# Principal Components Analysis Reveals the Correlation Structure of Resting-State fMRI Data

**H. HE[1], AND T. T. LIU[2]**
[1]ZHEJIANG UNIVERSITY, HANGZHOU, ZHEJIANG, CHINA, PEOPLE'S REPUBLIC OF, [2]CENTER FOR FUNCTIONAL MRI AND DEPARTMENT OF RADIOLOGY, UC SAN DIEGO, LA JOLLA, CALIFORNIA, UNITED STATES

***INTRODUCTION:*** Resting-state functional magnetic resonance imaging (fMRI) is proving to be an effective tool for mapping the long-range functional connections of the brain in both health and disease. A common but controversial pre-processing step in the analysis of resting-state fMRI data is the removal of a global time-course that is the average of all the time-courses within the brain. It has been argued that global signal regression (GSR) imposes spurious negative correlations between brain networks, specifically between the default mode (DMN) and task-positive networks (TPN) [1,2]. In this abstract we use principal components analysis to decompose resting-state correlation maps and present a method for ranking the principal components according to their relative contributions to a specified functional connectivity map. The results reveal the structure of resting-state correlation maps: the first ranked component is essentially identical to the global signal, and the second component shows the anti-correlation between the DMN and TPN.

***METHODS:*** The dataset used in this paper was originally analyzed by [3] and downloaded from www.brainscape.org (dataset BS002). This dataset comprises 17 normal right-handed young adults (9 females). Image preprocessing steps included slice timing correction, head-motion correction, spatial normalization to Talairach space, and spatial smoothing (FWHM = 6mm). Nuisance terms (including temporal mean, linear trends, and motion parameters) were regressed out of each voxel time-course. A low pass filter (0.1Hz cutoff) was applied to the residual data. Finally, the data were normalized to be unit-norm. Seed regions were defined using spheres with a radius of 6-mm. A correlation map was then computed by correlating the mean seed signal in the PCC region ([0,-51,26]) with the time-course from every voxel in the brain. An additional correlation map after GSR was also formed. For principal components analysis, we organized the data from each run as a $M \times N$ data matrix $\mathbf{Y}$ where the columns are the voxel time courses. Since columns were normalized to have zero mean and unit norm, the correlation matrix can be written as $\mathbf{C} = \mathbf{Y}^T\mathbf{Y}$. Using the singular value decomposition (SVD) $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^T$ we can further obtain $\mathbf{C} = \mathbf{V}\Sigma^2\mathbf{V}^T = \sum_{i=1}^{M}\sigma_i^2\mathbf{v}_i\mathbf{v}_i^T$, where $\mathbf{U}$ is a matrix composed of the left singular vectors, $\Sigma$ is a diagonal matrix composed of the singular values $\sigma_i$, and $\mathbf{V}$ is composed of the right singular vectors $\mathbf{v}_i$. Note that the matrix $\mathbf{C}$ contains the pair-wise correlations between all pairs of voxel time-courses. The *jth* column of $\mathbf{C}$ corresponds to the correlation map describing the correlations of all voxels with the *j*th voxel. This map may be expressed as the sum of component correlation maps $\mathbf{c}_j = \sum_{i=1}^{M}\sigma_i^2 v_{i,j}\mathbf{v}_i$, where $v_{i,j}$ is the *jth* element of the *ith* component vector. We rank each component by its absolute weight $\left|\sigma_i^2 v_{i,j}\right|$, which depends on both the singular value and the component coefficient $v_{i,j}$. We can then form a low dimensional approximation $\tilde{\mathbf{c}}_j = \sum_{k=1}^{K}\sigma_{i(k)}^2 v_{i(k),j}\mathbf{v}_{i(k)}$, where $K$ is the number of components retained and *i(k)* indicates the component index with the *kth* rank, with $k = 1$ corresponding to the largest weight $\max_i\left(\left|\sigma_i^2 v_{i,j}\right|\right)$. In practice, due to the similarity between the global signal and the first component (see Results), it is useful to modify this ranking so that the first principal component (as ranked by its singular value) has the first rank. In addition, the ranking is readily modified to handle seed signals formed from a region of interest by replacing the component coefficients $v_{i,j}$ with the spatial average of the coefficients over a specified region.

