement dégradée. Dans le cas d'un processus ARMA ou Markov, $\lim_{N \rightarrow \infty} \sum_{i \neq j}^N \rho_{i,j}$ existe et celle-ci est finie. On peut donc montrer que la variance de l'estimateur de la moyenne empirique tend vers 0, avec une constante qui peut détériorer l'estimation mais qui est calculable. Par contre, dans le cas de processus à longue mémoire comme définis dans 1, $\lim_{N \rightarrow \infty} \sum_{i \neq j}^N \rho_{i,j}$ est infinie, et la convergence vers 0 de la variance n'est plus assurée quand le nombre d'observations tend vers l'infini. De manière similaire, ces mêmes remarques peuvent être faites avec l'estimateur de covariance ou de corrélation avec les mêmes contraintes sur le contrôle de la dépendance des observations.

Dans le cas de processus à longue mémoire univarié, une manière de prendre en compte cette difficulté est d'utiliser une décomposition en ondelettes. Celles-ci sont très bien adaptées à cause des propriétés de décomposition d'échelle. De nombreux auteurs se sont intéressés aux propriétés des coefficients d'ondelettes en présence de longue mémoire. Les références sont nombreuses, on peut citer les livres (FLANDRIN 1998), et (PERCIVAL et WALDEN 2000) et la revue (TORRENCE et COMPO 1998).

L'analyse en ondelettes pour la caractérisation de la variance ou covariance est utilisée comme une analyse échelle par échelle et le but est de faire des tests d'homogénéité de variance ou bien d'identifier des échelles qui présentent un comportement différent (PERCIVAL et WALDEN 2000).

2.2.1 Définition ondelettes et corrélation

Soit $(\phi(\cdot), \psi(\cdot))$ respectivement une ondelette père et une ondelette mère. Leurs transformées de Fourier sont définies par $\widehat{\phi}(\lambda) = \int_{-\infty}^{\infty} \phi(t)e^{-i\lambda t} dt$ et par $\widehat{\psi}(\lambda) = \int_{-\infty}^{\infty} \psi(t)e^{-i\lambda t} dt$.

Pour une résolution donnée, $j \geq 0$, pour $k \in \mathbb{Z}$, les fonctions dilatées et translatées sont définies par, $\phi_{j,k}(\cdot) = 2^{-j/2} \phi(2^{-j} \cdot -k)$ et $\psi_{j,k}(\cdot) = 2^{-j/2} \psi(2^{-j} \cdot -k)$. La même convention que dans MOULINES et al. (2007) et MOULINES et al. (2008) est utilisée dans la suite (avec

cette convention, les petites échelles d'ondelettes correspondent aux hautes fréquences et réciproquement, les grandes échelles aux basses fréquences) .

Soit $\tilde{\mathbf{X}}(t) = \sum_{k \in \mathbb{Z}} \mathbf{X}(k) \phi(t - k)$. Les coefficients d'ondelettes du processus \mathbf{X} sont définis par

$$\mathbf{W}_{j,k} = \int_{\mathbb{R}} \tilde{\mathbf{X}}(t) \psi_{j,k}(t) dt \quad j \geq 0, k \in \mathbb{Z}. \quad (6)$$

Pour $j \geq 0$ et $k \in \mathbb{Z}$, $\mathbf{W}_{j,k}$ est un vecteur de dimension p ,

$$\mathbf{W}_{jk} = (W_{j,k}(1) \quad W_{j,k}(2) \quad \dots \quad W_{j,k}(p))$$

où $W_{j,k}(\ell) = \int_{\mathbb{R}} \tilde{X}_\ell(t) \psi_{j,k}(t) dt$.

Les hypothèses suivantes expriment les conditions vérifiées par la transformation d'ondelettes, elles seront supposées vraies dans la suite.

(W1) Les fonctions $\phi(\cdot)$ et $\psi(\cdot)$ sont intégrables, à support compact, $\int_{\mathbb{R}} \phi(t) dt = 1$ et $\int \psi^2(t) dt = 1$;

(W2) Il existe $\alpha > 1$ tel que $\sup_{\lambda \in \mathbb{R}} |\hat{\psi}(\lambda)| (1 + |\lambda|)^\alpha < \infty$, *i.e.* l'ondelette est α -régulière ;

(W3) L'ondelette mère $\psi(\cdot)$ a $M > 1$ moments nuls.

(W4) La fonction $\sum_{k \in \mathbb{Z}} k^\ell \phi(\cdot - k)$ est polynomiale de degré ℓ pour tout $\ell = 1, \dots, M - 1$.

(W5) Pour tout $i = 1, \dots, p$, $(1 + \beta)/2 - \alpha < d_i \leq M$.

Comme expliqué dans MOULINES et al. (2007), ces conditions ne sont pas restrictives et de nombreuses définitions classiques d'ondelettes vérifient ces hypothèses. Parmi celles-ci, on peut citer les ondelettes de Daubechies qui ont un support compact paramétrisées par le nombre de moments nuls. Elles sont α -régulières, avec α fonction croissante de M tendant vers l'infini (voir DAUBECHIES (1992)). Par conséquent, les ondelettes de Daubechies vérifient (W1)-(W5) pour M suffisamment grand.

Remarque. Le couple de fonctions $(\phi(\cdot), \psi(\cdot))$ peut-être associé à une analyse multirésolution, mais cette condition n'est pas nécessaire. De même, l'orthogonalité des ondelettes $\{\psi_{j,k}(\cdot)\}$ n'est pas requise (voir MOULINES et al. (2007), Section 3).

Plus précisément en suivant les arguments donnés dans (MOULINES et al. 2007), les coefficients d'ondelettes sont obtenus par convolution discrète et sous-échantillonnage, pour tout ℓ , $0 \leq \ell \leq p$,

$$W_{j,k}(\ell) = \sum_{r \in \mathbb{Z}} X_\ell(r) h_{j,2^j k - r} = (\downarrow^j [h_{j,\cdot} \star X_\ell])_k, \quad (7)$$

où \star est la convolution de séries discrètes, et

$$h_{j,k} = 2^{-j/2} \int \phi(t + k) \psi(2^{-j} t) dt,$$

où \downarrow^j est l'opérateur de sous-échantillonnage défini par $(\downarrow^j c)_k = c_{2^j k}$ avec $\{c_k\}_{k \in \mathbb{Z}}$ une série discrète.

Par les conditions (W3) et (W4), on montre que les coefficients d'ondelettes peuvent être obtenus directement à partir de $\Delta^M X_\ell$, le processus des différences d'ordre M .

Pour cela, il faut remarquer que par (W3) et (W4), la transformée de Fourier discrète de $\{h_{j,l}\}_{l \in \mathbb{Z}}$ s'écrit comme le produit suivant,

$$H_j(\lambda) = (1 - e^{-i\lambda})^M \tilde{H}_j(\lambda),$$

où $\tilde{H}_j(\lambda)$ est aussi un polynôme trigonométrique. Et finalement,

$$W_{j,k}(\ell) = (\downarrow^j [\tilde{h}_{j,\cdot} \star \Delta^M X_\ell])_k, \quad (8)$$

où $\tilde{h}_{j,\cdot}$ sont les coefficients du polynôme trigonométrique \tilde{H}_j .

En pratique, les observations sont de taille finie égale à N , en supposant que le support de ψ est inclus dans $[0, T_\psi]$ avec $T_\psi > 1$, les coefficients d'ondelettes sont définis pour tout k tel que $0 < k < n_j$ où $n_j = \max(0, 2^{-j}(N - T_\psi + 1))$. En dehors de ces conditions, les coefficients d'ondelettes sont pris égaux à zéro.

Les propriétés des ondelettes sont donc particulièrement intéressantes et adaptées dans le cadre de processus $M(d)$. En effet, les premiers résultats obtenus par ABRY et VEITCH (1998) et MOULINES et al. (2007) dans le cadre de signaux univariés puis prolongés dans le cas de signaux multivariés permettent d'obtenir la stationnarité des coefficients d'ondelettes pour des processus $M(d)$ (ACHARD et GANNAZ 2016), avec des hypothèses supplémentaires entre \mathbf{d} et M . La proposition suivante est tirée de MOULINES et al. (2007) et ACHARD et GANNAZ (2016).

Proposition 1 (stationnarité des coefficients d'ondelettes d'un processus $M(d)$). *Soit \mathbf{X} un processus $M(d)$ de densité spectrale (5). Alors sous les hypothèses (W1)-(W5), pour tout ℓ , $1 \leq \ell \leq p$, $W_{j,k}(\ell)$ est un processus stationnaire de covariance égale à*

$$\text{Cov}(W_{j,k}(\ell), W_{j,k'}(\ell)) = \int_{-\pi}^{\pi} D_{0;0}^{(j)}(\lambda; (\ell, \ell)) e^{i\lambda(k-k')} d\lambda.$$

où $D_{0;0}^{(j)}(\lambda; (\ell, \ell))$ est la densité spectrale de $W_{j,k}(\ell)$.

Pour tous ℓ, m , $1 \leq \ell, m \leq p$, le processus bivarié $\{W_{j,k}(\ell), W_{j,k'}(m)\}_{k \in \mathbb{Z}}$ est stationnaire avec,

$$\text{Cov}(W_{j,k}(\ell), W_{j,k'}(m)) = \int_{-\pi}^{\pi} D_{0;0}^{(j)}(\lambda; (\ell, m)) e^{i\lambda(k-k')} d\lambda.$$

où $D_{0;0}^{(j)}(\lambda; (\ell, m))$ est la densité spectrale jointe de $\{W_{j,k}(\ell), W_{j,k}(m)\}_{k \in \mathbb{Z}}$.

Sous l'hypothèse (W3), la transformée d'ondelettes agit sur le processus en produisant une différentiation à l'ordre M . Il est donc tout à fait adéquate d'appliquer cette méthode pour des processus non stationnaires. Ceci n'est pas le cas avec la transformée de Fourier. Il est alors nécessaire d'appliquer des procédures dites de *fenêtrage* (ou *tapering*) (VELASCO et ROBINSON 2000). Des alternatives ont aussi été proposées récemment, différentiation des séries avant estimation (HURVICH et CHEN 2000) ou bien une extension de Fourier au cas non stationnaire (ABADIR et al. 2007; NIELSEN 2011).

La proposition 1 permet de déduire des résultats de convergence sur les variances, covariances ou corrélations des coefficients d'ondelettes à chaque échelle. Nous définissons donc tout d'abord ces quantités.

Définition 6 (variance, covariance et corrélation). Soit \mathbf{X} un processus $M(d)$ de densité spectrale (5). Les coefficients d'ondelettes sont notés $\mathbf{W}_{jk} = (W_{j,k}(1) \ W_{j,k}(2) \ \dots \ W_{j,k}(p))$ (voir formule (6)).

La variance par ondelettes à l'échelle j du processus $X_\ell = \{X_\ell(k); k \in \mathbb{Z}\}$ est définie par la variance des coefficients d'ondelettes $\{W_{j,k}(\ell)\}_{k \in \mathbb{Z}}$ à l'échelle j .

On notera la variance par

$$\sigma_\ell^2(j) = \mathbb{V}(W_{j,k}(\ell)).$$

$\theta_{\ell,m}(j)$ désigne la covariance par ondelettes à l'échelle j entre les processus X_ℓ et X_m , et elle est définie par

$$\theta_{\ell,m}(j) = \text{Cov}(W_{j,k}(\ell), W_{j,k}(m)).$$

Enfin, la corrélation à l'échelle j , entre les processus X_ℓ et X_m est définie par,

$$\rho_{\ell,m}(j) = \frac{\theta_{\ell,m}(j)}{\sigma_\ell(j)\sigma_m(j)}$$

2.2.2 Estimation et propriétés asymptotiques

D'après les travaux de (PERCIVAL 1995) et (WHITCHER et al. 2000) pour des processus à différences finies, les résultats de convergence sur la variance, covariance ou corrélation des coefficients d'ondelettes à chaque échelle sont détaillés. Ces résultats permettent ainsi d'établir des intervalles de confiance et donc des tests d'hypothèses qui seront exploités dans la section 4.2.

Tout d'abord, concernant la convergence asymptotique de la variance,

Proposition 2 (Convergence de la variance par ondelettes). *Soit $X = \{X(k), k \in \mathbb{Z}\}$ un processus à valeurs réelles, $M(d)$. Notons $\{W_{j,k}\}_{k \in \mathbb{Z}}$ les coefficients d'ondelettes à l'échelle j de X .*

Sous les hypothèses (W1)-(W5), et sous l'hypothèse que le processus $\{W_{j,k}\}_k$ est un processus gaussien de densité spectrale $D_{0;0}^{(j)}$ telle que $D_{0;0}^{(j)}(f) > 0$ presque partout, l'estimateur empirique de la variance des coefficients d'ondelettes à l'échelle j est non biaisé et suit asymptotiquement une loi normale :

$$\hat{\sigma}^2(j) := \frac{1}{n_j} \sum_{k=0}^{n_j} W_{j,k}^2 \xrightarrow{\mathcal{L}} \mathcal{N}\left(\sigma^2(j), \frac{2A_j}{n_j}\right) \quad (9)$$

où $A_j = \int_{-\pi}^{\pi} (D_{0;0}^{(j)}(f))^2 df$.

En ce qui concerne la covariance, il est nécessaire de contrôler le comportement du processus produit, $\{W_{j,k}(\ell)W_{j,k}(m)\}_{k \in \mathbb{Z}}$. Or d'après WECKER (1978), le produit de deux séries stationnaires est stationnaire avec des hypothèses supplémentaires sur les moments d'ordre 3 et 4, ou bien en supposant simplement que les processus sont gaussiens. Nous nous placerons dans ce dernier cas pour la suite (WHITCHER et al. 2000).

Proposition 3 (Convergence de la covariance par ondelettes). *Soit $\mathbf{X} = \{\mathbf{X}(k), k \in \mathbb{Z}\}$ un processus à valeurs réelles, $M(d)$. Notons, pour tout ℓ , $1 \leq \ell \leq p$, $\{W_{j,k}(\ell)\}_{k \in \mathbb{Z}}$ les coefficients d'ondelettes à l'échelle j de X_ℓ . Nous faisons les hypothèses que, pour tout ℓ , $1 \leq \ell \leq p$, le processus $\{W_{j,k}(\ell)\}_k$ est un processus gaussien de densité spectrale $D_{0;0}^{(j)}(\cdot; \ell, \ell)$ telle que $D_{0;0}^{(j)}(f; \ell, \ell) > 0$ presque partout, et le processus bivarié $\{W_{j,k}(\ell), W_{j,k}(m)\}_k$ est gaussien de densité spectrale $D_{0;0}^{(j)}(\cdot; \ell, m)$ telle que $D_{0;0}^{(j)}(f; \ell, m) > 0$ presque partout.*

Sous les hypothèses (W1)-(W5), l'estimateur empirique de la covariance des coefficients d'ondelettes à l'échelle j est non biaisé et suit asymptotiquement une loi normale :

$$\hat{\theta}_{\ell,m}(j) := \frac{1}{n_j} \sum_{k=0}^{n_j} (W_{j,k}(\ell)W_{j,k}(m)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\theta_{\ell,m}(j), \frac{2AA_j}{n_j}\right) \quad (10)$$

où $AA_j = \int_{-\pi}^{\pi} D_{0;0}^{(j)}(f; \ell, \ell)D_{0;0}^{(j)}(f; m, m)df + \int_{-\pi}^{\pi} (D_{0;0}^{(j)}(f; \ell, m))^2 df$.

Par conséquent, WHITCHER et al. (2000) en déduit la convergence de l'estimateur des corrélations.

Proposition 4 (Convergence de la corrélation par ondelettes). *Soit $\mathbf{X} = \{\mathbf{X}(k), k \in \mathbb{Z}\}$ un processus à valeurs réelles, $M(d)$. Notons, pour tout ℓ , $1 \leq \ell \leq p$, $W_{j,k}(\ell)_{k \in \mathbb{Z}}$ les coefficients d'ondelettes à l'échelle j de X_ℓ . Nous faisons les hypothèses que, pour tout ℓ , $1 \leq \ell \leq p$, le processus $\{W_{j,k}(\ell)\}_k$ est un processus gaussien de densité spectrale $D_{0;0}^{(j)}(\cdot; \ell, \ell)$ telle que $D_{0;0}^{(j)}(f; \ell, \ell) > 0$ presque partout, et le processus bivarié $\{W_{j,k}(\ell), W_{j,k}(m)\}_k$ est gaussien de densité spectrale $D_{0;0}^{(j)}(\cdot; \ell, m)$ telle que $D_{0;0}^{(j)}(f; \ell, m) > 0$ presque partout.*

Sous les hypothèses (W1)-(W5), l'estimateur empirique de la corrélation des coefficients d'ondelettes à l'échelle j est non biaisé et suit asymptotiquement une loi normale :

$$\widehat{\rho}_{\ell,m}(j) := \frac{\widehat{\theta}_{\ell,m}}{\widehat{\sigma}_{\ell}(j)\widehat{\sigma}_m(j)} = \frac{\sum_{k=0}^{n_j} (W_{j,k}(\ell)W_{j,k}(m))}{(\sum_{k=0}^{n_j} W_{j,k}^2(\ell) \sum_{k=0}^{n_j} W_{j,k}^2(m))^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(\rho_{\ell,m}(j), \Sigma_j^{cc}) \quad (11)$$

où Σ_j^{cc} est explicitée dans la formule (5.4) de WHITCHER (1998).

L'expression de la variance asymptotique de la corrélation par ondelettes est complexe, mais en utilisant la transformée de Fisher, l'expression exacte de Σ_j^{cc} n'est pas utilisée. En effet, la transformée de Fisher permet de stabiliser la variance de l'estimation de la corrélation et dans notre cas d'obtenir des intervalles de confiance particulièrement simples (ANDERSON 2003). Notons z la transformée de Fisher, pour tout x , $-1 \leq x \leq 1$,

$$z(x) := \frac{1}{2} \log \frac{1+x}{1-x} = \tanh^{-1}(x).$$

Proposition 5 (Convergence de la corrélation avec Fisher). *Sous les hypothèses précédentes (voir proposition 4), l'estimateur de la corrélation empirique vérifie,*

$$\sqrt{(n_j - 3)}(z(\widehat{\rho}_{\ell,m}(j)) - z(\rho_{\ell,m}(j))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (12)$$

Cette dernière proposition permet d'établir des intervalles de confiance et des tests d'hypothèses sur les valeurs de corrélations à chaque échelle d'ondelettes. Ces résultats sont utilisés dans la suite du manuscrit en particulier dans la section 4 où est abordée la représentation sous forme de graphe d'observations multivariées.

3 Connectivité fractale

Nous avons vu dans la section 2 que les ondelettes sont un outil d'analyse particulièrement adapté à l'étude de processus $M(d)$. Les résultats de convergence des estimateurs de corrélation par ondelettes permettent en particulier d'étudier des phénomènes échelle par échelle. Pourtant l'extension au cas multivarié n'est pas direct et nécessite quelques précautions dans l'interprétation de la corrélation en particulier.

D'après la proposition 1, la covariance par ondelettes dans le cas du modèle $M(d)$ s'écrit,

$$\theta_{\ell,m}(j) = \int_{-\pi}^{\pi} (1 - e^{-i\lambda})^{-d_\ell} (1 - e^{i\lambda})^{-d_m} \Omega_{\ell,m} f_{\ell,m}^S(\lambda) |\mathbb{H}_j(\lambda)|^2 d\lambda,$$

où \mathbb{H}_j est la fonction de gain du filtre par ondelettes (ACHARD et GANNAZ 2016).

Il est alors évident que la corrélation estimée par les coefficients d'ondelettes est fonction des paramètres Ω et \mathbf{d} . Les figures suivantes illustrent ce phénomène qui est appelé

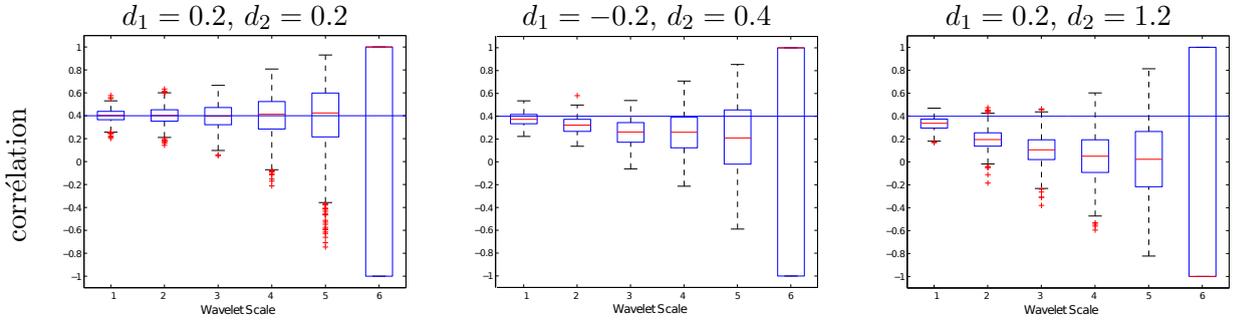


FIGURE 3 – Illustration de l’estimation de la corrélation par ondelettes quand les paramètres longues mémoires des séries sont différents. Chaque graphique représente la simulation de processus bivariés $M(d)$ avec $\Omega_{12} = 0.4$. Plus la différence entre les paramètres d est proche de 1, plus la corrélation par ondelettes tend vers 0 aux grandes échelles.

l’effet de la différence de phase dans le cas multivarié.

Nous avons en particulier montré la proposition suivante (ACHARD et GANNAZ 2016).

