Wavelet-based graph inference using multiple testing

Sophie Achard, Pierre Borgnat, Irène Gannaz, Marine Roux

Wavelet-based graph inference using multiple testing

Sophie Achard[a], Pierre Borgnat[b], Irène Gannaz[c], and Marine Roux[a]

[a]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[b]Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France
[c]Univ Lyon, INSA de Lyon, CNRS UMR 5208, Institut Camille Jordan, F-69621 Villeurbanne, France

ABSTRACT

Graph-based representation enables to outline efficiently interactions between sensors and as such has encountered a growing interest. For example in neurosciences, the graph of interactions between brain regions has shed lights on evolution of diseases. In this paper, we describe a whole procedure which estimates the graph from multivariate time series. First correlations using wavelet decomposition of the signals are estimated. Bonferroni (1935)'s procedure on multiple correlation testing is then used. We prove theoretically that the Family Wise Error Rate (FWER) is asymptotically controlled for any graph structures. We implement our approach on small-world graph structures, with signals possibly having long-memory properties. This structure is inspired by real data examples from resting-state functional magnetic resonance imaging. The control is confirmed graphically. Numerical simulations illustrate the behavior of the bias and the power of our proposed approach.

Keywords: wavelets, multiple testing, correlation, graph inference

1. INTRODUCTION

Brain connectivity analysis is very important in order to understand the mechanism implicated in brain pathologies such as for example schizophrenia,[1] Parkinson,[2] strokes,[3] traumatic brain injury[4]... With the use of non invasive recording of brain functioning it is nowadays easy to collect data on groups of patients covering the whole brain. Usually, the brain is parcellated in regions of interest with the objective to identify the possible pairs of brain regions dependent to each other.[5] This creates a graph where the nodes are the brain regions and the edges represent the dependence between two brain regions. The parcellation of the brain is well documented with several papers proposing different parcellation schemes.[6] The choice for the definition of dependence between nodes is more complicated. As temporal signals are extracted from brain regions, the dependence is measured via a statistical estimation facing classical statistical challenges such as number of available samples, long-memory time series, multiple comparisons... A poor performance of statistical estimations may have dramatic effect on the resulting extracted graphs and consequently on the graph metrics used to compare groups of patients and controls.[7–10]

In this paper, we provide a whole method from signals to graphs in order to control the misdetection of edges. The dependence is estimated using wavelet correlations.[11,12] For a given frequency band, pairs of correlations are estimated using the wavelet coefficients. Then, thanks to the good convergence properties of the estimator, p-values are extracted for each pair. The effect of making error on identification of edges is displayed in Figure 1. Two different paradigms on classical graph metrics are illustrated *via* global efficiency and local efficiency. Adding wrong edges at random has a dramatic effect on the graph metrics, however, missing true edges is less dramatic. The selection of significant pairs of correlation is then derived so as to control asymptotically the family wise error rate, *i.e.* the probability of having more than one wrong edge. In practice, we choose to apply the classical Bonferroni procedure on the p-values of wavelet correlations. The paper is organized in 4 different parts. The

Further author information: (Send correspondence to Sophie Achard)
Sophie Achard: E-mail: sophie.achard@univ-grenoble-alpes.fr, Telephone: +33 (0)4 76 57 43 52

first part is dedicated to the process to start from brain signals and end up with correlation matrices. The second part gives details on the multiple testing approach in order to extract the graphs. Simulations are proposed in a third part, where we took the classical example of small-world graphs. Finally, our approach is applied to a real fMRI dataset on rats.

2. FROM SIGNALS TO CORRELATION MATRICES USING WAVELETS

Understanding multivariate time series is the main motivation of our paper. It is now well recognized that the time series observed from brain recordings are well modeled using long-memory time series.[13,14] Also, from a neuroscience point of view it is common to study the low frequencies part of the signals. Wavelets are perfect tools to both take into account the long-memory and focus on low frequencies. Indeed, wavelets are acting as frequency decompositions of the signals. Besides several results on the good performances of wavelets estimations for long-memory time series have been obtained in a univariate context.[15,16] The estimation of interactions between brain regions is described in this part using wavelet correlation estimations.

2.1 Multivariate long-memory models

Let $\mathbf{X} = \{X_\ell(k), k \in \mathbb{Z}, \ell = 1, \ldots, p\}$ be a multivariate stochastic process. Each process X_ℓ is not necessarily stationary. Denote by ΔX_ℓ the first order difference, $(\Delta X_\ell)(k) = X_\ell(k) - X_\ell(k-1)$, and by $\Delta^D X_\ell$ the D-th order difference. For every component X_ℓ, there exists $D_\ell \in \mathbb{N}$ such that the D_ℓ-th order difference $\Delta^{D_\ell} X_\ell$ is covariance stationary. Following,[17–19] we consider a long-memory process $\mathbf{X}$ with memory parameters $\mathbf{d} = (d_1, d_2, \ldots, d_p)$. For any $\mathbf{D} > \mathbf{d} - 1/2$, we suppose that the multivariate process $\mathbf{Z} = \mathrm{diag}(\Delta^{D_\ell}, \ell = 1, \ldots, p)\mathbf{X}$ is covariance stationary with a spectral density matrix given by

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

where the long-memory parameters are given by $d_m^S = d_m - D_m$ for all m. The functions $f_{\ell,m}^S(\cdot)$ correspond to the short-memory behavior of the process. We assume that $f_{\ell,m}^S(0) = 1$ and that there exist Ł > 0 and $\beta \in (0, 2)$ such that for all (ℓ, m), $\max_{\lambda \in [-\pi,\pi]} |f_{\ell,m}(\lambda) - 1| \leq L|\lambda|^\beta$.

The generalized cross-spectral density of processes X_ℓ and X_m can be written as

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

A process $\mathbf{X}$ satisfying the above conditions will be denoted in the following as a $M(\mathbf{d})$ process.

2.2 Wavelets decomposition

Let $(\phi(\cdot), \psi(\cdot))$ be respectively a father and a mother wavelets. Their Fourier transforms are given by $\widehat{\phi}(\lambda) = \int_{-\infty}^{\infty} \phi(t) e^{-i\lambda t} dt$ and $\widehat{\psi}(\lambda) = \int_{-\infty}^{\infty} \psi(t) e^{-i\lambda t} dt$.

At a given resolution $j \geq 0$, 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 $\psi_{j,k}(\cdot) = 2^{-j/2}\psi(2^{-j}\cdot - k)$. Thoughout the paper, we adopt the same convention as in[18] and,[15] that is, large values of the scale index j correspond to coarse scales (low frequencies).

Let $\tilde{\mathbf{X}}(t) = \sum_{k \in \mathbb{Z}} \mathbf{X}(k)\phi(t - k)$. The wavelet coefficients of the process $\mathbf{X}$ are defined by

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

For given $j \geq 0$ and $k \in \mathbb{Z}$, $\mathbf{W}_{j,k}$ is a p-dimensional vector $\mathbf{W}_{jk} = \begin{pmatrix} W_{j,k}(1) & W_{j,k}(2) & \ldots & W_{j,k}(p) \end{pmatrix}$, where $W_{j,k}(\ell) = \int_{\mathbb{R}} \tilde{X}_\ell(t)\psi_{j,k}(t)dt$.

