

RESTING STATE FMRI FUNCTIONAL CONNECTIVITY ANALYSIS USING SOFT COMPETITIVE LEARNING ALGORITHMS

Alberto A. Vergani^{*}, Elisabetta Binaghi^{*}, Samuele Martinelli^{*}, Sabina Strocchi[†]

^{*}University of Insubria, Department of Theoretical and Applied Science (DiSTA)
Varese

aavergani@uninsubria.it
elisabetta.binaghi@uninsubria.it
smartinelli@uninsubria.it

[†]Ospedale di Circolo e Fondazione Macchi, Medical Physics
Varese

sabina.strocchi@asst-settelaghi.it

Keywords: RS-fMRI, Functional Connectivity, Clustering, Soft Competitive Learning.

Abstract: *RS-fMRI data analysis for functional connectivity explorations is a challenging topic in computational neuroimaging. Several approaches have been investigated to discover whole-brain data features. Among these, clustering techniques based on Soft Competitive Learning (SCL) have been shown effective in providing useful information in various contexts. However, although significant achievements have been reached, these techniques still present critical aspects that require further investigations. We selected three clustering algorithms, i.e. Self-Organizing Maps (SOM), Neural Gas (NG) and Growing Neural Gas (GNG), to study the intrinsic functional properties of images coming from a shared repository of resting state fMRI experiments (1000 Functional Connectome Project, i.e. Oxford dataset). To compare the functional connectivity based on soft clustering, we calculated the Seed Based Linear Correlation (SBLC) to study the Default Mode Network (DMN) functionality, i.e. we found that Precuneus L/R has the higher Correlations Coefficients with its contralateral part and with the posterior division of Cingulate Gyrus. The differences among the three soft clustering algorithms adopted were measured basing on Jaccard Similarity Coefficient (JSC), whereas the quality of clusters has been evaluated with Davies-Bouldin Index (DBI). The optimal clustering computation was with 2 partitions for all the algorithms. We obtained the following results: a) clusters differentiated the amplitude of BOLD signals for both Males and Females, i.e. low level signal vs high level signal; b) clusters also differentiated the quality of seed-based correlations, i.e. strong (positive) associations vs weakly associations. These multivariate outcomes highlighted the complementary usage of clustering algorithms with statistical signal processing: the first made the partitions, the last explain the partitions.*

1 INTRODUCTION

The main goal of our study is to integrate different methodologies useful to discover and to explore the inner properties of brain signals, with application to resting state BOLD time series in healthy subjects. The motivation of our study emerges in relation to a recent work published by Biswal et al. [1], that highlighted a universal architecture of functional connections in the brain resting state networks, with age and sex as significant determinants. The specific goal of our analysis was to evaluate the following points: 1) if there is a between gender functional variability, i.e. if there is a statistical BOLD signal difference between males and females, 2) if there is a within gender functional variability, i.e. if male and female exams have different spread, and 3) if there is a confirmation of some interesting functional connectivity networks. In particular, our intent was to extend this conclusion with clustering algorithms to find similarities in time series (signatures) or in activation patterns (exams).

Approaching the general fMRI signal processing with data-driven methods is a challenging application (cfr. works by Lachiche [2] and Liao [3] for an overview and Vergani et al [4] for our recent work with clustering techniques); data-driven methods are used as well in the resting state paradigm in functional neuroimaging (cfr. papers by Margulies [5], Van Den Heuvel [6], Lee [7], and Wang [8]). For this purpose, we adopted as data driven methods the soft competitive learning algorithms to explore the *natural partitions* of the data (cfr. [9] and [10]). We choose Self Organizing Map (SOM), Neural Gas (NG) and Growing Neural Gas (GNG), that are a soft class of unsupervised artificial neural networks.

SOM are models initially proposed by Kohonen [11] and they are widely used because they allow the representation of data in a low-dimensional space, preserving the topological properties of the entrance space. SOMs are single-layer feedforward neural networks where output neurons are organized into low-dimensional grids (typically 2D or 3D spaces). The number of clusters that will be created is defined a priori.

NG is an alternative approach to SOM networks [12] [13]. The name derives from the fact that the neurons in the data space are moved as particles of a gaseous element, all negatively charged. Neurons repel each other, occupying the surrounding space, but they are attracted by areas of high data density as if the latter are positively charged particles. The NG algorithm is part of the soft competitive learning family, where not only the winning unit is adapted after the presentation of an input data, but also the remaining units. Unlike the SOM algorithm, no fixed topology is imposed on the network, the neurons are not arranged on the grid (they are free in space). Learning is performed according to a leaky learning strategy, by updating not only weight vectors of the winner neurons but also weight vectors of all losing neurons with a smaller rate that decreases in function of the increasing distance with the current input data.

