ORIGINAL PAPER

CrossMark

A Comparative Study of the Robustness of Frequency-Domain Connectivity Measures to Finite Data Length

Sara Sommariva[1] · Alberto Sorrentino[1] · Michele Piana[1] · Vittorio Pizzella[2,3] · Laura Marzetti[2,3]

Abstract

In this work we use numerical simulation to investigate how the temporal length of the data affects the reliability of the estimates of brain connectivity from EEG time-series. We assume that the neural sources follow a stable MultiVariate AutoRegressive model, and consider three connectivity metrics: imaginary part of coherency (IC), generalized partial directed coherence (gPDC) and frequency-domain granger causality (fGC). In order to assess the statistical significance of the estimated values, we use the surrogate data test by generating phase-randomized and autoregressive surrogate data. We first consider the ideal case where we know the source time courses exactly. Here we show how, expectedly, even exact knowledge of the source time courses is not sufficient to provide reliable estimates of the connectivity when the number of samples gets small; however, while gPDC and fGC tend to provide a larger number of false positives, the IC becomes less sensitive to the presence of connectivity. Then we proceed with more realistic simulations, where the source time courses are estimated using eLORETA, and the EEG signal is affected by biological noise of increasing intensity. Using the ideal case as a reference, we show that the impact of biological noise on IC estimates is qualitatively different from the impact on gPDC and fGC.

Introduction

The idea that the synergic cooperation of several regions is required for the brain to be able to instantiate specific functions and behaviors has, in the recent years, become central to neuroscience. Largely interconnected brain networks have indeed been reported to act as building blocks for the dynamic segregation and integration of brain areas

✉ Sara Sommariva
sommariva@dima.unige.it

[1] Dipartimento di Matematica, Università di Genova, Genoa, Italy

[2] Department of Neuroscience, Imaging and Clinical Sciences, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy

[3] Institute for Advanced Biomedical Technologies, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy

during task execution or at rest in a wide range of spatial and temporal scales (Engel et al. 2013). Understanding brain connectivity, both at structural and functional level, is thus a prerequisite for understanding brain functioning as well as its alterations (Stam 2010). In this framework, magnetoencephalography (MEG) and electroencephalography (EEG) have significantly contributed to unravel the functional wiring of the brain by putting emphasis on its temporal aspects, and looking for mechanisms of oscillatory coupling (in the range of about 1–100 Hz) as well as for coupled slower aperiodic fluctuations of brain activity (Engel et al. 2013). To robustly measure functional coupling in task related or ongoing brain activity, a lot of effort has thus been made in the recent years in the development of methods for MEG and EEG connectivity with the aim of capturing one aspect or the other (Marzetti et al. 2008; Stam and Straaten 2012; Hillebrand et al. 2012; Ewald et al. 2012; Marzetti et al. 2013; Chella et al. 2014; Brookes et al. 2011a, b, 2012; de Pasquale et al. 2010, 2012). Several time-domain or frequency-domain metrics have thus been designed to capture the statistical dependencies or the causal relationships

among EEG and MEG time series (Pereda et al. 2005; Sakkalis 2011; Van Diessen et al. 2015). Despite the flourishing of such methods, still several issues affect the robustness of the estimation (Schoffelen and Gross 2009; Baccalá and Sameshima 2014) which require systematic investigation and comparison between the different metrics in order to assess their performance in terms of statistical robustness. To date, only few studies have faced this issue.

For instance, Wendling et al. (2009) and Silfverhuth et al. (2012) study the robustness of some connectivity measures with respect to the relative strength of the interaction and of the innovation term in the underlying MultiVariate AutoRegressive (MVAR) model; Brookes et al. (2011a) compare different connectivity measures in simulations and on resting state MEG data to test for robustness and similarities with respect to fMRI resting state networks; Chella et al. (2014) introduce a novel third-order spectral connectivity measure and compare its statistical robustness to traditional frequency-domain measures by means of simulated and experimental EEG data; more recently Hincapié et al. (2016) use synthetic MEG data to compare the best regularization parameter required in order to estimate either the auto-spectral density function or the coherence function from the neural sources reconstructed by means of the minimum norm estimate. Even fewer studies explicitly investigate the robustness of connectivity estimates with respect to the length of the data; Astolfi et al. (2007) use synthetic data to compare three connectivity measures all based on an MVAR model; they exploit knowledge of the theoretical values of the considered connectivity measures to compute relative estimation errors; among other results, they quantify the minimum number of (possibly non-consecutive) time-points that are necessary to obtain good estimates of the connectivity patterns. Bonita et al. (2014) compare the robustness to epoch length of four time-domain connectivity measures, using experimental resting state data: the criterion of comparison is their ability to discriminate between eyes open and eyes closed conditions. Among other results, they show on a toy example how considering too long data can result in a wrong estimation of the connectivity measures, due to the changes in the underlying connectivity pattern. Similarly, Fraschini et al. (2016) use resting-state EEG recordings to investigate the impact of the epoch length on two different measures of functional connectivity. They show that this impact can be reduced by performing connectivity analysis in the source space and using a proper metric to characterize the topology of the estimated network. Wang et al. (2014) present a Matlab toolbox that allows to compare the performances of many connectivity measures, using synthetic data generated with several different models, and use this toolbox to find the minimum data length that is necessary to obtain reliable estimates of fourty-two connectivity measures; to provide a global evaluation of the reconstructed connectivity

networks, the authors use the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC), computed through a sliding window approach: they divided the data in overlapping windows and average the AUC estimated from them. More recently, Liuzzi et al. (2017) identify the length of the recorded data as a critical parameter that can compromise intra-subject repeatability of MEG findings, when connectivity is studied in the source space, by means of five different connectivity measures, grouped depending on whether they analyze the phase or the amplitude of the source time courses.

In the present study we investigate the impact of the length of the recorded data on three widely-used connectivity metrics, namely imaginary part of coherency (IC) (Nolte et al. 2004), generalized partial directed coherence (gPDC) (Baccalá et al. 2007) and frequency-domain granger causality (fGC) (Geweke 1982). Our aim is to characterize the reliability of these connectivity metrics in the source space, i.e. when they are computed between the neural source time courses estimated from EEG time series. We posit that there are at least three distinct contributions to error in these connectivity estimates: one comes from the finiteness of the data length; one from the fact that the source time courses are, in turn, estimated from EEG time series through an inverse procedure; the last one comes from the presence of biological noise, that provides an additional source of error. In order to investigate separately the contributions of such error terms, we start by assuming knowledge of the exact source time courses, and investigate the effect of reducing the number of samples in the time series. Then we compute their EEG signals, affected by different levels of biological noise, and the corresponding estimated source time courses, computed with exact low resolution brain electromagnetic tomography (eLORETA) (Pascual-Marqui 2007; Pascual-Marqui et al. 2011). We use surrogate data to assess the significance of the estimated connectivity values.

Materials and Methods

Before describing in detail the materials and the procedures used in this work, we introduce the general methodological framework that has been used to assess the impact of the length of the input data on connectivity analysis performed in the source space. To this end, we rely on thee metrics, namely IC, gPDC and fGC. These three metrics have been chosen as they are representative of different classes of connectivity measures commonly used in the literature (Baccalá and Sameshima 2014; Faes et al. 2012; Eichler 2006). Indeed, the former is an antisymmetric, bivariate measure computed from the Fourier transform of the time series in input, and is thus completely data-driven. Conversely, gPDC and fGC are multivariate, directed measures which assume a MultiVariate AutoRegressive (MVAR) model for the data;

gPDC only measures direct connections, while fGC detects all (direct and indirect) connections.

In order to explore the differences among the considered connectivity measures, we devise the following simulation scheme in which a connectivity pattern is designed as four time series following a stable MVAR model, with information flowing only from the first to the second and to the third signal. This particular simulation scheme has been chosen because it allows us to compute also the theoretical values of the connectivity measures under investigation. Moreover, the common situation (Bastos and Schoffelen 2015) of the presence of a single signal influencing two different unconnected sources is simulated, and the difference in the behaviour of bivariate (IC) and multivariate (fGC and gPDC) measures when facing this situation is investigated.

We interpret these simulated time series as the time courses of four dipolar sources and we use them to generate synthetic EEG data with different levels of superimposed biological noise. Then, we perform connectivity estimation from sub-samples of increasing length extracted either from the original source time courses or from the neural activity reconstructed using eLORETA. The results obtained from the true source time courses are used as a reference for the results obtained from the EEG data. More specifically, using the surrogate data test we can compute how many of the estimated connectivity values turn out to be significant, i.e., large enough for rejecting the null hypothesis of no connectivity. We do the same for pairs of interacting sources (exploiting knowledge of the true connectivity values to asses if the values that pass the statistical test are actually relevant) and for pairs of independent sources (so as to quantify the ratio of false positives). Surrogate data for statistical significance are generated by means of two different procedures: phase randomization (PR) (Theiler et al. 1992) and generation through an autoregressive model (AR) (Schreiber and Schmitz 2000). Both of these surrogates share the auto-spectral density function of the original signals, the main difference being that PR data are generated by randomly changing the phase of the signals in input, whereas AR surrogate data are generated by making use of MVAR models. Finally, we compute a correlation coefficient between the estimated and the true connectivity values, as function of the frequency, to characterize the ability to recognise the frequency at which interaction occurs. In the following, we first define the mathematical instruments used throughout the paper, i.e. the connectivity metrics, the statistical tests and the inverse model. Then, we present how we generate and analyse the data and eventually we describe the criteria used to evaluate the obtained results.

Everywhere in the paper, except where explicitly specified, the connectivity measures and the source reconstructions were computed using the Fieldtrip Matlab toolbox (Oostenveld et al. 2011).

Connectivity Metrics

Let $\{\boldsymbol{x}(t_p)\}_{p=1}^{P}$ be a set of N time series sampled at P time points representing the activity of N brain sources:

$$\boldsymbol{x}(t_p) = \left(x_1(t_p), \dots, x_N(t_p) \right)^T \quad p = 1, \dots, P. \tag{1}$$

In this study we consider three connectivity measures, all defined in the frequency-domain: IC, gPDC, and fGC. We notice that these measures are strictly related to the ones used by Haufe et al. (2013) for the simulated data available from the workshop titled *Controversies in EEG source imaging*, held in August 2014 at the University of Electronic Science and Technology in Chengdu, China, with the aim of discussing the major issues in estimating brain activity and interaction properties from electrical potentials. All the simulations and data are available from the website http://neuroinformation.incf.org.

Imaginary Part of Coherency

Imaginary part of coherency is a bivariate, antisymmetric measure of the linear relationship between two time series as function of frequency.

For each pair of signals $(x_i(t_p),\, x_j(t_p))$, $i, j = 1, \dots, N$, let $(\hat{x}_i(f_q),\, \hat{x}_j(f_q))$ be the corresponding Fourier transform. The coherency (Nunez et al. 1997) between them is defined as the ratio between their cross-spectral density function and the product of their individual auto-spectral density functions, i.e.

$$COH_{ij}(f_q) = \frac{S_{ij}(f_q)}{\sqrt{S_{ii}(f_q)S_{jj}(f_q)}} \tag{2}$$

where

$$S_{ij}(f_q) = \,< \hat{x}_i(f_q)\hat{x}_j(f_q)^* >, \tag{3}$$

$x_j(f_q)^*$ being the complex conjugate of $x_j(f_q)$. Coherency is a complex valued quantity, the magnitude of which is strongly influenced by volume conduction and source leakage effects in MEG and EEG (M/EEG) which can induce spurious connectivity (Schoffelen and Gross 2009). However, Nolte et al. (2004) showed that for independent signals, only the real part of the cross-spectral density function can be non-zero, while its imaginary part vanishes (a part from random fluctuations around zero). For this reason, in M/EEG it is common to use the IC as a connectivity metric robust to spurious connectivity. It is defined as (Nolte et al. 2004)

$$IC_{ij}(f_q) = \frac{\mathrm{Im}(S_{ij}(f_q))}{\sqrt{S_{ii}(f_q)S_{jj}(f_q)}}, \tag{4}$$

Im($S_{ij}(f_q)$) being the imaginary part of $S_{ij}(f_q)$. We notice that the detection of functional coupling by the IC is in fact influenced by the concurrent presence of independent sources at the frequency of interest, due to the normalization by the auto-spectral density functions; however, the effect of independent sources is a reduction of the value of $IC_{ij}(f_q)$.

From the computational viewpoint, the most delicate step for calculating $IC_{ij}(f_q)$ is the estimation of the cross-spectral density function, $S_{ij}(f_q)$, and of the auto-spectral density functions, $S_{ii}(f_q)$ and $S_{jj}(f_q)$. In the simulations below we use the Welch's method with a Hann window (Welch 1967): the measured signals are subdivided into N_{win} epochs of fixed length (1 s) and with an overlap of 50%. For each epoch and for each pair of signals, $i, j = 1, \ldots, N$, we compute the FFT of the time series $\{x_i(t_p), x_j(t_p)\}_{p=1}^{P}$; the estimated cross-spectral density function is then given by Eq. (3), where $< \cdot >$ stands for the average over epochs.