The regularity conditions on the wavelet transform are expressed in the following assumptions. They will be needed throughout the paper.

(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 $\ell = 1, \ldots, M-1$.

(W5) For all $i = 1, \ldots, p$, $(1+\beta)/2 - \alpha < d_i \leq M$.

For all $j \geq 0$, we will denote also by $\mathcal{H}_j$ the squared gain function associated to wavelet filters at scale j.

In practice, a finite number of realization of the process $\mathbf{X}$, say $\mathbf{X}(1), \ldots \mathbf{X}(n)$, is observed. Assumption (W1) ensures that the wavelets have a compact support. Thus at each scale j only a finite number n_j of coefficients are non zero.

2.3 Wavelet correlation

Statistical convergence results of variance, covariance and correlation have already been established *e.g* by[20] and[11] for finite difference processes. Confidence intervals are then derived and statistical hypotheses tests can be constructed.

PROPOSITION 2.1 (CONVERGENCE OF WAVELET CORRELATIONS). *Let $\mathbf{X}(1), \ldots, \mathbf{X}(n)$ be n observations of a p-dimensional real valued $M(d)$ process with a spectral density $\boldsymbol{f}$. Define for all ℓ, $1 \leq \ell \leq p$, $\{W_{j,k}(\ell),\, k \in \mathbb{Z}\}$ the wavelets coefficients at scale j of X_ℓ. Assume that for all ℓ, $1 \leq \ell \leq p$, the process $\{W_{j,k}(\ell)\}_k$ is a Gaussian process of spectral density $D^{(j)}_{0;0}(.;\ell,\ell)$ so that $D^{(j)}_{0;0}(f;\ell,\ell) > 0$ almost everywhere, and the bivariate process $\{W_{j,k}(\ell), W_{j,k}(m)\}_k$ is Gaussian with spectral density $D^{(j)}_{0;0}(.;\ell,m)$ so that $D^{(j)}_{0;0}(f;\ell,m) > 0$ almost everywhere.*

Under the hypotheses (W1)-(W5), the empirical estimator of correlation of wavelets coefficients at scale j is unbiased and is asymptotically Gaussian distributed:

$$\widehat{\rho}_{\ell,m}(j) := \frac{\widehat{\theta}_{\ell,m}}{\widehat{\sigma}_\ell(j)\widehat{\sigma}_m(j)} = \frac{\sum_{k=0}^{n_j}(W_{j,k}(\ell)W_{j,k}(m))}{(\sum_{k=0}^{n_j} W^2_{j,k}(\ell)\sum_{k=0}^{n_j} W^2_{j,k}(m))^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\rho_{\ell,m}(j), \Sigma^{cc}_j\right) \tag{2.1}$$

where

$$\rho_{\ell,m}(j) := \frac{\gamma_{\ell,m}(j)}{(\gamma_{\ell,\ell}(j)\gamma_{m,m}(j))^{1/2}},\ \textit{with}\ \gamma_{\ell,m}(j) := \int \mathcal{H}_j(\lambda)\, f_{\ell,m}(\lambda)\, d\,\lambda. \tag{2.2}$$

The matrices Σ^{cc}_j are given in formula (5.4) of.[21] The symbol $\xrightarrow{\mathcal{L}}$ corresponds to the convergence in distribution when n tends to infinity.

Brain recordings often present long-range dependence, with non homogeneity of long-memory parameters $\mathbf{d}$. In these cases the tested values $\{\rho_{\ell,m}(j),\, 1 \leq \ell < m \leq p,\, j \geq 0\}$ depend on the parameters $\mathbf{d}$. We refer to Proposition 1 and Proposition 2 of.[22] More precisely, suppose that at a first order approximation the spectral density satisfies, when $\lambda \to 0^+$,

$$\boldsymbol{f}(\lambda) \sim \boldsymbol{\Sigma}(\lambda), \quad \boldsymbol{\Sigma}(\lambda) = \boldsymbol{\Lambda}(\mathbf{d})^* \boldsymbol{\Omega} \boldsymbol{\Lambda}(\mathbf{d})\,,\ \text{with}\ \boldsymbol{\Lambda}(\mathbf{d}) = \mathrm{diag}(\lambda^{-\mathbf{d}} e^{-i\pi\mathbf{d}/2})\,,$$

where the $*$ exponent denotes the complex conjugate. This approximation holds for $M(\mathbf{d})$ processes defined in Section 2.2. Let r be the correlation coefficients of $\boldsymbol{\Sigma}(\lambda)$, $r_{\ell,m} = \Sigma_{\ell,m}(\lambda)/\sqrt{\Sigma_{\ell,\ell}(\lambda)\Sigma_{m,m}(\lambda)} = \Omega_{\ell,m}\, e^{i\,\pi(d_\ell - d_m)/2}$. When ψ is real, Proposition 2 of[22] shows that asymptotically for all (ℓ, m), $\rho_{\ell,m}(j)$ tends to the real part of r when j tends to infinity, that is,

$$\rho_{\ell,m}(j) \underset{j\to\infty}{\longrightarrow} Re(r_{\ell,m}) = \Omega_{\ell,m}\, \cos(\pi(d_\ell - d_m)/2). \tag{2.3}$$

An interesting perspective is to build an estimation procedure on $\{\Omega_{\ell,m}\}_{(\ell,m)}$ rather then $\{r_{\ell,m}\}_{(\ell,m)}$. In particular, notice that when $d_\ell - d_m$ is close to 1 *modulo* 2, the correlation of the wavelet coefficients $\rho_{\ell,m}(j)$ tends to 0 even if $\Omega_{\ell,m} \neq 0$.

2.4 Fisher transform of wavelet correlation

The exact formula of the asymptotic variance of wavelet correlation 2.1 is complicated and it is tempting to get rid of this computation. It is possible by using the Fisher transform of the wavelet coefficients. Simply as in the usual case, the Fisher transform is stabilizing the variance of the estimator. In our case, this facilitates the computation of the confidence intervals where the variance is now simply a function of the number of samples.[23] Let z be the Fisher transform: for all x, $-1 \leq x \leq 1$,

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

PROPOSITION 2.2 (CONVERGENCE OF CORRELATION WITH FISHER TRANSFORM). *Under the same hypotheses of proposition 2.1, the estimator of empirical correlation verifies*

$$\sqrt{(n_j - 3)}(z(\widehat{\rho_{\ell,m}}(j)) - z(\rho_{\ell,m}(j))) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\,; \tag{2.4}$$

In the following, we will apply this result for constructing the multiple testing framework to test a large sample of pairwise correlations at a given wavelet scale j, that is,

$$H_{0,\ell m}(j) : \rho_{\ell,m}(j) = 0 \;\text{ against }\; H_{1,\ell m}(j) : \rho_{\ell,m}(j) \neq 0, \quad 1 \leq \ell < m \leq p, \quad j \geq 0. \tag{2.5}$$

3. MULTIPLE TESTING FRAMEWORK

In the previous section, we proposed a procedure which gives correlation matrices, where each entry $\widehat{\rho}_{\ell,m}(j)$ is the estimated correlation between brain regions ℓ and m at scale j. To infer the dependence graph, there is a need to select significant correlations based on a correct statistical approach.