GNG algorithm is an extension of NG and it was developed by Fritzke [14] [15]. Given a certain distribution of input data in the real domain, GNG incrementally creates a graph, or a network of nodes, where each node in the graph has a position in \mathbb{R}^n . GNG is an adaptive algorithm because if the distribution of input data changes over time, GNG can adapt, that is to move the nodes in order to adapt to the new distribution. In this graph the number of nodes is increased incrementally starting from two initial nodes. The nodes are considered neighbors if they are connected by an edge, and the neighborhood information is maintained during the execution of the algorithm basing on a variant of the standard Competitive Hebbian learning

(CHL). The big difference compared to SOM and NG is that it is not necessary to establish previously the number of *a priori* nodes (clusters) since the nodes are added incrementally during execution. An edge is associated to each node that, through a vector, represents the position in the node space. The edge has an associated age variable and a local error variable that has the purpose to indicate the insertion point of a new node. GNG is an algorithm with many parameters and it is complex. Its strong point is the adaptation of nodes that can also be deleted. This allows to free users of the burden of choosing *a priori* the number of clusters. The weak point is the difficulty in finding the optimal value for the all the parameters involved.

Keeping in mind the peculiarities that have these clustering algorithms, we want to understand with more details the features of the elements partitioned. In other word, we decided to complement the clustering outcomes with the classical signal processing methodologies adopting methods able to analyse the temporal dynamic of the BOLD brain signals and the spatial features related to specific regions of interest. Also, we studied the functional associations across the brain regions about the spontaneous signal fluctuations depending by the resting state fMRI experimental paradigm.

Therefore, we first approach the resting state signals with standard tools for image processing making filtering, motion correction, standard registrations, labelling and data reduction procedures. Then we integrated measures of strength/weakly signals association to investigate – in general – the cross correlations between all-ROIs with all-ROIs, and – precisely – the cross correlation with all-ROIs and two seeds: Left and Right Precuneus, that it is a bilateral region that has a role of central hub in the so-called Default Mode Network (DMN) (cfr. the historical work by Biswal [16] [17] and for the anatomo-functional details about the brain resting network cfr. Raichle et al [18] [19] or Utevsky et al [20] and also the work by Van Den Heuvel where he shown alternative to DMN [21] and work by Irajiloo [22] for technicalities about the resting state *connectivity-domain* analysis).

The general aim of this study is to address the functional connectivity problem in the resting state neuroimaging using both classical signal processing methods and soft clustering techniques. In the next sections, we propose and justify the type of data we have selected from a repository specialized in resting state functional neuroimages; then we present methods, results, discussion and conclusions.

2 DATA

Within the NITRC repository (<https://www.nitrc.org/>) and the 1000 Functional Connectome Project (http://fcon_1000.projects.nitrc.org/), we selected the Oxford dataset with 22 healthy subjects (12M /10F; ages 20-35). The fMRI parameters were the following: TR = 2, slices = 34, time-points = 175, magnet = 3 [T]. The selection of this dataset is motivated by the nice age balance and the small age spread that have the subjects. Furthermore, we selected this dataset because was one used by Biswal et al to discover resting state functional properties and their gender determinants. Therefore, our approach is also a confirmatory data analysis.

3 METHODS

The methods we used followed this pipeline: image processing, data reduction, statistical analysis and algebraic measurements, functional connectivity investigations with linear correlation and soft competitive clustering techniques.

Image Processing Image preprocessing was done with the functions for resting state image analysis contained in the software FSL [23]: we did spatial filtering with 3 [mm] of smoothing, frequency filtering with a high pass filter having 1/100 [Hz] as cutoff frequency, motion correction and standard registration with a reference atlas MNI152 (2mm). The ROIs labelling was done with the Harvard-Oxford atlas with 96 lateralized labels.

Data Reduction For both Females and Males, we did temporal signal reduction, spatial signal reduction, and whole brain signal reduction; i.e. the temporal data reduction was done with the extraction of mean and standard deviation of BOLD signals according to each time points, whereas the spatial reduction was the same but according to each atlas ROIs; the whole brain reduction is the global average of mean and standard deviation obtained by temporal reduction, with the aim to have two macro-signals, one for Females and one for Males.

Descriptive Statistics To investigate if Females and Males are samples coming from different populations, we tested the mean and the standard deviation of the whole brain signals with parametric (one-way ANOVA) and non-parametric test (Kruskall-Wallis); we choose both kind of tests because we have globally 22 subjects and some ANOVA assumptions are difficult to sustain; therefore, we preferred to compare the parametric results with the non-parametric outcomes, that had mild assumptions.

Algebraic Distances To compute metrics, we needed another step of data reduction: we averaged the spatial reduction results to have one value for each ROI, i.e. each exam became a vector with 96 components; then we measured how far are exams in vectorial forms from each other using Euclidean metric and Manhattan/Taxi-cub metric.

Correlations To study the brain functional connectivity, we used the Pearson's linear Correlation Coefficient (CC) applied to all-ROIs *versus* all-ROIs and applied to seed *versus* all-ROIs; the seeds we used were Precuneus Left and Right, according to the anatomical architecture of Default Network Mode, as a model for the brain resting state paradigm; we selected only the higher or the lesser seed correlation results according to specific cut-off, i.e. $CC > 0.8$ or $CC < -0.8$ and $-0.2 < CC < 0.2$, respectively, in order to evaluate strong (positive/negative) associations and weakly/absent associations with the seed.

