Structure recovery from time dependent data

Irène Gannaz

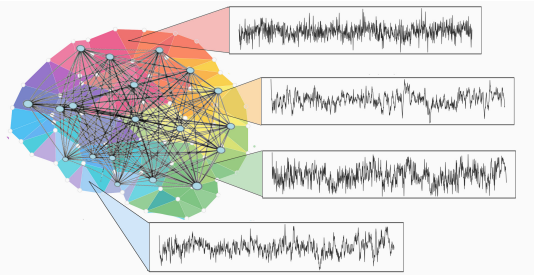
INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

UNIVERSITÉ DE LYON

Lyon 1

École doctorale INFOMATHS – ED 512

Structure recovery from time dependent data



HABILITATION À DIRIGER DE RECHERCHES

En mathématiques appliquées

par

Irène GANNAZ

– Maître de conférences à Grenoble INP –

Soutenue le 27 janvier 2023 devant le jury composé de

M. Jean-Marc Bardet	Université Paris 1 (SAMM)	Rapporteur
M. Clément Marteau	Université Lyon 1 (ICJ)	
Mme Voichiţa Maxim	INSA Lyon (Creatis)	
Mme Anne Philippe	Université de Nantes (LMJL)	Rapportrice
M. Dimitri Van de Ville	University of Geneva (Medical Image Processing) EPFL (Institute of Bioengineering)	Rapporteur
M. Olivier Wintenberger	Sorbonne Université (LPSM)	

Math

Institut Camille Jordan
UMR 5208

G-SCOP
UMR 5272

G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

Contents

Introduction

J'ai choisi de faire l'introduction en français et le reste du manuscrit en anglais. J'ai hésité à tout écrire dans ma langue natale, ce qui aurait amélioré la fluidité de lecture et diminué les imprécisions de langage. Mon choix a finalement été motivé par la facilité de transcription des articles rédigés en anglais.

Ma recherche comporte principalement des développements motivés par des problématiques d'extraction d'information de données réelles. Les travaux [2] sont issus d'une collaboration industrielle et l'article [15] concerne l'analyse de la réponse cérébrale dans le domaine de l'audition et les travaux [1] [4] [5] [6] [10] [28] et [12] [13] [14] [23] [24] ont été motivés par l'analyse de données d'IRM fonctionnelle ou de MEG. Afin de répondre aux problèmes posés, j'ai développé des modélisations théoriques, avec le développement de méthodologies spécifiques, l'étude de leur validité et leur implémentation. Par exemple, une problématique en neurosciences est l'inférence des graphes de connectivité cérébrale. Souhaitant mener à bien une recherche de qualité, je me suis intéressée à toutes les étapes de cette inférence afin de mener des données au résultat final, explorant théorie et application. Ainsi, j'ai proposé un modèle de séries temporelles avec une modélisation de la structure adaptée aux données, développé une procédure d'estimation de ce modèle et étudié des procédures de tests multiples pour en déduire un graphe de dépendance. Les domaines abordés sont donc variés : séries temporelles, théorie des ondelettes, matrices de covariance et régularisation en grande dimension, tests multiples, théorie des graphes… l'objectif étant de développer les méthodologies adaptées aux données. Ces développements sont également associés à des implémentations dans des packages.

Le dénominateur commun à une grande majorité de mes travaux est la nature temporelle des données, avec une dépendance entre les observations successives. Une modélisation sous la forme de données fonctionnelles ou de séries chronologiques est réalisée. Une représentation des données dans une base est ensuite effectuée, dans des bases de splines ou d'ondelettes par exemple. Je précise que je n'ai pas renié mon intérêt pour les ondelettes depuis leur découverte en master, et qu'ainsi beaucoup de mes travaux reposent sur une représentation en ondelettes.

Ce qui se justifie aussi par leurs bonnes propriétés pour les applications. J'identifie ensuite deux grandes catégories :

- la première consiste à prendre en compte la structure de dépendance des données lorsque celles-ci sont utilisées dans des modèles (modèles de régression, de classification supervisée ou non supervisée, etc),
- la seconde consiste à modéliser la structure de dépendance des données et à l'estimer.

Cette distinction m'a ainsi amené à rédiger ce manuscrit en deux parties.

La première partie concerne l'inférence à partir de données fonctionnelles. J'ai étudié trois modèles usuels d'inférence statistiques que sont l'estimation de densité, la régression et la classification. Dans chaque cas l'objectif était de proposer une procédure consistante avec des données présentant des dépendances.

Ainsi, je me suis intéressée à l'estimation de densité avec des données dépendantes [8], avec Olivier Wintenberger (LPSM, Univ. Sorbonne). Nous avons considéré des dépendances faibles, qui englobent un large spectre de dépendances et montré que l'estimation par ondelettes usuelle restait consistante dans ce contexte.

Durant ma thèse [29] j'ai étudié un modèle de régression dit partiellement linéaire. L'objectif est de décomposer des données observées en deux parties, la première mesurant l'influence (linéaire) de covariables et la seconde modélisant les données sous forme fonctionnelle. J'ai développé une estimation par ondelettes dans ce modèle [9]. Un parallèle a été établi avec une estimation robuste usuelle du paramètre de régression linéaire dans laquelle les coefficients d'ondelettes de la partie fonctionnelle sont considérés comme des valeurs aberrantes. J'ai également étendu cette procédure à un cadre fonctionnel généralisé dans [7]. Dans ces deux cadres, j'ai montré les propriétés théoriques des estimateurs, montrant que les parties linéaires et fonctionnelles peuvent être estimées à des vitesses quasi-optimales au sens minimax, indépendamment de la présence de l'autre partie.

Enfin, un axe de recherche de cette partie est la classification (supervisée ou non supervisée) de données fonctionnelles. Dans [15], j'ai proposé une procédure de classification de signaux d'EEG dans le domaine de l'audition, dans le cadre d'une collaboration avec Rafael Laboissière (LPNC, CNRS, Univ. Grenoble Alpes – alors CRNL, Inserm, CNRS, Univ. Lyon 1). La procédure reposait sur un modèle logistique avec une représentation par ondelettes des signaux et utilisait une double réduction de dimension (par pénalisation $\ell 1$ et analyse en composantes principales). J'ai aussi encadré un stage de Master 1 sur l'ANOVA fonctionnelle appliquée à des données de reconstruction osseuse en 2017, dans une problématique issue de discussions avec Segolen Geffray (IRMA, Univ. Strasbourg). Dans le cadre d'une collaboration industrielle

avec Julien Jacques (ERIC, Univ. Lyon 2) nous avons réalisé un projet de détection de dysfonctionnements d'un système multi-capteurs dans un contexte industriel [25]. Le projet initial se prolonge actuellement par la thèse CIFRE de Martial Amovin-Assagba, que je co-encadre. Dans un premier temps, nous avons considéré une approche basée sur la segmentation des données capteurs [19]. Au vu de l'hétérogénéité des données, nous avons ensuite développé un modèle de mélange fonctionnel avec prise en compte de courbes aberrantes dans le modèle [2]. Notre modèle permet conjointement de caractériser les différents régimes des capteurs et d'identifier les courbes issues d'un dysfonctionnement de façon automatisée.

La deuxième partie concerne l'estimation de structures spatio-temporelles de dépendance pour des séries temporelles multivariées (longue-mémoire, corrélation à long terme). L'objectif est de modéliser les dépendances entre différents points d'observation sous forme d'un graphe. Les nœuds du graphe correspondent soit aux capteurs (MEG-EEG) soit aux régions cérébrales (IRMf). Comme illustré en figure I.1, cette modélisation est faite en deux temps : premièrement une estimation d'une matrice de corrélation entre les différents signaux mesurés et ensuite une procédure de tests multiples permettant d'associer un graphe à cette matrice.

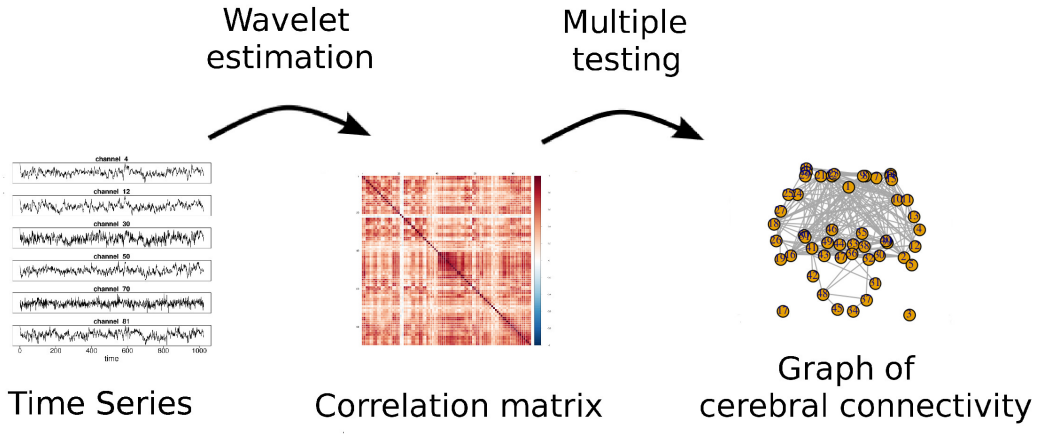


Figure I.1: Schéma d'estimation des graphes de connectivité cérébrale.

Les travaux [5] [6] [10] [13] [14] estiment une matrice de corrélation entre les mesures temporelles des différents capteurs/régions cérébrales. La spécificité de notre approche est de prendre en compte la structure temporelle des données (longue-mémoire) afin de contrôler l'estimation de cette structure inter-capteurs. En effet dans [6] nous avons mis en évidence un phénomène de phase induit par la présence de longue-mémoire dans les signaux. La procédure d'estimation construite dans [6] repose sur des ondelettes. Les aspects computationnels de la méthode sont présentés dans [5], avec un code `R` fourni dans [24]. [13] propose un exem-

ple d'application sur données réelles, avec mise en évidence de l'intérêt de la prise en compte de l'aspect multivarié des mesures. Au vu des limitations du modèle utilisé sur les données réelles, des ondelettes analytiques sont introduites dans l'estimation, [14]. Les propriétés des ondelettes utilisées sont étudiées dans [4]. Ces propriétés permettent d'obtenir des résultats de consistance pour l'approche choisie, [10]. Cette étape d'estimation de la matrice de corrélation a été réalisée en collaboration avec Sophie Achard (LJK, CNRS, Univ. Grenoble Alpes), Marianne Clausel (IECL, Université de Lorraine) et François Roueff (LTCI, Télécom-Paristech).

Ensuite une procédure de tests multiples permet d'en déduire le graphe associé. Une arête est présente entre deux nœuds si la corrélation entre les mesures correspondantes est significative. Dans les problématiques de connectivité cérébrale, plus de 1000 tests simultanés sont nécessaires, d'où l'importance d'une procédure de tests multiples. Ce projet a constitué notamment le sujet de doctorat de Marine Roux que j'ai co-encadré (soutenu en septembre 2018). L'étude des tests statistiques applicables dans ce contexte est réalisée dans [1]. L'article [28] présente diverses corrections de tests de corrélation, avec un package `R` associé, [23]. [12] décrit une application des tests multiples dans le domaine des séries temporelles, en illustrant les conséquences d'une mauvaise inférence sur une structure de graphe. Ces travaux ont été réalisés dans le cadre de collaborations avec Sophie Achard (LJK, CNRS, Univ. Grenoble Alpes), Pierre Borgnat (Laboratoire de Physique, CNRS, ENS de Lyon) et Etienne Roquain (LPMA, Univ. Paris 6).

En parallèle de ces thématiques, j'ai réalisé d'autres travaux qui ne seront pas présentés ici. Principalement, j'ai étudié l'utilisation de fonctions de profondeur dans les procédures de choix social [3], en collaboration avec Jean-Baptiste Aubin (ICJ, INSA Lyon), Samuela Leoni (ICJ, INSA Lyon) et Antoine Rolland (ERIC, Univ. Lyon 2). J'ai également proposé une procédure d'estimation de bruit dans un cadre Gaussien généralisé avec Fabien Millioz (CREATIS, Univ. Lyon 1) [21]. En collaboration avec Fabien Millioz, dans le cadre d'une collaboration industrielle, je me suis aussi intéressée à l'extraction de caractéristiques de signaux cardiaques et à l'aide au diagnostique de pathologies cardiaques, pour l'entreprise Cardiags, en 2016 et 2017 [26] [27].

Le manuscrit reprend les deux principales parties décrites ci-dessus. Ces travaux étant en grande partie motivés par des analyses de données réelles, chaque chapitre (excepté le premier) sera associé à un(des) jeu(x) de données réelles permettant d'illustrer la procédure développée. Ce jeu de données sera, le cas échéant, décrit au début de section ou de chapitre. De plus, les principales contributions et les publications associées seront rappelées en début de chapitre. Les publications, qui seront alors données entre crochets, font référence à mes publications (page 117). Des perspectives sont données en conclusion du manuscrit.

Part I

Dealing with dependent observations

Introduction

This part deals with classical statistical models such as density estimation, linear regression and (supervised and unsupervised) classification. The specificity of my work in these settings is the presence of chronological (or temporal) observations. Observations are not independent, and it has to be taken into account in the modeling.

First, in Chapter 1, I consider one recording, from which I want to make inference. Observations $(X_i)_{i=1,\dots,n}$ are reals which are not independent. Typically, X_i is the observation of a characteristic at time t_i. I want to extract from them some features, not linked with the dependence structure, but for which estimation may be modified by the presence of this dependence.

Two classical statistical models are studied: density estimation and regression. The first setting deals with a weak dependence modeling, while the regression handles a functional representation of the observations.

Second, Chapter 2 considers as an observation $(X_i)_{i=1,\dots,n}$ a sequence of recordings. Each observation X_i here is a signal, or a curve, $X_i = \{X_i(t_j),\ j = 1, \dots, N_i\}$, $N_i > 0$. The objective is to perform a (supervised or unsupervised) classification of these functional observations. A modeling is necessary to represent the data in a space where it will be possible to discriminate them.

CHAPTER 1

DENSITY ESTIMATION AND REGRESSION WITH DEPENDENT OBSERVATIONS

Motivation

Classical statistical models such as density estimation and regression often suppose that the observations are observed independently. Dealing with time recordings introduces a new paradigm. It is necessary to take into account the dependence structure, to conduct a consistent inference. This dependence structure can be modeled by a time series approach, or by a functional approach.

Contributions

Gannaz and Wintenberger (2010) [8] deal with a wavelet-based estimation of the probability distribution of the observations when they are dependent. The dependence is modeled using the notion of weak-dependence. We propose a non parametric estimation under our setting of dependence, and show the consistency of our approach.

During my PhD (Gannaz 2007a [29]), I have studied partial linear models. The idea is to infer the (linear) influence of covariates on a temporal recording, and to recover the remaining signal, which is nonparametric. The approach here to model the dependence is the functional modeling, rather than a time series setting. I proposed a wavelet-based procedure. Gannaz (2007b) [9] establishes a link between the thresholding of the wavelet coefficients of the nonparametric part and a robust estimation of the linear part. The consistency of the estimation is also proved. Next, Gannaz (2013) [7] extends the procedure to generalized partial linear models.

The sections of the chapter deal respectively with the density estimation studied in Gannaz and Wintenberger (2010) and with the results on partial linear regression obtained in Gannaz (2007a), Gannaz (2007b), and Gannaz (2013).

1.1 Wavelet decomposition – Notations

Let on $L^2(A)$ be the space of squared-integrable functions on A compact subset, $A \subseteq \mathbb{R}$. Typically, $A = [0,1]$. I will consider the usual inner product, which is defined as $\forall f, g \in L^2(A)$, $< f, g >= \int_{A]} f(t)g(t)\,dt$. Let us consider an α-regular ($\alpha \geqslant 0$) orthonormal multiresolution analysis of $\left(L^2(A), < \cdot, \cdot >\right)$, associated with a compactly supported scaling function ϕ and a compactly supported mother wavelet ψ. For a given primary resolution level j_0, the functions $\{\phi_{j,k} : x \mapsto 2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}}$ and $\{\psi_{j,k} : x \mapsto 2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}$ are such that the family

$$\{\phi_{j_0,k},\, k \in \mathbb{Z},\, \psi_{j,k}\, j \geqslant j_0,\, k \in \mathbb{Z}\}$$

is an orthonormal base of $L^2(A)$. For any $f \in L^2(A)$, denote by $c_{j_0k} =< f, \phi_{j_0k} >$ ($k \in \mathbb{Z}$) the scaling coefficients and by $d_{jk} =< f, \psi_{jk} >$ ($j \geqslant j_0;\ k \in \mathbb{Z}$) the wavelet coefficients of f for the orthonormal periodic wavelet basis defined above; the function $f \in L^2(A)$ can be decomposed as

$$f(t) = \sum_{k\in\mathbb{Z}} c_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(t), \quad t \in [0,1].$$

For simplicity in exposition, we work with periodic wavelet bases (see, e.g., Mallat (1999), Section 7.5.1). Then, for any $f \in L^2([0,1])$, the function is expressed in the form

$$f(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(t), \quad t \in [0,1].$$

The number of wavelet coefficients at each scale $j \geqslant 0$ is equal to 2^j, due to the use of periodic wavelet bases.

Denote by the exponent $\top$ the transpose operator. Let $\boldsymbol{f} = (f(t_1), \ldots, f(t_n))^\top$ be a vector of function values at equally spaced points t_i. The discrete wavelet transform (DWT) of $\boldsymbol{f}$ is given by $\boldsymbol{\theta} = W_{n\times n}\boldsymbol{f}$, where $\boldsymbol{\theta}$ is an $n \times 1$ vector comprising both discrete scaling coefficients, s_{j_0k}, and discrete wavelet coefficients, w_{jk}. $W_{n\times n}$ is an orthogonal $n \times n$ matrix associated with the orthonormal periodic wavelet basis chosen. The corresponding empirical coefficients s_{j_0k} and w_{jk} are related to their continuous counterparts c_{j_0k} and d_{jk} with a factor $n^{-1/2}$.

I will assume hereafter that f belongs to a (inhomogeneous) Besov spaces on the unit interval, $\mathcal{B}^s_{\pi,r}([0,1])$, with $s + 1/\pi - 1/2 > 0$. The last condition ensures in particular that evaluation of f at a given point makes sense. For a detailed study on (inhomogeneous) Besov spaces I refer to, *e.g.*, Donoho and Johnstone (1998). For $M_1 > 0$, the space $\mathcal{B}^s_{\pi,r}(A, M_1)$ denotes the set of

functions f defined on A such that $\|f\|_{s,\pi,r} \leq M_1$ where

$$\|f\|_{s,\pi,r} = |c_{0,0}| + \left(\sum_{j \in \mathbb{N}} \left(2^{j(s\pi+\pi/2-1)} \sum_{k \in \mathbb{Z}} |d_{j,k}|^\pi \right)^{r/\pi} \right)^{1/r}.$$

Remark. In this part, the convention used is that the frequencies increase when the scales j increase. In Part 2 of this manuscript, the inverse convention will be used.

1.2 Density estimation under weak-dependence

Let $(X_t)_{t \in \mathbb{Z}}$ be a real valued time series admitting a common marginal density f that is compactly supported on a set A. The general purpose of this work is to estimate f by wavelet estimators $\widehat{f_n}$ constructed from n observations $(X_1, \ldots, X_n)$. Nonlinear wavelet-based estimation provides estimators which adapt themselves to the unknown smoothness of f. In Gannaz and Wintenberger (2010), we have extended near minimax results of soft and hard-threshold estimators, obtained in Theorem 5 of Donoho, Johnstone, Kerkyacharian, and Picard (1996) in the independent framework, to time series observations. We have modeled the dependence with the notion of weak dependence.

Let us use the wavelet decomposition described in Section 1.1. Let B denote the compact support of the function ψ. The nonlinear estimator developed in Donoho, Johnstone, Kerkyacharian, and Picard (1996) is defined by the equation

$$\widehat{f_n} = \sum_{k \in \mathbb{Z}} \widehat{c}_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \gamma_{\lambda_j}(\widehat{d}_{j,k}) \psi_{j,k},$$

where $\widehat{c}_{j,k} = n^{-1} \sum_{i=1}^n \phi_{j,k}(X_i)$ and $\widehat{d}_{j,k} = n^{-1} \sum_{i=1}^n \psi_{j,k}(X_i)$, and where γ_λ is a threshold function of level λ. Both hard and soft thresholding functions are studied, corresponding respectively to $\gamma_\lambda(d) = d\, \mathbb{1}_{|d|>\lambda}$ and $\gamma_\lambda(d) = (|d| - \lambda)_+ \text{sign}(d)$.

1.2.1 The weak-dependence setting

The specificity of the setting in Gannaz and Wintenberger (2010) is to consider dependent observations. We have chosen to formulate the dependence assumption on the wavelet coefficient rather than directly on the observations, to set a general framework. I first give the main assumption, and, next, I illustrate it with some examples.

The symbol δ denotes with no distinction ϕ or ψ and $\widetilde{\delta}_{j,k}(x) = \delta_{j,k}(x) - \mathbb{E}\,\delta_{j,k}(X_0)$ for all integers $j \geq 0$ and k. Define for all positive integers u, v, for all $\Delta > 0$, the quantities

$$C^{j,k}_{u,v}(\Delta) = \sup_{\max s_{i+1} - s_i = s_{u+1} - s_u = \Delta} \{|\mathrm{Cov}\left(\widetilde{\delta}_{j,k}(X_{s_1}) \cdots \widetilde{\delta}_{j,k}(X_{s_u}), \widetilde{\delta}_{j,k}(X_{s_{u+1}}) \cdots \widetilde{\delta}_{j,k}(X_{s_{u+v}})\right)|\}.$$

Functions ϕ and ψ play a symmetric role through δ in this setting. As stressed in Doukhan and Louhichi (1999), bounds on covariance terms $C^{j,k}_{u,v}(\Delta)$ are useful to extend asymptotic results from the independent case.

The main dependence assumption of this work is the following:

(D) There exists a sequence $\rho(\Delta)$ such that for all $\Delta \geq 0$, all indexes j, k, u, v, we have

$$C^{j,k}_{u,v}(\Delta) \leq (u + v + uv)/2\,(2^{j/2} M_2)^{u+v-2} \rho(\Delta),$$

where M_2 is a constant satisfying $\|\delta\|_\infty \leq M_2$. Moreover, there exist real numbers $a, b, C_0 > 0$ depending only on δ, f and on the dependence properties of $(X_t)_{t\in\mathbb{Z}}$ such that

$$\rho(\Delta) \leq C_0 \exp(-a\Delta^b) \text{ for all } \Delta > 0.$$

This condition states that the dependence between the past and the future values of the process $(X_t)_{t\in\mathbb{Z}}$ is controlled when the gap Δ between past and future goes to infinity, as illustrated in Figure 1.1. I give hereafter some examples of processes satisfying **(D)**, based on $\widetilde{\varphi}$- and λ-weak dependence.

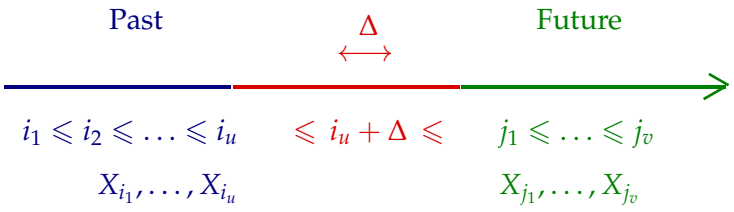


Figure 1.1: Illustration of the division of the observations used to define weak-dependence.

Example 1

The following example is based on $\widetilde{\varphi}$-weak dependence (Dedecker and Prieur 2007). I choose not to recall the definition of $\widetilde{\varphi}$-weak dependence here for simplicity. I refer to Gannaz and Wintenberger (2010) for a more precise description of the framework.

Let us define stationary Markov chains $(X_t)_{t\in\mathbb{Z}}$ associated with dynamical systems through a reversion of the time as non degenerate stationary solutions of the recurrent equation

$$X_t = T^i(X_{t-i}), \qquad \forall t \in \mathbb{Z}, i \in \mathbb{N},$$

where $T : [0,1] \to [0,1]$ is a deterministic function. A Markov chain $(X_t)_{t\in\mathbb{Z}}$ is associated with an expanding map through a reversion of the time if T satisfies

- *(Regularity)* The function T is differentiable, with a continuous derivative T' and there exists a grid $0 = a_0 \leq a_1 \cdots \leq a_k = 1$ such that $|T'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \ldots, k$.
- *(Expansivity)* For any integer i, let I_i be the set on which the first derivative of T^i, $(T^i)'$, is defined. There exists $a > 0$ and $s > 1$ such that $\inf_{x\in I_i}\{|(T^i)'(x)|\} > as^i$.
- *(Topological mixing)* For any nonempty open sets U, V, there exists $i_0 \geq 1$ such that $T^{-i}(U) \cap V \neq \varnothing$ for all $i \geq i_0$.

Under these three conditions $(X_t)_{t\in\mathbb{N}}$ satisfies Condition **(D)**.

For example, let X_i be obtained by the equation $X_i = F^{-1}(G(Y_i))$ for $i = 1, \ldots, n$ with $G(x) = 2\sqrt{x(1-x)}/\pi$ and $(Y_i)_{i=1,\ldots,n}$ given by $Y_1 = G^{-1}(U_1)$ and, recursively, $Y_i = T^{i-1}(Y_1)$ for $2 \leq i \leq n$ with $T(x) = 4x(1-x)$. The process is $\widetilde{\varphi}$-weakly dependent and satisfies Condition **(D)**. A realization is displayed on Figure 1.2.

Example 2

We establish in Gannaz and Wintenberger (2010) that condition **(D)** is satisfied by the following stationary processes $(X_t)_{t\in\mathbb{Z}}$, which are λ-weakly dependent. I refer to Doukhan and Wintenberger (2007) for a definition of λ-weakly dependence.

- **Infinite moving average**
  A Bernoulli shift is an infinite moving average process defined by

  $$X_t = \sum_{i\in\mathbb{Z}} a_i \xi_{t-i}.$$

  with ξ_t are independent and identically distributed (*i.i.d.*) random variables satisfying $\mathbb{E}|\xi_0| \leq 1$. $(X_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent with $\lambda(\Delta) \leq 4\sum_{|j|>[\Delta/2]} |a_j|$. Suppose that $a_j \leq C_a a_\infty^{|j|}$ for $j \neq 0$, $C_a > 0$ and $0 < a_\infty < 1$.

- **Affine model**
  Let us consider the stationary solution $(X_t)_{t\in\mathbb{Z}}$ of the equation

$$X_t = T(X_{t-1}, X_{t-2}, \ldots)\xi_t + g(X_{t-1}, X_{t-2}, \ldots),$$

  where M and f are both Lipschitz functions. This model contains various time series processes such as ARCH, GARCH, ARMA, ARMA-GARCH, etc. We suppose that the ξ_t are *i.i.d.* random variables with a bounded marginal density, and that the functions T and g have exponentially decreasing Lipschitz coefficients.

For example, Figure 1.2 displays the following λ-weakly dependent process. Let X_i, $i = 1, \ldots, n$, result from the transform $X_i = F^{-1}(G(Y_i))$ of variables $(Y_t)_{t\in\mathbb{Z}}$ which are solution of $Y_t = 2(Y_{t-1} + Y_{t+1})/5 + 5\xi_t/21$, where $(\xi_t)_{t\in\mathbb{Z}}$ is an *i.i.d.* sequence of Bernoulli variables with parameter $1/2$. The stationary solution of this equation admits the representation $Y_t = \sum_{j\in\mathbb{Z}} a_j \xi_{t-j}$, where $a_j = 1/3(1/2)^{|j|}$. The process $(X_t)_{t\in\mathbb{Z}}$ satisfies Condition **(D)**. A realization is displayed on Figure 1.2.

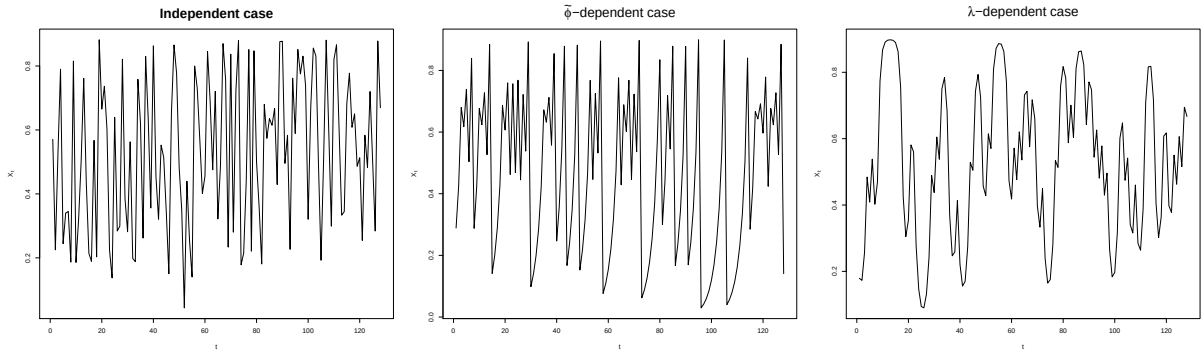


Figure 1.2: Examples of processes satisfying Condition **(D)**. From left to right, are displayed an independent process, a $\widetilde{\varphi}$-weak dependent process and a λ-weakly dependent process. The observations of each process follow the same distribution.

1.2.2 Consistency of the estimation

The minimax rate ν is determined in Donoho, Johnstone, Kerkyacharian, and Picard (1996) as:

$$\nu = \begin{cases} \nu_+ = (s/(1+2s) & \text{if } \varepsilon \geq 0, \\ \nu_- = (s - 1/\pi + 1/p)/(1 + 2s - 2/\pi) & \text{if } \varepsilon \leq 0, \end{cases} \quad \text{where } \varepsilon = s\pi - (p-\pi)/2. \quad (1.1)$$

I now formulate the main result, which is an extension to weak dependence settings of Theorem 5 of Donoho, Johnstone, Kerkyacharian, and Picard (1996).

Theorem 1.1 (Gannaz and Wintenberger 2010)**.** *Suppose that* $f \in \mathcal{B}^s_{\pi,r}(A, M_1)$ *with* $1/\pi < s < \alpha/2$ *where* α *is the regularity of the function* ψ*. If* **(D)** *holds, then for each* $1 \leq p < \infty$ *there exists a positive constant* C*, depending only on* $(\alpha, p, a, b, C_0, M_1, M_2, A, B)$*, such that*

$$\mathbb{E}[\|\widehat{f}_n - f\|_p^p] \leq C \begin{cases} \left(\frac{\log n}{n}\right)^{p\nu} & \text{if } \varepsilon \neq 0 \\ \left(\frac{\log n}{n}\right)^{p\nu} (\log n)^{(p/2-(1\wedge\pi)/r)_+} & \text{if } \varepsilon = 0, \end{cases}$$

where ν *and* ε *are given in* (1.1)*. Here* j_0 *is chosen as the smallest integer larger than* $\log(n)(1+N)^{-1}$*,* j_1 *is the largest integer smaller than* $\log(n \log^{-2/b-3}(n))$ *and* $\lambda_j = K\sqrt{j/n}$ *for a sufficiently large constant* $K > 0$*.*

The estimators $\widehat{f}_n$ are the same as in the independent case, given in Donoho, Johnstone, Kerkyacharian, and Picard (1996), except for the highest resolution level j_1 which is smaller here. This restriction is needed in the weak dependence context due to the Bernstein's type inequality which is not as sharp as the one in the independent case. But this restriction does not perturb the rate of convergence, which is the same as the one obtained by Donoho, Johnstone, Kerkyacharian, and Picard (1996).

The constant K plays a key role in the asymptotic behavior of $\widehat{f}_n$. It deeply depends on the dependence structure of observations $(X_1, \ldots, X_n)$. Contrarily to the independent case, we are not able to develop direct procedures based on the observations $(X_1, \ldots, X_n)$ that chose a convenient parameter K like in Juditsky and Lambert-Lacroix (2004).

Simulation results show that this procedure indeed provides satisfactory results. A cross validation procedure was proposed for application. I do not detail more in this manuscript, and I let the reader refer to Gannaz and Wintenberger (2010).

1.3 Partial linear models

Another usual problem in statistics is regression. During my PhD, I have studied a regression setting with functional data (Gannaz 2007a; Gannaz 2007b). I have next extended the model to generalized regression (Gannaz 2013). The idea here is to model the dependence through a functional representation. The main difference with the previous approach is that the data is supposed smooth. Hence, the functional representation allows a reduction of dimension. The key of the developments is to find a sparse representation.

Suppose that y_i, $i = 1, 2, \ldots, n$, is the i-th response at point t_i, where t is an index such as time or distance. The objective is to extract the influence of given covariates $\boldsymbol{X}_i \in \mathbb{R}^p$, , $i = 1, 2, \ldots, n$.

The partial linear model (PLM) writes as

$$y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + f(t_i) + u_i, \tag{1.2}$$

where $\boldsymbol{X}_i^\top$ are $p \times 1$ vectors, $t_i = \frac{i}{n}$ and $\boldsymbol{\beta}$ and f are respectively the parametric and nonparametric components. I will assume hereafter that the noise variables u_i are *i.i.d.* centered Gaussian variables with an unknown variance σ^2 and that the sample size writes as $n = 2^J$ for some positive integer J.

I have also studied in Gannaz (2007a, Chapter 7) an extension to non equidistant design $\{t_i,\ i = 1, \dots, n\}$, which I do not present here, and which was not published elsewhere.

We will assume that f belongs to a (inhomogeneous) Besov space on the unit interval, $\mathcal{B}^s_{\pi,r}([0,1])$, with $s + 1/\pi - 1/2 > 0$. As previously, we are working with a multiresolution analysis, described in Section 1.1. Applying the wavelet transform $W_{n\times n}$, we can define $\boldsymbol{Z} = W_{n\times n}\left(y_1, \dots, y_n\right)^\top$, $A = W_{n\times n}\left(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\right)$, $\boldsymbol{\theta} = W_{n\times n}\left(f(t_1), \dots f(t_n)\right)^\top$ and $\boldsymbol{\varepsilon} = W_{n\times n}\left(u_1, \dots, u_n\right)^\top$. We obtain the transformed model

$$\boldsymbol{Z} = A\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{1.3}$$

The orthogonality of the DWT matrix $W_{n\times n}$ ensures that the transformed noise vector $\boldsymbol{\varepsilon}$ is distributed as a Gaussian white noise with variance $\sigma^2 I_n$. The principle of inference is to make use of the structure of $\boldsymbol{\theta}$.

1.3.1 Soft thresholding and Huber's M-estimation

In Gannaz (2007b), I proposed estimating the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in model (1.3) by minimizing a usual penalized least squares objective function,

$$(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_n) = \underset{(\boldsymbol{\beta},\boldsymbol{\theta})}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{2}(z_i - \boldsymbol{A}_i^\top \boldsymbol{\beta} - \theta_i)^2 + \lambda \sum_{i=i_0}^{n} |\theta_i| \right\}, \tag{1.4}$$

for a given penalty parameter λ, where $i_0 = 2^{j_0} + 1$. The penalty term in the above expression penalizes only the empirical wavelet coefficients of the nonparametric part of the model and not its scaling coefficients. The choice l^1 of the penalty function produces the soft thresholding rule. It is motivated by the sparse nature of the wavelet coefficients.

The specificity of the approach of Gannaz (2007b) is to establish a link between the Penalized Least Square approach defined by (1.4) and the Huber M-estimator. The mathematical equivalence of the solution is stated in the following proposition.

Proposition 1.2 (Gannaz 2007b)**.** *If $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$ are solutions of the optimization problem* (1.4)*, then they satisfy*

$$\widehat{\boldsymbol{\beta}}_n = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=i_0}^{n} \rho_\lambda(z_i - \boldsymbol{A}_i^\top \boldsymbol{\beta}), \tag{1.5}$$

$$\widehat{\theta}_{i,n} = \begin{cases} z_i - \boldsymbol{A}_i^\top \widehat{\boldsymbol{\beta}}_n & \text{if } i < i_0 \\ \gamma_{soft,\lambda}(z_i - \boldsymbol{A}_i^\top \widehat{\boldsymbol{\beta}}_n) & \text{if } i \geqslant i_0, \end{cases} \quad i = 1, \dots, n, \tag{1.6}$$

with ρ_λ being Huber's cost functional defined by:

$$\rho_\lambda(u) = \begin{cases} u^2/2 & \text{if } |u| \leqslant \lambda, \\ \lambda|u| - \lambda^2/2 & \text{if } |u| > \lambda. \end{cases}$$

and $\gamma_{soft,\lambda}$ the soft-thresholding function with threshold λ defined by $\gamma_{soft,\lambda}(u) = sign(u)\,(|u| - \lambda)_+$.

A benefit of Proposition 1.2 is the nice interpretation of the estimators which is induced. Indeed, in the wavelet domain, the estimators can be interpreted as follows.

- The parameter $\boldsymbol{\beta}$ is Huber 's robust estimator. The nonparametric component in the PLM model, is considered as noise, and the robustness of Huber's approach allows that it does not influence much the estimation. In other words it results from considering the linear model $z_i = \boldsymbol{A}_i^\top \boldsymbol{\beta} + e_i$ with noise $e_i = \theta_i + \varepsilon_i$ with a robust method.
- The estimation of f is then obtained by applying the standard soft-thresholding wavelet-based nonparametric estimation on the residuals of the previous step. That is, f is a usual nonparametric estimator in the model:

  $$y_i - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}_n = f(t_i) + v_i, \quad i = 1, \dots, n,$$

  where $v_i = \boldsymbol{X}_i^\top (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) + u_i$.

