

# INVESTIGATING DIFFERENCES IN BRAIN FUNCTIONAL NETWORKS USING HIERARCHICAL COVARIATE-ADJUSTED INDEPENDENT COMPONENT ANALYSIS<sup>1</sup>

BY RAN SHI AND YING GUO

*Emory University*

Human brains perform tasks via complex functional networks consisting of separated brain regions. A popular approach to characterize brain functional networks in fMRI studies is independent component analysis (ICA), which is a powerful method to reconstruct latent source signals from their linear mixtures. In many fMRI studies, an important goal is to investigate how brain functional networks change according to specific clinical and demographic variabilities. Existing ICA methods, however, cannot directly incorporate covariate effects in ICA decomposition. Heuristic post-ICA analysis to address this need can be inaccurate and inefficient. In this paper, we propose a hierarchical covariate-adjusted ICA (hc-ICA) model that provides a formal statistical framework for estimating covariate effects and testing differences between brain functional networks. Our method provides a more reliable and powerful statistical tool for evaluating group differences in brain functional networks while appropriately controlling for potential confounding factors. We present an analytically tractable EM algorithm to obtain maximum likelihood estimates of our model. We also develop a subspace-based approximate EM that runs significantly faster while retaining high accuracy. To test the differences in functional networks, we introduce a voxel-wise approximate inference procedure which eliminates the need of computationally expensive covariance matrix estimation and inversion. We demonstrate the advantages of our methods over the existing method via simulation studies. We apply our method to an fMRI study to investigate differences in brain functional networks associated with post-traumatic stress disorder (PTSD).

**1. Introduction.** In the past decade, the field of neuroimaging has been moving toward a network-oriented view of brain functions. Functional magnetic resonance imaging (fMRI) is one of the most commonly used technologies to investigate brain functional networks (BFNs). In fMRI studies of functional connectivity, observed data are viewed as mixtures of signals generated from various BFNs. Each of these networks consists of a set of spatially distributed but functionally linked brain regions that present similar blood oxygenation level dependent (BOLD) signals during the scanning sessions. One of the major goals in fMRI

---

Received November 2014; revised April 2016.

<sup>1</sup>Supported in part by the National Institute of Mental Health of the National Institutes of Health under Award Numbers ROI MH105561 and R01MH079448 and by the Emory University Research Committee (URC/ACTSI grant.)

*Key words and phrases.* fMRI, blind source separation, brain functional networks, EM algorithm, subspace concentration.

data analysis is to decompose the observed fMRI data to identify the underlying functional networks and characterize their spatial distributions and temporal dynamics. Independent component analysis (ICA) has become the most widely used tool in the neuroscience community to investigate these functional networks. As a special case of blind source separation, ICA can separate observed fMRI signals into linear combinations of latent spatial source signals that are statistically as independent as possible. Each of these latent components correspond to a BFN.

Compared with alternative network methods, ICA has several major advantages. As a multivariate approach, ICA can jointly model the relationships among multiple voxels, and hence provide a tool for investigating whole brain connectivity. Unlike second-order statistical methods such as PCA, ICA takes into account higher-order statistics, and the spatial statistical independence assumption of ICA is well supported by the sparse nature in typical fMRI activation patterns [Beckmann and Smith (2004), Calhoun et al. (2001)]. Furthermore, ICA is a fully data-driven approach that does not need a priori temporal or spatial models. This makes ICA an important tool for analyzing resting-state fMRI where there is no experimental paradigm [Beckmann et al. (2005)]. Finally, compared with other whole brain connectivity methods such as graphical models, a distinctive feature of ICA is that it can partition the whole brain into functionally coherent networks.

ICA was initially used for analyzing single-subject fMRI data to either characterize spatially independent brain networks, that is, spatial ICA [Beckmann and Smith (2005), Biswal and Ulmer (1999), Calhoun et al. (2001), Mckeown et al. (1998)], or separate independent time courses, that is, temporal ICA [Lee et al. (2011)]. In this paper, we consider spatial ICA which is more suitable for our fMRI data example. Denote by  $\mathbf{Y}$  the  $T \times V$  fMRI data matrix for one subject, where  $T$  is the number of fMRI scans and  $V$  is the number of voxels in the 3D brain image acquired during each scan. Each row of  $\mathbf{Y}$  represents a vectorized 3D image. Classical noise-free spatial ICA decomposes the observed fMRI data for one subject as  $\mathbf{Y}_{T \times V} = \mathbf{A}_{T \times q} \mathbf{S}_{q \times V}$ , where  $q$  is the total number of source signals. Each row of  $\mathbf{S}$  represents a vectorized 3D image of a spatial source signal. The  $q$  spatial source signals are assumed to be statistically independent, and hence are called independent components (ICs).  $\mathbf{A}$  is the mixing matrix, the columns of which determine the temporal dynamics of the ICs.

