NeuroImage 82 (2013) 87–100

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Estimation of resting-state functional connectivity using random subspace based partial correlation: A novel method for reducing global artifacts

Tianwen Chen [a,*], Srikanth Ryali [a], Shaozheng Qin [a], Vinod Menon [a,b,c,*]

[a] Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA
[b] Program in Neuroscience, Stanford University School of Medicine, Stanford, CA 94305, USA
[c] Department of Neurology & Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA

ARTICLE INFO

*Article history:*
Accepted 24 May 2013
Available online 5 June 2013

*Keywords:*
fMRI
Resting state
Functional connectivity
Random subspace
Partial correlation
Global artifacts

ABSTRACT

Intrinsic functional connectivity analysis using resting-state functional magnetic resonance imaging (rsfMRI) has become a powerful tool for examining brain functional organization. Global artifacts such as physiological noise pose a significant problem in estimation of intrinsic functional connectivity. Here we develop and test a novel random subspace method for functional connectivity (RSMFC) that effectively removes global artifacts in rsfMRI data. RSMFC estimates the partial correlation between a seed region and each target brain voxel using multiple subsets of voxels sampled randomly across the whole brain. We evaluated RSMFC on both simulated and experimental rsfMRI data and compared its performance with standard methods that rely on global mean regression (GSReg) which are widely used to remove global artifacts. Using extensive simulations we demonstrate that RSMFC is effective in removing global artifacts in rsfMRI data. Critically, using a novel simulated dataset we demonstrate that, unlike GSReg, RSMFC does not artificially introduce anti-correlations between inherently uncorrelated networks, a result of paramount importance for reliably estimating functional connectivity. Furthermore, we show that the overall sensitivity, specificity and accuracy of RSMFC are superior to GSReg. Analysis of posterior cingulate cortex connectivity in experimental rsfMRI data from 22 healthy adults revealed strong functional connectivity in the default mode network, including more reliable identification of connectivity with left and right medial temporal lobe regions that were missed by GSReg. Notably, compared to GSReg, negative correlations with lateral fronto-parietal regions were significantly weaker in RSMFC. Our results suggest that RSMFC is an effective method for minimizing the effects of global artifacts and artificial negative correlations, while accurately recovering intrinsic functional brain networks.


## Introduction

Resting-state functional magnetic resonance imaging (rsfMRI) has emerged as a powerful technique for characterizing brain networks and functional connectivity (Beckmann et al., 2005; Biswal et al., 1995; Fox and Raichle, 2007; Fox et al., 2005; Greicius et al., 2003; Supekar et al., 2008; Van Dijk et al., 2010). One commonly used method for functional connectivity analysis is a seed-based investigation in which time series from a seed region of interest (ROI) is used as a covariate in a regression analysis with all other voxels in the brain. This approach has led to a number of important discoveries including the default mode network (DMN) (Greicius et al., 2003). Despite its widespread application to the characterization of intrinsic functional brain circuits in health and disease, the question of how global noise processes should be removed represents a significant and vexing problem (Birn, 2012; Weissenbacher et al., 2009).

Spontaneous fluctuations of rsfMRI signals contain multiple sources of noise that are, in general, hard to estimate and remove. For example, cardiac pulsation induces signal fluctuations in large vessels which then cause widespread BOLD signals changes in the brain (Dagli et al., 1999). Global noise also arises from respiration cycles that can cause head movements and variations in the static magnetic field, which subsequently impact signals across the entire brain (Raj et al., 2001). Additionally, variations in both respiration and heart rate can cause correlated signal changes throughout gray matter (Birn et al., 2006; Chang et al., 2009; Shmueli et al., 2007; Wise et al., 2004). Critically, due to the aliasing effects from long sampling times typically used in rsfMRI scanning, such physiological noise cannot be removed by filtering in the frequency domain (Lowe et al., 1998). Consequently, rsfMRI signal fluctuations arising from neurophysiological activity are confounded by multiple global noise processes, thereby leading to overestimation of intrinsic functional connectivity. Removal of these global artifacts from rsfMRI signals is

* Corresponding author at: Department of Psychiatry & Behavioral Sciences, 401 Quarry Rd. Stanford University School of Medicine, Stanford, CA 94305-5719, USA. Fax: +1 650 736 7200.
*E-mail addresses:* tianwenc@stanford.edu (T. Chen), menon@stanford.edu (V. Menon).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.05.118

therefore of paramount importance for accurate measurement of intrinsic functional connectivity.