Simulated observations are represented in Figure 1.3, with the two parts of the model. After the projection in the wavelet space, it can be seen in Figure 1.4 that the linear structure is more visible. The functional part appears as outliers and the robustness of Huber's estimator makes sense.

Additionally, I propose in Gannaz (2007b) a procedure to estimate the variance σ^2. The idea is that the function wavelet coefficients at high resolutions are sparse. Additionally, a QR decomposition on the regression matrix allows to eliminate the linear part. Hence, considering the highest frequencies, with a QR decomposition, a robust estimation provide satisfactory results.

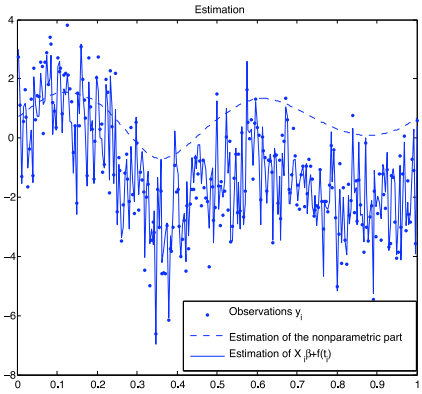


Figure 1.3: Example of a partial linear model. The figure represents the scatter plot of the observations, the estimated functional part (dash) and the partial linear fit (solid).

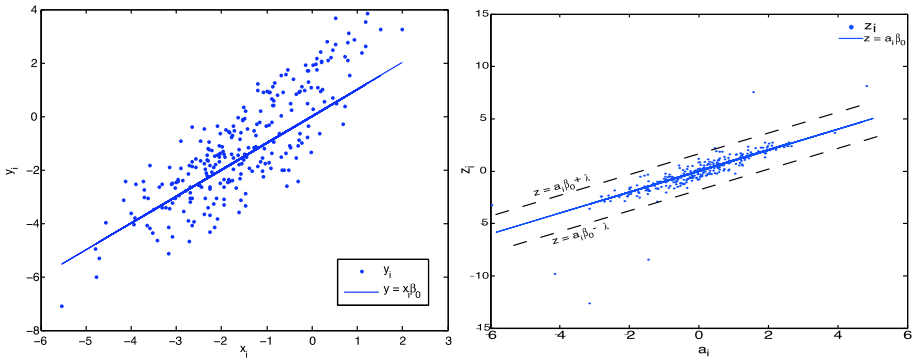


Figure 1.4: Figure (a) represents the scatter plot of the observations y_i versus the values of the covariates X_i for data of Figure 1.3. The line is the linear part of the model, of equation $y_i = X_i\boldsymbol{\beta}$. Figure (b) is the scatter plot in (a) after the Discrete Wavelet Transform: it represents the coefficients z_i versus $\boldsymbol{A}_i$. The solid line is the linear part of the model (equation $z_i = A_i\boldsymbol{\beta}$) and the dashed lines are the lines of equations $z_i = A_i\boldsymbol{\beta} \pm \lambda$.

1.3.2 Consistency of the estimation

We will assume:

(A1) The vector $\frac{1}{n}X^\top \boldsymbol{f}$ tends to 0 as n goes to infinity.

(A2) The matrix X is full rank, i.e. $\frac{1}{n}X^\top X$ converges towards an invertible matrix.

(A3) The series (K_n) defined by $K_n := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_i \boldsymbol{A}_i^\top \rho''_\lambda(\theta_i + \varepsilon_i)$, converges in the L^2-norm towards a non-singular matrix K_0.

(A4) The quantity $h := \max_{i=1,\dots,n} \boldsymbol{A}_i^\top (A^\top A)^{-1} \boldsymbol{A}_i$ tends to 0 when n goes to infinity.

(A5) $\forall j = 1,\dots,p,\ i = 1,\dots,n,\ X_{i,j} = g_j(t_i) + \xi_{i,j}$, with polynomial functions g_j of degree less or equal than the number of vanishing moments of the wavelet considered. For all $j = 1,\dots,p$, $(\xi_{i,j})_{i=1,\dots,n}$ is a n-sample such that $\max_{i=1,\dots,n} \mathbb{E}\left[\exp(\xi_{i,j}^2/a_j)\right] \leqslant a_j$, for given constants $a_j > 0$.

(A1) and **(A2)** ensure the identifiability of the model and **(A3)** the unicity of $\widehat{\boldsymbol{\beta}}_n$. Assumption **(A5)** controls the correlation between the linear and the function parts.

The main result is the following.

Theorem 1.3 (Gannaz 2007b)**.** *Let $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$ be the estimators defined by* (1.5) *and* (1.6) *in the model* (1.2)*. Consider that the penalty parameter λ is the universal threshold: $\lambda = \sigma\sqrt{2\log(n)}$. Under assumptions* **(A1) – (A5)***, we have*

$$\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \bigcirc_{\mathbb{P}}\left(\sqrt{\frac{\log(n)}{n}}\right).$$

If in addition we assume that the scaling function ϕ and the mother wavelet ψ belong to $\mathcal{C}^\alpha$ and that ψ has M vanishing moments, then, for f belonging to the Besov space $\mathcal{B}^s_{\pi,r}([0,1])$ with $0 < s - 1/2 + 1/\pi$ and $1/\pi < s < \min(\alpha, M)$, we have

$$\|\widehat{f}_n - f\|_2 = \bigcirc_{\mathbb{P}}\left(\left(\frac{\log(n)}{n}\right)^{\frac{s}{1+2s}}\right),$$

where $\|\widehat{f}_n - f\|_2^2 = \int_0^1 (\widehat{f}_n - f)^2$.

This theorem shows that the proposed inference attains parametric rates of convergence for the linear part and minimax rates for the nonparametric part, except an extra logarithmic term in the rate of the parametric part. In particular, the existence of a linear component does not changes the rates of convergence of the nonparametric component. Some comments on the accuracy of the assumption are given in the generalized framework below.

1.3.3 Implementation

Another advantage of the inference developped in Gannaz (2007b) is that it enables to estimate separately the linear and the functional part of the model. It is shown in Gannaz (2007b) that

this approach is more efficient than estimating conjointly $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using a backfitting algorithm (Fadili and Bullmore 2005).

The minimization problem for the linear part (1.5) can be solved using standard optimization tools such as relaxation, gradient, conjugated gradient and so on, but even if the loss function ρ_λ is convex, its second derivative is large near to zero, so the optimization may be slow. For this reason, I suggest *half-quadratic optimization*. This procedure leads to two algorithms, namely ARTUR and LEGEND, that are also referenced in the literature as *Iterative Reweighted Least Squares* and *Iterative Modified Residuals*. These algorithms are used for example in robust recognition, see *e.g* Vik (2004) which stresses the link between ARTUR and LEGEND and Huber's approach.

A real data application is provided in Gannaz (2007a), on a fMRI dataset from J. Fadili and described in Fadili and Bullmore (2005). I choose not to present it here.

1.3.4 Generalized partial linear models

In Gannaz (2013), I extend the results above to generalized settings. In a generalized regression setting, the response value y is drawn from a one-parameter exponential family of distributions, with a probabilistic density of the form:

$$\exp\left(\frac{y\eta(z) - b(\eta(z))}{u} + c(y,u)\right).$$

In this expression, $b(\cdot)$ and $c(\cdot)$ are known functions, which determine the specific form of the distribution. The parameter u is a dispersion parameter and is also supposed to be known in what follows. The unknown function $\eta(\cdot)$ is the natural parameter of the exponential family, which carries information from the explanatory variables. Given a random sample of size n drawn independently from a generalized regression model, the aim is to predict the function $\eta(\cdot)$.

In such a modeling, the conditional mean and variance of the i$^{\text{th}}$ response Y_i are given by:

$$\begin{aligned}\mathbb{E}[Y_i \mid z_i] &= \dot{b}(\eta(z_i)) \;=\; \mu(z_i),\\ Var[Y_i \mid z_i] &= \phi\, \ddot{b}(\eta(z_i)),\end{aligned}$$

where $\dot{b}(\cdot)$ and $\ddot{b}(\cdot)$ denote respectively the first and second derivatives of $b(\cdot)$. The function $G = \dot{b}^{-1}$ is called link function and one has $G(\mathbb{E}(Y_i \mid z_i)) = \eta(z_i)$.

Let $z = (\boldsymbol{X}, t)$, with $\boldsymbol{X}$ a p-dimensional vector and t a real-valued covariate. In the generalized partial linear model, the function $\eta(\cdot)$ is given by:

$$\eta(\boldsymbol{X}, t) = \boldsymbol{X}^\top \boldsymbol{\beta} + f(t),$$

where $\boldsymbol{\beta}$ is an unknown p-dimensional real parameter vector and f is an unknown real-valued function. I supposed that for all $i = 1, \ldots, n$, $t_i = i/n$.

The objective is to estimate simultaneously the parameter $\boldsymbol{\beta}$ and the function f, given the observed data $(Y_i, \boldsymbol{X}_i, t_i)_{i=1,\ldots,n}$. I proposed a penalized maximum loglikelihood estimation. Let $\mathcal{L}$ denotes the loglikelihood function, $\mathcal{L}(y, \eta) = y\eta - b(\eta)$. I considered estimators $\widehat{f}_n$ and $\widehat{\boldsymbol{\beta}}_n$ as solutions of

$$(\widehat{f}_n, \widehat{\boldsymbol{\beta}}_n) = \underset{\{f, \|f\|_\infty \leqslant C_\infty\}, \boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmax}} \sum_{i=1}^{n} \mathcal{L}\left(y_i, \boldsymbol{X}_i^\top \boldsymbol{\beta} + f(t_i)\right) - \lambda \, Pen(f).$$

The presence of the penalty $Pen(\cdot)$ corresponds to a constraint on f being in a given space. The parameter λ controls the degree of smoothness given on the estimation $\widehat{f}_n$.

Motivations

The introduction of generalized distributions was partly motivated by an application on recordings from olfactive cells on rats (Griff, Mafhouz, Perrut, and Chaput 2008). This work did not go to the final application (for personal reasons –including the death of Maryam Mafhouz who was the one working on these data– and because finally the modeling did not seem appropriate). Some examples of application can be found in the datasets of Schwartz (1994). For example, a dataset contains the number of pneumonia admissions in hospital Y_i, where the aim is to characterize the dependence to the pollution degree X_i. Schwartz (1994) proposes a generalized functional modeling, based on a Poisson distribution. Generalized partial linear regression seems adequate, considering a linear influence of the pollution and a time component of the recordings.

Theoretical results

Assumptions **(A1)** and **(A2)** ensure the identifiability of the model in the Gaussian framework. In the generalized framework, **(A1)** is unchanged but **(A2)** is replaced by

(A2) $\frac{1}{n}\sum_{i=1}^{n} \ddot{b}(\eta_i)\boldsymbol{X}_i\boldsymbol{X}_i^T$ converges to a strictly positive matrix when n goes to infinity

Assumptions **(A3)** and **(A4)** become

(A3) $\sup_{\{\tilde{\boldsymbol{\eta}} \in \mathbb{R}^n, \|\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}\|_n \leqslant 2C_\infty\}} \sup_{i=1,\ldots,n} \dddot{b}(\tilde{\eta}_i) \leqslant \|\dddot{b}\|_\infty < \infty$, where $\|g\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} g_i^2$..

(A4) $h = \max_{i=1,\dots,n} \boldsymbol{X}_i (\boldsymbol{X}^\top \mathrm{Diag}\left(\ddot{b}(\eta_i)\right) \boldsymbol{X})^{-1} \boldsymbol{X}_i^\top \to 0.$

Assumption **(A5)**, which controls the correlation between the covariates and the functional part, is unchanged.

We also need some assumptions on the distribution:

(A6.1) There exists a constant $a > 0$ such that $\max_{i=1,\dots,n} \mathbb{E}\left[\exp(\dot{\mathcal{L}}(Y_i, \eta_i)^2/a)\right] \leqslant a,$

(A6.2) There exist constants K, $\sigma_0^2 > 0$ such that $\max_{i=1,\dots,n} K^2 \left(\mathbb{E}[\exp\left(\left|\dot{\mathcal{L}}(Y_i, \eta_i)\right| / K^2\right)] - 1\right) \leqslant \sigma_0^2.$

Assumption **(A6.1)** corresponds to exponential tails and is weaker than assumption **(A6.2)**, which corresponds to sub-Gaussian tails.

I here chose to give only the results obtained with a wavelet-based penalty. Other results are obtained in Gannaz (2013), which I do not present here for brevity. The first result provides minimax optimality under weak conditions but with a non adaptive procedure.

Theorem 1.4 (Gannaz 2013)**.** *Suppose f belongs to a Besov ball $\mathcal{B}^s_{\pi,r}([0,1], C)$ with $C > 0$, $s > 1/2$, $0 < s + 1/\pi - 1/2$, $\pi > 2/(1+2s)$ and $1/\pi < s < \min(\alpha, M)$, where M denotes the number of vanishing moments of the wavelet ψ and α its regularity. Take the penalty $Pen(f) = \sum_{j=j_0}^{J-1} 2^{2js} \sum_k |\theta_{j,k}|^2$ where $\theta_{j,k}$ are the wavelet coefficients of f and $j_0 \geqslant 0$ a given resolution level. Assume conditions* **(A1)**–**(A5)** *and* **(A6.1)** *hold.*
If $\lambda \sim n^{-2s/(1+2s)}$, then

$$n^{s/(1+2s)} \|\widehat{f}_n - f\|_n = \bigcirc_{\mathbb{P}}(1)$$
$$\sqrt{n}\, \|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| = \bigcirc_{\mathbb{P}}(1).$$

Minimax optimality is obtained both for the linear predictor and the nonparametric estimator. This result is available for a wide class of distributions; *e.g.* compared to Mammen and Van der Geer (1997), assumptions on the distributions seem weaker. But the correlation assumption **(A5)** under which optimality is acquired, even if weaker than Rice (1986), is more restrictive than many of those encountered in literature, in particular comparing to Mammen and Van der Geer (1997).

Obviously, the estimation is not adaptive. An adaptive one is proposed in the theorem below. Yet, the cost for adaptivity is the assumption on the distribution. Indeed, Assumption **(A6.1)** is modified in its more restrictive version **(A6.2)**.

Theorem 1.5 (Gannaz 2013)**.** *Suppose assumptions* **(A1)**–**(A5)** *and* **(A6.2)** *hold. Suppose f belongs to a Besov ball $\mathcal{B}^s_{\pi,r}(C)$ with $s + 1/\pi - 1/2 > 0$ and $1/2 < s < \min(\alpha, M)$, where M denotes the number of vanishing moments of the wavelet ψ and α its regularity.*

Consider the penalty $Pen(f) = \lambda \sum_{i=i_0}^{n} |\theta_i|$ where (θ_i) are the wavelet coefficients of f and $i_0 \geqslant 0$ a given scale. Suppose $\lambda = c_0\sqrt{\log(n)}$. There exists a positive constant $C(K, \sigma_0^2)$ depending only on K and σ_0^2 given in assumption (A6.2) such that if $c_0 > C(K, \sigma_0^2)$, then, we have

$$\left(\frac{n}{\log(n)}\right)^{s/(1+2s)} \|\widehat{f}_n - f\|_n = \bigcirc_{\mathbb{P}}(1),$$
$$\sqrt{n}\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| = \bigcirc_{\mathbb{P}}(1).$$

The adaptivity is acquired. This offers the possibility of a computable procedure. Yet, the parameter λ is only chosen among an asymptotic condition and a finite sample application arises that the exact choice of this parameter is important. Observe finally that the link with soft-thresholding, which appears in the Gaussian setting, is not straightforward. Yet, I have established that it comes naturally during the iterative implementation of the estimators. The implementation is performed using a backfitting algorithm, as in Fadili and Bullmore (2005), since algorithms developed in Section 1.3.3 are not usable here.

References

Dedecker, J. and C. Prieur (2007). "An empirical central limit theorem for dependent sequences". In: *Stochastic Processes and their Applications* 117.1, pp. 121–142.

Donoho, D. and I. M. Johnstone (1998). "Minimax estimation via wavelet shrinkage". In: *The Annals of Statistics* 26.3, pp. 879–921.

Donoho, D., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1996). "Density estimation by wavelet thresholding". In: *Annals of Statistics* 24.2, pp. 508–539.

Doukhan, P. and S. Louhichi (1999). "A new weak dependence condition and applications to moment inequalities". In: *Stochastic Processes and their Applications* 84, pp. 313–342.

Doukhan, P. and O. Wintenberger (2007). "Invariance principle for new weakly dependent stationary models". In: *Probabability and Mathematical Statistics* 27, pp. 45–73.

Fadili, J. and E. Bullmore (2005). "Penalized partially linear modelsusing sparse representation with an application to fMRI time series". In: *IEEE Transactions on signal processing* 53.9, pp. 3436–3448.

Gannaz, I. (2007a). "Estimation par ondelettes dans les modèles partiellement linéaires". PhD thesis. Université Joseph Fourier, Grenoble.

Gannaz, I. (2007b). "Robust estimation and wavelet thresholding in partially linear models". In: *Statistics and Computing* 17.4, pp. 293–310.

Gannaz, I. (2013). "Wavelet penalized likelihood estimation in generalized functional models". In: *Test* 22.1, pp. 122–158.

Gannaz, I. and O. Wintenberger (2010). "Adaptive density estimation under weak dependence". In: *ESAIM: Probability and Statistics* 14, pp. 151–172.

Griff, E. R., M. Mafhouz, A. Perrut, and M. A. Chaput (2008). "Comparison of identified mitral and tufted cells in freely breathing rats: I. Conduction velocity and spontaneous activity". In: *Chemical senses* 33.9, pp. 779–792.

Juditsky, A. and S. Lambert-Lacroix (2004). "On minimax density estimation on $\mathbb{R}$." In: *Bernoulli* 10.2, pp. 187–220.

Mallat, S. (1999). *A wavelet tour on signal processing*. 2nd ed. Academic Press.

Mammen, E. and S. Van der Geer (1997). "Penalized Quasi-likelihood estimation in partial linear models". In: *The Annals of Statistics* 25.3, pp. 1014–1035.

Rice, J. (1986). "Convergence Rates for Partially Splined Models". In: *Statistics and Probability Letters* 4, pp. 203–208.

Schwartz, J. (1994). "Nonparametric smoothing in the analysis of air pollution and respiratory illness". In: *The Canadian Journal of Statistics, Special issue on the analysis of health and environmental data* 22.4, pp. 471–487.

Vik, T. (2004). "Modèles statistiques d'apparence non gaussiens. Application à la création d'un atlas probabiliste de perfusion cérebrale en imagerie médicale". PhD thesis. Université Strasbourg 1.

CHAPTER 2

CLASSIFICATION AND CLUSTERING OF FUNCTIONAL DATA

Motivations

I consider recordings from one or multiple sensors, on a fine equidistant grid of time. Functional data analysis methods are appropriate to take into account the nature of the data. Each observation here consists in a functional random variable $\boldsymbol{X} = (X_1(t), \ldots, X_p(t))_{\{t\in[a,b]\}}$, $p \geqslant 1$. The objective is to characterize the behavior of these (possibly multivariate) functional observations, by supervised or unsupervised classification models.

Contributions

In Gannaz (2014) [15], I work with EEG recordings concerning auditory activity. The objective is to discriminate the signal with respect to known stimuli. It is, hence, a supervised classification problem, which I chose to model with a wavelet-based logistic regression. I compared different reduction techniques. The study was experimental, without theoretical developments.

Amovin-Assagba, Gannaz, and Jacques (2022) [25] deal with an industrial application, where data is recorded simultaneously by four sensors. We are interested in automatic detection of abnormal behaviors of the sensors. Due to the heterogeneity of the measurements, we choose to due jointly a clustering and an outlier detection, where outliers are defined with respect to the clusters. Hence, an unsupervised classification modeling is proposed, which includes the presence of outliers, or contamination. An algorithm is developed to infer the parameters of the model and a simulation study shows its good performance. The application on the real dataset fulfills the objective. Unfortunately, due to a confidentiality contract with the enterprise, the code cannot be freely released.

The chapter is organized in two sections, dealing respectively with the supervised functional

classification studied in Gannaz (2014) [15], and with the non supervised functional classification developed in Amovin-Assagba, Gannaz, and Jacques (2022) [25].

2.1 Functional logistic linear model

This work was motivated by a collaboration with Rafael Laboissiere (CNRL, Univ. Lyon - now LPNC, Univ. Grenoble alpes). It relates to a neuroscience real data application which I present below. No theoretical results were developed, but different procedures of dimension reduction were compared empirically. Results are provided in Gannaz (2014).

Dataset 1

Data were obtained by an electroencephalography (EEG) which recorded the auditory evoked potentials from the scalp in a standard 32-electrodes mounting cap, on a single subject. Four stimuli, corresponding to sounds */ba/*, */pa/* and two intermediates obtained by modifying the offset of the plosives */b/* and */p/*, were emitted in a random order. For each of the four stimuli, we dispose of 1000 records, which we will study on a 2^{10} equally spaced grid at 5 kHz. See Bellier et al. (2013) for a complete description of the dataset. The objective is to see if EEG recordings discriminate the stimuli and to give an insight on where this discrimination is done. Only the evoked potentials between the frontal electrode and the right ear of a subject are considered. A high-pass filter at 80 Hertz was applied to keep only the frequencies corresponding to auditory activity. In addition, I have considered the averages of ten signals in order to get rid of a possible random effect. Hence, 100 signals are available for each sound.

2.1.1 Logistic functional regression

Let Y_i, $i = 1, \dots, n$, be independent labeled variables, with values 0 or 1. Suppose that the predictor variables $(\{X_i(t),\, t \in [0,1]\})_{i=1,\dots,n}$ belong to the separable Hilbert space $L^2([0,1])$, with the usual inner product (see Section 1.1).

Suppose that the observations are drawn from a functional logistic model :

$$\mathbb{P}\left(Y_i = 1 \mid X_i(\cdot)\right) = g(X_i(\cdot)),$$
$$\text{where } g(X_i(\cdot)) = \frac{e^{\eta(X_i(\cdot))}}{1 + e^{\eta(X_i(\cdot))}} \text{ and } \eta\left(X_i(\cdot)\right) = < X_i(\cdot), \beta(\cdot) > .$$

The function g is the logit link function. The function $\beta(\cdot)$ captures the distinctions between the curves $(X_i(\cdot))_{i=1,\dots,n}$. Intuitively when $|\beta(x_0)|$ is large the value taken by the signal at the point x_0 is discriminant.

2.1.2 Wavelet-based inference

As in Section 1.3, the idea is that the part of the data which captures the difference between the signals is sparse, in time and in frequency. Hence a wavelet representation will help to extract the significant characteristics.

Let us consider a multiresolution analysis of $L^2[0,1]$, associated with a compactly supported scaling function, φ, and a compactly supported mother wavelet, ψ, as described in Section 1.1. We suppose that each observation $X_i(\cdot)$, $i = 1,\dots,n$, is observed on a sample $t_j = \frac{j}{n_X}$, $\{X_i(t_1),\dots,X_i(t_{n_X})\}$. Suppose that there exists $J \in \mathbb{N}$ such that $n_X = 2^J$ (it is sufficient here to cut off the end of the signal which is of no interest). $W_{n_X \times n_X}$ denotes the DWT (see Section 1.1). Every signal $X_i(\cdot)$ is decomposed by $\boldsymbol{X}_i = W^\top_{n_X \times n_X}\boldsymbol{\theta}_i$, for $i = 1,\dots,n$, where the exponent $\top$ denotes the transpose operator. We introduce $\boldsymbol{\omega}$ the vector of wavelet coefficients of the function $\beta(\cdot)$. Hence, we can write $\boldsymbol{\beta} = W^\top_{n_X \times n_X}\boldsymbol{\omega}$ and thus $\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\Theta}\boldsymbol{\omega}$ with $\boldsymbol{\Theta}$ the matrix which ith column equals to $\boldsymbol{\theta}_i$. The functional regression model is expressed like a regression on the wavelet coefficients. To reduce the dimension of the resulting problem but also to match the functional nature of the regressor $\beta(\cdot)$, we impose the sparsity of the wavelet coefficients vector $\boldsymbol{\omega}$. This is usually done in literature thanks to a ℓ^1-penalization.

Following Reiss and Ogden (2007), we introduce an additional dimension reduction through a constraint on the wavelet coefficients $\boldsymbol{\omega}$. We impose that they belong to the space generated by the first components of the wavelet coefficients matrix of the curve predictors.

Let $a_1, a_2, \dots, a_{n_X}$ be the eigenvalues of $\boldsymbol{\Theta}$. We assume that the eigenvalues are sorted in the descending order $a_1 \geqslant a_2 \geqslant, \dots \geqslant a_{n_X}$. We introduce the matrix V_q of size $n_X \times q$ such that the ith column of the matrix V_q is the eigenvector associated with the eigenvalue a_i, for $i = 1,\dots,q$. We then search $\boldsymbol{\omega}$ such that there exists $\boldsymbol{\gamma} \in \mathbb{R}^q$ verifying $\boldsymbol{\omega} = V_q\boldsymbol{\gamma}$. An extension to Partial Least Squares reduction is easily done, similarly to Reiss and Ogden (2007). The idea is that V_q contain the linear combinations of $\boldsymbol{\Theta}$'s columns which better explain the labels $(Y_i)_{i=1,\dots,n}$. The additional reduction of dimension is straightforward: we impose the vector $\boldsymbol{\omega}$ to belong to a space of dimension q.

Consequently, the estimators are defined as follows:

$$\widetilde{\boldsymbol{\omega}}_n(q,\lambda) = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \; -\sum_{i=1}^{n} \mathcal{L}\left(Y_i,\, \theta_i^\top \boldsymbol{\omega}\right) + \lambda \sum_{\ell=2^{j_0}+1}^{n_X} |\omega_\ell|$$
$$s.t.\;\; \exists \boldsymbol{\gamma} \in \mathbb{R}^q,\, \boldsymbol{\omega} = \widetilde{V}_q \boldsymbol{\gamma} \tag{2.1}$$

where $\mathcal{L}$ is the loglikelihood. WCR estimator is defined by (2.1), with $\widetilde{V}_q = V_q$. While WLS estimator is defined by (2.1) but with $\widetilde{V}_q$ spanned by the eigen vectors obtained by Partial Least Squares.

Actually, I have compared the behavior of the following procedures, with a 5-folds cross validation:

- SPCR - A spline-based estimation, with Principal Component Reduction (CR) and a ℓ^2 penalization, initially described in Reiss and Ogden (2007).
- WNET - A wavelet-based estimation with a ℓ^1-penalty, studied in Zhao, Ogden, and Reiss (2012). It is available for elastic-net penalties but only the ℓ^1-penalty has been considered.
- WCR and WLS - A wavelet-based estimation, with respectively a sparse CR or a sparse Partial Least Squares (LS) reduction described in Johnstone and Lu (2009).
- WPCR and WPLS - Wavelet-based estimation described above, with respectively CR or LS and a ℓ^1-penalization.

2.1.3 Application

The dataset was cut in a training set and a validation set. The model was fitted using 75 signals for each sound and its accuracy was checked on the remaining data, that is, 25 signals by sounds.

Method	SPCR	**WNET**	WCR	WPCR	WLS	WPLS
Learning sample	0.586	**0.734**	0.808	0.744	1	1
Validation sample	0.533	**0.708**	0.659	0.648	0.580	0.572

Table 2.1: AUC for discrimination of EEG recordings in two classes. The learning set and the validation set contain respectively 75 curves and 25 curves for each stimulus.

No estimators succeed to discriminate the EEG recordings in four classes, with respect to the four stimuli. I then try to classify the signals in two classes, with /ba/ and the first intermediate

stimulus associated with a label 0 and /pa/ and the second intermediate stimulus with a label 1. Areas under ROC curves (AUC) are given in Table 2.1. The best result is obtained by WNET estimator, the discrimination of the EEG signals in two classes is validated. Spline-based SPCR procedure is not able to capture the differences in signals. Estimators with a principal component or a partial least squares step succeed in discriminating on the learning sample but not on the validation sample. The reduction of component seems too dependent on the learning set and it reduces the quality of prediction.

2.2 Functional mixture models

I present here a work on clustering and outlier detection on multivariate functional data, published in Amovin-Assagba, Gannaz, and Jacques (2022). This work is a part of the PhD of Martial-Amovin Assagba (Arpege MasterK, ERIC, Univ. Lyon 2), which I co-supervise with J. Jacques (ERIC, Univ. Lyon 2). It is motivated by an industrial application, which I present below.

Dataset 2

The data are time measurements coming simultaneously from various sensors in an industrial context. We consider a dataset from a material with 4 sensors, knowing that we also have other material with more than 4 sensors. The data are recorded every 10 milliseconds. We favor a functional data type approach, since the observations are smooth. One measurement is represented by 4 curves $x_i = \{(x_i^{(1)}(t_j), x_i^{(2)}(t_j), x_i^{(3)}(t_j), x_i^{(4)}(t_j)),\ t_j = 1/n_i, 2/n_i \ldots, 1,\ n_i \in \mathbb{N} \setminus \{0\}\}$, $i = 1, \ldots, n$. The objective is to detect automatically an abnormal behavior of the sensors. It can be due to a dysfunction of one sensor or a decoupling of the sensors recordings. The outliers are defined (manually) by an expert on the dataset. We have 569 measurements with 37 outliers. Our goal is to identify these aberrant curves, in an unsupervised way. This dataset is not freely available, due to the industrial context.

Figure 2.1 displays the behavior of some normal (figures on left) and abnormal (figures on right) curves. As shown in the figure, the observation durations differ as well as the amplitudes from one measurement to another. Abnormal curves cannot be determined using these characteristics. The figures on right show two of the abnormal measurements. Their shape are quite different from the measurements observed on the left. For our expert, a measurement is aberrant if at least the shape of one of the curves differs from the shape of normal curves (figures on left).

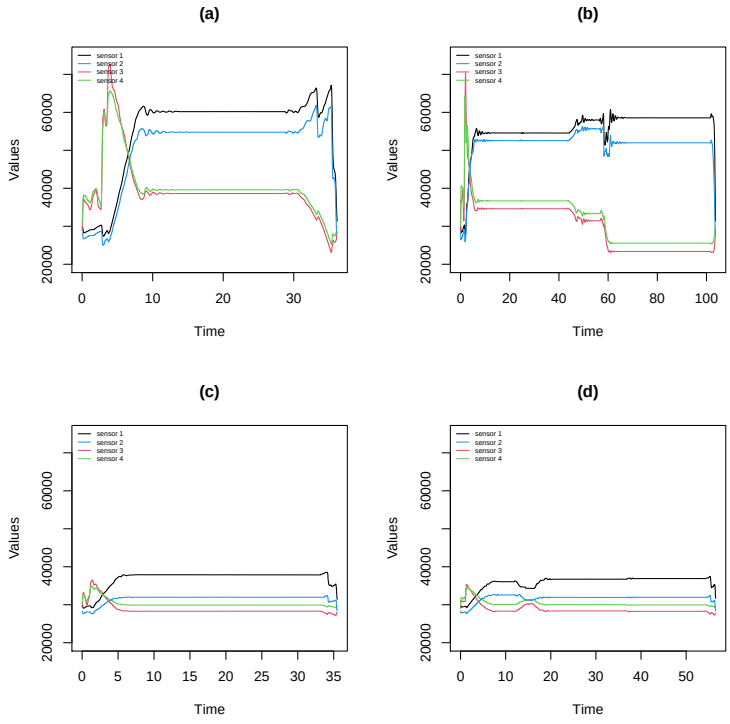


Figure 2.1: Representation of two normal recordings (curves on the left, (a) and (c)) and of two abnormal recordings (curves on the right, (b) and (d)).

The data can be considered as multivariate functional data. A functional data in multivariate case is represented by a stochastic process $\{\boldsymbol{X}(t),\, t \in [0,1]\}$, with $\boldsymbol{X}(t) = \left(X^{(1)}(t), \ldots, X^{(p)}(t)\right)$, $t \in [0,1]$, $p \in \mathbb{N} \setminus \{0\}$. For the real data application, we consider a normalization of the time duration to 1 for each recording. Each component of the process X belongs to $L^2[0,1]$. Let $\{X_1(t), X_2(t), \ldots, X_n(t)\}$ be an *i.i.d.* sample of X where each copy X_i of X is a multivariate curve $X_i(t) = \left(X_i^{(1)}(t), \ldots, X_i^{(p)}(t)\right)$ with $t \in [0,1]$, $p \in \mathbb{N} \setminus \{0\}$. Denote $x_i = (x_i^{(1)}(t), \ldots, x_i^{(p)}(t))$ the realizations of X_i. Our observations are, hence, observations on a time sample of curves x_i, $i = 1, \ldots, n$. For simplicity, I will often not distinguish observations $\{x_i(t),\, t \in [0,1]\}$ and observations of x_i on a time sample.

To sum up, we are interested by the study and definition of automatic outlier detection algorithms for functional data. Due to the heterogeneity of the measurement, we choose to couple the outlier detection with an automatic clustering. We are then considering a clustering of contaminated functional data, where the contamination corresponds to the outliers that we want

to identify.

2.2.1 Contaminated Gaussian mixture models for real data

Let me first describe the model for $(\mathbf{x}_i)_{i=1,\dots,n}$ independent observations in $\mathbb{R}^p$, $p \in \mathbb{N} \setminus \{0\}$. The objective of contaminated mixture models is to group observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ into K clusters and to detect simultaneously the outliers.

Following Punzo and McNicholas (2016), the idea is that the main distribution in each cluster is a Gaussian distribution with a large prior probability. And for each cluster, this distribution is slightly contaminated by another distribution with a small prior probability. The distribution of the contamination is Gaussian, with the same expectancy but a larger variance, corresponding to observations further to the center of the cluster than the main part of observations.

Let $g(\cdot, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ the density of a Gaussian distribution of mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. The Contaminated Gaussian mixture models suppose that, for each observation $\mathbf{x}_i$, the distribution is a mixture of multivariate contaminated normal distributions, with density

$$p(\mathbf{x}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \left[\delta_k g(x_i, \boldsymbol{\mu}_k, v\boldsymbol{\Sigma}_k) + (1-\delta_k) g(x_i, \boldsymbol{\mu}_k, \eta_k \boldsymbol{\Sigma}_k)\right],$$

where $\pi_k \in [0,1]$ is the proportion of observations in cluster k, $\sum_{k=1}^{K} \pi_k = 1$, $\delta_k \in [0,1]$ is the proportion of normal observations, and $\eta_k > 1$ is an inflation parameter that indicates the degree of outlyingness.

The likelihood function is, hence,

$$L(\{\mathbf{x}_1, \dots \mathbf{x}_n\}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{x}_i, \boldsymbol{\theta}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \left[\delta_k g(\mathbf{x}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + (1-\delta_k) g(\mathbf{x}_i, \boldsymbol{\mu}_k, \eta_k \boldsymbol{\Sigma}_k)\right],$$

with the vector of parameters $\boldsymbol{\theta} = \{(\pi_k, \delta_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \eta_k),\ k = 1, \dots, K\}$.

2.2.2 Extension to functional data

Probabilistic modeling of functional data

Let me now come back to the functional data which motivate this work. For functional data, the definition of a probability distribution cannot be done directly. According to Delaigle and Hall (2010), the modeling can be done on the coefficients of the decomposition of the curves on eigenfunctions. Following Jacques and Preda (2014), I detail how to propose a mixture of Gaussian distributions, based on this property.

Suppose the random curve $X_i = \{(X_i^{(1)}(t), X_i^{(2)}(t), \ldots, X_i^{(p)}(t)), t \in [0,1]\}$ centered with values in $L^2([0,1])$, assuming a covariance operator $\mathcal{V}$:

$$\forall f \in L_2[0,1], \forall t \in [0,1], \mathcal{V}f(t) = \int_0^1 V(s,t)f(s)\mathrm{d}s\,.$$

where V is the covariance function defined by: $\forall (s,t) \in [0,1]^2, V(s,t) = \mathbb{E}(X(t) \otimes X(s)) := \left(Cov(X^{(k)}(s), X^{(l)}(t)\right)_{k,l\in\{1,\ldots,p\}}$. Let us assume that $\mathbb{E}(\|X\|^2) < \infty$. Then the covariance operator $\mathcal{V}$ is an Hilbert-Schmidt operator, compact and self-adjoint. Hence it is diagonalizable. So, there exists a Hilbert basis of $(f_l)_{l\geqslant 1}$, formed by the eigenfunctions of the operator $\mathcal{V}$ associated with eigenvalues $(\lambda_l)_{l\geqslant 1}$. The principal components scores S_i are therefore defined as the projections of X_i on the eigenfunctions of $\mathcal{V}$, $\mathbf{S}_i = (\int_0^1 X_i(t)f_l(t)\mathrm{d}t)_{l\geqslant 1}$. Following Delaigle and Hall (2010), a probabilistic modeling of the scores $\{\mathbf{S}_i, i = 1, \ldots, n\}$ can be done.