Let V denote the set of the indexes of the brain regions, $V = \{1, \ldots, p\}$. For a given scale j, define graphs G_j as a pair $G_j = (V, E_j)$ where the set of edges E_j is a subset of $V \times V$. We consider undirected graphs, that is, if $(\ell, m) \in E_j$, then $(m, \ell) \in E_j$. The dependence graph is obtained by defining $E_j = \{(\ell, m),\, \rho_{\ell,m}(j) \neq 0\}$.

From the estimated correlation matrices obtained in Section 2, our objective is to estimate graphs G_j for each scale j. Denote $\widehat{G}_j = (V, \widehat{E}_j)$ the estimated graphs. We hence test $H_{0,\ell m}(j) : \rho_{\ell,m}(j) = 0$ against $H_{1,\ell m}(j) : \rho_{\ell,m}(j) \neq 0$, for all $(\ell, m) \in V$. The set of edges $\widehat{E}_j$ is defined as $E_j = \{(\ell, m) \in V \times V,\, H_{0,\ell m}(j) \text{ is rejected}\}$. For each graph G_j, estimation $\widehat{G}_j$ is obtained applying $\nu = p\,(p-1)/2$ tests (since sets E_j are symmetric). We now detail the procedure of testing.

3.1 Multiple hypothesis tests

Consider the specific case of correlations. Let $\mathbb{X}^{(n)} = \{X(1), \ldots, X(n)\}$ be independent realizations from a random vector $X = (X_{(\ell)})_{\ell=1,\ldots,p}$ with values in $\mathbb{R}^p$. Define $\Gamma = (\rho_{\ell,m})_{\ell,m=1,\ldots,p}$ the correlation matrix of X, with $\rho_{\ell,m} = \mathrm{Cor}(X_{(\ell)}, X_{(m)})$. Consider the two-sided testing problem:

$$H_{0,\ell m} : \rho_{\ell,m} = 0 \;\text{ against }\; H_{1,\ell m} : \rho_{\ell,m} \neq 0, \quad (\ell, m) \in \mathcal{H}, \tag{3.1}$$

where $\mathcal{H}$ is the set of indexes tested. In the following $\mathcal{H} = \{1 \leq \ell < m \leq p\}$. The number of tests corresponds to the cardinality of $\mathcal{H}$, that is, $\nu = p\,(p-1)/2$.

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

$$\mathcal{R} = \{(\ell, m) \in \mathcal{H},\; (H_{0,\ell m}) \text{ rejected}\},$$

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

$$\mathrm{FWER}(\mathcal{R}) = \mathbb{P}\left(\exists (\ell, m) \in \mathcal{R},\; \rho_{\ell,m} = 0\right).$$

For a given level α, the objective is to find $\mathcal{R}$ such that $\mathrm{FWER}(\mathcal{R}) \leq \alpha$. The rejection set is built on a p-value process associated to the tests. Note that in our particular case, the p-value processes associated to tests (3.1) present a dependence structure.[24]

In our proposed graph estimation, the rejection set $\mathcal{R}$ gives the estimated set of edges, $\widehat{E}$. That is, the estimated graph is $\widehat{G} = (V, \widehat{E})$ where $\widehat{E} = \mathcal{R}$. The FWER criterion corresponds to the probability to have a false additional edge in the estimated graph. As we will illustrate in Section 3.3, controlling this quantity is important in order that usual metrics on the resulting graph remains satisfactory.

3.2 Asymptotic control of FWER

The Bonferroni procedure[25] is the most classical example of FWER control without the necessity to have independence of tests. A difficulty of the application of multiple testing in the context of correlations is that the tests are asymptotic. The Bonferroni method has then to be validated in this context.

Method (Bonferroni) The Bonferroni multiple testing procedure applied on Fisher statistics is defined by

$$\mathcal{R}^{\mathrm{bonf}}_{n,\alpha} = \left\{ (\ell, m) \in \mathcal{H} : p_{n,\ell m} \leq \frac{\alpha}{\nu} \right\}, \tag{3.2}$$

where

$$p_{n,\ell m} = 2\left[1 - \Phi\left(\sqrt{n-3}\left|z\left(\widehat{\rho}_{\ell m}\big(\mathbb{X}^{(n)}\big)\right)\right|\right)\right]. \tag{3.3}$$

Proposition 3.1. *Let $\mathcal{P}$ denote a family of probability distributions of the random vector X. For the two-sided testing problem* (3.1) *based on the asymptotic p-value process* (3.3)*, the method $\mathcal{R}^{\mathrm{bonf}}_{n,\alpha}$ provides an asymptotic control of the FWER at level α. That is, for all $P \in \mathcal{P}$,*

$$\lim_{n\to+\infty} \mathrm{FWER}\Big(\mathcal{R}^{\mathrm{bonf}}_{n,\alpha}, P\Big) \leq \alpha. \tag{3.4}$$

Proof. Let $P \in \mathcal{P}$. Denote $\mathcal{H}_0(P)$ the set of indexes such that $\rho_{\ell,m} = 0$ under distribution P, and $\nu_0(P)$ its cardinality. We have

$$\mathrm{FWER}\Big(\mathcal{R}^{\mathrm{bonf}}_{n,\alpha}, P\Big) = \mathbb{P}\left(\bigcup_{(\ell,m)\in\mathcal{H}_0(P)} \{p_{n,\ell m} \leq \alpha/\nu\}\right).$$

Thus,

$$\begin{aligned}
\lim_{n\to+\infty} \mathrm{FWER}\Big(\mathcal{R}^{\mathrm{bonf}}_{n,\alpha}, P\Big) &\leq \sum_{(\ell,m)\in\mathcal{H}_0(P)} \alpha/\nu \\
&\leq \frac{\alpha\,\nu_0(P)}{\nu} \\
&\leq \alpha.
\end{aligned}$$

□

This proposition is valid for any tests but in this paper, we focus our attention to correlations.

An equivalent formulation of Proposition 3.1 is written in terms of graphs in the following lemma,

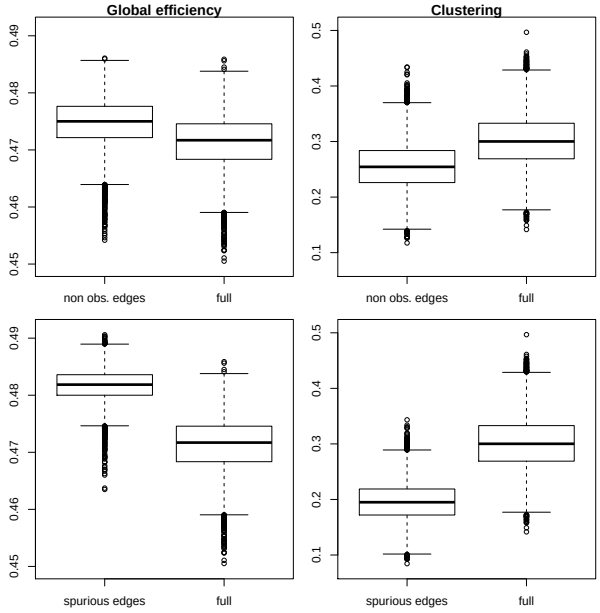
Lemma 3.1. *Let $\mathcal{P}$ denote a family of probability distributions of the random vector X. For $P \in \mathcal{P}$, let $G(P) = (V, E(P))$ be the dependence graph associated to the data, that is, $E(P) = \{(\ell, m) \in \mathcal{H},\, \rho_{\ell,m} \neq 0\}$. The estimated graph is $\widehat{G}_{n,\alpha} = (V, \widehat{E}_{n,\alpha})$ where $\widehat{E}_{n,\alpha} = \mathcal{R}^{\mathrm{bonf}}_{n,\alpha}$. Under the same hypotheses as in Proposition 3.1, for all $P \in \mathcal{P}$,*

$$\lim_{n\to+\infty} \mathrm{P}(\widehat{G}_{n,\alpha} \subseteq G(P)) \geq 1 - \alpha. \tag{3.5}$$

