Robust correlation for aggregated data with spatial characteristics

Sophie Achard[a], Jean-François Coeurjolly[a],
Pierre Lafaye de Micheaux[b] and Jonas Richiardi[c]

[a] Univ. Grenoble Alpes, CNRS, Inria, Grenoble-INP, LJK, 38000 Grenoble, France;

[b] School of Mathematics and Statistics, UNSW Sydney, NSW, Australia;

[c] Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

2020-11-18

Abstract

The objective of this paper is to study the robustness of computation of correlations under spatial constraints. The motivation of our paper is the specific case of functional magnetic resonance (fMRI) brain imaging data, where voxels are aggregated to compute correlations. In this paper we show that the way the spatial components are aggregating to compute correlation may have a strong influence on the resulting estimations. We then propose various estimators which take into account this spatial structure.

keywords correlation; aggregated data; familial correlations; serial correlations

1 Introduction

The use of aggregated data is particularly common in various field of sciences. In social sciences, both individual and organizational data are collected (Ostroff 1993). The objective is to evaluate the relationships among variables at different level of analysis. For example, in (Ostroff 1993), the author took the example of studying correlations between satisfaction and technology at two levels, individual scores, denoted individual correlation, and when individuals are grouped into organizations, also called organizational correlation. Depending on three main factors mainly the measurement errors, the variance within group and the variance-within ratios, the ratio of organizational correlation and individual correlation was shown to take values between -1 and 2. This means that it is crucial to identify the way the data are generated.

In the studies of familial data (Rosner et al. 1977), specific characteristics are obtained for different families with different sizes. The quantification of family resemblance is studied in this context, with the estimation of sibling correlations and/or parent-offspring correlations, see for a review (Donner and Eliasziw 1991). Each data is composed with characteristics extract for families with a mother and/or father and the children. The difficulties arise because of the dependence between the children and the different number of children per family. It was noted that *correlation between the average of time series* and the *average of correlations between all possible pairs of time series* are not equal in the majority of cases.

In psychology, (Vul et al. 2009) shows that the strength of relationship depends on the reliability of the measurements. Finally, the problem of correlation computation is also a common problem in ecology, geography, climate studies .... The data collected in this field are attached to a spatial position and usually with spatial correlation. This problem was first reported by Student (1914), and studied in (Clifford et al. 1989) for two spatial processes. Applications of these methods can be found for example in the study of meteorological data (Gunst 1995), in ecological data (Liebhold and Sharov 1998), in fMRI data (Ye et al. 2011). The main difficulty in these analyses is to take into account the spatial correlations in order to construct estimator of correlations and testing procedures when the averaged variables are not independent.

We are facing analogous difficulties in brain imaging, in particular when studying so-called functional connectivity. Functional connectivity of the brain is estimated from observations using non invasive techniques such as electroencephalography (EEG), magnetoencephalography (MEG) or fMRI. Each recording provides time series associated to spatial locations within regions of the brain. Functional connectomes, that is, graphs representing the estimated connectivity, are then constructed by computing dependence between the time series. It has already been shown that computation of connectomes is affected by three main parameters: the length of the acquisition, the number of regions and the chosen frequency band. In addition, the number of subjects available in the sample will also play a role in terms of group comparisons (Termenon et al. 2016). Using fMRI data, we record thousand of voxels. Each region of the brain is then associated to a given set of voxels. The idea is then to extract a representative of the set of voxels to attach one time series to each region. The most common approach is to take the average of the voxels. In this paper, we question this choice by studying statistically the computation of correlation of this average of voxels.

The paper is organized as follows. In a first part, we are motivating the definition of a new spatial model of fMRI. Then, computations of correlations are described. Based on simulations, we illustrate the good behavior of the newly introduced estimators.

2 Definition of the proposed spatial model for fMRI data

Let $\mathcal{C}$ denote a finite compact subset of indices of $\mathbb{Z}^d$. This subset $\mathcal{C}$ will play the role of the set of indices of voxels of the brain when $d = 3$. The J regions of interest of the brain are represented through their set of voxels denoted by $\mathcal{R}_j$ for $j = 1, \dots, N$. The number of voxels in each region is denoted by $\#\mathcal{R}_j = N_j$. So,

$$\mathcal{C} = \cup_{j=1}^J \mathcal{R}_j \text{ and } \#\mathcal{C} = \sum_{j=1}^J N_j.$$

For any $i \in \mathcal{C}$, we assume observing the signal $Y_i(\cdot)$ sampled at times $t = 1, \dots, T$ which can be decomposed as follows

$$Y_i(t) = X_i(t) + \varepsilon_i(t) + e(t), \tag{1}$$

where $X_i(\cdot)$ represents the signal of interest, $\varepsilon_i(\cdot)$ represents a noise contaminating locally the voxel i and the signal $e(\cdot)$ is a noise corrupting in the same way all the voxels $i \in \mathcal{C}$. We now make a few assumptions on these different components. First, we assume that all the random variables are centered, the signals $X_i(\cdot)$, $\varepsilon_i(\cdot)$ and $e(\cdot)$ are mutually independent and independent in time and that the global noise is homoscedastic. This implies that for any $i, i' \in \mathcal{C}$ and $s, t = 1, \dots, T$ $(s \neq t)$

$$\begin{aligned} E[X_i(t)] &= E[\varepsilon_i(t)] = E[e(t)] = 0 \\ E[X_i(s)X_i(t)] &= E[\varepsilon_i(s)\varepsilon_i(t)] = E[e(s)e(t)] = 0 \\ E[X_i(s)\varepsilon_{i'}(t)] &= E[X_i(s)e(t)] = E[\varepsilon_i(s)e(t)] = 0 \\ E[e(t)^2] &= \sigma_e^2, \end{aligned}$$