As the eigenfunctions differ for each cluster, we introduce a common basis where we decompose the eigenfunctions and the observed curves. Let $(\phi_b^{(j)})_{b\geqslant 1}$ be a basis function in $L_2[0,1]$, $j = 1, \ldots, p$. Each observation $x_i^{(j)}$ of $X_i^{(j)}$ is approximated by a linear combination of B_j functions, $B_j \in \mathbb{N} \setminus \{0\}$:

$$x_i^{(j)}(t) \simeq \sum_{b=1}^{B_j} c_{ib}^{(j)}\phi_b^{(j)}(t),\ t \in \mathbb{R},\ i = 1, \ldots, n,\ j = 1, \ldots, p.$$

Denote $\mathbf{C}$ the concatenation of the coefficients,

$$\mathbf{C} = (\mathbf{c}_{B_1}^{(1)}, \mathbf{c}_{B_2}^{(2)}, \ldots, \mathbf{c}_{B_p}^{(p)}) \text{ with } \mathbf{c}_{B_j}^{(j)} = (c_{i,b}^{(j)})_{\substack{i=1,\ldots,n,\\ b=1,\ldots,B_j}}.$$

$\mathbf{c}_{B_j}^{(j)}$ represents the coefficients of the j^{th} component of the multivariate curves. Then the curves $\mathbf{x} = (x_1, \ldots, x_n)$ can be written in the form:

$$\mathbf{x}(t) \simeq \mathbf{C}\mathbf{\Phi}(t)^\top,\ t \in [0,1],$$

with the matrix $\mathbf{\Phi}(t)$ defined by:

$$\mathbf{\Phi}(t) = \begin{pmatrix} \Phi_{B_1}^{(1)}(t) & 0 & \ldots & 0 \\ 0 & \Phi_{B_2}^{(2)}(t) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Phi_{B_p}^{(p)}(t) \end{pmatrix} \text{ with } \Phi_{B_j}^{(j)}(t) = (\phi_1^{(j)}(t), \ldots, \phi_{B_j}^{(j)}(t)).$$

Similarly, each eigenfunction f_l of the covariance operator $\mathcal{V}$ can be approximately decomposed in the form

$$f_l(t) \simeq \mathbf{A}_l\mathbf{\Phi}(t)^\top,\ t \in [0,1],$$

where $\mathbf{A}_l = (\mathbf{a}_{B_1}^{(1)}, \ldots, \mathbf{a}_{B_p}^{(p)})$ with $\mathbf{a}_{B_j}^{(j)} = (a_{i,b}^{(j)})_{\substack{i=1,\ldots,n,\\ b=1,\ldots,B_j}}$. Finally, the scores $\mathbf{S_i}$ can be approximated by

$$\mathbf{S}_i \simeq \mathbf{C}_i \mathbf{W} \mathbf{A},$$

where $\mathbf{W} = \int_0^1 \mathbf{\Phi}(t)^\top \mathbf{\Phi}(t) \mathrm{d}t$ denotes a symmetric block-diagonal matrix of the inner products between the basis functions, with size $\sum_{l=1}^p B_l \times \sum_{l=1}^p B_l$. $\mathbf{C}_i$ denotes the i^{th} row of $\mathbf{C}$.

The empirical estimator of the covariance function is

$$\widehat{V}(s,t) = \frac{1}{n-1} \sum_{i=1}^n x_i(s) x_i(t) = \frac{1}{n-1} \mathbf{x}(s)^\top \mathbf{x}(t) \simeq \frac{1}{n-1} \mathbf{\Phi}(s) \mathbf{C}^\top \mathbf{C} \mathbf{\Phi}(t)^\top .$$

The empirical version of the eigenproblem determining the eigenfunctions $(f_l)_{l \geqslant 0}$ can be written as

$$\frac{1}{n-1} \mathbf{\Phi}(s) \mathbf{C}^\top \mathbf{C} \mathbf{W} \mathbf{A}_l^\top = \lambda_l \mathbf{\Phi}(s) \mathbf{A}_l^\top, \ \ s \in [0,1].$$

This problem also writes as

$$\frac{1}{n-1} \mathbf{W}^{1/2} \mathbf{C}^\top \mathbf{C} \mathbf{W}^{1/2} \mathbf{U}_l^\top = \lambda_l \mathbf{Q}_l^\top, \, l \geqslant 1,$$

with $\mathbf{Q}_l = \mathbf{A}_l \mathbf{W}^{1/2}$. Consequently,

$$\mathbf{S}_i \simeq \mathbf{C}_i \mathbf{W}^{1/2} \mathbf{Q}^\top,$$

where $\mathbf{Q} = (\mathbf{Q}_l)$ is the matrix of the eigenvectors of $\mathbf{W}^{1/2} \mathbf{C}^\top \mathbf{C} \mathbf{W}^{1/2}$. Hence, the formulation of the scores is entirely determined by the coefficients $\mathbf{C}$ representing the data in a basis $\mathbf{\Phi}$. Making an assumption on the distribution of the scores or on the coefficients is equivalent. In particular, assuming that the scores $(\mathbf{S}_i)_{i=1,\ldots,n}$, follow a mixture of contaminated Gaussian distributions, is equivalent to assuming that the basis expansion coefficients $(\mathbf{C}_i)_{i=1,\ldots,n}$ also follow a mixture of contaminated Gaussian distributions.

Dimension reduction

If we consider a model like the one described in Section 2.2.1 on the scores $\mathbf{S}_i$, each of the K clusters has $2 + B + B \times B$ parameters to estimate, where $B = \sum_{j=1}^p B_j$ is the total number of coefficients associated to each multivariate curve x_i. Due to the high dimension of the model, we introduce a framework which reduces the dimension, following Schmutz et al. (2020). For $k \in \{1, \ldots, K\}$, we assume that the stochastic processes associated with the k^{th} cluster can be described in a low-dimensional functional latent subspace of $L^2[0,1]$ with dimension $d_k \leqslant B$. Let $\mathbf{Q}_k$ be the orthogonal matrix of size $B \times B$, which entries equal to the sorted eigenvectors of $\mathbf{W}^{1/2} \mathbf{C}^\top \mathbf{C} \mathbf{W}^{1/2}$ on the cluster k. We split into two parts: $\mathbf{Q}_k = [\mathbf{U}_k, \mathbf{V}_k]$ where $\mathbf{U}_k$ is a matrix of size $B \times d_k$ and $\mathbf{V}_k$ a matrix of size $B \times (B - d_k)$, and $\mathbf{U}_k$ and $\mathbf{V}_k$ are such that $\mathbf{U}_k^\top \mathbf{U}_k = \mathbf{I}_{d_k}$,

$\mathbf{V}_k^\top \mathbf{V}_k = \boldsymbol{I}_{B-d_k}$ and $\mathbf{U}_k^\top \mathbf{V}_k = 0$. Under this assumption, only the first d_k scores are significant on cluster k. The number of parameters of the k^{th} cluster is then $2 + B + d_k$.

Let z_{ik} be a latent variable associated to each observation $\mathbf{x}_i$, such that $z_{ik} = 1$ if the observation x_i belongs to the cluster k and $z_{ik} = 0$ otherwise. Let us similarly define v_{ik} such that $v_{ik} = 1$ if $\mathbf{x}_i$ in group k is a normal observation and $v_{ik} = 0$ if it is an outlier.

Conditionally to $z_{ik} = 1$, let us assume that the first d_k principal component scores $S_i^\top, 1 \leqslant i \leqslant n_k$, follow a contaminated Gaussian distribution. That is,

$$
\begin{aligned}
(S_{i1}, \dots S_{id_k}) \mid z_{ik} = 1, v_{ik} = 1 &\sim \mathcal{N}_{d_k}(\mathbf{m}_k, \boldsymbol{\Delta}_k), \\
(S_{i1}, \dots, S_{id_k}) \mid z_{ik} = 1, v_{ik} = 0 &\sim \mathcal{N}_{d_k}(\mathbf{m}_k, \eta_k \boldsymbol{\Delta}_k),
\end{aligned}
$$

with $\mathbf{m}_k \in \mathbb{R}^{d_k}$, $\eta_k > 1$ and $\boldsymbol{\Delta}_k = \text{Diag}\left(a_{k1}, a_{k2}, \dots, a_{kd_k}\right)$. Based on the previous discussion, we assume that the basis expansion coefficients satisfy

$$
\mathbf{C}_i \mid z_{ik} = 1, v_{ik} = 1 \sim \mathcal{N}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \mathbf{C}_i \mid z_{ik} = 1, v_{ik} = 0 \sim \mathcal{N}\left(\boldsymbol{\mu}_k, \eta_k \boldsymbol{\Sigma}_k\right),
$$

with

$$
\boldsymbol{\mu}_k = \mathbf{W}^{-1/2} \mathbf{U}_k m_k, \quad \boldsymbol{\Sigma}_k = \mathbf{W}^{-1/2} \mathbf{U}_k \boldsymbol{\Delta}_k \mathbf{U}_k^\top \mathbf{W}^{-1/2} + \boldsymbol{\Lambda}_k \,.
$$

Finally, following Schmutz et al. (2020), we suppose that the noise covariance matrix $\boldsymbol{\Lambda}_k$ is such that

$$
\mathbf{Q}_k^\top \mathbf{W}^{1/2} \boldsymbol{\Sigma}_k \mathbf{W}^{1/2} \mathbf{Q}_k = \left(\begin{array}{ccc|cc} \alpha_{k1} & & 0 & & \\ & \ddots & & & 0 \\ 0 & & \alpha_{kd_k} & & \\ \hline & & & \beta_k & 0 \\ & 0 & & & \ddots \\ & & & 0 & \beta_k \end{array}\right) \tag{2.2}
$$

with $\alpha_{k1} > \alpha_{k2} > \ldots > \alpha_{kd_k} > \beta_k$. In view of these notations, the data of the k^{th} cluster is decomposed by a main term, with a variance modeled by $\alpha_{k1}, \alpha_{k2}, \dots, \alpha_{kd_k}$, and by a noise component with a variance modeled by a unique parameter β_k.

Final model

To conclude, the contaminated mixture functional model of data $(x_1, \dots x_n)$ supposes that the associated coefficients $(\mathbf{C}_1, \dots, \mathbf{C}_n)$ when decomposing in a basis have the likelihood

$$
L(\{\mathbf{C}_1, \dots, \mathbf{C}_n\}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \left[\delta_k \, g(\mathbf{C}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + (1 - \delta_k) g(\mathbf{C}_i, \boldsymbol{\mu}_k, \eta_k \boldsymbol{\Sigma}_k)\right],
$$

with the dimension reduction assumption (2.2) on variances $\mathbf{\Sigma}_k$, $k = 1, \ldots, n$. The parameters of the model are $\boldsymbol{\theta} = \{(\pi_k, \delta_k, \boldsymbol{\mu}_k, (\alpha_{kl}, l = 1, \ldots, d_k), \beta_k, \eta_k),\ k = 1, \ldots, K\}$. This model is called C-funHDDC hereafter.

2.2.3 Implementation

An Expectation Conditional Maximization (ECM) algorithm allows to maximize the completed log-likelihood, and provides an estimation the parameters. I refer to Amovin-Assagba, Gannaz, and Jacques (2022) for a description of this algorithm. It needs an initialization. When using usual algorithms for the initialization (for example kmeans), it appears that they are perturbed by the presence of outliers in the data and that they lead to unsatisfactory results. Hence, a robust initialization, *trimmed kmeans* (Cuesta-Albertos, Gordaliza, Matrán, et al. 1997), is used.

Moreover, some hyperparameters need to be calibrated: the number K of clusters and the intrinsic dimensions $\{d_k, k = 1, \ldots, K\}$ of each cluster. We propose to consider the hyperparameters maximizing a BIC objective function, that is, introducing a penalization with respect to the number of parameters.

A simulation study shows that the procedure developed here obtains satisfactory results on simulated data. Indeed, it not only clusters adequately the curves, but also detects correctly the outliers. Yet, this behavior occurs when the proportion of outliers is low (at least below 5% per cluster in our simulation study). When this proportion increases, the algorithm behaves differently but still conveniently. When the proportion of outliers is (relatively) high, the clustering of normal curves is still adequate, independently of the presence of outliers. More surprising, the outliers are gathered into an additional cluster. Hence, outliers are not detected as abnormal by the model, but they are easily identifiable since they form a specific cluster. Observe that when applying alternative (non robust) algorithms, for example with *funHDDC* (Schmutz et al. 2020), this additional cluster of outliers is not created.

I refer to Amovin-Assagba, Gannaz, and Jacques (2022) for the description of the algorithm and of the simulation study.

2.2.4 Application

C-funHDDC is applied to the industrial dataset, Dataset 2, described above. The number K of clusters and the intrinsic dimension d_k are selected by BIC. The range of explored values is $\{1, 2, \ldots, 6\}$ for K and dimensions (d_k) are chosen by a Catell's method (see Amovin-Assagba, Gannaz, and Jacques 2022). The best number of clusters selected by the BIC criterion is $K = 1$

with $d_1 = 87$. The results are different from Amovin-Assagba, Gannaz, and Jacques 2022 due to a change in the code.

The results are presented in Table 2.2 as a confusion matrix. C-funHDDC detects 38/47 outliers, with 9 false positives and 6 false negatives (as labeled by the expert). The outliers are represented in Figure 2.2. Different kinds of outliers are detected.

Reference \ Predicted	normal	outlier
normal	522	9
outlier	6	38

Table 2.2: Confusion matrix using C-funHDDC model

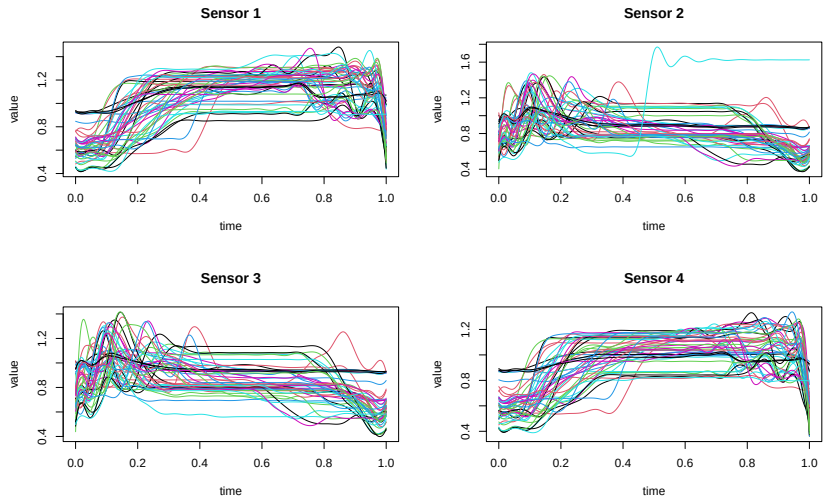


Figure 2.2: Detected outliers in Dataset 2.

We have compared with the results obtained by other methods in literature, namely with *Functional Trimmed* (FT) (García-Escudero, Gordaliza, Matrán, Mayo-Iscar, et al. 2008) and *Fonctional Isolation Forest* (FIF) (Staerman, Mozharovskyi, Clémençon, and d'Alché-Buc 2019). The problem with FT and FIF is that they need to fix the proportion of outliers, which is clearly unknown in real application. Nevertheless, even with the true proportion of outliers, the results are relatively poor with FT and FIF, in particular by comparison with the results of C-funHDDC.

References

Amovin-Assagba, M., I. Gannaz, and J. Jacques (2022). "Outlier detection in multivariate functional data through a contaminated mixture model". In: *Computational Statistics & Data Analysis* 174, p. 107496.

Bellier, L., M. Mazzuca, H. Thai-Van, A. Caclin, and R. Laboissière (2013). "Categorization of speech in early auditory evoked responses." In: *INTERSPEECH*, pp. 911–915.

Cuesta-Albertos, J. A., A. Gordaliza, C. Matrán, et al. (1997). "Trimmed k-means: An attempt to robustify quantizers". In: *The Annals of Statistics* 25.2, pp. 553–576.

Delaigle, A. and P. Hall (2010). "Defining probability density for a distribution of random functions". In: *The Annals of Statistics* 38.2, pp. 1171–1193.

Gannaz, I. (2014). "Classification of EEG recordings in auditory brain activity via a logistic functional linear regression model." In: *International Workshop on Functional and Operatorial Statistics*, p–125.

García-Escudero, L. A., A. Gordaliza, C. Matrán, A. Mayo-Iscar, et al. (2008). "A general trimming approach to robust cluster analysis". In: *The Annals of Statistics* 36.3, pp. 1324–1345.

Jacques, J. and C. Preda (2014). "Model-based clustering for multivariate functional data". In: *Computational Statistics & Data Analysis* 71, pp. 92–106.

Johnstone, I. M. and A. Y. Lu (2009). "On consistency and sparsity for principal components analysis in high dimensions". In: *Journal of the American Statistical Association* 104, pp. 682–693.

Punzo, A. and P. D. McNicholas (2016). "Parsimonious mixtures of multivariate contaminated normal distributions". In: *Biometrical Journal* 58.6, pp. 1506–1537.

Reiss, P. and R. Ogden (2007). "Functional principal component regression and functional partial least squares". In: *Journal of the American Statistical Association* 102, pp. 984–996.

Schmutz, A., J. Jacques, C. Bouveyron, L. Cheze, and P. Martin (2020). "Clustering multivariate functional data in group-specific functional subspaces". In: *Computational Statistics*, pp. 1–31.

Staerman, G., P. Mozharovskyi, S. Clémençon, and F. d'Alché-Buc (2019). "Functional Isolation Forest". In: *arXiv preprint arXiv:1904.04573*.

Zhao, Y., R. Ogden, and P. Reiss (2012). "Wavelet-based LASSO in functional linear regression". In: *Journal of Computational and Graphical Statistics* 21.3, pp. 600–617.

Part II

Extracting the dependence structures: multivariate long-range dependence models

Introduction

This part deals with statistical analysis of multivariate time series. On the contrary to the previous part, where the objective is to carry out inference while taking into account the dependence structures, the objective here is to extract the dependence structures. I am interested more specifically in the case where recordings present long-range dependence properties. The aim is to estimate the time-related memory of each components as well as the inter-components structure, modeled as a covariance matrix. This is realized by using (real and complex) wavelets representation of the time series. Once we have an estimate of this covariance matrix, a multiple testing approach makes possible to summarize the inter-signals dependence as a graph, which gives access to graph theory tools for analyzing or comparing the structure.

The procedures presented in this document were motivated by a neuroscience application. We dispose of non invasive recordings of brain activity (EEG, fIRM, MEG). Our objective is to extract the structure of the brain activity by applying statistical inferences described above. The two main features of interest are the long-range dependence properties and the graph of cerebral connectivity.

Figure 3.1 illustrates the continuity of the work described hereafter and its relation with neuroscience applications. The strength of this work is to go from the data to to a final structure of dependence, considering each step from the theoretical and the practical aspects : it proposes a modeling, an inference procedure, its implementation, proofs of consistency, statistical tests, *etc*.

This part is structured in three chapters, dealing respectively with

- Long-range dependence models and estimation of characteristics with a real wavelet representation,
- Asymptotic normality and multiple testing on the estimators,
- Extension to general phases modeling, with the introduction of complex wavelets.

Even if it would have been appropriate to gather the estimation with real and complex wavelets, I choose to keep this outline to increase the complexity progressively, in particular in the real data applications.

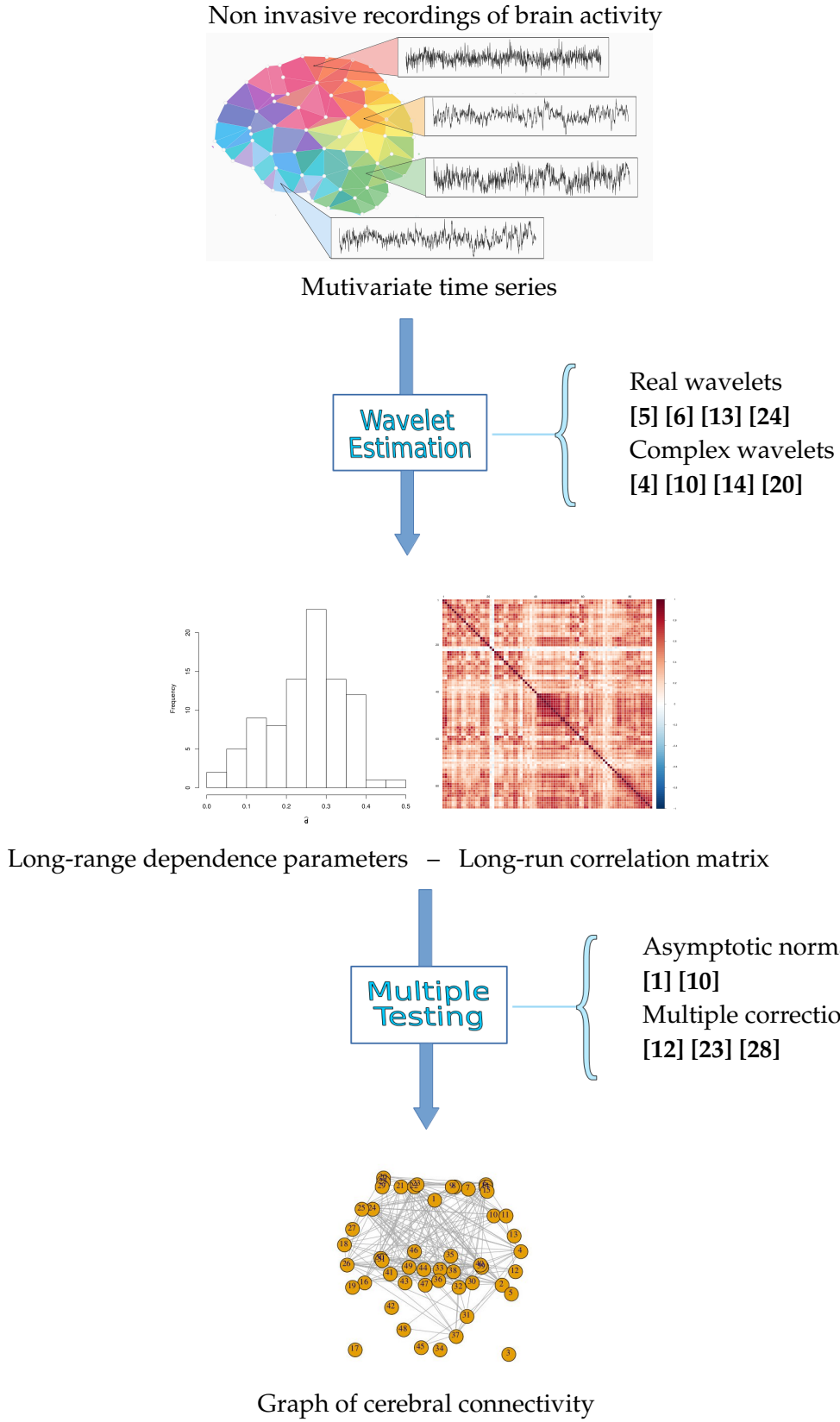


Figure 3.1: Inference of the graph of cerebral connectivity. Numbers into brackets refer to the list of publications (page 117).

CHAPTER 3

MULTIVARIATE WAVELET-BASED INFERENCE IN LONG-RANGE DEPENDENCE MODELS

Dataset 3

Functional magnetic resonance imaging (fMRI) is a non-invasive way to measure brain activity. Actually, it measures the BOLD (Blood Oxygen Level Dependent) effect. The fMRI recordings consist in time series for each voxel, where a voxel is a cube of a few millimeters, across the volume of the brain. A usual preprocessing, which will be applied for each dataset here, is to average the time series on brain regions. Brain regions are determined by anatomical properties. We hence obtain one time series for each brain region, rather than each voxel.

In this chapter, I will consider fMRI recordings for 100 subjects who were scanned twice. The whole description of this dataset is detailed in Termenon, Achard, Jaillard, and Delon-Martin (2016). 89 regions of interest were extracted for each scan, with time series of length 1200 time points, at 1.39Hz. Figure 3.2 displays 6 arbitrary signals from a subject in this data set.

Motivation

Let $X = \{X(t), t \in \mathbb{Z}\}$ be a real-valued second-order stationary process. Denote $\gamma(k) = \mathrm{Cov}\,(X(t), X(t+k))$, $t \in \mathbb{Z}$, $k \in \mathbb{Z}$, its autocovariance function. The process is said to have long-range dependence (LRD) if there exists $d \in (0, 1/2)$ such that $\gamma(k) = L_1(k)k^{2d-1}$, with L_1 a slowly varying function at infinity.

Let f denote the spectral density of X, $f(\lambda) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma(k)\mathrm{e}^{-\mathrm{i}k\lambda}$. Equivalently, the process X is LRD, with LRD parameter d, if $f(\lambda) = L_2(\lambda)\lambda^{-2d}$ as $\lambda \to 0^+$, with L_2 a slowly varying function at zero.

Figure 3.3 shows the autocorrelograms of two arbitrary fMRI signals for one individual of Dataset 3. It illustrates that fMRI signals present LRD properties and that the intensity of

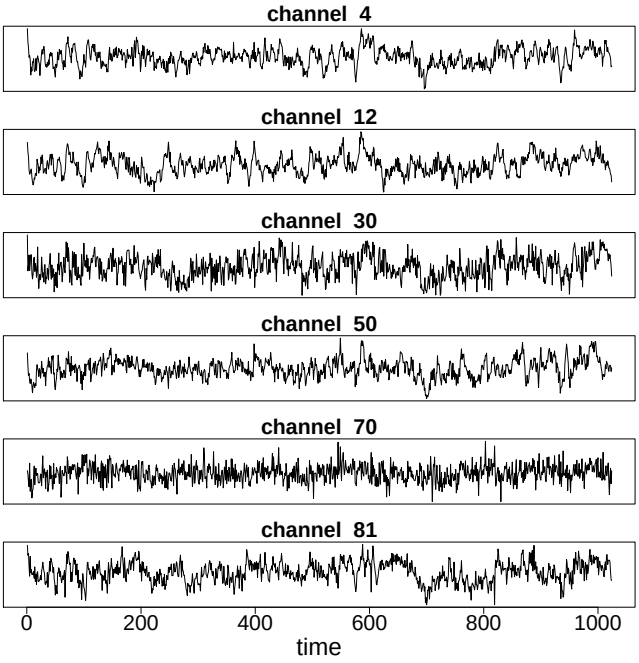


Figure 3.2: Plot of 6 arbitrary signals from one subject of Dataset 3.

this LRD is not homogeneous with respect to the brain region. A similar assessment was obtained for example in Maxim et al. (2005), where the authors highlight moreover that the LRD parameters are modified in presence of Alzheimer's disease.

A first motivation is, hence, to model and to estimate the LRD properties of the signals. Next, we are also interested in the inference of the cerebral connectivity, that is, the correlations between the activities of the different regions of the brain. We thus want the modeling to include inter-signals correlations.

Contributions

The objective is to define a modeling adequate for the data, on multivariate time series, with long-range dependence and coupling between the components. Next, we want to develop statistical inference in this setting. We choose to consider the multivariate time series modeling described in Lobato (1999), Shimotsu (2007), and Kechagias and Pipiras (2015). The two main characteristics in such models are the long-range dependence (LRD) parameters, which

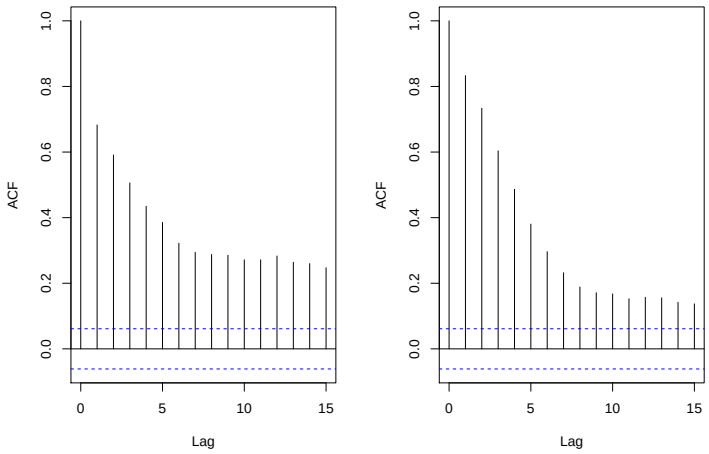


Figure 3.3: Autocorrelograms of 2 fMRI signals from one subject in Dataset 3.

captures the time dependence of each component, and the long-run covariance matrix, which measures the coupling between the components. My main contribution is to propose a wavelet-based estimation procedure in this context.

Due to their flexibility in real data applications, we have chosen to consider a wavelet representation of the multivariate time series. A natural approach for estimating the parameters is to consider the sample covariance of the wavelet coefficients. We first give some results about the behavior of the covariance of the wavelet coefficients in Achard and Gannaz (2016a) [6]. We have shown that a phase shift, caused by differences in LRD parameters, may introduce a bias in inference. At the beginning, we, hence, focus on cases where the phase between time series component is specified.

Achard and Gannaz (2016a) [6] propose a local Whittle approximation based on the wavelet coefficients. It enables to estimate jointly the time-related dependence parameters and the spatial dependence structure, that is, the long-range dependence parameters and the long-run covariance. Achard and Gannaz (2016a) [6] also establish the asymptotic consistency of the estimates and provide the rate of convergence.

A `R` package called `multiwave` [24] as well as a Matlab toolbox (available at `http://math.univ-lyon1.fr/~gannaz/research/MultiwaveMatlab.zip`) are freely available. They implement the wavelet-based Whittle estimation described in Achard and Gannaz (2016a) [6] and the Fourier-based Whittle estimation of Shimotsu (2007). Practical aspects of the estimation process are described in Achard and Gannaz (2019) [5]. A comparison between wavelet-based and Fourier-

based estimation is also displayed.

Achard and Gannaz (2016b) [13] provide an example of application on Dataset 3 and illustrate with a real data example that multivariate estimation improves univariate estimation for the long-memory parameters, and supplies additionally the long-run covariance.

The chapter is organized as follows. I first recall briefly the multivariate LRD model and present some (theoretical) examples. The second section describes the wavelet Whittle estimation procedure when the phase admits a parametric form. It introduces the (real) wavelet transform, presents the estimation procedure, and gives consistency results. The third section deals with the application on Dataset 3.

3.1 Multivariate long-memory setting

LRD models were used in a large scope of applications, for example finance (see *e.g.* Gençay, Selçuk, and Whitcher (2001) or references in Nielsen and Frederiksen (2005)), internet traffic analysis (Abry and Veitch 1998), physical sciences (Percival and Walden 2006; Papanicolaou and Sølna 2003), geosciences (Whitcher and Jensen 2000) and neuroimagery (Maxim et al. 2005).

Let $\boldsymbol{X} = \{\boldsymbol{X}(t), t \in \mathbb{Z}\}$ denote a multivariate long-memory dependence process, $\boldsymbol{X}(t) = \begin{bmatrix} X_1(t) & \dots & X_p(t) \end{bmatrix}^T$, $t \in \mathbb{Z}$ with long memory parameters $\boldsymbol{d} = (d_1, d_2, \dots, d_p)$, $\boldsymbol{d} \in (-0.5, +\infty)^p$. We will denote by $\mathbb{1}$ the identity operator and by $\mathbb{L}$ the lag operator, $(\mathbb{1} - \mathbb{L})\boldsymbol{X}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(t-1)$. The k^{th} difference operator, $(\mathbb{1} - \mathbb{L})^k$, $k \in \mathbb{N}$, is defined by k recursive applications of $(\mathbb{1} - \mathbb{L})$. For $\boldsymbol{D} = \lfloor \boldsymbol{d} + 1/2 \rfloor$, we assume that the multivariate process $\text{Diag}\left((\mathbb{1} - \mathbb{L})^{D_a}, a = 1, \dots, p\right) \boldsymbol{X}$ is covariance stationary with a spectral density matrix given by

(M-1) $$\boldsymbol{f}^{\boldsymbol{D}}(\lambda) = \left(\text{Diag}\left(|\lambda|^{-d_1^*}, \dots, |\lambda|^{-d_p^*}\right) \boldsymbol{\Theta} \, \text{Diag}\left(|\lambda|^{-d_1^*}, \dots, |\lambda|^{-d_p^*}\right)\right) \circ \boldsymbol{f}^S(\lambda), \text{for all } \lambda > 0,$$

where $\circ$ denotes the Hadamard product, and $d_a^* = d_a - D_a \in (-0.5, 0.5)$ for all a. The process $(\mathbb{1} - \mathbb{L})^{D_a} X_a$ is said to have long-memory if $d_a^* \in (0, 0.5)$, and to be anti-persistent if $d_a^* \in (-0.5, 0)$ (see for instance Lobato 1999; Shimotsu 2007). For simplicity of notation, I will use the term LRD parameters for $\boldsymbol{d}$ throughout the manuscript, whatever the values of $\boldsymbol{d}$.

Let $\boldsymbol{f}(\cdot)$ be the function defined by

$$\boldsymbol{f}(\lambda) = (\boldsymbol{\Lambda}(\lambda) \, \boldsymbol{\Theta} \, \boldsymbol{\Lambda}(\lambda)) \circ \boldsymbol{f}^S(\lambda), \quad \text{for all } \lambda > 0,$$

with $\boldsymbol{\Lambda}(\lambda) = \text{Diag}\left(|\lambda|^{-d_1}, \dots, |\lambda|^{-d_p}\right)$. Under the condition **(M-1)**, the function $\boldsymbol{f}(\cdot)$ is called the generalized spectral density of the multivariate process $\{\boldsymbol{X}(t), t \in \mathbb{Z}\}$.

The function $\boldsymbol{f}^S(\cdot)$ represents the short-range dependence of $\boldsymbol{f}(\cdot)$. In order to get identifiability, it is necessary to assume $\boldsymbol{f}^S(0) = \mathbf{1}$. The following assumption is also needed to control the regularity.

(M-2) There exists $C_f > 0$ and $\beta > 0$ such that $\sup_{0<\lambda<\pi} \sup_{a,b=1,\dots,N} \frac{\left|f^S_{a,b}(\lambda)-1\right|}{\lambda^\beta} \leqslant C_f$.

In particular, our definition agrees with the one given in Kechagias and Pipiras (2015) if $D_a = 0$ for all a. The authors also define long-range dependence in the time domain, through the behavior of the autocovariance function. They derive some examples, including some presented hereafter.

The major interest of this model is the introduction of the matrix $\boldsymbol{\Theta}$. This provides a generalisation of multivariate LRD models used in Lobato (1997) and Shimotsu (2007). Indeed, the matrix $\boldsymbol{\Theta}$ can be written as,

$$\Theta_{a,b} = \Omega_{a,b} \mathrm{e}^{\mathrm{i}\varphi_{a,b}} ,$$

with $\boldsymbol{\Omega} = (\Omega_{\ell,m})_{a,b=1,\dots,p}$ a real symmetric non-negative semi-definite matrix and with $\boldsymbol{\Phi} = (\varphi_{a,b})_{a,b=1,\dots,p}$ an anti-symmetric matrix. Let the bar above denote the conjugate operator. The matrix $\boldsymbol{\Theta}$ satisfies $\boldsymbol{\Theta}^T = \overline{\boldsymbol{\Theta}}$ since $\boldsymbol{f}^T(\cdot) = \overline{\boldsymbol{f}(\cdot)}$. We will use $\|\boldsymbol{\Omega}\|$ to denote the infinity norm, that is, $\|\boldsymbol{\Omega}\| = \max_{a,b=1,\dots,p} \Omega_{a,b}$. In Lobato (1997), Shimotsu (2007), and Achard and Gannaz (2016a), the phase term was defined by $\varphi_{a,b} = \pi(d_a - d_b)/2$.

Below, I sum up some examples of literature, from Kechagias and Pipiras (2015), Lobato (1997), Sela and Hurvich (2008), and Coeurjolly, Amblard, and Achard (2013).

3.1.1 Some multivariate linear time series

The extension of univariate LRD models to multivariate LRD models is quite recent in the time series literature. The first considerations on the multivariate extensions of ARFIMA models can be found in Lobato (1997), which I briefly sum up below. Since the composition of linear filters does not commute in the multivariate case, there have been multiple extensions of univariate ARFIMA to the multivariate framework afterwards. For example, Kechagias and Pipiras (2015) have proposed recently other (possibly non causal) multivariate linear processes with LRD properties.

Multivariate ARFIMA of Lobato (1997)

Denote by the exponent $\top$ the transpose operator. Let $\mathbf{u}$ be a p-dimensional white noise. with $\mathbb{E}[\mathbf{u}(t) \mid \mathcal{F}(t-1)] = 0$ and $\mathbb{E}[\mathbf{u}(t)\mathbf{u}(t)^\top \mid \mathcal{F}(t-1)] = \boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma}$ positive definite, where $\mathcal{F}(t-1)$ is the σ-field generated by $\{\mathbf{u}(s),\, s < t\}$. The spectral density of $\mathbf{u}$ is $\boldsymbol{f}_u(\lambda) = \Sigma/(2\pi)$, for all $\lambda \in \mathbb{R}$.

Let $(\boldsymbol{A}_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{p\times p}$ with $\boldsymbol{A}_0$ the identity matrix and $\sum_{k=0}^{\infty} \|\boldsymbol{A}_k\|^2 < \infty$. Let $\boldsymbol{A}(\cdot)$ be the discrete Fourier transform of the sequence, $\boldsymbol{A}(\lambda) = \sum_{k=0}^{\infty} \boldsymbol{A}_k e^{ik\lambda}$. We assume $|\boldsymbol{A}(\mathbb{L})|$ has all its roots outside the unit circle which ensures that $\boldsymbol{A}(\cdot)^{-1}$ is defined and smooth on $\mathbb{R}$. We are also given $(B_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{p\times p}$ with B_0 the identity matrix and $\sum_{k=0}^{\infty} \|B_k\|^2 < \infty$. Let $\boldsymbol{B}(\cdot)$ be the discrete Fourier transform of the sequence, $\boldsymbol{B}(\lambda) = \sum_{k=0}^{\infty} \boldsymbol{B}_k e^{ik\lambda}$.

