Brain networks construction using Bayes FDR and Average Power Function

Journal:	*Statistical Methods in Medical Research*
Manuscript ID	Draft
Manuscript Type:	Original Article
Keywords:	Average Power Function, Bayes FDR, Functional MRI, High Density EEG, Multiple hypothesis testing
Abstract:	Brain functional connectivity is a widely investigated topic in neuroscience. In recent years, the study of brain connectivity has been largely aided by graph theory. The link between time series recorded at multiple locations in the brain and the construction of a graph is usually an adjacency matrix. The latter converts a measure of the connectivity between two time series, typically a correlation coefficient, into a binary choice on whether the two brain locations are functionally connected or not. As a result, the choice of a threshold τ over the correlation coefficient is key. In the present work we propose a multiple testing approach to the choice of τ that uses the Bayes False Discovery Rate (FDR) and a new estimator of the statistical power called Average Power Function (APF) to balance the two types of statistical error. We show that the APF estimator substantially improves current methodology as it is unbiased, asymptotically robust in case of independence and stationary dependence of the tests and it is reliable under several simulated dependence conditions. Moreover, we propose a robust method for the choice of τ using the 5% and 95% bootstrap percentiles of the APF and FDR distributions respectively to improve stability. We applied our approach to functional Magnetic Resonance Imaging (fMRI) and High Density Electroencephalogram (HD-EEG) data.

Journal Title
XX(X):1–7

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Brain networks construction using Bayes FDR and Average Power Function

Piero Quatto[1] Nicolò Margaritella[2] Isa Costantini[3] Francesca Baglio[4] Massimo Garegnani[5] Raffaello Nemni[6] and Luigi Pugnetti[5]

Abstract

Brain functional connectivity is a widely investigated topic in neuroscience. In recent years, the study of brain connectivity has been largely aided by graph theory. The link between time series recorded at multiple locations in the brain and the construction of a graph is usually an adjacency matrix. The latter converts a measure of the connectivity between two time series, typically a correlation coefficient, into a binary choice on whether the two brain locations are functionally connected or not. As a result, the choice of a threshold τ over the correlation coefficient is key. In the present work we propose a multiple testing approach to the choice of τ that uses the Bayes False Discovery Rate (FDR) and a new estimator of the statistical power called Average Power Function (APF) to balance the two types of statistical error. We show that the APF estimator substantially improves current methodology as it is unbiased, asymptotically robust in case of independence and stationary dependence of the tests and it is reliable under several simulated dependence conditions. Moreover, we propose a robust method for the choice of τ using the 5% and 95% bootstrap percentiles of the APF and FDR distributions respectively to improve stability. We applied our approach to functional Magnetic Resonance Imaging (fMRI) and High Density Electroencephalogram (HD-EEG) data.

Keywords

Average Power Function, Bayes FDR, Functional MRI, High Density EEG, Multiple hypothesis testing

Introduction

Functional connectivity is defined as the temporal dependency within spatially remote neurophysiologic events[1]. In the past years an increasing body of neuroimaging studies explored functional connectivity by measuring the level of co-activation of time series between brain regions. Graph theory is increasingly used to define brain connectivity[2]; in the graphical representation of a brain network, a node corresponds to a brain region while an edge corresponds to an interaction between two brain regions[3]. Binary brain networks are defined only by the presence or absence of connections between brain areas. Although weighted graphs might carry more information, neuroscientists often rely on the detection of significantly active brain areas to prove their hypotheses or formulate new ones; furthermore, binary networks are simple to understand and explain and they have an easily defined null model for statistical comparsion[4]. Alternative methods to define thresholds for a connectivity measure such as Graphical LASSO and Mixed Graphical Models rely on multivariate assumptions of the time-series and are often computationally challenging[5]. On the other hand, the problem for binary networks is usually that of defining a threshold over a measure of connectivity that turns the matrix of all pairwise dependencies between brain areas into an adjacency matrix. This allows the calculation of several measures that summarise important global and local characteristics of brain connectivity, such as centrality, efficiency, density, and small worldness property[6,7].

By partitioning the brain into several anatomical regions of interest (ROI) and extracting time series from each ROI, we can subsequently employ a measure of dependence, namely the nonparametric Spearman's correlation between each pair of ROIs. The resulting correlation matrix is converted into an adjacency matrix by fixing a threshold τ on the Spearman's test statistics.

In a recent study Sala et al.[4] proposed a method to derive the adjacency matrix by using a multiple testing procedure on the correlation coefficients. The multiple testing problem arises from pairwise testing the correlation test statistics of all the time series recorded. Sala et al.[4] controlled type I and II error rates using the postive False Discovery Rate (pFDR) and False Nondiscovery Rate (pFNR) estimates and a method for balancing the two errors.

