Investigating Statistical Differences in Connectivity Patterns Properties at Single Subject Level: a New Resampling Approach*

J. Toppi*, A. Anzolin, M. Petti, F. Cincotti, D. Mattia, S. Salinari, F. Babiloni and L. Astolfi

Abstract— **Methods based on the multivariate autoregressive (MVAR) approach are commonly used for effective connectivity estimation as they allow to include all available sources into a unique model. To ensure high levels of accuracy for high model dimensions, all the observations are used to provide a unique estimation of the model, and thus of the network and its properties. The unavailability of a distribution of connectivity values for a single experimental condition prevents to perform statistical comparisons between different conditions at a single subject level. This is a major limitation, especially when dealing with the heterogeneity of clinical conditions presented by patients. In the present paper we proposed a novel approach to the construction of a distribution of connectivity in a single subject case. The proposed approach is based on small perturbations of the networks properties and allows to assess significant changes in brain connectivity indexes derived from graph theory. Its feasibility and applicability were investigated by means of a simulation study and an application to real EEG data.**

I. INTRODUCTION

In neuroscience, the concept of brain connectivity is central for the understanding of the organized behavior of cortical regions beyond the simple mapping of their activity [1]. Connectivity estimation techniques aim to describe interactions between cerebral areas as connectivity patterns describing direction and strength of the information flow between such areas. The quantitative characterization of the main properties of the networks allows to extract a set of indexes able to summarize the information provided by the connectivity estimator [2].

Among several connectivity estimators, those based on Granger Causality are extensively used in neuroelectrical studies. Several studies have proved the accuracy of methods, such as Partial Directed Coherence (PDC) [3], which are defined in frequency domain and based on the use of MVAR models built on original time-series. The simultaneous integration of all the signals in the same autoregressive model, reducing the hidden sources dilemma typical of the bivariate approach, leads to an increase of the estimation accuracy [4, 5]. However, the increase of model dimensions in multivariate approaches requires a considerable amount of observations (repetitions of the same phenomenon) for guaranteeing the correct estimation of the MVAR parameters. For this reason, all the available observations are used to provide a unique and accurate estimate of the investigated connectivity network and of the corresponding indexes.

The unavailability of a connectivity distribution for the single subject does not allow to assess statistically significant changes in the network properties at different conditions. Group analysis can provide a solution when it is possible to collect a homogeneous sample, which is not always true, especially in clinical applications, where patients usually present heterogeneous conditions.

The application of statistical resampling to data before the connectivity estimation represents a partial solution to such issue. The use of jackknife or bootstrapping [5] methods, to be applied directly to the available observations, provides distributions of connectivity patterns that can support statistical comparisons. However, the computational cost of such approaches is very high because the multivariate estimate is computed at each resampling iteration. The burden of such computations might be avoided by shifting from data resampling to the resampling of connectivity matrices, with the aim to alter the connectivity patterns by keeping, at the same time, their main properties, resulting in a distribution of specific brain indexes.

In the present study we proposed a new approach which, operating directly on the adjacency matrix that describes the networks structure, aims at building a specific distribution for each brain index, without iterating the time consuming connectivity estimation process. To evaluate the new approach, we first performed a simulation study in which we tested its capacity to take into account the variability induced in connectivity patterns by physiological, instrumental and modeling factors always present in a typical electroencephalographic (EEG) study. Then we validated the method on real EEG data recorded during a cognitive task.

II. MATERIAL AND METHODS

A. Graph Theory Approach

A graph consists of a set of vertices (or nodes) and a set of edges (or connections) indicating the interaction between the vertices. The adjacency matrix G contains the

*Research supported by the European ICT Program FP7-ICT-2009-4 Grant Agreement 287320 CONTRAST.

J. Toppi (corresponding author, phone: +39-0651501165; e-mail: jlenia.toppi@uniroma1.it), M. Petti, F. Cincotti and L. Astolfi are with Dept. of Computer, Control, and Management Engineering, University of Rome "Sapienza", Italy and IRCCS Fondazione Santa Lucia, Rome, Italy.