Proposition 6. *Pour tout $j \geq 0$, définissons $K(\cdot)$ par*

$$K(\delta) = \int_{-\infty}^{\infty} \lambda^{-\delta} |\hat{\psi}(\lambda)|^2 d\lambda, \quad \delta \in (-\alpha, M).$$

Sous les hypothèses (W1)-(W5), il existe une constante C_0 dépendant de β , $\min_i d_i$, $\max_i d_i$, ϕ et ψ tel que, pour tout $j \geq 0$,

$$|\theta_{\ell,m}(j) - \Omega_{\ell,m} f_{\ell,m}^S(0) 2^{j(d_\ell + d_m)} \cos(\pi(d_\ell - d_m)/2) K(d_\ell + d_m)| \leq C_0 f_{\ell,m}^S(0) L 2^{(d_\ell + d_m - \beta)j}. \quad (13)$$

Cette proposition montre en particulier que dans le cas bivarié, plus la différence entre les paramètres longues mémoires s’approchent de 1 modulo 2, plus la covariance (et par conséquent la corrélation) des coefficients d’ondelettes va tendre vers 0, même avec une matrice Ω non nulle.

Prenant en compte ces phénomènes, nous nous sommes intéressées à l’estimation des paramètres du modèle multivarié $M(d)$ à l’aide d’une décomposition en ondelettes.

3.1 Estimation des paramètres du modèle, et le retard de phase

L’estimation des paramètres de ces modèles a été étudiée depuis la fin des années 90 avec les travaux de ROBINSON (1995b), LOBATO (1999), SHIMOTSU (2007). Un paramètre important de l’extension au cas multivarié est appelé le retard de phase ou *phase shift*. Comme expliqué dans (SHIMOTSU 2007), il est possible d’approcher la matrice de densité spectrale de différentes manières.

En effet, en remarquant que $1 - e^{i\lambda} = 2 \sin(\lambda/2) e^{i(\pi-\lambda)/2}$, lorsque λ tend vers zéro, le deuxième terme de l’égalité peut être approché au premier ordre par $|\lambda|$ ou bien au second ordre par $\lambda e^{-i(\pi-\lambda)/2}$.

Dans (LOBATO 1999), l’auteur utilise l’approximation au premier ordre et donc il adopte la définition suivante,

$$f(\lambda) = \Omega \circ (\Lambda^0(\mathbf{d}) f^S(\lambda) \Lambda^0(\mathbf{d})^*), \quad \lambda \in [-\pi, \pi], \quad \text{avec } \Lambda^0(\mathbf{d}) = \text{diag}(|\lambda|)^{-\mathbf{d}} \quad (14)$$

Cette spécification de la densité spectrale est erronée par rapport au modèle général (5), mais comme montré par LOBATO (1999) et remarqué aussi par SHIMOTSU (2007), l'estimateur des paramètres \mathbf{d} défini à partir de l'approximation au premier ordre est consistant et converge asymptotiquement.

SHIMOTSU (2007) propose d'utiliser l'approximation au second ordre, et la densité spectrale est donc définie par,

$$\mathbf{f}(\lambda) = \mathbf{\Omega} \circ (\mathbf{\Lambda}^0(\mathbf{d})\mathbf{f}^S(\lambda)\mathbf{\Lambda}^0(\mathbf{d})^*), \quad \lambda \in [-\pi, \pi], \quad \text{avec } \mathbf{\Lambda}^0(\mathbf{d}) = \text{diag}(\lambda^{-\mathbf{d}}e^{-i(\pi-\lambda)\mathbf{d}/2}) \quad (15)$$

Dans (SHIMOTSU 2007), l'auteur définit l'estimateur des paramètres du modèle à longue mémoire par l'optimisation d'une fonction de vraisemblance basée sur la matrice de densité spectrale. Cette optimisation se fait en deux étapes, avec tout d'abord l'estimation des paramètres longues mémoires et ensuite de la matrice de covariance.

3.1.1 Approximation par Whittle et ondelettes

Les paramètres du modèle $M(d)$ sont donc les paramètres longues mémoires \mathbf{d} et la matrice $\mathbf{\Omega}$ qui déterminent les relations entre les séries temporelles. Les premiers travaux datent des années 90 avec des approches paramétriques (HURVICH. et RAY 1995), puis dans le domaine fréquentiel utilisant une transformée de Fourier (LOBATO 1999). Enfin les premiers résultats sur une convergence asymptotique en prenant en compte le phase-shift sont présentés dans (SHIMOTSU 2007). En utilisant la décomposition en ondelettes, nous avons défini et étudié un nouvel estimateur (ACHARD et GANNAZ 2016).

Soient $j_1 \geq j_0 \geq 1$ respectivement les niveaux de résolution maximal et minimal utilisés dans l'estimation. L'estimation est écrite à partir des vecteurs de coefficients d'ondelettes $\{\mathbf{W}_{j,k}, j_0 \leq j \leq j_1, k \in \mathbb{Z}\}$.

L'approximation par ondelettes de Whittle de la log-vraisemblance négative est notée $\mathcal{L}(\cdot)$. Ce critère correspond à une approximation de la log-vraisemblance négative du vecteur gaussien des coefficients d'ondelettes $(W_{j,k}(\ell))_{j,k,\ell}$. HANNAN (1973) et FOX et TAQQU (1986) montrent que cette approximation est également satisfaisante pour des processus non gaussiens.

Dans notre contexte, le critère de Whittle pour les ondelettes est défini par,

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log \det (\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})) + \sum_{k=0}^{n_j} \mathbf{W}_{j,k}^T (\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d}))^{-1} \mathbf{W}_{j,k} \right], \quad (16)$$

où $\mathbf{\Lambda}_j(\mathbf{d})$ et la matrice $\mathbf{G}(\mathbf{d})$ sont obtenues avec la proposition 6,

$$\mathbf{\Lambda}_j(\mathbf{d}) = \text{diag} \left(2^{j\mathbf{d}} \right)$$

et l'élément (ℓ, m) de la matrice $\mathbf{G}(\mathbf{d})$ est donné par $G_{\ell,m}(\mathbf{d}) = \Omega_{\ell,m} K(d_\ell + d_m) \cos(\pi(d_\ell - d_m)/2)$.

Pour chaque $j \geq 0$, la quantité $\sum_k \mathbf{W}_{j,k}^T (\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d}))^{-1} \mathbf{W}_{j,k}$ est de dimension 1 et elle est égale à sa trace. Aussi,

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log \det (\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})) + \text{trace} \left((\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d}))^{-1} \mathbf{I}(j) \right) \right], \quad (17)$$

où $\mathbf{I}(j) = \sum_{k=0}^{n_j} \mathbf{W}_{j,k} \mathbf{W}_{j,k}^T$.

Cette expression est très similaire à celle obtenue par SHIMOTSU (2007). Ici, le périodogramme utilisant la décomposition de Fourier est remplacé par le scalogramme $\mathbf{I}(j)$ construit à l'aide des ondelettes.

L'objectif est maintenant de minimiser (17) par rapport aux deux paramètres \mathbf{G} et \mathbf{d} . Par dérivation de l'expression (17) par rapport à la matrice \mathbf{G} , nous obtenons,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{G}}(\mathbf{G}, \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} [n_j \mathbf{G}^{-1} - \mathbf{G}^{-1} \mathbf{\Lambda}_j(\mathbf{d})^{-1} \mathbf{I}(j) \mathbf{\Lambda}_j(\mathbf{d})^{-1} \mathbf{G}^{-1}].$$

Par suite, le minimum pour \mathbf{d} fixé est atteint pour \mathbf{G} ,

$$\hat{\mathbf{G}}(\mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \mathbf{\Lambda}_j(\mathbf{d})^{-1} \mathbf{I}(j) \mathbf{\Lambda}_j(\mathbf{d})^{-1}. \quad (18)$$

En remplaçant, $\mathbf{G}(\mathbf{d})$ par $\hat{\mathbf{G}}(\mathbf{d})$, nous définissons le critère en \mathbf{d} par,

$$\begin{aligned} R(\mathbf{d}) &:= \mathcal{L}(\hat{\mathbf{G}}(\mathbf{d}), \mathbf{d}) - 1 \\ &= \log \det(\hat{\mathbf{G}}(\mathbf{d})) + \frac{1}{n} \sum_{j=j_0}^{j_1} n_j \log(\det(\mathbf{\Lambda}_j(\mathbf{d}) \mathbf{\Lambda}_j(\mathbf{d}))), \\ &= \log \det(\hat{\mathbf{G}}(\mathbf{d})) + 2 \log(2) \left(\frac{1}{n} \sum_{j=j_0}^{j_1} j n_j \right) \left(\sum_{\ell=1}^p d_\ell \right). \end{aligned} \quad (19)$$

Le vecteur \mathbf{d} des paramètres de mémoire est ainsi estimé par $\hat{\mathbf{d}} = \underset{\mathbf{d}}{\operatorname{argmin}} R(\mathbf{d})$. Remarquons que si la matrice $\mathbf{\Omega}$ est égale à l'identité, l'estimateur $\hat{\mathbf{d}}$ est exactement égal à l'estimateur univarié, introduit dans (MOULINES et al. 2008), qui consiste à estimer le paramètre de mémoire sur chaque composante du vecteur des données sans prendre en compte l'aspect multivarié.

Dans un deuxième temps, l'estimateur $\hat{\mathbf{G}}(\hat{\mathbf{d}})$ de $\mathbf{G}(\mathbf{d})$ est obtenu par la formule (18). Finalement, l'estimateur de $\mathbf{\Omega}$ est obtenu en appliquant la correction du phase-shift,

$$\hat{\Omega}_{\ell,m} = \hat{G}_{\ell,m}(\hat{\mathbf{d}}) / (\cos(\pi(\hat{d}_\ell - \hat{d}_m)/2) K(\hat{d}_\ell + \hat{d}_m)). \quad (20)$$

L'équation (20) est définie correctement car la probabilité que $\hat{d}_\ell - \hat{d}_m$ soit exactement congruente à 1 modulo 2 est nulle. Par conséquent, l'estimateur $\hat{\mathbf{\Omega}}$ est défini presque sûrement. Pourtant, quand $d_\ell - d_m$ est proche de 1 modulo 2, l'estimation de $\mathbf{\Omega}$ peut être fortement biaisée.

3.1.2 Résultats de convergence

Les résultats suivant montrent la convergence des estimateurs de \mathbf{d} dans un premier temps. Comme dans (LOBATO 1999), l'estimateur proposé présente de bonnes propriétés de convergence. Dans un deuxième temps, nous montrons les propriétés de convergence pour $\mathbf{\Omega}$.

Une première condition établit la convergence asymptotique du scalogramme,

Condition (C)

$$\text{Pour tous } \ell, m = 1, \dots, p, \quad \sup_n \sup_{j \geq 0} \frac{1}{n_j 2^{2j(d_\ell + d_m)}} \mathbb{V}(I_{\ell,m}(j)) < \infty$$

Cette condition est vérifiée pour une large classe de processus, et n'est donc pas restrictive pour les résultats de convergence présentés dans la suite (voir (ACHARD et GANNAZ 2016, proposition 4)).

Théorème 7. *Sous les hypothèses (W1)-(W5) et la condition (C). Si j_0 et j_1 sont choisis de telle sorte que $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \rightarrow 0$ et $j_0 < j_1 \leq j_N$ avec $j_N = \max\{j, n_j \geq 1\}$, alors*

$$\hat{\mathbf{d}} - \mathbf{d}^0 = o_{\mathbb{P}}(1).$$

Ce résultat est une généralisation de celui obtenu dans le cas univarié par MOULINES et al. (2008). Dans notre contexte multivarié, nous faisons les mêmes hypothèses sur le filtre d'ondelettes et le choix de l'échelle j_0 .

La convergence de $\hat{\mathbf{G}}(\hat{\mathbf{d}})$ vers \mathbf{G}^0 est montrée dans le théorème suivant 8. Cependant ce résultat est établi pour des conditions un peu plus restrictives que dans le cas univarié. Ceci est dû à une simplification dans la preuve.

Théorème 8. *Supposons que (W1)-(W5) et la condition (C) sont vérifiées. Si j_0 et j_1 sont choisis tels que $\log(N)^2(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}) \rightarrow 0$ et $j_0 < j_1 \leq j_N$ alors*

$$\hat{\mathbf{d}} - \mathbf{d}^0 = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}),$$

$$\begin{aligned} \forall (\ell, m) \in \{1, \dots, p\}^2, \hat{G}_{\ell, m}(\hat{\mathbf{d}}) - G_{\ell, m}(\mathbf{d}^0) &= O_{\mathbb{P}}(\log(N)(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})), \\ \hat{\Omega}_{\ell, m} - \Omega_{\ell, m} &= O_{\mathbb{P}}(\log(N)(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})). \end{aligned}$$

Le taux de convergence optimal peut être atteint en combinant les deux termes de la borne obtenue dans le résultat précédent (GIRAITIS et al. 1997).

Corollary 9. *Supposons que (W1)-(W5) et la condition (C) sont vérifiées. En posant $2^{j_0} = N^{1/(1+2\beta)}$,*

$$\hat{\mathbf{d}} - \mathbf{d}^0 = O_{\mathbb{P}}(N^{-\beta/(1+2\beta)}).$$

3.2 Intérêt du modèle

3.2.1 Comparaison approche univariée et multivariée

Nous avons comparé numériquement les performances des estimateurs univariés et multivariés pour l'estimation de l'exposant longue mémoire. Les résultats des simulations, tables 1 et 2 montrent que lorsque la corrélation entre deux processus est élevée, les performances de l'estimateur multivarié sont meilleures. Il est alors conseillé d'appliquer l'estimateur multivarié dans toutes les situations où les observations sont multivariées. Si les observations s'avèrent indépendantes, les deux estimateurs sont égaux.

3.2.2 Convergence asymptotique de la corrélation

Sous l'hypthèse du modèle à longue mémoire précédent (5), on peut interpréter la propriété des corrélations par ondelettes comme convergeant asymptotiquement vers une constante qui est donnée par Ω (ACHARD et al. 2008). Dans le cas d'observations de taille finie, il est possible d'identifier l'échelle d'ondelettes à partir de laquelle la corrélation par ondelettes atteint l'asymptote et ainsi inférer les réseaux aux plus grandes échelles. Ceci peut en particulier permettre de comparer les résultats obtenus en IRMf avec un échantillonnage de l'ordre de la seconde et en MEG où l'échantillonnage est de l'ordre de la milliseconde. Ces travaux ont été initiés par les observations faites sur des données

TABLE 1 – Estimation multivariée par ondelettes de \mathbf{d} pour un processus bivarié $ARFIMA(0, \mathbf{d}, 0)$ avec $\rho = 0.4$, $N = 512$ avec 1000 répétitions.

d_1	\mathbf{d}	j_0	biais	écart-type	RMSE	rapport M/U	
0.2	0.2	1	-0.0298	0.0428	0.0522	0.9631	
	0.0		-0.0002	0.0438	0.0438	0.9504	
	0.2	1	-0.0330	0.0456	0.0563	0.9713	
	0.2		-0.0333	0.0443	0.0554	0.9831	
	0.2	1	-0.0304	0.0429	0.0526	0.9583	
	0.4		-0.0571	0.0461	0.0734	0.9701	
1.2	1.2	2	-0.0380	0.0830	0.0913	0.9728	
	0.8		-0.0298	0.0775	0.0831	0.9643	
	1.2	2	-0.0360	0.0818	0.0894	0.9702	
	1.0		-0.0346	0.0808	0.0879	0.9626	
	1.2	2	-0.0463	0.0853	0.0970	0.9677	
	1.2		-0.0393	0.0850	0.0936	0.9688	
	1.2	2	-0.0369	0.0799	0.0880	0.9589	
	1.4		-0.0482	0.0863	0.0989	0.9648	
	2.2	2.2	2	-0.0452	0.0950	0.1052	0.9642
		1.8		-0.0418	0.0935	0.1025	0.9657
2.2		2	-0.0432	0.0962	0.1054	0.9712	
2.0			-0.0459	0.0930	0.1037	0.9524	
2.2		2	-0.0516	0.0953	0.1084	0.9815	
2.2			-0.0513	0.0955	0.1084	0.9812	
2.2		2	-0.0489	0.0962	0.1079	0.9840	
2.4			-0.0573	0.0986	0.1141	0.9797	

TABLE 2 – Estimation multivariée par ondelettes de \mathbf{d} pour un processus bivarié $ARFIMA(0, \mathbf{d}, 0)$ avec $\rho = -0.8$, $N = 512$ avec 1000 répétitions.

d_1	\mathbf{d}	j_0	biais	écart-type	RMSE	rapport M/U
0.2	0.2	1	-0.0161	0.0380	0.0413	0.7625
	0.0		0.0129	0.0371	0.0393	0.8980
	0.2	1	-0.0334	0.0384	0.0509	0.8780
	0.2		-0.0331	0.0391	0.0512	0.8966
	0.2	1	-0.0164	0.0392	0.0425	0.7742
	0.4		-0.0439	0.0387	0.0585	0.7836
1.2	1.2	2	-0.0130	0.0773	0.0784	0.8360
	0.8		-0.0048	0.0703	0.0704	0.8248
	1.2	2	-0.0277	0.0744	0.0794	0.8614
	1.0		-0.0255	0.0699	0.0744	0.8363
	1.2	2	-0.0440	0.0730	0.0852	0.8498
	1.2		-0.0411	0.0714	0.0824	0.8475
	1.2	2	-0.0321	0.0700	0.0770	0.8394
	1.4		-0.0371	0.0732	0.0820	0.8538
2.2	2.2	2	-0.0267	0.0898	0.0937	0.8587
	1.8		-0.0282	0.0866	0.0911	0.8266
	2.2	2	-0.0372	0.0856	0.0933	0.8599
	2.0		-0.0366	0.0838	0.0915	0.8337
	2.2	2	-0.0507	0.0812	0.0957	0.8669
	2.2		-0.0529	0.0838	0.0991	0.8762
	2.2	2	-0.0419	0.0832	0.0931	0.8492
	2.4		-0.0430	0.0869	0.0969	0.8395

de neuro-imagerie et de finances (ACHARD et al. 2008). Les observations empiriques sont souvent limitées en temps soit par le coût de l'acquisition soit par le fait que les paramètres du modèle peuvent être considérés comme constants seulement sur un temps limité. Les tailles d'échantillon sont alors souvent limitées et les estimations pour des échelles grandes (i.e. basse fréquence) sont compliquées.

La figure 4 illustre ces observations sur un réseau obtenu à partir de données MEG.

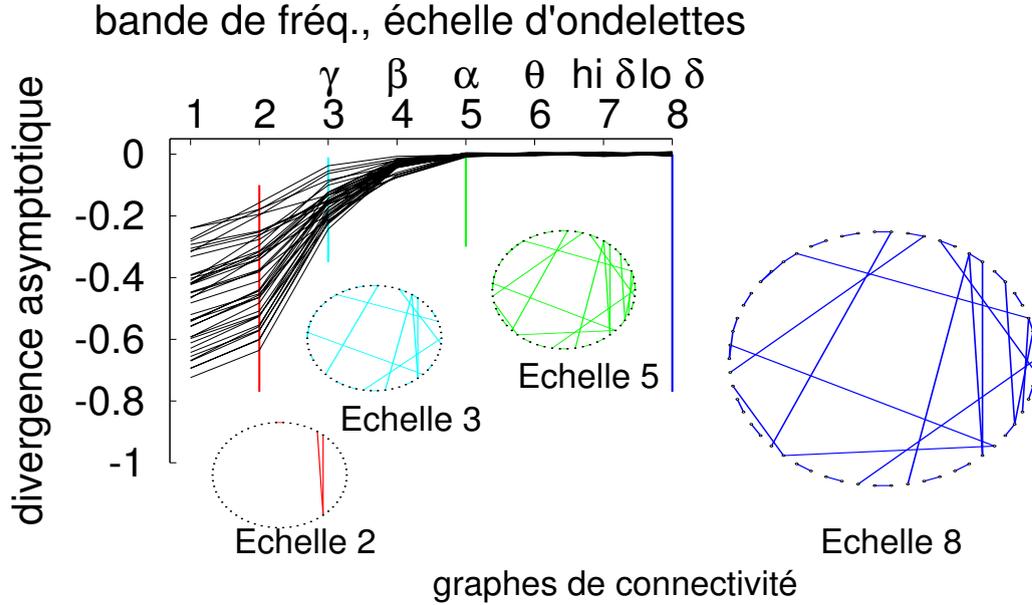


FIGURE 4 – Illustration de la convergence de la corrélation pour les basses fréquences ou grandes échelles d'ondelettes. Les réseaux obtenus aux échelles 5,6,7 ou 8 sont similaires. On peut alors inférer le réseau aux fréquences encore plus basses en faisant l'hypothèse que la corrélation ne sera pas modifiée d'après le modèle $M(d)$. Figure extraite de (ACHARD et al. 2008).

3.3 Perspectives

Du point de vue méthodologique, les prochaines étapes de l'étude de l'estimateur multivarié par ondelettes consistent à obtenir la convergence en loi des estimateurs (avec I. Gannaz). Ceci permettra en particulier de montrer théoriquement le bon comportement de l'estimateur multivarié par rapport à l'univarié lorsque la corrélation entre les processus augmente. Aussi, nous étudions d'autres choix de bases d'ondelettes afin de prendre en compte la différence de phase inhérente dans la définition du modèle multivarié (Avec M. Clausel, I. Gannaz et F. Roueff). Du point de vue application, nous avons un projet en cours avec l'Institut des Neurosciences de Grenoble sur l'étude de données d'IRMf acquises sur le rat anesthésié. Ces données ne souffrent pas du manque de points en temps et fournissent un cadre dans lequel la robustesse des méthodes peut être évaluée (avec E. Barbier, C. Delon-Martin, J. Warnking).

4 Estimation de structure de graphes

Dans le cadre de l'observation de séries multivariées, l'identification de la structure des interactions entre les composantes d'un système est un problème compliqué qui suscite de

nombreuses études actuellement.