3.3 Importance of detection of correct edges on graphs

Classical methods to compare graphs in neuroscience are using graph metrics such as minimum path length, clustering, and many others.[26,27] The precision of edge detections have to be quantified in terms of these classical metrics. We conducted 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 one illustrated on the first row of Figure 1 corresponds to type II error *i.e.* not all the correct edges are observed, but the one present in the graph are correct. The second one illustrated on the second row of Figure 1 corresponds to type I error *i.e.* additional edges are added at random to the observed graph, and they are not correct. In the simulations, we considered graphs with 51 nodes. For the first set of simulated graphs (corresponding to the first row of Figure 1), 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 1), 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,[28] 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 which is an efficient control of type I error.

Figure 1: 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 spuriousedges or 75% of only correctly observed edges, that is, in each case 51 edges misspecified.



3.4 Application to wavelet correlation

In our context, a multiple testing procedure is set for each wavelet scales. At each scale $j \geq 0$, the Bonferroni multiple testing procedure applied on Fisher statistics is defined by

$$\mathcal{R}^{\text{bonf}}_{n,\alpha}(j) = \left\{(\ell, m) \in \mathcal{H} : p_{n,\ell m}(j) \leq \frac{\alpha}{\nu}\right\}, \tag{3.6}$$

where

$$p_{n,\ell m}(j) = 2\left[1 - \Phi\left(\sqrt{n-3}\,|z\left(\widehat{\rho}_{\ell m}(j)\right)|\right)\right], \tag{3.7}$$

and $\widehat{\rho}_{\ell m}(j)$ is defined as 2.1. Similarly to general correlation tests, we have an asymptotic control of FWER.

PROPOSITION 3.2. *Let $\mathbf{X}(1), \dots \mathbf{X}(n)$ be n independent observation of a $\mathbb{R}^p$-valued $M(\mathbf{d})$ process. Consider the two-sided testing problem* (2.5) *based on the asymptotic p-value process* (3.7)*, evaluated with wavelet filters satisfying (W1)–(W5). Introduce a scale j_0 such that $n2^{-j_0}$ tends to infinity when n goes to infinity.*

Under the same assumptions of Proposition 2.1, the method $\mathcal{R}^{\text{bonf}}_{n,\alpha}(j)$ provides an asymptotic control of the FWER at level α. That is, for all $j \geq j_0$,

$$\lim_{n\to+\infty} \text{FWER}\Big(\mathcal{R}^{\text{bonf}}_{n,\alpha}(j)\Big) \leq \alpha. \tag{3.8}$$

The available number of samples is varying depending on the scales. Because of the asymtotic control of the FWER, when dealing with finite time series, it is not advisable to go for very small scales. The choice of the wavelet scale is complicated and will not be discussed in this paper.[29]

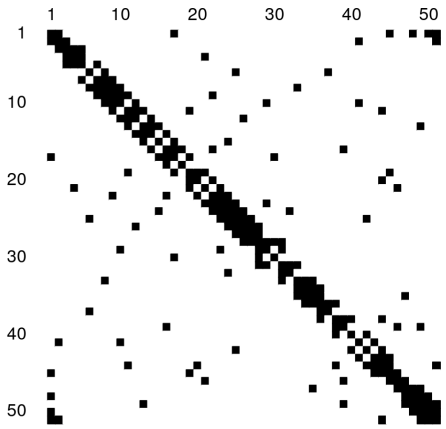
Besides the problem of size of available samples, the presence of short-memory affects the estimation of the graph. Indeed, for each scale $j \geq 0$ the graph G_j may be different. As the short-memory appears in the spectral density $\boldsymbol{f}$, it influences the values of $\{\rho_{\ell,m}(j),\, \ell, m = 1, \dots, p\}$ for low scales j. Note that, based on result (2.3), for sufficiently high scales, the graph G_j does not depend on j. Thus, the estimation of the graph $G^* = (V, E^*)$ where $E^* = \{(\ell, m) \in \{1, \dots, p\},\, \Omega_{\ell,m}) \neq 0\}$ may be better to consider. As shown in our previous paper,[29] the choice of wavelet scales for the estimation of Ω should include scales not affected by the short-memory and scales with enough remaining points. We illustrate this purpose on simulations in Section 4.

4. SIMULATIONS

In order to illustrate the different properties obtained theoretically in the previous section, we ran simulations on a small-world graph. The true graph is constructed with 51 nodes and 102 edges with small-world generative models as implemented in the `R` package `igraph`. The choice of the number of nodes and edges is motivated by the real data example of Section 5.

Let us denote $\mathbf{A}$ the adjacency matrix of this graph. An example is given in Figure 2. Then, let us define the correlation matrix as $\mathbf{C} = \mathbf{I} + \rho\mathbf{A}$, where $\mathbf{I}$ is the identity matrix of size p. The parameter ρ is chosen such that $\mathbf{C}$ is positive definite. This is the case when $|\rho| < 1/|\lambda_s|$, where λ_s is the smallest eigenvalues of $\mathbf{A}$.

Figure 2: Adjacency matrix of a small-world graph with 51 nodes and 102 edges.