A. Anzolin and D. Mattia are with IRCCS Fondazione Santa Lucia, Rome, Italy.

S. Salinari is with Dept. of Computer, Control, and Management Engineering, University of Rome "Sapienza", Italy.

F. Babiloni is with Dept. of Physiology and Pharmacology, University of Rome "Sapienza", Italy and IRCCS Fondazione Santa Lucia, Rome, Italy.

information about the connectivity structure of the graph and is built by comparing the values of connectivity matrix A with a statistical threshold representing the null-case [6]. Several indexes based on the elements of such matrix can be extracted for the characterization of the main properties of investigated networks: Global Efficiency, Local Efficiency [7], Degree, Characteristic Path Length and Clustering Coefficient [2].

B. The new Adjacency Matrix Resampling Approach

The new approach proposed in this paper is based on iterative small perturbations induced to the structure of the network whose properties we would like to characterize. The idea is to empirically reproduce the physiological, instrumental and modeling variability of real networks. Such perturbations are introduced by modifying the entries of the adjacency matrix according to three different criteria: a prefixed percentage of connections can randomly i) be deleted from the network, ii) be added to the pattern or iii) change position within the network. The random iteration of such perturbation process allows to build a distribution of the adjacency matrix and thus of all the indexes describing its properties. The verify if the validity of the proposed method and the selection of the best combination of the parameters to be set for employing it have been performed by means a simulation study in which the three approaches were statistically compared on the basis of the percentage of modified connections.

C. Simulation Study

The simulation study was conducted on 100 adjacency matrices extracted from real high resolution EEG data (51 channels) recorded on healthy subjects during 2 minutes of rest with eyes closed. The following steps were performed:

1. Each adjacency matrix was resampled by means of three different approaches (METHOD) in which a prefixed percentage of connections was randomly removed from the matrix (LEAVE-OUT), added to the matrix (ADD-ON) or moved in other positions (MOVED) respectively. The resampling was performed for different values of percentage of modified connections (%CONN: 5, 10, 20, 30, 50) and resampling iterations (RES-ITER: 50, 100, 250).

2. The following graph theory indices were extracted from each resampled adjacency matrix: Global Efficiency (GlobEff), Local Efficiency (LocEff), Path Length (PL), Clustering (Clust), Degree (deg) (INDEX).

3. In order to evaluate the properties of the resampled indexes distributions we defined two parameters, the relative polarization error (bias) and the distribution dispersion whose formulations are as follow:

$$bias = \frac{E[\hat{\vartheta}] - \vartheta}{\vartheta} \cdot 100 \quad (1)$$

where $E[\hat{\vartheta}]$ e ϑ represent the expected value of the estimate and the true value of the considered index respectively.

$$dispersion = \frac{\sqrt{\dfrac{\sum_{i=1}^{N} (\hat{\vartheta}_i - \bar{\vartheta})^2}{N}}}{\bar{\vartheta}} \quad (2)$$

where $\bar{\vartheta}$ and N represent the average value of index ϑ and the numerousness of the distribution respectively.

All the steps were repeated for each adjacency matrix and the results were subjected to a four-way repeated measures ANOVA computed considering as dependent variables the bias and the dispersion parameters and as within main factors the resampling methods, the percentage of modified connections, the resampling iterations and the graph indexes.

D. Application to Real EEG Data

The data used for testing the proposed methodology were recorded on a healthy subject who took part in a motor imagery experiment. Subject was asked to perform, according to the position of a red target on the screen, one of the following tasks: prolonged grasping of both hands (G) for the whole task length or just to relax (R). The experiment was divided into 6 runs of 24 trials each (randomly ordered). The task length was set to 15 seconds and the inter-trial interval to 2 seconds. EEG potentials were recorded by a 61-channel system by means of an electrode cap (BrainAmp, Brainproducts GmbH, Germany). Sampling rate was 200 Hz.

EEG data were pre-processed, segmented in the interval [-500;1000]ms according to the onset of the red target on the screen and were subjected to time-varying functional connectivity analysis [4], [8]. The connectivity patterns estimated for each time sample were averaged in six time intervals of 250ms each, defined according to the "GO" stimulus and in four frequency bands, defined according to the Individual Alpha Frequency (IAF) [9]. Connectivity patterns elicited during task and rest conditions were statistically compared for a significance level of 5%, False Discovery Rate corrected for multiple comparisons. The statistical threshold was also used for deriving the corresponding adjacency matrices on which different graph theory indexes were computed [6]. In order to compare, at single subject level, the indexes achieved in different conditions the LEAVE-OUT resampling method was applied. In particular we performed 50 resampling iterations by deleting each time 10% of connections.