Generalized Partial Directed Coherence

Generalized partial directed coherence is a multivariate directed measure, whose definition is based on the assumption that the time series under investigation follow a stable MVAR process. Indeed, let us assume that the signals $\{\boldsymbol{x}(t_p)\}_{p=1}^{P}$ can be modelled through the following MVAR process

$$\boldsymbol{x}(t_p) = \sum_{k=1}^{K} A(k)\boldsymbol{x}(t_{p-k}) + \boldsymbol{\varepsilon}(t), \tag{5}$$

where K is the order of the process, $A(1), \ldots, A(K)$ are proper coefficient matrices of size $N \times N$ and $\boldsymbol{\varepsilon}(t_p) = \left(\varepsilon_1(t_p), \ldots, \varepsilon_N(t_p)\right)^T$ is a N-dimensional innovation process, which is a zero-mean, uncorrelated white noise process whit covariance matrix Σ. For each pair of signals, $i, j = 1, \ldots, N$, the generalized partial directed coherence from $\{x_i(t_p)\}_{p=1}^{P}$ to $\{x_j(t_p)\}_{p=1}^{P}$ is defined as (Baccalá et al. 2007)

$$gPDC_{ij}(f_q) = \frac{\frac{1}{\sigma_{jj}}|\overline{A}_{ji}(f_q)|}{\sqrt{\sum_{n=1}^{N} \frac{1}{\sigma_{nn}^2} \overline{A}_{ni}(f_q)\overline{A}_{ni}^*(f_q)}} \tag{6}$$

where for each $n = 1, \ldots, N$, σ_{nn}^2 is the nth diagonal element of Σ and

$$\overline{A}_{ji}(f_q) = \delta_{ji} - \hat{A}_{ji}(f_q) \tag{7}$$

$\hat{A}_{ji}(f_q)$ being the Fourier transform of the sequence of coefficients $\{A_{ji}(k)\}_{k=1}^{K}$ and δ_{ji} the Kronecker delta function.

We observe that $gPDC_{ij}(f_q)$ is non-zero only if $A_{ji}(k) \neq 0$ for some k, i.e. only if the past of $x_i(t_p)$ directly influences $x_j(t_p)$. For this reason $gPDC_{ij}(f_q)$ is often referred to as a frequency-domain formulation of Granger Causality (see next paragraph for more details).

From the computational viewpoint, the most critical step for estimating the $gPDC_{ji}(f_q)$ is the fitting of the MVAR model. In the simulations below we estimate the model order K by means of the Bayesian information criterion (BIC), which is a consistent estimator (Lütkepohl 2007); specifically, we make use of the MVGC toolbox (Barnett and Seth 2014). The coefficient matrices, $A(1), \ldots, A(K)$, and the covariance matrix of the innovations process, Σ, are estimated through the Levinson, Wiggins, Robinson (LWR) algorithm, implemented in the BSMART toolbox (Haykin 2008; Cui et al. 2008).

Frequency-Domain Granger Causality

Like gPDC, fGC (Geweke 1982) is a directed connectivity measure, whose definition is based on the assumption that the the signals $\{\boldsymbol{x}(t_p)\}_{p=1}^{P}$ are drawn from a MVAR process.

An operational definition is as follows: given a pair of time series $\{x_i(t_p), x_j(t_p)\}_{p=1}^{P}, i, j = 1, \ldots N$, $\{x_i(t_p)\}_{p=1}^{P}$ is said to Granger-cause $\{x_j(t_p)\}_{p=1}^{P}$ if the past of the former process helps to predict the future of the latter (Wiener 1956). In this study we consider the following formalization of Granger Causality in the frequency domain, which makes use of MVAR processes (Granger 1969; Geweke 1982).

Let $H(f_q)$ be the *transfer matrix* of the MVAR model, i.e.

$$H(f_q) = \overline{A}(f_q)^{-1} \tag{8}$$

where $\overline{A}(f_q)$ is the $N \times N$ matrix whose (i, j)-th element is $\overline{A}_{ij}(f_q)$, as defined in Equation (7); then, for each pair of signals $\left(x_i(t_p), x_j(t_p)\right), i, j = 1, \ldots, N$, the fGC is given by

$$fGC_{ij}(f_q) = \ln\left(\frac{|S_{jj}(f_q)|}{|S_{jj}(f_q) - H_{ji}(f_q)(\Sigma_{ii} - \Sigma_{ij}\Sigma_{jj}^{-1}\Sigma_{ji})H_{ji}(f_q)^*|}\right). \tag{9}$$

Roughly speaking, $fGC_{ij}(f_q)$ quantifies the portion of the total auto-spectral density function $S_{jj}(f_q)$ that comes from a causal influence of $\{x_i(t_p)\}_{p=1}^{P}$ over $\{x_j(t_p)\}_{p=1}^{P}$ (Schelter et al. 2006).

From a computational point of view, we estimate $fGC_{ij}(f_q)$ by first fitting an MVAR model, as described in the previous paragraph, and then estimating the transfer matrix $H(f_q)$ through (8) and the cross-spectral density function by means of the following equation (Papoulis and Pillai 2002)

$$S(f_q) = H(f_q)\Sigma H(f_q)^*, \tag{10}$$

$H(f_q)^*$ being the Hermitian transpose of $H(f_q)$.

Statical Test

In principle, all the three connectivity metrics defined above vanish if the input time series are independent. In practice, though, their empirical values are never exactly zero, even for independent signals, mainly due to the finite length of the data. Therefore, once the connectivity values have been estimated, their significance has to be assessed, by quantifying the probability that those values could be generated in absence of connectivity. In other words, we need to test the estimated values against the null hypothesis of no-connection among the signals. However, the distribution under the null hypothesis is usually unknown and a threshold of statistical significance has to be set by studying the asymptotic distribution of the connectivity measures (Baccalá et al. 2013) or by means of Monte Carlo methods such as the *surrogate data approach* (Theiler et al. 1992; Schreiber and Schmitz 2000).

In this study, we use the latter approach. The idea of surrogate data is that of producing a large number, N_{surr}, of synthetic datasets which share given properties with the original signals, but in which the connectivity pattern under investigation is destroyed. For each pair of surrogate time series, $i, j = 1, \dots, N$, and for each frequency, f_q, the connectivity measures are computed and the distribution under the null hypothesis is approximated as the empirical distribution of the obtained values. Given a significance level $\alpha \in [0, 1]$, the threshold of statistical significance is then set at the $100(1 - \alpha)$ percentile of the empirical distribution, i.e. $\tau_{ij}(f_q)$ is the lowest value such that

$$\#\{ \widetilde{M}_{ij}^{(s)}(f_q) \ s.t. \ |\widetilde{M}_{ij}^{(s)}(f_q)| > \tau_{ij}(f_q), \ s = 1, \dots, N_{surr} \} \le (1 - \alpha) N_{surr},$$ (11)

where $\widetilde{M}_{ij}^{(s)}(f_q)$ is the value of connectivity measure under investigation estimated for the sth surrogate dataset and, given a set A, $\#A$ is the number of elements in A. Then, if the empirical value of the connectivity measure falls below the threshold, the probability that it has been generated in absence of connectivity is considered too high and the null hypothesis cannot be rejected.

Notice that, since the connectivity measures we use in this study are frequency dependent, the thresholds are frequency dependent too.

Clearly, the first step is to define a way to generate surrogate data. This topic has been extensively studied in the recent years, so that rigorous methods have been developed, specifically devised to test the significance of given connectivity measures (Prichard and Theiler 1994; Schreiber and Schmitz 1996; Faes et al. 2010; Liu and Molenaar 2016). Here we use two methods that are rather general and are

appropriate for all the three connectivity metrics considered in this study. Both methods produce independent time series which share the auto-spectral density function with the original signals. The first one, i.e. *phase-randomized surrogate data*, is the simplest one as it does not require model assumptions, whereas the second one, i.e. *autoregressive surrogate data*, makes use of MVAR models.

Phase-Randomized Surrogate Data

Let $\hat{\boldsymbol{x}}(f_q) = \left(\hat{x}_1(f_q), \dots, \hat{x}_1(f_q) \right)^T$ be the Fourier transform of the original time series $\{\boldsymbol{x}(t_p)\}_{p=1}^P$.

For each signal, $i = 1, \dots, N$, a *phase-randomized* (PR) surrogate data,[1] $\{\chi_i^{(PR)}(t_p)\}_{p=1}^P$ is generated as follows (Faes et al. 2004; Theiler et al. 1992).

First for each frequency, f_q, we define

$$\hat{\chi}_i^{(PR)}(f_q) := |\hat{x}_i(f_q)| \ e^{-i\varphi_i(f_q)}$$ (12)

where $|\hat{x}_i(f_q)|$ is the modulus of $\hat{x}_i(f_q)$ and $\varphi_i(f_q)$ is randomly drawn from a uniform distribution on $[-\pi, \pi]$ under the constraint

$$\varphi_i(-f_q) = -\varphi_i(f_q).$$ (13)

Secondly $\{\chi_i^{(PR)}(t_p)\}_{p=1}^P$ is computed as the inverse Fourier transform of $\hat{\chi}_i^{(PR)}(f_q)$. The constraint in equation (13) is required to obtain a real valued time series.

From a computational viewpoint, we generate PR surrogate data by means of the eMVAR toolbox (Faes et al. 2013).

Autoregressive Surrogate Data

For each signal, $i = 1, \dots, N$, we assume that $\{x_i(t_p)\}_{p=1}^P$ can be modelled as a unidimensional, autoregressive process of order K_i, coefficients $A_i(1), \dots A_i(K_i)$ and with covariance of the innovation σ_i^2.

An *autoregressive* (AR) surrogate data (Faes et al. 2004; Schreiber and Schmitz 2000), $\{\chi_i^{(AR)}(t_p)\}_{p=1}^P$, is a realization of such AR process, i.e. for each $p = 1, \dots, P$,

$$\chi_i^{(AR)}(t_p) = \sum_{k=1}^{K_i} A_i(k)\chi_i^{(AR)}(t_{p-k}) + \varepsilon_i(t_p)$$ (14)

where $\varepsilon_i(t_p) \sim \mathcal{N}(0, \sigma_i^2)$.

From a computational viewpoint, first we estimate K_i by means of the BIC criterion implemented in the MVGC toolbox (Barnett and Seth 2014); secondly we use FieldTrip to

---

[1] In literature, this type of surrogate data is also known as *Fourier transform surrogate data* (Faes et al. 2004).

estimate the other parameters of the AR model and to generate $\{\chi_i^{(AR)}(t_p)\}_{p=1}^P$.

Source Reconstruction

Let $\boldsymbol{y}(t_p)$ be the EEG time series, recorded by N_s sensors; we assume that the data have been produced by neural sources through a linear model of the form

$$\boldsymbol{y}(t_p) = G\boldsymbol{z}(t_p) + \boldsymbol{\eta}(t_p) \tag{15}$$

where G is the leadfield matrix, of size $N_s \times 3N_v$, N_v being the number of voxels; $\boldsymbol{z}(t_p)$ is a vector of length $3N_v$ containing the three orthogonal components of the neural current at each voxel and $\boldsymbol{\eta}(t_p)$ is the measurement noise.

To reconstruct the source time courses $\boldsymbol{z}(t_p)$ from the recorded potential $\boldsymbol{y}(t_p)$ we use the exact low resolution brain electromagnetic tomography (eLORETA) (Pascual-Marqui 2007; Pascual-Marqui et al. 2011). eLORETA belongs to the class of minimum norm solvers; specifically, the estimated source time courses are given by

$$\tilde{\boldsymbol{z}}(t_p) = \arg\min_{\boldsymbol{z}(t_p)} \left\{ ||\boldsymbol{y}(t_p) - G\boldsymbol{z}(t_p)||^2 + \lambda\, \boldsymbol{z}(t_p)^T\, W\, \boldsymbol{z}(t_p) \right\} \tag{16}$$

$$\tilde{\boldsymbol{z}}(t_p) = W^{-1} G^T \left(G W^{-1} G^T + \lambda I \right)^{-1} \boldsymbol{y}(t_p) \tag{17}$$

where λ is a regularization parameter and W is a *weight matrix*, of size $3N_v \times 3N_v$, defined in order to obtain zero localization error in the case of point-test sources and ideal noise-less condition. Pascual-Marqui (2007) shows that this can be achieved defining W as a block-diagonal matrix

$$W = \begin{pmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & W_{N_v} \end{pmatrix} \tag{18}$$

where each block W_v, $v = 1, \dots, N_v$, has size 3×3 and satisfies

$$W_v^2 = G(\boldsymbol{r}_v)^T \left(G W^{-1} G^T + \lambda I \right)^{-1} G(\boldsymbol{r}_v) \tag{19}$$

$G(\boldsymbol{r}_v)$ being the leadfield matrix of the voxel centred in $\boldsymbol{r}_v$.