The present study employs the Bayes False Discovery Rates[8] together with a new Bayesian estimator of the statistical

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy
[2]School of Mathematics, University of Edinburgh, UK
[3]Inria, Université Côte d'Azur, France
[4]Magnetic Resonance Laboratory, Scientific Institute (IRCCS) S. Maria Nascente, Italy
[5]Clinical Neurophysiology Laboratory, Scientific Institute (IRCCS) S. Maria Nascente, Italy
[6]Neurological Rehabilitation Unit, Scientific Institute (IRCCS) S. Maria Nascente, Italy

Corresponding author:
Nicolò Margaritella, University of Edinburgh, School of Mathematics, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK.
Email: N.Margaritella@sms.ed.co.uk

power called Average Power Function (APF). The latter is proven to be unbiased and to asymptotically approximate the actual value of the parameter both in the case of independent and stationary associated p-values. Simulation results show that APF has low bias and mean squared error (MSE) over its full range and also for several types and strengths of spatial dependence among tests.

Furthermore, as stability is an important feature of a testing procedure[9], we propose a robust method to find a suitable threshold τ for the Spearman's test statistics. We employ the 95^{th} and 5^{th} Bootstrap percentiles of FDR and APF respectively to find a threshold on the p-values that guarantees both a small type I error and a reasonably high power with 95% probability. We tested our approach with a Monte Carlo (MC) simulation study and both functional Magnetic Resonance Imaging (fMRI) and High Density Electroencephalogram (HD-EEG) data recorded from a healthy subject.

Multiple Testing

In order to deal with multiple testing we consider m pairs of hypotheses H_0 and H_1, with a priori probabilities defined by $\pi_0 = \mathrm{P}(H_0)$ and $\pi_1 = \mathrm{P}(H_1) = 1 - \pi_0$. Each pair is put through an hypothesis test that returns a p-value p_j for $j = 1, .., m$ which is assumed to be uniformly distributed under H_0.

Let us also consider the probability of false discoveries, called Bayes False Discovery Rate[8],

$$FDR(\gamma) = \mathrm{P}(H_0|p_j \leq \gamma) = \frac{\mathrm{P}(p_j \leq \gamma|H_0)\mathrm{P}(H_0)}{\mathrm{P}(p_j \leq \gamma)} = \frac{\gamma\pi_0}{F(\gamma)},$$

where γ and F represent a suitable threshold for the p-values and their cdf, respectively. Moreover, the probability of true discoveries called Average Power Function (APF) can be defined as

$$APF(\gamma) = \mathrm{P}(p_j \leq \gamma|H_1) = \frac{\mathrm{P}(H_1|p_j \leq \gamma)\mathrm{P}(p_j \leq \gamma)}{\mathrm{P}(H_1)}$$
$$= \frac{\big[1 - FDR(\gamma)\big]F(\gamma)}{1 - \pi_0} = \frac{F(\gamma) - \gamma\pi_0}{1 - \pi_0}.$$

We propose the following estimates

$$\widehat{FDR}(\gamma) = \frac{\gamma\pi_0}{\hat{F}(\gamma)},$$
$$\widehat{APF}(\gamma) = \frac{\hat{F}(\gamma) - \gamma\pi_0}{1 - \pi_0},$$

for which the choice of the p-values empirical cdf $\hat{F}(\gamma) = \#\{p_j \leq \gamma\}/m$ leads to the expected values,

$$E\big[\widehat{FDR}(\gamma)\big] \geq \frac{\gamma\pi_0}{E[\hat{F}(\gamma)]} = FDR(\gamma),$$
$$E\big[\widehat{APF}(\gamma)\big] = \frac{E[\hat{F}(\gamma)] - \gamma\pi_0}{1 - \pi_0} = APF(\gamma),$$

due to Jensen's inequality and $E[\hat{F}(\gamma)] = F(\gamma) = \gamma\pi_0 + APF(\gamma)\pi_1$.

It is worth noting that the FDR estimate is conservative while the APF estimate is unbiased.