III. RESULTS

A. Simulated Data

Results of the four way ANOVA computed on the two parameters extracted in the simulation study revealed a strong statistical influence of the main factors METHOD ($F=2903$, $p < 0.00001$), INDEX ($F=758$, $p < 0.00001$), and %CONN ($F=778$, $p<0.00001$), as well as their interactions METHOD x INDEX ($F=983$, $p<0.00001$), METHOD x %CONN ($F=7159$, $p<0.00001$), INDEX x %CONN ($F=887$, $p<0.00001$) and METHOD x %CONN x INDEX ($F=1707$, $p<0.00001$) on bias parameter. Similar results were found for the dispersion index (METHOD ($F=1255$, $p < 0.00001$), INDEX ($F=588$, $p < 0.00001$), and %CONN ($F=952$, p<0.00001), as well as their interactions METHOD x INDEX (F=506, p<0.00001), METHOD x %CONN (F=1740, p<0.00001), INDEX x %CONN (F=1017, p<0.00001) and METHOD x %CONN x INDEX (F=537, p<0.00001)). No significant effect of RES-ITER factor and its interactions with other factors was found on both parameters.

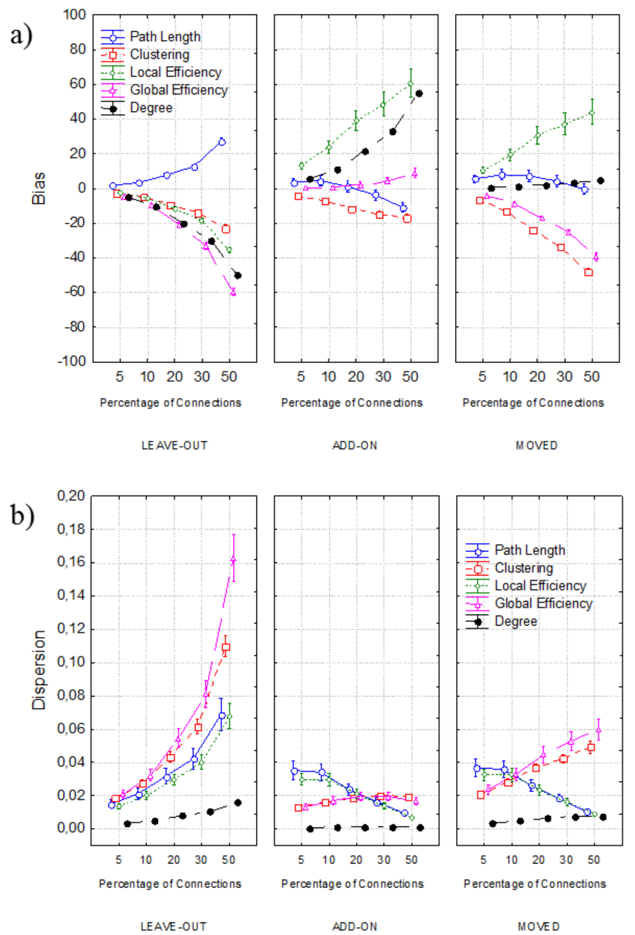

Figure 1. Results of ANOVA performed on Bias (panel a → F=1707, p<0.00001) and Dispersion (panel b → F=537, p<0.00001)) parameters computed on resampled indexes distributions, using METHOD x %CONN x INDEX as within main factor. The diagram shows the mean value of the investigated parameters. The bar represented their relative 95% confidence interval.

In Fig.1 we reported the results of ANOVA performed on Bias (panel a) and Dispersion (panel b) parameters computed on resampled indexes distributions, using METHOD x %CONN x INDEX as within factor. The polarization error increases its absolute value according to the increase of the percentage of modified connections for all the three methods and all the indexes. The error remains below 10% for all the indexes if the resampling is computed by deleting less than 10% of connections or by moving/adding less than 5% of arcs. The indexes which are underestimated by the LEAVE-OUT method are overestimated by the other two approaches and vice-versa, except for the clustering which is always underestimated. The dispersion index shows a different behavior for the three resampling methods. For LEAVE-OUT approach the dispersion increases according to the increase of the percentage of deleted connections for all the indexes. For the other two methods the dispersion of degree, global efficiency and clustering indexes increases according to the increase of added/moved arcs, while the dispersion of path length and local efficiency decreases.