Clustering To investigate brain resting functionality using unsupervised learning algorithms, we used Self Organizing Map (SOM), Growing Neural Gas (GNG) and Neural Gas (NG), i.e. to set GNG, we adopted 10 nodes, 1000 iterations, $\lambda = 2$, $\epsilon_b = 0.0005$, $\epsilon_n = 0.00001$, $\alpha = 0.05$, $\delta = 0.995$, age-node = 60. To set NG, we used 2 nodes, 500 iterations, $t_{max} = 8000$, $\epsilon_{initial} = 0.90$, $\epsilon_{final} = 0.50$, $\lambda_{initial} = 10$, $\lambda_{final} = 1$, $T_{initial} = 5$, $T_{final} = 10$. To evaluates the optimal partitions with the three clustering methods, we adopted Davies-Bouldin separation measure [24] and we compared the selected partitions with Jaccard similarity measures.

4 RESULTS

The results we obtained regard the statistical descriptions of resting state fMRI data, the algebraic measures of fMRI in vectorial forms, the functional connectivity studied with the Pearson linear coefficient correlation in both the conditions (all-ROIs vs all-ROIs and Seeds vs all-ROIs) and the general outcomes of the three clustering techniques with their comparison with the classical statistical signal processing approach.

Statistical descriptions of fMRI data The temporal analysis of RS-fMRI exams indicated that Females had the higher values for both the average and the standard deviation of the BOLD signals (cfr. Figure 1 - top); the spatial analysis revealed that Females and Males were similar for the average signals, but Females had more standard deviation (cfr. Figure 1 - bottom); both one-way ANOVA and Kruskal-Wallis test proved that there are statistical differences between males and females ($p\text{-value} \leq 0.05$) for the mean and the standard deviation of the whole brain RS-fMRI signals, i.e. Females have greater mean and a greater variance than Males (cfr. Figure 2 - top).

Algebraic measures of fMRI data Euclidean and Manhattan/Taxi-cub distances estimated that Females are more far from each other than Males, for both the mean and the standard deviation of the measures (cfr. Figure 2 - bottom).

Clustering validations and outcomes comparison Davies-Bouldin separation index (DB) indicated that SOM, NG and GNG reached the best data partitions with two clusters (cfr. Figure 3 - top). The Jaccard index computed for all the pairs (NG vs GNG, SOM vs GNG, SOM vs NG) in both Females and Males showed that the more similar clustering was in Males and Females between NG and GNG (cfr. Figure 3 - bottom).

Functional Connectivity with Linear Correlation Pearson's linear Correlation Coefficient (CC) applied to all-ROIs quantified that there is a difference between gender, i.e. Females had more negative correlations than Males, but they shared common positive correlations (cfr. Figure 4 – top – the left plot). Instead, the seed-based Correlation Coefficient (CC), with Left and Right Precuneus as seeds, shown that all the subjects had principally positive correlations, with a little presence of anti-correlation in Females (cfr. Figure 5 – top – the central and the right plot). Apart general comments about strength and weakly associations, it is remarkable the presence of the specific seeds weakly correlations, especially for the L/R Posterior Division of the Temporal Gyrus (ROI 29 and ROI 30) and for the L/R Anterior Division of Temporal Fusiform Cortexes (ROI 73 and ROI 74); also, it is remarkable the highest seeds positive correlation, especially for the L/R Cingulate Cortexes (ROI 59 and ROI 60) and with the contralateral part of the seed, the L/R Precuneus (ROI 61 and ROI 62) (cfr. Figure 5 - the correlation matrixes)

Functional Connectivity with Soft Competitive Clustering The two optimal clusters distinguished the quality of correlations for both the seeds (Precuneus Left and Right) analysis in Females and Males, i.e. strength (positive) coefficients and weakly coefficients were always mismatched in separated clusters (cfr. Figure 6 and Figure 7). We also noticed some overimposition for the central values of the coefficients. Furthermore, the two clusters discriminated the amplitude of BOLD signals, i.e. low and high intensities were mismatched as well in both Females and Males (cfr. Figure 4).