Information regarding the a priori probability π_0 can be acquired empirically from the data.This approach defines a conservative estimate of the a priori probability π_0 as showed by Storey (2002)[10],

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda\}}{m(1 - \lambda)} = \frac{1 - \hat{F}(\lambda)}{1 - \lambda},$$

whose expected value depends on λ and is defined as

$$E\big[\hat{\pi}_0(\lambda)\big] = \frac{1 - F(\lambda)}{1 - \lambda} = \pi_0 + \frac{1 - APF(\lambda)}{1 - \lambda}\pi_1 \geq \pi_0,$$

which is obtained through

$$E\big[1 - \hat{F}(\lambda)\big] = 1 - F(\lambda) = (1 - \lambda)\pi_0 + \big[1 - APF(\lambda)\big]\pi_1.$$

Therefore, the empirical Bayes estimates of FDR and APF are

$$\widehat{FDR}(\gamma) = \frac{\gamma\hat{\pi}_0(\lambda_1)}{\hat{F}(\gamma)},$$
$$\widehat{APF}(\gamma) = \frac{\hat{F}(\gamma) - \gamma\hat{\pi}_0(\lambda_2)}{1 - \hat{\pi}_0(\lambda_2)},$$

with $\lambda_1 \neq \lambda_2$. In order to derive the optimal value of λ for each estimate, we resample the m p-values with replacement B times, we calculate the bootstrap versions of $\hat{\pi}_0(\lambda)$ over a range of λ values (e.g. from 0 to 1 with step 0.05) and we minimise the bootstrap estimate of the corresponding mean square error (MSE) as in[4].

The optimal values of λ for the two estimates allow us to construct the one-sided $(1 - \alpha)$-confidence intervals for the parameters by taking, respectively, the $(1 - \alpha)$-quantile of the $\widehat{FDR}_{\lambda_1}(\gamma)$ bootstrap distribution as the upper confidence bound, and the α-quantile of the $\widehat{APF}_{\lambda_2}(\gamma)$ bootstrap distribution as the lower confidence bound of the corresponding parameters.

Since it is not sufficient to control the FDR alone, we propose a robust approach to balance the two types of error rate. The trade-off can be made by first choosing the alpha value for both the $(1 - \alpha)$-quantile of the bootstrap distribution of FDR and the α-quantile of the bootstrap distribution of APF and then evaluating these quantities over all the gamma range and identifying a suitable gamma threshold such that (first) the FDR is low and (second) the APF is reasonably high, both having $(1 - \alpha)$ probability.

Asymptotics

In the case of independent p-values, by the strong law of large numbers, we have $\hat{F}(\gamma) \to E[\hat{F}(\gamma)] = F(\gamma)$ almost surely, from which it follows that,

$$\hat{\pi}_0(\lambda) \to E[\hat{\pi}_0(\lambda)] = \pi_0\left[1 + \frac{1 - APF(\lambda)}{1 - \lambda}\frac{\pi_1}{\pi_0}\right] \geq \pi_0,$$
$$\widehat{FDR}(\gamma) \to \frac{\gamma E[\hat{\pi}_0(\lambda)]}{E[\hat{F}(\gamma)]} = \frac{\gamma\pi_0}{F(\gamma)}\left[1 + \frac{1 - APF(\lambda)}{1 - \lambda}\frac{\pi_1}{\pi_0}\right]$$
$$= FDR(\gamma)\left[1 + \frac{1 - APF(\lambda)}{1 - \lambda}\frac{\pi_1}{\pi_0}\right] \geq FDR(\gamma),$$

and

$$\begin{aligned}\widehat{APF}(\gamma) \to & \frac{E[\hat{F}(\gamma)] - \gamma E[\hat{\pi}_0(\lambda)]}{1 - E[\hat{\pi}_0(\lambda)]} \\ = & \frac{F(\gamma) - \gamma\pi_0\left[1 + \frac{1-APF(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right]}{1 - \pi_0\left[1 + \frac{1-APF(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right]} \\ = & \frac{APF(\gamma) - \gamma\frac{1-APF(\gamma)}{1-\lambda}}{1 - \frac{1-APF(\gamma)}{1-\lambda}} \\ = & \frac{1-\lambda}{APF(\lambda)-\lambda}APF(\gamma) - \gamma\frac{1-APF(\lambda)}{APF(\lambda)-\lambda},\end{aligned}$$

almost surely, by the continuous mapping theorem.
In case of stationary associated p-values, if $\sum_{i=1}^{m}\mathrm{Cov}(p_i, p_j) = o(m)$ for $m \to \infty$ then, by a result of Yu (1993)[11],

$$\sup_{\gamma}\{\hat{F}(\gamma) - F(\gamma)\} \to 0,$$

almost surely, from which it follows that

$$\begin{aligned}\hat{\pi}_0(\lambda) \to & E[\hat{\pi}_0(\lambda)] = \pi_0\left[1 + \frac{1-APF(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right], \\ \widehat{FDR}(\gamma) \to & \frac{\gamma E[\hat{\pi}_0(\lambda)]}{E[\hat{F}(\gamma)]} = FDR(\gamma)\left[1 + \frac{1-APF(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right],\end{aligned}$$

and

$$\begin{aligned}\widehat{APF}(\gamma) \to & \frac{E[\hat{F}(\gamma)] - \gamma E[\hat{\pi}_0(\lambda)]}{1 - E[\hat{\pi}_0(\lambda)]} \\ = & \frac{1-\lambda}{APF(\lambda)-\lambda}APF(\gamma) - \gamma\frac{1-APF(\lambda)}{APF(\lambda)-\lambda},\end{aligned}$$

almost surely, by the continuous mapping theorem.
Therefore, when λ is chosen so that $APF(\lambda)$ is close to 1 (for example, if λ is near 1), the empirical Bayes estimates asymptotically approximate the actual values of the corresponding parameters.