Data Simulation and Analysis Pipeline

Original Source Time Courses

We generate $N_d = 100$ dataset of length P = 10000 time points. Each dataset is composed by $N = 4$ signals, referred to as the *original* source time courses,

$$\boldsymbol{x}(t_p) = \left(x_1(t_p), x_2(t_p), x_3(t_p), x_4(t_p) \right)^T, \tag{20}$$

drawn from a stable MVAR process of order $K = 5$. The coefficient matrices of the process are defined in such a way that information flows only from $x_1(t_p)$ to $x_2(t_p)$ and to $x_3(t_p)$. Thus for each dataset we have

$$\begin{pmatrix} x_1(t_p) \\ x_2(t_p) \\ x_3(t_p) \\ x_4(t_p) \end{pmatrix} = \sum_{k=1}^K \begin{pmatrix} A_{11}(k) & 0 & 0 & 0 \\ A_{12}(k) & A_{22}(k) & 0 & 0 \\ A_{13}(k) & 0 & A_{33}(k) & 0 \\ 0 & 0 & 0 & A_{44}(k) \end{pmatrix} \begin{pmatrix} x_1(t_{p-k}) \\ x_2(t_{p-k}) \\ x_3(t_{p-k}) \\ x_4(t_{p-k}) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t_p) \\ \varepsilon_2(t_p) \\ \varepsilon_3(t_p) \\ \varepsilon_4(t_p) \end{pmatrix} \tag{21}$$

where the non zero coefficients, $A_{ij}(k)$, are drawn from a zero-mean normal distribution with variance of 0.01, and the innovation $\boldsymbol{\varepsilon}(t_p) = \left(\varepsilon_1(t_p),\ \varepsilon_2(t_p),\ \varepsilon_3(t_p),\ \varepsilon_4(t_p) \right)^T$ is drawn from a standard normal distribution ($\Sigma = I$).

The model in Eq. (21) is similar to the model proposed by Haufe et al. (2013),[2] the only difference being the presence of two more signals: (i) the third signal $x_3(t_p)$, that, like $x_2(t_p)$, is influenced by $x_1(t_p)$: the aim here is to have two non-directly connected brain areas with a common influence; (ii) the fourth signal $x_4(t_p)$, which is independent from the other three: the aim here is to have a control signal for detecting false positives.

As a first step, we investigate the impact of the data length on the estimates of the connectivity metrics computed from the original source time courses. While this is not of practical use, because the original time courses are seldom available in practice, it provides a sort of upper bound to the quality of the estimated values. To do this, we subdivide each dataset, $\{\boldsymbol{x}(t_p)\}_{p=1}^P$, in subsamples of increasing length.

We choose an increment of 500 time points for a total of 20 subsamples, so that the lth subsample contains $P_l = 500\, l$ time points. In practical terms, assuming a sampling frequency of 250 Hz, each dataset $\{\boldsymbol{x}(t_p)\}_{p=1}^P$ will consist of 40 s of data, the shortest subsample will contain 2 s of data and the lth subsample will contain $2\, l$ s of data.

For each subsample, and for each pair of signals $i, j = 1, \dots, N$, we estimate the value of the three connectivity measures (IC, gPDC and fGC) and generate surrogate data as described in "Statical Test" section; more specifically, we generate 100 realizations for each type of surrogate data, and choose a significance level $\alpha = 0.01$. As described in the previous section, we use these surrogate data to calculate the threshold of statistical significance corresponding to the α level.

Finally, for each dataset and for each subsample we compare the estimated connectivity measures with the corresponding thresholds: for each frequency, only the values that pass the threshold are considered significant.

---

[2] In order to simulate $\{\boldsymbol{x}(t_p)\}_{p=1}^P$ we modified the Matlab code available at http://neuroinformation.incf.org.
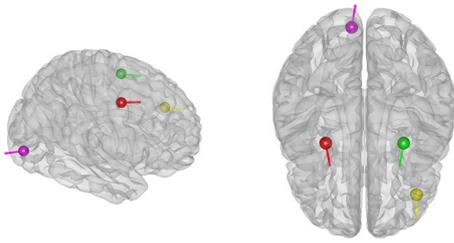
Fig. 1 Dipole locations. Left: lateral view from the right. Right: sagittal view from the top, front at the bottom. Colored points and vectors represent the positions and orientations of the dipoles used to simulate the scalp potential $\{y^{(i)}(t_p)\}_{p=1}^P$

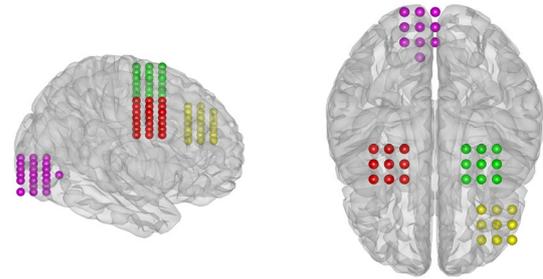**Fig. 2** Region of interests. Left: lateral view from the right. Right: sagittal view from the top, front at the bottom. Each ROI is defined as the set of voxels that are within 2 cm from the location of the corresponding dipole of interest

Estimated Source Time Courses

In order to investigate the impact of the EEG forward and inverse model on the connectivity estimates, we simulate EEG datasets as follows. For each dataset of original source time courses $\{x(t_p)\}_{p=1}^P$, we interpret the four signals as the time courses of four dipolar sources that mimic the connectivity pattern observed in the stimulus driven control network (Corbetta and Shulman 2002). In this network, a directed interaction between primary visual cortex and bilateral frontal areas has been reported. In our simulation dipole locations and orientations are set as in Fig. 1: the occipital purple source is located in the visual cortex and its time course is $\{x_1(t_p)\}_{p=1}^P$; the green and red dipoles are placed in the proximity of the left and right frontal eye fields, being their time courses $\{x_2(t_p)\}_{p=1}^P$ and $\{x_3(t_p)\}_{p=1}^P$, respectively. An additional frontal source (yellow) models frontal activity uncoupled with the above system.

We compute the scalp potential $y^{(i)}(t_p)$ generated by these four sources according to Nolte and Dassios (2005): the forward model is based on three prolate spheroid shells that model a realistically shaped three-compartment (brain, skull and skin) head obtained from Holmes et al. (1998). A nose reference is used for the leadfield calculation (Haufe et al. 2013). The presence of both *biological* and *sensor* noise is then simulated as follows.

Sensor noise $\eta(t_p)$ is modelled by a standard normal distribution with fixed variance. Biological noise $y^{(b)}(t_p)$ is simulated as the scalp potential generated by ten dipolar sources of random locations and orientations, and whose time courses follow each one a unidimensional autoregressive model[3] of order ten. The three contributions are scaled

in order to control the signal to noise ratio (SNR): the final simulated EEG data is given by

$$y(t_p) = (1 - \gamma_1 - \gamma_2) \frac{y^{(i)}(t_p)}{||y^{(i)}||} + \gamma_1 \frac{y^{(b)}(t_p)}{||y^{(b)}||} + \gamma_2 \frac{\eta(t_p)}{||\eta||} \qquad (22)$$

where γ_1 and γ_2 weight the contribution of biological noise and sensor noise, respectively, and, given a time series $\{z(t_p)\}_{p=1}^P$, $||z||$ is the Frobenius norm of the matrix $\left(z(t_1), \ldots, z(t_P)\right)^T$. In the simulations below we set

$$\gamma_2 = \frac{1}{8} \qquad \gamma_1 \in \left\{ 0, \frac{1}{4}(1 - \gamma_2), \frac{1}{2}(1 - \gamma_2) \right\}, \qquad (23)$$

i.e. we consider three different noise levels.

From each scalp potential we reconstruct the brain activity $\{\tilde{z}(t_p)\}_{p=1}^P$ using eLORETA, Eq. (17). In particular, the inverse operator is computed with the same leadfield matrix used to generate the data, and setting the regularization parameter equal to the variance of the simulated sensor noise, i.e.

$$\lambda = \left(\frac{\gamma_2}{||\eta||} \right)^2. \qquad (24)$$

Before studying connectivity we reduce the dimensionality of the problem by a two-step procedure. First we define four regions of interest (ROIs), each one containing the voxels that are within 2 cm from one of the sources of interest (Fig. 2); the activity of each ROI is estimated by summing up the activity of its voxels, thus reducing the overall dimension from $3N_v$ to 12. Then we use principal component analysis to project the activity of each area onto the direction of maximum power; we thus end up with four time series

$$\tilde{x}(t_p) = \left(\tilde{x}_1(t_p), \tilde{x}_2(t_p), \tilde{x}_3(t_p), \tilde{x}_4(t_p)\right)^T, \quad p = 1, \ldots, P \qquad (25)$$

that estimate the original time courses $\{x(t_p)\}_{p=1}^P$.

---

[3] We observe that either the frontal yellow dipole or the ten sources just described are not connected with anyone of the other dipoles of the networks. However, while the latter are used to simulate biological noise the former contributes to the scalp potential of interest and is used to evaluate the sensitivity of each connectivity measure to false positive.

To perform connectivity analysis from $\{\tilde{x}(t_p)\}_{p=1}^{P}$, we proceed analogously to what we have done with the original time courses: we subdivide the reconstructed activity in subsamples of increasing length, with the same lengths described in the previous paragraph. The values of the three connectivity measures, IC, gPDC and fGC, are estimated for each pair of reconstructed time courses, $\tilde{x}_i(t_p), \tilde{x}_j(t_p)$, $i, j = 1, \ldots, N$, and their statistical significance is tested making use of PR and AR surrogate data.

Evaluation of the Performances

Since the original source time courses are drawn from known MVAR processes, we can compute the *theoretical values* of the connectivity measures under investigation. Indeed, first the transfer matrix $H(f_q)$ and the cross-spectral density function $S(f_q)$ can be computed from the model parameters through Eqs. (8) and (10). Then IC, gPDC and fGC are given by Eqs. (4), (6) and (9), respectively.

In the following simulations, these theoretical values are used to evaluate and compare the results obtained performing connectivity analysis from the subsamples of different

length extracted either from the original datasets, $\{x(t_p)\}_{p=1}^{P}$, or from the reconstructed ROIs activity, $\{\tilde{x}(t_p)\}_{p=1}^{P}$.

First, we count how many estimated connectivity values pass the threshold of statistical significance. We do this for both the connected source pairs and for the independent source pairs. For connected source pairs we provide histograms, where the counts are grouped based on the theoretical value of the connectivity measure. For independent source pairs we provide the ratio of false positives.

Then, we quantify the impact of the data length and (only for connectivity values computed from the estimated time courses) of biological noise by means of the correlation coefficient between the theoretical and the estimated values, as functions of frequency. More precisely we defined the correlation distance

$$\delta_{ij} := 1 - \frac{\sum_{q=1}^{Q} \left(M_{ij}(f_q) - \overline{M}_{ij} \right)\left(\tilde{M}_{ij}(f_q) - \overline{\tilde{M}}_{ij} \right)}{\sqrt{\sum_{q=1}^{Q} \left(M_{ij}(f_q) - \overline{M}_{ij} \right)^2 \sum_{q=1}^{Q} \left(\tilde{M}_{ij}(f_q) - \overline{\tilde{M}}_{ij} \right)^2}}$$

(26)

where $M_{ij}(f_q)$ and $\tilde{M}_{ij}(f_q)$ are the theoretical and estimated values of one of the connectivity measures between the pair of signals $\left(x_i(t_p), x_j(t_p) \right)$, $\overline{M}_{ij} = \sum_{q=1}^{Q} M_{ij}(f_q)$ and
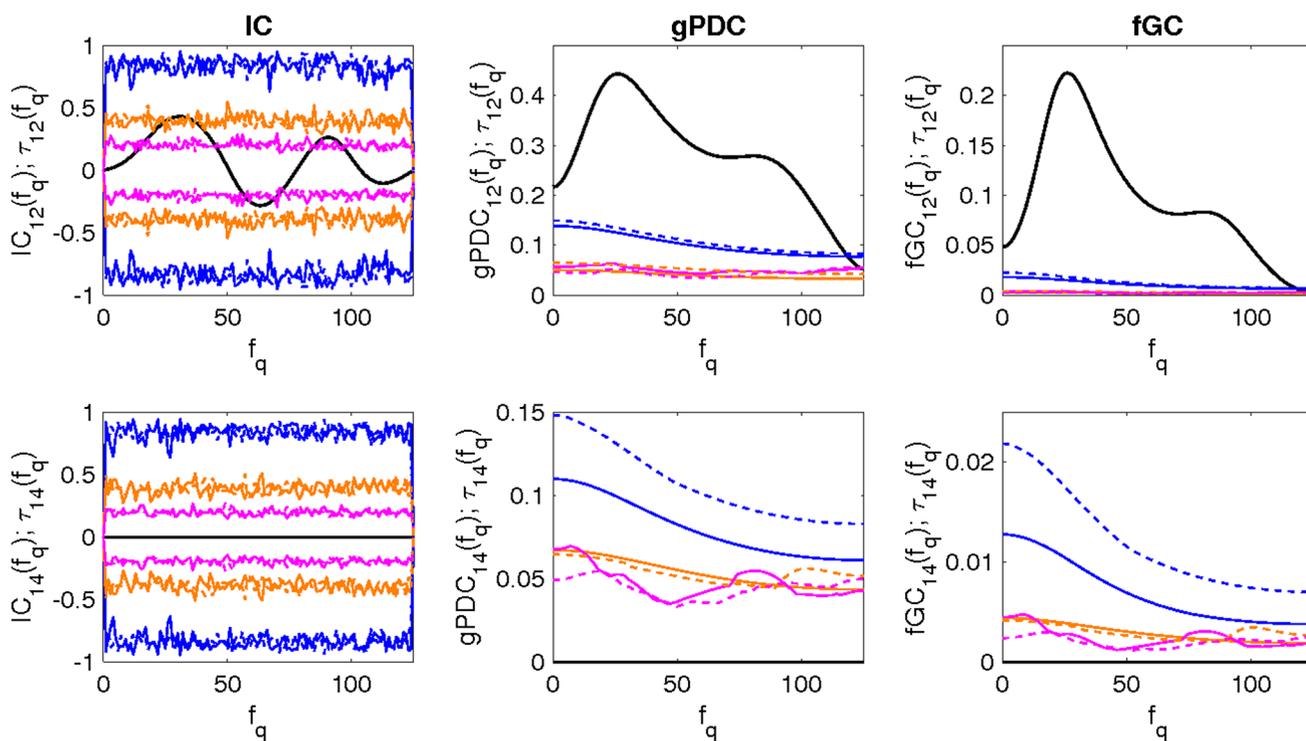
Fig. 3 Theoretical values and thresholds for a single dataset. For each connectivity measure we plot the theoretical value as a function of the frequency (black line) from $x_1(t_p)$ to $x_2(t_p)$ (first row) and from $x_1(t_p)$ to $x_4(t_p)$ (second row, independent signals). Then we superimpose the value of the threshold obtained from the sub-samples of length 500 (blue), 2500 (orange) and 10000 (magenta) time points. Dashed and solid lines are the thresholds obtained with PR and AR surrogate data, respectively. Notice the different scales on the y axes

$\overline{\tilde{M}_{ij}} = \sum_{q=1}^{Q} \tilde{M}_{ij}(f_q)$. We observe that $0 \leq \delta_{ij} \leq 2$ and $\delta_{ij} = 0$ when $\tilde{M}_{ij}(f_q) = \alpha M_{ij}(f_q)$ $(\alpha > 0)$, i.e. when the empirical values correctly reproduce the behaviour over frequencies of the theoretical values.