Lobato (1997) define two models:

Model A - FIVARMA. If $\boldsymbol{A}(\mathbb{L})\,\mathrm{Diag}\left((\mathbb{1}-\mathbb{L})^{\boldsymbol{d}}\right)\boldsymbol{X}(t) = \boldsymbol{B}(\mathbb{L})\mathbf{u}(t)$ the spectral density satisfies

$$\boldsymbol{f}(\lambda) = (1-e^{-i\lambda})^{-\boldsymbol{d}}\boldsymbol{A}(e^{-i\lambda})^{-1}\boldsymbol{B}(e^{-i\lambda})\boldsymbol{f}_u(\lambda)\boldsymbol{B}(e^{i\lambda})^\top \boldsymbol{A}(e^{i\lambda})^{\top\,-1}(1-e^{i\lambda})^{-\boldsymbol{d}}.$$

In particular

$$f_{a,b}(\lambda) \sim_{\lambda\to 0^+} \Omega_{a,b} e^{-i\pi/2(d_a-d_b)}\lambda^{-(d_a+d_b)}\,,$$

with $\boldsymbol{\Omega} = \boldsymbol{A}(0)^{-1}\boldsymbol{B}(0)\Sigma\boldsymbol{B}(0)^\top \boldsymbol{A}(0)^{\top\,-1}/(2\pi)$ with is a real valued matrix. Condition **(M-2)** is satisfied with $\beta = \min_{a=1,\dots,p}(d_a)$. In this case $\boldsymbol{f}(0^+) = \boldsymbol{f}(0^-)$.

Model B - VARFIMA. If $\mathrm{Diag}\left((\mathbb{1}-\mathbb{L})^{\boldsymbol{d}}\right)\boldsymbol{A}(\mathbb{L})\,\boldsymbol{X}(t) = \boldsymbol{B}(\mathbb{L})\mathbf{u}(t)$ the spectral density satisfies

$$\boldsymbol{f}(\lambda) = \boldsymbol{A}(e^{-i\lambda})^{-1}(1-e^{-i\lambda})^{-\boldsymbol{d}}\boldsymbol{B}(e^{-i\lambda})\boldsymbol{f}_u(\lambda)\boldsymbol{B}(e^{i\lambda})^\top(1-e^{i\lambda})^{-\boldsymbol{d}}\boldsymbol{A}(e^{i\lambda})^{\top-1}.$$

In particular

$$f_{a,b}(\lambda) \sim_{\lambda\to 0^+} \sum_{\ell,m} \beta_{\ell,m}\alpha_{a,\ell}\alpha_{b,m} e^{-i\pi/2(d_\ell-d_m)}\lambda^{-(d_\ell+d_m)},$$

with $\alpha_{\ell,m} = (\boldsymbol{A}(0)^{-1})_{\ell,m}$ and $\beta_{\ell,m} = (\boldsymbol{B}(0)\boldsymbol{f}_u(\lambda)\boldsymbol{B}(0)^\top)_{\ell,m}$. Condition **(M-2)** is satisfied with $\beta = \min_a(d_a)$. In this case $\boldsymbol{f}(0^+) \neq \boldsymbol{f}(0^-)$.

According to Lobato (1997) both models satisfy the definition of LRD processes **(M-1)** when $\boldsymbol{f}_u(\lambda) \sim_{\lambda\to 0^+} \boldsymbol{\Sigma}/(2\pi)$. It is straightforward in Model A. In Model B, Lobato argues that the spectral density will be equivalent to the term in $(a,b) = \mathrm{argmax}\{|d_a + d_b|,\ \beta_{a,b}\alpha_{\ell,a}\alpha_{m,b} \neq 0\}$. It gives a more general form $f_{a,b} \sim \Theta_{a,b}\lambda^{-d_{a,b}}$ with $|d_{a,b}| \leqslant (d_a+d_b)/2$. This case is called fractional cointegration. Setting $\Theta_{a,b} = 0$ if $|d_{a,b}| < (d_a+d_b)/2$, the LRD property (condition

(M-1)) holds. Lobato (1997) affirms that the LRD property includes any fractional model as a particular case.

Differences between the two models are highlighted in Sela and Hurvich (2008). The authors considered a subclass of Lobato (1997)'s with no MA part, called respectively FIVAR and VARFI models. FIVARMA and VARFIMA names are derived from this work. Models of Lobato (1997) are studied more extensivley in Sela and Hurvich (2012). In particular, an estimator for $d_{1,2}$ in a bivariate setting is proposed. Our approach does not enable this estimation. Next, Kechagias and Pipiras (2020) consider the general linear setting of Lobato (1997), that is, FIVARMA and VARFIMA models, extending Sela and Hurvich (2008)'s FIVAR and VARFI models by adding a MA part. The authors point out that all phases $\varphi_{a,b}$ can occur in the model, as summarized below. A parametric estimation based on Durbin-Levinson's theorem is proposed. Finally, Baek, Kechagias, and Pipiras (2020) propose a Fourier-based local Whittle estimation in such a framework.

To visualize the meaning of the different parameters, I represent some realizations of FIVARMA processes. Figure 3.4 displays examples of FIVARMA processes with the same LRD parameters but with different covariance values, and Figure 3.5 displays examples of FIVARMA processes with different LRD parameters and with the same covariance values. The figures illustrate that the higher the long-run correlation, the higher the coupling between the components (Figure 3.4). And the higher the difference between the LRD components, the higher the asymmetry of the cross-correlograms (Figure 3.5). The latest remark is related with the phase shift $\phi_{a,b} = \frac{-\pi}{2}(d_a - d_b)$ in the spectral density.

Extensions of Kechagias and Pipiras

Kechagias and Pipiras (2015), in Proposition 3.1, give examples of multivariate LRD linear time series. I choose to present only a generalization of multivariate ARFIMA described above.

Let $\{\mathbf{u}(s)\}_{s\in\mathbb{Z}}$ be an $\mathbb{R}^p$-valued white noise, satisfying $\mathbb{E}\big(\mathbf{u}(s)\big) = 0$ and $\mathbb{E}\big(\mathbf{u}(s)\mathbf{u}(s)^\top\big) = \boldsymbol{I}_p$ for all $s \in \mathbb{Z}$. Let also $\{\boldsymbol{\Psi}_s = (\Psi_{a,b,s})_{a,b=1,\dots,p}\}_{s\in\mathbb{Z}}$ be a sequence of real-valued matrices such that

$$\Psi_{a,b,s} = L_{a,b}(k)|k|^{d_a-1},\ s \in \mathbb{Z},$$

where $d_a \in (0, 1/2)$ and $\boldsymbol{L}(s)$, $s = 1, \cdots, p$, is an $\mathbb{R}^{p\times p}$-valued function satisfying

$$\boldsymbol{L}(s) \sim_{s\to+\infty} \boldsymbol{A}^+,\ \ \boldsymbol{L}(s) \sim_{s\to-\infty} \boldsymbol{A}^-$$

for some $\mathbb{R}^{p\times p}$-valued matrices $\boldsymbol{A}^+$ and $\boldsymbol{A}^-$. Then, the time series $\boldsymbol{X} = \{\boldsymbol{X}(t),\ t \in \mathbb{Z}\}$ given by

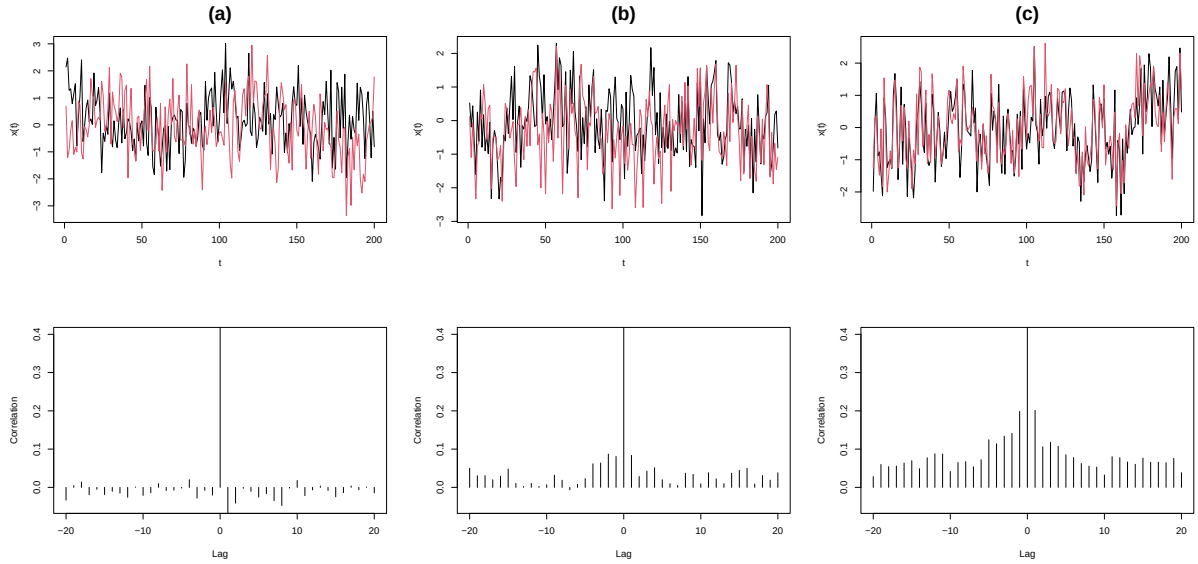


Figure 3.4: Realizations from 3 bivariate FIVARMA processes, with no MA nor AR parts. On the first line, the processes $\boldsymbol{X}(t)$ are represented with respect to t. On the second line, the cross-correlograms are displayed. At a lag $l \in \mathbb{Z}$, the value on the graph is the sample estimation of $Cor(\{(X_1(t), X_2(t+l)),\, t \in \mathbb{Z}\})$ with $N = 2000$ observations. Each bivariate fBM has parameters $d_1 = d_2 = 0.2$ and $\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. From left to right (a) $\rho = 0$ (b) $\rho = 0.4$ (c) $\rho = 0.8$.

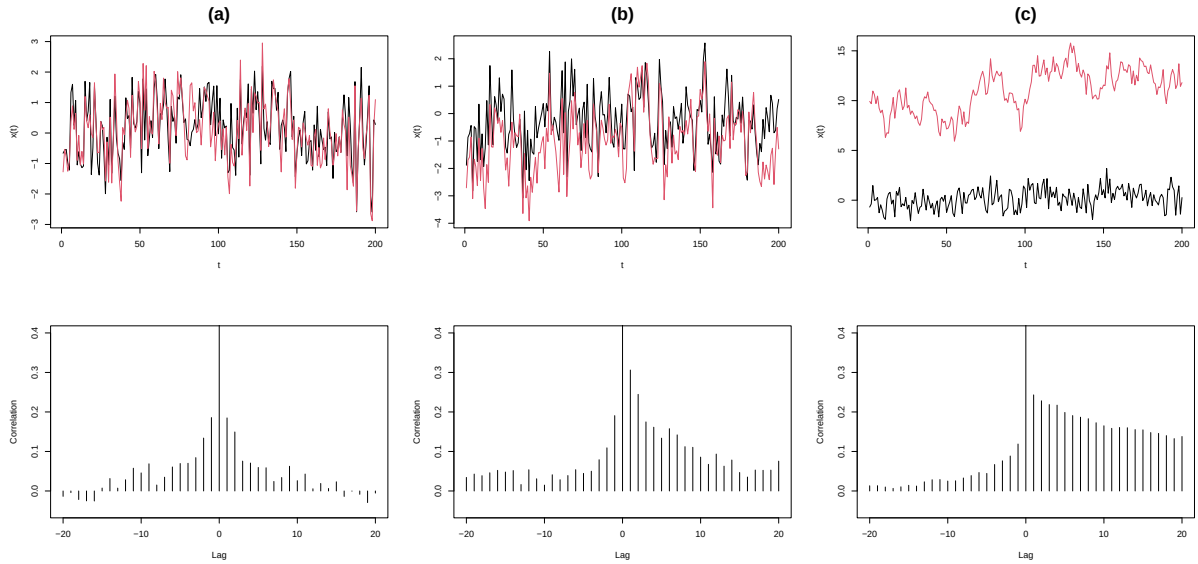


Figure 3.5: Realizations from 3 bivariate FIVARMA processes, with no MA nor AR parts. On the first line, the processes $\boldsymbol{X}(t)$ are represented with respect to t. On the second line, the cross-correlograms are displayed. At a lag $l \in \mathbb{Z}$, the value on the graph is the sample estimation of $Cor(\{(X_1(t), X_2(t+l)),\, t \in \mathbb{Z}\})$ with $N = 2000$ observations. Each bivariate fBM has parameters $d_1 = d_2 = 0.2$ and $\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho = 0.8$. $d_1 = 0.2$ and, from left to right, (a) $d_2 = 0.2$ (b) $d_2 = 0.4$ (c) $d_2 = 0.8$.

a linear representation

$$\boldsymbol{X}(t) = \sum_{s=-\infty}^{+\infty} \boldsymbol{\Psi}_s \mathbf{u}(t-s),\ t \in \mathbb{Z}$$

is LRD.

The spectral density satisfies **(M-1)**, with $\boldsymbol{\Theta}$ given by $\boldsymbol{\Theta} = (\overline{\boldsymbol{F}}^\top \boldsymbol{A}^+ + \boldsymbol{F}\boldsymbol{A}^-)\overline{(\overline{\boldsymbol{F}}^\top \boldsymbol{A}^+ + \boldsymbol{F}\boldsymbol{A}^-)}^\top/(2\pi)$ with $\boldsymbol{F} = \text{Diag}\left(\Gamma(d_a)e^{i\frac{\pi}{2}d_a},\ a = 1,\dots,p\right)$. It follows that any combinations of $(\boldsymbol{d}, \boldsymbol{G}, \boldsymbol{\Phi})$ can be attained choosing appropriate matrices $\boldsymbol{A}^+$ and $\boldsymbol{A}^-$.

Observe that if we consider causal linear representations, $\boldsymbol{X}(t) = \sum_{s=0}^{+\infty} \boldsymbol{\Psi}_s \mathbf{u}(t-s)$, with the same assumptions, we obtain that $\boldsymbol{X}$ is LRD with

$$\begin{aligned}\Omega_{a,b} &= \frac{\Gamma(d_a)\Gamma(d_b)}{2\pi}(\boldsymbol{A}^+\overline{(\boldsymbol{A}^+)}^\top)_{a,b},\\ \varphi_{a,b} &= -\frac{\pi}{2}(d_a - d_b).\end{aligned}$$

We recover the same phase as in the FIVARMA model described above.

3.1.2 A simple example

Subsequently, I will introduce wavelet representations of the time series, and develop an estimation procedure on them. To illustrate the behavior of the wavelet coefficients, I will have a look at a simple example, which I present here.

Let me consider a bivariate FIVARMA(0,$\boldsymbol{d}$,0) process defined as

$$X_a(t) = (\mathbb{1} - \mathbb{L})^{-d_a} u_a(t), \quad a = 1,2,\ t \in \mathbb{Z},$$

with $\begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}$ *i.i.d.* following a Gaussian distribution $\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega}\right)$, with $\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$. The spectral density of (X_1, X_2) satisfies

$$\boldsymbol{f}(\lambda) = \boldsymbol{\Lambda}(\boldsymbol{d})\boldsymbol{\Omega}\overline{\boldsymbol{\Lambda}(\boldsymbol{d})}^\top \text{ where } \boldsymbol{\Lambda}(\boldsymbol{d}) = \text{Diag}\left((1-\mathrm{e}^{-\mathrm{i}\lambda})^{-\boldsymbol{d}}\right).$$

A first order approximation, when $\lambda \to 0^+$, is

$$\begin{aligned}\boldsymbol{f}(\lambda) &\sim \boldsymbol{\Sigma}(\lambda)\\ \boldsymbol{\Sigma}(\lambda) &= \overline{\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d})}^\top \boldsymbol{\Omega}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d}), \text{ with } \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d}) = \text{Diag}\left(\lambda^{-\boldsymbol{d}}\mathrm{e}^{-\mathrm{i}\pi\boldsymbol{d}/2}\right).\end{aligned}$$

Hence, LRD assumption **(M-1)** is satisfied with $\varphi_{1,2} = \frac{\pi}{2}(d_1 - d_2)$.

Let ρ be the correlation coefficient of $\boldsymbol{\Sigma}(\lambda)$, $\rho = \boldsymbol{\Sigma}_{1,2}(\lambda)/\sqrt{\boldsymbol{\Sigma}_{1,1}(\lambda)\boldsymbol{\Sigma}_{2,2}(\lambda)} = \Omega_{1,2}\,\mathrm{e}^{\mathrm{i}\pi(d_1-d_2)/2}$, which does not depend on λ. Later, I will compare the correlation of the wavelet coefficients of $\boldsymbol{X}$ to ρ.

3.1.3 Multivariate fractional Brownian motion

Multivariate fractional Brownian motion (mFBM) is an example of LRD process which does not have a linear representation, even if it can be seen as the limit process of a linear representation (see Amblard, Coeurjolly, Lavancier, and Philippe 2013).

The p-multivariate fractional Brownian motion $\boldsymbol{X} = \{\boldsymbol{X}(t),\, t \in \mathbb{R}\}$ of long-memory parameter $\boldsymbol{d}$, for any $\boldsymbol{d} \in (0.5, 1.5)^p$ is a process satisfying the three following properties:

- $\boldsymbol{X}(t)$ is Gaussian for any $t \in \mathbb{R}$;
- $\boldsymbol{X}$ is self-similar with parameter $\boldsymbol{d} - 1/2$, *i.e.* for every $t \in \mathbb{R}$ and $\gamma > 0$, $(X_1(\gamma t), \dots, X_p(\gamma t))$ has the same distribution as $(\gamma^{d_1-1/2}X_1(t), \dots, \gamma^{d_p-1/2}X_p(t))$;
- the increments are stationary.

Another usual parametrization is the one with Hurst parameters, equal to $\boldsymbol{d} - 1/2$.

We introduce the following quantities, for $1 \leqslant a, b \leqslant p$:

$$\begin{aligned}
\sigma_a &= \mathbb{E}[X_a(1)^2]^{1/2} \\
\rho_{a,b} &= \rho_{a,b} = \mathrm{Cor}(X_a(1), X_b(-1)) \\
\eta_{a,b} &= -\eta_{a,b} = (\mathrm{Cor}(X_a(1), X_b(-1)) - \mathrm{Cor}(X_a(-1), X_b(1)))/c_{a,b} \\
&\qquad \text{with } c_{a,b} = \begin{cases} 2(1 - 2^{d_a+d_b-1}) & \text{if } d_a + d_b \neq 1, \\ 2\log(2) & \text{if } d_a + d_b = 1, \end{cases}
\end{aligned}$$

where $\mathrm{Cor}(X_1, X_2)$ denotes the Pearson correlation between variables X_1 and X_2. The quantities $(\eta_{a,b})_{a,b=1,\dots,p}$ measure the dissymmetry of the process. A mFBM is time reversible if the distribution of $\boldsymbol{X}(-t)$ is equal to the distribution of $\boldsymbol{X}(t)$ for every t. Didier and Pipiras (2011) established that zero-mean multivariate Gaussian stationary processes $\boldsymbol{X}$ is equivalent to $\mathbb{E}[X_a(t)X_b(s)] = \mathbb{E}[X_a(s)X_b(t)]$ for all (s, t), which corresponds to the definition of time reversibility used in Kechagias and Pipiras (2015). A mFBM is time-reversible if and only if $\eta_{a,b} = 0$ for all (a, b).

Coeurjolly, Amblard, and Achard (2013) characterize the spectral behaviour of the increments of a mFBM. If $f_{a,b}^{(1,1)}(\cdot)$ denotes the cross-spectral density of $((\mathbb{1}-\mathbb{L})X_a, (\mathbb{1}-\mathbb{L})X_b)$, then

$$f_{a,b}^{(1,1)}(\lambda) = 2\,\Omega_{a,b}\frac{1-\cos(\lambda)}{|\lambda|^{d_a+d_b}}\mathrm{e}^{\mathrm{i}\phi_{a,b}},$$

with

$$\Omega_{a,b} = \begin{cases} \frac{\sigma_a\sigma_b}{2\pi}\Gamma(d_a+d_b)\left(\rho_{a,b}^2\cos^2(\frac{\pi}{2}(d_a+d_b)) + \eta_{a,b}^2\sin^2(\frac{\pi}{2}(d_a+d_b))\right)^{1/2} & \text{if } d_a+d_b \neq 2 \\ \frac{\sigma_a\sigma_b}{2\pi}\Gamma(d_a+d_b)\left(\rho_{a,b}^2 + \eta_{a,b}^2\frac{\pi^2}{4}\right)^{1/2} & \text{if } d_a+d_b = 2 \end{cases}$$

$$\varphi_{a,b} = \begin{cases} \operatorname{atan}\left(\frac{\eta_{a,b}}{\rho_{a,b}}\tan(\frac{\pi}{2}(d_a+d_b))\right) & \text{if } d_a+d_b \neq 2 \\ \operatorname{atan}\left(\frac{\eta_{a,b}}{\rho_{a,b}}\frac{\pi}{2}\right) & \text{if } d_a+d_b = 2. \end{cases}$$

Let $\boldsymbol{\Theta}$ be given by $\boldsymbol{\Theta} = (\Omega_{a,b}\mathrm{e}^{\mathrm{i}\varphi_{a,b}})_{a,b=1,\ldots,p}$. When λ tends to 0, the spectral density $f_{a,b}^{(1,1)}(\lambda)$ is equivalent to $\Theta_{a,b}|\lambda|^{-(d_a+d_b-2)}$. Thus, assumption **(M-1)** holds. Assumption **(M-2)** is satisfied for any $0 < \beta < 2$. It can be easily verified that time-reversibility is still equivalent to $\varphi_{a,b} = 0$ in this setting.

Note that the set of parameters $\{d_a, \sigma_a, \rho_{a,b}, \eta_{a,b}, a, b = 1, \ldots, p\}$ is not identifiable. Indeed, for $0 < \gamma < 1$, $\{d_a, \sigma_a, \rho_{a,b}, \eta_{a,b}, a, b = 1, \ldots, p\}$ and $\{d_a, \sqrt{\gamma}\,\sigma_a, \rho_{a,b}/\gamma, \eta_{a,b}/\gamma, \ell, m = 1, \ldots, p\}$ lead to the same expressions of $f_{a,b}^{(1,1)}(\cdot)$. It thus seems reasonable to parameterize the fractional Brownian motion by $\{d_a, \Theta_{a,b},\ a, b = 1, \ldots, p\}$.

Discussion on the phase parameter.

The interpretation of the phase parameter is not easy. To better highlight the role of this parameter, I display on Figure 3.6 some bivariate fractional Brownian motions where all the parameters are equal (or very close), except the phase parameter. Since only the phase parameter differs, its influence can be observed in this example.

First, a phase $\phi_{a,b}$ is null when the process $\{(X_a(t), X_b(t)),\ t \in \mathbb{Z}\}$ is time reversible. That is, when the distribution of $(X_a(t), X_b(t))$ equals the distribution of $(X_a(-t), X_b(-t))$ for every $t \in \mathbb{Z}$. In Figure 3.6 (a), the two components of the process $\boldsymbol{X}$ are very similar, and no time drift appears. When the phase $\phi_{a,b}$ increases, even if the long-run correlation remains the same, an asymmetry appears in the cross-correlogram. The higher the phase, the more the two components may move aside one from each other. These phenomenons are well visible with a phase $\phi_{a,b}$ equal to $\pi/2$ as in Figure 3.6 (c).

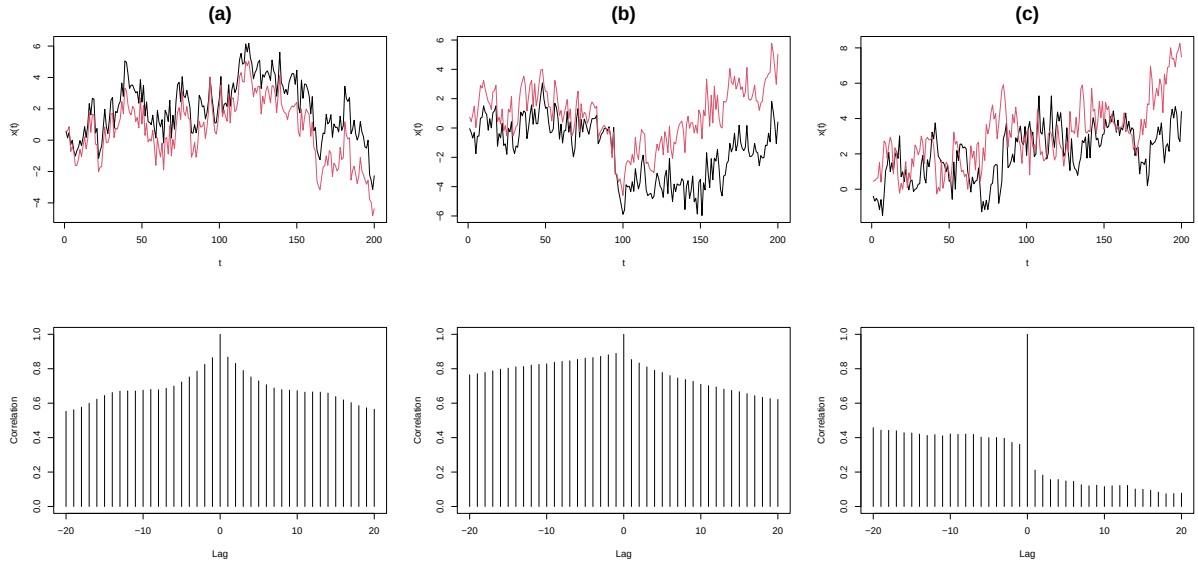


Figure 3.6: Realizations from 3 bivariate fractional Brownian motions (fBM). On the first line, the processes $\boldsymbol{X}(t)$ are represented with respect to t. On the second line, the cross-correlograms are displayed. At a lag $l \in \mathbb{Z}$, the value on the graph is the sample estimation of $Cor(\{(X_1(t), X_2(t+l)),\, t \in \mathbb{Z}\})$ with $N = 500$ observations. Each bivariate fBM has parameters $d_1 = d_2 = 1.25$ and $\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho \simeq 0.67$. Next, from left to right (a) $\phi_{1,2} = 0$ (b) $\phi_{1,2} \simeq 0.138\,\pi (\approx \pi/7)$ (c) $\phi_{1,2} = \pi/2$.

3.2 Wavelet-based estimation with a parametric phase

The objective is to define an estimation procedure in the model described above. To this aim, I introduce a wavelet representation of the time series. I first present the (real) wavelet transform and how it is applied on the data. The behavior of wavelet transform for long-range dependent time series has been studied in Moulines, Roueff, and Taqqu (2007) in a univariate setting. Present work generalizes their result in multivariate settings. I am interested in particular on how the presence of LRD influences the wavelet coefficients.

Next, we propose a local Whittle estimation of the vector $\boldsymbol{d}$ jointly with the estimation of the long-run covariance $\boldsymbol{\Omega}$. A similar Whittle estimation of the LRD parameter d has been derived in Moulines, Roueff, and Taqqu (2008) in a univariate setting. The proposed procedure generalizes it to multivariate time series.

Let $\boldsymbol{X} = \{\boldsymbol{X}(t),\, t \in \mathbb{Z}\}$, with $\boldsymbol{X}(t) = (X_a(t))_{a=1,\dots,p}$, be a multivariate stochastic process. In this

section, I suppose that the cross-spectral density satisfies **(M-1)** with phases $\varphi_{a,b} = \frac{\pi}{2}(d_a - d_b)$:

$$\boldsymbol{f}(\lambda) \sim \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d})\boldsymbol{\Omega}\overline{\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d})}^{\top}, \quad \text{when } \lambda \to 0, \quad \text{with } \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{d}) = \operatorname{Diag}\left(|\lambda|^{-\boldsymbol{d}} e^{-i\operatorname{sign}(\lambda)\pi\boldsymbol{d}/2}\right).$$

This approximation holds for Models A of Lobato (1997), but is not true in general LRD models, see Section 3.1.

3.2.1 Wavelet transform

Let me introduce the wavelet bases and how the time series are decomposed in the bases. Let $(\phi(\cdot), \psi(\cdot))$ be respectively a father and a mother wavelets, satisfying the following regularity conditions:

(W1) The functions $\phi(\cdot)$ and $\psi(\cdot)$ are integrable, have compact supports, $\int_{\mathbb{R}} \phi(t)dt = 1$ and $\int \psi^2(t)dt = 1$;

(W2) There exists $\alpha > 1$ such that $\sup_{\lambda\in\mathbb{R}} |\widehat{\psi}(\lambda)|(1+|\lambda|)^{\alpha} < \infty$, *i.e.* the wavelet is α-regular;

(W3) The mother wavelet $\psi(\cdot)$ has $M > 1$ vanishing moments.

(W4) The function $\sum_{k\in\mathbb{Z}} k^{\ell}\phi(\cdot - k)$ is polynomial with degree ℓ for all $\ell = 1, \dots, M-1$.

Additionally, suppose

(C-a) For all $i = 1, \dots, p$, $(1+\beta)/2 - \alpha < d_i \leqslant M$.

These conditions are not restrictive. Assumptions **(W1)-(W4)** and **(C-a)** hold for example for Daubechies wavelet basis with a sufficiently large number of vanishing moments M.

At a given resolution $j \geqslant 0$, for $k \in \mathbb{Z}$, we define the dilated and translated functions $\phi_{j,k}(\cdot) = 2^{-j/2}\phi(2^{-j}\cdot - k)$ and $\psi_{j,k}(\cdot) = 2^{-j/2}\psi(2^{-j}\cdot - k)$. The wavelet coefficients of the process $\boldsymbol{X}$ are defined by

$$\boldsymbol{W}(j,k) = \int_{\mathbb{R}} \widetilde{\boldsymbol{X}}(t)\psi_{j,k}(t)dt \quad j \geqslant 0, k \in \mathbb{Z},$$

where $\widetilde{\boldsymbol{X}}(t) = \sum_{k\in\mathbb{Z}} \boldsymbol{X}(k)\phi(t-k)$. For given $j \geqslant 0$ and $k \in \mathbb{Z}$, $\boldsymbol{W}(j,k)$ is a p-dimensional vector $\boldsymbol{W}(jk) = \begin{pmatrix} W_1(j,k) & W_2(j,k) & \dots & W_p(j,k) \end{pmatrix}$ where $W_a(j,k) = \int_{\mathbb{R}} \widetilde{X_a}(t)\psi_{j,k}(t)dt$.

In practice, a finite sample of $\{\boldsymbol{X}(t),\ t \in \mathbb{Z}\}$ is observed, say $\{bX(1), \dots \boldsymbol{X}(N)\}$. At each scale $j \geqslant 0$, denote by n_j the number of wavelet coefficients obtained from the observations.

Remark. In this part, the convention used is that the frequencies decrease when the scales j increase. In Part 1 of this manuscript, the inverse convention was used.

3.2.2 Behavior of the wavelet coefficients

I now study the behavior of the wavelet coefficients. For any $j \geqslant 0$, the process $(\boldsymbol{W}(j,k))_{k\in\mathbb{Z}}$ is covariance stationary (Achard and Gannaz 2016a). The wavelet scalogram is

$$\boldsymbol{I}(j) = \sum_{k\in\mathbb{Z}} \boldsymbol{W}(j,k)\boldsymbol{W}(j,k)^\top.$$

The scalogram is the equivalent of the periodogram for the wavelet transform. Yet the scalogram is not normalized, on the contrary of the periodogram. The following proposition is deriving a first order approximation.

Proposition 3.1 (Achard and Gannaz 2016a)**.** *Let $K^{MWW}(\cdot)$ be defined by*

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

Under assumptions **(W1)**-**(W4)** *and* **(C-a)**, *there exists a constant C depending on C_f, β, $\min_i d_i$, $\max_i d_i$, $\boldsymbol{\Omega}$, $\phi(\cdot)$ and $\psi(\cdot)$ such that, for all $j \geqslant 0$, for all $a, b = 1, \ldots, p$,*

$$\left|\mathrm{Cov}\left(W_a(j,k), W_b(j,k)\right) - \Omega_{a,b} 2^{j(d_a+d_b)} \cos(\pi(d_a-d_b)/2) K(d_a+d_b)\right| \leqslant C 2^{(d_a+d_b-\beta)j}. \quad (3.1)$$

Observe that the approximation (3.1) shows that the difficulty with the case of multivariate LRD processes is the appearance of a phase-shift that has to be taken into account for the estimation of the covariance $\boldsymbol{\Omega}$. Indeed, $\mathrm{Cov}\left(W_a(j,k), W_b(j,k)\right)$ is proved to be close to a term proportional to $\cos(\pi(d_a - d_b)/2)$. If the phase in the cosine term is unknown, it can be seen that identifiability problems occur.

Figure 3.7 illustrates the approximation on sample correlations, on simulated processes of Section 3.1.2. It shows the good quality of the approximation of Proposition 3.1 when the scales are high enough.

Define $\boldsymbol{G}^{MWW}(\boldsymbol{d})$, a $p \times p$-matrix with $(a,b)^{\text{th}}$ element equal to

$$G^{MWW}_{a,b}(\boldsymbol{d}) = \Omega_{a,b} K^{MWW}(d_a + d_b) cos(\pi(d_a - d_b)/2). \quad (3.2)$$

When j is high enough,

$$\log_2 \mathrm{Cov}\left(W_a(j,k), W_b(j,k)\right) \approx \log_2(G^{MWW}_{a,b}(\boldsymbol{d})) + (d_a + d_b)\, j\,.$$

A first idea is to replace $\mathrm{Cov}\left(W_a(j,k), W_b(j,k)\right)$ by an approximation, for example by its empirical version $n_j^{-1} I_{a,b}(j)$. See Achard and Gannaz (2019) for an illustration on simulated data. Parameters $\boldsymbol{d}$ can then be estimated with the regression of $(n_j^{-1} I_{a,b}(j))_{j\geqslant 0}$ with respect to the scales j. This approach corresponds to the generalization of the log-scalogram approach in univariate settings (which corresponds to the Wavelet version of the log-periodogram regression, see Table 3.1). Yet in high dimensional frameworks, this approach seems less accurate than Whittle approximation. That is why this estimation scheme has been chosen.

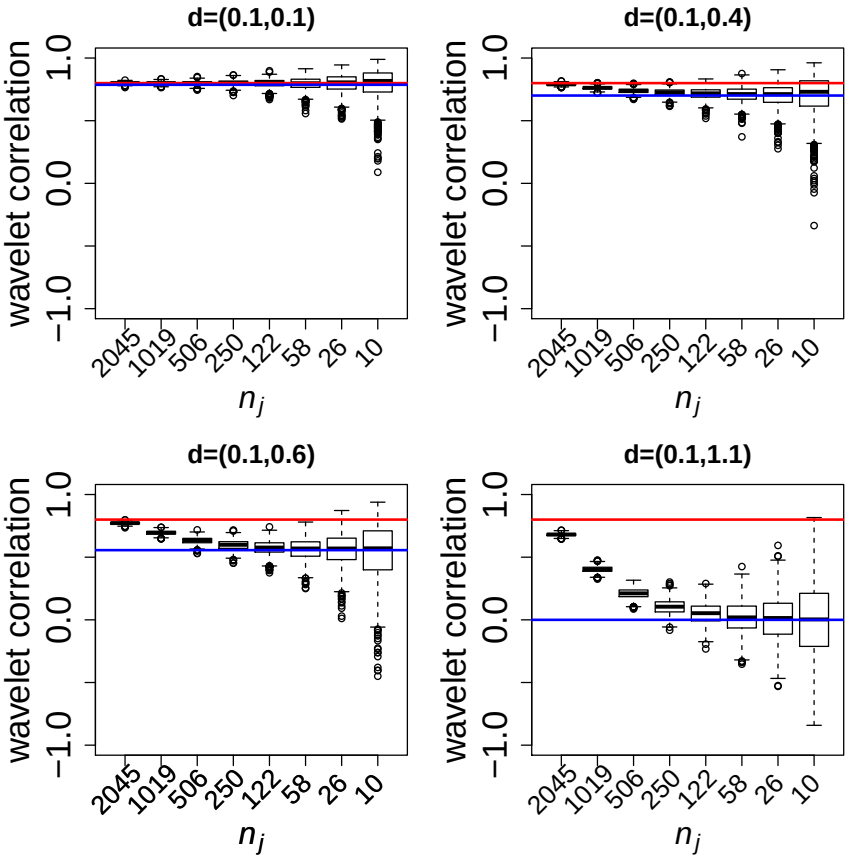


Figure 3.7: Boxplots of the sample wavelet correlations $I_{1,2}(j)/\sqrt{I_{1,1}(j)I_{2,2}(j)}$, at different scales j, for the FIVARMA(0,$\boldsymbol{d}$,0) described in Section 3.1.2, with different values of $\boldsymbol{d}$. Results were obtained on 1000 realizations of $(\boldsymbol{X}(1),\dots,\boldsymbol{X}(N))$ with $N = 2^{12}$, with a bivariate process $\boldsymbol{X}$ being a FIVARMA(0,$\boldsymbol{d}$,0). Number of coefficients available at each scale are given in indexes. The horizontal red line corresponds to the correlation $\rho = \Omega_{1,2}/\sqrt{\Omega_{1,1}\Omega_{2,2}}$ between the two innovations processes. The horizontal blue line corresponds to the first order approximation $\rho\cos(\pi(d_1-d_2)/2)\,K^{MWW}(d_1+d_2)/\sqrt{K^{MWW}(2d_1)K^{MWW}(2d_2)}$.