Etant donnée X_1, X_2, \dots, X_p des variables aléatoires. Le graphe G sous jacent est défini par le couple (V, E) où $V = (V_1, V_2, \dots, V_p)$ sont les nœuds (ou sommets) du graphe correspondant aux variables aléatoires considérées et E est l'ensemble des arêtes, c'est-à-dire l'ensemble des couples de nœuds dont les variables aléatoires correspondantes présentent un couplage. La notion de couplage n'est pas évidente à définir et cette notion est discutée dans cette section.

Prenons par exemple un vecteur aléatoire gaussien de matrice de covariance Σ , figure 5. Le graphe associé peut-être défini en mettant une arête pour les couples de variables aléatoires qui ont une corrélation différente de 0 ou bien plus grande qu'une certaine valeur qui est choisie par l'utilisateur.

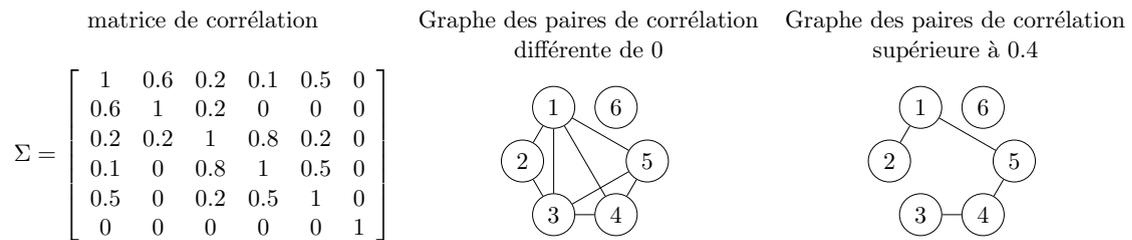


FIGURE 5 – Schéma de représentations d'un graphe à partir de la matrice de corrélation.

L'association entre corrélation non nulle et arête du graphe est complètement arbitraire et amènera du point de vue de l'application une interprétation particulière.

Nous abordons dans cette partie plusieurs choix de construction de la structure de graphe étant donné l'observation de variables aléatoires multivariées. Tout d'abord, le choix de la mesure de dépendance peut-être compliqué dans des cas de données réelles où nous n'avons aucun *a priori* sur la structure de graphe qui est recherchée, nous donnons un exemple simple de comparaison de la corrélation et corrélation partielle. Ensuite, nous établissons l'estimation de la structure de graphe en utilisant des tests d'hypothèses sur les corrélations obtenues par paires de variables aléatoires. Enfin, dans le cadre des modèles graphiques gaussiens, nous décrivons une approche gloutonne pour l'estimation de la structure du graphe. Pour conclure, nous évoquons deux autres approches. Dans une première approche, la structure de graphe est construite par la comparaison de deux états observés d'un même système, la présence d'arête dans le graphe modélise le fait que la mesure de dépendance varie significativement entre les deux états. Dans une deuxième approche, le but est d'identifier les ruptures dans des signaux multivariés, une arête dans le graphe est définie par la présence simultanée des ruptures entre deux signaux.

4.1 Corrélation ou corrélation partielle ?

La corrélation est une mesure très simple de dépendance entre variables aléatoires. Etudiée depuis de nombreuses années, de nombreux articles décrivent différentes manières d'estimer la corrélation et les variances des estimateurs sont connues (MUIRHEAD 1982). Pourtant cette mesure peut s'avérer trop simple dans différents contextes et fournir une estimation biaisée. C'est le cas par exemple quand deux variables sont liées à une troisième. Le fait de voir des parapluies dehors nous incitent à penser qu'il pleut et donc que le sol est mouillé. Pourtant le fait de voir des parapluies est complètement indépendant du fait que le sol est mouillé. La corrélation partielle va répondre à ce problème en calculant la corrélation entre les résidus de deux variables une fois régressées par rapport à une troisième variable. Cette notion s'étend aussi à des vecteurs de variables aléatoires. C'est

le cadre des modèles graphiques gaussiens où la non présence d'une arête dans le graphe est équivalente à une corrélation partielle nulle (WHITTAKER 1990).

Définition 7 (corrélation partielle). Soient X , Y et Z , vecteurs aléatoires. Notons $\epsilon_{X|Z}$ (resp. $\epsilon_{Y|Z}$) le résidu de la régression linéaire de X par Z (resp. Y par Z),

$$\begin{aligned}\epsilon_{X|Z} &= X - \beta_X^T Z, \\ \epsilon_{Y|Z} &= Y - \beta_Y^T Z,\end{aligned}$$

où β_X^T est défini par $\min_{\beta}(E[X - \beta^T Z]^2)$ et β_Y^T est défini par $\min_{\beta}(E[Y - \beta^T Z]^2)$.
la corrélation partielle entre X et Y étant donné Z est définie par

$$\rho_{XY|Z} = \text{Cor}(\epsilon_{X|Z}, \epsilon_{Y|Z}). \quad (21)$$

Pour des vecteurs gaussiens, la corrélation partielle peut être calculée à partir de l'inverse de la matrice de corrélation (WHITTAKER 1990),

Lemme 10. Soit $\mathbf{X} = (X_1, X_2, \dots, X_p)$, vecteur aléatoire gaussien, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Pour tous i, j , $1 \leq i \neq j \leq p$,

$$\rho_{X_i X_j | X_{V \setminus \{i, j\}}} = \frac{K_{ij}}{K_{ii} K_{jj}}$$

où $X_{V \setminus \{i, j\}} = \{X_\ell, 1 \leq \ell \leq p, \ell \neq i, \ell \neq j\}$ et K est la matrice de précision définie par l'inverse de la matrice de covariance Σ .

La question du choix de la mesure de dépendance est compliquée et dans certains cas mal posée de telle sorte que l'utilisation de la corrélation ou corrélation partielle permet par exemple d'identifier des propriétés différentes des graphes obtenus. Nous avons illustré ce phénomène dans (LEMOINE et al. 2009), en montrant des simulations de données simples qui mettent en lumière cette ambiguïté, voir figure 6.

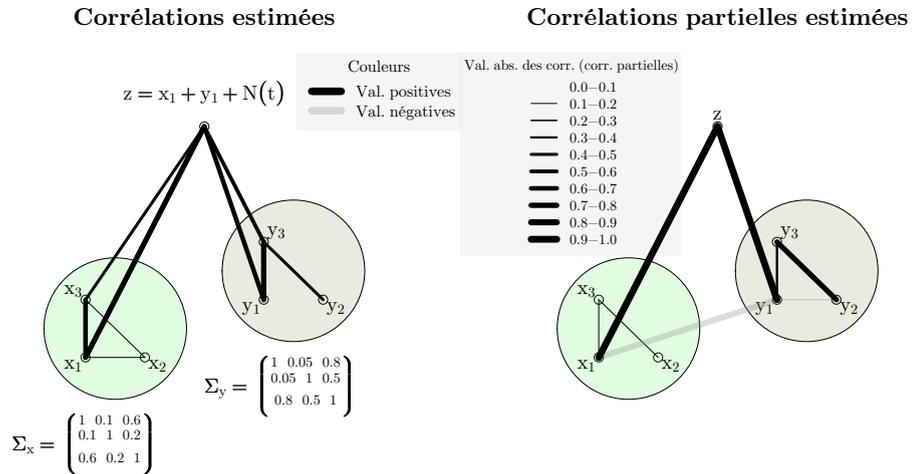


FIGURE 6 – Illustration de la notion de corrélation partielle (reproduite de (LEMOINE et al. 2009))

En effet, dans l'exemple de la figure 6, le graphe ne peut être construit à partir de la propriété de Markov par paire (i.e. deux variables non reliées par une arête sont indépendantes conditionnellement aux autres variables) (WHITTAKER 1990). Derrière ce problème se cache la notion de causalité et de moralisation du graphe qui a attiré beaucoup de réflexions du point de vue philosophique ou statistique (PEARL 2000). Le développement récent des réseaux bayésiens a permis de définir un cadre précis statistique pour l'étude de ces ambiguïtés, voir par exemple le livre de PEARL (2000). Cet exemple montre l'incertitude qu'il peut y avoir dans l'utilisation de la dépendance conditionnelle ou non. L'interprétation des résultats en pratique est donc conditionnée à l'utilisation d'une certaine définition de la structure de graphe associée aux données.

4.2 La corrélation comme outil pour extraire les graphes

Dans le cadre de l'extraction de structure de graphe, nous avons proposé de travailler avec la matrice de corrélation et d'associer une arête dans le graphe dès que la valeur de corrélation associée à une paire de nœuds est significativement différente de zéro.

Définition 8. Soient (X_1, X_2, \dots, X_p) un vecteur de variables aléatoires. Pour chaque couple de variables aléatoires (X_i, X_j) , on applique une décomposition en ondelettes, et l'estimateur de la corrélation à l'échelle k est définie par l'écriture classique de la corrélation empirique rappelée dans la formule (11).

Pour alléger les notations, le paramètre k ne sera pas mentionné dans la suite de l'exposé, et sera considéré fixé à une certaine valeur. L'analyse de la corrélation est donc menée pour une échelle d'ondelettes donnée. Le test d'hypothèse associé à ce couple de variables aléatoires se formule de la manière suivante,

$$\begin{aligned} H_{0,i,j} & : \rho_{i,j} = 0 \\ H_{1,i,j} & : \rho_{i,j} \neq 0 \end{aligned}$$

Après avoir appliqué la transformation de Fisher à la corrélation, on obtient le test suivant,

$$\begin{aligned} H_{0,i,j} & : z(\rho_{i,j}) = 0 \\ H_{1,i,j} & : z(\rho_{i,j}) \neq 0 \end{aligned}$$

Etant donnés les résultats mentionnés dans la section 2, propriété 5, pour un seuil α fixé pour chaque test d'hypothèse, il est possible d'en déduire la p -valeur correspondante et donc de rejeter ou non H_0 en contrôlant l'erreur de type I en particulier. Cependant, pour p variables, le nombre de tests d'hypothèses à établir est égal à $p(p-1)/2$. Le contrôle à un seuil α de chaque test d'hypothèse doit donc être relié au test global des $p(p-1)/2$ tests d'hypothèses. C'est le cadre de la théorie des tests multiples qui a connu un essor fulgurant ces dernières années avec le développement de méthodes d'acquisition toujours plus rapides et performantes qui donnent accès à un ensemble de données multivariées de grande dimension. On peut citer les données génomiques avec les micropuces d'ADN, les données de neuro-imagerie en EEG, MEG ou bien IRMf, les données en géosciences ou bien en astronomie (GOEMAN et SOLARI 2014).

Prenons un exemple où nous devons tester m hypothèses avec des tests indépendants. Il y a exactement m_0 tests pour lesquelles l'hypothèse nulle est vraie, notons alors \mathcal{H}_0 l'ensemble des indices i pour lesquels l'hypothèse nulle est vraie. Notons aussi \mathcal{R} l'ensemble des indices rejetés.

Alors, le *Family Wise Error Rate* est défini par le fait de contrôler la probabilité de rejeter au moins une vraie hypothèse, et dans les hypothèses d'indépendance des p -valeurs,

$$P(|\mathcal{R} \cap \mathcal{H}_0| \leq 1) = 1 - P(|\mathcal{R} \cap \mathcal{H}_0| = 0) = 1 - (1 - \alpha)^{m_0} \quad (22)$$

Pour $\alpha = 0.05$, et $m_0 = 20$, la probabilité de rejeter au moins une vraie hypothèse nulle est de 64%! Ce calcul montre que le contrôle de chacun des tests d'hypothèse sans correction ne permet pas le contrôle global. Dans notre cas particulier des corrélations, on peut citer l'exemple décrit dans HERO et RAJARATNAMA (2011) où les auteurs simulent des variables indépendantes, et font apparaître des valeurs de corrélations élevées, figure 7.

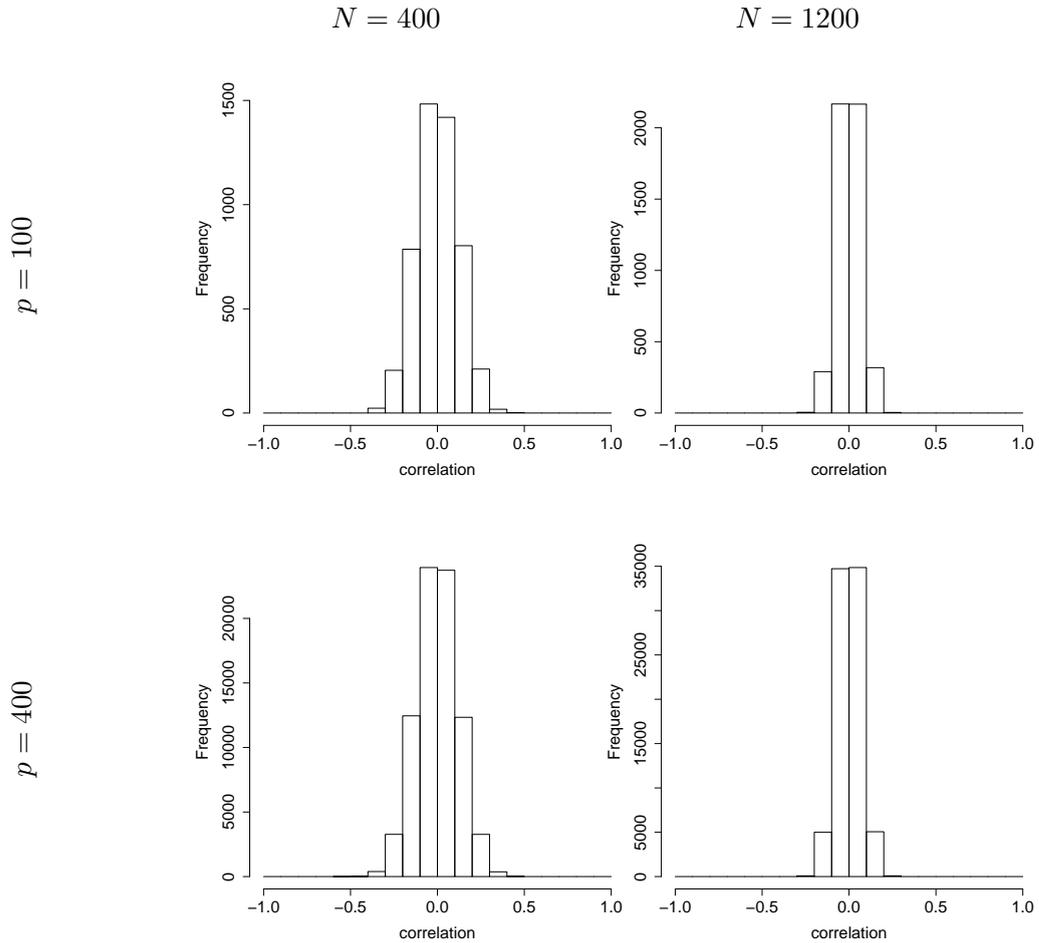


FIGURE 7 – Exemple de valeurs de corrélations par ondelettes obtenues pour des variables gaussiennes simulées indépendantes avec différents paramètres choisis pour le nombre de variables p et le nombre d'échantillons en temps N . L'échelle d'ondelettes 3 a été retenue pour calculer les corrélations. On peut voir que pour $p = 100$ et $N = 400$, les corrélations estimées peuvent être plus grandes que 0.5 même avec des variables indépendantes.

Pourtant, il est quand même possible de proposer une procédure qui permet de contrôler la probabilité de rejeter au moins une vraie hypothèse, c'est ce qu'on appelle la procédure de Bonferroni (voir par exemple GOEMAN et SOLARI (2014)). Nous noterons dans la suite m le nombre d'hypothèses à tester. Dans le cadre de l'identification de la structure de graphe, $m = p(p - 1)/2$

Proposition 11 (procédure Bonferroni). *Si le contrôle de chaque test d'hypothèse est ajusté en fonction du nombre total de tests d'hypothèses effectués, c'est-à-dire*

$$\mathcal{R} = \{i, 1 \leq i \leq m, p_i \leq q_{bonf} = \alpha/m\}$$

alors,

$$FWER = P(|\mathcal{R} \cap \mathcal{H}_0| \leq 1) \leq \alpha \quad (23)$$

La démonstration de cette proposition utilise l'inégalité de Boole, et elle est vraie même dans le cas de tests dépendants.

$$P(|\mathcal{R} \cap \mathcal{H}_0| \leq 1) = P\left(\bigcup_{i \in \mathcal{H}_0} (p_i \leq q_{bonf})\right) \leq \sum_{i \in \mathcal{H}_0} P(p_i \leq q_{bonf}) \leq \sum_{i \in \mathcal{H}_0} \alpha = M_0 q_{bonf} \leq \frac{m_0}{m} \alpha \leq \alpha \quad (24)$$

Cette procédure permet donc le contrôle global du *Family Wise Error Rate*, mais lorsque le nombre de tests à effectuer est grand, cette procédure s'avère vite très conservatrice et très peu de tests vont rejeter H_0 .

Dans notre cas, il est évident que les tests sont dépendants ainsi que les p -valeurs. HERO et RAJARATNAMA (2011) ont proposé le contrôle FWER pour des tests multiples de corrélations en utilisant une procédure ad-hoc avec la transformation des estimations de la corrélation en U-statistique.

Une alternative au FWER est le contrôle du taux de fausse détection appelé FDR (*False Discovery Rate*). BENJAMINI et HOCHBERG (1995) ont proposé une procédure simple à mettre en place qui assure ce contrôle.

Définition 9 (FDR). Le FDR est défini par,

$$FDR = E \left[\frac{|\mathcal{R} \cap \mathcal{H}_0|}{|\mathcal{R}| \vee 1} \right] \quad (25)$$

La procédure de (BENJAMINI et HOCHBERG 1995) consiste à trier les p -valeurs et à sélectionner les p -valeurs plus petites qu'un certain seuil qui s'ajuste en rejetant des hypothèses nulles.

Définition 10 (Procédure Benjamini et Hochberg). Soient $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$, les p -valeurs ordonnées. Le seuil de rejet des hypothèses nulles est défini par,

$$\hat{k} = \max\{0 \leq k \leq m, p_{(k)} \leq \frac{k}{m} q\} \quad (26)$$

où $p_{(0)} = 0$.

Et alors

$$\mathcal{R}^{BH} = \{0 \leq k \leq m, p_{(k)} \leq \frac{\hat{k}}{m} q\}$$

Cette procédure permet le contrôle du FDR pour des tests univariés au seuil q sous l'hypothèse d'indépendance des p -valeurs (voir l'article de revue (ROQUAIN 2010)). BENJAMINI et YEKUTIELI (2001) ont étendu le contrôle à des tests dépendants avec une dépendance appelé PRDS (positive regression dependence on the subset). Dans le cas de dépendance quelconque, le contrôle est obtenu en remplaçant $\frac{k}{m} q$ par $\frac{k}{\sum_{k=1}^m 1/k} q$. Dans le cas de l'étude de corrélation, la dépendance entre les p -valeurs est compliquée à caractériser. DRTON et

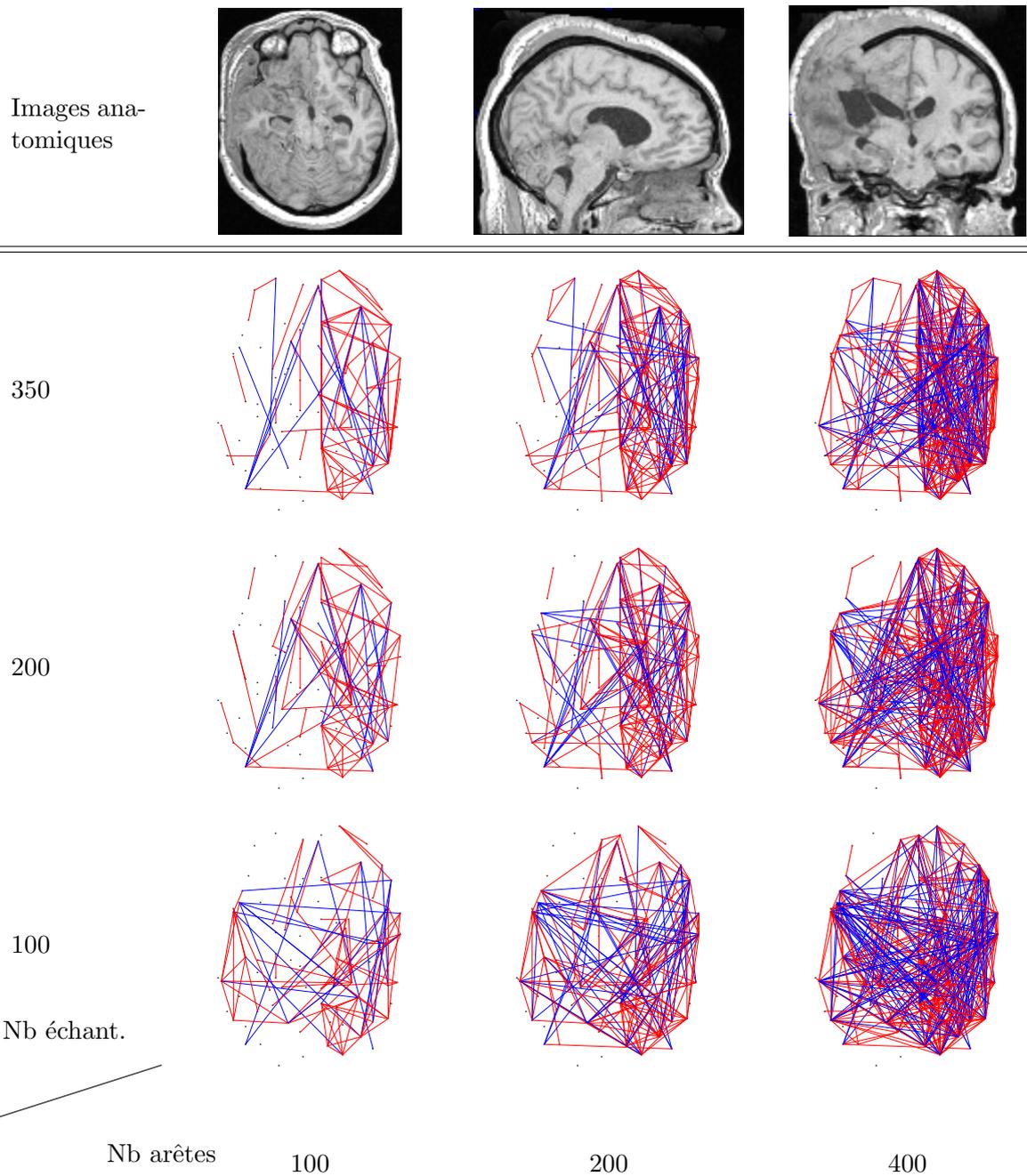


FIGURE 8 – Positionnement géométrique des 100, 200 ou 400 arêtes (abscisse) présentant les plus fortes corrélations pour 100, 200 ou 350 points en temps (ordonnée). Les graphes sont représentés par une vue de dessus avec l'hémisphère gauche à gauche de l'image et l'hémisphère droit à droite de l'image. La patiente scannée présente une anomalie sur l'hémisphère gauche, on s'attend donc à ce que les corrélations soient altérées pour l'hémisphère gauche. C'est bien le cas sur les représentations quand le nombre d'arêtes est inférieur aux nombres de points en temps. Ces résultats dépendent aussi du nombre de nœuds du graphe, ici ce nombre est égal à 90. Cette figure illustre la nécessité de prendre en compte la correction multiple pour éviter la sélection d'arêtes non significatives.