Results

Threshold as a Function of the Data Length

In Fig. 3, we exemplify the behaviour of the threshold for statistical significance. The results here are obtained from a single realization of the original time courses, for illustrative purposes. For each connectivity measure we show two panels, one for the pair $(x_1(t_p), x_2(t_p))$ that has non-zero connectivity, one for the pair of independent signals $(x_1(t_p), x_4(t_p))$. The results from the other signal pairs are almost identical. In each panel we plot the theoretical values of the connectivity measure and the thresholds produced by PR and AR surrogate data as functions of the frequency. Different colors correspond to different sub-sample lengths.

As expected, the value of the threshold decreases as the number of samples in the time series increases, suggesting that lower connectivity values can be reliably estimated from longer time series. The two types of surrogate data produce almost identical thresholds for IC; in addition, the threshold values of IC appear to be frequency-independent. For gPDC and fGC the behaviour is quite different: the threshold do vary with the frequency, and the two types of surrogate data produce different thresholds, although there does not seem to be any systematic difference. Finally, the picture suggests that IC can hardly be estimated from the shortest sub-sample considered in this study, made of 500 time points, where the threshold is very close to 1; one can be more optimistic for gPDC and fGC, where the thresholds remain lower than the true value even for the shortest sub-sample.
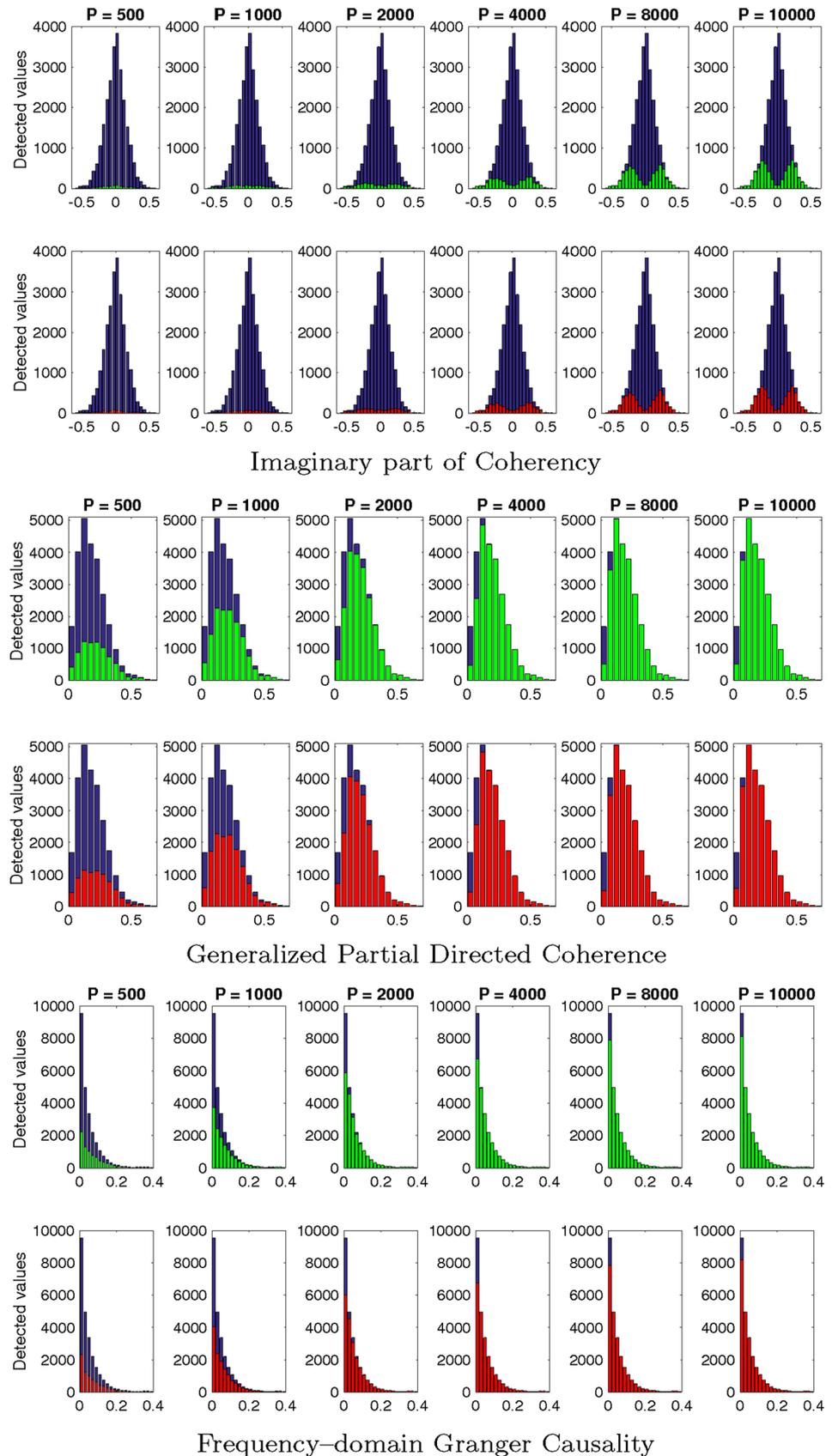
To show that these behaviours do not depend on the specific choice of the source time courses, in Fig. 4 we plot mean and standard deviation of the threshold across the 100 datasets, as functions of the data length. This is in fact the result of a double averaging procedure: first, for each dataset we average across different frequencies; then, we average across the 100 datasets. As before, for each connectivity measure we show two panels, corresponding to the pair $(x_1(t_p), x_2(t_p))$ and $(x_1(t_p), x_4(t_p))$, respectively. The plots confirm that the thresholds decrease as the data length increases and that there is no systematic difference between PR and AR surrogate data. Moreover they show an almost zero variance for IC and a decreasing variance for gPDC and
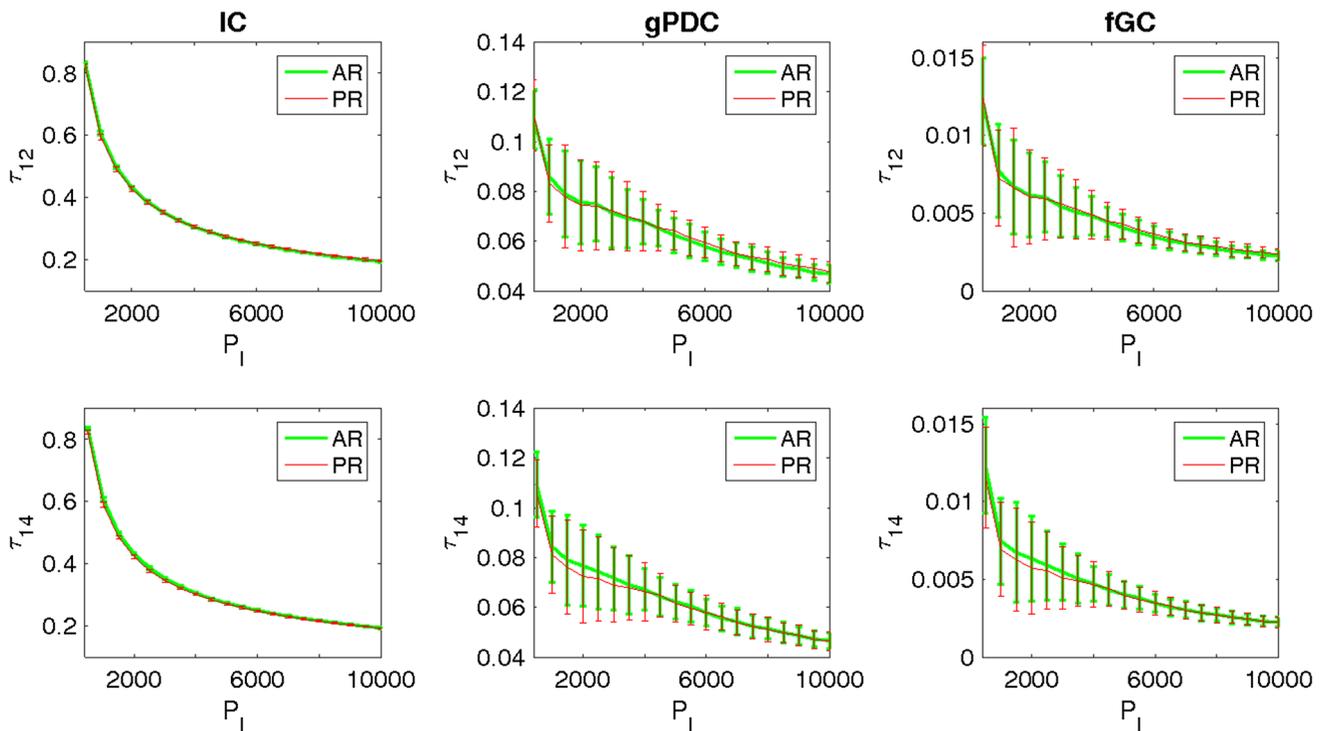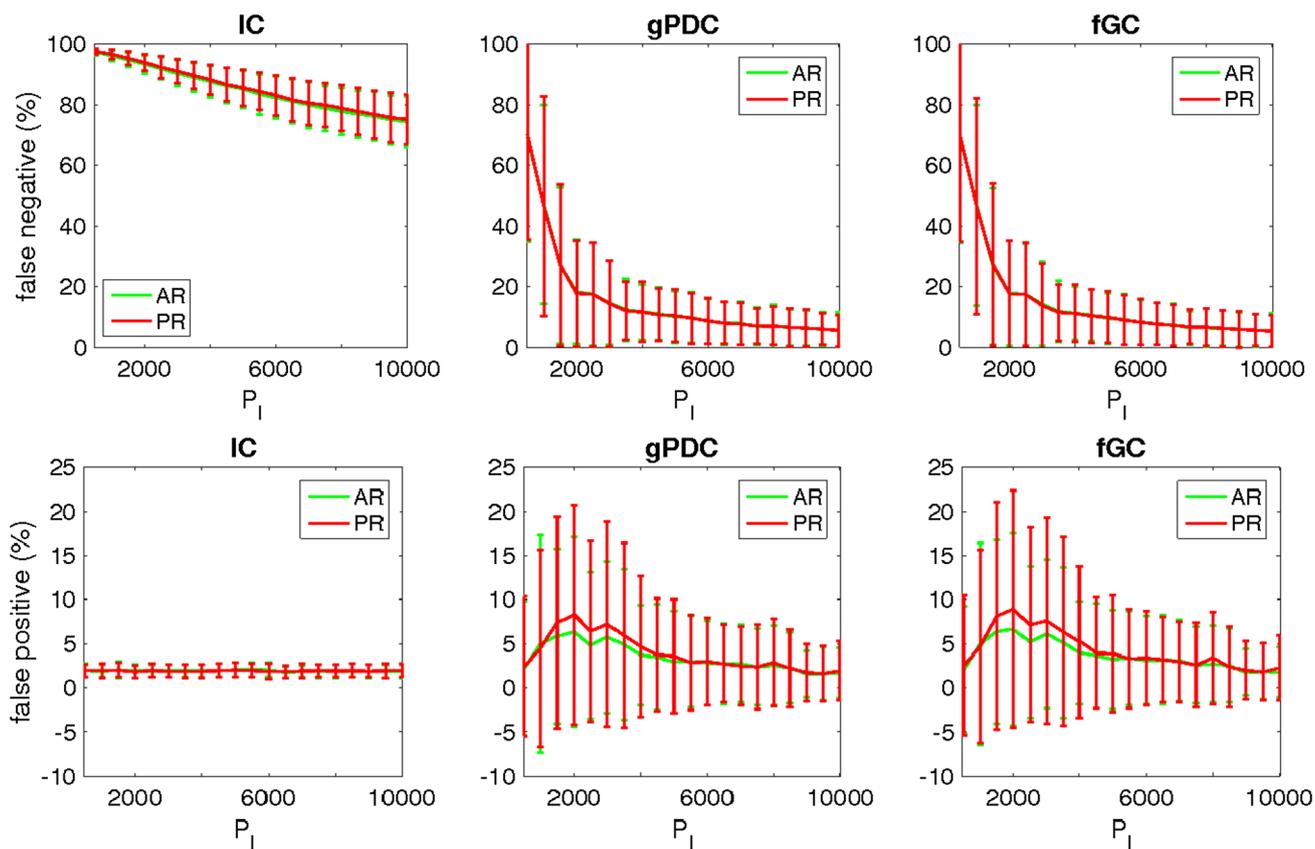

Fig. 4 The threshold as a function of the length of the data: mean and standard deviation over frequencies and over test dataset for the pair $(x_1(t_p), x_2(t_p))$ (first row) and for the pair $(x_1(t_p), x_4(t_p))$ (second row).