The case of nonparametric independence testing

Let us consider the case of m pairs of hypotheses H_0: independence vs. H_1: dependence, with a priori probabilities defined by $\pi_0 = \mathrm{P}(H_0)$ and $\pi_1 = \mathrm{P}(H_1) = 1 - \pi_0$. We apply the approach introduced in Section 2 to this scenario where dependence is measured through the Spearman's test statistics which is

$$t_j = \frac{r_j}{\sqrt{1 - r_j^2}}\sqrt{n-2},$$

where n is the number of sampled points in the time series, r_j represents the Spearman's rank correlation coefficient and t_j, $j = 1, ..., m$ are approximately distributed under the null hypothesis as a Student's t with $n-2$ degrees of freedom. The corresponding m p-values are approximated by

$$p_j = 2 - 2\mathrm{F}_{n-2}(|t_j|),$$

where F_{n-2} is the Student's cumulative distribution function (cdf) with $n-2$ degrees of freedom. A threshold γ on the p-values corresponds to a threshold τ on the test statistics.
The idea of basing multiple testing procedures on correlation coefficients is common in the biological literature[12]. Information regarding the a priori probability π_0 can be either acquired from previous studies or empirically from the data which is the approach we employed here in the application to fMRI and EEG data. We will always refer to the pair of hypotheses defined above in the following sections where our multiple testing procedure is applied to a simulation study and the construction of fMRI and HD-EEG brain networks.

Simulation study

We performed a Monte Carlo (MC) simulation study to assess the performance of the proposed APF estimator together with the FDR estimator. Multiple tests of the form $\mathrm{H}_0 : \mu_0 = 0$; $\mathrm{H}_1 : \mu_1 = 2$ were simulated on the Spearman's test statistics by using its normal asymptotic distribution with $\sigma^2 = 1$.We define one hundred tests with $\pi_0 = 0.3$ as the proportion of true null-hypotheses (treated as known) and $\pi_1 = 1 - \pi_0$; this corresponds to 100 nodes and 7000 edges graph. Tests were repeatedly simulated $B = 1000$ times by drawing from a multivariate normal with parameters $\underline{\mu} = (0_1, ..., 0_{30}, 2_{31}, ..., 2_{100})$ and $\Sigma = I_{100}$. For each test the p-value is defined as $p_{i,b} = P\{N(0,1) \geq z_i\}$ for the b-th iteration, where z_i is the i-th observed value of the vector $\underline{z}$ drawn from $N_{100}(\underline{\mu}, \Sigma)$.
The multivariate normal distribution allows to study the performance of the estimators when the independence between tests is violated; by modifying the correlation structure in Σ, we employed typical forms of spatial dependence, namely the first-order autoregressive structure $\rho^{|d|}$ [13] and the Matérn class of covariance functions $C_v(d)$ for $v = (1/2, \infty)$ and $\rho = (0.2, 0.4, 0.7)$,

$$C_v(d) = \sigma^2\frac{2^{1-v}}{\Gamma(v)}\left(\sqrt{2v}\frac{d}{\rho}\right)^v K_v\left(\sqrt{2v}\frac{d}{\rho}\right),$$

where d is is the absolute distance between two tests, Γ is the gamma function and K_v is the modified Bessel function of the second kind[14]. To assess the overall performance of the estimators we computed the MC bias and MSE as follows:

$$\begin{aligned}\widehat{Bias}_{APF}(\gamma) = & \frac{1}{B}\sum_{b=1}^{B}\widehat{APF^{*b}}(\gamma) - APF(\gamma), \\ \widehat{MSE}_{APF}(\gamma) = & \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{APF^{*b}}(\gamma) - APF(\gamma)\right)^2,\end{aligned}$$

and similarly for the FDR. Monte Carlo Bias and MSE were reported for a sensible set of γ values in Table 1.