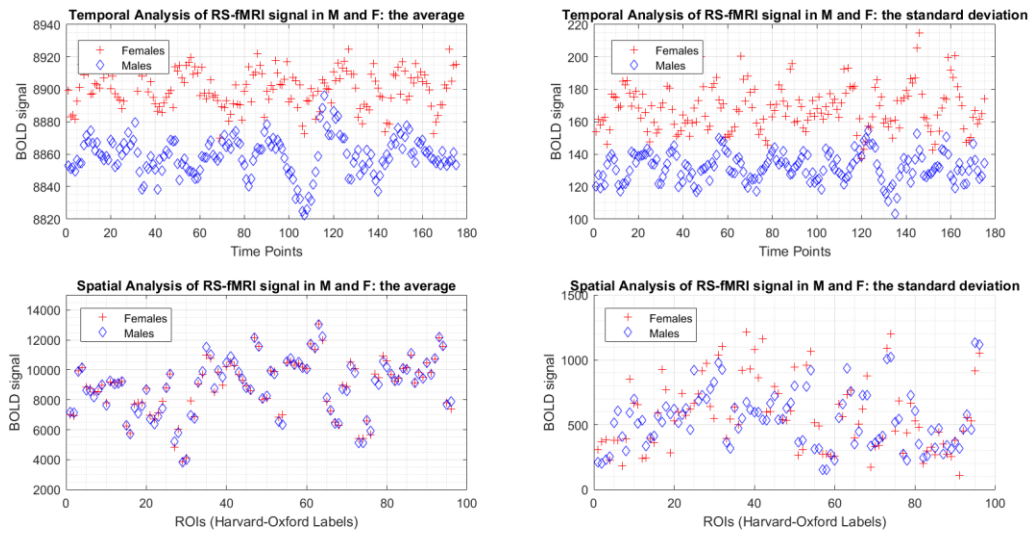


Figure 1. Plots of the average and the standard deviation of BOLD signals in Males and Females in the temporal (175 time-points *versus* BOLD signals) and spatial domains (96 ROIs *versus* BOLD signals). Both the average and the standard deviation of BOLD signals in the temporal domain are quite different between gender: Females have higher values than Males; whereas, in the spatial domain, both Males and Females have similar average BOLD signals, but Females have more standard deviation of BOLD signal.

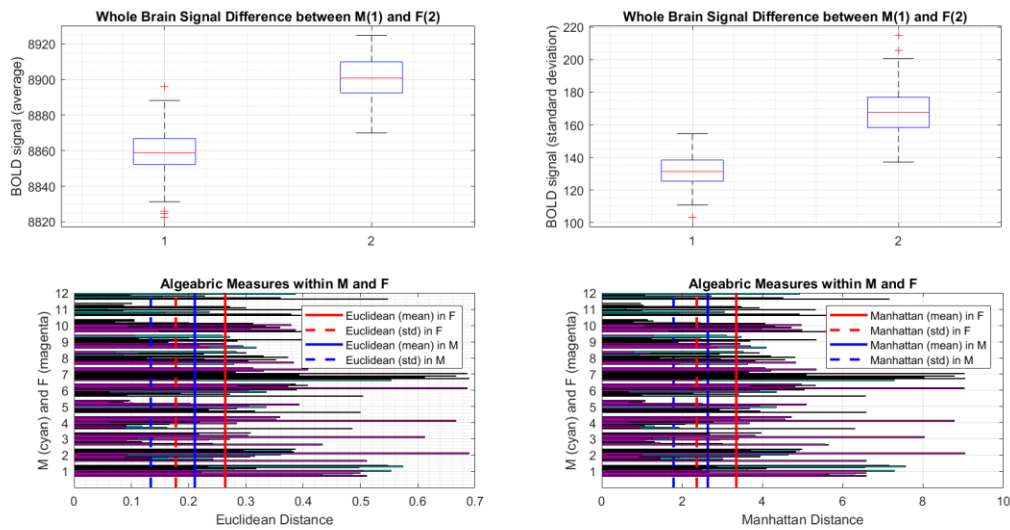


Figure 2. In the top part of the figure, there are the boxplots of whole brain signals in Males (1) and Females (2), showing the between gender statistical difference of the average (left) and the standard deviation (right) about the whole brain signals: in the both cases, Females have higher values than Males. In the bottom part of the figure, there are the within gender distance measures of the exams: with both Euclidean and Manhattan distances, Females have higher values than Males for the mean and the standard deviation of the distances computed.

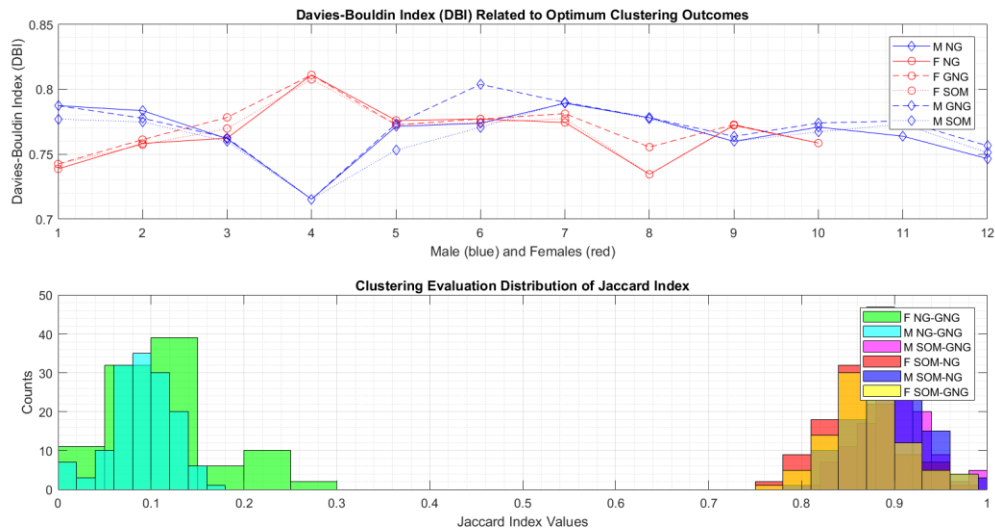


Figure 3. The top diagram shows the optimum Davies-Bouldin index associated to SOM, NG and GNG algorithms, for both Females and Males: all the DB are referred to 2 clusters as optimal clusters number for each algorithm; all the best indexes are under 1; the lesser is related to the fourth Males subject, and the higher to the fourth Female subject. The bottom histogram represents the discrete distribution of the Jaccard index computed for all the algorithm pairs: the more similar clustering outcomes are with the NG-GNG pairs in both Females and Males (Jaccard < 0.3), and the other clustering combinations are very different (Jaccard > 0.7).