Let us now describe the spatial nature of the signal and the local noise. For any $i, i' \in \mathcal{C}$ $j, j' = 1, \dots, J$ $(j \neq j')$ and for all $t = 1, \dots, T$, we assume that there exists $\sigma_j > 0$, $\sigma_e \geq 0$, $r_{jj'} \in [-1, 1]$, $\rho_{ii'} \in (0, 1]$, $\eta_{ii'} \in [-1, 1]$ such that

$$E[X_i(t)X_{i'}(t)] = \begin{cases} \sigma_j \sigma_{j'} r_{jj'} & \text{if } i \in \mathcal{R}_j, i' \in \mathcal{R}_{j'}, j \neq j' \\ \sigma_j^2 \rho_{ii'} & \text{if } i, i' \in \mathcal{R}_j \end{cases}$$

and

$$E[\varepsilon_i(t)\varepsilon_{i'}(t)] = \sigma_\varepsilon^2 \eta_{ii'}.$$

The parameter $r_{jj'}$ represents the correlation between two signals of two different regions $\mathcal{R}_j$ and $\mathcal{R}_{j'}$ and is called inter-correlation between regions $\mathcal{R}_j$ and $\mathcal{R}_{j'}$ in the following. The parameter $\rho_{ii'}$ (resp. $\eta_{ii'}$) represents the intra-correlation between two signals (resp. the spatial correlation between two local noises) inside a common region. We assume that inside each region, the signals of interest have positive intra-correlation and that for each time t and for $j = 1, \dots, J$, $(X_i(t), i \in \mathcal{R}_j)$ (resp. $(\varepsilon_i(t), i \in \mathcal{C})$) is a stationary random field observed in $\mathcal{R}_j$ (resp. $\mathcal{C}$). We furthermore assume that both the correlations $\rho_{ii'}$ (for any $i, i' \in \mathcal{R}_j$ for some j) and $\eta_{ii'}$ (for $i, i' \in \mathcal{C}$) depend only on the (uniform) distance between the two voxels i and i'. For brevity, we still denote $\rho_{|i'-i|}$ by $\rho_{ii'}$ and $\eta_{|i'-i|}$ by $\eta_{ii'}$ where for $x \in \mathbb{Z}^d$, the notation $|x|$ stands for the uniform norm. Our a priori is that the intra-correlation ρ_d is close to 1 for moderate distances d, meaning that close neighbors are strongly (positively) correlated. For the local noise, we assume that the spatial correlation function is such that $\eta_0 = 1$ and $\eta_d = 0$ for $d \geq p$. When $p = 1$, this of course means that for any $i, i' \in \mathcal{C}, i \neq i'$, $\varepsilon_i(t)$ and $\varepsilon_{i'}(t)$ are uncorrelated.

Remarque 2.1. *We assume that the random variables are independent in time. This is not a too restrictive hypothesis, in particular, if the random variables have long memory hypothesis, after a wavelet decomposition, the random variables can be approximated to be decorrelated in time for large wavelet scales. In addition, assuming that the X_i's are centered is coherent as it is a well-known fact that a wavelet decomposition based on a wavelet mother with K vanishing moments cancels out every polynomial trend with degree $K - 1$.*

2.1 Objectives and issues

Given a parcellation of the brain, the objective is to estimate inter-correlations for each pair of parcel. The estimator has to be non parametric, robust and fast. In particular we do not want to focus on the intra-correlations, nor the estimations of variances and correlations of the additional noises.

3 Inter-correlation estimators

3.1 Notation

Let $\mathbf{Y}_1 = (Y_1(1), \ldots, Y_1(T))$ and $\mathbf{Y}_2 = (Y_2(1), \ldots, Y_2(T))$ denote two samples of length T and let $\widehat{\mathrm{cov}}(\mathbf{Y}_1, \mathbf{Y}_2)$, $\widehat{\mathrm{cor}}(\mathbf{Y}_1, \mathbf{Y}_2)$ and $\widehat{\sigma}(\mathbf{Y}_1)$ denote respectively the sample covariance between $\mathbf{Y}_1$ and $\mathbf{Y}_2$, the sample correlation between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ and the standard deviation of $\mathbf{Y}_1$.

In this section, we recall some known estimates of the inter-correlation parameter $r_{jj'}$ (that we simply denote by r for brevity) for two regions of interest $\mathcal{R}_j$ and $\mathcal{R}_{j'}$. We also introduce some new ones which can handle the problems occurring when dealing with spatio-temporal data such as fMRI data.

To understand the limit of each estimator, we need some additional notation. For any $j = 1, \ldots, J$, we first define a ν-neighborhood as a subset of $n_\nu := (2\nu + 1)^d$ indices for which all indices are at distance $\leq \nu$ from the center.

Then, for any ν-neighborhood $\mathcal{V}$ ($\nu \geq 1$) we let

$$\bar{\rho}_\nu = \frac{1}{n_\nu^2} \sum_{i,i' \in \mathcal{V}} \rho_{|i'-i|} \quad \text{and} \quad \bar{\eta}_\nu = \frac{1}{n_\nu^2} \sum_{i,i' \in \mathcal{V}} \eta_{|i'-i|} = \frac{1}{n_\nu^2} \sum_{\substack{i,i' \in \mathcal{V} \\ |i'-i| < p}} \eta_{|i'-i|}.$$

We also need the related notation where $\mathcal{V}$ corresponds to the entire region of interest $\mathcal{R}_j$

$$\bar{\rho}^{(j)} = \frac{1}{N_j^2} \sum_{i,i' \in \mathcal{R}_j} \rho_{|i'-i|} \quad \text{and} \quad \bar{\eta}^{(j)} = \frac{1}{N_j^2} \sum_{\substack{i,i' \in \mathcal{R}_j \\ |i'-i|<p}} \eta_{|i'-i|}$$