Application

A thirty year old healthy woman from the research team of the Scientific Institute Santa Maria Nascente of the Don Gnocchi Foundation (Milan, Italy) volunteered for the study. She underwent resting state functional Magnetic Resonance Imaging (fMRI) and High Density electroencephalogram (HD-EEG) recordings. Each exam lasted 20 minutes and was recorded at the same hour of the day in a darkened room with

the subject laid in supine position with eyes closed. She was instructed to keep alert and relaxed; no specific mental task was requested.

fMRI

The resting state fMRI was carried out at the Department of Radiology using a 1.5 T Siemens Magnetom Avanto (Erlangen, Germany) MRI scanner with 8-channel head coil. BOLD EPI images were collected at rest for approximately 8 minutes. High resolution T1-weighted 3D scans were also collected to be used as anatomical references for fMRI data analysis. Standard pre-processing involved the following steps: motion and EPI distortion corrections, non-brain tissues removal, high-pass temporal filtering (cut-off 0.01 Hz) and artefacts removal using the FMRIB ICA-based Xnoiseifier (FIX) toolbox [15].

After the pre-processing, the resulting 4D dataset was aligned to the subject's high-resolution T1-weighted image, registered to MNI152 standard space and subsequently resampled to $2 \times 2 \times 2\ mm^3$ resolution. One hundred ninety volumes were available for successive analyses. fMRI time series were then extracted as the average signal within each of 84 human functional Brodmann's Areas (BA) as regions of interest (ROIs) using the Resting-State fMRI Data Analysis Toolkit REST [16].

HD-EEG

The high density EEG (HD-EEG) was recorded in the Neurophysiology Lab using a BrainVision Recorder 1.20 (Brain Products GmbH, Germany) and a pre-cabled EEG recording cap equipped with 64 Ag/AgCl electrodes with FCz as the reference. Analog signals were digitalized at 500 Hz sampling rate and bandpass filtered from 0.1 to 100 Hz. Raw data were further notch filtered at 50 Hz and band-pass filtered (1-30 Hz) off-line. Before segmentation, both visual inspection and Independent Component Analysis (ICA) were used for semi-automated removal of ocular artefacts [17]. Data were then segmented into consecutive non overlapping 2.5-seconds epochs yielding 120 epochs available for successive analyses.

EEG time series for each ROI were obtained by first applying the standard procedures for the computation of mean spectral density. The cross-spectral matrix was used as input for sLORETA source analysis [18]. Source activities were combined into 84 regions of interest (ROIs). Each ROI centre was placed at the respective BA centroid and then the time series of the electric neuronal activity at the ROIs were extracted.

Results

Figure 1 and Table 1 report the results of the Monte Carlo simulation study. Figure 1 shows the difference between the true values of APF (blue) and their point estimates computed throughout the full γ range for different covariance functions and correlation intensities. The estimates are close to the true values of APF for the full γ range and in almost every scenario tested, the highest variability being observed with the most correlated spatial structure. Table 1 reports MC Bias and MSE of APF and FDR for a meaningful set of γ values and for different correlation patterns and intensities. All the estimates of FDR and APF show low bias and MSE. The APF bias turns out to be always conservative for γ values equivalent to the range of power most useful in applications (0.4 to 0.9). The MSE tends to grow, especially for the APF, as γ increases or the spatial correlation structure becomes stronger.

fMRI and HD-EEG brain network construction

We computed the 95^{th} percentile of the bootstrap distribution of FDR ($95^{th}\widehat{FDR^{*b}}$) together with the 5^{th} percentile of the bootstrap distribution of APF ($5^{th}\widehat{APF^{*b}}$) to identify a suitable threshold τ for the construction of the fMRI and HD-EEG networks. Figure 2 shows the selected quantiles of FDR and APF over a range of τ values. Both $95^{th}\widehat{FDR^{*b}}$ and $5^{th}\widehat{APF^{*b}}$ decrease as the threshold τ increases. This allowed us to draw the trade-off between type I error and power for the two networks (Figure 3). Therefore, it was possible to balance power and type I error by considering the set of pairs $(95^{th}\widehat{FDR^{*b}}(\gamma),\ 5^{th}\widehat{APF^{*b}}(\gamma))$ and choosing a suitable pair. An example for both the fMRI and HD-EEG is reported in Table 2. In order to find a suitable trade-off for these data we considered the standard experimental framework where priority is on controlling the type I error; however, we also added the APF to the decision-making process. In particular, we chose τ so as to achieve at least 50% of APF (with 95% probability) while keeping the FDR low (at most 10% with 95% probability).

The FDR–APF trade-off for the fMRI network (Figure 3, left) did not provide alternatives: to control both errors sensibly we had to select a threshold τ returning estimates of FDR and APF not greater than 10% and of at least 50% respectively with 95% probability. In this scenario, a threshold τ returning an FDR of at most 5% would not guarantee an APF of at least 50% with 95% probability (Figure 3, left).