Thresholds have been obtained with PR (red line) and AR (green line) surrogate data

Fig. 5 True (blue) values summed over different frequencies and datasets and over the two pairs $(x_1(t_p), x_2(t_p))$ and $(x_1(t_p), x_3(t_p))$. Different columns correspond to different sub-sample lengths, namely [500, 1000, 2000, 4000, 8000 10000] time points. Green and red bars represent the number of estimated values passing the statistical test with AR and PR surrogate data, respectively

Imaginary part of Coherency

Generalized Partial Directed Coherence

Frequency–domain Granger Causality

Fig. 6 Top/bottom: mean and standard deviation over datasets of the total percentage of false negatives/positives detected between the pairs of signals $(x_1(t_p), x_2(t_p))$ and $(x_1(t_p), x_3(t_p))$ / $(x_1(t_p), x_4(t_p))$, $(x_2(t_p), x_4(t_p))$ and $(x_3(t_p), x_4(t_p))$ as a function of the sub-sample length. The tresholds of statistical significance have been computed making use of AR (green line) and PR (red line) surrogate data. Each column correspond to a different connectivity measure
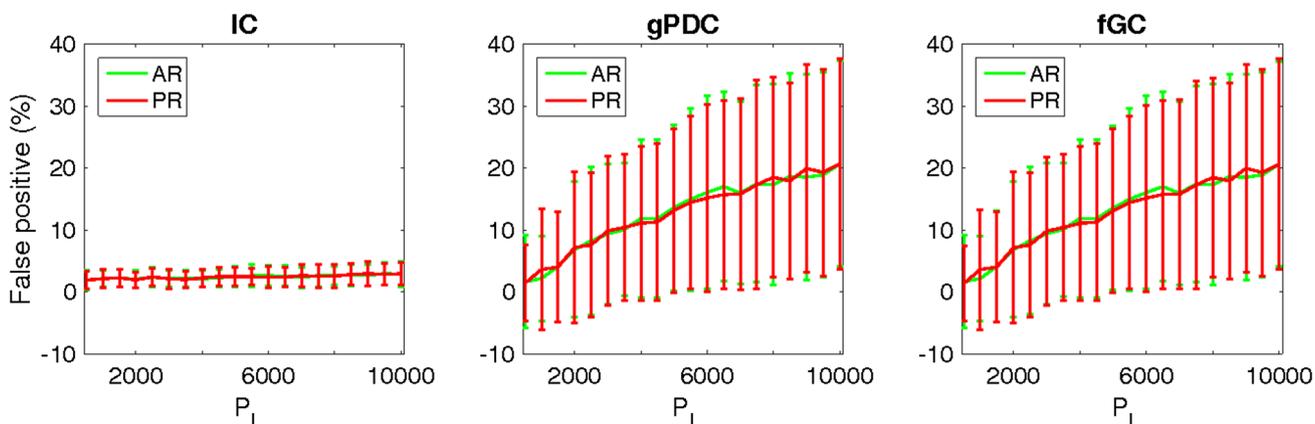

Fig. 7 Mean and standard deviation over the dataset of the total percentage of false positives between the order pair of signals $(x_2(t_p), x_3(t_p))$ and $(x_3(t_p), x_2(t_p))$ when the connectivity analysis is performed without considering the signals $x_1(t_p)$ and $x_4(t_p)$

fGC. We notice that the higher variance of the thresholds for gPDC and fGC, compared to those for IC, is partly a consequence of the fact that they are frequency-dependent; in this sens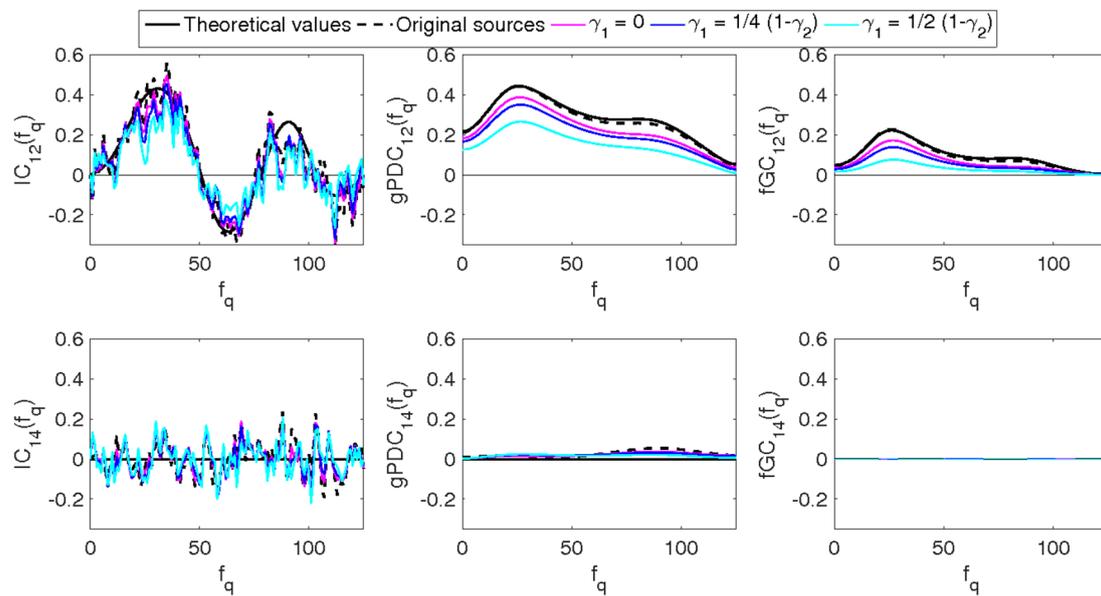e, such higher variability is not necessarily a limitation, but may reflect a higher capability of capturing the main features of the original signals. In order to investigate these aspects, in the following section these thresholds are used
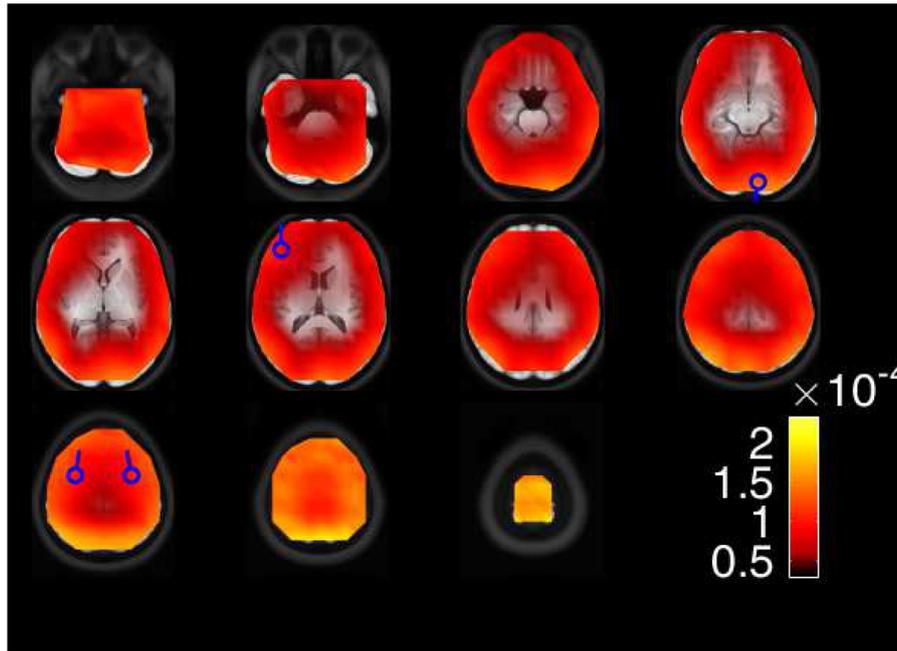
Step 1: source reconstruction
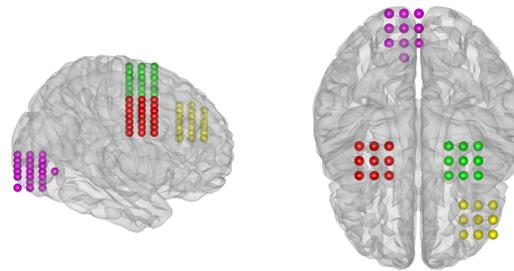
Step 2: dimensionality reduction

Step 3: estimation of the connectivity measures

◄ **Fig. 8** Fixed one of the original source time courses $\{\boldsymbol{x}(t_p)\}_{p=1}^{P}$, in the top panel, we plot the voxelwise amplitude averaged over time of the neural sources reconstructed from the EEG data generated setting $\gamma_1 = 1/4(1 - \gamma_2)$, where $\gamma_2 = 1/8$. Blue circles and lines represent locations and orientations of the dipoles of interest. In the second panel, we plot the ROIs defined to reduce the number of time series used for the connectivity analysis. Eventually, in the last two rows we show the estimated connectivity measures for two pairs of ROIs. In each panel we plot the theoretical values of the measure (black line) and the values estimated from the original source time courses (dotted line) and from the reconstructed source activities obtained from EEG data simulated with different values of γ_1 (coloured lines). We observe that when there is no connectivity the computed values of fGC are close to zero in all the conditions (bottom right panel)

to assess the statistical significance of the empirical values of each connectivity measure.

Connectivity Analysis from the Original Source Time Courses

In this section we aim to investigate the capability of the different measures to detect the correct functional relationship among signals. As described in "Data Simulation and Analysis Pipeline" section, for each sub-sample extracted from each dataset we estimate the empirical values of the connectivity measures and we assess their statistical significance by means of the threshold previously defined and analysed.

In Figure 5 we quantify the probability of correctly identifying the presence of connectivity as a function of the theoretical strength of such connectivity.

Indeed, we consider the (ordered) pairs of signals $(x_1(t_p), x_2(t_p))$ and $(x_1(t_p), x_3(t_p))$, which have non-null connectivity; for each measure, we compute the histogram of the theoretical connectivity values for all frequencies, summed up for all datasets and for the two pairs of signals; then, for each bar of such histogram we count how many of the corresponding empirical values pass the statistical test when the threshold is computed by using AR (green bar) or PR (red bar) surrogate data. In the figure we plot the results concerning the sub-samples of length [500, 1000, 2000, 4000, 8000, 10000] time points. The plots confirm that IC is the most "conservative" metric, i.e. IC values pass the threshold much more rarely than gPDC and fGC values. However, we also observe that, while above $P = 2000$ the most of the high connectivity values are correctly recognised, below $P = 2000$ the estimated values appear to pass the threshold more randomly, i.e. independently on the true underlying connectivity value. The results for gPDC and fGC are quite different: in fact, most estimated values pass the statistical test even at $P = 2000$, and a large portion of them seems significant even at $P = 500$.

In order to quantify the differences between the three measures, in Figure 6 we plot the percentage of false

negatives as a function of the data length. This percentage is computed as follows. At each frequency a false negative is considered to occur if the empirical value of the connectivity measure does not pass the statistical test, even though the corresponding theoretical value is non-zero. Then we sum up the number of false negatives over frequencies and over the two pairs of signals $(x_1(t_p), x_2(t_p))$ and $(x_1(t_p), x_3(t_p))$. The Figure confirms our previous considerations. Indeed, for all the connectivity measures the number of false negatives decreases for increasing data length. Moreover, the false negative percentage for IC is systematically higher.