If ε_i and $\varepsilon_{i'}$ are spatially uncorrelated for any $i, i' \in \mathcal{C}$ $(i \neq i')$, then $\bar{\eta}_\nu = 1/n_\nu$ and $\bar{\eta}^{(j)} = 1/N_j$. Otherwise since we assume that $\eta_d = 0$ for $d \geq p$, $\bar{\eta}_\nu = \mathcal{O}(1/n_\nu)$ and $\bar{\eta}^{(j)} = \mathcal{O}(1/N_j)$. For moderate ν, it is expected that the intra-correlations satisfy $\bar{\rho}_\nu$ is close to 1. However, $\bar{\rho}^{(j)}$ can be significantly far from 1, especially for large regions of interest. Finally for more complex estimators developed hereafter, we need the additional notation

$$\bar{\rho}_{\nu,\delta} = \frac{1}{n_\nu^2} \sum_{i \in \mathcal{V}, i' \in \mathcal{V}'} \rho_{|i'-i|}$$

for two ν-neighborhoods $\mathcal{V}, \mathcal{V}'$ (from the same region of interest) distant from δ (i.e. $d(\mathcal{V}, \mathcal{V}') = \delta$). Again, if ν and δ are not too large, we can expect that $\bar{\rho}_{\nu,\delta}$ is close to 1.

From Section 2, we may derive the following calculations. Let $j, j' \in \{1, \dots, J\}$ two different indices. Let $i \in \mathcal{R}_j, i' \in \mathcal{R}_j \cup \mathcal{R}_{j'}$.

$$\operatorname{cov}(Y_i(t), Y_{i'}(t)) = \begin{cases} \sigma_j \sigma_{j'} r_{jj'} + \sigma_\varepsilon^2 \eta_{|i'-i|} + \sigma_e^2 & \text{if } i \in \mathcal{R}_j \text{ and } i' \in \mathcal{R}_{j'} \\ \sigma_j^2 \rho_{|i'-i|} + \sigma_\varepsilon^2 \eta_{|i'-i|} + \sigma_e^2 & \text{if } i, i' \in \mathcal{R}_j. \end{cases} \tag{2}$$

If $\bar{Y}_j(t)$ denotes the spatial average signal at time t over the region $\mathcal{R}_j$, i.e. $\bar{Y}_j(t) = N_j^{-1} \sum_{i \in \mathcal{R}_j} Y_i(t)$, then for $t = 1, \dots, T$

$$\operatorname{Var}(\bar{Y}_j(t)) = \sigma_j^2 \bar{\rho}_j + \sigma_\varepsilon^2 \bar{\eta}_j + \sigma_e^2. \tag{3}$$

Since, we assume the independence in time of the signals, the local noise and the global noise most of the results presented in the next sections are straightforward. Without loss of generality, we intrinsically assume that $d(\mathcal{R}_j, \mathcal{R}_{j'}) \geq p$ which ensures that for any $i \in \mathcal{R}_j$ and $i' \in \mathcal{R}_{j'}$, $\varepsilon_i(t)$ and $\varepsilon_{i'}(t)$ are uncorrelated. This slightly simplifies the numerator of r^{CA} and r^{AC} defined below by (5) and (7).

3.2 Classical estimator : correlation of averages (method CA)

In order to increase the signal-to-noise ratio, classical methods in fMRI or EEG use to average or convolve with Gaussian kernel the signal in space. The aggregated correlation estimator corresponds to the classic estimator considered for example in Achard et al. (2006) by taking the correlation between the aggregated variables corresponding to a pair of given groups:

$$\widehat{r}^{\text{CA}} = \frac{\widehat{\operatorname{cov}}(\bar{\mathbf{Y}}_j, \bar{\mathbf{Y}}_{j'})}{\widehat{\sigma}(\bar{\mathbf{Y}}_j)\widehat{\sigma}(\bar{\mathbf{Y}}_{j'})} \tag{4}$$

Using (2) and (3), it can be shown that $\widehat{r}^{\,\text{CA}}$ is a strongly consistent estimator of $r^{\,\text{CA}}$ as $T \to \infty$ where

$$r^{\,\text{CA}} = \frac{\sigma_j \sigma_{j'} r + \sigma_e^2}{(\sigma_j^2 \bar{\rho}^{(j)} + \sigma_\varepsilon^2 \bar{\eta}^{(j)} + \sigma_e^2)^{1/2} (\sigma_{j'}^2 \bar{\rho}^{(j')} + \sigma_\varepsilon^2 \bar{\eta}^{(j')} + \sigma_e^2)^{1/2}}. \tag{5}$$

Even in absence of noise ($\sigma_\varepsilon = \sigma_e = 0$), $\widehat{r}^{\,\text{CA}}$ has the drawback to bias the estimation because the average of correlations is not equal to the correlation of the averages. In Achard et al. (2011), we demonstrated that, even without noise, the correlation of averages is multiplied by the number of points taken in the average. So when the group of voxels in the brain have different sizes, the largest ones have higher correlation simply because of this bias mainly due to the fact that for large regions $\bar{\rho}^{(j)}$ can be far from 1.

3.3 Average of correlations (method AC)

This estimator defined by

$$\widehat{r}^{\,\text{AC}} = \frac{1}{N_j N_{j'}} \sum_{i \in \mathcal{R}_j, i' \in \mathcal{R}_{j'}} \widehat{\text{cor}}(\mathbf{Y}_i, \mathbf{Y}_{i'}). \tag{6}$$

clearly tends (almost surely) as $T \to \infty$ toward $r^{\,\text{AC}}$ given by

$$r^{\,\text{AC}} = \frac{\sigma_j \sigma_{j'} r + \sigma_e^2}{(\sigma_j^2 + \sigma_\varepsilon^2 + \sigma_e^2)^{1/2} (\sigma_{j'}^2 + \sigma_\varepsilon^2 + \sigma_e^2)^{1/2}}. \tag{7}$$