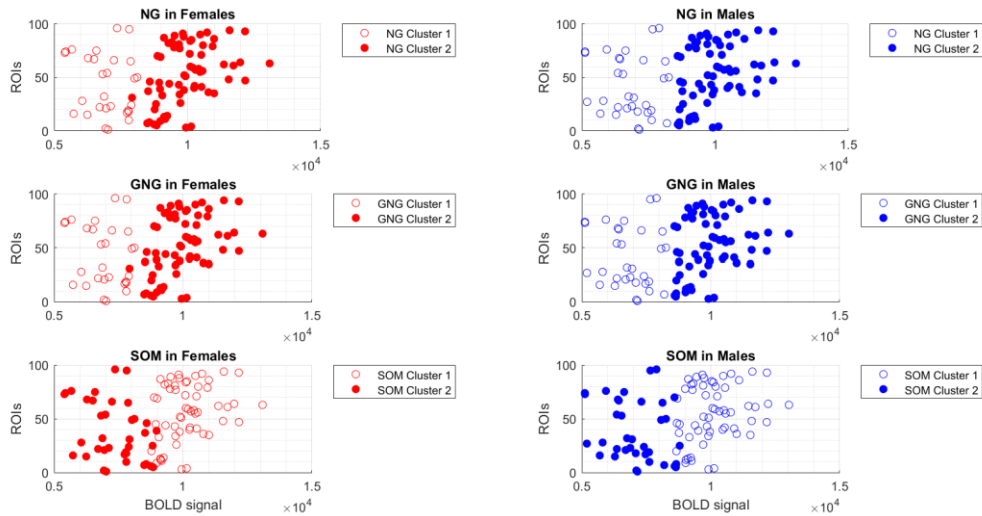


Figure 4. The scatter plots show the clustering outcomes for NG, GNG and SOM in Females (red points) and Males (blue points): the main result is that, using two clusters as optimum partitions number, the clusters detect the amplitude information of BOLD signals in Females and Males, i.e. clusters differentiate low levels and high levels in BOLD signals.