Clearly, these results feature a higher sensitivity of gPDC and fGC, i.e. a higher capability of the two connectivity measures to recognize the presence of connectivity, which in turn may also cause a higher number of false positives. To investigate this aspect, we evaluate the false positive ratio by making use of the pairs of unconnected signals $(x_1(t_p), x_4(t_p))$, $(x_2(t_p), x_4(t_p))$ and $(x_3(t_p), x_4(t_p))$. In the second row of Fig. 6, for each connectivity measure we plot the mean and the standard deviation over datasets of the percentage of false positives as a function of the sub-sample length. This percentage is computed by summing up the number of empirical values that pass the statistical test when the connectivity measures are computed between the three aforementioned pairs of signals. Please notice that, here and in figures below, the standard deviation has been plotted symmetrically with respect to the mean, though clearly the false positive ratio cannot assume negative values. The figure shows that for each connectivity measure the results obtained with either PR or AR surrogate data are almost identical. On the other hand, gPDC and fGC seem to be much more affected by false positives than IC. Indeed, while for IC the percentage of false positives remains very low for all the sub-sample lengths, for gPDC and fGC such percentage, and in particular the standard deviation across dataset, is considerably higher; while such percentage decreases as the sub-sample length increases, it never gets quite as low as that of IC.

Common Input Problem

As described in "Data Simulation and Analysis Pipeline" section, the original source time courses are simulated in such a way that there is no direct connection between $x_2(t_p)$ and $x_3(t_p)$; however, both are influenced by $x_1(t_p)$, the common input, and this induces some correlation between $x_2(t_p)$ and $x_3(t_p)$. The three connectivity metrics behave differently in this respect. Since IC is a bivariate measure, there is no way to explicitly account for the presence of the common input in the estimation process. As a consequence, the theoretical value of $IC_{23}(f_q)$ is non-zero; but, since the two sources are not actually connected, all the estimated values passing the threshold are to be regarded as false positives. In

Fig. 7, left panel, we plot the false positive rate for the pair $(x_2(t_p), x_3(t_p))$: the values are somewhat larger with respect to those of purely independent signals (cfr. Fig. 6), but not dramatically so. For gPDC and fGC, instead, there are two possibilities depending on whether the source $x_1(t_p)$ is taken into account or missed, due to, e.g., poor visibility in the data, as in the case of a radial source in MEG. If the common input is part of the MVAR model that is used in the estimation process, then the rate of false positives for the pair $(x_2(t_p), x_3(t_p))$ is the same of that for independent signals, depicted in Fig. 6. If, however, $x_1(t_p)$ is not part of the MVAR model, gPDC and fGC detect a spurious connectivity from $x_2(t_p)$ to $x_3(t_p)$ or vice versa, and the rate of false positives increases substantially. In Fig. 7 we plot the rate of false positives obtained by neglecting the common input, i.e., by applying the analysis pipeline (MVAR model estimation, connectivity estimation, surrogate data generation) to the pair $(x_2(t_p), x_3(t_p))$. Remarkably, the rate of false positives increases with the data length, because the spurious connectivity becomes more and more evident in the data.

Connectivity Analysis from the Estimated Source Time Courses

The results of the previous subsection are to be considered as "optimal", in the sense that they have been obtained from the original time series. We now proceed to investigate the effect of the EEG forward/inverse model, using eLORETA, and of biological noise. In Fig. 8, we summarize the analysis pipeline for connectivity analysis from the synthetic EEG data. In the last row of the same figure we exemplify the effect of biological noise on the connectivity estimates, when the connectivity values are estimated from the longest sub-sample. The values of gPDC and fGC are systematically and increasingly under-estimated and over-smoothed: the two bumps of the original profile become less and less visible. Conversely IC seems to be less influenced by the presence of biological noise: for the lower γ_1 also the estimated IC values are slightly underestimated but still similar to those estimated from the original source time courses (black dotted line). We will quantify this similarity later by making use of the covariance distance δ_{ij}.

In Figs. 9, 10, 11 we show the histograms of statistically significant values,[4] grouped according to the true value of the underlying connectivity; different rows correspond to different levels of biological noise. A comparison with the plots in Fig. 5 suggests that the effect of increasing the noise level is similar to that of shortening the data length. For example,

the results with $P = 10000$ with measurement noise only (top row, right panel of Figs. 9, 10, 11) are similar to those from the original time courses with $P = 8000$; results for high biological noise (bottom row) with $P = 10000$ are similar to those for measurement noise only (top row) with $P = 4000$.

In Fig. 12, we show the false negative and false positive rate for different values of the SNR. For the sake of comparison, we also superimpose the results from the original time courses. Expectedly, the false negative rate increases with increasing biological noise for all the connectivity measures. More interestingly, the plots obtained from the reconstructed sources show a higher number of false positives at increasing data length only for gPDC and fGC. These results seem to confirm that the effect of the biological noise we simulated is not qualitatively different from the effect of having less data.

Finally, in Fig. 13 we plot the mean and standard deviation over test dataset of the correlation distance δ_{12} defined in Eq. (26), for the pair of correlated sources $(x_1(t_p), x_2(t_p))$. We recall that this correlation distance does not depend on the significance of the estimated values; instead, it is an attempt at quantifying how much of the true connectivity pattern, as a function of frequency, is retained in the estimated connectivity pattern. Here we show the values of the correlation distance for connectivity estimated from the original time courses (black dotted line) and from the estimated time courses (coloured lines). The plots for IC are easier to interpret. The effect of increasing the data length is a monotonic decrease of the correlation distance; noise, on the other hand, monotonically increases such distance. For gPDC and fGC, the connectivity estimates from the original time courses become quickly fairly good, as the data length increases. However, the EEG forward/inverse models and biological noise disrupt such good behaviour more dramatically than for IC. The plots suggest that gPDC and fGC are still preferable for longer data lengths, where they have smaller variance and smaller/comparable correlation distance. The case of high biological noise, however, provides rather similar performances for all three metrics.

Discussion

In the present work we studied the impact of the length of the time series on the estimates of connectivity from EEG data, using three frequency-domain connectivity measures, IC, gPDC and fGC, and numerical simulations. We first studied the ideal case, in which the source time courses are known exactly, to provide a reference value and isolate the effects due solely to the finiteness of the data. Then we went on to consider the case where connectivity is computed from source time courses that are, in turn, estimated from EEG time series, possibly affected by biological noise. We

---

[4] We show only the results obtained when the statistical test is performed using AR surrogate data. The results with PR surrogate data are almost equal.

quantified the impact of the various elements in multiple ways: by assessing the significance of the estimated connectivity values, using surrogate data; by computing the false negative and false positive rates; by computing a correlation distance with the theoretical connectivity value, which is known exactly.

As mentioned in the introduction, there are few studies aimed at quantifying the impact of the data length on the quality of the connectivity estimates. However, to the best of our knowledge, this is the first study that takes into account

the effect of the EEG forward/inverse model and of biological noise, and compares connectivity from the original time courses with connectivity from the estimated time courses.

Our numerical results from the exact time courses confirm most expectations concerning qualities and drawbacks of the connectivity metrics under investigation. Specifically, all the simulations confirm that IC is much more "conservative" than gPDC and fGC, meaning that the estimated IC values are less likely to pass the threshold of statistical significance, compared to the estimated gPDC and fGC values.
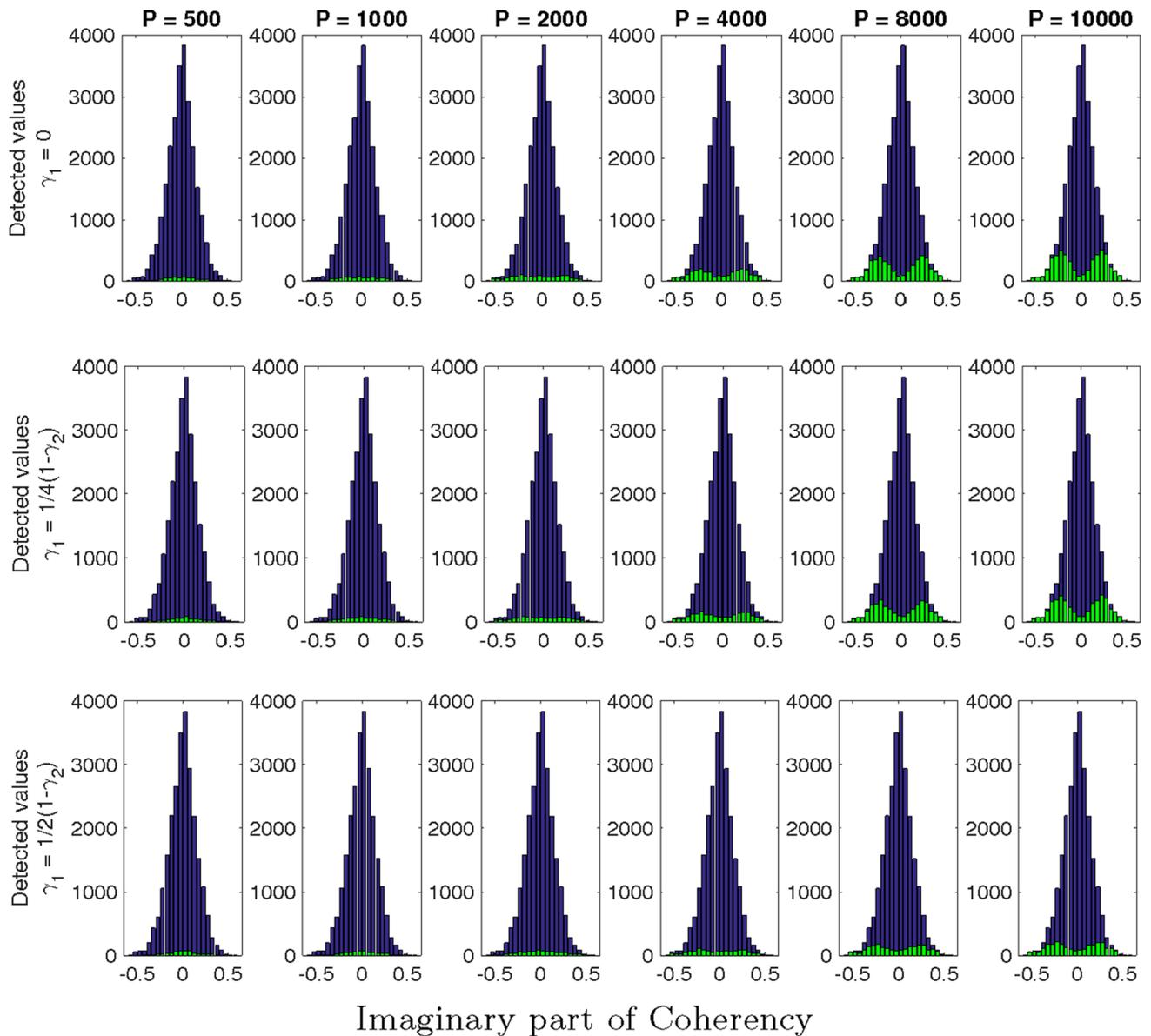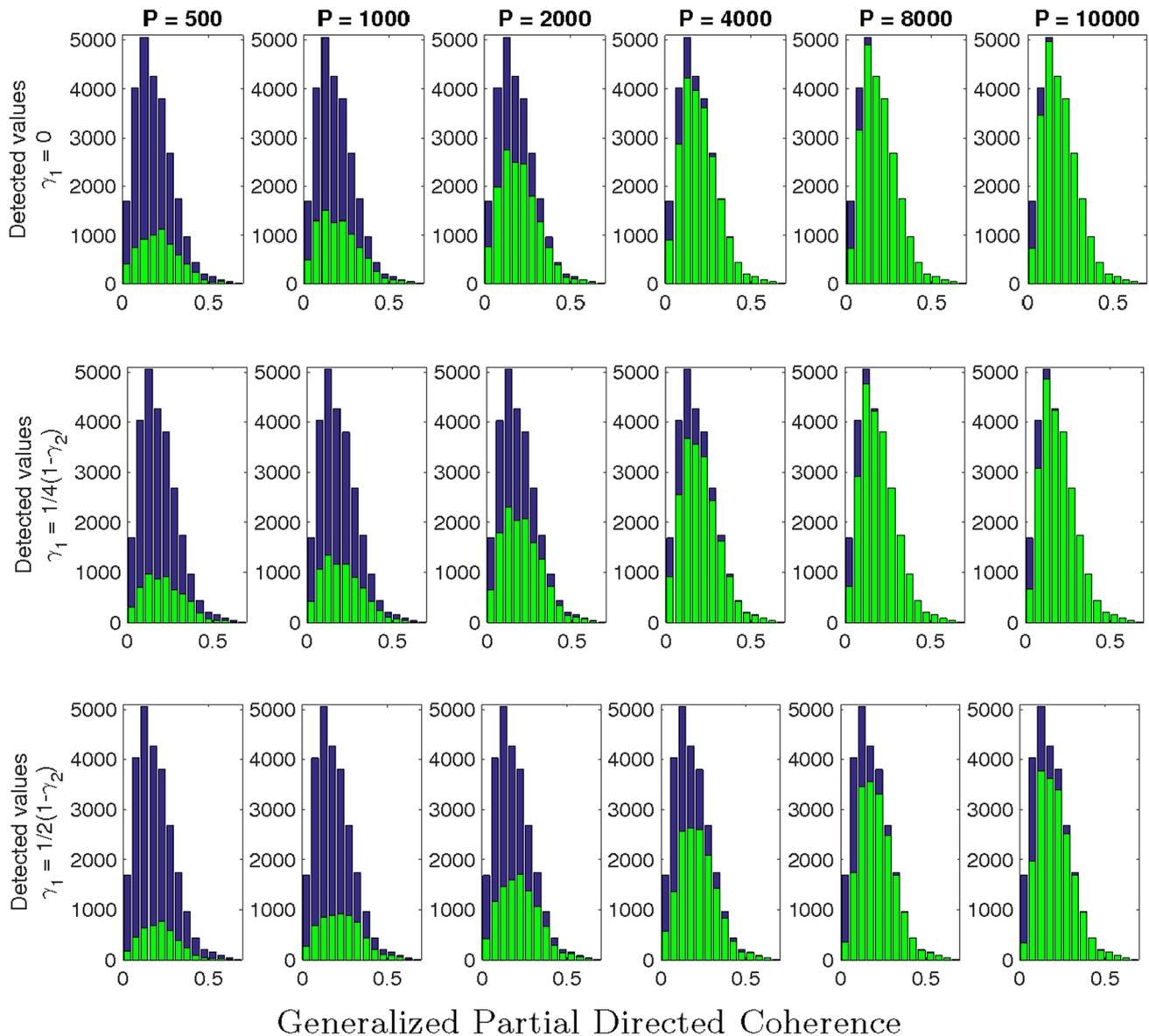