Another way to correct the size effect is to compensate the inter-correlation by the intra-correlation. This leads to the following estimator:

$$\widehat{r}^{\,\widetilde{\text{AC}}} = \frac{1}{N_j N_{j'}} \left(\sum_{i,i' \in \mathcal{R}_j} \widehat{\text{cor}}(\mathbf{Y}_i, \mathbf{Y}_{i'}) \sum_{i,i' \in \mathcal{R}_{j'}} \widehat{\text{cor}}(\mathbf{Y}_i, \mathbf{Y}_{i'}) \right)^{1/2} \widehat{r}^{\,\text{AC}}. \tag{8}$$

The two estimators (6) and (8) have the important property to remove the size effect (since when $\sigma_\varepsilon = \sigma_e = 0$, $r^{\,\text{AC}} = r$). As it can be straightforwardly shown that these two estimators tend to the same limit, we only focus in Sections 4, we skip the estimator $\widehat{r}^{\,\widetilde{\text{AC}}}$.

We can underline that both estimators are very sensitive to (local and global) noise. Indeed, with the presence of local noise, the variance of noise appears directly in the denominator which is decreasing the values of correlation estimations. In the sequel of this section, we introduce first estimators to be robust to the local noise using either local average or replicated data. Then, we introduce estimators to compensate the global noise by using 2 other decorrelated groups of variables, and then we will combine the two to define estimators robust to both local and global noise.

3.4 Local correlation of averages (method ℓCA)

The following estimator is defined by averaging correlations between average of variables belonging to two different regions on small neighborhood inside the regions. Its aim is to reduce the local noise. As the choice of neighborhood to average and the choice of pairs of neighborhood are arbitrary, we propose to repeat draws of small neighborhood and to take the average of all computed correlations. For $b = 1, \dots, B$, let $\mathcal{V}_j^{(b)}$ (resp. $\mathcal{V}_{j'}^{(b)}$) be a ν-neighborhood of $\mathcal{R}_j$ (resp. $\mathcal{R}_{j'}$), then we define

$$\widehat{r}^{\,\ell\mathrm{CA}} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_j^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'}^{(b)}}). \tag{9}$$

Since $d(\mathcal{R}_j, r_{j'}) \geq p$ then $d(\mathcal{V}_j^{(b)}, \mathcal{V}_{j'}^{(b)}) \geq p$ for any b and $\widehat{r}^{\,\ell\mathrm{CA}}$ tends (almost surely) as $T \to \infty$ towards $r^{\,\ell\mathrm{CA}}$

$$r^{\,\ell\mathrm{CA}} = \frac{\sigma_j \sigma_{j'} r + \sigma_e^2}{(\sigma_j^2 \bar{\rho}_\nu + \sigma_e^2 \bar{\eta}_\nu + \sigma_e^2)^{1/2} (\sigma_{j'}^2 \bar{\rho}_\nu + \sigma_e^2 \bar{\eta}_\nu + \sigma_e^2)^{1/2}}. \tag{10}$$

When there is no global noise ($\sigma_e = 0$), we observe that $\widehat{r}^{\,\ell\mathrm{CA}}$ is more robust to the local noise since $\bar{\eta}_\nu = \mathcal{O}(n_\nu^{-1})$. However, it is important to choose moderate ν otherwise $\bar{\rho}_\nu$ can deviate significantly from 1.

3.5 Replicates for correlations (method R)

In order to suppress the effect of local noise, we introduce a new estimator computed using replicates within the same region and the estimator $\widehat{r}^{\,\mathrm{AC}}$ is corrected by the correlation between the chosen replicates. This estimator, denoted by $\widehat{r}^{\,\mathrm{R}}$ (R for replicates), was first introduced by Bergholm et al. (2010), in the context of image analysis. It is defined by

$$\widehat{r}^{\,\mathrm{R}} = \frac{1}{B}\sum_{b=1}^{B} \frac{\frac{1}{4}\sum_{\alpha,\beta=1}^{2} \widehat{\mathrm{cor}}(\mathbf{Y}_{i_\alpha^{(b)}}, \mathbf{Y}_{i'^{(b)}_\beta})}{\sqrt{\widehat{\mathrm{cor}}(\mathbf{Y}_{i_1^{(b)}}, \mathbf{Y}_{i_2^{(b)}})\ \widehat{\mathrm{cor}}(\mathbf{Y}_{i'^{(b)}_1}, \mathbf{Y}_{i'^{(b)}_2})}} \tag{11}$$

where for $b = 1, \dots, B$, $i_1^{(b)}$ and $i_2^{(b)}$ belong to $\mathcal{R}_j$, and are expected to verify $\rho_{i_1^{(b)} i_2^{(b)}} \approx 1$. Respectively, $i'^{(b)}_1$ and $i'^{(b)}_2$ belong to $\mathcal{R}_{j'}$ and are expected to verify $\rho_{i'^{(b)}_1 i'^{(b)}_2} \approx 1$.

Let us detail the convergence of $\widehat{r}^{\,\mathrm{R}}$. We assume that any $b = 1, \dots, B$, $|i_2^{(b)} - i_1^{(b)}| = |i'^{(b)}_2 - i'^{(b)}_1| = \delta$ for some fixed integer $\delta \geq p$. Since

$$\frac{1}{4}\sum_{\alpha,\beta=1}^{2} \widehat{\mathrm{cor}}(\mathbf{Y}_{i_\alpha^{(b)}}, \mathbf{Y}_{i'^{(b)}_\beta}) \xrightarrow{a.s.} \frac{\sigma_j \sigma_{j'} r_{jj'} + \sigma_e^2}{\left(\sigma_j^2 + \sigma_\varepsilon^2 + \sigma_e^2\right)^{1/2} \left(\sigma_{j'}^2 + \sigma_\varepsilon^2 + \sigma_e^2\right)^{1/2}}$$

and

$$\widehat{\mathrm{cor}}(\mathbf{Y}_{i_1^{(b)}}, \mathbf{Y}_{i_2^{(b)}}) \overset{a.s.}{\to} \frac{\sigma_j^2 \rho_\delta + \sigma_\varepsilon^2 \eta_\delta + \sigma_e^2}{\sigma_j^2 + \sigma_\varepsilon^2 + \sigma_e^2},$$

and since $\eta_\delta = 0$, we obtain that $\widehat{r}^{\,\mathrm{R}}$ converges almost surely towards $r^{\,\mathrm{R}}$ as $T \to \infty$ where

$$r^{\,\mathrm{R}} = \frac{\sigma_j \sigma_{j'} r + \sigma_e^2}{(\sigma_j^2 \rho_\delta + \sigma_e^2)^{1/2} (\sigma_{j'}^2 \rho_\delta + \sigma_e^2)^{1/2}} \tag{12}$$