PERLMAN (2004) proposent un contrôle du FWER pour l'identification de la structure du graphe à partir de corrélation (voir par exemple (MESKALDJI et al. 2011) dans le cadre de réseaux de connectivité cérébrale).

Comme nous l'avons mentionné dans (FALLANI et al. 2014), les valeurs de corrélation significativement différentes de zéro dépendent du nombre de points de la série temporelle et du nombre de variables à prendre en compte.

La figure 8 illustre la sélection des corrélations avec des données réelles d'IRMf de repos et montre la nécessité de corriger les p -valeurs quand on est face à un problème de tests multiples. La patiente a été scannée alors qu'elle était dans le coma, sur l'image anatomique, il est clair que l'hémisphère gauche est gravement atteint et d'après un examen clinique approfondi, il s'avère que l'hémisphère gauche ne fonctionne plus normalement. Il est donc attendu que le graphe des connexions de cette patiente présente une dissymétrie droite-gauche avec plus de connexions à droite qu'à gauche. La figure illustre le fait que si on ne prend pas en compte la correction pour tests multiples, les corrélations seuillées peuvent être complètement erronées. La méthode retenue dans ce travail consiste à tout d'abord sélectionner les paires de corrélations différentes de zéros en contrôlant le FWER, notons F cet ensemble.

$$F = \{(i, j), H_{0,i,j} \text{ est rejetée}\}$$

Ensuite, dans le but de comparer les réseaux de connectivité entre différents sujets, les L paires de plus fortes corrélations, $L \subseteq F$, sont retenues pour définir le graphe. Le contrôle du FWER n'est pas très satisfaisant car la puissance du test peut être alors très faible avec peu de tests rejetés. Il serait alors préférable de pouvoir contrôler le FDR, c'est le travail de Marine Roux dans le cadre de sa thèse.

4.3 Méthode gloutonne avec la corrélation partielle

Dans le cadre de la thèse d'Aude Costard, nous avons proposé une approche gloutonne par vraisemblance bayésienne asymptotique gaussienne pour l'estimation de la structure de graphe basée sur la corrélation partielle.

Définition 11 (modèles graphiques gaussiens). Soit \mathbf{X} un vecteur aléatoire gaussien dans \mathbb{R}^p . \mathbf{X} satisfait un modèle graphique gaussien attaché au graphe $G = (V, E)$, si \mathbf{X} suit une loi gaussienne multivariée de moyenne nulle qui vérifie la propriété de Markov par paire, i.e. étant donnée Σ une matrice $p \times p$, définie positive,

$$\begin{aligned} X &\sim \mathcal{N}(0, \Sigma) \\ (\Sigma^{-1})_{ij} &= 0, \forall (i, j) \notin E \end{aligned}$$

Pour plus de détails, les livres (WHITTAKER 1990; LAURITZEN 1996; EDWARDS 2012) donnent une vue d'ensemble sur les modèles graphiques gaussiens. L'estimation de G est équivalente à l'estimation des zéros de la matrice de précision, $K := \Sigma^{-1}$. La vraisemblance de K est donnée par,

$$P(X|\Sigma_G, G) = \left(\frac{\det K}{(2\pi)^p}\right)^{n/2} \exp\left(-\frac{n}{2}\text{trace}(KS)\right), \quad (27)$$

où S est l'estimateur de la matrice de covariance empirique, $S = \mathbf{X}\mathbf{X}^T/N$.

Plusieurs approches ont alors été développées dans le but d'estimer la matrice de précision K et surtout les zéros de celle-ci. *gLasso* (graphical lasso) (FRIEDMAN et al. 2008)

consiste à minimiser la vraisemblance en pénalisant par la norme ℓ_1 sur K . GIRAUD et al. (2012) ont étudié l'estimation de K par régression avec un terme de pénalisation. Enfin, des approches bayésiennes ont été développées (ROVERATO et WHITTAKER 1998 ; GIUDICI et GREEN 1999 ; DONNET et MARIN 2012) supposant le plus souvent le graphe décomposable pour faciliter l'expression de la densité des observations.

Dans la thèse d'Aude Costard, nous développons une approche qui ne suppose pas la décomposabilité du graphe. A la différence de l'approche par test, cette approche permet d'attacher à chaque configuration de graphe une probabilité.

Soit \mathbf{X} un vecteur aléatoire gaussien, associé à la la structure de graphe $G = (V, E)$. Nous nous sommes basés sur la propriété suivante qui permet d'écrire la vraisemblance de G étant donné \mathbf{X} (MARRELEC et BENALI 2006) :

$$P(G|\mathbf{X}) = C(\mathbf{X}) \times \frac{\varphi_{\mathbf{p}_{\bar{E}}, W_{\bar{E}\bar{E}}}(\mathbf{0})}{V(G)} \quad (28)$$

où $C(\mathbf{X})$ est une constante dépendant seulement des données \mathbf{X} , W est la matrice d'Is-
serlis (ROVERATO et WHITTAKER 1998) de la corrélation empirique S . $\varphi_{\mathbf{p}_{\bar{E}}, W_{\bar{E}\bar{E}}}$ est la
gaussienne de moyenne \mathbf{p} et de covariance $W_{\bar{E}\bar{E}}$ avec \mathbf{p} le vecteur des corrélations par-
tielles, \bar{E} et $\bar{E}\bar{E}$ signifient que l'on considère seulement les arêtes non présentes dans G .
 $V(G)$ représente le volume des matrices définies positives avec des zéros pour les paires de
nœuds qui ne sont pas reliés dans G . Le calcul de $V(G)$ est très couteux même pour des
valeurs de p petites, de l'ordre de 10 par exemple.

La difficulté de ce genre d'approche est de connaître la normalisation $C(\mathbf{X})$ qui permet
de calculer précisément la probabilité a posteriori $P(G|\mathbf{X})$. En effet, ici la seule possibilité
est de calculer les probabilités pour toutes les configurations possibles de graphes et de
normaliser. Or pour un graphe à 6 nœuds, il existe 2^{15} configurations, et pour 10 nœuds,
 2^{45} . Cette approche est donc complètement impraticable pour des graphes de plus de 6
nœuds.

Cependant pour des graphes plus petits, le calcul de $P(G|\mathbf{X})$ permet d'illustrer la
variance des résultats pour un faible nombre de données ou pour une faible valeur de
corrélations partielles, figures 9 et 10.

Dans (COSTARD et al. 2013), nous avons proposé une stratégie de recherche de la so-
lution qui minimise la vraisemblance (28) sans avoir à parcourir l'ensemble des structures
de graphes possibles, cette méthode est appelée *ABiGlasso*. Celle-ci commence par l'ini-
tialisation de la recherche pour la solution donnée par le graphical lasso FRIEDMAN et al.
(2008) puis consiste à étendre l'espace de recherche autour de ce point en ajoutant ou
enlevant des arêtes de la structure initiale.

Dans la thèse d'Aude Costard, nous avons montré que cette méthode fournit des résul-
tats similaires aux méthodes utilisées dans l'état de l'art. Dans la figure 12, nous présentons
les résultats d'identification de la structure de graphe pour différents types de graphes pré-
sentés dans la figure 11 et différentes tailles d'échantillons. La figure 12 illustre la difficulté
des méthodes testées d'identifier la vraie structure quand un faible nombre d'échantillons
est disponible. Les comparaisons de graphes sont quantifiées grâce à la distance de Ham-
ming, qui correspond aux nombres d'arêtes qui diffèrent entre les deux graphes. Les ré-
sultats de la méthode proposée sont comparés avec trois autres approches. La méthode
SIN consiste à appliquer des tests sur les valeurs de corrélations partielles (DRTON et
PERLMAN 2008). Une approche par pénalisation est appelée *BAGlasso* (WANG 2012) avec

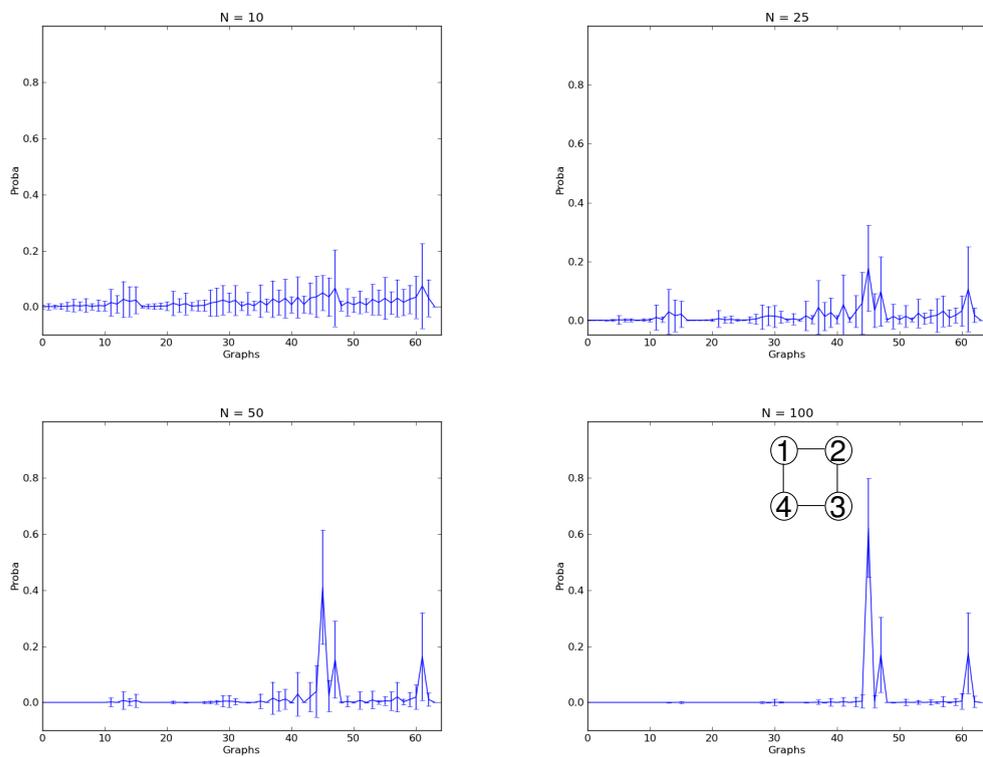


FIGURE 9 – Illustration de la taille des observations N sur le calcul de $P(G|\mathbf{X})$. En abscisse sont représentés les 64 graphes possibles pour un graphe à 4 nœuds. $P(G|\mathbf{X})$ est tracée en ordonnée. Chaque simulation correspond à \mathbf{X} , un vecteur aléatoire gaussien de moyenne nulle et de matrice de précision $K = \begin{pmatrix} 1 & 0.4 & 0 & 0.4 \\ 0.4 & 1 & 0.4 & 0 \\ 0 & 0.4 & 1 & 0.4 \\ 0.4 & 0 & 0.4 & 1 \end{pmatrix}$. Le graphe exact choisi pour la simulation est tracé pour $N = 100$ correspondant au maximum de $P(G|\mathbf{X})$.

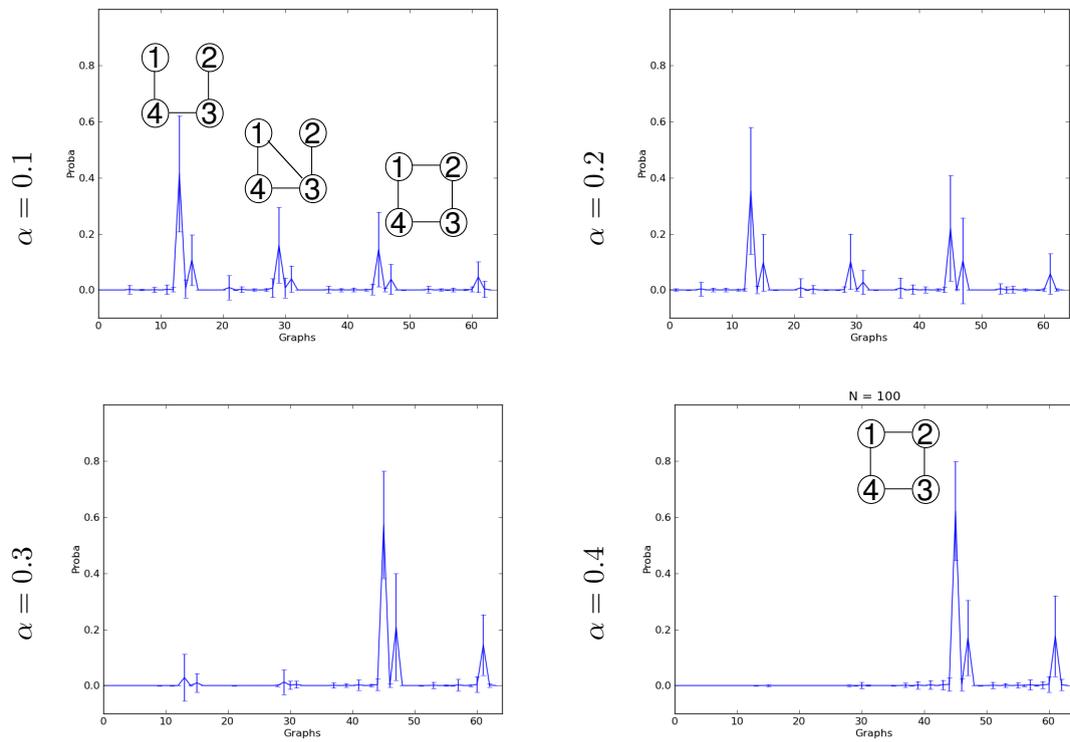


FIGURE 10 – Illustration de la variation du coefficient d’une arête sur le calcul de $P(G|\mathbf{X})$. En abscisse sont représentés les 64 graphes possibles pour un graphe à 4 nœuds. $P(G|\mathbf{X})$ est tracée en ordonnée. Chaque simulation correspond à \mathbf{X} , un vecteur aléatoire gaussien de moyenne nulle et de matrice de précision $K = \begin{pmatrix} 1 & \alpha & 0 & 0.4 \\ \alpha & 1 & 0.4 & 0 \\ 0 & 0.4 & 1 & 0.4 \\ 0.4 & 0 & 0.4 & 1 \end{pmatrix}$. Le graphe exact choisi pour la simulation est tracé pour $\alpha = 0.4$ correspondant maximum de $P(G|\mathbf{X})$. Pour $\alpha = 0.1$ les trois graphes les plus probables sont tracés explicitement.

une procédure bayésienne pour le choix du paramètre de pénalisation. Enfin, la méthode *GGMselect* (GIRAUD et al. 2012) propose une estimation par régression et pénalisation.

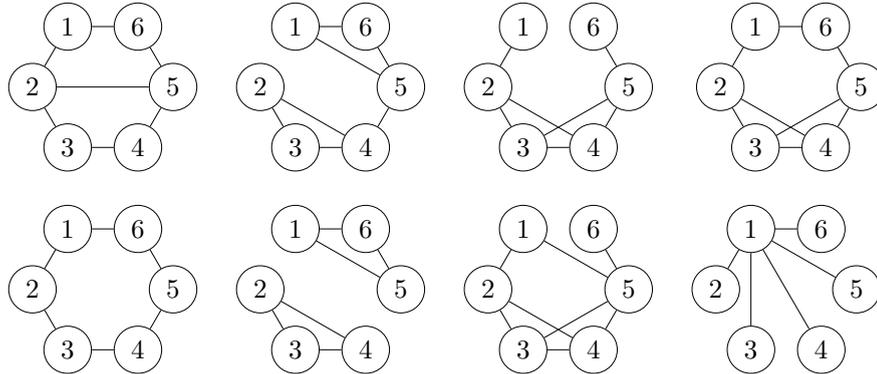


FIGURE 11 – Structures de graphes choisies pour les comparaisons des performances des méthodes

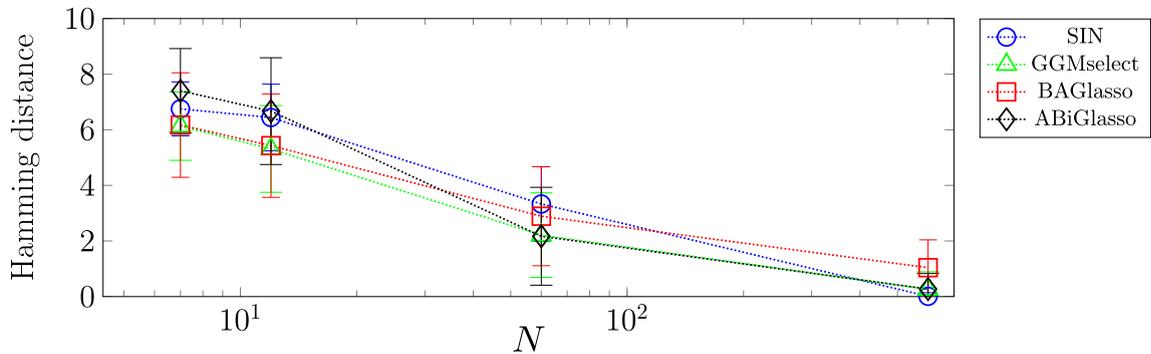


FIGURE 12 – Moyenne et écart-type de la distance de Hamming calculée entre la structure de graphe estimée et la vraie pour différentes tailles d'échantillons n . Les résultats sont similaires pour les 4 méthodes testées. Pour $N = 7$ et $N = 12$ les graphes estimés sont très loin de la vraie structure. Puis les résultats deviennent vraiment acceptables pour $n = 60$. *ABiGlasso* (COSTARD et al. 2013); *SIN* (DRTON et PERLMAN 2008); *BAGlasso* (WANG 2012); *GGMselect* (GIRAUD et al. 2012).

4.4 Graphes différentiels

Lors de l'étude de signaux d'EEG, nous avons proposé d'utiliser des mesures basées sur la fonction d'inter-corrélation ρ définie pour deux séries stationnaires X_t et Y_t par

$$\rho(\tau) = \text{Cor}(X_t, Y_{t+\tau}) \quad (29)$$

On extrait alors le maximum de la fonction d'inter-corrélation. Cette mesure a été mise au point dans le contexte de la détection de foyers épileptiques. Pour cela, nous avons à notre disposition pour chaque patient un jeu de données qui présente des intervalles de temps associés à une décharge épileptique et d'autres associés à un signal sans décharge épileptique. Ces étiquettes sont placées avant tout traitement par les neurologues qui suivent ces patients. Enfin, on applique des méthodes de tests par permutations pour

extraire les paires d'électrodes où la corrélation a été significativement différente entre les deux conditions, voir figure 13.

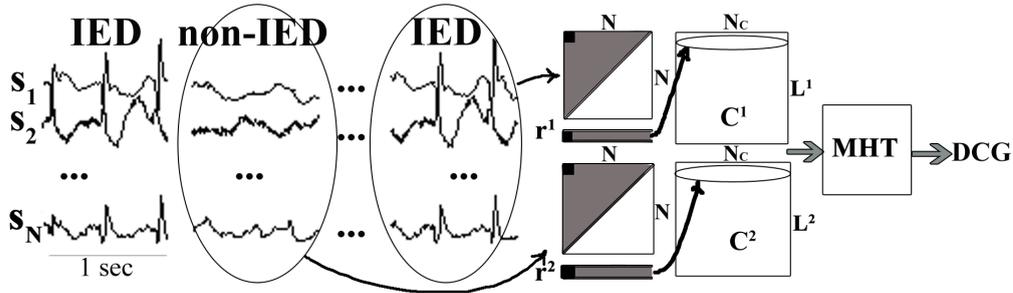


FIGURE 13 – Description de la méthode de construction des graphes différentiels pour des enregistrements électro-encéphalogrammes de patients épileptiques

Les graphes ainsi construits permettent de mettre en évidence les réseaux de connexions qui se construisent lors de décharges épileptiques. Partant de l’hypothèse que lors d’une décharge épileptique, les neurones ont tendance à se synchroniser fortement dans la région du foyer épileptique, les graphes de connexions résultant de la décharge nous permettent d’identifier les “hubs” correspondant vraisemblablement aux foyers épileptiques (AMINI et al. 2010; AMINI et al. 2011). Cette représentation est une première étape vers les réseaux dynamiques (LEONARDI et al. 2013).

4.5 Détection de rupture pour des signaux multivariés

Comme illustré dans la figure 14, l’objectif est de détecter des changements de rupture dans un signal en observant des données multivariées. Chaque nœud du graphe est un signal. Les arêtes du graphe sont construites quand il y a la même rupture dans les signaux. La présence d’une arête dans le graphe reflète donc la présence d’une rupture simultanée entre les deux signaux correspondants (voir figure 15). Nous avons développé un nouvel estimateur appelé Bernoulli Detector qui permet d’évaluer des ruptures dans les signaux dans un cadre multivarié.