3.2.3 Local Whittle estimation

Now that the behavior of the wavelet coefficients is known, it is possible to develop a procedure based on them. I present hereafter a multivariate wavelet local Whittle procedure, introduced in Achard and Gannaz (2016a). Only scales $j = j_0,\dots,j_1$ are used in estimation, with j_0 and j_1 to be determined. Let $\boldsymbol{\Lambda}_j(\boldsymbol{d}) = \text{Diag}\left(2^{j\boldsymbol{d}},\ j_0 \leqslant j \leqslant j_1\right)$. The estimators $(\widehat{\boldsymbol{d}}^{MWW}, \widehat{\boldsymbol{G}}^{MWW})$ minimize

	Regression of the log-periodogram	Whittle approximation
Fourier	Geweke and Porter-Hudak (1983)	Fox and Taqqu (1986)
Wavelets	Abry and Veitch (1998) Moulines, Roueff, and Taqqu (2007)	Moulines, Roueff, and Taqqu (2008)

Table 3.1: Main estimation schemes in univariate settings

$\mathcal{L}(\boldsymbol{d}, \boldsymbol{G})$, with

$$\mathcal{L}(\boldsymbol{d}, \boldsymbol{G}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log\det\left(\boldsymbol{\Lambda}_j(\boldsymbol{d}) \boldsymbol{G} \boldsymbol{\Lambda}_j(\boldsymbol{d})\right) + \sum_k \boldsymbol{W}(j,k)^\top \left(\boldsymbol{\Lambda}_j(\boldsymbol{d}) \boldsymbol{G} \boldsymbol{\Lambda}_j(\boldsymbol{d})\right)^{-1} \boldsymbol{W}(j,k) \right]. \quad (3.3)$$

The estimation is here based on a first order approximation of the spectral density matrix at the neighborhood of the zero frequency.

Achard and Gannaz (2016a) establish that the solutions of this problem satisfy

$$\begin{aligned} \widehat{\boldsymbol{d}}^{MWW} &= \underset{\boldsymbol{d}}{\operatorname{argmin}} \log\det(\widehat{\boldsymbol{G}}^{MWW}(\boldsymbol{d})) + 2\log(2) \left(\frac{1}{n}\sum_{j=j_0}^{j_1} j n_j\right)\left(\sum_{a=1}^{p} d_a\right), \\ \widehat{\boldsymbol{G}}^{MWW}(\boldsymbol{d}) &= \frac{1}{n}\sum_{j=j_0}^{j_1} \boldsymbol{\Lambda}_j(\widehat{\boldsymbol{d}})^{-1} \boldsymbol{I}(j) \boldsymbol{\Lambda}_j(\boldsymbol{d})^{-1}, \end{aligned} \quad (3.4)$$

with $n = \sum_{j_0}^{j_1} n_j$. $\boldsymbol{G}^{MWW}(\boldsymbol{d})$ and the long-run covariance matrix can then be estimated by

$$\begin{aligned} \widehat{\boldsymbol{G}}^{MWW} &= \widehat{\boldsymbol{G}}^{MWW}(\widehat{\boldsymbol{d}}^{MWW}) \\ \widehat{\Omega}^{MWW}_{a,b} &= \widehat{G}^{MWW}_{a,b} / (\cos(\pi(\widehat{d}^{MWW}_a - \widehat{d}^{MWW}_b)/2) K^{MWW}(\widehat{d}^{MWW}_a + \widehat{d}^{MWW}_b)), \end{aligned} \quad (3.5)$$

with $a, b = 1, \ldots, p$.

On the contrary to Fourier Whittle estimation (Lobato 1999; Shimotsu 2007), the procedure does not recover directly the long-run covariance matrix $\boldsymbol{\Omega}$. A second step in the estimation procedure is needed in order to correct the phase-shift (given by (3.5)). Indeed, the wavelet-based procedure cannot be achieved in a one-step estimation because wavelet coefficients are real, on the contrary to Fourier coefficients. Hence, the phase-shift is expressed as a multiplicative cosine term in the covariance of the coefficient, which implies a two-step procedure.

Additionnaly, if Lobato (1999) uses the same first-order approximation, Shimotsu (2007) proposes a Fourier-based method which is based on a second-order approximation of the cross-spectral density. This is done through a complex term in $\mathbf{\Lambda}_j(\boldsymbol{d})$. A similar approximation is not possible with a real wavelets procedure.

Let me introduce a condition on the wavelet scalogram:

Condition (C)

$$\text{For all } a,b=1,\dots,p, \quad \sup_n \sup_{j\geqslant 0} \frac{1}{n_j 2^{2j(d_a+d_b)}} Var\left(I_{a,b}(j)\right) < \infty.$$

In Achard and Gannaz (2016a), the consistency and the rate of convergence of the estimators are established.

Theorem 3.2 (Achard and Gannaz 2016a)**.** *Assume that* **(W1)**-**(W4)**, **(C-a)** *and* **Condition (C)** *hold. If j_0 and j_1 are chosen such that $\log(N)^2(2^{-j_0\beta}+N^{-1/2}2^{j_0/2})\to 0$ and $j_0<j_1\leqslant j_N$ then*

$$\widehat{\boldsymbol{d}}^{MWW}-\boldsymbol{d}=O_{\mathbb{P}}(2^{-j_0\beta}+N^{-1/2}2^{j_0/2}),$$

$$\forall (a,b)\in\{1,\dots,p\}^2,\ \widehat{G}^{MWW}_{a,b}-G^{MWW}_{a,b}(\boldsymbol{d})=O_{\mathbb{P}}(\log(N)(2^{-j_0\beta}+N^{-1/2}2^{j_0/2})),$$
$$\widehat{\Omega}^{MWW}_{a,b}-\Omega_{a,b}=O_{\mathbb{P}}(\log(N)(2^{-j_0\beta}+N^{-1/2}2^{j_0/2})).$$

Recall that the influence of the short-range dependence is calibrated by β. The parameter j_0 must be chosen high enough, so that the approximation of Proposition 3.1 is satisfactory (see Figure 5.1). Moreover the highest frequencies are corrupted by the presence of short-range memory, and taking j_0 sufficiently high enables that they do not weight much in the estimation (see Achard and Gannaz 2019). Yet, it also must be taken small enough to ensure that sufficiently information is used in estimation. The optimal rate is then expressed by balancing the two terms appearing in the bound above.

Corollary 3.3 (Achard and Gannaz 2016a)**.** *Assume that* **(W1)**-**(W4)**, **(C-a)** *and* **Condition (C)** *hold. Taking $2^{j_0}=N^{1/(1+2\beta)}$,*

$$\widehat{\boldsymbol{d}}^{MWW}-\boldsymbol{d}=O_{\mathbb{P}}(N^{-\beta/(1+2\beta)}).$$

This corresponds to the optimal rate (Giraitis, Robinson, and Samarov 1997). When $p=1$ we recover the result of Moulines, Roueff, and Taqqu (2008). Fourier Whittle estimators in Lobato (1999) and Shimotsu (2007) obtain the rate $m^{1/2}$ where m is the number of discrete frequencies used in the Fourier transform. When $m\sim cN^{\zeta}$ with a positive constant c, the convergence

is obtained for $0 < \zeta < 2\beta/(1+2\beta)$. Wavelet estimators thus give a slightly better rate of convergence.

Achard and Gannaz (2016a) also provide a setting satisfying **Condition (C)**, which I do not detail here for the sake of simplicity.

3.3 Application

I present here the application on Dataset 3. Results presented in this section are published in Achard and Gannaz (2016b).

For the application on real data, the wavelet Whittle estimation described above is implemented in the `R` package `multiwave`. Achard and Gannaz (2019) describe the practical aspects of the procedure. In particular, the paper highlights how to choose appropriate scales. The idea is to evaluate the distribution of the empirical correlations of the wavelet correlation at each scale, using a temporal bootstrap. As displayed in Figure 3.7, the behaviour of the boxplots of empirical wavelet correlations enables to identify scales where the approximation of Proposition 3.1 are not consistent. These scales should be removed from estimation. Here, the estimation procedure uses scales between 3 and 6, that is to remove frequencies above 0.087Hz. Similar graphical approach on real data is proposed in Achard and Gannaz (2019), using a bootstrap scheme, as illustrated in Figure 3.8.

Figure 3.9 displays the estimated $\boldsymbol{d}$ and $\boldsymbol{\Omega}$ for one subject on one recording. It illustrates that fMRI data indeed has long-range memory and that the estimated long-run covariance matrices are not diagonal.

As the dataset consists of a test-retest paradigm with two recordings for each subject, a way to evaluate the accuracy of the estimator is to evaluate the reproductibility using intra-class correlation. Following Shrout and Fleiss (1979), intra-class correlation (ICC) is computed. ICC is a coefficient smaller than 1 that takes into account the variance within subject in comparison to the variance between subject, defined as,

$$ICC = \frac{s_b - s_w}{s_b + (k-1)s_w} \tag{3.6}$$

where s_b is the variance between subjects, s_w is the variance within subjects and k is the number of sessions per subject. ICC is close to 0 when the reliability is low, and close to 1 when the reliability is high. ICC defined as (3.6) can have negative values but this reflects a wrong behaviour of the data set.

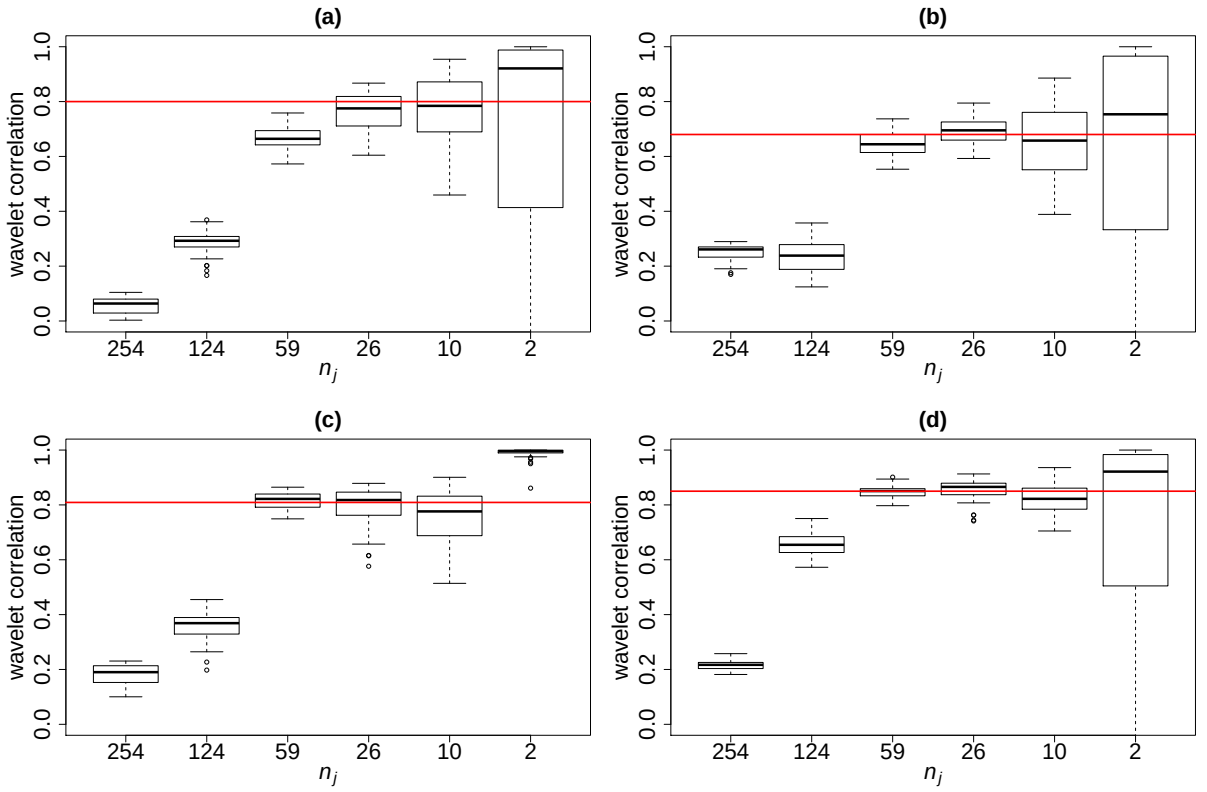


Figure 3.8: Boxplots of the correlation of the wavelet coefficients at different scales for real time series from on subject of fMRI Dataset 3: (a) Time series 1 and 2; (b) Time series 13 and 14; (c) Time series 31 and 32; (d) Time series 47 and 48. Boxplots were obtained using sub-time series with N points, extracted from two fMRI time series with length equal to 1200 points, from a single subject. The estimated long parameters d of the two time series are equal. The index of the horizontal axis displays the number of coefficients available. With this graphic, scales 3 to 6 were chosen for estimation. The horizontal red lines represent the estimated long-run correlation. Calculation was done on $N = 512$ observations and among 100 replications using sliding windows (with overlap).

Figure 3.10 shows the results obtained using the univariate estimator and mutivariate estimator. This result suggests that multivariate estimations are more reproducible in a test-retest paradigm than univariate estimations.

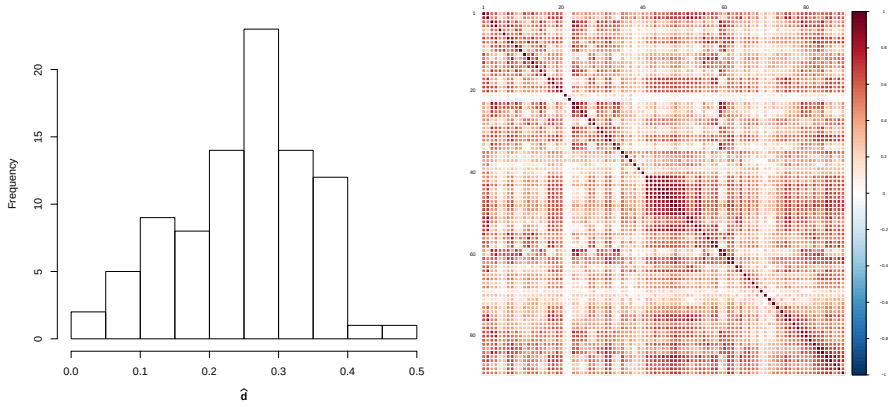


Figure 3.9: Histogram of $\widehat{\mathbf{d}}^{MWW}$ and of $\widehat{\mathbf{\Omega}}^{MWW}$ from one subject of fMRI Dataset 3.

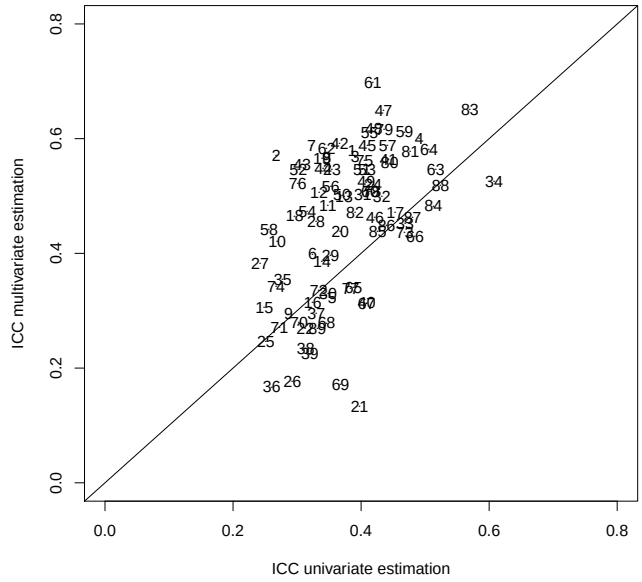


Figure 3.10: Estimation of ICC using multivariate or univariate estimations in fMRI Dataset 3.

References

Abry, P. and D. Veitch (1998). "Wavelet analysis of long-range-dependent traffic". In: *Information Theory, IEEE Transactions on* 44.1, pp. 2–15.

Achard, S. and I. Gannaz (2016a). "Multivariate Wavelet Whittle Estimation in Long-range Dependence". In: *Journal of Time Series Analysis* 37.4, pp. 476–512.

Achard, S. and I. Gannaz (2016b). "Wavelet Whittle Estimation in Multivariate Time Series Models: Application to fMRI Data". In: *Conference of the International Society for Non-Parametric Statistics*. Springer, pp. 271–285.

Achard, S. and I. Gannaz (2019). "Wavelet-Based and Fourier-Based Multivariate Whittle Estimation: multiwave". In: *Journal of Statistical Software* 89.6, pp. 1–31.

Amblard, P.-O., J.-F. Coeurjolly, F. Lavancier, and A. Philippe (2013). "Basic properties of the Multivariate Fractional Brownian Motion". In: *Séminaires et congrès*. Vol. 28, pp. 65–87.

Baek, C., S. Kechagias, and V. Pipiras (2020). "Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity". In: *Journal of Statistical Planning and Inference* 205, pp. 245–268.

Coeurjolly, J.-F., P.-O. Amblard, and S. Achard (2013). "Wavelet analysis of the multivariate fractional Brownian motion". In: *ESAIM: Probability and Statistics* 17, pp. 592–604.

Didier, G. and V. Pipiras (2011). "Integral representations and properties of operator fractional Brownian motions". In: *Bernoulli* 17.1, pp. 1–33.

Fox, R. and M. S. Taqqu (1986). "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series". In: *The Annals of Statistics*, pp. 517–532.

Gençay, R., F. Selçuk, and B. Whitcher (2001). *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press.

Geweke, J. and S. Porter-Hudak (1983). "The estimation and application of long memory time series models". In: *Journal of Time Series Analysis* 4.4, pp. 221–238.

Giraitis, L., P. M. Robinson, and 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, pp. 49–60.

Kechagias, S. and V. Pipiras (2015). "Definitions and Representations of Multivariate Long-Range Dependent Time Series". In: *Journal of Time Series Analysis* 36.1, pp. 1–25.

Kechagias, S. and V. Pipiras (2020). "Modeling bivariate long-range dependence with general phase". In: *Journal of Time Series Analysis* 41.2, pp. 268–292.

Lobato, I. N. (1997). "Consistency of the averaged cross-periodogram in long memory series". In: *Journal of Time Series Analysis* 18.2, pp. 137–155.

Lobato, I. N. (1999). "A semiparametric two-step estimator in a multivariate long memory model". In: *Journal of Econometrics* 90.1, pp. 129–153.

Maxim, V., L. Şendur, M. J. Fadili, J. Suckling, R. Gould, R. Howard, and E. T. Bullmore (2005). "Fractional Gaussian noise, functional MRI and Alzheimer's disease". In: *NeuroImage* 25, pp. 141–158.

Moulines, E., F. Roueff, and M. S. Taqqu (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, pp. 155–187.

Moulines, E., F. Roueff, and M. S. Taqqu (2008). "A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series". In: *The Annals of Statistics* 36.4, pp. 1925–1956.

Nielsen, M. Ø. and P. H. Frederiksen (2005). "Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration". In: *Econometric Reviews* 24.4, pp. 405–443.

Papanicolaou, G. C. and K. Sølna (2003). "Wavelet based estimation of local Kolmogorov turbulence". In: *Theory and Applications of Long-range Dependence*, pp. 473–505.

Percival, D. B. and A. T. Walden (2006). *Wavelet methods for time series analysis*. Vol. 4. Cambridge University Press.

Sela, R. J. and C. M. Hurvich (2008). "Computationaly efficient methods for two multivariate fractionnaly integrated models". In: *Journal of Time Series Analysis* 30, p. 6.

Sela, R. J. and C. M. Hurvich (2012). "The averaged periodogram estimator for a power law in coherency". In: *Journal of Time Series Analysis* 33.2, pp. 340–363.

Shimotsu, K. (2007). "Gaussian semiparametric estimation of multivariate fractionally integrated processes". In: *Journal of Econometrics* 137.2, pp. 277–310.

Shrout, P. E. and J. L. Fleiss (1979). "Intraclass correlations: uses in assessing rater reliability." In: *Psychological bulletin* 86.2, p. 420.

Termenon, M., S. Achard, A. Jaillard, and C. Delon-Martin (2016). "The "Hub Disruption Index", a Reliable Index Sensitive to the Brain Networks Reorganization. A Study of the Contralesional Hemisphere in Stroke". In: *Frontiers in Computational Neuroscience* 10.

Whitcher, B. and M. J. Jensen (2000). "Wavelet estimation of a local long memory parameter". In: *Exploration Geophysics* 31.1/2, pp. 94–103.

CHAPTER 4

TESTING IN MULTIVARIATE LONG-RANGE DEPENDENCE MODELS FOR WAVELET-BASED WHITTLE ESTIMATORS

Dataset 4

In this chapter, I will consider fMRI recordings on rats. Dead rats and live rats under anesthetics were scanned. I will focus on one dead rat and one live rat (under anesthetic). The data are freely available at https://zenodo.org/record/2452871 (Becq, Barbier, and Achard 2020; Becq, Habet, et al. 2020). The duration of the scanning is 30 minutes with a time repetition of 0.5 second so that 3,600 time points are available at the end of experience. After preprocessing as described in Pawela et al. (2008), 51 time series were extracted, for each rat, each time series being associated with a brain region of rats.

fMRI recordings of brain activity are based on the hemodynamic response to a magnetic field, which may create some temporal and spatial dependence. They suffer from different sources of noise, including system-related instabilities, subject motion, or physiological fluctuations (Liu 2016). Additionally, the preprocessing step aggregates the time series of each voxel to obtain a unique time series for each brain region. This aggregation step may create LRD properties (Leipus, Philippe, Puplinskaite, and Surgailis 2013). My claim is that long-range dependence and long-run covariance are closely related to brain activity and not to recording artifacts or preprocessing. I would like to check this assertion on the dataset.

Motivation

In the previous chapter, a scheme for the estimation of the long-run correlation matrices between time recordings is described. The next step is to deduce a graph modeling of dependencies between cerebral regions. Such a graph characterizes the links between the anatomical

parts of the brain, called the brain connectivity. Numerous clinical results have been obtained for the characterization of loss of connectivity in different pathologies (*e.g.* for Alzheimer's disease, Buckner et al. 2009), or of changes in the network characteristics (*e.g.* for disorders of consciousness, Achard, Delon-Martin, et al. 2012).

Let V denote the set of the indexes of brain regions, $V = \{1, \dots, p\}$. Suppose that we have estimated a correlation matrix between the activity of brain regions $(r_{a,b})_{(a,b)\in V\times V}$. The objective is to infer the dependence graph $\mathcal{G} = (V, E)$ with vertices V and where the set of edges E is a subset of $V \times V$ defined by $E = \{(a, b) : r_{a,b} \neq 0\}$.

A first approach consists in penalizing the local Whittle objective function (3.3). Baek, Kechagias, and Pipiras (2017) propose a ℓ^1-penalty. Yet, Lasso regression and graphical Lasso estimation suffer from bad support recoveries in simulation studies (see *e.g.* respectively Meinshausen, Meier, and Bühlmann 2009; Villers, Schaeffer, Bertin, and Huet 2008). That is, many estimated edges are false positives. Theoretically, the control is ensured under an incoherence assumption (Wainwright 2009; Ravikumar, Wainwright, Raskutti, Yu, et al. 2011) which is very restrictive in practice. For example Meinshausen and Bühlmann (2006) and Ravikumar, Wainwright, Raskutti, Yu, et al. (2011) put in evidence some graph structures where Lasso type estimation fails. Moreover, our dataset does not suffer from too high dimension and the estimation is consistent. Consequently, a multiple testing method is preferred.

To estimate the edges, we consider, for all $(a, b) \in V \times V$, the tests

$$(H_{0,a,b})\ r_{a,b} = 0 \ \text{ against } \ (H_{1,a,b})\ r_{a,b} \neq 0 \tag{4.1}$$

where $r_{a,b}$ denotes a correlation between a variable a and a variable b, corresponding to nodes V. The set of edges is then estimated as $\{(a, b) \in V \times V \mid (H_{0,a,b}) \text{ is rejected}\}$. For a given graph $\mathcal{G}$, estimation $\widehat{\mathcal{G}}$ is obtained applying $m = p\,(p-1)/2$ tests (since the set E is symmetric by symmetry of the correlations – the edges are undirected). Such an approach for graph inference has been proposed in Drton and Perlman (2007).

The main point for this step is to develop a testing procedure. Hence, first, asymptotic normality of the estimators is proved. Next, multiple testing corrections are proposed, since a high number of tests have to be done. For example, in Dataset 4, $m = 1275$ tests have to be done for each rat.

Contributions

The objective is to develop testing procedures, on the LRD parameters and on the long-run covariance, in the setting of Chapter 3. To develop testing on the parameters, the first step is to

obtain the (asymptotic) distribution of the estimators. Gannaz (2023) [1] proves the asymptotic normality of the sample wavelet covariances, and of the estimators $(\widehat{\boldsymbol{d}}^{MWW}, \widehat{\boldsymbol{G}}^{MWW})$ defined in Chapter 3. The expression of the asymptotic variance matrices are explicitly given, allowing to build test procedures. Implementation has been done (with the use of `RcppParallel` since calculating of the formulas is time consuming) and I plan to add the code to the `R` package *multiwave* [24].

Next, in the neuroscience application, many tests have to be done (in Dataset 4, 51 tests for LRD parameters and 51.50/2=1275 for long-run correlations, for each rat, since there are 51 brain regions). It is, hence, necessary to consider multiple testing corrections. Achard, Borgnat, and Gannaz (2020) [28] review existing Family Wise Error Rate (FWER) controlling multiple testing procedures for Gaussian statistics, where p-value processes are correlated. They establish that the control holds asymptotically. The application to correlation testing is specified, in view of the desired application. A simulation study highlights the behavior of the procedures, among other their stability with respect to the sparsity of the correlation matrix. The `R` package `TestCor` [23] implements the multiple testing procedures described in Achard, Borgnat, and Gannaz (2020) [28] in the specific case of correlations testing.

This work was partly carried out during Marine Roux's PhD, Roux (2018), which I co-supervised. In her PhD, she also explored the control of the False Discovery Rate (FDR). She studied in particular the behavior of Benjamini and Hochberg (1995)'s procedure in two-sided tests for the mean of a vector-valued Gaussian variable. Some extensions about Reiner-Benaim (2007) and Cohen (2016) are given.

In parallel, since we want to apply the procedures on time series data, with dependence, Achard, Borgnat, Gannaz, and Roux (2019) [12] explore the influence of short-range and long-range properties when testing wavelets correlations.

The chapter is organized as follows. First, I give some results about the asymptotic normality of the estimators developed in Chapter 3, which allow to build tests on the parameters. The second section reviews multiple testing procedures that can be apply in this context. These two sections are used to infer connectivity graphs, and a short discussion on this construction is given in Section 4.3. The last section presents the application of the tests procedure to a dead rat and a live rat of Dataset 4.

4.1 Asymptotic normality

The main step for developing testing procedures is to establish the asymptotic distribution of the estimators. It can be shown that the wavelet-based Whitthle estimators described Chapter 3 are asymptotically Gaussian. This result is obtained in Gannaz (2023). I here sum up some results. They are an extension to a multivariate setting of Roueff and Taqqu (2009a). The proofs are very similar, and they are based on the results on decimated linear processes given in Roueff and Taqqu (2009b).

Let me consider again the framework of Chapter 3. That is, let $\boldsymbol{X}$ be a p-multivariate process satisfying **(M-1)** and **(M-2)**. $\boldsymbol{X}$ is a LRD process with LRD parameters $\boldsymbol{d}$ and a complex long-run covariance $\boldsymbol{\Theta} = \boldsymbol{\Omega} e^{i\boldsymbol{\Phi}}$, with $\boldsymbol{\Omega}$ non-negative real symmetric matrix and $\boldsymbol{\Phi} = (\frac{\pi}{2}(d_a - d_b))_{a,b=1,\dots,p}$. Asymptotic normality is obtained for LRD processes admitting a linear representation. Let me thus introduce an additional assumption on the process $\boldsymbol{X}$.

(M-3) There exists a sequence $\{\boldsymbol{A}^{(\mathbf{D})}(s)\}_{s\in\mathbb{Z}}$ in $\mathbb{R}^{p\times p}$ such that $\sum_{s\in\mathbb{Z}} \max_{a,b=1,\dots,p} |A^{(\mathbf{D})}_{a,b}(s)|^2 < \infty$ and

$$\forall t \in \mathbb{Z},\ \left((\mathbb{1} - \mathbb{L})^{D_a} X_a(t)\right)_{a=1,\dots,p} = \sum_{s\in\mathbb{Z}} \boldsymbol{A}^{(\mathbf{D})}(t+s)\mathbf{u}(s)$$

with $\mathbf{u}(t)$ weak white noise process, in $\mathbb{R}^p$. Let $\mathcal{F}(t-1)$ denote the σ-field of events generated by $\{\mathbf{u}(s),\, s \leqslant t-1\}$. Assume that $\mathbf{u}$ satisfies $\mathbb{E}[\mathbf{u}(t)|\mathcal{F}(t-1)] = 0$, $\mathbb{E}[u_a(t)u_b(t)|\mathcal{F}(t-1)] = \mathbb{1}_{a=b}$ and $\mathbb{E}[u_a(t)u_b(t)u_c(t)u_d(t)|\mathcal{F}(t-1)] = \mu_{a,b,c,d}$ with $|\mu_{a,b,c,d}| \leqslant \mu_\infty < \infty$, for all $a, b, c, d = 1, \dots, p$.

For all $(a, b) \in \{1, \dots, p\}^2$, for all $\lambda \in \mathbb{R}$, the sequence $(2^{-j d_a} |A^{(\mathbf{D})*}_{a,b}(2^{-j}\lambda)|)_{j\geqslant 0}$ is convergent as j goes to infinity.

The estimation process of $(\boldsymbol{d}, \boldsymbol{\Omega})$ is based on the scheme defined in Chapter 3. A wavelet pair $(\phi(\cdot), \psi(\cdot))$ satisfying assumptions **(W1)**–**(W4)** are used. Recall that the approximation of the covariance of the wavelet coefficients is given by Proposition 3.1, with a key parameter $\boldsymbol{G}^{MWW}(\boldsymbol{d})$ given by (3.2). Estimators $(\widehat{\boldsymbol{d}}^{MWW}, \widehat{\boldsymbol{G}}^{MWW}, \widehat{\boldsymbol{\Omega}}^{MWW})$ are the local wavelet Whittle estimation described in Section 3.2. I suppose hereafter that the assumptions of Theorem 5.4 are satisfied, so that the estimators are consistent.

Next, I need to define some quantities which appear into the asymptotic expression of the distributions. Even if the expressions are very complicated, I choose to present them, to show that they are calculable, and because their form brings information on the between-scale correlations.

Technical quantities.
For $u \geqslant 0$, $(\delta_1, \delta_2) \in (-\alpha, M)^2$, let me define $\widetilde{I}_u(\delta_1, \delta_2)$ as

$$\widetilde{I}_u(\delta_1, \delta_2) = \frac{2\pi}{K(\delta_1)K(\delta_2)} \int_{-\pi}^{\pi} \overline{\widetilde{D}_{u,\infty}(\lambda; \delta_1)} \widetilde{D}_{u,\infty}(\lambda; \delta_2) \, \mathrm{d}\lambda \,,$$

where $\widetilde{D}_{u,\infty}(\lambda; \delta) = \sum_{\tau=0}^{2^{-u}-1} D_{u,\tau}(\lambda; \delta)$, and

$$D_{u,\tau}(\lambda; \delta) = \sum_{t \in \mathbb{Z}} |\lambda + 2t\pi|^{-\delta} \overline{\widehat{\psi}(\lambda + 2t\pi)} \, 2^{u/2} \widehat{\psi}(2^u(\lambda + 2t\pi)) \, \mathrm{e}^{-\mathrm{i}2^u\tau(\lambda+2t\pi)}.$$

Expression of the variance of the LRD parameter estimators $\widehat{d}$.
I introduce

$$\mathcal{I}_\Delta^d(\delta_1, \delta_2) = \frac{2}{\kappa_\Delta} \widetilde{I}_0(\delta_1, \delta_2) \tag{4.2}$$
$$+ \frac{2}{\kappa_\Delta^2} \sum_{u=1}^{\Delta} (2^{u\delta_1} + 2^{u\delta_2}) \, 2^{-u} \frac{2 - 2^{-\Delta+u}}{2 - 2^{-\Delta}} ((u + \eta_{\Delta-u} - \eta_\Delta)(\eta_{\Delta-u} - \eta_\Delta) + \kappa_{\Delta-u}) \, \widetilde{I}_u(\delta_1, \delta_2)$$
$$\text{if } \Delta < \infty,$$

$$\mathcal{I}_\infty^d(\delta_1, \delta_2) = \widetilde{I}_0(\delta_1, \delta_2) + \sum_{u=1}^{\infty} (2^{u\delta_1} + 2^{u\delta_2}) \, 2^{-u} \, \widetilde{I}_u(\delta_1, \delta_2) \,, \quad \text{if } \Delta = \infty. \tag{4.3}$$

The terms in the summation part in the expressions above result from the between-scale correlation of the wavelet representation. These correlation is supposed negligible in some works, for example in Whitcher, Guttorp, and Percival (2000). Define now

$$\boldsymbol{G} \boldsymbol{.} \mathcal{I}^{\boldsymbol{d}} \boldsymbol{.} \boldsymbol{G}(\Delta) = \mathrm{Diag}\,(\mathrm{vec}(\boldsymbol{G})) \left(\mathcal{I}_\Delta^d(d_a + d_b, d_{a'} + d_{b'})_{(a,b),(a',b') \in \{1,\dots,p^2\}}\right) \mathrm{Diag}\,(\mathrm{vec}(\mathbf{G})) \,, \tag{4.4}$$

where $\mathrm{vec}(\boldsymbol{M})$ denotes the operation that transforms a matrix $\boldsymbol{M}$ in $\mathbb{R}^{p \times p}$ to a vector in $\mathbb{R}^{p^2}$. Similarly, the terms in the summation part in the expressions above result from the between-scale correlation. The asymptotic variance of the LRD estimators $\widehat{\boldsymbol{d}}$ will be of the form $\mathbf{V}^{(\boldsymbol{d})}(\Delta)$ with

$$\mathbf{V}^{(\boldsymbol{d})}(\Delta) = \frac{1}{2 \log(2)^2} (\boldsymbol{G}^{-1} \circ \boldsymbol{G} + \boldsymbol{I}_p)^{-1} \, \mathbf{Y}(\Delta) \, (\boldsymbol{G}^{-1} \circ \boldsymbol{G} + \boldsymbol{I}_p)^{-1}, \tag{4.5}$$

where $\boldsymbol{I}_p$ is the identity matrix in $\mathbb{R}^{p \times p}$ and with entry (a, a') of $\mathbf{Y}^{(\Delta)}$, for $(a, a') \in \{1, \dots, p\}^2$, given by

$$\mathrm{Y}_{a,a'}(\Delta) = \sum_{b,b'=1,\dots,p} (\boldsymbol{G}^{-1})_{a,b} (\boldsymbol{G}^{-1})_{a',b'} \left(\boldsymbol{G} \boldsymbol{.} \mathcal{I}^{\boldsymbol{d}} \boldsymbol{.} \boldsymbol{G}_{(a,a'),(b,b')}(\Delta) + \boldsymbol{G} \boldsymbol{.} \mathcal{I}^{\boldsymbol{d}} \boldsymbol{.} \boldsymbol{G}_{(a,b'),(a',b)}(\Delta)\right). \tag{4.6}$$

Expression of the variance of the normalized long-run covariance estimators $\widehat{G}$.
Denote

$$\mathcal{I}_{\Delta}^{G}(\delta_1,\delta_2) = \widetilde{I}_0(\delta_1,\delta_2) + \sum_{u=1}^{\Delta}(2^{u\delta_1}+2^{u\delta_2})2^{-u}\,\frac{2-2^{-\Delta+u}}{2-2^{-\Delta}}\,\widetilde{I}_u(\delta_1,\delta_2) \qquad \text{if } \Delta<\infty, \quad (4.7)$$

$$\mathcal{I}_{\infty}^{G}(\delta_1,\delta_2) = \widetilde{I}_0(\delta_1,\delta_2) + \sum_{u=1}^{\infty}(2^{u\delta_1}+2^{u\delta_2})2^{-u}\,\widetilde{I}_u(\delta_1,\delta_2) \qquad \text{if } \Delta=\infty. \quad (4.8)$$

Let me also define

$$\boldsymbol{G}\,\boldsymbol{.}\,\mathcal{I}^{\boldsymbol{G}}\,\boldsymbol{.}\,\boldsymbol{G}(\Delta) \;=\; \mathrm{Diag}\left(\mathrm{vec}(\boldsymbol{G})\right)\left(\mathcal{I}_{\Delta}^{G}(d_a+d_b,d_{a'}+d_{b'})_{(a,b),(a',b')\in\{1,\dots,p^2\}}\right)\mathrm{Diag}\left(\mathrm{vec}(\boldsymbol{G})\right).$$

The asymptotic variance of the normalized long-run covariance estimators $\widehat{\boldsymbol{G}}$ will be of the form $\mathbf{V}^{(\boldsymbol{G})}(\Delta)$ with

$$V^{(\boldsymbol{G})}_{(a,b),(a',b')}(\Delta) = \boldsymbol{G}\,\boldsymbol{.}\,\mathcal{I}^{\boldsymbol{G}}\,\boldsymbol{.}\,\boldsymbol{G}_{(a,a'),(b,b')}(\Delta) + \boldsymbol{G}\,\boldsymbol{.}\,\mathcal{I}^{\boldsymbol{G}}\,\boldsymbol{.}\,\boldsymbol{G}_{(a,b'),(a',b)}(\Delta) \quad (4.9)$$

The asymptotic normality of the estimators of the long-range dependence modeling is established by the next theorem.