To decompose multi-subject fMRI data, ICA has been extended for group analysis, which is referred to as group ICA [Calhoun et al. (2001)]. One commonly used group ICA framework in fMRI analysis is the temporal concatenation group ICA (TC-GICA). In TC-GICA, the  $T \times V$  fMRI data matrices from  $N$  subjects are stacked in the temporal domain to form a tall  $TN \times V$  group data matrix. The concatenated group data are then decomposed into the product of a  $TN \times q$  group mixing matrix and a  $q \times V$  spatial source matrix with independent rows. Many existing group ICA methods [Beckmann and Smith (2005), Calhoun et al. (2001), Guo (2011), Guo and Pagnoni (2008)] were developed under the TC-GICA framework. A notable restriction of the TC-GICA models is that they assume the same spatial distribution of BFNs across subjects.

In recent years, neuroscience literature has provided evidence that BFNs can vary considerably due to subjects' clinical, biological and demographic characteristics. For example, neuroimaging studies have shown that neural activity and connectivity in specific functional networks are significantly associated with mental disorders and their responses to treatment regimes [Anand et al. (2005), Chen et al. (2007), Greicius et al. (2007), Sheline et al. (2009)]. Other studies have found activity patterns in major functional networks vary with demographic factors including age and gender [Cole et al. (2010), Quiton and Greenspan (2007)]. These findings call for statistical methods that can quantify the effects of subjects' characteristics on the BFNs and can evaluate the differences in BFNs between subject groups (e.g., diseased vs. normal).

The data example in this paper demonstrates the need for incorporating covariate effects when investigating differences in BFNs using group ICA. The data come from a post-traumatic stress disorder (PTSD) study conducted by the Grady Memorial Hospital and Emory University in Atlanta. This PTSD study is one of the largest NIH sponsored ongoing research projects on PTSD in an urban population. In our data example, a subgroup of African-American female subjects from the Grady PTSD study were recruited for fMRI acquisitions. One of the main goals of the fMRI study is to investigate PTSD-related differences in BFNs. A major challenge in achieving this goal is that the Grady PTSD study is an observational study in which the PTSD positive and PTSD negative groups were not matched on their demographic or clinical variables. Therefore, between-group comparisons are prone to be biased due to potential confounding factors. For example, it is well known that PTSD is often co-morbid with other mental problems such as major depression disorders (MDD) [Campbell et al. (2007), Kessler et al. (1995)]. Another example is that the heterogeneity of age distribution between the two PTSD groups can affect the BFNs, according to findings in Bullmore and Sporns (2009). Thus, to assess PTSD-related brain network alterations, it is necessary to adjust for these potential confounding factors.

Existing group ICA methods, which often assume the same spatial patterns of BFNs across subjects, do not directly incorporate covariate information in the ICA decomposition. Currently, differences in brain functional networks and their associations with subjects' covariates are assessed through two kinds of heuristic approaches. The first approach is through conducting single-subject ICA separately on each subject's, selecting matched ICs and then performing group analysis on the selected subject-level IC maps [Greicius et al. (2007)]. A major problem with this approach is that it is often challenging to match ICs across subjects since ICA results are only identifiable up to a permutation of the ICs. Furthermore, since most ICA algorithms are stochastic [Himberg, Hyvärinen and Esposito (2004)], the levels of ICs extracted from separate ICA runs for different subjects are often not comparable to each other. The second approach is via two-stage analysis based on TC-GICA. Two representative methods in this category are the back construction [Calhoun et al. (2001)] and dual regression [Beckmann et al. (2009), Filippini

et al. (2009)]. These methods first perform TC-GICA to extract common IC maps at the group level and then reconstruct subject-specific IC maps by post-ICA steps. The covariate effects are evaluated via secondary hypothesis testing or regression analysis on the reconstructed subject-specific maps. These methods do not take into account the random variabilities introduced in reconstructing subject-specific IC maps, which could lead to loss of accuracy and efficiency in estimating and testing covariate effects on functional networks.