La méthode se base sur un modèle bayésien non-paramétrique utilisant les p -valeurs obtenues par test de Wilcoxon. On montre le contrôle du FDR dans le cadre d’une seule rupture. Notre approche apporte une souplesse par rapport aux méthodes classiquement développées en permettant de faire des groupes de nœuds et de ne pas forcer une segmentation unique pour toutes les observations (HARLÉ et al. 2014).

4.6 Perspectives

L’estimation précise de la structure de graphe est compliquée sur les données réelles en IRMf, en particulier à cause du faible nombre d’observations en temps en comparaison de la complexité spatiale du cerveau. Nous nous concentrons maintenant sur l’approche par tests d’hypothèses sur la matrice de corrélation ou corrélation partielle en proposant un contrôle du *false discovery rate* avec des hypothèses sur la parcimonie du graphe. C’est le travail actuel effectué dans le cadre de la thèse de Marine Roux.

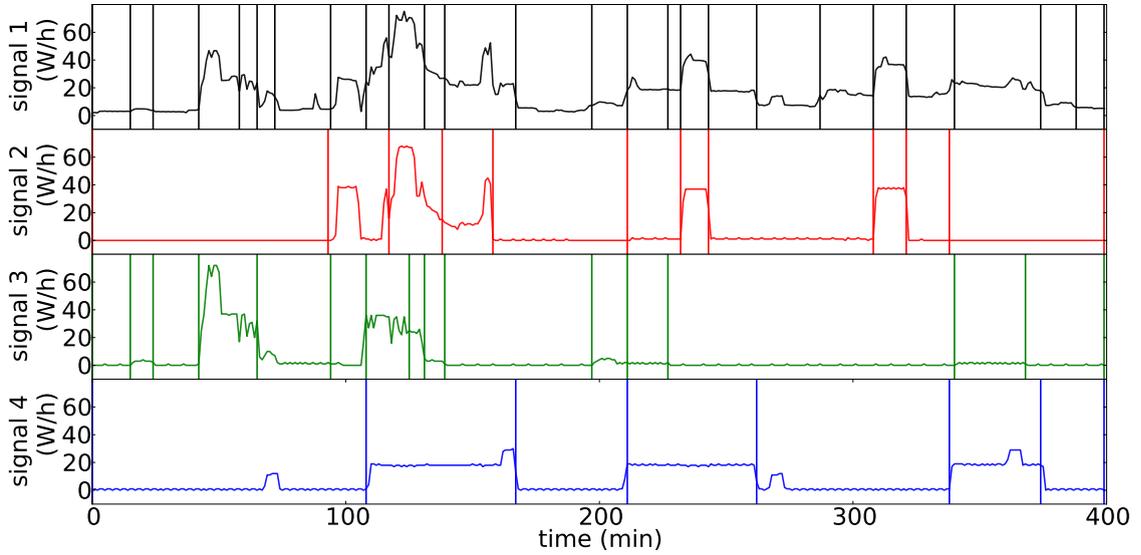


FIGURE 14 – Segmentation de données de consommation électrique dans une maison. Les ruptures sont représentées par des lignes verticales. Le signal 1 représente le compteur général et les autres signaux sont pris dans différentes parties de la maison. Les ruptures trouvées dans les signaux 2,3 et 4 ont été répercutées sur le signal 1. Mais le signal 1 présente des ruptures qui ne sont pas dans les autres signaux. Figure extraite de (HARLÉ et al. 2014).

5 Méthodes statistiques de comparaisons des graphes

La représentation sous forme de graphes des données de neuro-imagerie permet tout d’abord une visualisation adaptée pour favoriser les échanges avec les médecins. La figure 16 présente la visualisation graphique du réseau de connectivité pour un sujet. Cependant, pour procéder à des comparaisons statistiques de groupes, la visualisation seule du graphe ne suffit pas et il est crucial de pouvoir extraire des paramètres quantitatifs de ces graphes. Je me suis donc attachée à promouvoir l’utilisation de métriques topologiques qui permettent d’extraire des caractéristiques des graphes et donc de les comparer (ACHARD et al. 2006 ; ACHARD et BULLMORE 2007). L’enjeu ici est non seulement d’extraire des caractéristiques différenciant les groupes de sujets, mais aussi de pouvoir donner une interprétation neuroscientifique aux différences détectées et de permettre une évaluation pour chaque patient afin de confronter ces données à un diagnostic clinique.

Je présente dans une première partie, les métriques de graphes bien connues et développées par ailleurs que nous avons utilisées. Puis lors de l’étude de patients dans le coma, après avoir fait le constat que les métriques existantes étaient insuffisantes, nous avons proposé un nouvel index de comparaison des graphes qui nous a permis de mettre en évidence une réorganisation cérébrale chez ces patients (ACHARD et al. 2012). Enfin, dans une dernière partie, nous proposons une solution originale qui permet de comparer des graphes à l’aide de formulations bayésiennes. Je poursuis actuellement cette thématique avec l’encadrement de Maite Termenón en co-direction entre GIPSA-lab et GIN sur l’étude de la robustesse des graphes de connectivités cérébrales mesurés par IRMf.

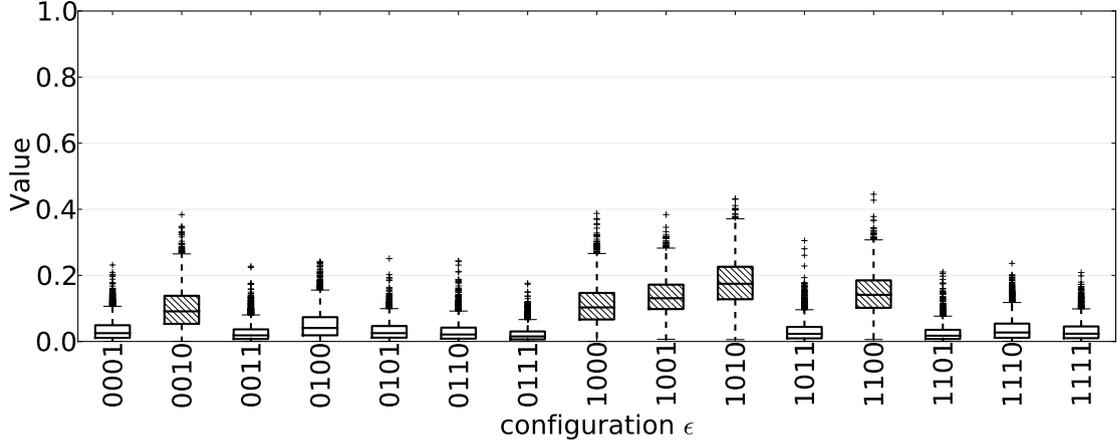


FIGURE 15 – La distribution a posteriori P_ϵ des configurations possibles pour les données de consommation électrique dans une maison. Les configurations qui sont hachurées sont celles qui sont les plus significatives. Pour chaque configuration, on représente la présence du signal avec 1 et sa non présence avec 0. P_{1001} correspond à la configuration du graphe où le signal 1 et le signal 4 sont reliés par une arête. Chaque configuration compte donc quatre chiffres pour les 4 signaux étudiés. De manière prépondérante, l’algorithme sélectionne les configurations contenant le signal 1. Figure extraite de (HARLÉ et al. 2014).

5.1 Quels outils pour comparer des graphes ?

Une fois les graphes obtenus, il est intéressant de savoir comment les comparer. Plusieurs stratégies sont possibles (voir figure 17) : utilisation directe des matrices de corrélation ou des matrices d’adjacence (obtenues en mettant un 1 quand il y a une arête), ou bien extraction de métriques topologiques (voir (RICHIARDI et al. 2013) pour une revue des différentes possibilités). En neurosciences beaucoup de chercheurs se sont focalisés sur ce problème (BULLMORE et SPORNS 2009 ; RUBINOV et SPORNS 2010). Trois outils sont couramment utilisés : le degré qui compte le nombre d’arêtes par nœud du graphe, le plus court chemin, ou encore le clustering. Nous avons appliqué ces résultats par exemple dans la problématique de l’effet de l’âge sur la connectivité cérébrale au repos (ACHARD et BULLMORE 2007).

Pour un graphe $G = (V, E)$ (V est l’ensemble des nœuds et E est l’ensemble des arêtes), nous pouvons calculer des paramètres quantitatifs associés à chaque nœud du graphe ou bien des paramètres moyens sur tout le graphe en prenant la moyenne sur tous les nœuds. Je me focalise ici sur trois métriques, il en existe beaucoup d’autres qui peuvent aussi être utilisées de manière similaire (RUBINOV et SPORNS 2010).

Le degré d’un nœud i correspond au nombre de connexions entre ce nœud et tous les autres du graphes.

$$D_i = \sum_{j \in G} A_{ij}.$$

où A est la matrice d’adjacence du graphe, $A_{ij} = 0$ quand $(i, j) \notin E$, et $A_{ij} = 1$ quand $(i, j) \in E$. Les mesures d’efficacité développées par Latora et ses collaborateurs (LATORA et MARCHIORI 2001) s’adaptent à une quantification de chaque nœud du graphe. Pour

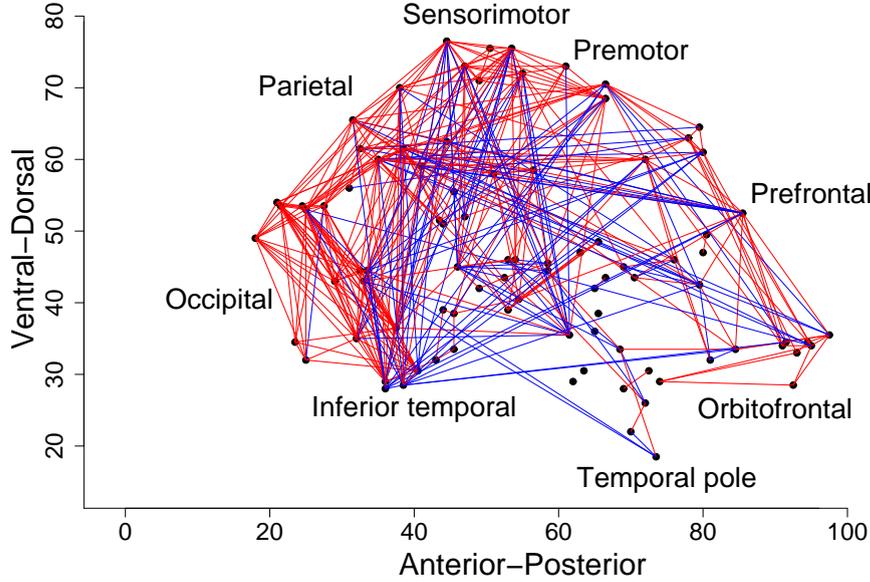


FIGURE 16 – Exemple de représentation d’un réseau du cerveau obtenu à partir d’images d’IRMf. Les arêtes rouges représentent les connexions courtes distances (la distance euclidienne est inférieure à 7.5cm). Les arêtes bleues représentent les connexions longues distances (la distance euclidienne est supérieure à 7.5cm). Figure extraite de (ACHARD et al. 2006).

un graphe $G = (V, E)$, nous caractérisons le rôle de “hub” d’un nœud i par son “efficacité régionale”, définie de la manière suivante :

$$E_{glob_i} = \frac{1}{p-1} \sum_{j \in G} \frac{1}{L_{i,j}} \quad (30)$$

où p est le nombre de nœuds total du graphe G , et $L_{i,j}$ est le plus court chemin entre le nœud i et le nœud j . L’efficacité globale correspond à la moyenne sur tous les nœuds du graphe des efficacités régionales. La figure 18 présente un exemple de deux graphes ayant le même nombre de sommets et d’arêtes, mais une efficacité globale très différente.

Pour chaque nœud, nous caractérisons la densité de connexions entre ses voisins directs par son “efficacité locale”, définie de la manière suivante, et souvent dénotée clustering :

$$Clust_i = \frac{1}{(p_{G_i}-1)p_{G_i}} \sum_{j,k \in G_i} \frac{1}{L_{j,k}} \quad (31)$$

où G_i est le sous-graphe formé par les voisins du nœud i , en enlevant le nœud i et p_{G_i} est le nombre de nœuds du sous graphe G_i . La figure 19 présente un exemple de deux graphes ayant le même nombre de nœuds et d’arêtes, mais une efficacité locale très différente.

Nous avons utilisé ces mesures pour caractériser les régions cérébrales et identifier les régions ayant le plus grand nombre de connexions, dites “hubs” ou régions pivots. Ces résultats sont présentés dans plusieurs publications (ACHARD et al. 2006 ; BASSETT et al. 2006 ; ACHARD et BULLMORE 2007 ; MEUNIER et al. 2009). Récemment, nous avons comparé ces résultats avec des méthodes à base de support vecteur machine directement sur les corrélations de la matrice, et nous avons montré sur un jeu de données réelles que

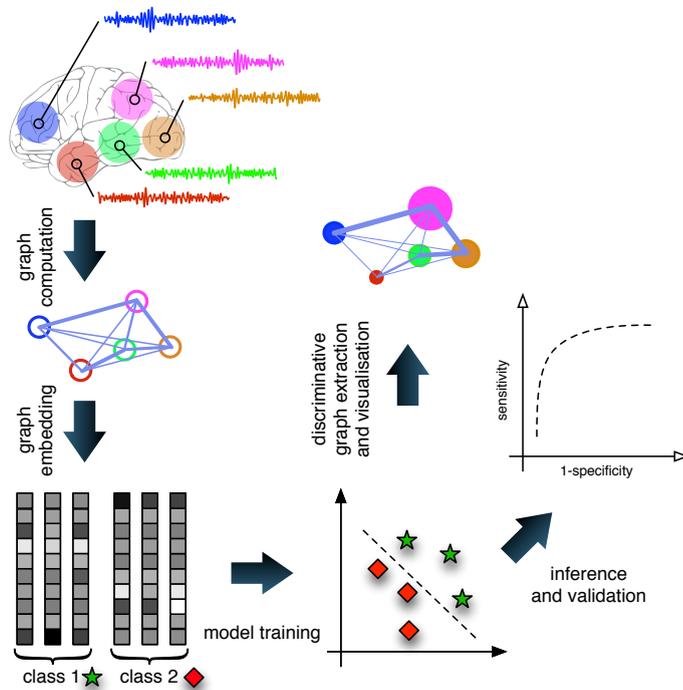


FIGURE 17 – Schéma de comparaison de données de graphes de connectivité cérébrale, figure extraite de (RICHIARDI et al. 2013).

l'extraction des métriques de graphes permet d'avoir une plus grande précision dans la discrimination des deux groupes (RICHIARDI et al. 2011 ; RICHIARDI et al. 2013).

5.2 Introduction d'un nouvel index

Lors de récentes recherches sur les possibles modifications des réseaux de connectivité chez des patients présentant des désordres de la conscience après un arrêt cardio-respiratoire, nous avons montré que ces indicateurs globaux ne sont pas représentatifs des changements pour des patients dans le coma, figure 20. Il y a conservation des propriétés topologiques globales des graphes : existence de régions “hubs”, et de régions moins connectées même chez les patients. Nous avons donc cherché à savoir pourquoi, et l'explication vient du fait que pour certains nœuds on observe une décroissance des index alors que pour d'autres, on observe une croissance, figure 23. En prenant la moyenne il n'y a donc aucune différence alors que les patients présentent des désordres de la conscience et ne sont pas sensibles à l'environnement qui les entoure par exemple.

Nous avons donc introduit un index qui mesure la différence entre un patient et la moyenne sur un groupe contrôle des indices obtenus au niveau des nœuds du réseau, figure 21. Cet index appelé *coefficient de perturbation des hubs* et noté κ , nous a permis de discriminer parfaitement les deux groupes patients et contrôles figure 22.

5.3 Validation du coefficient de perturbation des hubs

Une première validation consiste à comparer le pouvoir discriminant de cet index par rapport à d'autres méthodes non supervisées comme LDA (linear discriminant analysis) (FISHER 1936 ; ROBINSON et al. 2010), FS (feature selection) (PEDREGOSA et al. 2011) pour lesquels nous avons obtenus des résultats similaires (voir table 3). En appliquant des

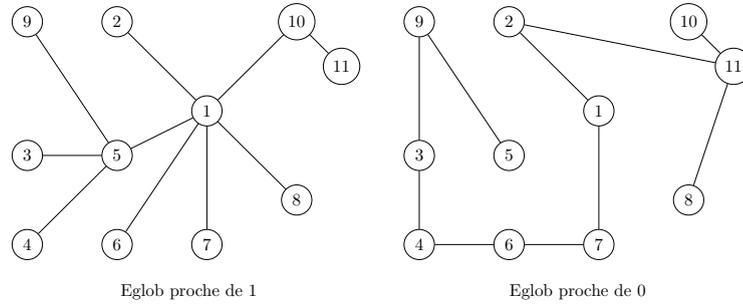


FIGURE 18 – Exemple de deux graphes ayant le même nombre de nœuds et d’arêtes, mais une efficacité globale très différente.

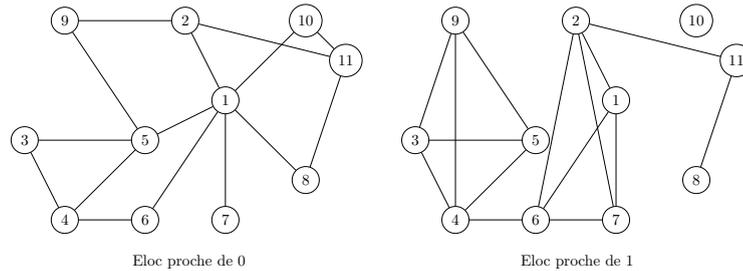


FIGURE 19 – Exemple de deux graphes ayant le même nombre de nœuds et d’arêtes, mais une efficacité locale très différente.

méthodes ISOMAP (TENENBAUM et al. 2000), nous avons pu voir que pour ce jeu de données, un des axes obtenus lors de la réduction de dimension correspond au coefficient de perturbation des hubs, figure 24. Ceci nous a permis d’introduire un nouvel algorithme d’ISOMAP en proposant à l’utilisateur de pouvoir fixer un des axes avec une caractéristique recherchée du jeu de données. Ceci peut permettre en particulier une meilleure interprétation du résultat de classification, figure 25. Dans un deuxième temps, nous avons appliqué ces résultats sur d’autres jeux de données réelles pour des patients (RIDLEY et al. 2015; HEMMERT et al. 2013) ou des volontaires sains en état de méditation (JAO et al. 2015).

Coefficients	<i>Eglob</i> (<i>p</i> -valeur)	<i>Clust</i> (<i>p</i> -valeur)	<i>D</i> (<i>p</i> -valeur)
κ	1.9×10^{-9}	7×10^{-9}	5.3×10^{-9}
<i>LDA</i>	8.7×10^{-10}	9.8×10^{-6}	5.1×10^{-8}
<i>FS</i>	2.1×10^{-6}	3.2×10^{-7}	1.1×10^{-6}
<i>MEAN</i>	0.58	0.43	

TABLE 3 – Comparaison de la *p*-valeur d’un t-test entre le groupe de volontaires sains et le groupe de patients dans le coma avec différentes méthodes d’extraction de caractéristiques.

5.4 Comparaison par profil de probabilités

Toujours dans le but de comparer des populations de sujets, au lieu d’extraire des paramètres topologiques des graphes, nous avons proposé dans la thèse d’Aude Costard de

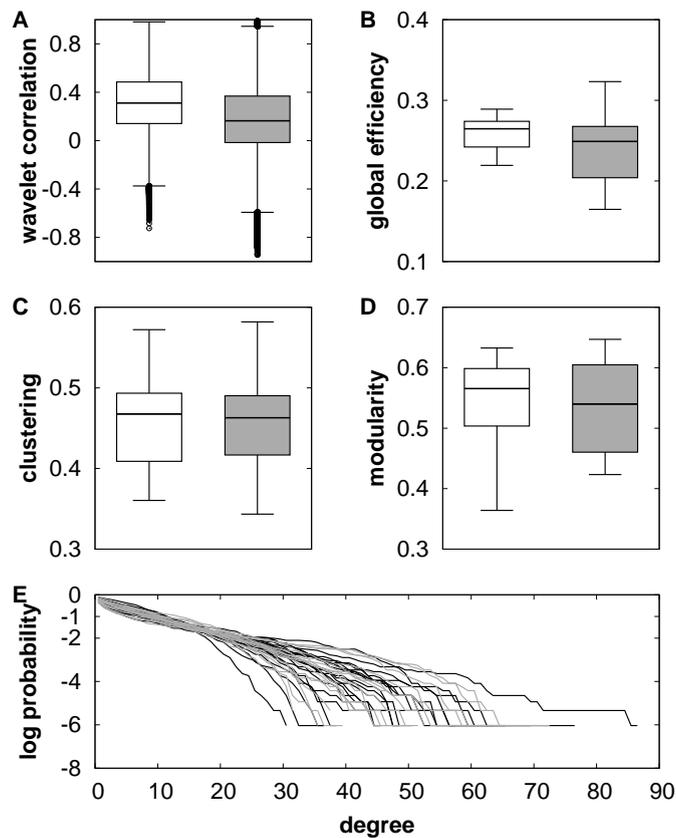


FIGURE 20 – **Comparaison des propriétés topologiques des réseaux de connectivité de patients dans le coma (gris) et de volontaires sains (blanc).** (A) Moyenne des corrélations par ondelettes, une mesure de la connectivité fonctionnelle sur toutes les paires de régions cérébrales. (B) Efficacité globale, une mesure topologique de la diffusion du transfert d'information reliée à la notion de plus court chemin. (C) Clustering, une mesure topologique liée à la robustesse du graphe. (D) Modularité, une mesure globale qui quantifie la possibilité de décomposer le graphe en communautés. (E) distribution des degrés, la fonction de répartition du degré des nœuds du graphe. Figure extraite de (ACHARD et al. 2012).