Theorem 4.1 (Gannaz 2023)**.** *Suppose that conditions of Theorem 3.2 are satisfied and that assumption* **(M-3)** *hold. Let $j_0 < j_1 \leqslant j_N$ with $j_N = \max\{j, n_j \geqslant 1\}$ such that*

$$j_1 - j_0 \to \Delta \in \{1,\dots,\infty\},\ \log(N_X)^2(N_X 2^{-j_0(1+2\beta)} + N_X^{-1/2}2^{j_0/2}) \to 0.$$

Let $n = \sum_{j=j_0}^{j_1} n_j$ be the total number of (real) wavelet coefficients used in estimation. Then,

- *$\sqrt{n}(\widehat{\boldsymbol{d}}^{MWW} - \boldsymbol{d})$ converges in distribution to a centered Gaussian distribution with a variance equal to $\mathbf{V}^{MWW\,(\boldsymbol{d})}(\Delta)$, defined by* (4.5) *with $\mathbf{G} = \boldsymbol{G}^{MWW}(\boldsymbol{d})$, whenever it appears.*
- *$\mathrm{vec}\left(\sqrt{n}\left(\widehat{\boldsymbol{G}}^{MWW} - \boldsymbol{G}^{MWW}(\boldsymbol{d})\right)\right)$ converges in distribution to a centered Gaussian distribution with a variance equal to $\mathbf{V}^{MWW,\boldsymbol{G}(\Delta)}$, with $\mathbf{V}^{MWW\,\boldsymbol{G}(\Delta)}$ defined by* (4.9) *with $\mathbf{G} = \boldsymbol{G}^{MWW}(\boldsymbol{d})$, whenever it appears.*

A similar result on long-run correlations, rather than long-run covariances, can be deduced.

Corollary 4.2 (Gannaz 2023)**.** *Let $(a,b) \in \{1,\dots,p\}^2$, $a \neq b$. Define*

$$\widehat{r}^{MWW}_{a,b} = \frac{\widehat{G}^{MWW}_{a,b}}{\sqrt{\widehat{G}^{MWW}_{a,a}\widehat{G}^{MWW}_{b,b}}} \text{ and } r^{MWW}_{a,b} = \frac{G^{MWW}_{a,b}(\boldsymbol{d})}{\sqrt{G^{MWW}_{a,a}(\boldsymbol{d})G^{MWW}_{b,b}(\boldsymbol{d})}}.$$

Then, under conditions of Theorem 4.1,

$$\sqrt{n}\,\left(\widehat{r}_{a,b}^{MWW} - r_{a,b}^{MWW}\right) \xrightarrow[j\to\infty]{\mathcal{L}} \mathcal{N}\left(0, V_{a,b}^{(\mathbf{r})}(\Delta)\right)$$

with

$$\begin{aligned} V_{a,b}^{(\mathbf{r})}(\Delta) = {} & \mathcal{I}_\Delta^G(2d_a, 2d_b) + \mathcal{I}_\Delta^G(d_a+d_b, d_a+d_b)(r_{a,b}^2 + r_{a,b}^4) \\ & - \left(\mathcal{I}_\Delta^G(2d_a, 2d_b) + \mathcal{I}_\Delta^G(2d_b, d_a+d_b)\right) 2\, r_{a,b}^2 - \left(I_\Delta^G(2d_a, 2d_a) + \mathcal{I}_\Delta^G(2d_b, 2d_b)\right) r_{a,b}^2/2, \end{aligned}$$

with$\mathcal{I}_\Delta^G(\cdot,\cdot)$ *defined in* (4.2)-(4.3)*. When all off-diagonal entries of* $\boldsymbol{G}^{MWW}(\boldsymbol{d})$ *are equal to 0,*

$$\sqrt{n}\ \mathrm{vec}\Big(\widehat{r}_{a,b}^{MWW},\ 1 \leqslant a < b \leqslant p\Big) \xrightarrow[j\to\infty]{\mathcal{L}} \mathcal{N}_{p(p-1)/2}\Big(0, \mathrm{Diag}\left(vec\big(I_\Delta^G(2d_a, 2d_b),\ 1 \leqslant a < b \leqslant p\big)\right)\Big).$$

Quantities $\mathcal{I}_\Delta(\delta_1, \delta_2)$ are (very) intricate, but they are computable for given δ_1, δ_2, Δ. Hence, by plugging into (4.5)–(4.6) consistent estimators of $\boldsymbol{d}$ and $\boldsymbol{G}^{MWW}(\boldsymbol{d})$, for example $\widehat{\boldsymbol{d}}^{MWW}$ and $\widehat{\boldsymbol{G}}^{MWW}$ as proved in Theorem 3.2, a test procedure on parameters $\boldsymbol{d}$ can be built. Similarly, tests on $\{\widehat{r}_{a,b}^{MWW},\ 1 \leqslant a < b \leqslant p\}$ follow from Corollary 4.2.

4.2 Review of FWER controlling procedures

Based on the results of previous section, it is possible to test both the LRD parameters and the log-run correlation, as presented in the motivations. Let me give uniform notations.

Let $\boldsymbol{\theta} = (\theta_\ell)_{\ell=1,\dots,m}$ denote indifferently $(d_a)_{a=1,\dots,p}$ or $(r_{a,b}^{MWW})_{1\leqslant a<b\leqslant p}$.

Let $\mathbb{W}^{(n)} = (W_1, \dots, W_n)$ be the wavelet coefficients of a LRD process. Observing $\mathbb{W}^{(n)}$, the aim is to test, for all $\ell \in \{1, \dots, m\}$,

$$(H_{0,\ell})\ \theta_\ell = 0 \ \text{ against } \ (H_{1,\ell})\ \theta_\ell \neq 0. \tag{4.10}$$

For all $\ell \in \{1, \dots, m\}$, let $T_{n,\ell}$ be test statistics built on $\mathbb{W}^{(n)}$. If $(\theta_\ell)_{\ell=1,\dots,m}$ denote the LRD parameters $(d_a)_{a=1,\dots,p}$, the statistics are $T_{n,\ell} = \sqrt{n}\,\widehat{d}_\ell^{MWW} \cdot \widehat{\mathbf{V}}^{MWW\,\boldsymbol{d}}(\Delta)_{\ell,\ell}^{-1/2}$, with $\widehat{\mathbf{V}}^{MWW\,\boldsymbol{d}}(\Delta)$ estimator of $\mathbf{V}^{MWW\,\boldsymbol{d}}(\Delta)$, obtained by plugging $\widehat{\boldsymbol{d}}^{MWW}$ and $\widehat{\boldsymbol{G}}^{MWW}$ into (4.5). Alternatively, if $(\theta_\ell)_{\ell=1,\dots,m}$ denote the correlation coefficients $(r_{a,b}^{MWW})_{1\leqslant a<b\leqslant p}$, the statistics are $T_{n,\ell} = \sqrt{n}\,\widehat{r}_{a,b}^{MWW} \cdot I_\Delta^G(2\widehat{\boldsymbol{d}}_a, 2\widehat{\boldsymbol{d}}_b)^{-1/2}$ with $I_\Delta^G(\cdot,\cdot)$ defined in (4.7) - (4.8), and a re-indexing from$\{(a,b),\ 1 \leqslant a < b \leqslant p\}$ to $\{\ell = 1, \dots, m\}$.

For all $\ell \in \{1, \dots, m\}$, test statistics $T_{n,\ell}$ are of the form $T_{n,i} = \sqrt{n}\,\widehat{\theta}_{n,\ell}\big(\mathbb{W}^{(n)}\big)$. Statistics $\widehat{\theta}_{n,\ell}\big(\mathbb{W}^{(n)}\big)$ are consistent estimator of quantities $\widetilde{\theta}_\ell$, for $\ell = 1, \dots, m$, and they have an asymptotic Gaussian distribution. Namely,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{n,\cdot}(\mathbb{W}^{(n)}) - \widetilde{\boldsymbol{\theta}}\right) \xrightarrow{d} \mathcal{N}_m(0, \boldsymbol{\Sigma}), \quad \text{when } n \to +\infty, \tag{4.11}$$

where $\xrightarrow{d}$ denotes the convergence in distribution. Assume that $\boldsymbol{\Sigma}$ is invertible. Quantities $\widetilde{\boldsymbol{\theta}}$ are renormalizations of $\boldsymbol{\theta}$, given respectively by $d_a \cdot V^{MWW\,d}(\Delta)_{a,a}^{-1/2}$ and $(\mathbf{r}_{a,b} \cdot I_\Delta^G(d_a, d_b)^{-1/2}$, $1 \leqslant a < b \leqslant p$. It is hence equivalent to test (4.10) considering $\boldsymbol{\theta}$ or $\widetilde{\boldsymbol{\theta}}$. The statistics have been built such that $\Sigma_{ii} = 1$ when $\theta_i = 0$. That is, every statistic is normalized under the null hypothesis of (4.10), so that $\mathrm{Var}(T_{n,\ell}) = 1$, for all $\ell \in \{l,\ H_{0l} \text{ holds}\}$.

In Dataset 4, which motivates this work, there are $p = 51$ brain regions. Hence, there are $m = p = 51$ tests (4.10) when considering LRD parameters $\{d_\ell,\ \ell = 1, \dots, p\}$, and $m = p(p-1)/2 = 1275$ tests when considering long-run correlations $\{r_{a,b}^{MWW},\ 1 \leqslant a < b \leqslant p\}$. If no multiple correction is applied, the probability of having false positives is very high, in particular for the correlation tests. This motivates a review of multiple testing procedures which can be applied in our setting.

The objective of multiple testing procedure is to give a rejection set

$$\mathcal{R} = \{\ell,\ 1 \leqslant \ell \leqslant m : (H_{0,\ell}) \text{ rejected}\},$$

such that the error is controlled. I will consider here the type I error called Family Wise Error Rate (FWER), defined as

$$\mathrm{FWER}(\mathcal{R}) = \mathbb{P}\left(\exists \ell \in \mathcal{R} : \theta_\ell = 0\right).$$

To control the FWER for a given level $\alpha \in [0, 1]$, the objective is to find a procedure yielding a rejection set $\mathcal{R}$ such that $\mathrm{FWER}(\mathcal{R}) \leqslant \alpha$.

Let $(p_{n,\ell})_{1 \leq \ell \leq m}$ be a family of p-values resulting from each m individual tests. The asymptotic Gaussian property (4.11) gives rise to the asymptotic p-value process:

$$\forall \ell \in \{1, \dots, m\},\ p_{n,\ell} = 2\left(1 - F_{\mathcal{N}}\left(\left|T_{n,i}\right|\right)\right), \tag{4.12}$$

where $F_{\mathcal{N}}$ is the standard Gaussian cumulative distributive function. Multiple testing procedures will be based on this p-value process.

Let $\mathcal{H}_0 = \{\ell \in \{1, \dots, m\} : \theta_\ell = 0\}$ be the index set of true null hypotheses, that is, the index set of all ℓ such that $H_{0,\ell}$ is satisfied. Denote $m_0 = |\mathcal{H}_0|$, its cardinality. The FWER corresponds to the probability of rejecting at least one true null hypothesis, namely,

$$\mathrm{FWER}(\mathcal{R}, P) = \mathbb{P}(|\mathcal{R} \cap \mathcal{H}_0| \geq 1).$$

The FWER depends also on the (unknown) distribution of the observations, here the distribution of $\mathbb{W}^{(n)}$, but this dependence is omitted for clarity.

Since we consider an asymptotic p-value process, we can only get asymptotic results in terms of control of the errors.

Definition 4.3. *A multiple testing procedure $\mathcal{R}$ is said to asymptotically control the FWER at level α if*

$$\limsup_{n\to+\infty} FWER(\mathcal{R}) \leq \alpha.$$

I review three multiple testing procedures which asymptotically control FWER for the two-sided testing problem (4.10), based on the asymptotic p-value process (4.12). I also present an additional procedure, useful in the general framework of tests on correlations.

Bonferroni

The Bonferroni procedure, Bonferroni (1935), is the most classical example of FWER control.

Method 1 (Bonferroni)**.** *The Bonferroni multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^{bonf} = \left\{\ell \in \{1,\dots,m\} : p_{n,\ell} \leq \frac{\alpha}{m}\right\}.$$

This procedure does not require any assumption on the dependence structure of the p-values, however under strong dependence the Bonferroni correction is known to be conservative, see *e.g.* Bland and Altman (1995).

Šidák

As mentioned by Westfall and Young (1993), an asymptotic FWER controlling procedure can be derived by Šidák's inequality (Šidák 1967). For the specific case of correlation testing, Drton and Perlman (2004) use this inequality to construct a procedure that asymptotically controls the FWER for the problem (4.10).

Method 2 (Šidák)**.** *Let $c_\alpha^s = \Phi^{-1}\left(\frac{1}{2}(1-\alpha)^{1/m} + \frac{1}{2}\right) > 0$. The Šidák's multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^s = \left\{\ell \in \{1,\dots,m\} : |T_{n,\ell}| > c_\alpha^s\right\}.$$

The procedure is valid for any dependencies, but only in Gaussian setting. The Šidák's procedure is less conservative than the Bonferroni procedure.

Parametric bootstrap

Drton and Perlman (2007) detail a parametric bootstrap method for testing (4.10) on partial correlation coefficients. This method differs from Romano and Wolf (2005). Indeed, the quantile is here evaluated on the asymptotic distribution rather than the empirical distribution. An estimation of the matrix $\mathbf{\Sigma}$ (defined in (4.11)) is needed. Denote by $\widehat{\mathbf{\Sigma}}_n$ such an estimator.

Method 3 (MaxT)**.** *Let $t_{n,\alpha}(\widehat{\mathbf{\Sigma}}_n)$ be the $(1-\alpha)$-quantile of the distribution of $\|Z(\widehat{\mathbf{\Sigma}}_n)\|_\infty$, with $Z(\widehat{\Sigma}_n) \sim \mathcal{N}_m(0, \widehat{\mathbf{\Sigma}}_n)$. The MaxT multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^{MaxT} = \left\{ \ell \in \{1, \dots, m\} : |T_{n,\ell}| > t_{n,\alpha}\left(\widehat{\mathbf{\Sigma}}_n\right) \right\},$$

where $t_{n,\alpha}\left(\widehat{\mathbf{\Sigma}}_n\right)$ is computed using (simulated) samples of $\mathcal{N}_m(0, \widehat{\mathbf{\Sigma}}_n)$.

Procedure MaxT is available with any consistent estimation of $\mathbf{\Sigma}$. However, in practice, the quality of estimation may influence the quality of the procedure for a given number n of observations.

Non parametric bootstrap

I here present a method which is also reviewed in Achard, Borgnat, and Gannaz (2020). Yet, if it can be applied in a usual correlation test framework, the application here is not possible. Indeed, it supposes that the observations are independent. Romano and Wolf (2005) propose an asymptotic FWER controlling procedure which only requires a monotonic assumption on the family of thresholds.

Method (BootRW)**.** *Suppose that $\mathbb{W}^{(n)}$ is a sample of n independent observations. Let $t_{n,\alpha}(\Sigma)$ be the $(1-\alpha)$-quantile of the probability distribution of $\|Z(\Sigma)\|_\infty$ for $Z(\Sigma)$ random variable following a $\mathcal{N}_m(0, \Sigma)$.*

The Romano-Wolf's multiple testing procedure is defined by

$$\mathcal{R}_\alpha^{BootRW} = \left\{ i \in \{1, \dots, m\} : |T_{n,i}| > \widehat{t}_{n,\alpha}(\Sigma) \right\},$$

where $\widehat{t}_{n,\alpha}(\Sigma)$ is an estimation of $t_{n,\alpha}(\Sigma)$ computed using bootstrap resamples of $\mathbb{W}^{(n)}$. A bootstrap resample from independent and identically distributed (i.i.d.) $\mathbb{W}^{(n)}$ is defined as an n i.i.d. sample from the empirical distribution of $\mathbb{W}^{(n)}$.

In high dimensional settings, Chernozhukov, Chetverikov, and Kato (2013) suggest to use rather a so-called wild bootstrap approach, which I do not present here. As stated before,

this approach is not applicable directly in our neuroscience setting, since the observations $\mathbb{W}^{(n)}$ are not independent. A bootstrap procedure adapted to the functional nature of the data must be provided.

Methods 1 to 3 control asymptotically the FWER for problems (4.10). That is, $\limsup_{n\to+\infty} \text{FWER}(\mathcal{R}) \leq \alpha$ for $\mathcal{R} \in \{\mathcal{R}_\alpha^{bonf}, \mathcal{R}_\alpha^{s}, \mathcal{R}_\alpha^{MaxT}\}$.

Step-down algorithms are also proposed to improve these single-step methods (see Section 2.2 of Achard, Borgnat, and Gannaz 2020). The principle is to iterate multiple testing on the non-rejected hypothesis. Such methods increase the power of the procedures, still preserving FWER control.

A special focus is given in Achard, Borgnat, and Gannaz (2020) on testing on correlation coefficients from Gaussian samples. The four methods described above can be applied in this context. Various test statistics are also available, and we provide a simulation study to compare the alternative choices. Since we are motivated by the inference of a graph structure, the simulation study is built on consequence. We also established a link with Bickel and Levina (2008)'s thresholding for regularization in high-dimensional settings.

4.3 Discussion on graph inference

Recall that tests on the long-run correlations (4.1) lead to a graph representation of the connections between the brain regions. Let $\mathcal{G}$ denote the graph of connections between brain regions $V = \{1, \dots, p\}$; $\mathcal{G} = (V, E)$ with the set of edges $E \subseteq V \times V$ corresponding to regions which are connected given by $E = \{(a, b),\ r_{a,b}^{MWW} \neq 0\}$. The brain connectivity graph $\mathcal{G}$ in this framework is an undirected graph.

Remark that, due to the cosine term in $\boldsymbol{G}^{MWW}(\boldsymbol{d})$ (see (3.2)), r_{ab}^{MWW} can be equal to 0 if the phase $\phi_{a,b}$ is equal to $\pi/2$, for $a, b = 1, \dots, p$. Hence the graph does not correspond to the graph with edges $\{(a, b),\ \Omega_{a,b} \neq 0\}$, which seems more natural. This bias will be corrected in Chapter 5.

Our procedure is based on correlations, rather than partial or conditional correlations. Recall that, by Hammersley & Clifford's theorem, the graph based on conditional correlations is the only dependence graph which satisfies the pairwise Markov property (when the variables admit a positive density – and the three usual Markov properties if the density is continuous by Pearl & Paz's theorem), see Lauritzen (1996). Hence, this motivates to consider conditional correlations rather than correlations. The problem is that estimating conditional correlation is not as easy as estimating correlations. Actually, graphical models are often based on partial correlations (Rothman, Bickel, Levina, Zhu, et al. 2008; Ravikumar, Wainwright, Raskutti, Yu,

et al. 2011). Recall that a partial correlation between variables $X^{(a)}$ and $X^{(b)}$ is the correlation between the projections of $X^{(a)}$ and $X^{(b)}$ on the space spanned by $\{X^{(l)}, l \in V \setminus \{a,b\}\}$. The partial covariance matrix, called precision matrix, is, in fact, equal to the inverse of the covariance matrix. It is, hence, natural in a Gaussian setting, since the likelihood writes more easily with the precision matrix. Moreover, with Gaussian distributions, resulting partial correlations are equal to conditional correlations. Yet, this is not the case with other distributions (Baba, Shibata, and Sibuya 2004). It is, nevertheless, possible to build graphs based on partial correlations rather than correlations, considering the inverse of the matrix $\mathbf{\Omega}$. *A priori*, the step from correlations to partial correlations is not high and tests can also be built. I have chosen to focus here on correlations, which are easier to handle with. Observe that it would be interesting to estimate both graphs, since they bring different (and complementary) information.

Next, the graph is estimated by the control of the FWER, which is a type I error. Controlling the FWER at level α means that the probability to have the estimated graph $\widehat{\mathcal{G}}$ included in $\mathcal{G}$ is asymptotically greater than $1-\alpha$. The probability to have an additional false edge is controlled, but there is no guarantee about the number of missing edges. This later would corresponds to a type II error. This choice of FWER seems adequate in our context due to the fact that graph characteristics appear more perturbed by additional edges than missing edges (see Perspectives and other ongoing works, Figure P.2).

Finally, I just give an informal insight on the interpretation of the graph $\mathcal{G}$. What follows is not rigorous. Suppose that the process $\boldsymbol{X}$ is associated to a second-order stationary innovation process $\mathbf{u}$, obtained by getting rid of the LRD properties,

$$\mathrm{Diag}\left((\mathbb{1}-\mathbb{L})^{\boldsymbol{d}}\right)\boldsymbol{X}(t) = \mathbf{u}(t),\ t \in \mathbb{Z}.$$

The phase term is here $\varphi_{a,b} = \frac{\pi}{2}(d_a - d_b)$, $a,b = 1,\dots,p$ and the long-run covariance matrix $\mathbf{\Omega}$ only depends on $\mathbf{u}$. Next, consider the case when $\mathbf{u}$ is a (stationary) Markov process $\mathbf{u}(t+1) = \boldsymbol{A}\mathbf{u}(t)$. The matrix $\boldsymbol{A}$ corresponds to the weights of a directed graph, and provides the temporal dependence between $\mathbf{u}(t)$ and $\mathbf{u}(t+1)$. The autocorrelation writes as $\gamma(k) = \mathbb{E}(\mathbf{u}(t)\mathbf{u}(t+k)^\top) = \mathbb{E}(\mathbf{u}(0)\mathbf{u}(0)^\top)\boldsymbol{A}^{k\top}$, for $k \in \mathbb{N}$. The exponent $\top$ denotes the transpose operator. Hence, for $(a,b) \in \{1,\dots,p\}^2$,

$$\Omega_{a,b} = \frac{\mathbb{E}(\mathbf{u}(0)\mathbf{u}(0)^\top)}{2\pi}\Big(\boldsymbol{I}_p + \sum_{k\in\mathbb{N}\setminus\{0\}}\big((\boldsymbol{A}^k)_{a,b} + (\boldsymbol{A}^k)_{b,a}\big)\Big).$$

The value $\Omega_{a,b}$ is the sum of the influences of all paths from a to b, and of paths from b to a. In conclusion, this illustrates that the graph obtained from $\{r_{a,b},\ 1 \leqslant a < b \leqslant p\}$ is indeed a long-run graph. It corresponds intuitively to the dependence summed over all time between the components of the innovation process, that is, if long-run dependence was erased from the time series.

4.4 Application

Consider Dataset 4. The objective is to test either there is long-range dependence or cerebral connectivity in dead rats recordings. I expect that $\boldsymbol{d} = 0$ and that $\boldsymbol{G}$ is a diagonal matrix for a dead rat but not for a live one. This would mean that LRD properties and long-run correlation are linked with the brain activity.

As explained above, cerebral connectivity is inferred by testing on long-run correlation coefficients. Parameters $\boldsymbol{d}$ and $\boldsymbol{G}$ are inferred by wavelet-based Whittle estimation, using the procedure defined in Chapter 3. The scales in the estimation are chosen with the bootstrap method described in Section 3.3. Estimation is performed taking $j_0 = 4$ and $j_1 = 9$, which is the maximal scale; that is, the frequencies above 0.12 Hz are removed.

Based on Theorem 4.1, for each rat, one can test if the LRD parameters are significant for each brain region. That is, for all $a = 1, \dots, p$, I test

$$(H_{0\,a}^{(\boldsymbol{d})})\ d_a = 0 \ \text{ against } \ (H_{1\,a}^{(\boldsymbol{d})})\ d_a \neq 0\,,$$

replacing $\boldsymbol{d}$ and $\boldsymbol{G}$ respectively by $\widehat{\boldsymbol{d}}^{MWW}$ and $\widehat{\boldsymbol{G}}^{MWW}$ in $\mathbf{V}^{MWW(\boldsymbol{d})}$. As described in Section 4.2, three methods are available, Bonferroni, Šidák, and parametric bootstrap. I choose here to apply Bonferroni's correction. Even if it is conservative, it is the simplest approach. Hence, I consider a level $\alpha' = 5\%$ and apply Bonferroni's multiple testing correction, *i.e.* each test is applied with a level α'/p to ensure that the probability to have a false positive on the p tests is equal to α'.

Next Corollary 4.2 allows to test the significance of the long-run correlation between each pair of brain regions. For all $1 \leqslant a < b \leqslant p$, I test

$$(H_{0\,a,b}^{(\mathbf{r})})\ r_{a,b}^{MWW} = 0 \ \text{ against } \ (H_{1\,a,b}^{(\mathbf{r})})\ r_{a,b}^{MWW} \neq 0\,.$$

Similarly, I apply Bonferroni's multiple testing correction and I consider a level $\alpha'/\ (p(p-1)/2)$ for each test.

The tests have been applied on one dead rat and one live rat. The results are displayed in Figure 4.1 as graphs. Figure 4.1 shows that, indeed, we can conclude that $\boldsymbol{d} = 0$ and that off-diagonal entries of $\boldsymbol{G}$ are equal to zero for the dead rat. For the live rat, six brain regions (over $p = 51$) have a significant LRD parameter, and 483 correlations (over the 1275 values of $\{r_{a,b},\ 1 \leqslant a < b \leqslant p\}$) are significant. These observations tend to confirm that long-range dependence and long-run covariance result from brain activity.

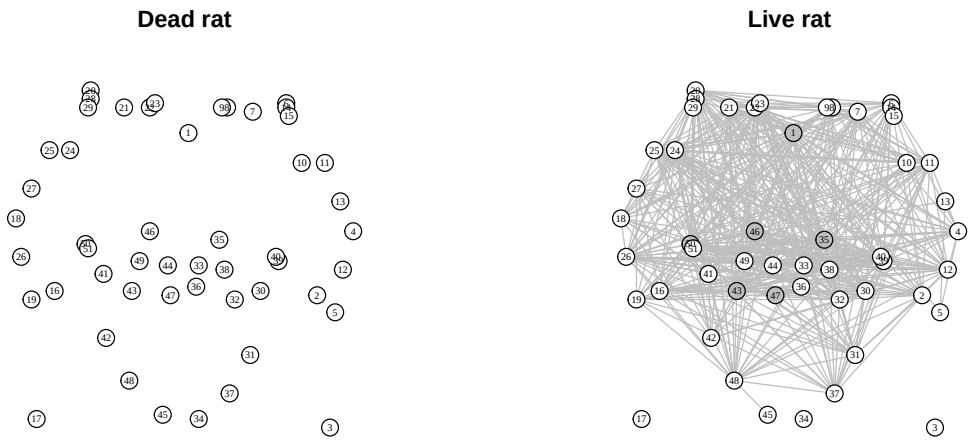


Figure 4.1: Inferred graphs of cerebral activity for a dead rat (left) and a live rat (right). Each vertex of the graph corresponds to a brain region. Colored vertices are regions where the LRD parameter d_a is significant, *i.e.* where the null hypothesis $(H_{0\,a}^{(\mathbf{d})})$ is rejected. Two vertices a, b are connected by an edge if the long-run correlation is significant, *i.e.* if the null hypothesis $(H_{0\,a,b}^{(\mathbf{r})})$ is rejected.

Remark on the quality of the recordings

Dataset 4 contains 4 dead rats. Hence, it is possible to apply the procedure for each of the dead rats. Surprisingly, for some rats, identifiability problems appear, due to some high values of LRD. Looking at the recordings, it appears that the quality is not satisfactory for some rats. An example of an abnormal recording for a dead rat is displayed in Figure 4.2. In [28], we have focused on wavelet scale 4, corresponding to the frequency interval [0.06 ; 0.12] Hz. This reduces the influence of measurement errors. Nevertheless, when applying the wavelet-based Whittle estimation, results may not be accurate. This is why, in this manuscript, only the results for one dead rat are presented, where the recordings look reliable. In my opinion, other experiments should be conducted, with no brain activity, to confirm the assertion that LRD is only due to brain activity and that fMRI recordings do not create artificial connections.

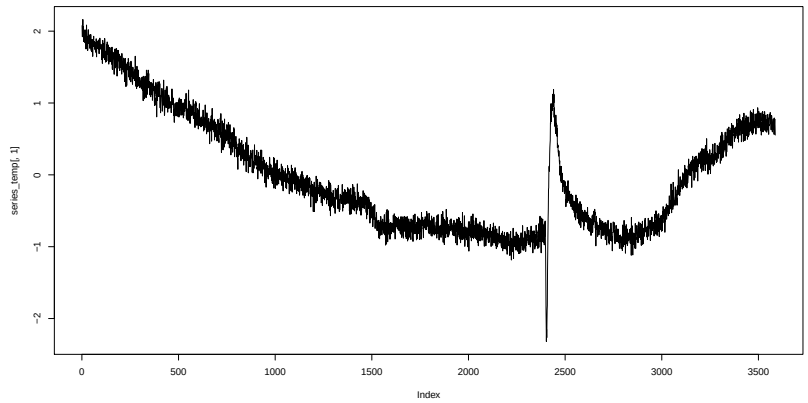


Figure 4.2: Example of an abnormal time series obtained for the cerebral region of a dead rat.

References

Achard, S., P. Borgnat, and I. Gannaz (2020). "Asymptotic control of FWER under Gaussian assumption: application to correlation tests". working paper or preprint.

Achard, S., P. Borgnat, I. Gannaz, and M. Roux (2019). "Wavelet-based graph inference using multiple testing". In: *Wavelets and Sparsity XVIII*. Vol. 11138. International Society for Optics and Photonics, p. 1113811.

Achard, S., C. Delon-Martin, P. E. Vértes, F. Renard, M. Schenck, F. Schneider, C. Heinrich, S. Kremer, and E. T. Bullmore (2012). "Hubs of brain functional networks are radically reorganized in comatose patients". In: *Proceedings of the National Academy of Sciences* 109.50, pp. 20608–20613.

Baba, K., R. Shibata, and M. Sibuya (2004). "Partial correlation and conditional correlation as measures of conditional independence". In: *Australian & New Zealand Journal of Statistics* 46.4, pp. 657–664.

Baek, C., S. Kechagias, and V. Pipiras (2017). "Semiparametric, parametric, and possibly sparse models for multivariate long-range dependence". In: *Wavelets and Sparsity XVII*. Vol. 10394. International Society for Optics and Photonics, 103941S.

Becq, G., E. Barbier, and S. Achard (2020). "Brain networks of rats under anesthesia using resting-state fMRI: comparison with dead rats, random noise and generative models of networks". In: *Journal of Neural Engineering*.

Becq, G., T. Habet, N. Collomb, M. Faucher, C. Delon-Martin, V. Coizet, S. Achard, and E. L. Barbier (2020). "Functional connectivity is preserved but reorganized across several anesthetic regimes". In: *NeuroImage* 219, p. 116945.

Benjamini, Y. and 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)*, pp. 289–300.

Bickel, P. J. and E. Levina (2008). "Covariance regularization by thresholding". In: *The Annals of Statistics* 36.6, pp. 2577–2604.

Bland, J. M. and D. G. Altman (1995). "Multiple significance tests: the Bonferroni method". In: *Bmj* 310.6973, p. 170.

Bonferroni, C. E. (1935). *Il calcolo delle assicurazioni su gruppi di teste*. Tipografia del Senato.

Buckner, R. L., J. Sepulcre, T. Talukdar, F. M. Krienen, H. Liu, T. Hedden, J. R. Andrews-Hanna, R. A. Sperling, and K. A. Johnson (2009). "Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease". In: *Journal of neuroscience* 29.6, pp. 1860–1873.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". In: *The Annals of Statistics* 41.6, pp. 2786–2819.

Cohen, R. (2016). *Finding maximximum FDR, at two tailed hypotheses testing in a cas that the test statistics are normally distributed*. internal communication.

Drton, M. and M. D. Perlman (2004). "Model selection for Gaussian concentration graphs". In: *Biometrika* 91.3, pp. 591–602.

Drton, M. and M. D. Perlman (2007). "Multiple testing and error control in Gaussian graphical model selection". In: *Statistical Science*, pp. 430–449.

Gannaz, I. (2023). "Asymptotic normality of wavelet covariances and multivariate wavelet Whittle estimators". In: *Stochastic Processes and their Applications* 155, pp. 485–534.

Lauritzen, S. L. (1996). *Graphical models*. Vol. 17. Clarendon Press.

Leipus, R., A. Philippe, D. Puplinskaite, and D. Surgailis (2013). "Aggregation and long memory: recent developments". In: *Journal of the Indian Statistical Association* 52.1, pp. 71–101.

Liu, T. T. (2016). "Noise contributions to the fMRI signal: An overview". In: *NeuroImage* 143, pp. 141–151.

Meinshausen, N. and P. Bühlmann (2006). "High-dimensional graphs and variable selection with the Lasso". In: *Annals of Statistics* 34.3, pp. 1436–1462.

Meinshausen, N., L. Meier, and P. Bühlmann (2009). "P-values for high-dimensional regression". In: *Journal of the American Statistical Association* 104.488, pp. 1671–1681.

Pawela, C. P., B. B. Biswal, Y. R. Cho, D. S. Kao, R. Li, S. R. Jones, M. L. Schulte, H. S. Matloub, A. G. Hudetz, and J. S. Hyde (2008). "Resting-state functional connectivity of the rat brain". In: *Magnetic Resonance in Medicine* 59.5, pp. 1021–1029.

Ravikumar, P., M. J. Wainwright, G. Raskutti, B. Yu, et al. (2011). "High-dimensional covariance estimation by minimizing $\ell_1$1-penalized log-determinant divergence". In: *Electronic Journal of Statistics* 5, pp. 935–980.

Reiner-Benaim, A. (2007). "FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis". In: *Biometrical Journal* 49.1, pp. 107–126.

Romano, J. P. and M. Wolf (2005). "Exact and approximate stepdown methods for multiple hypothesis testing". In: *Journal of the American Statistical Association* 100.469, pp. 94–108.

Rothman, A. J., P. J. Bickel, E. Levina, J. Zhu, et al. (2008). "Sparse permutation invariant covariance estimation". In: *Electronic Journal of Statistics* 2, pp. 494–515.

Roueff, F. and M. S. Taqqu (2009a). "Asymptotic normality of wavelet estimators of the memory parameter for linear processes". In: *Journal of Time Series Analysis* 30.5, pp. 534–558.

Roueff, F. and M. S. Taqqu (2009b). "Central limit theorems for arrays of decimated linear processes". In: *Stochastic processes and their applications* 119.9, pp. 3006–3041.

Roux, M. (2018). "Graph inference by multiple testing with application to Neuroimaging". Ph.D. Université Grenoble Alpes, France.

Šidák, Z. (1967). "Rectangular confidence regions for the means of multivariate normal distributions". In: *Journal of the American Statistical Association* 62.318, pp. 626–633.

Villers, F., B. Schaeffer, C. Bertin, and S. Huet (2008). "Assessing the validity domains of graphical Gaussian models in order to infer relationships among components of complex biological systems". In: *Statistical Applications in Genetics and Molecular Biology* 7.2.

Wainwright, M. J. (2009). "Sharp thresholds for High-Dimensional and noisy sparsity recovery using ℓ_1-Constrained Quadratic Programming (Lasso)". In: *IEEE transactions on information theory* 55.5, pp. 2183–2202.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279. John Wiley & Sons.

Whitcher, B., P. Guttorp, and D. B. Percival (2000). "Wavelet Analysis of Covariance with Application to Atmospheric Time Series". In: *Journal of Geophysical Research* 105(D11).14, pp. 941–962.

CHAPTER 5

COMPLEX WAVELET-BASED INFERENCE FOR MULTIVARIATE LONG-RANGE TIME SERIES WITH A GENERAL PHASE

Dataset 4 - extension

Hereafter, I still will consider the fMRI recordings on rats presented in the previous chapter, but I will consider additional data on various anesthetics. 4 rats were scanned dead, and 7 live rats were scanned for each of the 4 anesthetics considered in the study. Recall that the duration of the scanning is 30 minutes with a time repetition of 0.5 second so that $3,600$ time points are available, and that 51 time series, each time series being associated with a brain region of rats, are extracted.

Motivations

A limitation of the estimation procedure developed in the two previous chapters is that it is available only for when the phase is equal to $\varphi_{a,b} = \frac{\pi}{2}(d_a - d_b)$, for all $(a, b) \in \{1, \dots, p\}^2$, in the spectral approximation **(M-1)**. The modeling described in Section 3.1, yet, shows the limitation of this assumption. Moreover, it is expected that the anesthetics influence not only the magnitude but also the *shape* of the coupling between brain regions. This may have a consequence on the phases $\varphi_{a,b}$ appearing in **(M-1)**. An extension to general phases would be welcome. As the wavelet transform applied in Chapter 3 is real, it is not possible to recover the imaginary part of the spectral density, on the contrary to Fourier transform. As a consequence, we want to use a complex-valued representation of time-series. The use of wavelets is motivated by their flexibility for real data application, including their ability to consider non stationary time series. We therefore introduce complex wavelets. In a first step, we study the properties of the *common-factor* complex wavelets. Next we use them for inference on time series.