Fig. 9 Histograms of the statistically significant values (green bar) grouped according to the theoretical values (blue bars), summed over different frequencies and datasets and over the two pairs of time series $(\tilde{x}_1(t_p), \tilde{x}_2(t_p))$ and $(\tilde{x}_1(t_p), \tilde{x}_3(t_p))$ reconstructed from EEG data with different levels of biological noise. Different columns correspond to different sub-sample lengths, namely [500, 1000, 2000, 4000, 8000 10000] time points. The statistical test is performed by means of AR surrogate data

In fact, the empirical values of IC tend to pass the statistical test only in correspondence of the greater theoretical values, providing a systematically higher number of false negatives. On the other hand, estimated gPDC and fGC values pass the threshold very often and tend to be more affected by false positives. In general, it can be said that these two model-based measures tend to provide more accurate values than IC when the MVAR model is correctly estimated, but are more prone to errors when it is not. This is particularly manifest in the case of a common input: if the common input is correctly addressed in the model, gPDC and fGC provide better results than IC, which is a bi-variate measure; if it is not, results from gPDC and fGC are less reliable. This issue might be particularly relevant in MEG studies, where radial sources might not show up in the sensor level data.

The numerical results from the estimated time courses clearly indicate that the effects of the EEG forward/inverse plus measurement noise, and of biological noise are not negligible. Even in the absence of biological noise, the results obtained from the estimated time courses are worse than those obtained from the exact time courses; in particular, they appear similar to those obtained from the exact time courses but with shorter data. This is reasonable because measurement noise lowers the signal-to-noise ratio without disrupting connectivity. Additionally, the use of ROIs (centred in the correct dipole locations) in place of pointlike

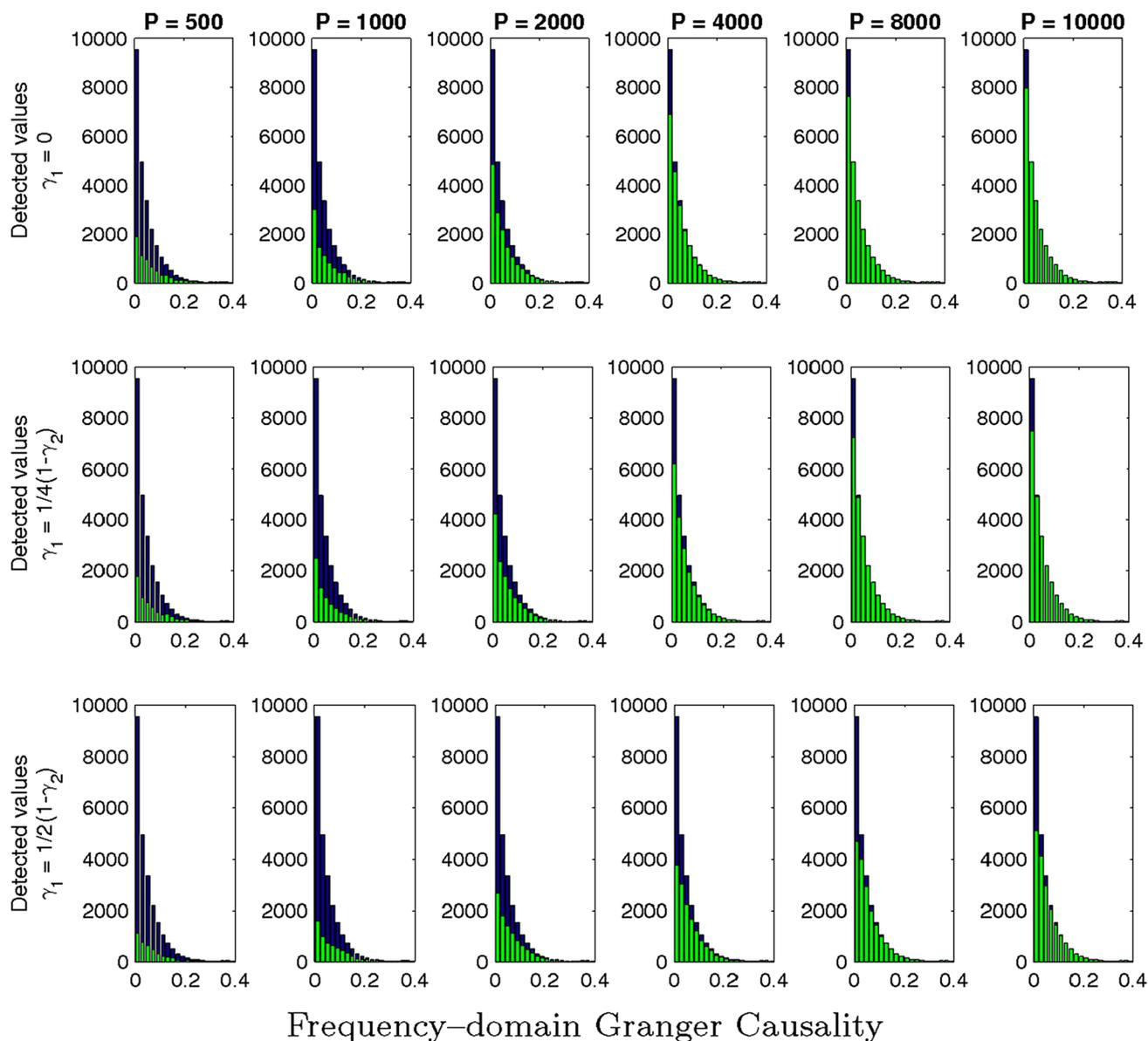**Fig. 10** Histograms of the statistically significant values obtained as described in Fig. 9

sources only partially compensates for the lowering of the signal-to-noise ratio in the reconstructed source induced by the spatial dispersion of the forward/inverse procedure.

The effect of biological noise does not seem to be qualitatively different, i.e. mostly has the same effect of having shorter time series, but is quantitatively relevant, yielding substantial increased false negative and false positive rates; the only exception here appears to be the false positive rate of IC, which remains systematically low.

Our results seem to indicate that explicitly modelling the forward/inverse procedure is key for a quantitative assessment of the reliability of connectivity measures. We speculate that this is particularly relevant when comparing data-driven and model-based measures. Indeed, in our simulations the results obtained from the estimated sources and those obtained from the original sources appear to be more diverse for gPDC and fGC than for IC. Comparative studies that do not explicitly model the forward/inverse procedure might thus be prone to misleading conclusions.

Our results are of course limited by the type of simulations we performed, and might still be optimistic. For instance, biological noise is simulated as the activity of ten independent sources, and the regularization parameter in eLORETA has been chosen based on the knowledge of the noise variance. Even more so, our results show that the impact of the EEG forward/inverse and of biological noise

Fig. 11 Histograms of the statistically significant values obtained as described in Fig. 9

Fig. 12 First/last three rows: mean and standard deviation over the 100 original data-sets of the total percentage of false negatives/positives detected between the pairs of signals $(\tilde{x}_1(t_p), \tilde{x}_2(t_p))$ and $(\tilde{x}_1(t_p), \tilde{x}_3(t_p))/(\tilde{x}_1(t_p), \tilde{x}_4(t_p))$, $(\tilde{x}_2(t_p), \tilde{x}_4(t_p))$ and $(\tilde{x}_3(t_p), \tilde{x}_4(t_p))$. In each panel we compare the results obtained from the original source time courses with those from the ROIs activity reconstructed from EEG data with different levels of biological noise

Percentage of false negatives

Percentage of false positives
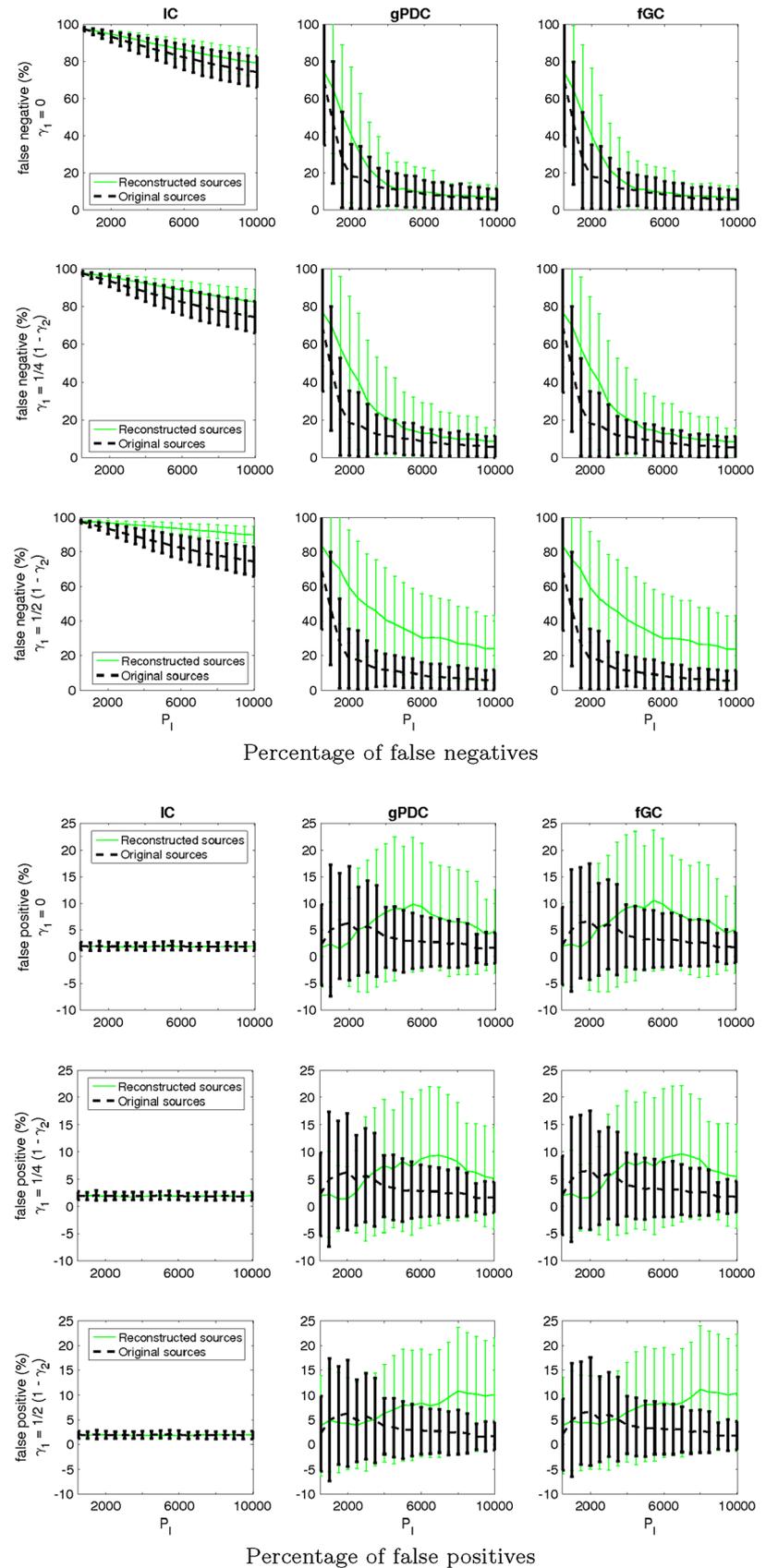
should be taken into account when one wants to quantitatively assess the sensitivity of the connectivity metrics to the data length; if not, results are overly optimistic.