When $\sigma_e = 0$ and then choosing $\delta = p$ (so 1 if the local noise is spatially independent) leads to $r^{\,\mathrm{R}} = \frac{r}{\rho_p}$ and we may hope that in this situation, ρ_p is very close to 1. In other words, $\widehat{r}^{\,\mathrm{R}}$ is an estimator robust to the size of the regions of interest and robust to a local noise.

In the case of low signal-to-noise ratio, the quantity in the denominator of (11) may be negative and the square root becomes undefined. We propose then to combine the local averaging on small neighborhood and the introduction of replicates to compensate for local noise corruption of signals.

3.6 Local average of replicates (method ℓR)

This estimator consists in combining the idea of replicates with the one consisting in averaging locally the signals.

$$\widehat{r}^{\,\ell\mathrm{R}} = \frac{1}{B} \sum_{b=1}^{B} \frac{\frac{1}{4}\sum_{\alpha,\beta=1}^{2} \widehat{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j_\alpha}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'_\beta}^{(b)}})}{\sqrt{\widehat{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j_1}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j_2}^{(b)}})\, \widehat{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j'_1}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'_2}^{(b)}})}}. \tag{13}$$

Using arguments developed in Sections 3.4 and 3.5 then if we assume that the distance between ν-neighborhoods involved in (13) is at least p, then it can be shown $\widehat{r}^{\,\ell\mathrm{R}}$ is a strongly consistent estimator of $r^{\,\ell\mathrm{R}}$ defined by

$$r^{\,\ell\mathrm{R}} = \frac{\sigma_j \sigma_{j'} r + \sigma_e^2}{(\sigma_j^2 \bar{\rho}_{\nu,\delta} + \sigma_e^2)^{1/2} (\sigma_{j'}^2 \bar{\rho}_{\nu,\delta} + \sigma_e^2)^{1/2}} \tag{14}$$

When $\sigma_e = 0$, then letting $\delta = p$ reduces $r^{\,\ell\mathrm{R}}$ to $\frac{r}{\bar{\rho}_{\nu,p}}$.

3.7 Use of a priori disconnected regions (method D based on differences)

We now present an estimator which handles the problem of global noise, based on the selection of two extra regions that are uncorrelated. This assumption is realistic in the context of fMRI data where we are interested in the correlations between cortical regions. Indeed, the volume recorded by the scanning may be bigger, and the definition of extra regions is possible. In this case, if we want to estimate the correlation between two regions, say $\mathcal{R}_j$ and $\mathcal{R}_{j'}$, the idea is to assume the extra information that neither $\mathcal{R}_j$ nor $\mathcal{R}_{j'}$ is connected to some regions $\mathcal{R}_k$ and $\mathcal{R}_{k'}$ where $k, k' \in \{1, \dots, J\} \setminus \{j, j'\}$ with $k \neq k'$ and we assume in addition that $\mathcal{R}_k$ and $\mathcal{R}_{k'}$ are also not connected. Then, we propose the following estimator: for $b = 1, \dots, B$ let $i^{(b)}$, ${i'}^{(b)}$, $k^{(b)}$ and ${k'}^{(b)}$ be voxels of $\mathcal{R}_j$, $\mathcal{R}_{j'}$, $\mathcal{R}_k$ and $\mathcal{R}_{k'}$.

$$\widehat{r}^{\,\mathrm{D}} = \frac{1}{B} \sum_{b=1}^{B} \widetilde{\mathrm{cor}}(\mathbf{Y}_{i^{(b)}}, \mathbf{Y}_{i'^{(b)}}; \mathbf{Y}_{k^{(b)}}, \mathbf{Y}_{k'^{(b)}}), \tag{15}$$

where for four vectors $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{Y}_3$ and $\mathbf{Y}_4$ (with same length)

$$\widetilde{\mathrm{cor}}(\mathbf{Y}_1, \mathbf{Y}_2; \mathbf{Y}_3, \mathbf{Y}_4) = \frac{\widehat{\mathrm{cov}}(\mathbf{Y}_1 - \mathbf{Y}_3, \mathbf{Y}_2 - \mathbf{Y}_4)}{\widehat{s}(\mathbf{Y}_1, \mathbf{Y}_3, \mathbf{Y}_4)\ \widehat{s}(\mathbf{Y}_2, \mathbf{Y}_3, \mathbf{Y}_4)} \tag{16}$$

and where for three vectors $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ with same length

$$\widehat{s}^2(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \left(\widehat{\sigma}^2(\mathbf{U} - \mathbf{V}) + \widehat{\sigma}^2(\mathbf{U} - \mathbf{W}) - \widehat{\sigma}^2(\mathbf{V} - \mathbf{W}) \right) /2.$$

The intuition of this estimate is quite simple. Assume that the local noise has variance null. Since the noise $e(\cdot)$ is global, subtracting from $Y_{i^{(b)}}(t)$ the value $Y_{k^{(b)}}(t)$ and from $Y_{i'^{(b)}}(t)$ the value $Y_{k'^{(b)}}(t)$ discards the global noise. And since, the regions $\mathcal{R}_k$ and $\mathcal{R}_{k'}$ are not correlated and not correlated to the other ones. The numerator (for each b) is an estimate of $\sigma_j \sigma_{j'} r$. Then, we just have to divide by estimates of σ_j (and $\sigma_{j'}$). We observe that this cannot be done using simply $\widehat{\sigma}^2(\mathbf{Y}_{i^{(b)}} - \mathbf{Y}_{k^{(b)}})$ which estimates $\sigma_j^2 + \sigma_k^2$. This justifies the introduction of $\widehat{s}^2$.