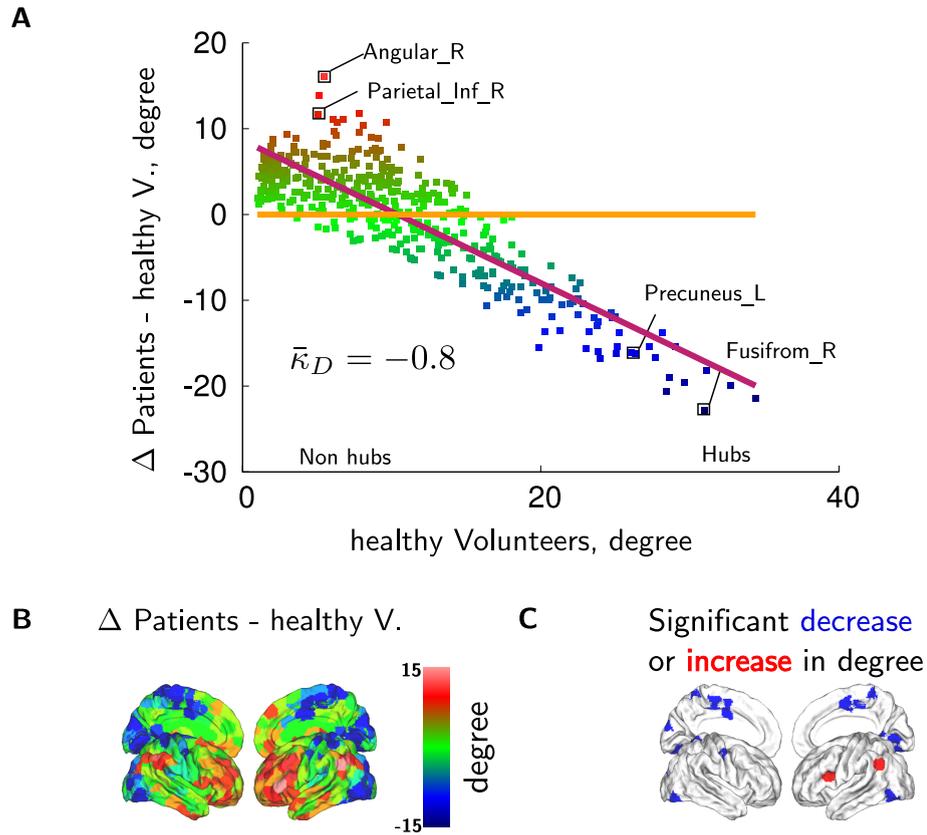


FIGURE 21 – **Propriétés du degré des réseaux de connectivité chez les patients dans le coma et les volontaires sains.** (A) Illustration de la perturbation du degré des nœuds des réseaux de connectivité cérébrale chez les patients dans le coma. (B) On représente ici la croissance et décroissance anormales des degrés des nœuds en comparant la différence entre les degrés des graphes du groupe de patients et ceux du groupe de volontaires sains en fonction de la moyenne pour chaque nœud des graphes des volontaires sains. La droite en trait plein est la représentation de la droite de régression obtenue sur le nuage de points, on note ce coefficient $\bar{\kappa}_D$. (C) Représentation sur la surface corticale de la différence de la moyenne des degrés entre les patients et les volontaires sains. (D) Représentation sur la surface corticale des régions du groupe de patients significativement différente par rapport au groupe de volontaires sains (après correction pour comparaisons multiples). Figure extraite de (ACHARD et al. 2012). Figure extraite de (ACHARD et al. 2012).

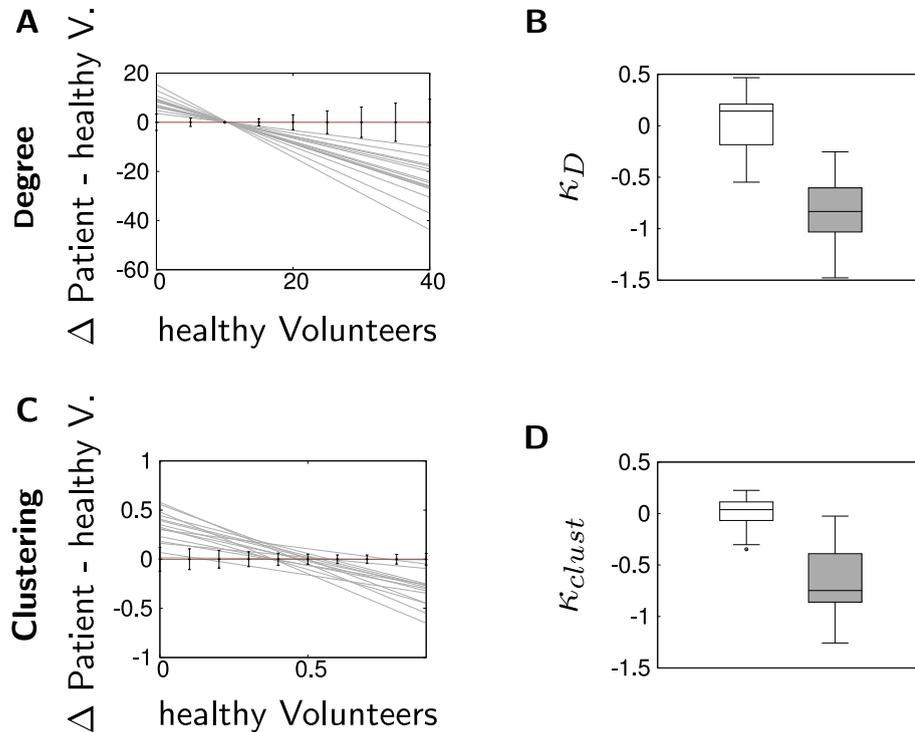


FIGURE 22 – **Comparaison des deux groupes en utilisant le coefficient de perturbation des hubs (A) et (C)** Représentation graphique des coefficients de perturbation des hubs κ_D et κ_{clust} estimés pour chaque patient (lignes grises). La ligne noire horizontale indique les mêmes coefficients obtenus pour chaque individu du groupe des volontaires sains (barre d'erreur = 1 écart-type). (B) et (D) Comparaison des deux groupes, les valeurs pour le groupe des volontaires sains sont autour de zéro, cependant les valeurs obtenues pour le groupe des patients sont autour de -1. Les deux groupes sont statistiquement différents quelle que soit la métrique utilisée : ici pour le degré ($p < 10^{-5}$) et pour le clustering ($p < 10^{-5}$). Figure extraite de (ACHARD et al. 2012).

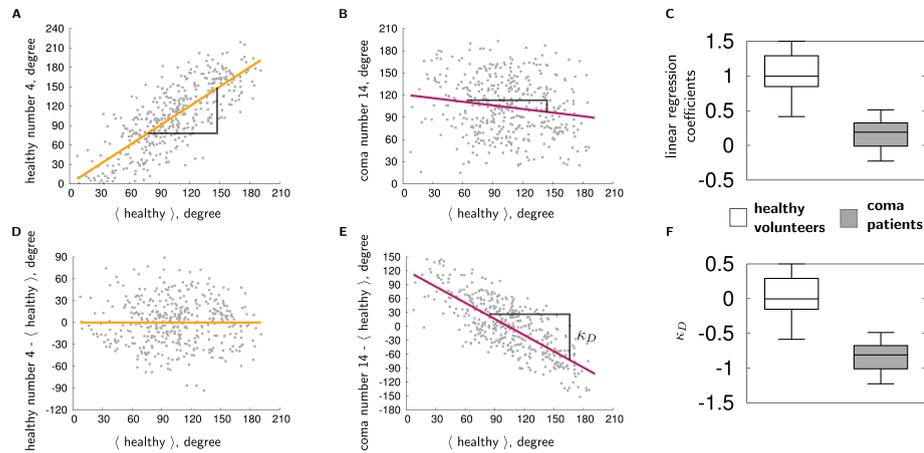


FIGURE 23 – **Détails sur la construction du coefficient de perturbation des hubs** (Première ligne) Les valeurs obtenues au niveau des nœuds du graphe en fonction de la moyenne sur les données des volontaires sains (A) pour un volontaire sain (B) pour un patient dans le coma. (C) Estimation de la pente de la droite de régression sur les données. Les coefficients de pente pour les volontaires sains oscillent autour de la valeur 1, ce qui montre que les sujets sont assez similaires entre eux. Par contre, pour les patients, les valeurs des pentes oscillent autour de 0 ce qui montre que les valeurs obtenues pour les patients ne peuvent pas être induites à partir des valeurs moyennes obtenues sur les volontaires sains. (Deuxième ligne) Dans l'article (ACHARD et al. 2012) nous avons choisi de soustraire la moyenne obtenue pour tous les volontaires sains à chacun des sujets avant de calculer la pente de la droite de régression. Ceci consiste simplement à soustraire 1 aux valeurs obtenues sans la soustraction. Les données pour un volontaire sain (D) sont réparties autour de la ligne horizontale zéro, tandis que les données pour un patient dans le coma (E) sont distribuées autour d'une droite de pente négative autour de -1. (F) regroupe les valeurs de la pente de régression pour tous les sujets de l'étude. Figure extraite de (ACHARD et al. 2012).

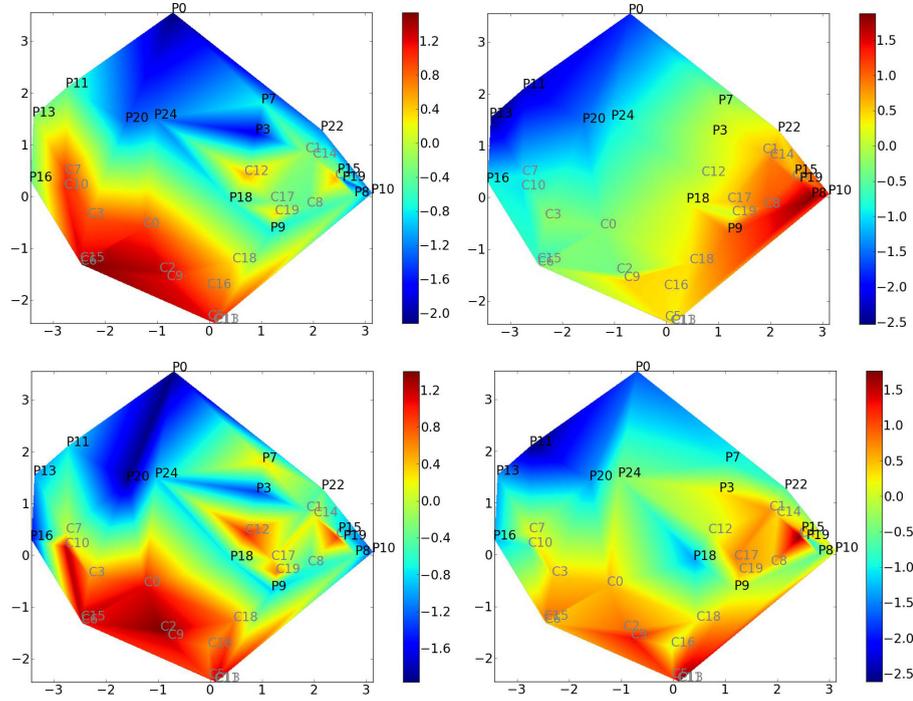


FIGURE 24 – Superposition d’un mapping non linéaire ISOMAP avec les coefficients de perturbation des hubs (en haut à gauche), moyenne (en haut à droite), LDA (en bas à gauche) et FS (en bas à droite).

comparer les probabilités $P(G|\mathbf{X})$ introduites dans la section 4.3. Sur la base de l’expression de la probabilité a posteriori (28), nous avons défini un score z qui mesure l’adéquation entre une structure de graphe donnée et les observations.

$$z(G|\mathbf{X}) = \frac{\varphi_{\mathbf{p}_{\bar{E}}, W_{\bar{E}\bar{E}}}(\mathbf{0})}{V(G)} \quad (32)$$

où W est la matrice d’Isserlis (ROVERATO et WHITTAKER 1998) de la corrélation empirique S . $\varphi_{\mathbf{p}_{\bar{E}}, W_{\bar{E}\bar{E}}}$ est la gaussienne de moyenne \mathbf{p} et de covariance $W_{\bar{E}\bar{E}}$ avec \mathbf{p} le vecteur des corrélations partielles. \bar{E} et $\bar{E}\bar{E}$ signifient que l’on considère seulement les arêtes non présentes dans G .

Nous appliquons d’abord la méthode décrite dans la section 4.3 qui effectue une recherche guidée autour d’un point initial obtenu par *glasso* (FRIEDMAN et al. 2008). Ensuite, nous utilisons le critère (32) pour le développement de méthodes de comparaison de deux groupes de données.

A la différence de la figure 12 où il est difficile d’identifier la structure des graphes, la figure 27 illustre les bonnes propriétés de notre méthode même pour de faibles tailles d’échantillons, les structures de graphes utilisées sont rappelées dans la figure 26.

5.5 Perspectives

Une étude théorique approfondie du coefficient de perturbation des hubs est nécessaire en terme de robustesse et de reproductibilité. C’est en effet une étape cruciale pour promouvoir l’utilisation de données quantitatives en clinique. Celle-ci consiste tout d’abord

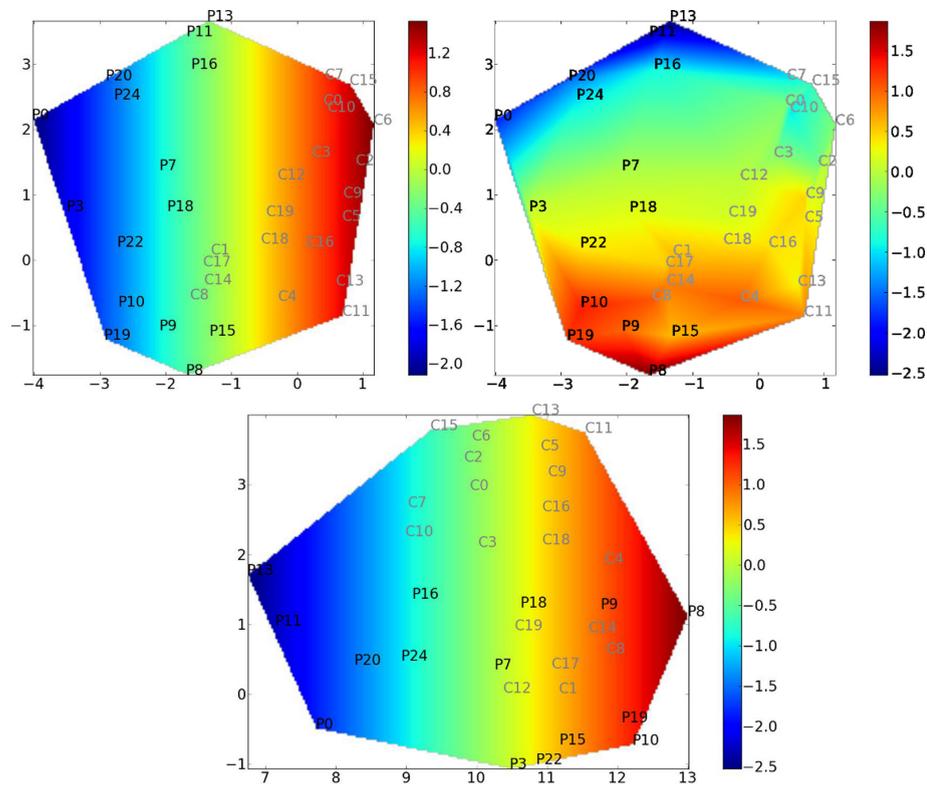


FIGURE 25 – Nouvelle décomposition en fixant l’axe principal selon une certaine caractéristique, le coefficient de perturbation des hubs (en haut), la moyenne (en bas). La carte de couleur indique la superposition du résultat avec soit le coefficient de perturbation des hubs pour la carte en haut à gauche soit la moyenne pour les cartes en haut à droite et du bas.

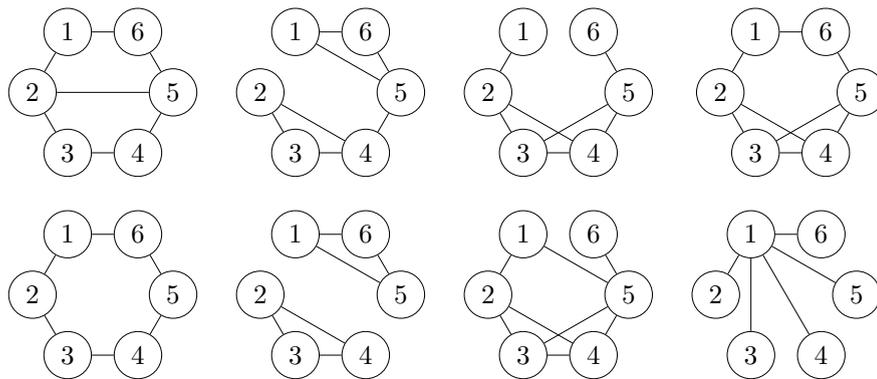


FIGURE 26 – Structures de graphes choisies pour les comparaisons des performances des méthodes

en l’étude statistique de l’estimateur et en la validation des résultats sur des données de patients avec des pathologies différentes et des données simulées. Puis, nous proposons de tester la reproductibilité de ce paramètre sur des données dites de *test-retest* qui se composent par l’acquisition de deux jeux de données d’IRMf sur plusieurs sujets. Cette étude est en cours dans le cadre de la thèse de Maite Termenón en utilisant des données disponibles en ligne par le Human Connectome Project sur 100 sujets.

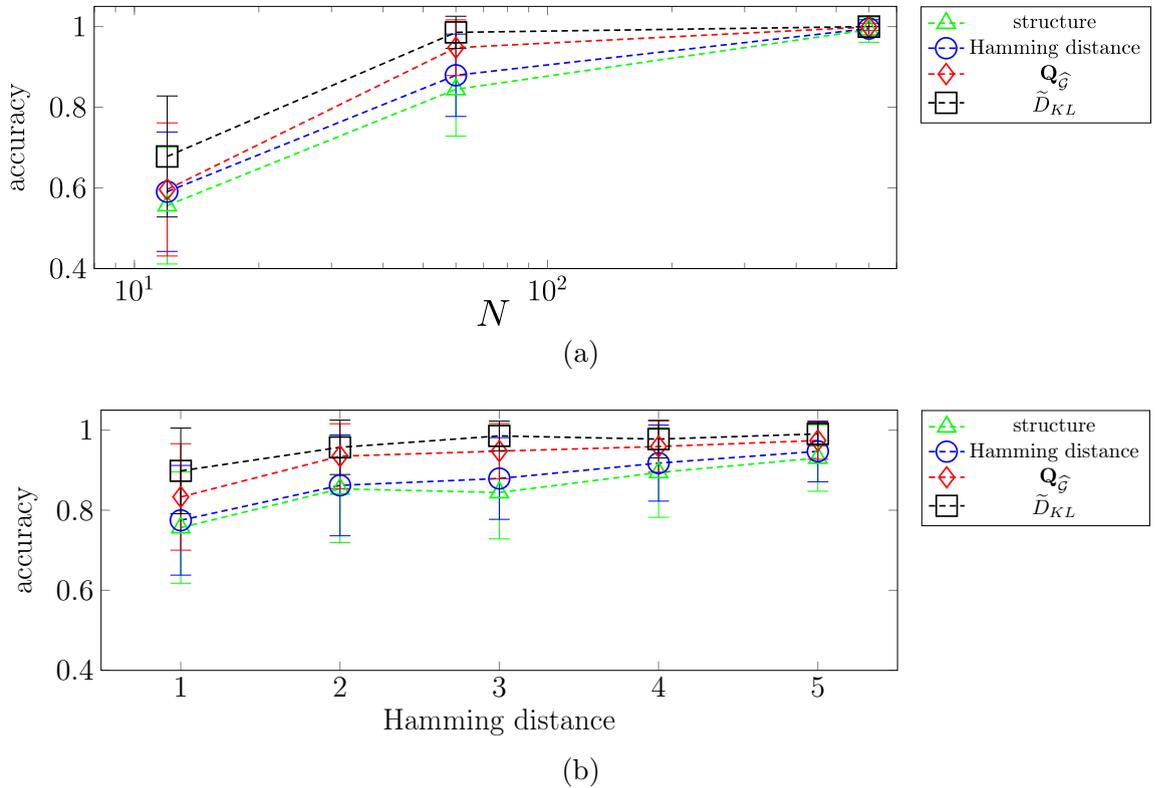


FIGURE 27 – Au lieu de comparer les graphes estimés par rapport à la structure attendue, nous nous sommes concentrés ici sur la capacité de classification du score z . Nous avons simulé deux groupes de 20 jeux de données provenant de deux structures de graphes différentes. La précision reportée sur la figure mesure la performance du classifieur à séparer les deux groupes, plus elle est proche de 1, plus la méthode est performante. (a) illustre la précision de la méthode en fonction de la taille de l'échantillon. Pour $N = 60$, on remarque que même si la structure de graphe n'est pas bien estimée, la méthode permet de séparer les deux groupes avec de bonnes performances. (b) illustre la précision de la méthode quand les structures de graphes choisies pour les deux groupes sont très similaires (distance de Hamming de 1) ou bien très éloignées (distance de Hamming de 5).

6 Logiciels

Dans un souci de reproductibilité des résultats, j'ai mis à disposition trois logiciels en accès libre.

6.1 brainwaver

Le logiciel *brainwaver* (sous R) est dédié à l'analyse des données d'IRMf. Il permet à partir d'un ensemble de séries temporelles d'IRMf par exemple de construire les matrices de corrélations par ondelettes et d'extraire les matrices d'adjacence en sélectionnant les valeurs de corrélations suffisamment grandes. Puis, le logiciel permet de calculer des paramètres topologiques des graphes classiquement utilisés dans le cadre de réseaux de connectivité cérébrale. Ce logiciel a été utilisé dans les publications (ACHARD et al. 2006 ; ACHARD et BULLMORE 2007 ; ACHARD et al. 2012).