On the other hand, without any a priori knowledge about the HD-EEG network, the FDR–APF trade-off allowed different reasonable choices of the threshold τ (Figure 3, right). For instance, the researcher could arguably favour a low upper bound for the FDR and guarantee no more than at least 50% of APF with 95% probability (Table 2, HD-EEG(1)), although, in this case it would be better choosing a less conservative FDR in order to gain a much more desirable lower bound for the power (Table 2, HD-EEG(2)). It is worth noting the macroscopic impact these choices have on the properties of the subsequent brain networks; for instance, the resulting HD-EEG networks have density, i.e. number of links over all possible connections [19], of 0.46 and 0.72 respectively (Table 2).

Discussion

The present study addresses the problem of setting a threshold τ on the correlation coefficient between time series of brain activity recorded from several brain areas. An adjacency matrix is consequently defined for the construction of networks which are widely employed by the neuroscientific community for the study of the brain activity. To this purpose, we paired the Bayes FDR [8] with a new

estimator called Average Power Function (APF) to take into account both type I error and statistical power in the choice of τ. As pointed out in Sala et al.[4], combining the two types of error helps the construction of reliable brain networks.

The APF has features which improves substantially the current literature[4] as we were able to prove its unbiasedness and its almost sure convergence to the actual value of the parameter, either when assuming independent or stationary associated p-values. As the spatial dependence of multiple tests can affect the selection of the threshold τ, the behaviour of the APF and FDR estimator under different types and strengths of dependence of the tests should be considered. Here we employed standard forms of spatial dependence and thus our results might not be the same in other scenarios. Nonetheless, our results showed that both FDR and APF have low Monte Carlo bias and MSE, even when different structures of spatial dependence among tests are considered. Furthermore, as reproducibility of results is of major concern in neuroscientific studies employing testing procedures[9], we proposed the combined use of 95% and 5% percentiles of FDR and APF bootstrap distributions respectively to account for sampling variability in the choice of τ and hence returning more robust results in terms of network stability. This approach allows also a straightforward interpretation of the threshold as the 95% probability of achieving at least the power and at most the type I error estimated by the pair of APF and FDR percentiles chosen.

The results of the application to the construction of fMRI and HD-EEG networks supported the added value of our method: When there is only one sensible choice for the threshold τ, as in the fMRI example, the pair FDR-APF informs researchers on both the statistical errors they are willing to accept. On the other hand, when multiple choices of τ are possible, as for the HD-EEG example, the addition of the APF enables a more informed choice of the threshold than the FDR alone. Our method is not limited to fMRI and EEG networks; we believe the additional information on power helps researchers who employ multiple tests to strengthen their results.

It is worth noting that, in the case of HD-EEG, a highly dense brain network was expected as a result of a well-known phenomenon called volume conduction[20]. Nonetheless, there were different sensible choices of τ which returned even moderately dense HD-EEG networks. In such example, the combined use of FDR and APF proved to be a helpful tool in selecting the threshold which most effectively captures the correct density structure of the underlying phenomenon.

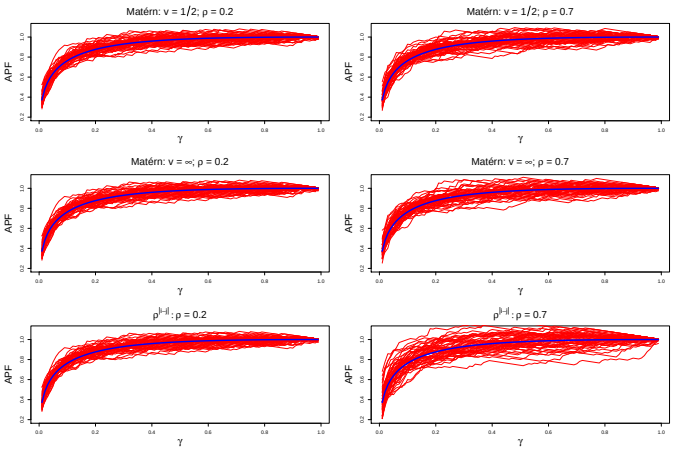

Figure 1. True APF values (dark line) and 50 replications of their point estimates for different covariance functions and correlation intensities. Results are shown over all range of γ (x-axis).