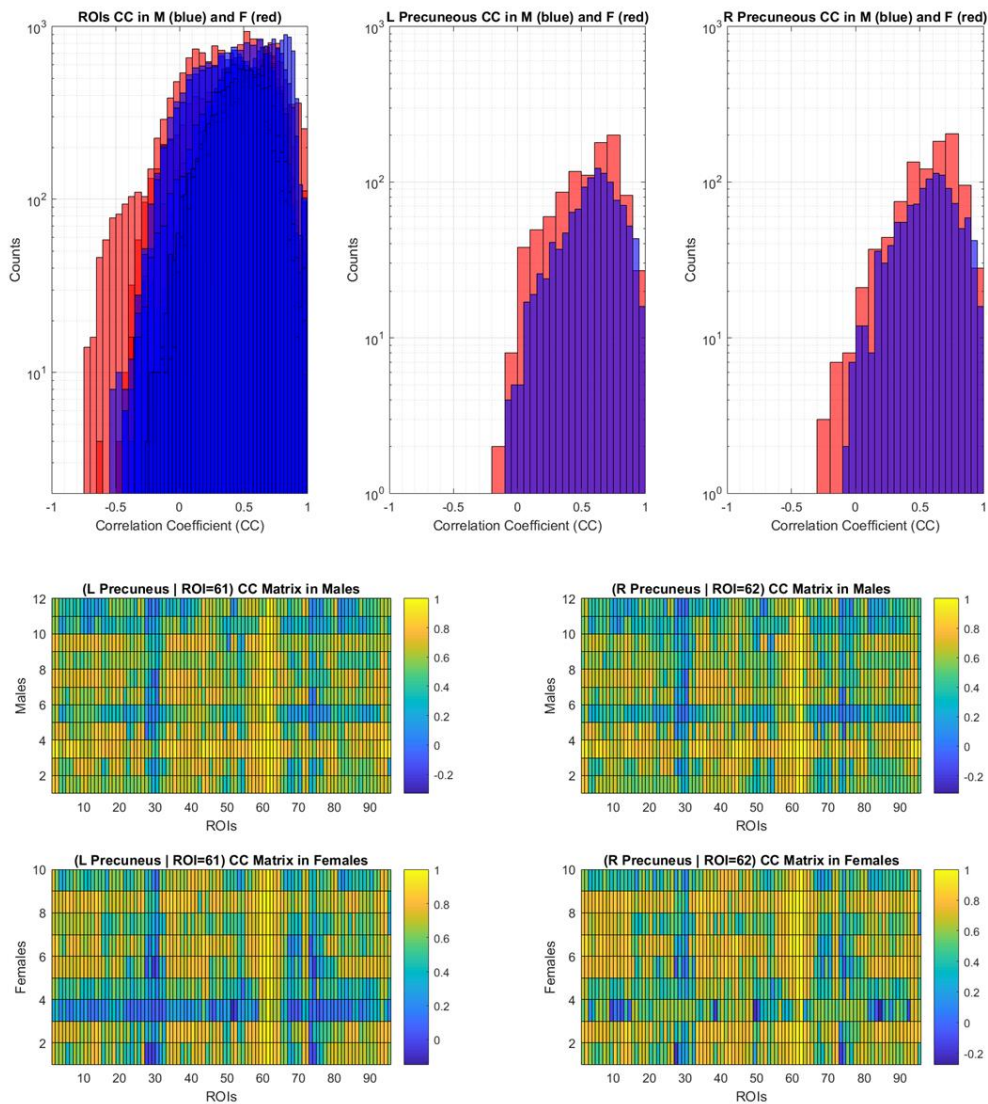


Figure 5. In the upper part of the figure, there are the Correlation Coefficients (CC) distributions between all-ROIs *versus* all-ROIs in Males and Females, and the correlation coefficients distributions between seeds (L/R Precuneus) *versus* all-ROIs. In the first discrete distribution, there are present positive and negative correlations, both in Males and Females; whereas, in the seeds based correlations distributions, there are less anti-correlations, with always, a bit gender difference. In the bottom part of the figure, there are the seed-based correlation matrix between Left and Right Precuneus in Males and Females: it is evident the strong correlation with many brain regions, as well as weak correlations with few brain regions. The highest positive correlations (CC > 0.8) are with seed-controlateral region (ROI 61 and ROI 62) and L/R Posterior Division of Cingulated Gyrus (ROI 59 and ROI 60). The lowest correlations (CC +/- 0.2) are with L/R Posterior Division of Temporal Gyrus (ROI 29 and ROI 30) and with the L/R Anterior Division of Temporal Fusiform Cortexes (ROI 73 and ROI 74).