From a theoretical point of view, we can show that almost surely $\widehat{r}^{\,\mathrm{D}}$ converges almost surely as $T \to \infty$ towards $r^{\,\mathrm{D}}$ given by

$$r^{\,\mathrm{D}} = \frac{\sigma_j \sigma_{j'} r + \tau_{jj'}}{\left(\sigma_j^2 + 2\sigma_\varepsilon^2 + \tau_j\right)^{1/2} \left(\sigma_{j'}^2 + 2\sigma_\varepsilon^2 + \tau_{j'}\right)^{1/2}}$$

where

$$\tau_{jj'} = \sigma_j \sigma_{k'} r_{jk'} + \sigma_{j'} \sigma_k r_{j'k} \quad \text{and} \quad \tau_j = -\sigma_j \sigma_k r_{jk} - \sigma_j \sigma_{k'} r_{jk'} + \sigma_k \sigma_{k'} r_{kk'}.$$

Since we assume a priori that the regions $\mathcal{R}_k$ and $\mathcal{R}_{k'}$ are not connected each other and to the other regions then $\tau_j = \tau_{j'} = \tau_{jj'} = 0$ which reduces the previous expression to

$$r^{\,\mathrm{D}} = \frac{\sigma_j \sigma_{j'} r}{\left(\sigma_j^2 + 2\sigma_\varepsilon^2\right)^{1/2} \left(\sigma_{j'}^2 + 2\sigma_\varepsilon^2\right)^{1/2}} \tag{17}$$

which in case of absence of local noise is nothing else than r.

3.8 Combinations of previous estimators: methods ℓD, RD, ℓRD

The three following estimators extend respectively the methods ℓCA, R, ℓR using the idea developed above to get rid of the global noise. We use the notation presented in the previous sections. Theoretical results are also derived along similar lines.

Local averages and use of a priori disconnected regions: method ℓD

This estimator given by

$$\widehat{r}^{\,\ell\mathrm{D}} = \frac{1}{B}\sum_{b=1}^{B} \widetilde{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_j^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'}^{(b)}}; \bar{\mathbf{Y}}_{\mathcal{V}_k^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{k'}^{(b)}}) \tag{18}$$

is a strongly consistent estimator of $r^{\,\ell\mathrm{D}}$ given by

$$r^{\,\ell\mathrm{D}} = \frac{\sigma_j \sigma_{j'} r}{(\sigma_j^2 + 2\sigma_\varepsilon^2)^{1/2}(\sigma_{j'}^2 + 2\sigma_\varepsilon^2)^{1/2}}. \tag{19}$$

Replicates and use of a priori disconnected regions: method RD

This estimator given by

$$\widehat{r}^{\,\mathrm{RD}} = \frac{1}{B}\sum_{b=1}^{B} \frac{\frac{1}{4}\sum_{\alpha,\beta=1}^{2} \widetilde{\mathrm{cor}}(\mathbf{Y}_{i_\alpha^{(b)}}, \mathbf{Y}_{i'^{(b)}_\beta}; \mathbf{Y}_{i_3^{(b)}}, \mathbf{Y}_{i_4^{(b)}})}{\sqrt{\widetilde{\mathrm{cor}}(\mathbf{Y}_{i_1^{(b)}}, \mathbf{Y}_{i_2^{(b)}}; \mathbf{Y}_{k^{(b)}}, \mathbf{Y}_{k'^{(b)}})\ \widetilde{\mathrm{cor}}(\mathbf{Y}_{i'^{(b)}_1}, \mathbf{Y}_{i'^{(b)}_2}; \mathbf{Y}_{k^{(b)}}, \mathbf{Y}_{k'^{(b)}})}} \tag{20}$$

is a strongly consistent estimator of $r^{\,\mathrm{RD}}$ given by

$$r^{\,\mathrm{RD}} = \frac{r}{\bar{\rho}_\delta}. \tag{21}$$

Replicates, local averages and use of a priori disconnected regions: method ℓRD

This estimator given by

$$\widehat{r}^{\,\ell\mathrm{RD}} = \frac{1}{B}\sum_{b=1}^{B} \frac{\frac{1}{4}\sum_{\alpha,\beta=1}^{2} \widetilde{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j_\alpha}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'_\beta}^{(b)}}; \bar{\mathbf{Y}}_{\mathcal{V}_k^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{k'}^{(b)}})}{\sqrt{\widetilde{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j_1}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j_2}^{(b)}}; \bar{\mathbf{Y}}_{\mathcal{V}_k^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{k'}^{(b)}})\ \widetilde{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_{j'_1}^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'_2}^{(b)}}; \bar{\mathbf{Y}}_{\mathcal{V}_k^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{k'}^{(b)}})}} \tag{22}$$

is a strongly consistent estimator of $r^{\,\ell\mathrm{RD}}$ given by

$$r^{\,\ell\mathrm{RD}} = \frac{r}{\bar{\rho}_{\nu,\delta}}. \tag{23}$$

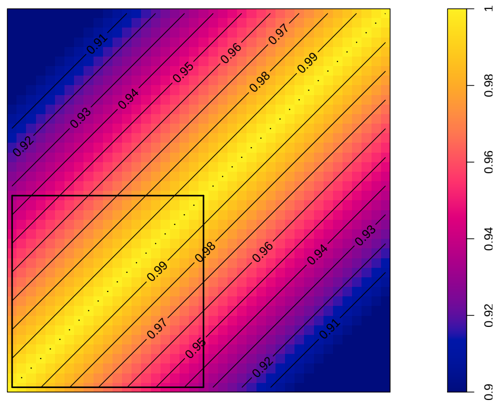
4 Simulations

In this section we investigate the finite sample properties of the estimators of r according to different settings. We focus on the planar case ($d = 2$) and only on two regions of interest, say $\mathcal{R}_j$ and $\mathcal{R}_{j'}$ which contain respectively