6.2 multiwave

Le logiciel *multiwave* (disponible sous R et matlab <http://math.univ-lyon1.fr/~gannaz/recherche.html#codes>) implémente les estimations multivariées par décomposition en ondelettes ou transformée de Fourier présentées dans la section 3. Pour un jeu de données multivariées, le logiciel permet d’extraire les estimations des paramètres longues mémoires et de la matrice de connectivité fractale. Ce logiciel a été utilisé dans la publication (ACHARD et GANNAZ 2016).

6.3 modèles graphiques

Une boîte à outils (disponible sous matlab) qui rassemble les codes nécessaires à l’analyse de l’indépendance conditionnelle par modèles graphiques gaussiens est disponible sur la page d’Aude Costard <http://www.gipsa-lab.grenoble-inp.fr/~aude.costard/toolbox.html>. Cette boîte à outils est composée de

- méthode ABiGlasso (Thèse Aude Costard (COSTARD 2014))
- méthode Bayesian Adaptive Graphical lasso (WANG 2012)
- méthode GGMselect (GIRAUD et al. 2012)
- méthode SIN (DRTON et PERLMAN 2008)

7 Conclusion et projets futurs

Nous avons présenté dans ce manuscrit une approche multivariée originale pour l’estimation de la structure de graphes avec les paramètres longues mémoires et la connectivité fractale. La mise en pratique de celle-ci sur des données réelles en IRMf en particulier a abouti à la mise en place d’un protocole qui a généré de nombreuses collaborations et des publications avec le CHU de Grenoble, de Strasbourg et de Marseille.

Ces recherches se poursuivent dans un cadre interdisciplinaire très enrichissant, avec à la fois des collègues statisticiens, de traitement du signal, des physiciens, des neurologues, réanimateurs, psychiatres

L’enjeu crucial de ces recherches est la possibilité d’aller de l’étude théorique des estimateurs jusqu’aux patients pour une meilleure compréhension de certaines pathologies. Du point de vue méthodologique, les données complexes obtenues par IRMf ne peuvent être exploitées directement, un pré-traitement est obligatoire pour permettre l’extraction et l’exploitation des réseaux de connectivité cérébrale. De plus, il a déjà été montré que le signal BOLD est sujet à des variations provenant de plusieurs paramètres : rythme cardiaque, respiratoire, vasodilatation (BLANCHARD et al. 2011 ; MUTCH et al. 2012). Ces variations, particulièrement importantes dans le cas d’étude de patients, peuvent influencer la détection de zones activées (JIANG et al. 2010).

L’objectif de mon projet de recherche est de mieux comprendre le signal BOLD enregistré en IRMf afin de permettre un apprentissage statistique des réseaux de connectivité cérébrale précis et robuste. Une connaissance approfondie du signal BOLD d’un point de vue biophysique et traitement du signal est donc nécessaire au développement de nouvelles approches de connectivité qui pourront être plus précises et moins sensibles à la présence d’artéfacts ou de bruits physiologiques, inhérents à ce type de données. Cet aspect théorique sera complété par l’application et le test de ces résultats sur des données petit animal

comme le rat, grâce à la possibilité de répéter les expériences pour une pré-évaluation clinique fiable. Enfin, la confrontation avec des données réelles sur l'homme sera cruciale.

Le programme scientifique se composera de trois axes principaux : l'étude des propriétés statistiques du signal BOLD, le développement d'outils d'exploration des réseaux, et l'application sur données réelles. La figure 28 illustre le déroulement de ce projet. Le lien entre développement méthodologique et données réelles est un point crucial du projet, et les approches proposées seront menées en concertation permanente entre les statisticiens pour le côté méthodologique, les médecins pour la compréhension du signal BOLD et les médecins pour les applications cliniques.

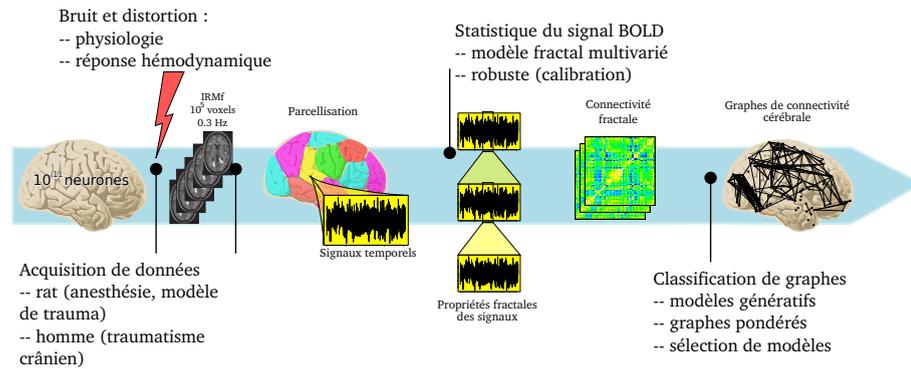


FIGURE 28 – Le processus d'apprentissage statistique de réseaux de connectivité cérébrale et description des axes du programme scientifique.

7.1 Axe 1 : propriétés statistiques du signal BOLD

Nous avons montré dans le manuscrit l'importance de la prise en compte de la longue mémoire pour l'étude des séries multivariées. Il a déjà été montré que le signal BOLD présente de la longue mémoire (MAXIM et al. 2005). Cette modélisation a été déjà validée en montrant une amélioration pour la détection de zones activées (BULLMORE et al. 2004). Par ailleurs, les propriétés fractales du signal BOLD sont modulées par l'âge et la prise de médicaments (WINK et al. 2006 ; SUCKLING et al. 2008). Cependant, le phénomène sous-jacent à ces changements est encore méconnu. Récemment, en 2012, MUTCH et al. (2012) ont montré que les propriétés fractales du signal BOLD changent avec le taux de dioxyde de carbone, en parallèle, CANTIN et al. (2011) ont montré l'influence d'une étude de la vaso-réactivité pour des patients Alzheimer. Ces études montrent que les propriétés statistiques du signal BOLD (et de la modélisation de la fonction hémodynamique (CIUCIU et al. 2003)) sont un enjeu pour l'interprétation des données acquises en IRMf. Cet axe consiste en l'étude théorique de l'estimateur de connectivité fractale et le test de cet estimateur sur les données sur le rat qui peuvent offrir suffisamment d'échantillons pour fournir une estimation efficace (ACHARD et COEURJOLLY 2010).

7.2 Axe 2 : Développement d'outils d'exploration des réseaux

Une fois établis les graphes de connectivité pour chaque jeu de données, il est nécessaire d'établir une stratégie de comparaison de ces graphes. Les matrices de connectivité sont souvent trop complexes pour permettre une interprétation, et une étude approfondie de

la topologie du graphe est souvent plus robuste et plus précise (FALLANI et al. 2014). Nous avons déjà travaillé sur la mise au point d'indicateurs de perturbations des hubs sur les réseaux de connectivité (ACHARD et al. 2012). Une perspective est de développer des méthodes d'apprentissage statistique à l'aide de modèles génératifs de graphes. Les travaux de VÉRTES et al. (2012) ont montré la pertinence du point de vue clinique de développer des modèles de génération de graphes qui permettent d'inférer des caractéristiques topologiques très différentes en fonction de la pathologie étudiée. De plus, les estimateurs de l'axe 1 nous fournissent un scalaire entre 0 et 1 de force du lien existant entre chaque paire de régions. Ces pondérations pourront être utilisées dans le développement d'extraction de caractéristiques pour définir le poids des arêtes des graphes (RICHIARDI et al. 2013).

7.3 Axe 3 : Développement en lien avec les données réelles

Afin que les développements méthodologiques soient utilisables dans le cas de données réelles provenant de patients par exemple, les différentes étapes de ce projet seront en lien direct avec deux types de données : des données acquises sur le petit animal, et des données sur des patients présentant des troubles de la conscience après un traumatisme crânien. Enfin, l'acquisition de signaux physiologiques en parallèle des acquisitions d'IRMf permettra de mieux comprendre l'apparition de l'effet longue mémoire et de le corriger.

Données sur le petit animal, application pré-clinique

Dans le cadre d'étude pré-clinique, l'équipe 5 du GIN gère la plate-forme IRMaGE dans le domaine de l'imagerie RMN et des explorations fonctionnelles cérébrales. En particulier, la plate-forme petit animal permet le développement et le test de méthodologies innovantes dans un cadre pré-clinique. Nous avons déjà procédé à l'acquisition de données d'IRMf au repos sur le rat avec deux doses d'anesthésique différentes. Dans le cadre d'un premier stage de master, il a fallu procéder aux réglages de ces acquisitions avec le contrôle du rapport signal sur bruit, de la présence d'artéfact et du conditionnement des rats. Nous envisageons maintenant de nous attaquer plus précisément au traitement des données par les méthodes développées dans les axes 1 et 2. Il est attendu aussi que les données déjà acquises seront insuffisantes à la validation des méthodes, et nous envisageons donc l'acquisition de nouvelles données en adéquation avec les résultats des axes 1 et 2. Le choix de l'anesthésique engendre aussi des complications, et nous envisageons de tester d'autres possibilités plus adaptés aux phénomènes de déconnexion que nous souhaitons mettre en lumière. Enfin, l'équipe 5 du GIN a acquis un savoir-faire reconnu dans la simulation in vivo de traumatisme crânien chez le rat. Il est alors possible de connaître de manière très précise l'emplacement des lésions et aussi d'évaluer les problèmes dus à la perturbation de l'oxygénation et donc du signal BOLD. Ces modèles fiables et reproductibles seront utilisés dans le projet pour permettre la calibration et la validation des estimateurs définis dans l'axe 1.

Données sur l'homme dans des cas cliniques

Les données sur l'homme proviendront d'un projet en cours financé par l'association "Gueules cassées". Ce projet qui est porté par Lydia Oujamaa (Praticien Hospitalier responsable du service Soins de Réadaptation Post Réanimation du CHU de Grenoble) a pour but l'exploration de la connectivité fonctionnelle cérébrale après coma traumatique. Ces méthodes d'exploration du cerveau entier par IRMf de repos sont nouvelles et permettront aux médecins de compléter leurs diagnostics en apportant une vision complémentaire des examens cliniques faits par ailleurs. L'hypothèse médicale avancée est que l'altération de la conscience d'un sujet cérébrolésé est caractérisée par une perte de connectivité fonctionnelle dans des régions cérébrales nécessaires à la réémergence de la conscience. Nous

voulons décrire ces régions dans le cadre d'une étude pilote. Nous allons comparer la connectivité fonctionnelle du traumatisé crânien un mois après l'accident et ensuite environ trois mois après. Nous rechercherons également un lien entre l'évolution clinique du niveau de conscience et l'évolution de la connectivité fonctionnelle au cours du temps. Pour cela nous recruterons 15 patients cérébrolésés traumatiques conscients et autant en état de conscience altéré lors de leur hospitalisation en réanimation au CHU de Grenoble. La période de recrutement devrait durer deux ans pour cette étude mono-centrique. Les sujets bénéficieront d'une évaluation clinique de la conscience et d'une IRMf, réalisées avant leur sortie de réanimation. Un suivi longitudinal sera effectué après 60 jours. Notre but serait d'identifier un marqueur neuro-physiologique issu des réseaux de connectivité cérébrale observés. Ceci pourrait améliorer la précision du diagnostic d'état de conscience altéré et en conséquence la prise en charge rééducative précoce des traumatisés crâniens graves.

Références

- ABADIR, K. M., W. DISTASO et L. GIRAITIS (2007). "Nonstationarity-extended local whittle estimation". In : *Journal of Econometrics* 141.2, p. 1353-1384.
- ABRY, P. et D. VEITCH (jan. 1998). "Wavelet analysis of long-range-dependent traffic". In : *IEEE Transactions on Information Theory* 44.1, p. 2-15.
- ACHARD, S., D. BASSETT, A. MEYER-LINDENBERG et E. BULLMORE (2008). "Fractal connectivity of long-memory networks". In : *Physical Review E* 77, p. 036104.
- ACHARD, S. et E. BULLMORE (2007). "Efficiency and cost of economical human brain functional networks". In : *PLoS Computational Biology* 3, e17.
- ACHARD, S. et J.-F. COEURJOLLY (2010). "Discrete variations for the fractional Brownian motion in presence of outliers and/or an additive noise". In : *Statistics Surveys* 4, p. 117-147.
- ACHARD, S., C. DELON-MARTIN, P. VÉRTES, F. RENARD, M. SCHENCK, F. SCHNEIDER, C. HEINRICH, S. KREMER et E. T. BULLMORE (déc. 2012). "Hubs of brain functional networks are radically reorganized in comatose patients." In : *Proc. Natl. Acad. Sci.* 109.50, p. 20608-20613. DOI : [10.1073/pnas.1208933109](https://doi.org/10.1073/pnas.1208933109).
- ACHARD, S. et I. GANNAZ (2016). "Multivariate wavelet whittle estimation in long-range dependence". In : *Journal of Time Series Analysis* 37.4. arXiv preprint arXiv :1412.0391, p. 476-512. DOI : [10.1111/jtsa.12170](https://doi.org/10.1111/jtsa.12170).
- ACHARD, S., R. SALVADOR, B. WHITCHER, J. SUCKLING et E. BULLMORE (jan. 2006). "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs". In : *The Journal of Neuroscience* 26.1. 00831, p. 63-72. ISSN : 0270-6474, 1529-2401. DOI : [10.1523/JNEUROSCI.3874-05.2006](https://doi.org/10.1523/JNEUROSCI.3874-05.2006).
- AMINI, L., C. JUTTEN, S. ACHARD, O. DAVID, P. KAHANE, L. VERCUEIL, L. MINOTTI, G. A. HOSSEIN-ZADEH et H. SOLTANIAN-ZADEH (nov. 2010). "Comparison of five directed graph measures for identification of leading interictal epileptic regions." In : *Physiol Meas* 31.11, p. 1529-1546. DOI : [10.1088/0967-3334/31/11/009](https://doi.org/10.1088/0967-3334/31/11/009).
- AMINI, L., C. JUTTEN, S. ACHARD, O. DAVID, H. SOLTANIAN-ZADEH, G. A. HOSSEIN-ZADEH, P. KAHANE, L. MINOTTI et L. VERCUEIL (avr. 2011). "Directed differential connectivity graph of interictal epileptiform discharges." In : *IEEE Trans Biomed Eng* 58.4, p. 884-893. DOI : [10.1109/TBME.2010.2099227](https://doi.org/10.1109/TBME.2010.2099227).
- ANDERSON, T. (2003). *An introduction to multivariate statistical analysis*. Third edition. John Wiley & Sons.
- BASSETT, D. S., A. MEYER-LINDENBERG, S. ACHARD, T. DUKE et E. T. BULLMORE (2006). "Adaptive reconfiguration of fractal small-world human brain functional networks". In : *Proceedings National Academy of Sciences (USA)* 103, p. 19518-19523.
- BENJAMINI, Y. et Y. HOCHBERG (1995). "Controlling the false discovery rate : a practical and powerful approach to multiple testing". In : *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, p. 289-300. ISSN : 00359246.
- BENJAMINI, Y. et D. YEKUTIELI (2001). "The control of the false discovery rate in multiple testing under dependency". In : *Annals of statistics*, p. 1165-1188.
- BERAN, J. (1994). *Statistics for long-memory processes*. New York : Chapman et Hall.
- BERAN, J. (2010). "Long-range dependence". In : *Wiley Interdisciplinary Reviews : Computational Statistics* 2.1, p. 26-35.
- BLANCHARD, S., T. PAPADOPOULOU, C.-G. BÉNAR, N. VOGES, M. CLERC, H. BENALI, J. WARNKING, O. DAVID et F. WENDLING (mar. 2011). "Relationship between flow and metabolism in bold signals : insights from biophysical models." In : *Brain Topogr* 24.1, p. 40-53. DOI : [10.1007/s10548-010-0166-6](https://doi.org/10.1007/s10548-010-0166-6).
- BOX, G. E., G. M. JENKINS et G. C. REINSEL (2011). *Time series analysis : forecasting and control*. T. 734. John Wiley & Sons.
- BUCKNER, R. L., J. SEPULCRE, T. TALUKDAR, F. M. KRIENEN, H. LIU, T. HEDDEN, J. R. ANDREWS-HANNA, R. A. SPERLING et K. A. JOHNSON (2009). "Cortical hubs revealed by intrinsic functional connectivity : mapping, assessment of stability, and relation to alzheimer's disease". In : *J Neurosci*. 29.6, p. 1860-73.
- BULLMORE, E. et O. SPORNS (mar. 2009). "Complex brain networks : graph theoretical analysis of structural and functional systems." In : *Nat Rev Neurosci* 10.3, p. 186-198. DOI : [10.1038/nrn2575](https://doi.org/10.1038/nrn2575).

- BULLMORE, E., J. FADILI, V. MAXIM, L. SENDUR, B. WHITCHER, J. SUCKLING, M. BRAMMER et M. BREAKSPEAR (2004). "Wavelets and functional magnetic resonance imaging of the human brain". In : *NeuroImage* 23.Sup 1, p. 234-249.
- BUZSÁKI, G. (2006). *Rhythms of the brain*. Oxford University Press.
- CANTIN, S., M. VILLIEN, O. MOREAUD, I. TROPRES, S. KEIGNART, E. CHIPON, J.-F. LE BAS, J. WARNKING et A. KRAINIK (sept. 2011). "Impaired cerebral vasoreactivity to co2 in alzheimer's disease using bold fMRI." In : *Neuroimage* 58.2, p. 579-587. DOI : [10.1016/j.neuroimage.2011.06.070](https://doi.org/10.1016/j.neuroimage.2011.06.070).
- CAUDA, F., B. M. MICON, K. SACCO, S. DUCA, F. D'AGATA, G. GEMINIANI et S. CANAVERO (avr. 2009). "Disrupted intrinsic functional connectivity in the vegetative state." In : *J Neurol Neurosurg Psychiatry* 80.4, p. 429-431. DOI : [10.1136/jnnp.2007.142349](https://doi.org/10.1136/jnnp.2007.142349).
- CHAMBERS, M. (1995). "The simulation of random vector time series with given spectrum". In : *Mathematical and Computer Modelling* 22.2, p. 1-6.
- CIUCIU, P., J.-B. POLINE, G. MARRELEC, J. IDIER, C. PALLIER et H. BENALI (2003). "Unsupervised robust nonparametric estimation of the hemodynamic response function for any fmri experiment". In : *Medical Imaging, IEEE Transactions on* 22.10, p. 1235-1251.
- COEURJOLLY, J.-F., P.-O. AMBLARD et S. ACHARD (2013). "Wavelet analysis of the multivariate fractional brownian motion." In : *ESAIM : Probability and Statistics* 17, p. 592-604. DOI : [doi:10.1051/ps/2012011](https://doi.org/10.1051/ps/2012011).
- COSTARD, A., S. ACHARD, O. MICHEL, P. BORGNAT et P. ABRY (2013). "Estimation bayésienne asymptotique de la structure d'un graphe initialisée par graphical lasso". In : *GRETSI'13*. Brest, France.
- (2014). "Data comparison using gaussian graphical models". In : *12th international Conference on Signal Processing (ICSP)*. China, p. 1346-1351.
- COSTARD, A. (nov. 2014). "Estimation de la structure d'indépendance conditionnelle d'un réseau de capteurs. application à l'imagerie médicale". Thèse de doctorat. Université de Grenoble.
- DAHLHAUS, R. (1989). "Efficient parameter estimation for self-similar processes". In : *The Annals of Statistics*, p. 1749-1766.
- DAUBECHIES, I. (1992). *Ten lectures on wavelets*. T. 61. SIAM.
- DEMPSTER, A. P. (1972). "Covariance selection". In : *Biometrics*, p. 157-175.
- DONNET, S. et J.-M. MARIN (2012). "An empirical bayes procedure for the selection of gaussian graphical models". In : *Statistics and Computing* 22.5, p. 1113-1123.
- DRTON, M. et M. D. PERLMAN (2004). "Model selection for Gaussian concentration graphs". In : *Biometrika* 91.3, p. 591-602.
- (2008). "A SINful approach to Gaussian graphical model selection". In : *Journal of Statistical Planning and Inference* 138.4, p. 1179-1200.
- EDWARDS, D. (2012). *Introduction to graphical modelling*. Springer Science & Business Media.
- FALLANI, F. D. V., J. RICHIARDI, M. CHAVEZ et S. ACHARD (2014). "Graph analysis of functional brain networks : practical issues in translational neuroscience". In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 369.1653, p. 20130521. DOI : [10.1098/rstb.2013.0521](https://doi.org/10.1098/rstb.2013.0521).
- FISHER, R. A. (1936). "The use of multiple measurements in taxonomic problems". In : *Annals of Eugenics* 7.2, p. 179-188. ISSN : 2050-1439. DOI : [10.1111/j.1469-1809.1936.tb02137.x](https://doi.org/10.1111/j.1469-1809.1936.tb02137.x).
- FLANDRIN, P. (1992). "Wavelet analysis and synthesis of fractional Brownian motion". In : *IEEE Trans. Inform. Theory* 38, p. 910-17.
- FLANDRIN, P. (1998). *Time-frequency/time-scale analysis*. T. 10. Academic press.
- FOX, R. et M. S. TAQQU (1986). "Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series". In : *The Annals of Statistics*, p. 517-532.
- FRIEDMAN, J., T. HASTIE et R. TIBSHIRANI (2008). "Sparse inverse covariance estimation with the graphical lasso". In : *Biostatistics* 9.3, p. 432-441. DOI : [10.1093/biostatistics/kxm045](https://doi.org/10.1093/biostatistics/kxm045). eprint : <http://biostatistics.oxfordjournals.org/content/9/3/432.full.pdf+html>.
- GEWEKE, J. et S. PORTER-HUDAK (1983). "The estimation and application of long memory time series models". In : *Journal of Time Series Analysis* 4.4, p. 221-238.
- GIRAITIS, L., P. M. ROBINSON et A. SAMAROV (1997). "Rate optimal semiparametric estimation of the memory parameter of the gaussian time series with long-range dependence". In : *Journal of Time Series Analysis* 18.1, p. 49-60.
- GIRAUD, C., S. HUET et N. VERZELEN (2012). "Graph selection with GGMselect". In : *Statistical applications in genetics and molecular biology* 11.3. DOI : [10.1515/1544-6115.1625](https://doi.org/10.1515/1544-6115.1625).
- GIUDICI, P. et P. J. GREEN (1999). "Decomposable graphical Gaussian model determination". In : *Biometrika* 86.4, p. 785-801.
- GOEMAN, J. J. et A. SOLARI (mai 2014). "Multiple hypothesis testing in genomics." In : *Stat Med* 33.11, p. 1946-1978. DOI : [10.1002/sim.6082](https://doi.org/10.1002/sim.6082).
- GRANGER, C. W. J. et R. JOYEUX (1980). "An introduction to long-memory time series models and fractional differencing". In : *Journal of Time Series Analysis* 1, p. 15-29.
- HANNAN, E. J. (1973). "The asymptotic theory of linear time-series models". In : *Journal of Applied Probability*, p. 130-145.
- HARLÉ, F., F. CHATELAIN, C. GOUY-PAILLER et S. ACHARD (2014). "Rank-based multiple change-point detection in multivariate time series". In : *Proc. 22th European Signal Processing Conference, EUSIPCO 2014*.
- (2015). "Bayesian model for multiple change-points detection in multivariate time series". submitted to IEEE Transaction on Signal Processing.
- HEMMERT, C., S. ACHARD, F. RENARD, C. DELON-MARTIN, J. DIETMANN, J.-P. ARMSPACH, J. DE SÈZE, S. KREMER et al. (2013). "Resting-state functional MRI demonstrates brain network reorganisation in neuromyelitis optica". In : *Multiple sclerosis journal*. T. 19. 11, p. 167-167.