Conclusions

The great advantage of EEG and MEG in the study of brain connectivity resides mainly in their temporal resolution, that allows, in principle, monitoring of dynamic changes in the connectivity patterns. However, this requires connectivity to be estimated from short time windows. This study
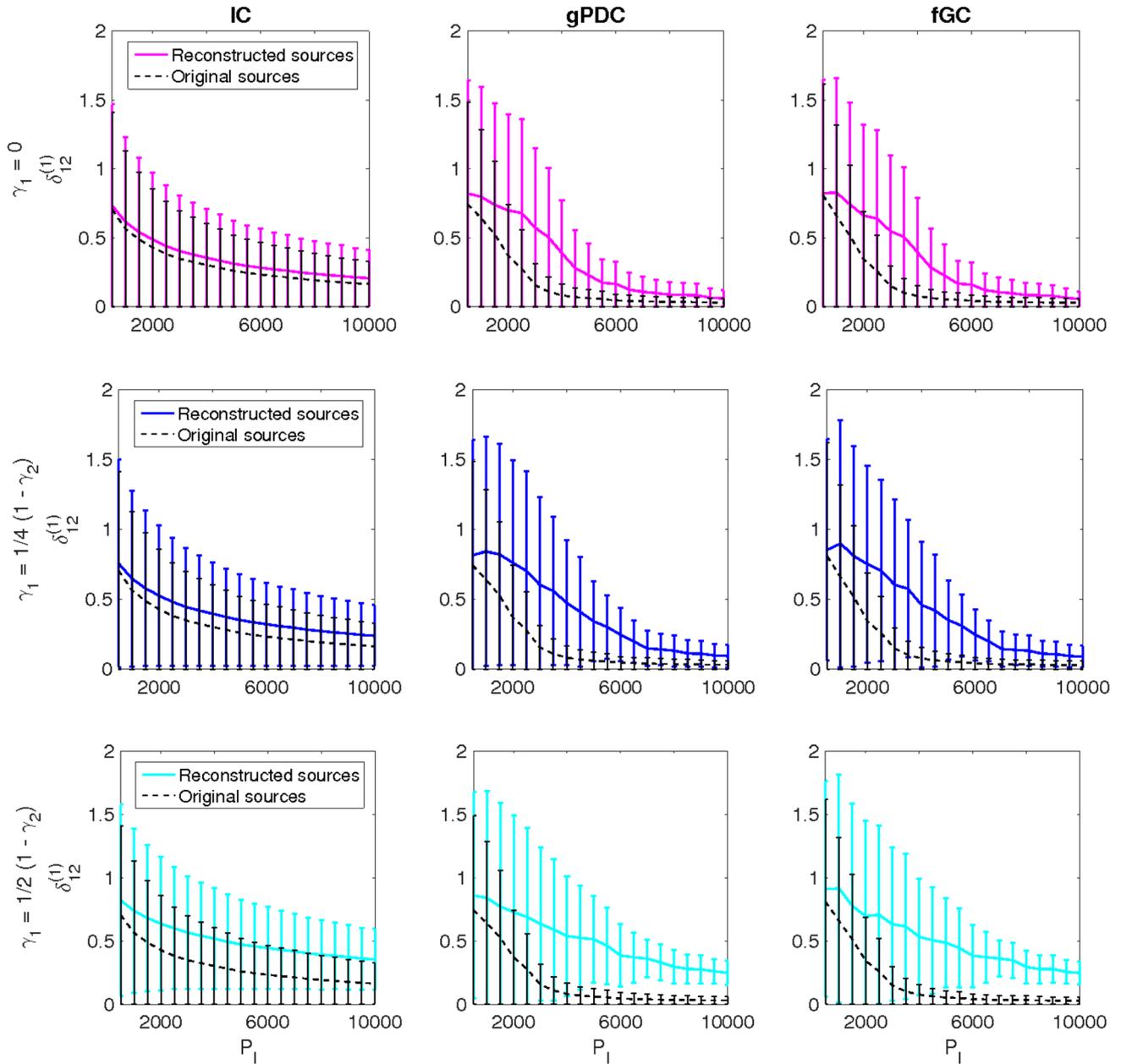
Fig. 13 Correlation coefficient between the true and the estimated values of the connectivity measures. In each panel we plot the mean and standard deviation over the original datasets of δ_{12} as a function of the sub-samples length. For each panel the black dotted line represents the correlation coefficient between the true values of the con-

nectivity measure and the empirical values estimated from the true source time courses whereas coloured lines are obtained by estimating the connectivity measured from the reconstructed ROIs activity. Each row concerns a different level of biological noise

confirms that such estimation is possible, but with some caution, because using less data makes the results more prone to errors. In addition, some contributions to error, like neglecting a common input in the estimate of MVAR based connectivity metrics, in fact increase with increasing data length. While none of these empirical results is particularly surprising from a theoretical perspective, their consequences should be taken into account when studying dynamic connectivity from EEG and MEG time series. Finally, our numerical results suggest that an analytic characterization of the impact of the various error causes on the quality of the connectivity estimates would prove useful. Future studies will be devoted to address this point.

References

Astolfi L, Cincotti F, Mattia D, Marciani MG, Baccala LA, de Vico Fallani F, Salinari S, Ursino M, Zavaglia M, Ding L et al (2007) Comparison of different cortical connectivity estimators for high-resolution EEG recordings. Hum Brain Mapp 28(2):143–157

Baccalá L, Sameshima K (2014) Multivariate time-series brain connectivity: a sum up. Methods in brain connectivity inference through multivariate time series analysis. CRC Press, Boca Raton, pp 245–251

Baccalá LA, Sameshima K, Takahashi DY (2007) Generalized partial directed coherence. In: 15th International Conference on Digital Signal Processing (DSP'07), IEEE, pp 163–166

Baccalá LA, De Brito CS, Takahashi DY, Sameshima K (2013) Unified asymptotic theory for all partial directed coherence forms. Philos Trans R Soc A 371(1997):20120, 158

Barnett L, Seth AK (2014) The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. J Neurosci Methods 223:50–68

Bastos A, Schoffelen J (2015) A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front Syst Neurosci 9:175

Bonita JD, Ambolode LCC II, Rosenberg BM, Cellucci CJ, Watanabe TAA, Rapp PE, Albano AM (2014) Time domain measures of inter-channel EEG correlations: a comparison of linear, nonparametric and nonlinear measures. Cogn Neurodyn 8(1):1–15

Brookes MJ, Hale JR, Zumer JM, Stevenson CM, Francis ST, Barnes GR, Owen JP, Morris PG, Nagarajan SS (2011a) Measuring functional connectivity using MEG: methodology and comparison with fcMRI. NeuroImage 56(3):1082–1104

Brookes MJ, Woolrich M, Luckhoo H, Price D, Hale JR, Stephenson MC, Barnes GR, Smith SM, Morris PG (2011b) Investigating the electrophysiological basis of resting state networks using magnetoencephalography. Proc Natl Acad Sci 108(40):16783–16788

Brookes MJ, Woolrich MW, Barnes GR (2012) Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. NeuroImage 63(2):910–920

Chella F, Marzetti L, Pizzella V, Zappasodi F, Nolte G (2014) Third order spectral analysis robust to mixing artifacts for mapping cross-frequency interactions in EEG/MEG. NeuroImage 91:146–161

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3(3):201–215

Cui J, Xu L, Bressler SL, Ding M, Liang H (2008) BSMART: a Matlab/C toolbox for analysis of multichannel neural time series. Neural Netw 21(8):1094–1104

de Pasquale F, Della Penna S, Snyder AZ, Lewis C, Mantini D, Marzetti L, Belardinelli P, Ciancetta L, Pizzella V, Romani GL, Corbetta M (2010) Temporal dynamics of spontaneous MEG activity in brain networks. Proc Natl Acad Sci 107(13):6040–6045

de Pasquale F, Della Penna S, Snyder AZ, Marzetti L, Pizzella V, Romani GL, Corbetta M (2012) A cortical core for dynamic integration of functional networks in the resting human brain. Neuron 74(4):753–764

Eichler M (2006) On the evaluation of information flow in multivariate systems by the directed transfer function. Biol Cybern 94(6):469–482

Engel AK, Gerloff C, Hilgetag CC, Nolte G (2013) Intrinsic coupling modes: multiscale interactions in ongoing brain activity. Neuron 80(4):867–886

Ewald A, Marzetti L, Zappasodi F, Meinecke FC, Nolte G (2012) Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space. NeuroImage 60(1):476–488

Faes L, Pinna GD, Porta A, Maestri R, Nollo G (2004) Surrogate data analysis for assessing the significance of the coherence function. IEEE Trans Biomed Eng 51(7):1156–1166

Faes L, Porta A, Nollo G (2010) Testing frequency-domain causality in multivariate time series. IEEE Trans Biomed Eng 57(8):1897–1906

Faes L, Erla S, Nollo G (2012) Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis. Comput Math Methods Med 2012

Faes L, Erla S, Porta A, Nollo G (2013) A framework for assessing frequency domain causality in physiological time series with instantaneous effects. Philos Trans R Soc A 371(1997):20110,618

Fraschini M, Demuru M, Crobe A, Marrosu F, Stam C, Hillebrand A (2016) The effect of epoch length on estimated EEG functional connectivity and brain network organisation. J Neural Eng 13(3):036,015

Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. J Am Stat Assoc 77(378):304–13

Granger C (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 424–438

Haufe S, Nikulin VV, Müller KR, Nolte G (2013) A critical assessment of connectivity measures for EEG data: a simulation study. NeuroImage 64:120–133

Haykin SS (2008) Adaptive filter theory. Pearson Education India

Hillebrand A, Barnes GR, Bosboom JL, Berendse HW, Stam CJ (2012) Frequency-dependent functional connectivity within resting-state networks: an atlas-based MEG beamformer solution. NeuroImage 59(4):3909–3921

Hincapié A, Kujala J, Mattout J, Daligault S, Delpuech C, Mery D, Cosmelli D, Jerbi K (2016) MEG connectivity and power detections with minimum norm estimates require different regularization parameters. Comput Intell Neurosci 2016:19

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC (1998) Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr 22(2):324–333

Liu S, Molenaar P (2016) Testing for granger causality in the frequency domain: A phase resampling method. Multivar Behav Res 51(1):53–66

Liuzzi L, Gascoyne LE, Tewarie PK, Barratt EL, Boto E, Brookes MJ (2017) Optimising experimental design for MEG resting state functional connectivity measurement. NeuroImage 155:565–576

Lütkepohl H (2007) New introduction to multiple time series analysis. Springer, New York

Marzetti L, Del Gratta C, Nolte G (2008) Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. NeuroImage 42(1):87–98

Marzetti L, Della Penna S, Snyder AZ, Pizzella V, Nolte G, de Pasquale F, Romani GL, Corbetta M (2013) Frequency specific interactions of MEG resting state activity within and across brain networks as revealed by the multivariate interaction measure. NeuroImage 79:172–183

Nolte G, Dassios G (2005) Analytic expansion of the EEG lead field for realistic volume conductors. Phys Med Biol 50(16):3807

Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M (2004) Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin Neurophysiol 115(10):2292–2307

Nunez P, Srinivasan R, Westdorp A, Wijesinghe R, Tucker D, Silberstein R, Cadusch P (1997) EEG coherency. I. Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr Clin Neurophysiol 103(5):499–515

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011

Papoulis A, Pillai S (2002) Probability, random variables, and stochastic processes. McGraw-Hill, New York

Pascual-Marqui RD (2007) Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:07103341

Pascual-Marqui RD, Lehmann D, Koukkou M, Kochi K, Anderer P, Saletu B, Tanaka H, Hirata K, John ER, Prichep L, Biscay-Lirio R, Kinoshita T (2011) Assessing interactions in the brain with exact low-resolution electromagnetic tomography. Philos Trans R Soc A 369(1952):3768–3784

Pereda E, Quiroga RQ, Bhattacharya J (2005) Nonlinear multivariate analysis of neurophysiological signals. Prog Neurobiol 77(1):1–37

Prichard D, Theiler J (1994) Generating surrogate data for time series with several simultaneously measured variables. Phys Rev Lett 73(7):951

Sakkalis V (2011) Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Comput Biol Med 41(12):1110–1117

Schelter B, Winterhalder M, Timmer J (2006) Handbook of time series analysis: recent theoretical developments and applications. Wiley, Weinheim

Schoffelen JM, Gross J (2009) Source connectivity analysis with MEG and EEG. Hum Brain Mapp 30(6):1857–65

Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77(4):635

Schreiber T, Schmitz A (2000) Surrogate time series. Physica D 142(3):346–382

Silfverhuth MJ, Hintsala H, Kortelainen J, Seppänen T (2012) Experimental comparison of connectivity measures with simulated EEG signals. Med Biol Eng Comput 50(7):683–688

Stam CJ (2010) Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders. J Neurol Sci 289(1–2):128–234

Stam CJ, van Straaten EC (2012) Go with the flow: use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. NeuroImage 62(3):1415–28

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58(1):77–94

Van Diessen E, Numan T, van Dellen E, van der Kooi AW, Boersma M, Hofman D, van Lutterveld R, van Dijk BW, van Straaten ECW, Hillebrand A, Stam CJ (2015) Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. Clin Neurophysiol 126(8):1468–81

Wang HE, Bénar CG, Quilichini PP, Friston KJ, Jirsa VK, Bernard C (2014) A systematic framework for functional connectivity measures. Front Neurosci 8:405

Welch P (1967) The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE unified asyptotics on audio and electroacoustics 15(2):70–73

Wendling F, Ansari-Asl K, Bartolomei F, Senhadji L (2009) From EEG signals to brain connectivity: a model-based evaluation of interdependence measures. J Neurosci Methods 183(1):9–18

Wiener N (1956) The theory of prediction. Modern mathematics for engineers 1:125–139