Contributions

The objective is to propose a flexible representation of the signals, which can deal with non stationarity, but do not loss information about the phase. Fourier periodograms well recover the structure of the long-run covariance $\mathbf{\Theta}$, but cannot be used under with stationarity. With real wavelets, non stationarity can be considered. But as shown in Chapter 3, the covariances of the wavelet coefficient only captures the real part of the long-run covariance matrix. Hence, we have proposed a complex wavelet based procedure, which combines advantages of both representations above.

Selesnick (2001) proposes complex wavelets, called *Common-factor wavelets*, which are quasi-analytic. They present good empirical properties (in particular relatively to quasi-analyticity). Nevertheless, for their use in a statistical procedure, a theoretical control of these properties is needed. This is done in Achard, Clausel, Gannaz, and Roueff (2020) [4].

Next, Gannaz, Achard, Clausel, and Roueff (2017) [14] illustrate on simulations that Common-factor wavelets indeed provide a promising transform for inference with multivariate time series.

Finally, a statistical application of Common-factor wavelets is provided in Achard and Gannaz (2022) [10], where they are used for the inference of time series characteristics. An approximation of the sample wavelet covariance similar to the one established in Achard and Gannaz (2016) [6] is proved, and the Whittle procedure is shown to be consistent. The results are obtained under more restrictive conditions, since the choice of the wavelet filters is not adaptive. But this approach allows to consider non stationary multivariate time series within a larger scope than real wavelets. The procedure is in particular adequate to handle with multivariate fractional Brownian motions. The procedure has been implemented. I plan to add the code of the procedure described in this chapter to package `multiwave` [24].

The chapter is organized as follows. The first section presents Selesnick (2001)'s *common-factor* complex wavelets. The definition is recalled as well as our main result about the quality of the analytic approximation of the filters. The second section gives the estimation procedure and its consistency. Finally, I describe the application on Dataset 4.

5.1 Common-factor complex wavelets

The problem encountered with the estimation developed in Section 3.2 is that the approximation of the covariance of the coefficients corresponds to the real part of the long-run covariance.

Using a complex transform will enable to recover also its imaginary part. We hence introduce complex wavelets.

Let $(h^{(L)}, h^{(H)})$ and $(g^{(L)}, g^{(H)})$ denote two pairs of respectively low-pass and high-pass filters. Let $(\phi_h(\cdot), \psi_h(\cdot))$ be respectively the father and mother wavelets associated to $(h^{(L)}, h^{(H)})$. They can be defined through their Fourier transforms as

$$\widehat{\phi}_h(\omega) = 2^{-1/2} \prod_{j=1}^{\infty} \left[2^{-1/2} \widehat{h}^{(L)}(2^{-j}\omega) \right] \quad \text{and} \quad \widehat{\psi}_h(\omega) = 2^{-1} \widehat{h}^{(H)}(\omega/2)\, \widehat{\phi}_h(\omega/2). \tag{5.1}$$

Let us define similarly (ϕ_g, ψ_g) the father and the mother wavelets associated with the wavelet filters $g^{(L)}$ and $g^{(H)}$. Their Fourier transforms are

$$\widehat{\phi}_h(\omega) = 2^{-1/2} \prod_{j=1}^{\infty} \left[2^{-1/2} \widehat{g}^{(L)}(2^{-j}\omega) \right] \quad \text{and} \quad \widehat{\psi}_h(\omega) = 2^{-1} \widehat{g}^{(H)}(\omega/2)\, \widehat{\phi}_g(\omega/2). \tag{5.2}$$

The complex father and mother wavelets $(\phi(\cdot), \psi(\cdot))$ are then defined by

$$\widehat{\phi}(\omega) = \widehat{\phi}_h + \mathrm{i}\, \widehat{\phi}_g \quad \text{and} \quad \widehat{\psi}(\omega) = \widehat{\psi}_h + \mathrm{i}\, \widehat{\psi}_g. \tag{5.3}$$

To recover efficiently the long-run covariance, we would like to have an analytic wavelet ψ. A wavelet $\psi(\cdot)$ is said analytic if its Fourier transform is only supported on the positive frequency semi-axis. It is sufficient that the pair $(\psi_g(\cdot), \psi_h(\cdot))$ forms a Hilbert pair, that is, if $\widehat{\psi}_g(\lambda) = -\mathrm{i}\ \mathrm{sign}(\lambda)\widehat{\psi}_h(\lambda)$, for all $\lambda \in \mathbb{R}$, where $\mathrm{sign}(\lambda)$ denotes the sign function taking values $-1, 0$ and 1 for $\lambda < 0$, $\lambda = 0$ and $\lambda > 0$, respectively.

In our application, it is preferable that the wavelets have a compact support (to control the number of coefficients). From Paley-Wiener's theorem, analytic filters with finite support do not exist. Selesnick (2001) proposes compact wavelet filters with a relaxation of the strict analytic condition. The wavelets are said quasi-analytic, and satisfy an approximation of the form $\widehat{\psi}_g(\lambda) \approx -\mathrm{i}\ \mathrm{sign}(\lambda)\widehat{\psi}_h(\lambda)$, for all $\lambda \in \mathbb{R}$. Details on the properties of the wavelets, including a quantification of the approximation, are provided below.

5.1.1 Definition

Selesnick (2001) proposes to look for filters $(\widehat{h}^{(L)}, \widehat{h}^{(L)})$, and $(\widehat{g}^{(L)}, \widehat{g}^{(L)})$ on the form

$$\widehat{h}^{(L)}(\lambda) = 2^{-M+1/2} \left(1 + e^{-i\lambda}\right)^M \widehat{q}_{L,M}(\lambda)\, \widehat{d}_L(\lambda) \quad \text{and} \quad \widehat{h}^{(H)}(\lambda) = \overline{\widehat{h}^{(L)}(\lambda + \pi)} e^{-i\lambda}, \tag{5.4}$$

$$\widehat{g}^{(L)}(\lambda) = 2^{-M+1/2} \left(1 + e^{-i\lambda}\right)^M \widehat{q}_{L,M}(\lambda)\, \overline{\widehat{d}_L(\lambda)} e^{-i\lambda L} \quad \text{and} \quad \widehat{g}^{(H)}(\lambda) = \overline{\widehat{g}^{(L)}(\lambda + \pi)} e^{-i\lambda}, \tag{5.5}$$

with $\widehat{q}_{L,M}(\lambda)$ a real polynomial of $(e^{-i\lambda})$ such that $\widehat{q}_{L,M}(0) = 1$. The filter $\widehat{d}_L(\lambda)$ is defined by

$$\widehat{d}_L(\lambda) = \mathrm{e}^{\mathrm{i}\,\lambda(-L/2+1/4)} \left[\cos(\lambda/4)^{2L+1} + \mathrm{i}\,(-1)^{L+1} \sin(\lambda/4)^{2L+1}\right].$$

Common factor wavelets of Selesnick (2001) are introduced with $\widehat{q}_{L,M}$ such that the wavelet decomposition satisfies the perfect reconstruction condition. This condition is classically used for deriving wavelet bases $2^{1/2}\psi_{gj,k} = 2^{-1/2}2^{j/2}\psi_g(2^j \cdot -k)$ and $2^{1/2}\psi_{hj,k} = 2^{-1/2}2^{j/2}\psi_h(2^j \cdot -k)$, $j, k \in \mathbb{Z}$, which are orthonormal bases of $L^2(\mathbb{R})$. In that case, $\widehat{q}_{L,M}$ is defined as a solution of

$$|\widehat{q}_{L,M}(\lambda)|^2 s(\lambda) + |\widehat{q}_{L,M}(\lambda+\pi)|^2 s(\lambda+\pi) = 1\,, \tag{5.6}$$

where we have set $s(\lambda) = \frac{2^{4L-1}}{(2L+1)^2} 2^{-M}(1+\cos(\lambda))^M \left|\widehat{d}_L(\lambda)\right|^2$. Achard, Clausel, Gannaz, and Roueff (2020) prove the existence of $\widehat{q}_{L,M}$.

Another possibility is to consider that $\widehat{q}_{L,M}$ is a constant equal to 1. Perfect reconstruction is not ensured but it is not necessary for estimation procedures. The properties of the filters are then easier to establish, since $\widehat{q}_{L,M}$ does not have an explicit expression with perfect reconstruction.

Definition 5.1 (Common Factor Wavelets (CFW))**.** *Let M, L be strictly positive integers. Let (ϕ, ψ) be a family of Common Factor wavelets defined by equation* (5.1), (5.2), (5.3), *and* (5.4), (5.5). *If filter $\widehat{q}_{L,M}$ satisfies perfect reconstruction condition* (5.6), *the pair (ϕ, ψ) will be denoted as CFW-PR(M,L). If $\widehat{q}_{L,M}$ is a constant polynomial equal to 1, (ϕ, ψ) will be denoted as CFW-C(M,L).*

In practice, both CFW(PR(M,L) and CFW-C(M,L) wavelets have a compact support, with respective length $2M + 2L$ and $M + L + 2$. This can be proved theoretically for CFW-C(M,L) filters.

It is also possible to quantify the analytic approximation of the filters.

Theorem 5.2 (Achard, Clausel, Gannaz, and Roueff 2020)**.** *For all $\lambda \in \mathbb{R}$, for all $\widehat{q}_{L,M}$ real polynomial of $(e^{-i\lambda})$,*

$$\begin{aligned} \widehat{\psi}(\lambda) &= \widehat{\psi}_h(\lambda) + \mathrm{i}\,\widehat{\psi}_g(\lambda) = \left(1 - \mathrm{e}^{\mathrm{i}\eta_L(\lambda)}\right) \widehat{\psi}_h(\lambda)\,, \\ \textit{with}\quad \alpha_L(\lambda) &= 2(-1)^L \operatorname{atan}\left(\tan^{2L+1}(\lambda/4)\right)\,, \\ \eta_L(\lambda) &= -\alpha_L(\lambda/2+\pi) + \sum_{j=1}^{\infty} \alpha_L(2^{-j-1}\lambda)\,. \end{aligned} \tag{5.7}$$

Additionally, for all $\lambda \in \mathbb{R}$,

$$\left|\widehat{\psi}_h(\omega) + \mathrm{i}\,\widehat{\psi}_g(\omega) - 2\mathbb{1}_{\mathbb{R}_+}(\omega)\,\widehat{\psi}_h(\omega)\right| = U_L(\omega)\left|\widehat{\psi}_h(\omega)\right|\,,$$

where U_L *is a* $\mathbb{R} \to [0,2]$ *function satisfying, for all* $\lambda \in \mathbb{R}$,

$$U_L(\lambda) \leqslant 2\sqrt{2}\left(\log_2\left(\frac{\max(4\pi, |\lambda|)}{2\pi}\right) + 2\right)\left(1 - \frac{\delta(\lambda, 4\pi\mathbb{Z})}{\max(4\pi, |\lambda|)}\right)^{2L+1} .$$

and, for all $\lambda \in \mathbb{R}$ *and* $A \subset \mathbb{R}$, $\delta(\lambda, A)$ *denotes the distance of* λ *to* A *defined by* $\delta(\lambda, A) = \inf\{|\lambda - x|\,,\ x \in A\}$.

Equation (5.7) uses the convention $\operatorname{atan}(\pm\infty) = \pm\pi/2$ so that α_L is well defined on $\mathbb{R}$.

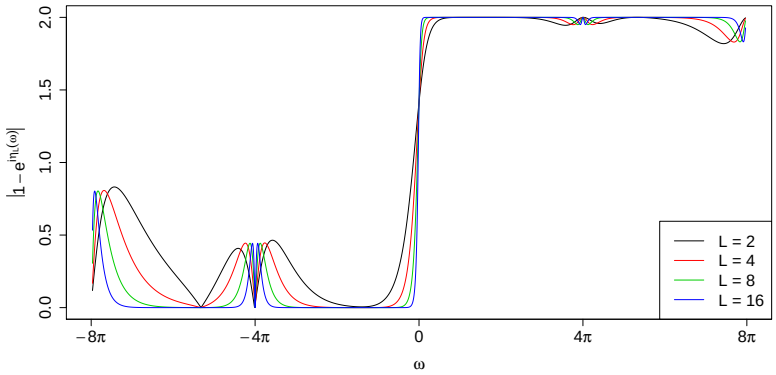


Figure 5.1: Plots of the function $\omega \mapsto |1 - e^{i\eta_L(\omega)}|$ for $L = 2, 4, 8, 16$.

A remarkable result of Theorem 5.2 is that the relation between $\widehat{\psi}_h$ and $\widehat{\psi}_g$ does not depend on M but only on L. Figure 5.1 represents the behavior of the multiplicative term, $\omega \mapsto |1 - e^{i\eta_L(\omega)}|$, for different values of L. As $L \to \infty$, the relative difference $U_L = \left|\widehat{\psi} - 2\mathbb{1}_{\mathbb{R}_+}\widehat{\psi}_h\right| / \left|\widehat{\psi}_h\right|$ converges to zero exponentially fast on any compact subsets that do not intersect $4\pi\mathbb{Z}$. Nevertheless, the regularity of the function $\widehat{\psi}_h$ is related to parameter M and may influence the quality of the analytic approximation.

Figure 5.2 displays the overall shapes of the Fourier transforms $\widehat{\psi}_h$ of CFW-PR(M,L) wavelets for various values of M and L. Their quasi-analytic counterparts $\widehat{\psi}_h(\omega) + i\,\widehat{\psi}_g(\omega)$ are plotted below in the same scales. It illustrates the satisfactory quality of the analytic approximation.

This study shows good properties of the CFW(PR(M,L) and CFW-C(M,L) wavelets. In particular, they have a compact support, with known length, which will allow to control the number of coefficients. And the quality of the analytic approximation can be quantified using the parameter L, independently from the parameter M. They are, hence, good candidates for developing an estimation procedure.

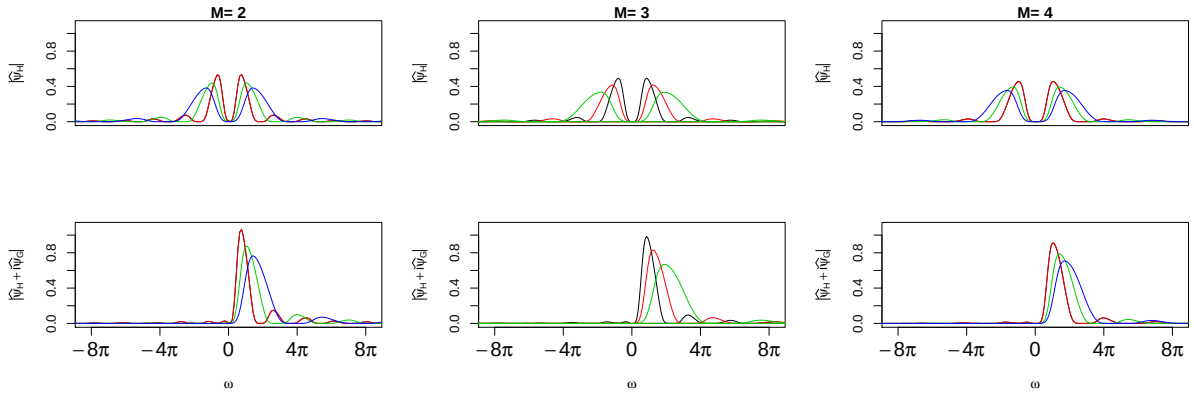


Figure 5.2: Top row: Plots of $|\widehat{\psi}_h|$ for $M = 2$(left), 3 (center), 4 (right) and $L = 2$ (black), 4 (red), 8 (green). Bottom row: same for $|\widehat{\psi}_h + \mathrm{i}\widehat{\psi}_g|$.

5.1.2 Approximation of the sample covariance

Consider $\boldsymbol{X} = \left\{ \left[X_1(t), \dots X_p(t) \right], t \in \mathbb{Z} \right\}$ a multivariate LRD process, satisfying **(M-1)** and **(M-2)**. Recall that $\boldsymbol{X}$ admits a generalized spectral density $\boldsymbol{f}(\cdot)$ satisfying

$$\boldsymbol{f}(\lambda) = (\boldsymbol{\Lambda}(\lambda)\, \boldsymbol{\Theta}\, \boldsymbol{\Lambda}(\lambda)) \circ \boldsymbol{f}^S(\lambda), \quad \text{for all } \lambda > 0,$$

where $\boldsymbol{\Lambda}(\lambda) = \mathrm{Diag}\left(|\lambda|^{-d_1}, \dots, |\lambda|^{-d_p} \right)$, and $\boldsymbol{\Theta} = \boldsymbol{\Omega} \mathrm{e}^{\mathrm{i}\boldsymbol{\Phi}}$ with $\boldsymbol{\Omega}$ symmetric non-negative semi-definite matrix and $\boldsymbol{\Phi}$ anti-symmetric matrix. The parameter β measures the regularity of the short-range dependence function $\boldsymbol{f}^S(\cdot)$.

Suppose that $\boldsymbol{X}$ is observed at $\{\boldsymbol{X}(1), \dots \boldsymbol{X}(N)\}$. Let $\boldsymbol{W}(jk) = \begin{pmatrix} W_1(j,k) & W_2(j,k) & \dots & W_p(j,k) \end{pmatrix}$ denote the wavelet coefficients of $\boldsymbol{X}$, for $j \geqslant 0$ and $k \in \mathbb{Z}$.

As in Section 3.2, the behavior of Cov $(\boldsymbol{W}(j,k))$ with complex wavelets can be obtained theoretically. This result consists in the extension of Proposition 3.1 to quasi-analytic wavelets. The results are obtained hereafter only for CFW-C(M,L) filters. Indeed, the same results are more difficult to obtain for CFW-PR(M,L) filters because no explicit expression of $\widehat{q}_{L,M}$ satisfying (5.6) is available.

I introduce an assumption, similar to **(C-a)**. .

(C-b) $-M + \beta/2 + 1/2 < d_\ell < M + \frac{1}{4} + \frac{2-3\beta}{2(2-\beta)}$ for all $a = 1, \dots, p$, $M \geqslant 2$ and $0 < \beta < 2$.

The approximation of the covariance of the wavelet coefficients is stated in the following proposition.

Proposition 5.3 (Achard and Gannaz 2022)**.** *Let $\boldsymbol{X}$ be a p-multivariate long range dependent process with long-range dependence parameters $d_1, \dots, d_p$ with normalized spectral density satisfying* **(M-1)**–**(M-2)**. *Consider $\{W_a(jk), (j,k) \in \mathbb{Z}, a = 1, \dots, p\}$ the wavelet coefficients obtained with CFW-C(M,L) filters. Suppose that* **(C-b)** *hold and that $L2^{-2j} + L^{-(2M-d_a-d_b+1)} \to 0$ when j goes to infinity. Then, for all $a, b = 1, \dots, p$,*

$$\lim_{j\to\infty} 2^{-j\,(d_a+d_b-\beta)}\mathrm{Cov}\left(W_a(jk), W_b(jk)\right) = G_{a,b}^{MCW}, \text{ with } G_{a,b}^{MCW} = 2^{j(d_a+d_b)}\,\Omega_{a,b}\,K^{MCW}(d_a+d_b),$$

$$K^{MCW}(\delta) = 4\int_0^\infty |\lambda|^{-\delta}\ \left|\widehat{\psi}_h(\lambda)\right|^2\,\mathrm{d}\lambda\,.$$

The specificity of CFW-C(M,L) filters is that the quality of the analyticity approximation is based only on parameter L, as written in Proposition 5.3. Nevertheless, to have an approximation with the same quality than the one obtained with real wavelets, the choice of L is more constrained (see Proposition 3 of Achard and Gannaz 2022). This tradeoff is due to the fact that the greater L, the better analyticity approximation, but the larger the length of the wavelets support. In practice, due to numerical instability, choosing high values (*i.e.* $\geqslant 8$) is not tractable. Nevertheless, the results have a good quality even with a smaller value of L.

Illustration

I consider again the bivariate FIVARMA(0,$\boldsymbol{d}$,0) process, $\boldsymbol{X} = (X_1, X_2)$, defined in Section 3.1.2. Recall that the spectral density of (X_1, X_2) satisfies LRD assumption **(M-1)** (given in Section 3.1) is satisfied with $\varphi_{\mathbb{1},2} = \frac{\pi}{2}(d_1 - d_2)$. Denote $\rho = \boldsymbol{\Omega}_{1,2}\mathrm{e}^{\mathrm{i}\,\varphi_{1,2}}$ the correlation term at the 0^+ frequency.

Define $\{W_a(j,k),\, j \geqslant 0,\, k \in \mathbb{Z}\}$ for $a = 1, 2$ the wavelet coefficients of the process $\boldsymbol{X}$ obtained from a wavelet transform. Let $\theta(j)$ denotes the empirical correlation between $\{W_1(j,k),\, k \in \mathbb{Z}\}$ and $\{W_2(j,k),\, k \in \mathbb{Z}\}$, that is,

$$\theta(j) = Cor(\{\boldsymbol{W}_1(j,k),\, k \in \mathbb{Z}\}, \{\boldsymbol{W}_2(j,k),\, k \in \mathbb{Z}\})$$

for a given scale $j \geqslant 0$. Proposition 3.1 shows that, if the wavelet transform is real, asymptotically θ_j tends to the real part of ρ when j tends to infinity. In particular, when the phase is equal to $\pi/2$, *i.e.* when $d_1 - d_2$ is close to 1 *modulo* 2, the correlation of the real wavelet coefficients tends to 0. With complex wavelets, Proposition 5.3 states that asymptotically $\theta(j)$ tends to ρ when j tends to infinity. Hence, the magnitude of the correlation is asymptotically equal to $\Omega_{1,2}$, and the phase $\varphi_{\ell,m}$ can be recovered through the argument.

Let me simulate such processes with a LRD parameter $\boldsymbol{d}$ equal to $(0.2, 1.2)$. The phase is here equal to $\pi/2$ and the approximation of the spectral density at the zero frequency contains only

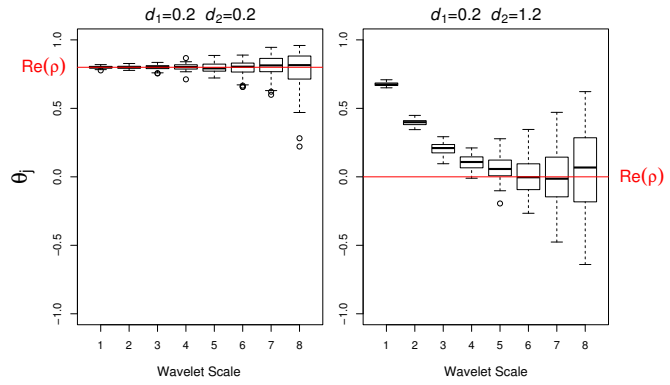


Figure 5.3: Boxplots of the sample correlations of the wavelet coefficients $\{\theta(j),\, j \geqslant 0\}$ for real wavelets, at different scales j, for the FIVARMA(0,$\boldsymbol{d}$,0) described in Section 3.1.2, with different values of $\boldsymbol{d}$. Results were obtained on 1000 realizations of $(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(N))$ with $N = 2^{12}$. The red horizontal lines correspond to the real part of the theoretical value of ρ of the ARFIMA model.

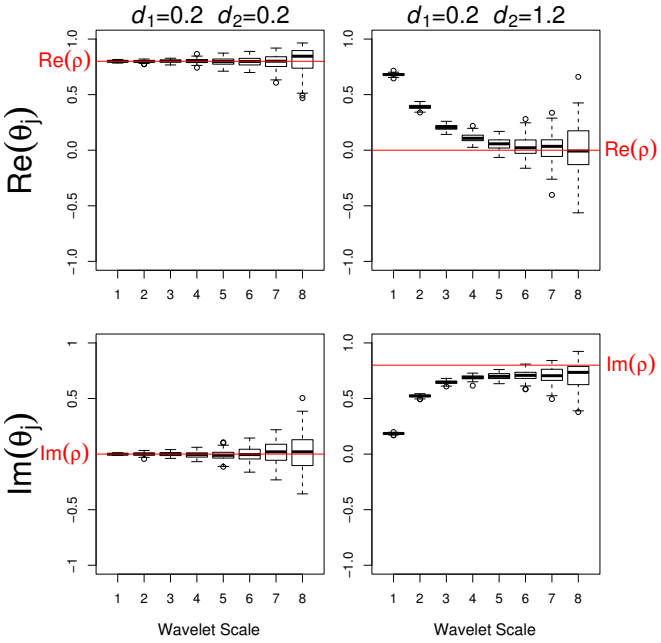


Figure 5.4: Boxplots of the real and the imaginary parts of the sample correlations of the wavelet coefficients $\{\theta(j),\, j \geqslant 0\}$ for analytic wavelets CFW-PR(4,4), at different scales j, for the FIVARMA(0,$\boldsymbol{d}$,0) described in Section 3.1.2, with different values of $\boldsymbol{d}$. Results were obtained on 1000 realizations of $(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(N))$ with $N = 2^{12}$. The red horizontal lines correspond to the real and the imaginary parts of the theoretical value of ρ of the ARFIMA model.

the imaginary part. In this example, Figure 5.3 illustrates the cancellation of the correlation of the real wavelet coefficients and consequently the impossibility to identify ρ. With complex wavelets, Figure 5.4 reveals that the imaginary part of the wavelet coefficient correlations is, as expected in this case, converging in mean to ρ.

5.2 Inference

Recall that inference in Chapter 3 deals with a parametric phase term, $\phi_{\ell,m} = \pi(d_\ell - d_m)/2$. In this case, the estimation of both the long-run covariance structure and the LRD parameters was proposed by Lobato (1999) and Shimotsu (2007), with a Fourier-based local Whittle estimation, and by Achard and Gannaz (2016) with a similar procedure based on a real wavelet representation. The objective is to extend the estimation to non parametric phases. For a general phase term, Sela and Hurvich (2012) proposed an estimation based on the average periodogram and Robinson (2008) and Baek, Kechagias, and Pipiras (2020) developed a Fourier-based local Whittle estimation. I now present a (complex) wavelet-based procedure. It has the advantage, compared to Fourier-based procedures, to allows for non stationarity. For example, our method can be applied to multivariate fractional Brownian motions (mFBMs, see Section 3.1.3). It is, to my knowledge, the only non parametric approach that can deal with mFBMs.

Similarly to Section 3.2, a multivariate wavelet-based local Whittle procedure can be built. Let $\boldsymbol{\Lambda}_j(\boldsymbol{d}) = \mathrm{Diag}\left(2^{j\boldsymbol{d}}\right)$, $j_0 \leqslant j \leqslant j_1$. The estimators $(\widehat{\boldsymbol{d}}^{MCW}, \widehat{\boldsymbol{G}}^{MCW})$ minimize $\mathcal{L}(\boldsymbol{d}, \boldsymbol{G})$, with

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

where n_j is the number of wavelet coefficients at scale j. The exponent $\top$ denotes the transpose operator. As with real wavelets, the consistency of the estimators and their asymptotic normality can be proved. Since the approximation of the covariance $\boldsymbol{G}^{MCW}(\boldsymbol{d})$ is complex, the real and the imaginary parts of $\boldsymbol{\Theta}$ are both recovered. This allows to consider non parametric phases in modeling **(M-1)**.

Let me first state the consistency of the estimators.

Theorem 5.4 (Achard and Gannaz 2022)**.** *Suppose that assumptions of Proposition 5.3 and* **Condition (C)** *are satisfied. Let j_0 and j_1 be such that $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$ and $j_0 < j_1 \leqslant j_N$ with $j_N = \max\{j, n_j \geqslant 1\}$.*

Consider CFW-C(M,L) filters with $M \geqslant 2$ and

$$2^{-2j_0}L + N^{-1}2^{j_1}L + \log(N)^3\, L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell)} \to 0.$$

Then, $\forall (a,b) \in \{1,\dots,p\}^2$,

$$\begin{aligned}
\widehat{\boldsymbol{d}}^{MCW} - \boldsymbol{d} &= O_{\mathbb{P}}(L2^{-2j_0} + \log(N)L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell)} + 2^{-j_0\beta} + N^{-1/2}2^{j_0/2}),\\
\widehat{G}^{MCW}_{a,b} - G^{MCW}_{a,b}(\boldsymbol{d}) &= O_{\mathbb{P}}(\log(N)(L2^{-2j_0} + \log(N)L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell)} + 2^{-j_0\beta} + N^{-1/2}2^{j_0/2})),\\
\widehat{\Theta}^{MCW}_{a,b} - \Theta_{a,b} &= O_{\mathbb{P}}(\log(N)(L2^{-2j_0} + \log(N)L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell)} + 2^{-j_0\beta} + N^{-1/2}2^{j_0/2})).
\end{aligned}$$

Taking $2^{j_0} = N^{\frac{1}{1+2\beta}}$ *and* $L = N^{\frac{2-\beta}{2+\beta}\frac{\beta}{1+2\beta}}$,

$$\widehat{\boldsymbol{d}}^{MCW} - \boldsymbol{d} = O_{\mathbb{P}}(N^{-\beta/(1+2\beta)}).$$

The condition on parameter j_0, which determines the scales used in estimation, is the same as the one for real wavelets, given in Theorem 3.2. Yet, the highest scale j_1 also has to satisfy $N^{-1/2}2^{j_1/2} \to 0$. The assumption can be formulated only on j_0, taking $j_1 = j_0 + \Delta$, with $\Delta < \infty$.

The parameters M and L in CFW-C(M,L) filters are subject to the conditions **(C-b)** and $2^{-2j_0}L + N^{-1}2^{j_0}L + \log(N)^3\, L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell)} \to 0$. Condition **(C-b)** depends only on M. It is stronger than the one given in Section 3.2 with real filters. It imposes the number of vanishing moments M to be high enough. In particular, it yields high values of M when β is close to 2.

The parameter L quantifies the quality of the analytic approximation of CFW-C(M,L) filters.The assumption $2^{j_1}N^{-1}L \to 0$ allows n_j to be equivalent to $2^{-j}N$ as N goes to infinity. This facilitates the translation of the proofs from real wavelets to complex wavelets. The condition $\log(N)^3\, L^{-(2M+1-2\max_{\ell=1,\dots,p} d_\ell} \to 0$ says that L must be high enough so that the analytic approximation is satisfactory. Alternatively, L should not be too high, since the length of the support of the wavelet depends on L. Condition $2^{-2j_0}\, L \to 0$ ensures that the size of the support of the wavelets remains reasonable. Observe that, in practice, the choice of L is not critical, but this condition influences the choice of j_0. It appears that it must be higher than the usual choice for real filters. See the simulation study of Achard and Gannaz (2022) for details.

Finally, asymptotic convergence can be established, similarly to real wavelets (Section 4.1). I only give the main result on estimators, but the asymptotic distributions of the long-run correlation can also be deduced easily.

Theorem 5.5 (Achard and Gannaz 2022)**.** *Suppose that conditions of Theorem 5.4 are satisfied and that assumption* **(M-3)** *hold. Let* $j_0 < j_1 \leqslant j_N$ *with* $j_N = \max\{j, n_j \geqslant 1\}$ *such that*

$$j_1 - j_0 \to \Delta \in \{1,\dots,\infty\},\ \log(N)^2(N2^{-j_0(1+2\beta)} + N^{-1/2}2^{j_1/2}) \to 0.$$

Define $n = \sum_{j=j_0}^{j_1} n_j$.

Consider CFW-C(M,L) filters with $M \geqslant 2$ and

$$N^{-1}2^{j_1}L + \log(N)^3\, N^{1/2}2^{-j_0/2}(L2^{-2j_0} + L^{-(2M+1-\max_{\ell=1,\dots,p} d_\ell)}) \to 0.$$

Define $n = \sum_{j=j_0}^{j_1} n_j$. Then,

- *$\sqrt{n}(\widehat{\boldsymbol{d}}^{MCW} - \boldsymbol{d})$ converges in distribution to a centered Gaussian distribution with a variance equal to $\mathbf{V}^{MCW\,(\boldsymbol{d})}(\Delta)$, defined by (4.4) with $\mathbf{G} = \boldsymbol{G}^{MCW}(\boldsymbol{d})$, whenever it appears.*
- *$\mathrm{vec}\Big(\sqrt{n}\left(\widehat{\boldsymbol{G}}^{MCW} - \boldsymbol{G}^{MCW}(\boldsymbol{d})\right)\Big)$ converges in distribution to a centered Gaussian distribution with a variance equal to $\mathbf{V}^{MCW\,\boldsymbol{G}(\Delta)}$, with $\mathbf{V}^{MCW\,\boldsymbol{G}(\Delta)}$ defined by (4.9) with $\mathbf{G} = \boldsymbol{G}^{MCW}(\boldsymbol{d})$, whenever it appears.*

Similarly to Chapter 4, tests can be built from this theorem. A test which may be interesting is the test of

$$(H_{0,a,b})\ \varphi_{a,b} = \frac{\pi}{2}(d_a - d_b) \ \text{ against } \ (H_{1,a,b})\ \varphi_{a,b} \neq \frac{\pi}{2}(d_a - d_b).$$

Indeed, it would allow to see if the parametric modeling of Chapter 3 is possible. Simulations show that real wavelet estimation yields more accurate estimation. If the test concludes that $(H_{0,a,b})$ is not rejected, a real wavelet estimation could be preferred. This test has not been done in practice yet.

Let $(a,b) \in \{1,\dots,p\}^2$, $a \neq b$. Define

$$\widehat{r}_{a,b}^{MCW} = \frac{\widehat{G}_{a,b}^{MCW}}{\sqrt{\widehat{G}_{a,a}^{MCW}\widehat{G}_{b,b}^{MCW}}} \text{ and } r_{a,b}^{MCW} = \frac{G_{a,b}^{MCW}(\boldsymbol{d})}{\sqrt{G_{a,a}^{MCW}(\boldsymbol{d})G_{b,b}^{MCW}(\boldsymbol{d})}}.$$

As a corollary of Theorem 5.5, the asymptotic normality and tests on $\mathbf{r}^{MCW}$ follow. Observe that testing

$$(H_{0\,a,b}^{(\mathbf{r})})\ r_{a,b}^{MCW} = 0 \ \text{ against } \ (H_{1\,a,b}^{(\mathbf{r})})\ r_{a,b}^{MCW} \neq 0$$

is not biased as the same test on $\mathbf{r}^{MWW}$. Indeed, no cosine term appears in $\boldsymbol{G}^{MCW}(\boldsymbol{d})$ (see (3.4)). Having $r_{a,b}^{MCW} = 0$ is equivalent to $\Omega_{a,b} = 0$.

Note that the multiple testing procedures proposed in Section 4.2 are still available in this context. That is, the corrections of Bonferroni, Šidák, or MaxT, can be applied to control asymptotically the FWER.

5.3 Application

I now present the application on Dataset 4. Recall that four dead rats and twenty-one live rats were scanned, seven for each of three anesthetics, called ISO_W, ETO_L and MED_L. Estimation was done using CFW-PR(4,4) filters. The scales above $j_0 = 4$ were removed from estimation. This choice is based on the simulation results of Achard and Gannaz (2022).

Figure 5.5 shows the empirical distribution of the estimated empirical estimators $\widehat{d}$. As expected, the long-memory parameters for dead rats are close to zero. The distributions are centered around zero, with a Gaussian-like shape. For rats under anesthetics, densities are not centered around zero and the variance between brain regions are higher than what is observed for dead rats. Long-memories for rats under anesthetic ISO_W are higher than under other anesthetics.

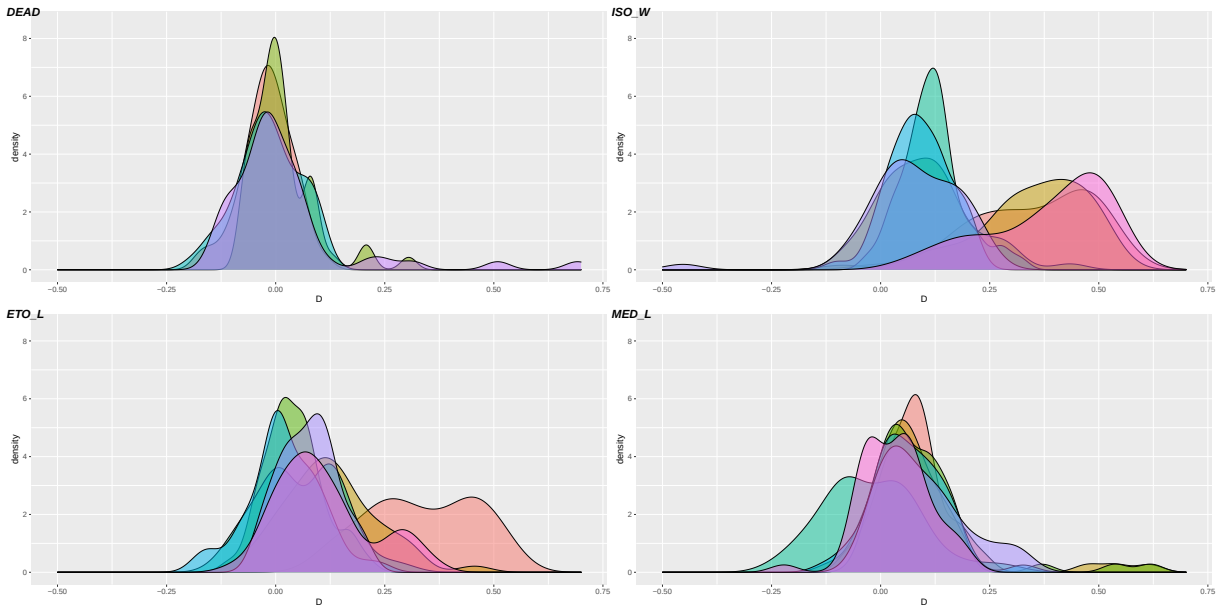


Figure 5.5: Plot of the empirical distribution of the long memory parameters $\widehat{d}$ obtained for the 4 groups of rats in Dataset 4: DEAD, ISO_W, ETO_L and MED_L. Each color corresponds to a rat.