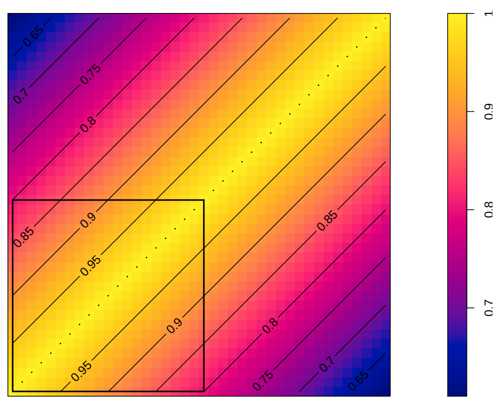
20^2 and 40^2 voxels. To implement the estimators $\widehat{r}^{\bullet}$ for $\bullet = \mathrm{D}, \ell\mathrm{D}, \mathrm{RD}, \ell\mathrm{RD}$ we also assume to observe the signals in two other regions $\mathcal{R}_k$ and $\mathcal{R}_{k'}$ which are disconnected each other and to the two regions of interest. We assume that the signal, the local noise and the global noise follow Gaussian distribution. The variances are set to the following values: $\sigma_j = 1$, $\sigma_{j'} = 2$, $\sigma_k = \sigma_{k'} = 1$. Throughout the simulations the parameter $r = r_{jj'}$ is fixed to 0.6, the sample size T is set to 1000 and we set $\eta_k = 0$ for $k \geq 1$ (i.e. $\varepsilon_i(t)$ and $\varepsilon_{i'}(t)$ are uncorrelated). Finally, we consider the following spatial function for the intra-correlation: $\rho_{|i'-i|} = \max(1 - |i'-i|/K_{\max}, r_{\min})$. In particular we consider two different models for the intra-correlation:

- Model 1: $K_{\max} = 300$, $r_{\min} = 0.9$.
- Model 2: $K_{\max} = 100$, $r_{\min} = 0.6$.

Model 1 is model ling a region of interest where the voxels are well connected while model 2 will be used to show the problems that could occur when two voxels inside a common region are not as connected as they should be. Figure 1 shows the intra-correlation matrix for both these models.



(a) Model 1



(b) Model 2

Figure 1: Intra-correlation matrices for the models 1 and 2 and for the largest region containing 40^2 voxels. The bottom-left square corresponds to the intra-correlation matrix of the smallest region containing 20^2 voxels.

4.1 Influence of the intra-correlation in absence of noise

We first study the properties of the estimators in absence of noise ($\sigma_\varepsilon = \sigma_e = 0$). Figure 2 shows boxplots based on 500 replications of the general model (1) with parameters described previously. Figure 2 (with $\sigma_\varepsilon = \sigma_e = 0$) shows that when the intra-correlation within each region $\mathcal{R}_j$ and $\mathcal{R}_{j'}$ is high for any pair of voxels, then the estimates of r are quite satisfactory even if we can already observe a slight bias for the method CA. However, when the model 2 is considered, the bias of the latter method is significant. It is to be noticed that in terms of variance the methods based on replicates (i.e. methods R, ℓR) are as efficient as the methods AC and ℓCA. The methods based on the a priori knowledge of two other disconnected regions (methods based on "differences", i.e. the methods D, ℓD, RD and ℓRD) have higher dispersion than the other ones. Finally, we also observe that for the "complex" methods ℓR and ℓRD based on local averages generate a more important bias than other methods. This bias is clearly smaller than the one observed for the method CA and illustrates the following fact: when we suspect that the intra-correlation matrix is not very well concentrated on the diagonal then the size of the neighborhood (ν in the definition of the estimators) should be chosen sufficiently small.