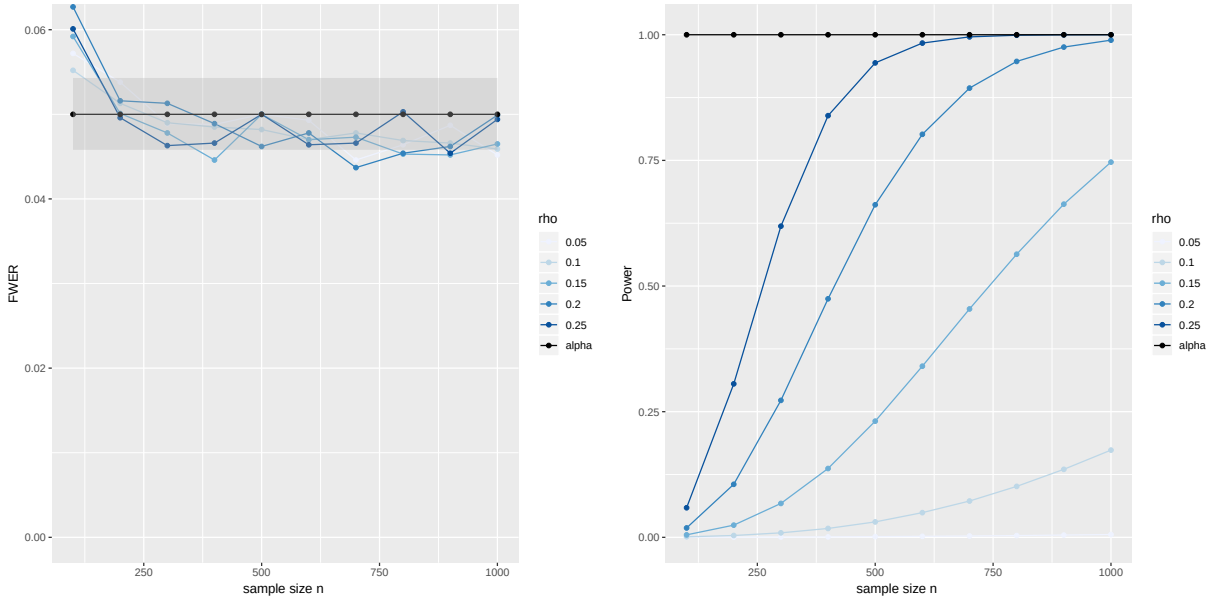


Simulations were done on a given graph adjacency matrix. Some extensions to numerical studies with different simulations of adjacency matrices are in progress. They are not displayed here for the sake of clarity.

4.1 Good behavior with a small-world generative models

Figure 3 is illustrating what we are expecting from theory. Indeed, the FWER is controlled as soon as the sample size is large enough. It is well known that controlling the FWER is not sufficient to have good results and the power of the procedure must be satisfactory. We also computed the power of our procedure. As illustrated in Figure 3 when the true correlation is too small, it is impossible to get an accurate solution, and no edges in the graph are detected. In order to get a correct solution, different parameters have to be taken into account: the size of the sample, the number of nodes in the graphs and the values of correlations in the correlation matrix. Similar conclusions were obtained *e.g.* by.[30]

Figure 3: FWER of edges inference with respect to the sample size for different values of ρ. 10000 simulations were done on 51 nodes small-world graphs. The grey interval around the value α corresponds to the confidence interval of a binomial distribution at level 95%. Schematically, there is now contradiction with the theory $FWER \leq \alpha$ even if the observed FWER is greater than α as long as it belongs to this interval. Power of edges inference : proportion of true edges detected with respect to the sample size for different values of ρ.



4.2 Wavelet correlation testing and influence of short-memory on FWER

As an example, we simulate FIVAR(1,**d**) processes[31] $\mathbf{X}$, defined as

$$\mathbf{X}(k) = \mathrm{diag}\big((1-\mathbb{L})^{-d_\ell}, \ell = 1, \ldots, p\big)\, \mathbf{u}(k),$$

where $\mathbb{L}$ is a lag operator and $\mathbf{u}$ follows a $\mathbb{R}^p$-valued VAR model:

$$\big(\mathbf{I}_p + \mathbf{\Phi}\,\mathbb{L}\big)\mathbf{u}(k) = \epsilon(k)$$

with $\epsilon(k) = \begin{pmatrix} \epsilon_1(k) \\ \vdots \\ \epsilon_p(k) \end{pmatrix}$ *i.i.d.* with a Gaussian distribution $\mathcal{N}_p\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{C}\right)$.

The covariance matrix $\mathbf{C}$ satisfies a small-world model, as described at the beginning of the section. Similarly we consider $p = 51$ components. The matrix $\mathbf{I}_p$ is the identity matrix in $\mathbb{R}^{p\times p}$ and $\mathbf{\Phi}$ is a matrix in $\mathbb{R}^{p\times p}$.

As stated previously, we are interested in the estimation of the graph $G^* = (V, E^*)$ where $E^* = \{(\ell, m) \in \{1, \ldots, p\},\ \Omega_{\ell,m}) \neq 0\}$. In presence of long-memory, the values $\{\rho_{\ell,m}(j),\ 1 \leq \ell < m \leq p,\ j \geq 0\}$ depend on the long-memory parameters $\mathbf{d}$. Due to equation (2.3) it is possible to apply a bias correction of the multiplicative

term $\cos(\pi(d_\ell - d_m)/2)$. However, for clarity, the influence of the heterogeneity of long-range dependence is not studied in our simulations. Then, we will take d_ℓ constant, equal here to 0.2 in all our simulations.

To study the influence of short-range memory properties, we will consider three cases for the VAR part of the model:

(a) First the matrix $\mathbf{\Phi}$ in the FIVAR model will be taken equal to the zero matrix, giving simulations with no short-range memory: $\Phi_{\ell,m} = 0$ for all $1 \leq \ell, m \leq p$.

(b) Next, we will consider a matrix were all components are concerned by a VAR part: $\mathbf{\Phi} = \begin{pmatrix} 0.8 & & \\ & \ddots & 0 \\ & & \ddots \\ 0.01 & & 0.8 \end{pmatrix}$.

(c) Finally, we provide simulations with a matrix $\mathbf{\Phi}$ such that only a part of the components are concerned with short-memory: $\mathbf{\Phi} = \begin{pmatrix} 0.8 & & 0 & \\ & \ddots & & 0 \\ 0.01 & & 0.8 & \\ \hline 0 & & 0 & \end{pmatrix}$, with only the upper 10×10 block matrix with non-zero values.

As we want to recover $E^* = \{(\ell, m) \in \{1, \ldots, p\},\ \Omega_{\ell,m}) \neq 0\}$, the short-memory effect of FIVAR model has a dramatic effect on FWER, Figure 4. This is due to the fact that $\rho_{\ell,m}(j)$ depends of the short-range memory and thus $\rho_{\ell,m}(j) \neq 0$ does not necessarily implies that $\Omega_{\ell,m} \neq 0$. The FWER of (2.5) is still controlled but the inferred graph is far from G^*. This problem does not come from the test procedure but from a bias in the statistics estimation. Notice that if the probability to make a false edge detection is not controlled in the first scales, due to short-memory, the number of false positives remains low. Figure 5 illustrates that except for the first scale, only 1 or 2 edges are wrongly put in the rejection set. Finally, when less components are concerned by the VAR part, we observe that much less scales are perturbed by the presence of short-memory.

Additionally, since the procedure is asymptotic, the FWER control is not ensured for small values of n_j. This explains why for highest scales (coarsest frequencies) the control is not acquired.

5. FMRI DATA WITH DEAD RATS

The dataset consists in fMRI recordings on rats. Four rats are scanned dead (rats 1 to 4), and seven rats are scanned alive under aneasthetic (rats 5 to 11). 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,[32] we extracted 51 time series, each time series being associated with a brain region of rats.

As explained in Section 2, we first estimate the correlation using a wavelet decomposition. Relatively to previous discussions, we will focus on wavelet scale 4 corresponding to the frequency interval [0.06 ; 0.12] Hz. The number of wavelet coefficients available is $n_4 = 122$. As seen in simulations of Section 4.2 the expected power is not far from 1 with such a number of coefficients (see Figure 6). Examples of correlation matrices obtained are given in Figure 7. Even if dead rats correlation matrices are much closer to identity matrices than correlation matrices on alive rats, some non zero values appear which need a test of significance.