The distributions of magnitudes and phases of the estimated correlations, ρ and ϕ, for each rats are displayed respectively in Figure 5.6 and Figure 5.7. First, as expected, the magnitudes obtained for dead rats seem significantly different from live rats. For dead rats, distributions have a small support, that is, only 9 on the 5100 values (0.18%) satisfy $\widehat{\rho} > 0.3$. Then, ISO_W and ETO_L present quite similar distributions, with possibly high magnitudes. By contrast,

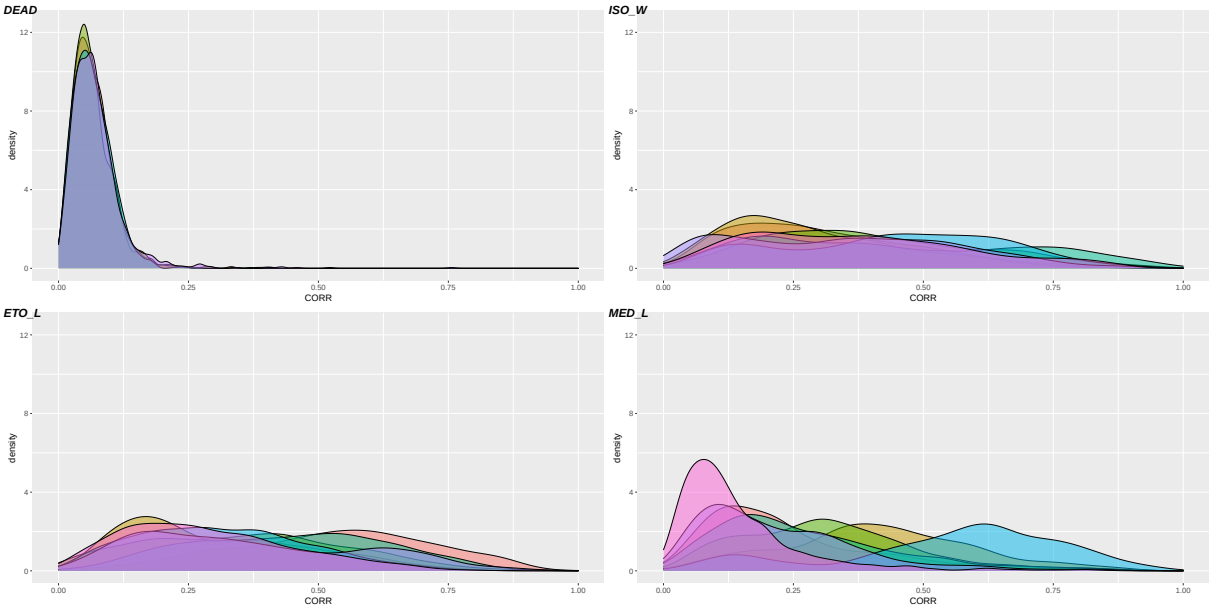


Figure 5.6: Plot of the empirical distribution of the correlation magnitudes $(\widehat{\rho}_{a,b})_{1\leqslant a<b\leqslant p}$ obtained for the 4 groups of rats in Dataset 4: DEAD, ISO_W, ETO_L and MED_L. Each color corresponds to a rat.

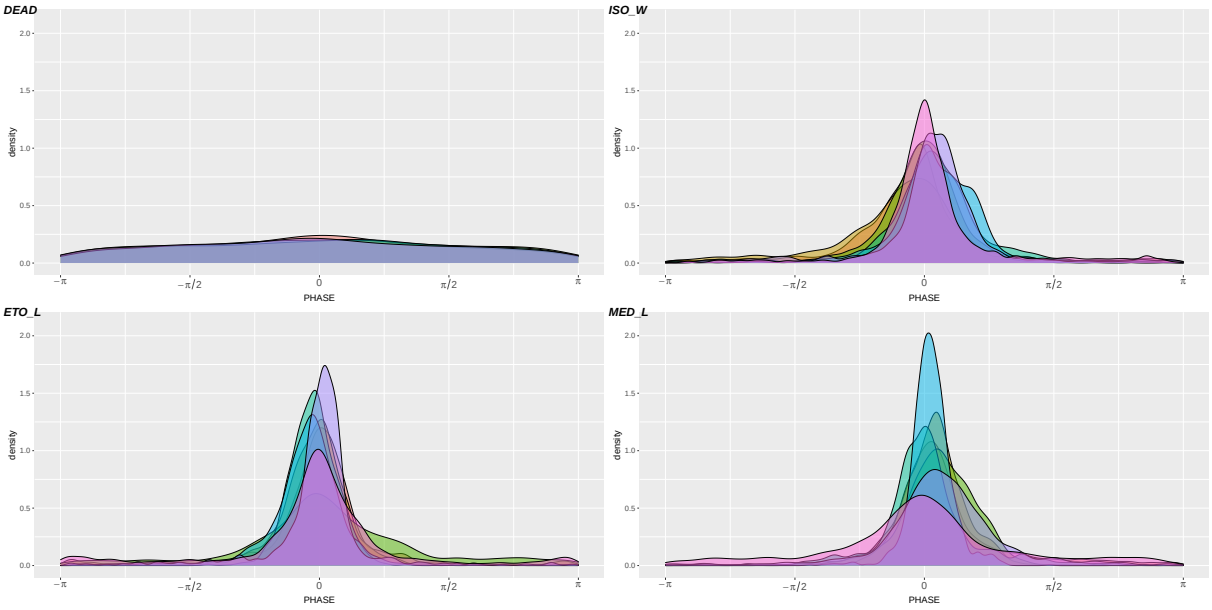


Figure 5.7: Plot of the empirical distribution of the phases $(\widehat{\varphi_{a,b}})_{a,b=1,\ldots,p}$ obtained for the 4 groups of rats in Dataset 4: DEAD, ISO_W, ETO_L and MED_L. Each color corresponds to a rat.

correlations for MED_L anesthetic are lower. These results tend to show that MED_L anesthetic is stronger than the other anesthetics, leading to fewer connections between brain regions.

The phase parameter can be seen as a measure of the asymmetry in the properties at large lags among the components of the signals. See *e.g.* Section 3.1. For the dead rats, we observe mainly uniform distributions. For the live rats, Figure 5.7 shows that the distributions have heavy tails. The tails are heavier for MED_L anesthetic than for other anesthetics. This can be explained by the fact that the phase is non-informative when the magnitude is close to zero. The lower the correlation magnitudes, the higher the tails of the phase distributions.

The fact that the phase is non informative when the correlations are close to zero is related to an identifiability problem. Observe that when a magnitude is equal to zero, $\Omega_{a,b} = 0$, $a, b = 1, \dots, p$, the corresponding phase, $\varphi_{a,b}$, is not identifiable. It occurs for example in case of fractional co-integration (VARFIMA models, Section 3.1.1). To counter this problem, Baek, Kechagias, and Pipiras (2020) propose to use another parametrization, where $\Theta_{a,b}$ is written with its real and its imaginary part. The parametrization chosen here, nevertheless, seems more appropriate for graph inference, which is important in our neuroscience framework.

Graph representation

Let us compute the adjacency matrix obtained for each rat within each group. Edges correspond to a magnitude higher than 0.3. The value of the threshold is motivated by the observation of the supports obtained for dead rats. We then select the edges which are present in all the graphs of the rats of the group. One graph is then obtained per group. For each group, we compute the average of the estimated phase for each detected edge. Figure 5.8 illustrates the graphs obtained for the 4 different groups.

Each edge is colored based on the average phase when it satisfies $|\varphi_{a,b}| > 1.1|\varphi^*_{a,b}|$ where $\varphi^*_{a,b} = -\frac{\pi}{2}(d_a - d_b)$, $(a, b) \in \{1, \dots, p\}^2$. The value $\varphi^*_{a,b}$ corresponds to the phase considered in Chapter 3 and Chapter 4. The more colored edges, the more the phase behavior differs from the modeling of Section 3.2. And the less the real wavelet-based procedure is adequate.

The DEAD group has indeed no edges. The MED_L group has less edges than the two other groups of anesthetic. It hence seems that MED_L anesthetic inhibits more the activity. Next, ETO_L group and ISO_W group have a similar number of edges (respectively 133 and 145). The phases sightly differ. Around 50% of mean phases are outside the interval $[-1.1|\varphi^*|, 1.1|\varphi^*|]$ for ETO_L group and 46% ISO_W group. The two anesthetics seem very similar. The high proportion of phases outside the interval illustrates that the modeling of these data is complex. The introduction of a general phase enables to take into account this complexity.

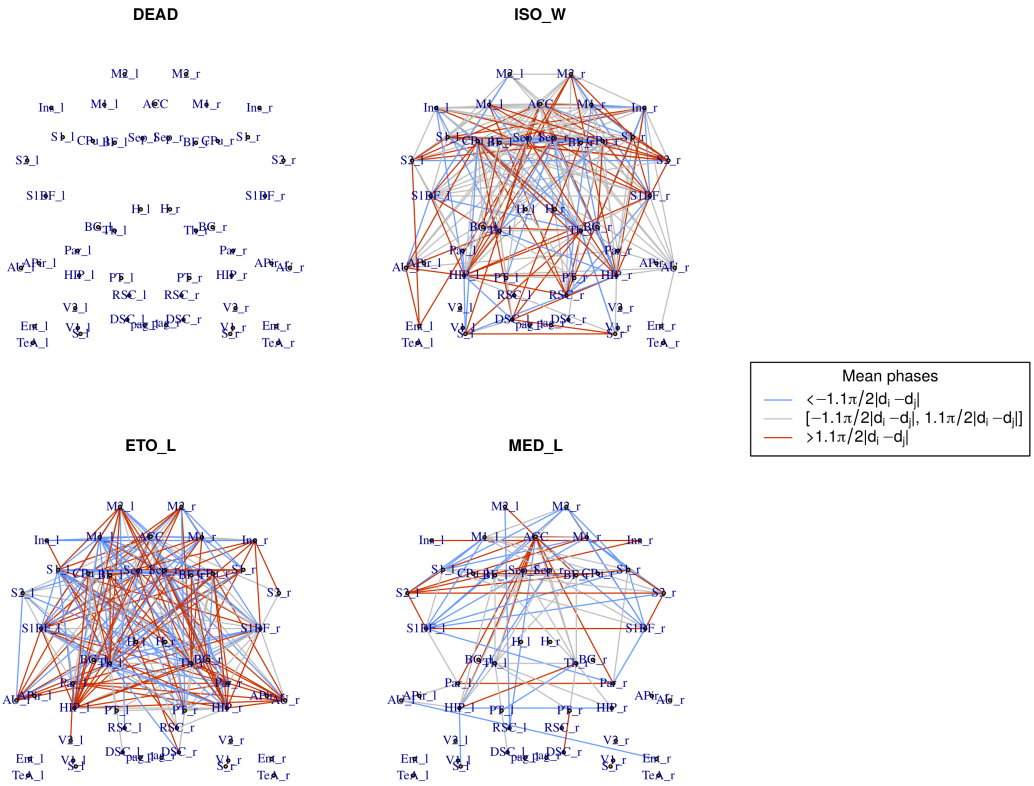


Figure 5.8: Plot of the average graphs with correlations and phases obtained for 4 groups of rats in Dataset 4: DEAD, ISO_W, ETO_L and MED_L. Only edges corresponding to an average correlation's magnitude higher than 0.3 are displayed. Red edges correspond to positive average phases higher than $1.1|\varphi^*|$, blue edges correspond to negative average phases lower than $-1.1|\varphi^*|$, and grey edges to average phases between $-1.1|\varphi^*|$ and $1.1|\varphi^*|$. The quantities φ^* are equal to $\varphi^*_{a,b} = -\frac{\pi}{2}(d_a - d_b)$, $a, b = 1, \ldots, p$.

Concerning the physical interpretation, no easy conclusion can be given. Considering the different time scales involved in the production of the BOLD response, we may hypothesize that lags are not the underlying phenomenon that produces phase differences in the fMRI signals. However, as stated in Buxton (2013), the time scale can vary in the same subject depending on the physiological baseline state, which is known to be modified under anesthetization.

Remark on the quality of the recordings

Recall that, as explained in page 84, the recordings of dead rats in Dataset 4 are not satisfactory. This appears in Figure 5.5 where the densities of $\widehat{\boldsymbol{d}}$ for some dead rats present significant non zero values. Observe that the graph for dead rats, in Figure 5.8, is empty since I only consider the edges that are present on the four dead rats. The problems in the recording of the data do not appear in the graph due to some randomness in the measurement errors. The comparison between dead and live rats is, with this approach, still accurate. Indeed, significant differences appear, which are not the cause only of artifacts in recordings. But a finer study would need recordings with a better quality.

References

Achard, S., M. Clausel, I. Gannaz, and F. Roueff (2020). "New results on approximate Hilbert pairs of wavelet filters with common factors". In: *Applied and Computational Harmonic Analysis* 49.3, pp. 1025–1045.

Achard, S. and I. Gannaz (2016). "Multivariate Wavelet Whittle Estimation in Long-range Dependence". In: *Journal of Time Series Analysis* 37.4, pp. 476–512.

Achard, S. and I. Gannaz (2022). "Whittle estimation with (quasi-)analytic wavelets". working paper or preprint.

Baek, C., S. Kechagias, and V. Pipiras (2020). "Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity". In: *Journal of Statistical Planning and Inference* 205, pp. 245–268.

Buxton, R. B. (2013). "The physics of functional magnetic resonance imaging (fMRI)". In: *Reports on Progress in Physics* 76.9, p. 096601.

Gannaz, I., S. Achard, M. Clausel, and F. Roueff (2017). "Analytic wavelets for multivariate time series analysis". In: *Wavelets and Sparsity XVII*. Vol. 10394. International Society for Optics and Photonics, p. 103941X.

Lobato, I. N. (1999). "A semiparametric two-step estimator in a multivariate long memory model". In: *Journal of Econometrics* 90.1, pp. 129–153.

Robinson, P. M. (2008). "Multiple local Whittle estimation in stationary systems". In: *The Annals of Statistics* 36.05, pp. 2508–2530.

Sela, R. J. and C. M. Hurvich (2012). "The averaged periodogram estimator for a power law in coherency". In: *Journal of Time Series Analysis* 33.2, pp. 340–363.

Selesnick, I. W. (2001). "Hilbert transform pairs of wavelet bases". In: *Signal Processing Letters, IEEE* 8.6, pp. 170–173.

Shimotsu, K. (2007). "Gaussian semiparametric estimation of multivariate fractionally integrated processes". In: *Journal of Econometrics* 137.2, pp. 277–310.

Perspectives and other ongoing works

I present below some possible perspectives of the work described in this manuscript. Some trails are ideas which I may never follow, due to lack of time or to the will to discover new paths. Some perspectives given here are currently in progress.

Density estimation under dependence, optimality?

In Section 1.2, we have proposed an estimation procedure of density with dependent data, where the rate is similar to the independent case. Other works have been done on density estimation under weak dependence assumptions, for example Ragache and Wintenberger (2006) and Lerasle (2009). It would be interesting to study the sharpness of the weak-dependence condition which allows a (near) optimal rate. Is the rate decreased for more dependent observations? How? Moreover, to my knowledge, no adaptive functional regression procedure has been made when the design is weak-dependent. Bühlmann, Doukhan, and Nze (2002) propose a kernel-based procedure, but it is not adaptive.

Online functional clustering and outlier detection

In Section 2.2, I have presented a clustering and outlier detection problem with functional data, motivated by an industrial context. A first perspective is to extend the Gaussian model based clustering proposed in Amovin-Assagba, Gannaz, and Jacques (2022) to other distributions, following the work of Forbes and Wraith (2014).

Since outliers represent a small proportion of measurements, a characterization of different types of outliers would need a huge amount of data. A work in progress is to build a model which re-evaluate the parameters when new data is recorded, without keeping all previous data. This online estimation can be done following the works of Bellas, Bouveyron, Cottrell, and Lacaille (2013) and Bouveyron (2014). With Martial Amovin (Arpege MasterK, ERIC, Univ.

Lyon2) and Julien Jacques (ERIC, Univ. Lyon 2), we are currently building an incremental process, which adjusts the parameters of C-funHDDC model (described in Section 2.2) every day, using new data collected by the sensors. This step answers an important issue for the implementation in industry, which is the storage of the recordings.

An extension, next, would be the consideration of dynamic functional clustering. In the work presented in Section 2.2, it would be interesting if the dysfunction of a sensor could be detected before its failure. This would allow to change the sensor before a breakdown. Studying the dynamics over time of a clustering can put in evidence a trend in the parameters, or eventually a rupture, which would warn about the dysfunction.

High dimensional settings for multivariate long-range dependence

As explained at the beginning of Chapter 4, the neuroscience application does not deal with a high dimensional setting (nor with sparse graphs). Yet, for other applications, it may be interesting to propose a penalty-based estimation. For instance, a Lasso estimation with multivariate wavelet Whittle procedure, following what was done for Fourier-based procedure by Baek, Kechagias, and Pipiras (2017) and Düker and Pipiras (2019).

Since we are often interested in the latent graph structure of the covariance, another penalization, based on graph metrics, could be proposed. It was used for example by Maretic and Frossard (2020) in a clustering approach. Yet, to my knowledge, such a penalization has not been tried in a graph inference setting.

Study of the graph properties

Once graphs are inferred, as described in Chapter 4, an objective is to extract the structure characteristics. In particular, a main perspective is to compare graphs of patients presenting a pathology with respect to a control group. Classical methods in neuroscience are using graph metrics such as minimum path length, clustering, and many others (Rubinov and Sporns 2010; Bullmore and Sporns 2009). For instance, I estimate the connectivity graph on the dataset of Achard, Delon-Martin, et al. (2012). The data consist in fMRI recordings of 17 healthy individuals and 20 individuals in a coma. 90 time series are given for each individual, each one corresponding to a brain region. Real wavelet based Whittle estimation as described in Section 3.2 is applied. Daubechies' wavelets with $M = 4$ vanishing moments were used, and the three highest frequency scales were removed ($j_0 = 4$).

Graphs with 10% of edges are built, taking the highest long-run correlation values. Since the

objective is to compare the graph characteristics, not only the degree, considering graph with the same number of edges is more appropriate. Many classical characteristics, such as global efficiency or clustering, are not able to discriminate people under coma in this dataset. But the first eigenvectors of the graphs are different. Figure P.1 displays the first plan of a Principal Component Analysis on the first eigenvectors of the adjacency matrices. It illustrates the ability of the graph representation to discriminate healthy and coma connectivity structures. Lohmann et al. (2010) also conclude that the first eigenvectors of the adjacency matrices are adequate for studying the brain connectivity. Here, this result can be interpreted as a different organization of the connectivity, where the much connected nodes are not dispatched similarly for people in coma and healthy people. Identifying which characteristics are able to describe the graphs is, hence, an interesting step after graph inference.

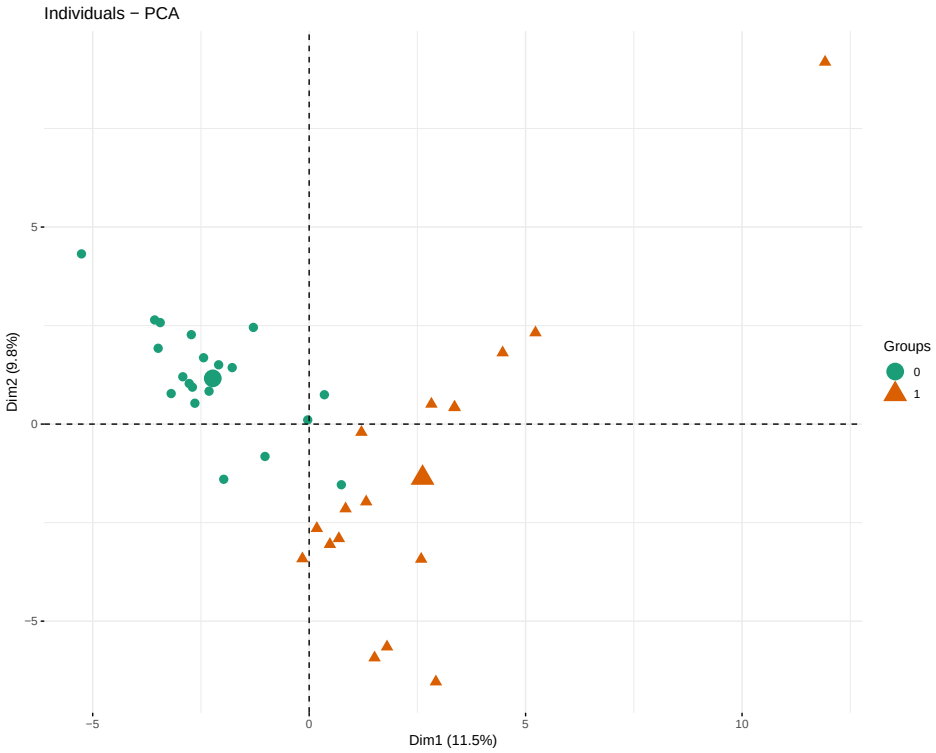


Figure P.1: First plan of Principal Component Analysis on the absolute value of the first eigenvector of the adjacency matrices inferred in Achard, Delon-Martin, et al. (2012). 0 - green circles correspond to healthy individuals and 1 - orange triangles to individuals under coma.

Quantification of the error on graph metrics

In relation with the above paragraph, a natural question is the consequence of edge missidentifications on graph structures and on classical graph metrics. In Achard, Borgnat, Gannaz, and Roux (2019), we conduct simulations to observe the effect of missidentifications of edges on classical small-world graphs. We consider two different ways of missidentification of edges. The first missidentification, illustrated on the first row of Figure P.2, corresponds to type I error *i.e.* not all the correct edges are observed, but the ones present in the graph are correct. The second missidentification, illustrated on the second row of Figure P.2, corresponds to type II error *i.e.* additional edges are added at random to the observed graph, and they are not correct. In the simulations, we take graphs with 51 nodes (motivated by Dataset 4). For the first set of simulated graphs (corresponding to the first row of Figure P.2), 51 edges of a small-world graph with 204 edges are not observed. For the second set of simulated graphs (corresponding to the second row of Figure P.2), 51 edges are added spuriously from a small-world graph with 102 edges. All generated graphs have finally 153 edges. We then compare the obtained graph with fully generated small-world graphs of 51 nodes and 153 edges. The objective is to observe the difference in metrics of the graphs under these two sets of simulations. The number of edges is determined by the necessity to have the same number of vertices and edges to make the comparison possible. Based on the computation of two graph metrics, global efficiency and clustering (Achard and Bullmore 2007), we can observe that the effect on graph metrics is more important when type I errors are made. Indeed, the difference in the distribution of global efficiency and clustering coefficient is less important between graphs with non observed edges and small-world graphs than between graphs with additional edges and small-world graphs. This justifies also our choice to control the FWER in Chapter 4, which is an efficient control of type I error.

This simulation illustrates that misspecification may indeed influence graph metrics, but a more detail study would be interesting to carry out.

Simulation of correlation matrices

Simulation of realistic dependence structures is an ongoing challenge (see Pourahmadi and Wang (2015) and references therein) and imposing a latent graph structure is a difficult task (Córdoba, Varando, Bielza, and Larrañaga 2020). In the work described in Chapter 4, we were limited in our simulation studies, and the validation of graph inference on simulation is compromised. Indeed, to ensure the positive definiteness of the covariance matrix, non-zero off-diagonal elements are generated with small values. In usual simulation schemes, dealing with a uniform distribution, the correlations distribution is not realistic (see [16]). With Kevin

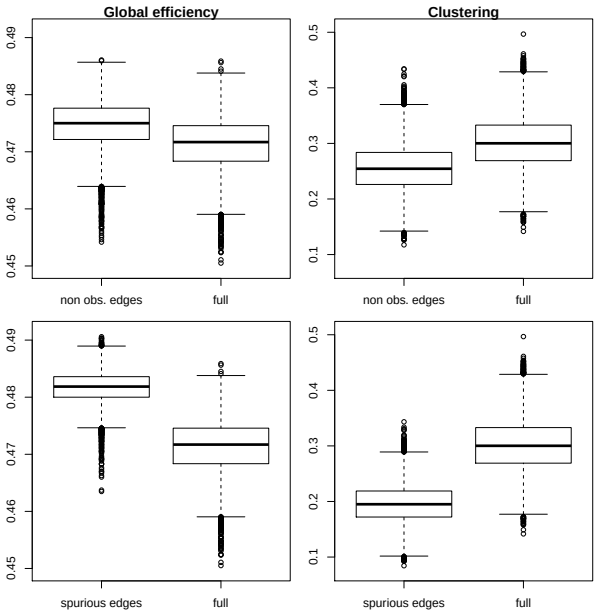


Figure P.2: Influence of missidentifications of edges on Global Efficiency and on Clustering coefficient. 10000 simulations were done on 51 nodes small-world graphs with 153 edges. Missidentifications consist in 33% of spurious edges or 75% of only correctly observed edges, that is, in each case 51 edges misspecified.

Polisano (LJK, Univ. Grenoble Alpes) and Sophie Achard (LJK, Univ. Grenoble Alpes) we have studied in [16] the influence of the graph structure on the distribution of the correlations. We would like to develop now algorithms which simulate correlation matrices with latent graph structures in a wider variety of distributions.

Partial Least Squares regression

When I have applied functional partial least squares regression (Section 2.1), I have noted how much theoretical work still has to be done in this model. Interested by this topic, I currently co-supervise the PhD of Luca Castelli on this modeling with Clément Marteau (ICJ, Univ. Lyon 1). The starting point is the work of Cook and Forzani (2018) and Cook and Forzani (2019), which gives interesting results in such models. We are not considering for now time series or dependent data, but this can be a perspective.

Grading-based votes

I am currently working on other topics, which are not included in this manuscript. For example, I have studied voting procedures based on evaluations rather than preferences in Aubin, Gannaz, Leoni-Aubin, and Rolland (2022) [3]. The idea is to represent the grades of each voter on d candidates as a point in $\mathbb{R}^d$ and to define the winner of the vote using the deepest point of the scatter plot. The deepest point is obtained by the maximization of a depth function (Zuo and Serfling 2000). This allows to present grading-based voting processes in a unified framework. By continuity, I am working on simulations for social choice theory with Jean-Baptiste Aubin (ICJ, INSA Lyon), Samuela Leoni (ICJ, INSA Lyon), and Antoine Rolland (ERIC, Univ. Lyon 2). The objective is to propose simulation procedures for grading-based votes, in order to study their empirical properties.

And...

...I hope to discover many more applications and various fields of research!

References

Achard, S. and E. Bullmore (2007). "Efficiency and cost of economical human brain functional networks". In: *PLoS Computational Biology* 3, e17.

Achard, S., P. Borgnat, I. Gannaz, and M. Roux (2019). "Wavelet-based graph inference using multiple testing". In: *Wavelets and Sparsity XVIII*. Vol. 11138. International Society for Optics and Photonics, p. 1113811.

Achard, S., C. Delon-Martin, P. E. Vértes, F. Renard, M. Schenck, F. Schneider, C. Heinrich, S. Kremer, and E. T. Bullmore (2012). "Hubs of brain functional networks are radically reorganized in comatose patients". In: *Proceedings of the National Academy of Sciences* 109.50, pp. 20608–20613.

Amovin-Assagba, M., I. Gannaz, and J. Jacques (2022). "Outlier detection in multivariate functional data through a contaminated mixture model". In: *Computational Statistics & Data Analysis* 174, p. 107496.

Aubin, J.-B., I. Gannaz, S. Leoni-Aubin, and A. Rolland (2022). "Deepest Voting: a new way of electing". In: *Mathematical Social Sciences* 116, pp. 1–16.

Baek, C., S. Kechagias, and V. Pipiras (2017). "Semiparametric, parametric, and possibly sparse models for multivariate long-range dependence". In: *Proceedings Volume 10394, Wavelets and Sparsity XVII*.

Bellas, A., C. Bouveyron, M. Cottrell, and J. Lacaille (2013). "Model-based clustering of high-dimensional data streams with online mixture of probabilistic PCA". In: *Advances in Data Analysis and Classification* 7.3, pp. 281–300.

Bouveyron, C. (2014). "Adaptive mixture discriminant analysis for supervised learning with unobserved classes". In: *Journal of Classification* 31.1, pp. 49–84.

Bühlmann, P., P. Doukhan, and P. A. Nze (2002). "Weak dependence beyond mixing and asymptotics for nonparametric regression". In: *The Annals of Statistics* 30.2, pp. 397–430.

Bullmore, E. and O. Sporns (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature reviews neuroscience* 10.3, pp. 186–198.

Cook, R. D. and L. Forzani (2018). "Big data and partial least-squares prediction". In: *Canadian Journal of Statistics* 46.1, pp. 62–78.

Cook, R. D. and L. Forzani (2019). "Partial least squares prediction in high-dimensional regression". In: *The Annals of Statistics* 47.2, pp. 884–908.

Córdoba, I., G. Varando, C. Bielza, and P. Larrañaga (2020). "On generating random Gaussian graphical models". In: *International Journal of Approximate Reasoning* 125, pp. 240–250.

Düker, M.-C. and V. Pipiras (2019). "Asymptotic results for multivariate local Whittle estimation with applications". In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, pp. 584–588.

Forbes, F. and D. Wraith (2014). "A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering". In: *Statistics and computing* 24.6, pp. 971–984.

Lerasle, M. (2009). "Adaptive density estimation of stationary β-mixing and τ-mixing processes". In: *Mathematical Methods of statistics* 18.1, pp. 59–83.

Lohmann, G., D. S. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, and R. Turner (2010). "Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain". In: *PloS one* 5.4, e10232.

Maretic, H. P. and P. Frossard (2020). "Graph Laplacian mixture model". In: *IEEE Transactions on Signal and Information Processing over Networks* 6, pp. 261–270.

Pourahmadi, M. and X. Wang (2015). "Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor". English. In: *Statistics and Probability Letters* 106.Complete, pp. 5–12.

Ragache, N. and O. Wintenberger (2006). "Convergence rates for density estimators of weakly dependent time series". In: *Dependence in probability and statistics*. Springer, pp. 349–372.

Rubinov, M. and O. Sporns (2010). "Complex network measures of brain connectivity: uses and interpretations". In: *NeuroImage* 52.3, pp. 1059–1069.

Zuo, Y. and R. Serfling (2000). "General notions of statistical depth function". In: *Annals of statistics*, pp. 461–482.

List of Figures

Chapter 4

Chapter 5

Perspectives

Mes publications

Les liens fournis ci-après sont ceux des prépublications HAL.

Articles dans des revues internationales à comité de lecture

[1] I. Gannaz (2023). *Asymptotic normality of wavelet covariances and of multivariate wavelet Whittle estimators.* Stochastic Processes and their Applications, Vol 155, pp 485-534. <hal-03068460v2>

[2] M. Amovin-Assagba, I. Gannaz, J. Jacques (2022). *Outlier detection in multivariate functional data through a contaminated mixture model*. Computational Statistics and Data Analysis, Vol 174, pp 107496. <hal-03255849>

[3] J.-B. Aubin, I. Gannaz, S. Leoni-Aubin, A. Rolland (2022). *Deepest Voting a new way of electing.* Mathematical Social Sciences, Vol 116, pp 1-16. <hal-03192793>

[4] S. Achard, M. Clausel, I. Gannaz, F. Roueff (2019). *New results on approximate Hilbert pairs of wavelet filters with common factor structure.* Applied and Computational Harmonic Analysis, Vol 49, N°3, pp 1025-1045. <hal-01613583>

[5] S. Achard, I. Gannaz (2019). *Wavelet-based multivariate Whittle estimation, comparison with Fourier multiwave.* Journal of Statistical Software, Vol 69, N°6, pp 1-31. <10.18637/jss.v089.i06>

[6] S. Achard, I. Gannaz (2016). *Multivariate wavelet Whittle estimation in long-range dependence*. Journal of Time Series Analysis, Vol. 37, No 4, p. 476–512. <hal-01079645>

[7] I. Gannaz (2013). *Wavelet penalized likelihood estimation in generalized functional models*. TEST, Vol. 22, n^o 1, p. 122-158. <hal-00577733>

[8] I. Gannaz, O. Wintenberger (2010). *Adaptive density estimation with dependent observations*. ESAIM P&S, Vol. 14, p. 151-172. <hal-00591694>

[9] I. Gannaz (2007). *Robust estimation and wavelet thresholding in partially linear models*. Statistics and Computing, Vol. 17, p. 293-310. <hal-00118237>

Articles soumis dans des journaux à comité de lecture

[10] S. Achard, I. Gannaz (2022). *Whittle estimation with (quasi-)analytic wavelets.* <hal-03272326v2>

[11] A. Rolland, J.-B. Aubin, I. Gannaz, S. Leoni (2022). *A Note on Data Simulations for Voting by Evaluation.* <hal-03194218>

Articles de conférences internationales avec actes ou chapitres de livres

[12] S. Achard, P. Borgnat, I. Gannaz, M. Roux (2019). *Wavelet-based graph inference using multiple testing.* In Wavelets and Sparsity XVIII (Vol. 11138, p. 1113811). International Society for Optics and Photonics. <hal-02399391>

[13] S. Achard, I. Gannaz (2018) *Wavelet Whittle Estimation in Multivariate Time Series Models: Application to fMRI Data.* In: Bertail P., Blanke D., Cornillon PA., Matzner-Løber E. (eds) Nonparametric Statistics. ISNPS 2016. Springer Proceedings in Mathematics & Statistics, vol 250. Springer, Cham.

[14] I. Gannaz, S. Achard, M. Clausel, F. Roueff (2017). *Analytic wavelets for multivariate time series analysis.* In Wavelets and Sparsity XVII, Vol. 10394, p. 103941X. International Society for Optics and Photonics. <hal-01618447>

[15] I. Gannaz (2014). *Classification of EEG recordings in auditory brain activity via a logistic functional linear regression model.* Contributions in infinite-dimensional statistics and related topics, International Workshop on Functional and Operatorial Statistics. <hal-00830313>

Articles de conférences nationales avec actes

[16] S. Achard, I. Gannaz, K. Polisano (2022). *Génération de modèles graphiques*. XXVII-Ième colloque GRETSI. <hal-03811642>

[17] M. Amovin-Assagba, I. Gannaz, J. Jacques (2022). *Clustering et détection d'anomalies dans les données fonctionnelles.* 53èmes Journées de Statistique de la Société Française de Statistique. <hal-03649201>

[18] J.B. Aubin, I. Gannaz, S. Leoni, A. Rolland (2022). *Votes par évaluation avec des fonctions de profondeur.* 53èmes Journées de Statistique de la Société Française de Statistique. <hal-03810253>

[19] M. Amovin-Assagba, J. Jacques, I. Gannaz, F. Fossi, J. Mozul (2020). *Détection*

d'anomalies dans des données fonctionnelles multivariées. 52èmes Journées de Statistiques de la SFdS. <hal-02987148>

[20] I. Gannaz, S. Achard, M. Clausel, F. Roueff (2017). *Ondelettes analytiques, application à l'analyse des processus multivariés à longue mémoire.* XXVIème colloque GRETSI. <hal-01840082>

[21] F. Millioz, I. Gannaz (2017). *Estimation des paramètres d'un bruit gaussien généralisé basée sur le kurtosis des statistiques minimales.* XXVIème colloque GRETSI. <hal-01618926>

[22] I. Gannaz (2012). *Estimation par ondelettes dans des modèles fonctionnels généralisés*. 44èmes Journées de Statistique, Bruxelles : Belgique. <hal-00863163>

Packages R

[23] I. Gannaz, M. Roux (2019). *TestCor: FWER and FDR Controlling Procedures for Multiple Correlation Tests.* R package. <lien>

[24] S. Achard, I. Gannaz (2016). *multiwave: Estimation of Multivariate Long-Memory Models Parameters.* R package. <lien>

Rapports de recherche industrielle

[25] I. Gannaz, J. Jacques (2017). *Projet MasterK.* Rapport de recherche (confidentiel).

[26] I. Gannaz, F. Millioz (2017). *Extraction de caractéristiques cardiaques sur des signaux multimodaux.* Rapport de recherche (confidentiel).

[27] I. Gannaz, F. Millioz (2016). *Aide au diagnostique de pathologies cardiaques.* Rapport de recherche (confidentiel).

Rapport de recherche académique

[28] S. Achard, P. Borgnat, I. Gannaz (2020). *Asymptotic control of FWER under Gaussian assumption: application to correlation tests.* <hal-02883720v1>

Thèse

[29] I. Gannaz (2007). *Estimation par ondelettes dans les modèles partiellement linéaires.* Université Grenoble 1, sous la direction d'A. Antoniadis. <tel-00197146>