- HERO, A. et B. RAJARATNAMA (2011). "Large-scale correlation screening". In : *Journal of the American Statistical Association* 106.496, p. 1540-1552.
- HOSKING, J. R. M. (1981). "Fractional differencing". In : *Biometrika* 68, p. 165-176.
- HURVICH, C. M. et W. W. CHEN (2000). "An efficient taper for potentially overdifferenced long-memory time series". In : *Journal of Time Series Analysis* 21.2, p. 155-180.
- HURVICH, C. M. et K. RAY (1995). "Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes". In : *Journal of Time Series Analysis* 16.1.
- JAO, T., C.-W. LI, P. E. VÉRTES, C. W. WU, S. ACHARD, C.-H. HSIEH, C.-H. LIOU, J.-H. CHEN et E. BULLMORE (2015). "Large-scale functional brain network reorganization during taoist meditation". In : *Brain connectivity*. DOI : [0.1089/brain.2014.0318](https://doi.org/10.1089/brain.2014.0318).
- JIANG, Z., A. KRAINIK, O. DAVID, C. SALON, I. TROPRES, D. HOFFMANN, N. PANNETIER, E. L. BARBIER, E. R. BOMBINI, J. WARNKING, C. PASTERIS, S. CHABARDES, F. BERGER, S. GRAND, C. SEGEBARTH, E. GAY et J.-F. LE BAS (août 2010). "Impaired fMRI activation in patients with primary brain tumors." In : *Neuroimage* 52.2, p. 538-548. DOI : [10.1016/j.neuroimage.2010.04.194](https://doi.org/10.1016/j.neuroimage.2010.04.194).
- JOLLIFFE, I. (2002). *Principal component analysis. 2nd ed.* Springer Series in Statistics. New York.
- JORDAN, M. I. (2004). "Graphical models". In : *Statistical Science*, p. 140-155.
- LATORA, V. et M. MARCHIORI (2001). "Efficient behavior of small-world networks". In : *Physical Review Letters* 87, p. 198701.
- LAURITZEN, S. L. (1996). *Graphical models*. Oxford University press.
- LEMOINE, J., C. GOUY-PAILLER, S. ACHARD et P.-O. AMBLARD (2009). "Recherche de la connectivité de réseaux complexes. Application en IRMf." In : *Actes du 22ème Colloque sur le Traitement du Signal et des Images, GRETSI 2009*. Département Images et Signal. Dijon France.
- LEONARDI, N., J. RICHIARDI, M. GSCHWIND, S. SIMIONI, J.-M. ANNONI, M. SCHLUEP, P. VUILLEUMIER et D. VAN DE VILLE (2013). "Principal components of functional connectivity : a new approach to study dynamic brain connectivity during rest". In : *NeuroImage* 83, p. 937-50.
- LOBATO, I. N. (1999). "A semiparametric two-step estimator in a multivariate long memory model". In : *Journal of Econometrics* 90.1, p. 129-153. DOI : [http://dx.doi.org/10.1016/S0304-4076\(98\)00038-4](http://dx.doi.org/10.1016/S0304-4076(98)00038-4).
- LYNALL, M.-E., D. S. BASSETT, R. KERWIN, P. J. MCKENNA, M. KITZBICHLER, U. MULLER et E. BULLMORE (juil. 2010). "Functional connectivity and brain networks in schizophrenia." In : *J Neurosci* 30.28, p. 9477-9487. DOI : [10.1523/JNEUROSCI.0333-10.2010](https://doi.org/10.1523/JNEUROSCI.0333-10.2010).
- MANDELBROT, B. B. et J. W. V. NESS (1968). "Fractional Brownian motions, fractional noises and applications". In : *SIAM Review* 10, p. 422-437.
- MARRELEC, G. et H. BENALI (2006). "Asymptotic bayesian structure learning using graph supports for gaussian graphical models". In : *Journal of multivariate analysis* 97.6, p. 1451-1466.
- MAXIM, V., L. ŞENDUR, M. J. FADILI, J. SUCKLING, R. GOULD, R. HOWARD et E. T. BULLMORE (2005). "Fractional Gaussian noise, functional MRI and Alzheimer's disease". In : *NeuroImage* 25, p. 141-158.
- MESKALDJI, D., M. OTTET, L. CAMMOUN, P. HAGMANN, R. MEULI, S. E. J. THIRAN et S. MORGENTHALER (2011). "Adaptive strategy for the statistical analysis of connectomes". In : *PLoS One* 6.8, e23009. DOI : [10.1371/journal.pone.0023009](https://doi.org/10.1371/journal.pone.0023009).
- MEUNIER, D., S. ACHARD, A. MORCOM et E. BULLMORE (2009). "Age-related changes in modular organization of human brain functional networks." In : *Neuroimage* 44.3, p. 715-23. DOI : [10.1016/j.neuroimage.2008.09.062](https://doi.org/10.1016/j.neuroimage.2008.09.062).
- MOULINES, E., F. ROUEFF et M. S. TAQU (2007). "On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter". In : *Journal of Time Series Analysis* 28.2, p. 155-187.
- (2008). "A wavelet whittle estimator of the memory parameter of a nonstationary gaussian time series". In : *The Annals of Statistics* 36.4, p. 1925-1956.
- MUIRHEAD, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley et Sons.
- MUTCH, W. A. C., D. M. MANDELL, J. A. FISHER, D. J. MIKULIS, A. P. CRAWLEY, O. PUCCI et J. DUFFIN (2012). "Approaches to brain stress testing : bold magnetic resonance imaging with computer-controlled delivery of carbon dioxide." In : *PLoS One* 7.11, e47443. DOI : [10.1371/journal.pone.0047443](https://doi.org/10.1371/journal.pone.0047443).
- NIELSEN, F. S. (2011). "Local whittle estimation of multi-variate fractionally integrated processes". In : *Journal of Time Series Analysis* 32.3, p. 317-335.
- PEARL, J. (2000). *Causality: models, reasoning and inference*. T. 29. Cambridge Univ Press.
- PEDREGOSA, F., G. VAROQUAUX, A. GRAMFORT, V. MICHEL, B. THIRION, O. GRISEL, M. BLONDEL, P. PRETTEHOFER, R. WEISS, V. DUBOURG, J. VANDERPLAS, A. PASSOS, D. COURNAPEAU, M. BRUCHER, M. PERROT et E. DUCHESNAY (2011). "Scikit-learn : machine learning in Python". In : *Journal of Machine Learning Research* 12, p. 2825-2830.
- PERCIVAL, D. B. et A. T. WALDEN (2000). *Wavelet methods for time series analysis*. Cambridge, UK : Cambridge University Press.
- PERCIVAL, D. B. (1995). "On estimation of the wavelet variance". In : *Biometrika* 82.3, p. 619-631.
- RICHIARDI, J., S. ACHARD, H. BUNKE et D. VAN DE VILLE (2013). "Machine learning with brain graphs : predictive modeling approaches for functional imaging in systems neuroscience". In : *IEEE Signal Process. Mag.* 30.3, p. 58-70.
- RICHIARDI, J., S. ACHARD, E. BULLMORE et D. VAN DE VILLE (2011). "Classifying connectivity graphs using graph and vertex attributes". In : *Proc. IEEE International Workshop on Pattern Recognition in NeuroImaging*. Seoul, Korea, p. 45-48.
- RIDLEY, B. G. Y., C. ROUSSEAU, J. WIRSICH, A. LE TROTIER, E. SOULIER, S. CONFORT-GOUNY, F. BARTOLOMEI, J.-P. RANJEVA, S. ACHARD et M. GUYE (sept. 2015). "Nodal approach reveals differential impact of lateralized focal epilepsies on hub reorganization." In : *Neuroimage* 118, p. 39-48. DOI : [10.1016/j.neuroimage.2015.05.096](https://doi.org/10.1016/j.neuroimage.2015.05.096).

- ROBINSON, E. C., A. HAMMERS, A. ERICSSON, A. D. EDWARDS et D. RUECKERT (avr. 2010). “Identifying population differences in whole-brain structural networks : a machine learning approach”. In : *NeuroImage* 50.3, p. 910-919.
- ROBINSON, P. M. (1994). “Semiparametric analysis of long-memory time series”. In : *The Annals of Statistics*, p. 515-539.
- (1995a). “Gaussian semiparametric estimation of long range dependence”. In : *The Annals of Statistics* 23.5, p. 1630-1661.
- (1995b). “Log-periodogram regression of time series with long range dependence”. In : *The Annals of Statistics*, p. 1048-1072.
- (2005). “Robust covariance matrix estimation : hac estimates with long memory/antipersistence correction”. In : *Econometric Theory* 21.01, p. 171-180.
- ROQUAIN, E. (2010). “Type i error rate control for testing many hypotheses : a survey with proofs”. In : *Journal de la Société Française de Statistique* 152.2.
- ROVERATO, A. et J. WHITTAKER (1998). “The issleris matrix and its application to non-decomposable graphical Gaussian models”. In : *Biometrika* 85.3, p. 711-725.
- RUBINOV, M. et O. SPORNS (sept. 2010). “Complex network measures of brain connectivity : uses and interpretations”. In : *NeuroImage* 52.3, p. 1059-1069. ISSN : 1095-9572. DOI : [10.1016/j.neuroimage.2009.10.003](https://doi.org/10.1016/j.neuroimage.2009.10.003).
- SHIMOTSU, K. (2007). “Gaussian semiparametric estimation of multivariate fractionally integrated processes”. In : *Journal of Econometrics* 137.2, p. 277-310. DOI : [10.1016/j.jeconom.2006.01.003](https://doi.org/10.1016/j.jeconom.2006.01.003).
- SUCKLING, J., A. WINK, F. BERNARD, A. BARNES et E. BULLMORE (sept. 2008). “Endogenous multifractal brain dynamics are modulated by age, cholinergic blockade and cognitive performance.” In : *J Neurosci Methods* 174.2, p. 292-300. DOI : [10.1016/j.jneumeth.2008.06.037](https://doi.org/10.1016/j.jneumeth.2008.06.037).
- TENENBAUM, J. B., V. de SILVA et J. C. LANGFORD (déc. 2000). “A global geometric framework for nonlinear dimensionality reduction”. In : *Science* 290.5500, p. 2319-2323.
- TORRENCE, C. et G. P. COMPO (1998). “A practical guide to wavelet analysis”. In : *Bulletin of the American Meteorological society* 79.1, p. 61-78.
- VELASCO, C. et P. M. ROBINSON (2000). “Whittle pseudo-maximum likelihood estimation for nonstationary time series”. In : *Journal of the American Statistical Association* 95.452, p. 1229-1243.
- VÉRTES, P., A. ALEXANDER-BLOCH, N. GOGTAY, J. GIEDD, J. RAPOPORT et E. BULLMORE (2012). “Simple models of human brain functional networks”. In : *Proc. Natl. Acad. Sci. (USA)* 109.15, p. 5868-73. DOI : [10.1073/pnas.1111738109](https://doi.org/10.1073/pnas.1111738109).
- WANG, H. (2012). “Bayesian graphical lasso models and efficient posterior computation”. In : *Bayesian Analysis* 7.4, p. 867-886.
- WANG, L., C. ZHU, Y. HE, Y. ZANG, Q. CAO, H. ZHANG, Q. ZHONG et Y. WANG (fév. 2009). “Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder.” In : *Hum Brain Mapp* 30.2, p. 638-649. DOI : [10.1002/hbm.20530](https://doi.org/10.1002/hbm.20530).
- WECKER, W. E. (1978). “A note on the time series which is the product of two stationary time series”. In : *Stochastic Processes and their Applications* 8.2, p. 153-157.
- WHITCHER, B. (1998). “Assessing nonstationary time series using wavelets”. Thèse de doct. Univ. of Wash. Press, Seattle.
- WHITCHER, B., P. GUTTORP et D. B. PERCIVAL (2000). “Wavelet analysis of covariance with application to atmospheric time series”. In : *Journal of Geophysical Research* 105(D11).14, p. 941-962.
- WHITTAKER, J. (1990). *Graphical models in applied multivariate statistics*. Wiley.
- WINK, A.-M., F. BERNARD, R. SALVADOR, E. T. BULLMORE et J. SUCKLING (2006). “Age and cholinergic effects on hemodynamics and functional coherence of human hippocampus”. In : *Neurobiology of Aging* 27, p. 1395-1404.
- YAGLOM, A. (1958). “Correlation theory of processes with random stationary nth increments”. In : *American Mathematical Society Translations (Series 2)* 8, p. 87-141.
- YAJIMA, Y. (1985). “On estimation of long-memory time series models”. In : *Australian journal of Statistics* 27.3, p. 303-320.

A Annexe : texte complet des publications mentionnées dans le manuscrit

A.1 (Achard et al. 2006)

Pour télécharger l'article : <http://www.jneurosci.org/content/26/1/63.short>

Summary

Small-world properties have been demonstrated for many complex networks. Here, we applied the discrete wavelet transform to functional magnetic resonance imaging (fMRI) time series, acquired from healthy volunteers in the resting state, to estimate frequency-dependent correlation matrices characterizing functional connectivity between 90 cortical and subcortical regions. After thresholding the wavelet correlation matrices to create undirected graphs of brain functional networks, we found a small-world topology of sparse

connections most salient in the low-frequency interval 0.03–0.06 Hz. Global mean path length (2.49) was approximately equivalent to a comparable random network, whereas clustering (0.53) was two times greater; similar parameters have been reported for the network of anatomical connections in the macaque cortex. The human functional network was dominated by a neocortical core of highly connected hubs and had an exponentially truncated power law degree distribution. Hubs included recently evolved regions of the heteromodal association cortex, with long-distance connections to other regions, and more cliquishly connected regions of the unimodal association and primary cortices; paralimbic and limbic regions were topologically more peripheral. The network was more resilient to targeted attack on its hubs than a comparable scale-free network, but about equally resilient to random error. We conclude that correlated, low-frequency oscillations in human fMRI data have a small-world architecture that probably reflects underlying anatomical connectivity of the cortex. Because the major hubs of this network are critical for cognition, its slow dynamics could provide a physiological substrate for segregated and distributed information processing.

A.2 (Achard et al. 2008)

Pour télécharger l'article : <http://journals.aps.org/pre/abstract/10.1103/PhysRevE.77.036104>

Summary

Using the multivariate long memory (LM) model and Taylor expansions, we find the conditions for convergence of the wavelet correlations between two LM processes on an asymptotic value at low frequencies. These mathematical results, and a least squares estimator of LM parameters, are validated in simulations and applied to neurophysiological (human brain) and financial market time series. Both brain and market systems had multivariate LM properties including a “fractal connectivity” regime of scales over which wavelet correlations were invariantly close to their asymptotic value. This analysis provides efficient and unbiased estimation of long-term correlations in diverse dynamic networks.

A.3 (Achard et al. 2012)

Pour télécharger l'article : <http://www.pnas.org/content/109/50/20608.abstract>

Summary

Human brain networks have topological properties in common with many other complex systems, prompting the following question : what aspects of brain network organization are critical for distinctive functional properties of the brain, such as consciousness ? To address this question, we used graph theoretical methods to explore brain network topology in resting state functional MRI data acquired from 17 patients with severely impaired consciousness and 20 healthy volunteers. We found that many global network properties were conserved in comatose patients. Specifically, there was no significant abnormality of global efficiency, clustering, small-worldness, modularity, or degree distribution in the patient group. However, in every patient, we found evidence for a radical reorganization of high degree or highly efficient “hub” nodes. Cortical regions that were hubs of healthy brain networks had typically become nonhubs of comatose brain networks and vice versa. These results indicate that global topological properties of complex brain networks

may be homeostatically conserved under extremely different clinical conditions and that consciousness likely depends on the anatomical location of hub nodes in human brain networks.

A.4 (Richiardi et al. 2013)

Pour télécharger l'article : <http://dx.doi.org/10.1109/MSP.2012.2233865>

Summary

The observation and description of the living brain has attracted a lot of research over the past centuries. Many noninvasive imaging modalities have been developed, such as topographical techniques based on the electromagnetic field potential [i.e., electroencephalography (EEG) and magnetoencephalography (MEG)], and tomography approaches including positron emission tomography and magnetic resonance imaging (MRI). Here we will focus on functional MRI (fMRI) since it is widely deployed for clinical and cognitive neurosciences today, and it can reveal brain function due to neurovascular coupling (see “From Brain Images to fMRI Time Series”). It has led to a much better understanding of brain function, including the description of brain areas with very specialized functions such as face recognition. These neuroscientific insights have been made possible by important methodological advances in MR physics, signal processing, and mathematical modeling.

A.5 (Fallani et al. 2014)

Pour télécharger l'article : <http://dx.doi.org/10.1098/rstb.2013.0521>

Summary

The brain can be regarded as a network : a connected system where nodes, or units, represent different specialized regions and links, or connections, represent communication pathways. From a functional perspective, communication is coded by temporal dependence between the activities of different brain areas. In the last decade, the abstract representation of the brain as a graph has allowed to visualize functional brain networks and describe their non-trivial topological properties in a compact and objective way. Nowadays, the use of graph analysis in translational neuroscience has become essential to quantify brain dysfunctions in terms of aberrant reconfiguration of functional brain networks. Despite its evident impact, graph analysis of functional brain networks is not a simple toolbox that can be blindly applied to brain signals. On the one hand, it requires the know-how of all the methodological steps of the pipeline that manipulate the input brain signals and extract the functional network properties. On the other hand, knowledge of the neural phenomenon under study is required to perform physiologically relevant analysis. The aim of this review is to provide practical indications to make sense of brain network analysis and contrast counterproductive attitudes.

A.6 (Costard et al. 2014)

Pour télécharger l'article : <http://dx.doi.org/10.1109/ICOSP.2014.7015219>

Summary

This paper focuses on estimated Gaussian Graphical Models (GGM) from sets of experimental data. Some extension of known Bayesian methods are proposed, allowing to introduce score functions to measure the relevance of the obtained GGM structure to describe the data. These score functions form the basic measurement to derive a new dissimilarity matrix based on the GGM structure. This latter is then exploited for classification purpose. Examples are provided using both simulated and real experimental functional Magnetic Resonance Imaging (fMRI) data.

A.7 (Harlé et al. 2014), (Harlé et al. 2015)

Pour télécharger l'article : http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6952467

Summary

In this paper, we propose a Bayesian approach for multivariate time series segmentation. A robust non-parametric test, based on rank statistics, is derived in a Bayesian framework to yield robust distribution-independent segmentations of piecewise constant multivariate time series for which mutual dependencies are unknown. By modelling rank-test p-values, a pseudo-likelihood is proposed to favour change-points detection for significant p-values. A vague prior is chosen for dependency structure between time series, and a MCMC method is applied to the resulting posterior distribution. The Gibbs sampling strategy makes the method computationally efficient. The algorithm is illustrated on simulated and real signals in two practical settings. It is demonstrated that change-points are robustly detected and localized, through implicit dependency structure learning or explicit structural prior introduction.

A.8 (Achard et Gannaz 2016)

Pour télécharger l'article : <http://onlinelibrary.wiley.com/doi/10.1111/jtsa.12170/abstract>

Summary

Multivariate processes with long-range dependent properties are found in a large number of applications including finance, geophysics and neuroscience. For real-data applications, the correlation between time series is crucial. Usual estimations of correlation can be highly biased owing to phase shifts caused by the differences in the properties of autocorrelation in the processes. To address this issue, we introduce a semiparametric estimation of multivariate long-range dependent processes. The parameters of interest in the model are the vector of the long-range dependence parameters and the long-run covariance matrix, also called functional connectivity in neuroscience. This matrix characterizes coupling between time series. The proposed multivariate wavelet-based Whittle estimation is shown to be consistent for the estimation of both the long-range dependence and the covariance matrix and to encompass both stationary and nonstationary processes. A simulation study and a real-data example are presented to illustrate the finite-sample behaviour.