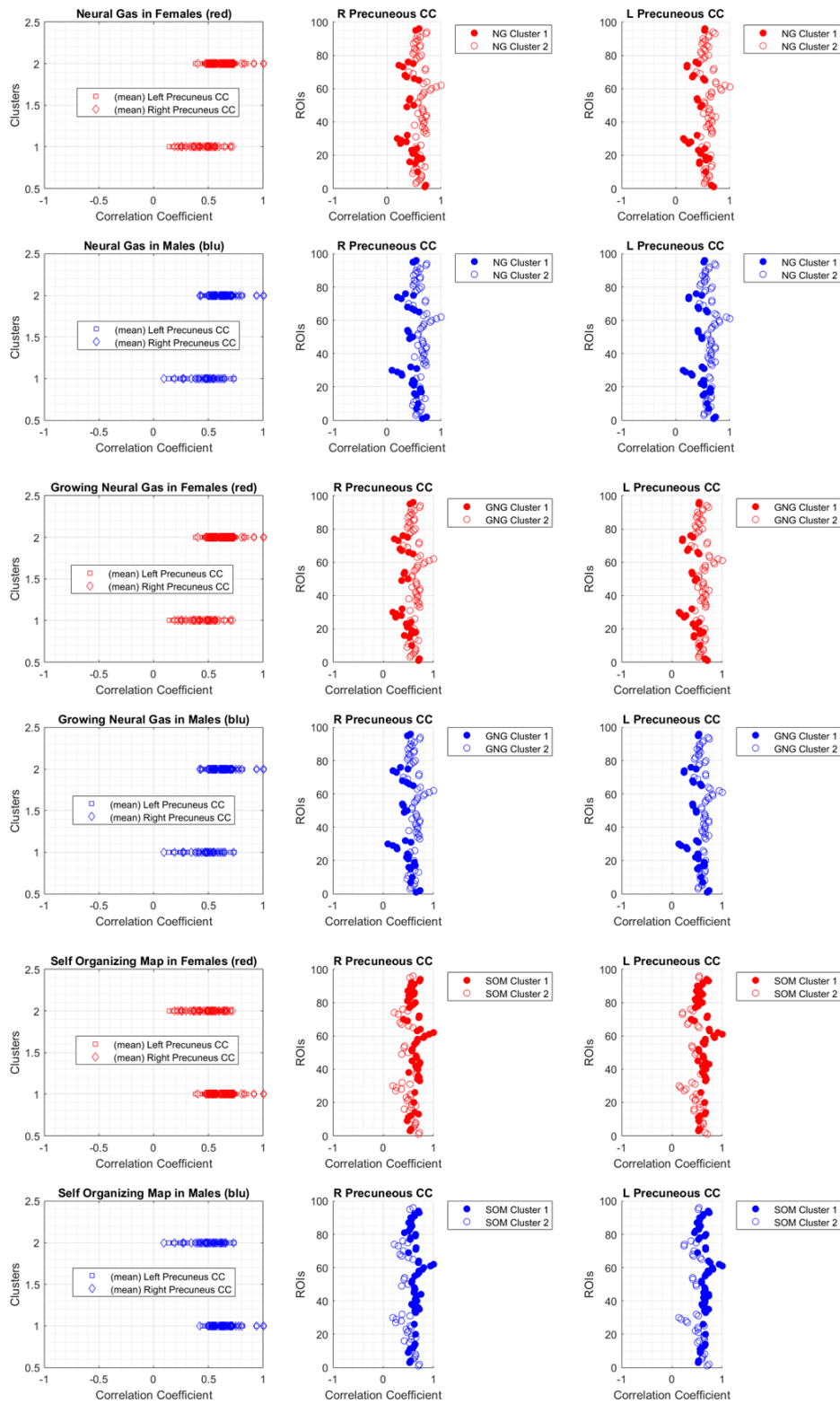


Figure 6. These figures represent the organization of the two optimal clusters in Males and Females in relation to the seeds based (L/R Precuneus) Correlation Coefficient. In the left column, it is reported the evidence that low and high correlations belong to different clusters, with some overlap between them for the central values. The central and the right columns show the relation between ROIs, correlation coefficients and clusters, showing precisely that low and high correlations are within different clusters in Males and Females.

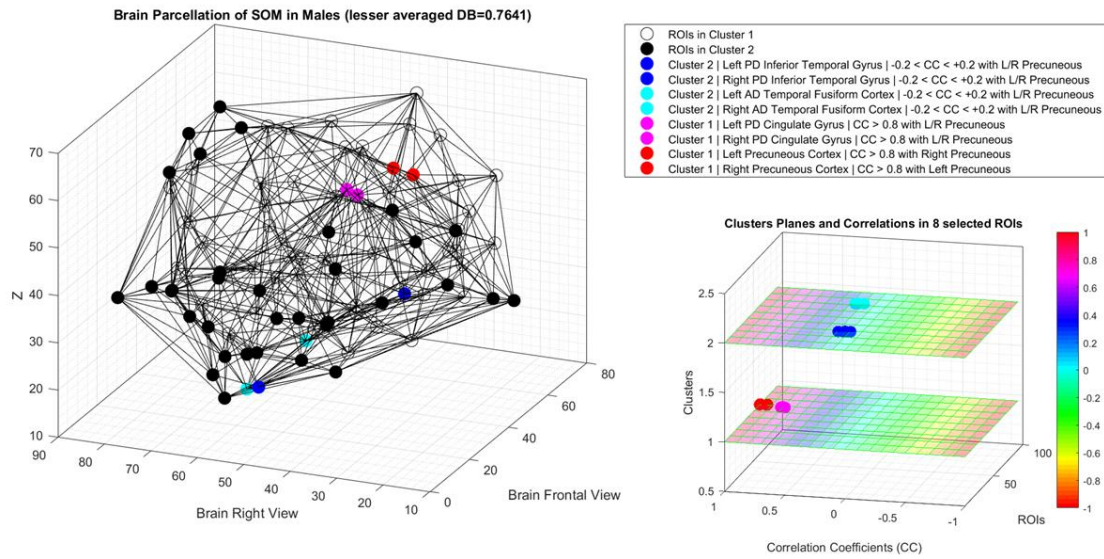


Figure 7. The image shows a detailed result of SOM algorithm in the Males case, with a specification about the higher and the lower correlations with both the seeds (Left and Right Precuneus), and their organization within the two clusters. In the left side of the figure, the brain is represented with the centre-of-mass of each regions of interest (the 96 ROIs labelled with Harvard-Oxford atlas), that are filled or unfilled if they belong to cluster 1 or 2. The coloured circles are the height selected regions that are, respectively, the four with the higher correlations (hot colours), and the four with the lower correlation (cool colours). In the right side of the figure, there is a plot with the clusters planes in relation to their elements: the four higher correlations regions are on the plane of cluster 1, whereas the four lower correlation are on the plane of cluster 2. We choose to plot the brain parcellation relative to the SOM algorithm in Males because was the better optimized algorithm, having as a Davies-Bouldin (DB) separation measure a value equal to 0.7641 (the lesser value respect the other algorithms).