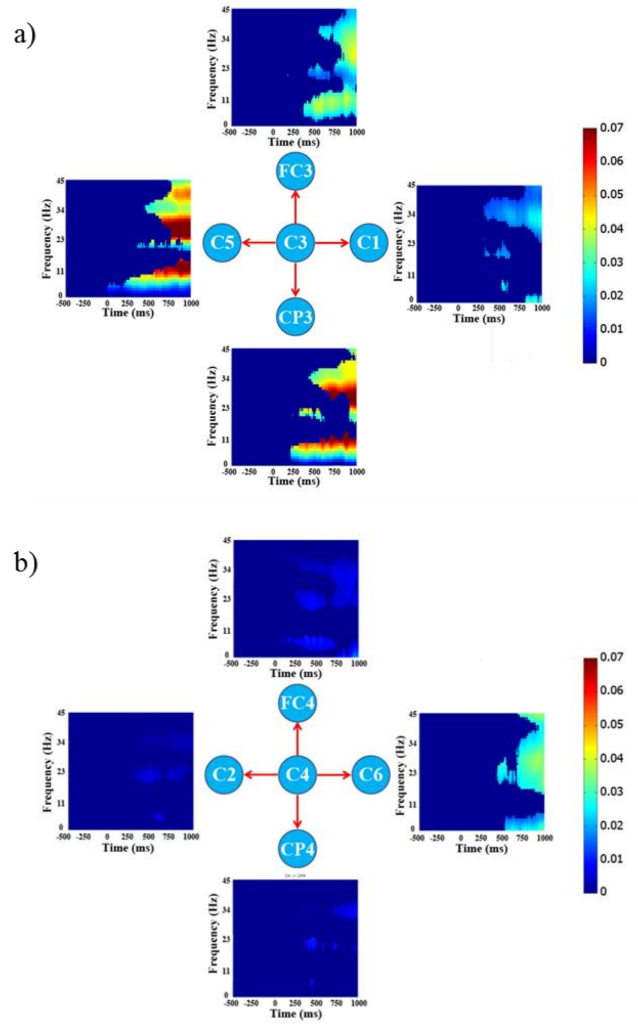

Figure 2. Time-frequency connectivity distribution for causal links spreading from C3 (panel a) and C4 (panel b) electrodes and directed to their nearest neighbors. In each matrix we reported the PDC values for the corresponding causal link achieved along all the time samples in [-500;1000] ms considered window and over all the included frequency samples in the range [1-45]Hz.

B. Real EEG Data

Time-varying connectivity was applied to the EEG data of a representative subject performing a grasping imagery task. Fig.2 shows the time-frequency connectivity distribution for causal links spreading out from C3 (panel a) and C4 (panel b) electrodes and directed to their nearest neighbors. No significant activations resulted before the “GO” onset between Grasping and Rest conditions for both channels. Since 250 ms after the beginning of imagery task a connection directed from C3 to its neighbors was found in the frequency range of Theta and Alpha bands. The strength of the significant connections for C4 was weaker than the one achieved for causal links spreading out from C3, confirming the important role of the left hemisphere (dominant hemisphere because the subject is right-handed) in motor imagery tasks [10]. In order to statistically confirming such difference in the involvement of the two electrodes during the motor imagery we applied the proposed resampling approach. In Fig.3 we reported a bar diagram with the trend of C3 (blue bars) and C4 (red bars) degree index across four different time intervals in Alpha band. The statistical analysis on resampled degree distributions revealed significant differences between the degree of the two electrodes since 250ms after the GO stimulus. In particular the degree of C3 resulted statistically higher than the one computed for C4.