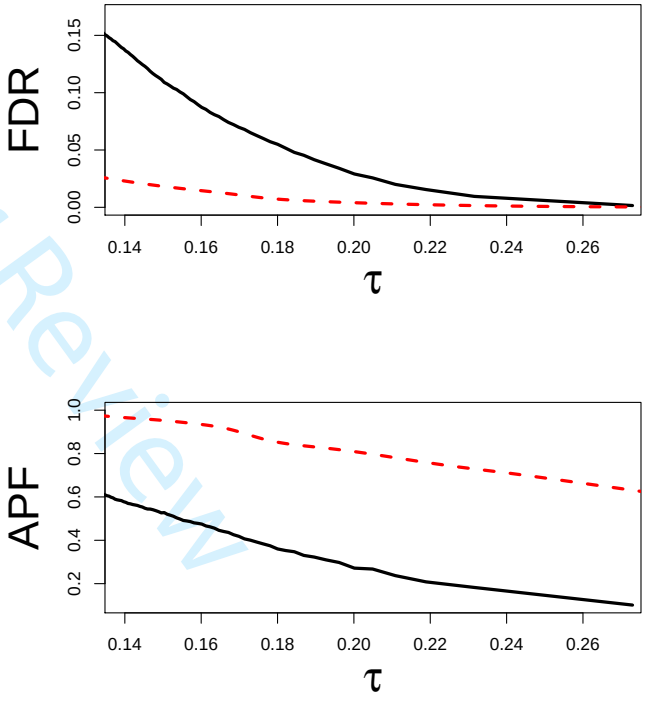

Figure 2. 95^{th} bootstrap percentile of the FDR (upper) and 5^{th} bootstrap percentile of the APF (lower) for a range of τ values. Bootstrap quantiles are reported for both the fMRI (line) and HD-EEG (dashed) data.

References

1. Friston KJ. Functional and effective connectivity: A review. *Brain Connect* 2011; 1: 13–36.
2. Van den Heuvel MP and Hulshoff Pol HE. Exploring the brain network: A review on resting state fMRI functional connectivity. *Eur Neuropsychopharmacol* 2010; 20: 519–534.
3. Rubinov M and Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 2010; 52: 1059–1069.
4. Sala S, Quatto P, Valsasina P, et al. pFDR and pFNR for brain networks construction. *Stat Med* 2014; 33: 158–169.
5. Lee JD and Strimmer TJ. Learning the structure of mixed graphical models. *J Comput Graph Stat* 2015; 24: 230–253.
6. Latora V and Marchiori M. Economic small-world behavior in weighted networks. *Eur Phys J B* 2003; 32: 249–263.
7. Watts D and Strogatz S. Collective dynamics of small-world networks. *Nature* 1998; 393: 440–442.
8. Efron B. *Large-Scale Inference. Empirical Bayes Methods for Estimation, Testing and Prediction*. Cambridge: Cambridge University Press, 2010.

Table 1. Monte Carlo Bias and MSE for a sensible range of γ values, different covariance functions and correlation intensities. *d* is the absolute distance between two nodes. Results are scaled by 10^5.

Cov Function	*Monte Carlo Bias(MSE)* $[*10^{-5}]$				
Independence	$\gamma = 0.0001$	0.001	0.01	0.1	0.2
FDR	31.4(0.08)	34.6(0.22)	33.7(0.43)	28.2(1.53)	50.5(2.28)
APF	176.6(52.6)	133.7(177.2)	−106.6(335.4)	−10.2(327.6)	−276.0(254.5)
$\rho^d;\ \rho = 0.2$					
FDR	34.9(0.09)	47.2(0.43)	45.7(0.64)	46.5(1.92)	54.0(2.82)
APF	225.1(61.0)	149.5(218.6)	−139.5(441.9)	−183.1(408.2)	−251.6(313.6)
$C_{1/2}(d);\ \rho = 0.4$					
FDR	32.7(0.08)	40.5(0.31)	38.4(0.52)	34.2(1.67)	52.9(2.49)
APF	202.3(55.3)	118.0(194.9)	−132.3(370.8)	−63.1(356.5)	−277.4(278.1)
$C_\infty(d);\ \rho = 0.4$					
FDR	31.7(0.08)	38.1(0.30)	35.7(0.47)	27.6(1.60)	53.7(2.32)
APF	199.4(55.0)	160.8(187.5)	−108.1(353.9)	18.3(343.2)	−306.0(259.7)
$\rho^d;\ \rho = 0.4$					
FDR	40.6(0.10)	62.9(0.62)	61.7(1.00)	63.2(2.59)	53.4(3.84)
APF	283.7(76.9)	120.9(268.8)	−63.8(613.4)	−271.7(529.1)	−127.4(423.8)
$C_{1/2}(d);\ \rho = 0.7$					
FDR	35.8(0.09)	50.2(0.46)	47.6(0.67)	51.6(2.02)	55.1(3.0)
APF	246.6(63.7)	132.3(225.8)	−126.6(468.5)	−235.9(426.2)	−243.1(333.1)
$C_\infty(d);\ \rho = 0.7$					
FDR	37.2(0.09)	53.5(0.55)	45.4(0.72)	52.5(2.10)	46.2(3.1)
APF	236.6(67.1)	190.9(237.0)	−13.8(472.2)	−225.9(447.0)	−134.5(342.5)
$\rho^d;\ \rho = 0.7$					
FDR	60.8(0.15)	160.0(2.47)	151.7(4.76)	118.3(6.22)	122.0(8.5)
APF	386.6(132.9)	169.5(548.1)	−10.9(1282)	−271.7(1116)	−380.2(849.6)