Next, we define the rejection set of (2.5) applying Bonferroni's multiple testing procedure on Fisher statistics. As we have to deal with $50 \times 51/2 = 1275$ tests for each of the 11 rats, a multiple testing procedure is necessary. Without a multiple testing correction on p-values, the inferred graph is highly biased, as illustrated by the clearly non-empty graph for a dead rat displayed in Figure 8.

Figure 4: Empirical probability to make a false edge detection with respect to the scale. The horizontal plain line corresponds to the threshold $\alpha = 0.05$. 1000 simulations were done with FIVAR models, with (a) the VAR part is null (b) all the components are concerned with a non null VAR part (c) 10 of the 51 components are concerned with a non null VAR part.

(a) (b) (c)

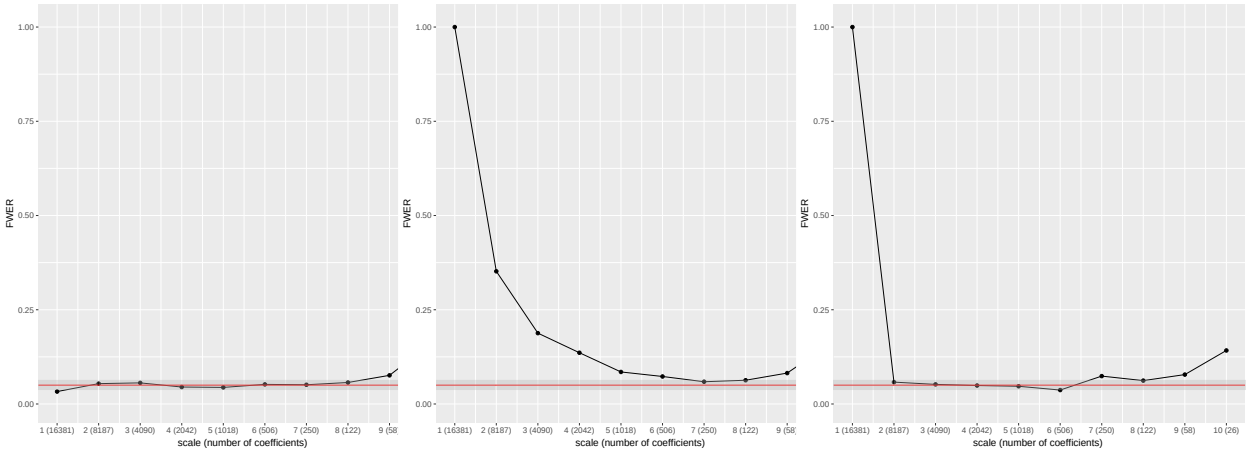
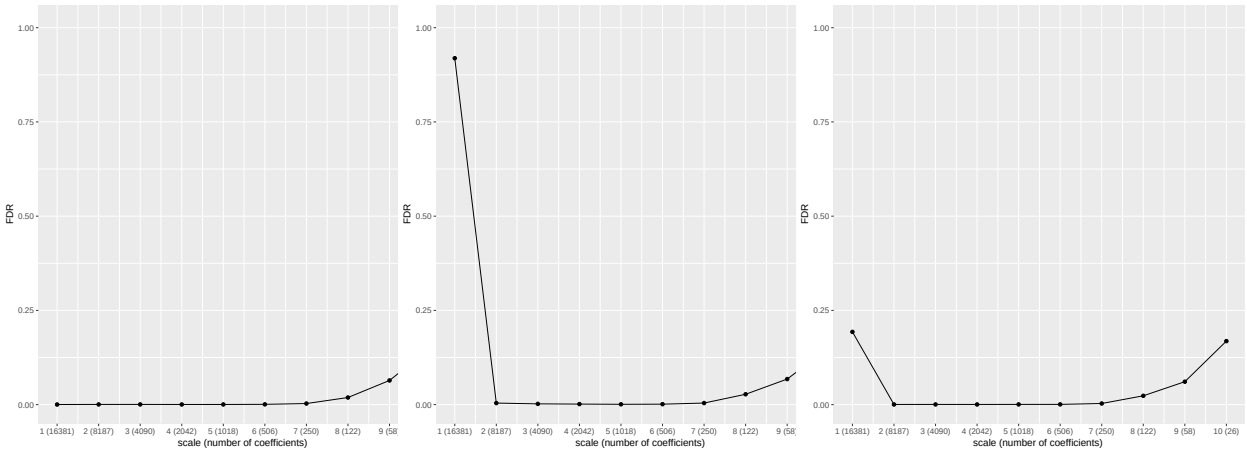


Figure 5: Empirical proportion of false edge detections in the rejection set with respect to the scale. 1000 simulations were done with FIVAR models, with (a) the VAR part is null (b) all the components are concerned with a non null VAR part (c) 10 of the 51 components are concerned with a non null VAR part.

(a) (b) (c)



It is not possible in general to evaluate the accuracy of procedures on real dataset because of the lack of ground truth. However, since there is actually no cerebral activity in a brain of a dead rat, all null hypotheses to be tested are true nulls and the number of rejected hypotheses must be equal to zero in this case. Indeed, we obtain that the number of significant correlations is zero or near to zero for dead rats (see Table 1). We can still observe a few remaining connections in the graphs from very close brain regions. This is due to the fact that the fMRI scanner is introducing bias in the neighborhood parts of the brain.

For alive rats, the proportion of rejected null hypotheses is always higher than 5%. Note that there is a high variability with respect to rats. The highest number of detections is obtained for the rats 5 and 8 with more than 33% of edges. For the others, the number of detections is between 15% and 25%.

We applied a bootstrap procedure on time series using moving windows in time (with overlapping). Results are given in Figure 9. We observe an important difference between alive and dead rats. Dead rats have a mean proportion of rejected null hypotheses equal to zero or really near to zero. The highest mean of rejected

Figure 6: Proportion of true edges detected with respect to the scale. 1000 simulations were done with FIVAR models, with (a) the VAR part is null (b) all the components are concerned with a non null VAR part (c) 10 of the 51 components are concerned with a non null VAR part.

(a) (b) (c)