In recent years, several methods have been developed to remove different components of these global artifacts. RETROICOR (Glover et al., 2000) removes time-locked cardiac and respiratory artifacts, and RVHRCOR (Chang et al., 2009) regresses out signal changes related to respiration and heart rate variations. Both methods require independent and accurate external measurements of heart rate and respiration; data that is often difficult to acquire in pediatric and clinical participants. Furthermore, most public domain rsfMRI datasets from sources such as the 1000 Functional Connectomes Project and Autism Brain Imaging Data Exchange (ABIDE) do not contain measures of heart rate and respiration thereby precluding the use of existing global artifact removal methods for these important publically available datasets. Thus, alternate and accurate methods are needed for global artifact removal in rsfMRI data. Most commonly used methods to achieve this goal are based on estimation and removal of global noise derived from the rsfMRI data itself. These approaches are much more flexible and researchers have used a variety of methods to estimate non-neurophysiological noise in the data. For example, some studies have used principal components from white matter and cerebrospinal fluid (CSF) fMRI signals as nuisance regressors that presumably do not contain signals from neurophysiological sources (Behzadi et al., 2007; Chai et al., 2012). However, because respiration also impacts gray matter (Birn et al., 2006; Wise et al., 2004), signals from white matter alone do not fully represent global artifacts, and consequently functional connectivity between brain regions may still be overestimated. To overcome this issue researchers have used various types of global signal regression (GSReg) procedures based on either the global mean signal computed across the whole brain (Desjardins et al., 2001; Greicius et al., 2003; Macey et al., 2004) or a linear combination of signals computed from voxels in grey matter, white matter and CSF (Fox et al., 2005). GSReg has been the most widely used approach because early studies revealed a more consistent and focal pattern of functional brain connectivity (Fox et al., 2005, 2009; Greicius et al., 2003). For example, analysis of PCC connectivity using GSReg has consistently identified major nodes of the DMN consistent with other approaches such as ICA (Seeley et al., 2007). One problem with GSReg is that it also identifies strong negative correlations. The validity of GSReg has recently been questioned because it introduces artificial anti-correlations in ways that can be unambiguously demonstrated mathematically (Murphy et al., 2009; Weissenbacher et al., 2009). Thus, observed anti-correlation between brain systems in experimental data might arise as an artifact of the procedures currently used to estimate and remove the global artifacts. It currently remains unclear how to derive optimal nuisance regressors that can produce the most robust and accurate functional connectivity map.

A different approach is to use partial correlation based methods that can remove the effects of global artifacts by measuring the connectivity between the seed region and every voxel in the brain after removing the (linear) dependence of other voxels. Partial correlations between the seed region and all brain voxels can be computed by inverting and appropriately scaling the sample covariance matrix (Edwards, 2000) based on the time series of the seed region and all brain voxels. Unfortunately, since the number of features ($p$, number of voxels) is larger than the number of samples ($N$, number of time points or scans), the sample covariance matrix is singular and is not invertible (Ryali et al., 2012). In such cases, pseudo-inverse methods are often used. The pseudo-inverse is constructed from nonzero eigenvalues of the sample covariance matrix and corresponding eigenvectors. However, pseudo-inverse solutions suffer from significant estimation error when $p \gg N$ because components corresponding to nonzero eigenvalues of the sample covariance matrix may be eliminated even though they contain useful information (Hoyle, 2010). To overcome this problem, Hoyle (2010) proposed a random subspace method (RSM) to reduce estimation errors of standard pseudo-inverse methods. In RSM, multiple subsets of features are randomly sampled from the feature space, and partial correlations between features within each subset are computed using a pseudo-inverse. RSM provides a more accurate estimate of the partial correlation matrix because the sample-to-feature ratio is higher in each random subspace compared to the original feature space, thus shifting the estimation error curve towards the direction of a larger effective sample size.

Here, we develop a novel RSM-based method to remove global artifacts and estimate whole-brain functional connectivity in rsfMRI data—an approach we refer to as RSM functional connectivity or RSMFC. We first evaluate our methods on a carefully constructed simulated dataset in which there are no inherent negative correlations. We then use this dataset to examine the performance of RSMFC and compare its performance with results from GSReg. Critically, we demonstrate that unlike GSReg, RSMFC does not artificially introduce negative correlations in data in which there are no inherent negative correlations. Finally, we examine functional connectivity of the posterior medial cortex based on experimental rsfMRI data from 22 healthy adults and show that our method effectively removes global artifacts and recovers the DMN with better anatomical specificity than GSReg.

## Methods

### *Estimation of partial correlations in seed-based functional connectivity analysis*

Let $Y_{N\times p}$ be BOLD fMRI time series of $p$ voxels. Observations (rows of $Y$) are sampled from a multivariate normal distribution $N(\mu_{1\times p},\Sigma_{p\times p})$. A partial correlation value $\Pi_{ij}$ is a measure of the direct linear interaction between brain voxels $i$ and $j$ that cannot be explained by influence of the remaining $(p-2)$ voxels. It can be shown that the partial correlation matrix $\Pi$ can be computed from the covariance matrix $\Sigma$ of $p$ voxels by using the following relations (Edwards, 2000)

$$\Theta = \Sigma^{-1}, \tag{1}$$

$$\Pi_{i,j} = -\frac{\Theta_{i,j}}{\sqrt{\Theta_{i,i}\Theta_{j,j}}}. \tag{2}$$

Typically, $\Sigma$ is estimated by the sample covariance matrix

$$\hat{\Sigma} = \frac{(Y-\overline{Y})^T(Y-\overline{Y})}{N-1}, \tag{3}$$

and $\Pi$ is estimated by using the sample estimate of $\Sigma\left(\hat{\Sigma}^{-1}\right)$. However, the inversion of $\Sigma$ in Eq. (1) is problematic for high-dimensional fMRI data because when the number of time points ($N$) is less than the number of voxels ($p$), $\hat{\Sigma}$ becomes singular and is not invertible. To circumvent this issue, Moore–Penrose pseudo-inverse (here, referred to pseudo-inverse) is commonly used, which is constructed from the eigenvectors of corresponding nonzero eigenvalues of the sample covariance matrix $\hat{\Sigma}$. In Eqs. (1) and (2), $\Theta$ and $\Pi$ are now estimated as

$$\hat{\theta} = \hat{\Sigma}^{+}. \tag{4}$$

$$\hat{\Pi}_{i,j} = -\frac{\hat{\theta}_{i,j}}{\sqrt{\hat{\theta}_{i,i}\hat{\theta}_{j,j}}}. \tag{5}$$

where $\hat{\Sigma}^{+}$ denotes the pseudo-inverse of $\hat{\Sigma}$.