Table 2. Examples of thresholds γ on the p-values and τ on the Spearman's test statistics for the fMRI and HD-EEG networks. The 95^{th} bootstrap percentile of FDR and the 5^{th} bootstrap percentile APF are reported together with the resulting networks density. Two examples of threshold are proposed for the HD-EEG network that preserve a small FDR while return different values of APF and network density.

	$95^{th}\widehat{FDR^{*b}}$	$5^{th}\widehat{APF^{*b}}$	Network density
fMRI			
$\gamma = 0.0331$	0.103	0.503	0.22
$\tau = 0.154$			
HD-EEG (1)			
$\gamma = 0.0001$	0.00003	0.512	0.46
$\tau = 0.273$			
HD-EEG (2)			
$\gamma = 0.0191$	0.004	0.802	0.72
$\tau = 0.168$			

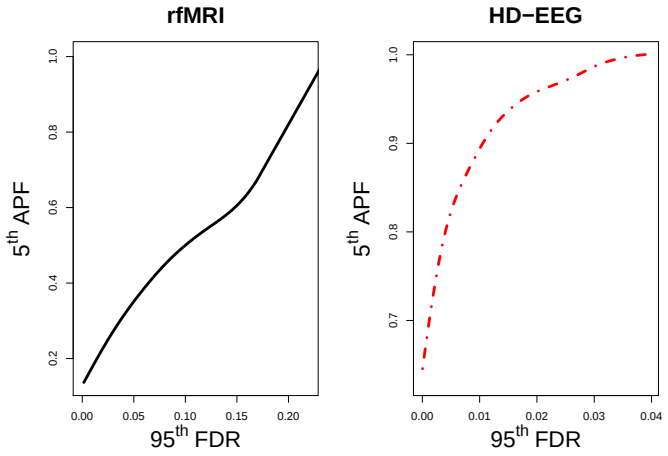

Figure 3. Trade-off between $95^{th}\widehat{FDR^{*b}}$ and $5^{th}\widehat{APF^{*b}}$ for the fMRI (left) and HD-EEG (right, dot-dashed) networks.

9. Dunrez J, Roels SP and Moerkerke B. Multiple testing in fMRI: an empirical case study on the balance between sensitivity, specificity, and stability. *Biom J* 2014; 56: 649–661.
10. Storey J. A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol* 2002; 64: 479–498.
11. Yu H. A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probab Theory Rel* 1993; 95: 337–370.
12. Schäfer J and Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 2005; 21: 754–764.
13. Myers RH, Montgomery DC and Vining GG. *Generalized Linear Models with Applications in Engineering and the Sciences*. New York: Wiley, 2002.
14. Rasmussen CE and Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: The MIT press, 2006.
15. Griffanti L, Salimi-KhorshidiVG, Beckmann CF, et al. ICA-based artefact removal and accelerated fMRI acquisition for

improved resting state network imaging. *Neuroimage* 2014; 95: 232–247.
16. Song XW, Dong ZY, Long XY, et al. REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *Plos One* 2011; 6: e25031.
17. Jung TP, Makeig S, Humphries C, et al. Removing electroencephalographic artefacts by blind source separation. *Psychophysiology* 2000; 37: 163–178.
18. Pascual-Marqui RD, Michel CM and Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiol* 1994; 18: 49–65.
19. Kolaczyk ED. *Statistical Analysis of Network Data. Methods and Models*. New York: Springer Series in Statistics, 2009.
20. Khadem A and Hossein-Zadeh GA. Quantification of the effects of volume conduction on the EEG/MEG connectivity estimates: an index of sensitivity to brain interactions. *Physiol Meas* 2014; 35: 2149–2164.

Declaration of conflicting interests

The authors have declared no conflict of interest.

Funding

The study was supported by a grant of the Italian Ministry of Health, Ricerca Corrente funding program 2015-2016 [RC2015].

Supplemental material

An R-package for implementing the FDR-APF robust threshold estimation will be available online.