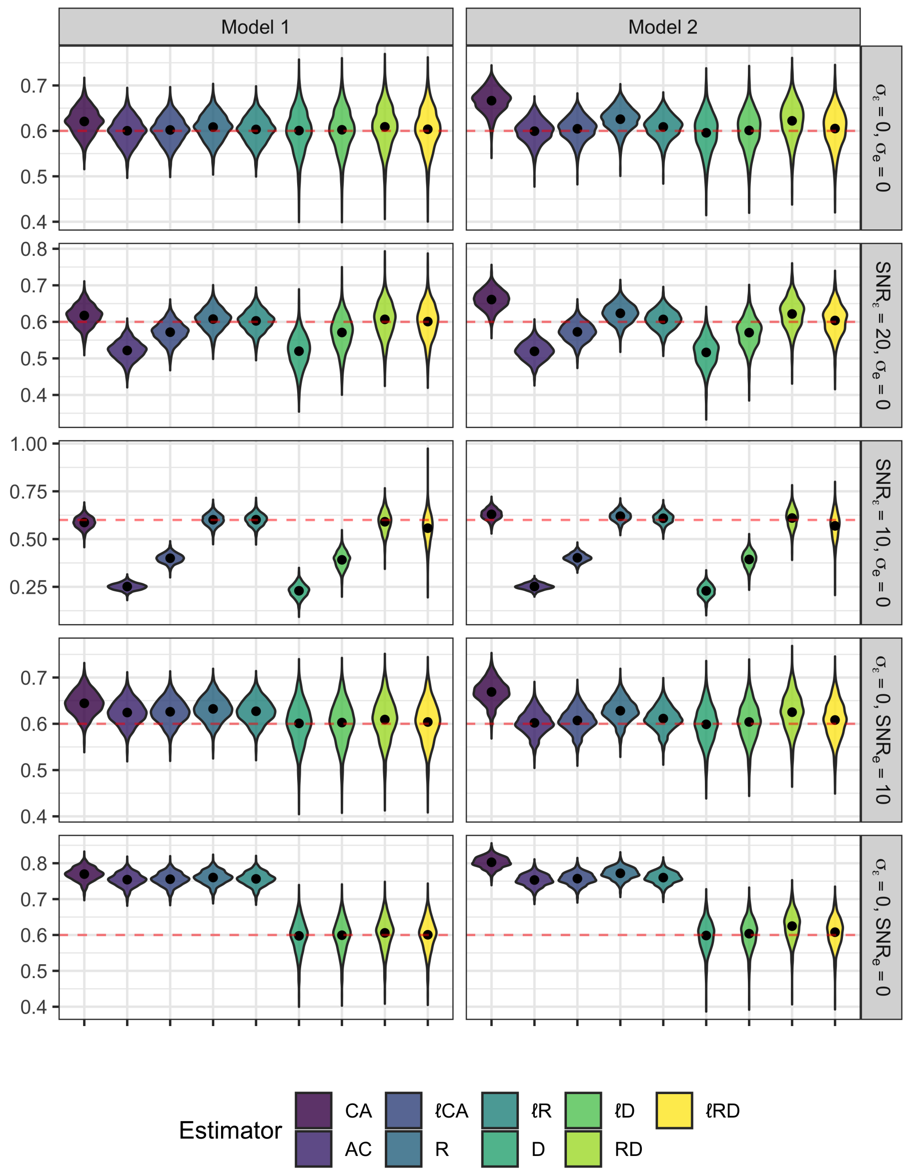


Figure 2: Boxplots of estimates of the inter-correlation parameter r based on 500 replications of the general model (1) for the models 1 and 2 of the intra-correlation matrix. Situations with no noise, local noise or global noise are considered with different levels of signal to noise ratio.

4.2 Influence of the local noise and global noise

We now turn to study the influence of a local noise or a global noise are also illustrated in Figure 2. The variances of the local noise or the global noise are parameterized by fixing the signal-to-noise ratio: when for instance $\sigma_e = 0$, we fixed σ_ε^2 as follows

$$\mathrm{SNR}_\varepsilon = 10 \log_{10} \left(\frac{\min(\sigma_j^2, \sigma_{j'}^2)}{\sigma_\varepsilon^2} \right) \Leftrightarrow \sigma_\varepsilon^2 = 10^{-\mathrm{SNR}_\varepsilon/10} \min(\sigma_j^2, \sigma_{j'}^2).$$

When $\sigma_\varepsilon \neq 0$ and $\sigma_e = 0$, as expected, Figure 2 shows that the methods based on replicates are able to estimate correctly the inter-correlation parameter r and these methods remain efficient whatever the value of the SNR and for both models of the intra-correlation matrix. The methods AC, ℓCA, D and ℓD are strongly affected by this additional local noise and exhibit a high negative bias. As explained in Section 3.2, the method CA which averages the signals in the regions $\mathcal{R}_j$ and $\mathcal{R}_{j'}$ is able to reduce the effect of the local noise.

Now, focusing on Figure 2 when $\sigma_\varepsilon = 0$ and $\sigma_e \neq 0$, which corresponds to situations where there is only a global noise, we observe that the higher the SNR_e the higher the positive bias for all the methods except the ones which were expected to handle the global noise, namely the methods D, ℓD, RD and ℓRD. It is also to be noticed that the dispersion of all the methods do not seem to be very affected by this extra noise.

5 Conclusion

In this paper we illustrate the effect of averaged data on estimations of correlation when two noises are present, local and global noise. The use of the classical correlation of averages is perturbed by the presence of these noises in addition to the presence of within correlations. We proposed alternative estimators including correction terms to compensate the intra-correlations, local and global noises. The performances of these estimators are illustrated on simulations.

References

Achard, S., Coeurjolly, J., Marcillaud, R., Richiardi, J., 2011. fMRI functional connectivity estimators robust to region size bias, in: IEEE Workshop on Statistical Signal Processing, SSP2011, Nice, France.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. Journal of Neuroscience 26, 63–72.

Bergholm, F., Adler, J., Parmryd, I., 2010. Analysis of bias in the apparent correlation coefficient between image pairs corrupted by severe noise. J Math Imaging Vis 37, 204–219.

Clifford, P., Richardson, S., Hemon, D., 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45, 123–134.

Donner, A., Eliasziw, M., 1991. Methodology for inferences concerning familial correlations: a review. J Clin Epidemio 44, 449–455.

Gunst, R.F., 1995. Estimating spatial correlations from spatial-temporal meteorological data. Journal of climate 8, 2454–69.

Liebhold, A., Sharov, A., 1998. Testing for correlation in the presence of spatial autocorrelation in insect count data. Population and Community Ecology for Insect Management and Conservation , 1–8.

Ostroff, C., 1993. Comparing correlations based on individual-level and aggregated data. Journal of Applied Psychology 78, 569.

Rosner, B., Donner, A., Hennekens, C., 1977. Estimation of interclass correlation from familial data. Applied statistics 26, 179–187.

Student, 1914. The elimination of spurious correlation due to position in time and space. Biometrika 10, 179–181.

Termenon, M., Jaillard, A., Delon-Martin, C., Achard, S., 2016. Reliability of graph analysis of resting state fMRI using test-retest dataset from the human connectome project. Neuroimage 142, 172–187.

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. Perspectives on Psychological Science 4, 274–290.

Ye, J., Lazar, N., Li, Y., 2011. Sparse geostatistical analysis in clustering fMRI time series. Journal of Neuroscience Methods 199, 336–345.