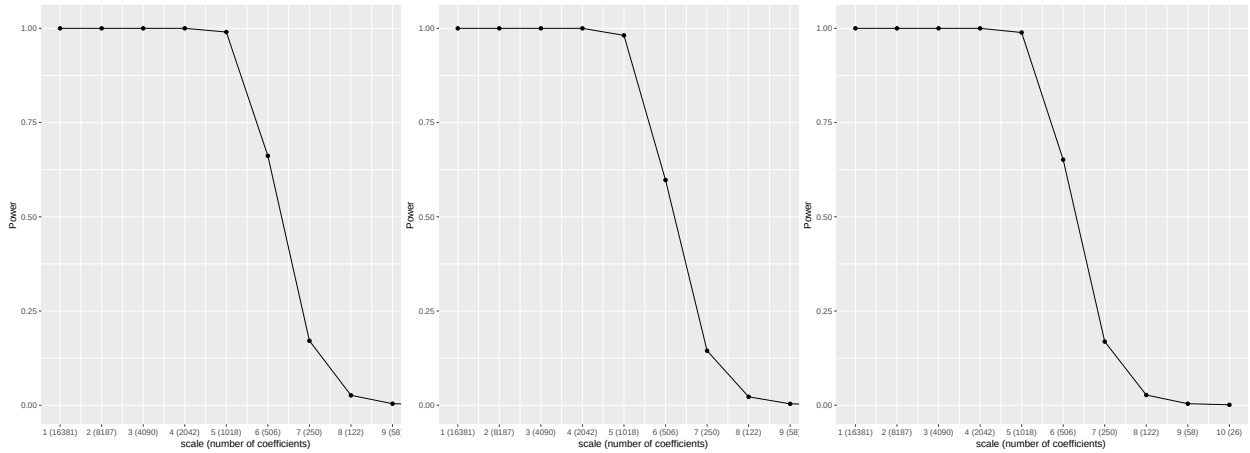
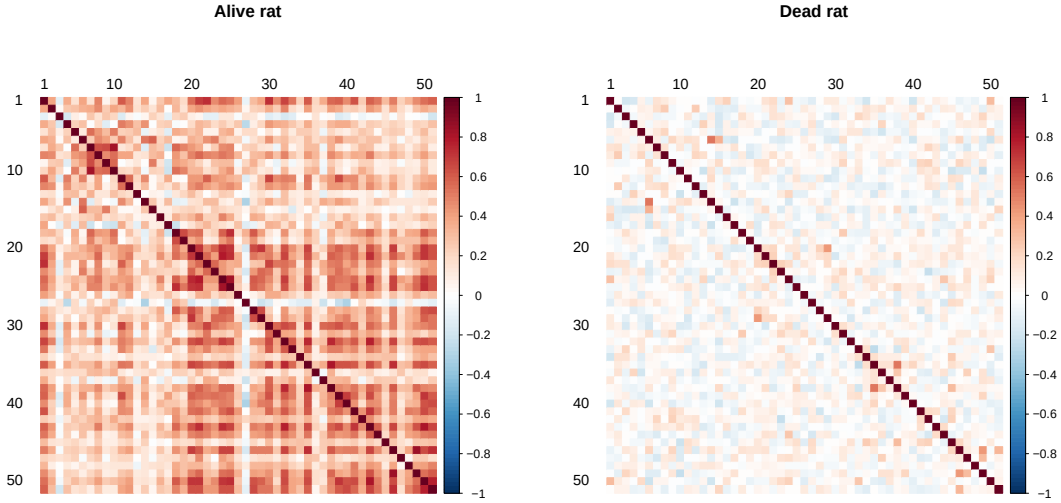


Figure 7: Examples of two correlation matrices between wavelet coefficients obtained at scale 4 for an alive rat (left) and a dead rat (right).



Dead rat	1	2	3	4
Number of edges	4	1	0	2

Alive rat	5	6	7	8	9	10	11
Number of edges	433	230	196	567	366	387	296

Table 1: Number of rejected null hypotheses (or estimate edges) obtained by Bonferroni's procedure on Fisher statistics, evaluated at scale 4.

Figure 8: Illustration of the necessity to make mutiple correction for a given dead rat.

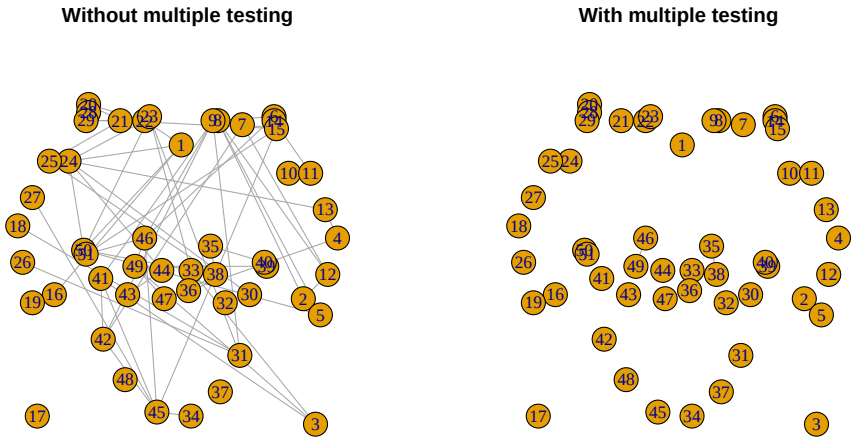
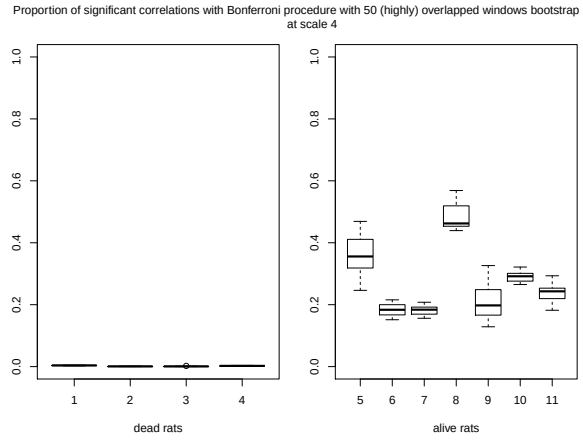


Figure 9: Proportion of significant edges with Bonferroni procedure with overlapped time windows bootstrap, for dead rats (left) and alive rats (right).



hypotheses is indeed 4 for 1275 tests. On the contrary for alive rats we observe always more than 10% of null hypotheses rejected.

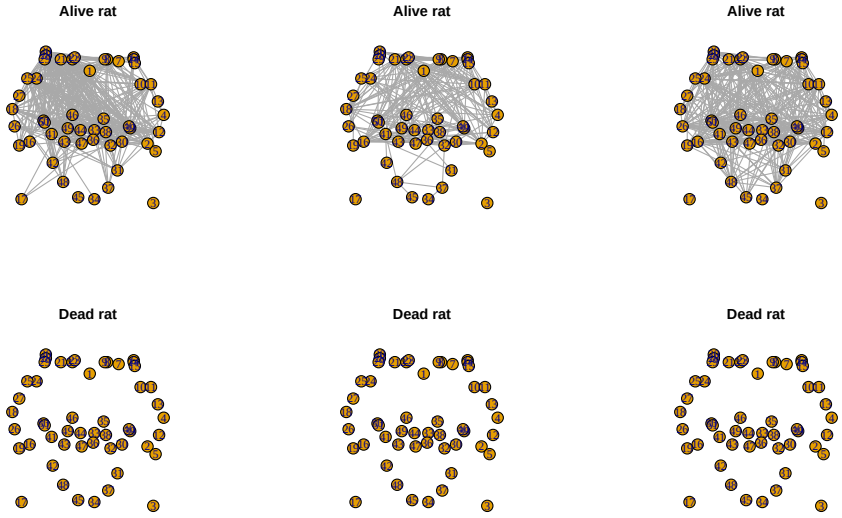
Some examples of estimated graphs for dead rats and alive rats are given in Figure 10.