5 DISCUSSION AND CONCLUSIONS

We obtained two types of results. The first-type result globally confirms the gender determinants in RS-fMRI functionality found by Biswal et al. Our approach is also related to the anatomo-functional correlation of a seeds (L/F Precuneus - DMN) with other regions, and specially with their contralateral part and their associated bilateral cingulated regions; this is a confirmation because we found brain functionality noted in the scientific literatures [25]. We also added information about Male and Females peculiarities using algebraic distances to measure the within gender variability. Globally, we can claim that Females had more amplitude and more variability than Males. The second-type result regards the integration of clustering techniques, with classical statistical processing for signal analysis. We can affirm that, once the clustering algorithm differentiates data in different clusters, it became necessary to explicate the inner property that determine their inclusion: in our case, we found that clusters differentiated the intensity of the brain signal, i.e. low *versus* high level BOLD signal, and clusters also differentiated the quality of the brain functional connectivity, i.e. strength (positive) associations *versus* weakly associations. There are limitations regard this study: the number of samples we adopted are limited – 22 subjects – but, interestingly, the statistical tests demonstrated that the two subsamples (10 Females and 12 Males) are significantly different (they did not come from the same population, and then they are nice comparable but not useful to infer general population properties). Next study will attempt to address these limitations, using also other soft competitive learning algorithms to cluster fMRI data, e.g. using fuzzy algorithms to better handle the complexity of the of Resting-State fMRI data.

REFERENCES

- [1] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe and others, "Toward discovery science of human brain function," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4734-4739, 2010.
- [2] N. Lachiche, J. Hommet, J. Korczak and A. Braud, "Neuronal clustering of brain fMRI images," in *International Conference on Pattern Recognition and Machine Intelligence*, 2005.
- [3] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857-1874, 2005.
- [4] A. A. Vergani, S. Martinelli and E. Binaghi, "Cluster Analysis of Functional Neuroimages Using Data Reduction and Competitive Learning Algorithms," in *European Congress on Computational Methods in Applied Sciences and Engineering*, 2017.
- [5] D. S. Margulies, J. B. Tettger, X. Long, Y. Lv, C. Kelly, A. Schfer, D. Goldhahn, A. Abbushi, M. P. Milham, G. Lohmann and others, "Resting developments: a review of fMRI post-processing methodologies for spontaneous brain activity," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, no. 5-6, pp. 289-307, 2010.
- [6] M. Van Den Heuvel, R. Mandl and H. H. Pol, "Normalized cut group clustering of resting-state FMRI data," *PloS one*, vol. 3, no. 4, p. e2001, 2008.
- [7] M. P. Van Den Heuvel and H. E. H. Pol, "Exploring the brain network: a review on resting-state fMRI functional connectivity," *Eur. Neuropsychopharmacol.*, vol. 20, no. 8, pp. 519-534, 2010.
- [8] M. H. Lee, C. D. Hacker, A. Z. Snyder, M. Corbetta, D. Zhang, E. C. Leuthardt and J. S. Shimony, "Clustering of resting state networks," *PloS one*, vol. 7, no. 7, p. e40370, 2012.
- [9] Y. Wang and T.-Q. Li, "Analysis of whole-brain resting-state fMRI data using hierarchical clustering approach," *PLoS One*, vol. 8, no. 10, p. e76315, 2013.
- [10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.
- [11] P. E. H. D. G. S. Richard O. Duda, *Pattern classification*, 2nd ed ed., Wiley, 2001.
- [12] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological cybernetics*, vol. 43, no. 1, pp. 59-69, 1982.
- [13] E. Dimitriadou, M. Barth, C. Windischberger, K. Hornik and E. Moser, "A quantitative comparison of functional MRI cluster analysis," *Artificial intelligence in medicine*, vol. 31, no. 1, pp. 57-71, 2004.
- [14] T. Martinez, K. Schulten and others, "A "neural-gas" network learns topologies," 1991.
- [15] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in neural information processing systems*, 1995.
- [16] B. Fritzke, "Let it grow-self-organizing feature maps with problem dependent cell structure," in *Artificial neural networks*, 1991.

- [17] B. Biswal, F. Zerrin Yetkin, V. M. Haughton and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537-541, 1995.
- [18] B. B. Biswal, J. V. Klyen and J. S. Hyde, "Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps," *NMR Biomed.*, vol. 10, no. 45, pp. 165-170, 1997.
- [19] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard and G. L. Shulman, "A default mode of brain function," *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 676-682, 2001.
- [20] M. E. Raichle and A. Z. Snyder, "A default mode of brain function: a brief history of an evolving idea," *Neuroimage*, vol. 37, no. 4, pp. 1083-1090, 2007.
- [21] A. V. Utevsky, D. V. Smith and S. A. Huettel, "Precuneus is a functional core of the default-mode network," *Journal of Neuroscience*, vol. 34, no. 3, pp. 932-940, 2014.
- [22] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782-790, 2012.
- [23] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on neural networks*, vol. 11, no. 3, pp. 586-600, 2000.
- [24] A. Iraj, V. D. Calhoun, N. M. Wiseman, E. Davoodi-Bojd, M. R. N. Avanaki, E. M. Haacke and Z. Kou, "The connectivity domain: Analyzing resting state fMRI data using feature-based data-driven and model-based methods," *Neuroimage*, vol. 134, pp. 494-507, 2016.
- [25] J. R. Andrews-Hanna, J. Smallwood and R. N. Spreng, "The default network and self-generated thought: component processes, dynamic control, and clinical relevance," *Annals of the New York Academy of Sciences*, vol. 1316, no. 1, pp. 29-52, 2014.