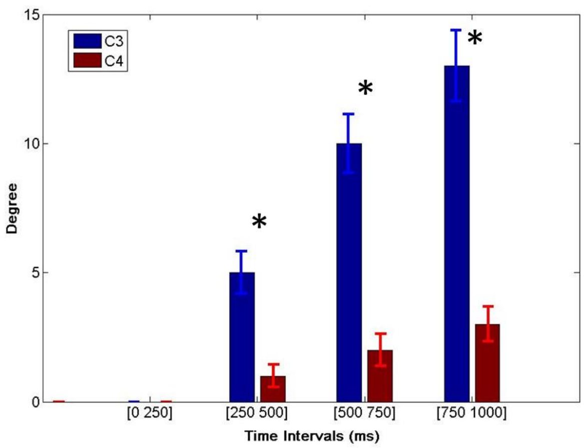

Figure 3. Bar diagrams reporting the trend of C3 (blue bars) and C4 (red bars) degree index across the four different time intervals ([0:250]ms, [250:500]ms, [500:750]ms, [750:1000]ms) in Alpha band. The symbol (*) highlights a statistical difference between C3 and C4 electrodes (unpaired t-test, $p<0.05$).

IV. Discussion

The results of the simulation study provided some guidelines for the use of the proposed method. As expected, both bias and polarization parameters increase according to the increase of the percentage of modified connections. In fact, the greater the perturbation applied to the network, the greater is the error in estimating the expected value of the indexes. Considering percentages of modified connections below 10%, we achieve distribution with dispersion around 3% and polarization error below 10%. High perturbations of the networks (%CONN > 20%) led to inaccurate distributions of graph indexes. Among the three different types of perturbations, the LEAVE-OUT approach achieved the best performances, because it allows, on equal conditions, to achieve distributions with higher dispersion and lower polarization error for all the considered indexes. The use of LEAVE-OUT approach with 10% of deleted connections is confirmed as valid choice by the application of the method to real data. The results achieved for the hand grasping imagery task and, in particular, the unbalance (controlateral > ipsilateral) between the two hemispheres, in agreement with the physiology of movement execution [11], [12] and imagination, confirmed the feasibility of a statistical analysis, allowing the evaluation of single subject significant changes in connectivity properties.

V. Conclusion

The results of the simulation study and the application to high density EEG suggest that the proposed method provides accurate distributions of the indexes describing the networks properties, with a procedure that requires reduced computational and time resources with respect to classical resampling approaches. The new method here proposed can thus provide an effective tool for assessing significant modifications in brain network properties at a single subject level.

Acknowledgment

Research supported by the European ICT Program FP7-ICT-2009-4 Grant Agreement 287320 CONTRAST.

References

[1] B. Horwitz, “The elusive concept of brain connectivity,” vol. 19, no. 2, pp. 466–470, 2003.

[2] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, “Organization, development and function of complex brain networks,” *Trends Cogn Sci*, vol. 8, no. 9, pp. 418–425, 2004.

[3] L. A. Baccalá and K. Sameshima, “Partial directed coherence: a new concept in neural structure determination,” *Biol. Cybern.*, vol. 84, pp. 463–474, May 2001.

[4] J. Toppi, F. Babiloni, G. Vecchiato, F. De Vico Fallani, D. Mattia, S. Salinari, T. Milde, L. Leistritz, H. Witte, and L. Astolfi, “Towards the time varying estimation of complex brain connectivity networks by means of a General Linear Kalman Filter approach,” *Conf Proc IEEE Eng Med Biol Soc*, vol. 2012b, pp. 6192–6195, 2012.

[5] B. Efron, “Bootstrap methods: another look at the Jackknife,” vol. 7, pp. 1–26, 1979.

[6] J. Toppi, F. De Vico Fallani, G. Vecchiato, A. G. Maglione, F. Cincotti, D. Mattia, S. Salinari, F. Babiloni, and L. Astolfi, “How the statistical validation of functional connectivity patterns can prevent erroneous definition of small-world properties of a brain connectivity network,” *Comput Math Methods Med*, p. 130985, 2012.

[7] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Phys Rev Lett*, vol. 87, no. 19, p. 198701, Nov. 2001.

[8] T. Milde, L. Leistritz, L. Astolfi, W. H. R. Miltner, T. Weiss, F. Babiloni, and H. Witte, “A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials,” *Neuroimage*, vol. 50, no. 3, pp. 960–969, Apr. 2010.

[9] W. Klimesch, “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis,” vol. 29, no. 2–3, pp. 169–195, Apr. 1999.

[10] G. Pfurtscheller and F. H. Lopes da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clin Neurophysiol*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[11] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[12] G. Pfurtscheller and A. Berghold, “Patterns of cortical activation during planning of voluntary movement,” *Electroencephalogr Clin Neurophysiol*, vol. 72, no. 3, pp. 250–258, Mar. 1989.