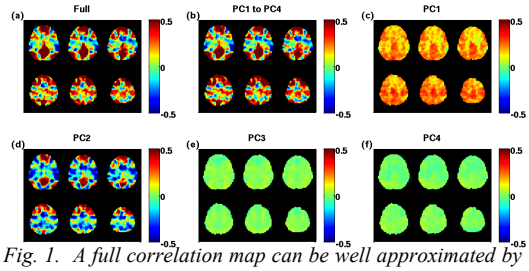


*Fig. 1. A full correlation map can be well approximated by the correlation maps of a few ranked principal components.*

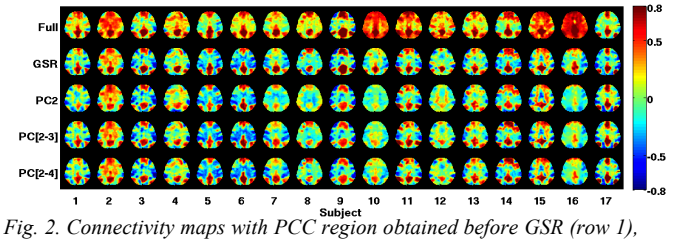


*Fig. 2. Connectivity maps with PCC region obtained before GSR (row 1), after GSR (row 2), and three low-dimensional approximations (rows 3 to 5) for a representative slice from each subject.*

***RESULTS:*** Figure 1 shows a correlation map obtained using a seed signal from the PCC region (Fig. 1a) and a low-dimensional approximation using the top 4 components (Fig. 1b). Note the high degree of similarity between the maps. We found that the global signal was highly correlated with the first principal component across all subjects and runs (r > 0.94). The correlation map associated with this component is shown in Fig. 1c. The second component map (Fig. 1d) exhibits the anti-correlated networks of the default mode network and the task positive network. Fig. 2 shows the functional connectivity maps with PCC seed region for a representative slice from each subject (z = 32 mm). Rows 1 and 2 shows the correlation maps obtained prior to and after GSR, respectively, with the anti-correlated networks clearly evident in the maps obtained with GSR. Rows 3 through 5 show the low dimensional correlation maps obtained when using just the 2nd component, the 2nd and 3rd components, and the 2nd through 4th components, respectively. The 2nd component alone captures most of the correlation structure of the anticorrelated networks. Furthermore, we present low-dimensional visualizations of the resting-state data (for 2 subjects) projected onto either the top 3 (Figs. 3a and 3b) or top 2 ranked components (Figs 3c and 3d). In these low-dimensional views, the brain regions associated with the default mode network (DMN) (i.e. PCC and MPF) are clearly in a different part of the signal space than the regions associated with the task positive network (TPN) (i.e. IPS, FEF, and MT+).

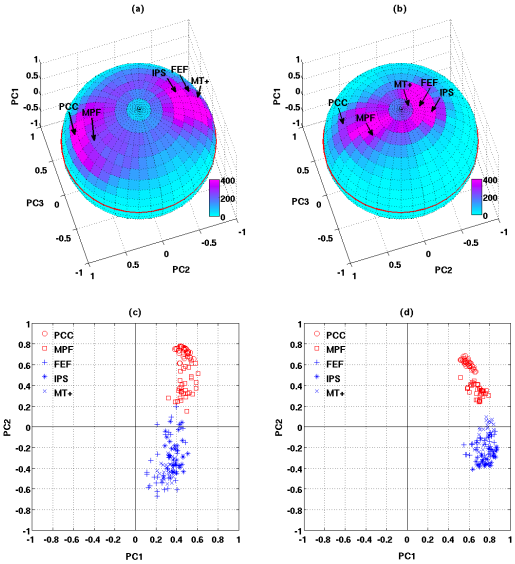


*Fig. 3. (a,b) Low-dimensional visualization of data structure on 3-d sphere surfaces spanned by the first three ranked components; (c,d) projection of data from voxels within the selected ROIs onto the first two ranked components.*

***DISCUSSION:*** We have shown that the structure of resting-state correlation maps (with a seed signal from the PCC) is well represented by a few principal components, using the proposed ranking scheme. As compared to a conventional ranking scheme based only on the singular values, the proposed scheme has the advantage of identifying those components that contribute the most to a specified correlation map. Reflecting this advantage, the anti-correlated relation between the DMN and TPN is well captured by the map associated with the second ranked component. The relation between the DMN and TPN can also be seen in the projections of the data onto the top few ranked components (Fig 3.) Due to the tight relation between the first component and the global signal, the process of global signal regression is equivalent to viewing the data within the frame of reference of the upper hemisphere of the spheres in Fig. 3. Note the relation of these two regions within this frame of reference is not an artifact of our analysis, but represents a relationship inherent in the data. Overall, our results support the existence of an anti-correlated relation between the DMN and TPN and the general validity of global signal regression.

***REFERENCES:*** **[1]** *Murphy et al., Neuroimage, 2009.* **[2]** *Fox et al., J Neurophysiol, 2009.* **[3]** *Fox et al., PNAS, 2005.*