6. CONCLUSION

A whole range of graphical models have been proposed for graph inference, using either correlation or partial correlation as a measure of dependence. Most of these methods relies on assumptions on the structure of the estimated graphs: sparsity for gLasso, conditions on the degree distribution, independence between blocks for stochastic block model. In this paper we propose to use a multiple testing strategy to avoid making such assumptions. Our approach is valid and control the FWER for any graphs' structures. However, quantifying the power of the approach is more complicated and will be the objective of future works.

We considered wavelet-based graph inference by multiple testing. We highlighted some specificity of this approach. The number of available coefficients at each scale is the first limitation due to the asymptotic testing.

Figure 10: Examples of graphs obtained at scale 4 for some alive rats and some dead rats. Bonferroni's correction was used with Fisher statistics.



Next the memory properties influences the estimated graph. Long-range dependence implies a specific discrepancy (with respect to graph G^*) which can be corrected if needed. Short-range memory has for consequence that the inferred graph is dependent to the high frequency scales. A perspective is to develop a multiple testing approach to infer the graph G^* with a correction of all memory effects.

ACKNOWLEDGMENTS

The authors would like to thank Etienne Roquain for very interesting scientific discussions. We also would like to thank Emmanuel Barbier and Guillaume Becq for providing us the data of the resting state fMRI on the rats.

REFERENCES

[1] Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., and Bullmore, E., "Functional connectivity and brain networks in schizophrenia.," *J Neurosci* **30**, 9477–9487 (Jul 2010).

[2] Kahan, J., Urner, M., Moran, R., Flandin, G., Marreiros, A., Mancini, L., White, M., Thornton, J., Yousry, T., Zrinzo, L., Hariz, M., Limousin, P., Friston, K., and Foltynie, T., "Resting state functional mri in parkinson's disease: the impact of deep brain stimulation on 'effective' connectivity," *Brain* **137**(4), 1130–1144 (2014).

[3] Termenon, M., Achard, S., Jaillard, A., and Delon-Martin, C., "The "hub disruption index", a reliable index sensitive to the brain networks reorganization. a study of the contralesional hemisphere in stroke," *Frontiers in Computational Neuroscience* **10** (2016).

[4] Silva, S., De Pasquale, F., Vuillaume, C., Riu, B., Loubinoux, I., Geeraerts, T., Seguin, T., Bounes, V., Fourcade, O., Demonet, J.-F., and Péran, P., "Disruption of posteromedial large-scale neural communication predicts recovery from coma," *Neurology* **85**(23), 2036–2044 (2015).

[5] Richiardi, J., Achard, S., Bunke, H., and Van De Ville, D., "Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience," *IEEE Signal Process. Mag.* **30**(3), 58–70 (2013).

[6] Margulies, D. S., Böttger, J., Long, X., Lv, Y., Kelly, C., Schäfer, A., Goldhahn, D., Abbushi, A., Milham, M. P., Lohmann, G., and Villringer, A., "Resting developments: a review of fmri post-processing methodologies for spontaneous brain activity," *Magnetic Resonance Materials in Physics, Biology and Medicine* **23**(5-6), 289–307 (2010).

[7] Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E., "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *The Journal of Neuroscience* **26**, 63–72 (Jan. 2006). 00831.

[8] Fallani, F. D. V., Richiardi, J., Chavez, M., and Achard, S., "Graph analysis of functional brain networks: practical issues in translational neuroscience," *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1653), 20130521 (2014).

[9] Patel, A. X. and Bullmore, E. T., "A wavelet-based estimator of the degrees of freedom in denoised fMRI time series for probabilistic testing of functional connectivity and brain graphs," *NeuroImage* (2015).

[10] Malagurski, B., Péran, P., Sarton, B., Vinour, H., Naboulsi, E., Riu, B., Bounes, F., Seguin, T., Lotterie, J. A., Fourcade, O., Minville, V., Ferr, F., Achard, S., and Silva, S., "Topological disintegration of resting state functional connectomes in coma," *NeuroImage* **195**, 354–361 (2019).

[11] Whitcher, B., Guttorp, P., and Percival, D. B., "Wavelet analysis of covariance with application to atmospheric time series," *Journal of Geophysical Research* **105(D11)**(14), 941–962 (2000).

[12] Achard, S. and Gannaz, I., "Multivariate wavelet whittle estimation in long-range dependence," *Journal of Time Series Analysis* **37**(4), 476–512 (2016). arXiv preprint arXiv:1412.0391.

[13] Meyer, F. G., "Wavelet-based estimation of a semiparametric generalized linear model of fmri time-series," *IEEE transactions on medical imaging* **22**(3), 315–322 (2003).

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

[15] Moulines, E., Roueff, F., and Taqqu, M. S., "A wavelet Whittle estimator of the memory parameter of a nonstationary gaussian time series," *The Annals of Statistics* , 1925–1956 (2008).

[16] Veitch, D. and Abry, P., "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Transactions on Information Theory* **45**(3), 878–897 (1999).

[17] Chambers, M., "The simulation of random vector time series with given spectrum," *Mathematical and Computer Modelling* **22**(2), 1–6 (1995).

[18] Moulines, E., Roueff, F., and Taqqu, M. S., "On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter," *Journal of Time Series Analysis* **28**(2), 155–187 (2007).

[19] Achard, S., Bassett, D. S., Meyer-Lindenberg, A., and Bullmore, E., "Fractal connectivity of long-memory networks," *Physical Review E* **77**(3), 036104 (2008).

[20] Percival, D. B., "On estimation of the wavelet variance," *Biometrika* **82**(3), 619–631 (1995).

[21] Whitcher, B., *Assessing nonstationary time series using wavelets*, PhD thesis, Univ. of Wash. Press, Seattle (1998).

[22] Achard, S. and Gannaz, I., "Multivariate wavelet whittle estimation in long-range dependence," *Journal of Time Series Analysis* **37**(4), 476–512 (2016).

[23] Anderson, T., [*An Introduction to Multivariate Statistical Analysis*], John Wiley & Sons, third edition ed. (2003).

[24] Romano, J. P., Shaikh, A. M., and Wolf, M., "Control of the false discovery rate under dependence using the bootstrap and subsampling," *Test* **17**(3), 417 (2008).

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

[26] Rubinov, M. and Sporns, O., "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage* **52**, 1059–1069 (Sept. 2010).

[27] Bullmore, E. and Sporns, O., "Complex brain networks: graph theoretical analysis of structural and functional systems.," *Nat Rev Neurosci* **10**, 186–198 (Mar 2009).

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

[29] Achard, S. and Gannaz, I., "Wavelet-based and fourier-based multivariate whittle estimation: multiwave," *Journal of Statistical Software* **89** (2019).

[30] Hero, A. and Rajaratnama, B., "Large-scale correlation screening," *Journal of the American Statistical Association* **106**(496), 1540–1552 (2011).

[31] Sela, R. J. and Hurvich, C. M., "Computationaly efficient methods for two multivariate fractionnaly integrated models," *Journal of Time Series Analysis* **30**(6), 631–651 (2009).

[32] Pawela, C. P., Biswal, B. B., Cho, Y. R., Kao, D. S., Li, R., Jones, S. R., Schulte, M. L., Matloub, H. S., Hudetz, A. G., and Hyde, J. S., "Resting-state functional connectivity of the rat brain," *Magnetic Resonance in Medicine* **59**(5), 1021–1029 (2008).