In this paper, we propose a hierarchical covariate-adjusted ICA (hc-ICA) model that directly incorporates covariate effects in group ICA decomposition to investigate differences in BFNs. The hc-ICA model decomposes each subject's fMRI data into linear mixtures of subject-specific spatial source signals (ICs). These distinct subject-specific ICs are then modeled in terms of population-level baseline source signals, covariate effects and between-subject random variabilities. To the best of our knowledge, hc-ICA is the first model-based group ICA method that captures variabilities in BFNs due to covariates effects. Compared with existing group ICA methods, hc-ICA has several advantages. hc-ICA is more accurate and powerful in terms of detecting brain network differences due to the primary effects of interest, such as disease status, while controlling for other confounding factors. For example, application of hc-ICA to the Grady PTSD study reveals important differences in the brain networks of the two PTSD groups, while the existing group ICA method cannot detect these differences effectively. Results from our simulation studies also corroborate that hc-ICA has better performance than the existing method in terms of both estimation accuracy and statistical power. In addition, hc-ICA can provide model-based estimation or prediction of brain functional networks for subpopulations defined by specific clinical or demographic characteristics. This will promote understandings of both commonalities and distinctions in brain networks across various subgroups within a study cohort.

Our hc-ICA model is developed under the hierarchical probabilistic ICA modeling framework first introduced in Guo and Tang (2013) which proposed a hierarchical random effects ICA model for relaxing the spatial homogeneity assumption in TC-GICA. The hc-ICA model, as well as its estimation and inference procedures, provides several important contributions to hierarchical ICA modeling. First, hc-ICA provides the first statistical framework to evaluate how subjects' demographic and clinical characteristics can affect their brain functional networks. This is not available in any existing group ICA methods, including the random effect model in Guo and Tang (2013). Second, we propose a novel subspace-based approximate expectation-maximization (EM) algorithm for obtaining maximum likelihood estimates. The approximate EM algorithm scales linearly with the number of ICs, which is significantly faster than the exponential growth of the exact EM algorithms used by Guo (2011) and Guo and Tang (2013). Third, our work provides an efficient voxel-wise approximate inference procedure for testing covariate effects on ICs. Such statistical inference procedures are not available in existing group ICA methods, including Guo and Tang (2013).

We introduce in Section 2 the hc-ICA framework including data preprocessing, model specification, estimation and inference. Section 4 presents an analysis of the PTSD dataset using our method. Section 3 reports simulation results for comparing hc-ICA to the existing TC-GICA two-stage method, comparing the subspace-based EM to the exact EM algorithms and comparing the proposed inference method to the existing TC-GICA two-stage method for testing covariate effects. Conclusions and discussions are presented in Section 5. Derivations, proofs, additional simulation studies and details for the analysis of the PTSD data are provided in the web Supplementary Materials [Shi and Guo (2016)].

**2. Methods.** This section introduces the hc-ICA framework, which includes the preprocessing step, the hc-ICA model, estimation algorithms and the inference procedure.

*2.1. Preprocessing prior to ICA.* Prior to an ICA algorithm, some preprocessing steps such as centering, dimension reduction and whitening of the observed data are usually performed to facilitate the subsequent ICA decomposition [Hyvärinen, Karhunen and Oja (2001)]. Suppose that the fMRI study consists of  $N$  subjects. For each subject, the fMRI signal is acquired at  $T$  time points across  $V$  voxels. Let  $\tilde{\mathbf{y}}_i(v) \in \mathbb{R}^T$  be the centered time series recorded for subject  $i$  at voxel  $v$ ;  $\tilde{\mathbf{Y}}_i = [\tilde{\mathbf{y}}_i(1), \dots, \tilde{\mathbf{y}}_i(V)]$  is the  $T \times V$  fMRI data matrix for subject  $i$ .

Under the paradigm of group ICA, we perform the following dimension reduction and whitening procedure on the original fMRI data: for  $i = 1, \dots, N$ ,

$$(1) \quad \mathbf{Y}_i = (\mathbf{\Lambda}_{i,q} - \tilde{\sigma}_{i,q}^2 \mathbf{I}_q)^{-\frac{1}{2}} \mathbf{U}'_{i,q} \tilde{\mathbf{Y}}_i,$$

where  $\mathbf{U}_{i,q}$  and  $\mathbf{\Lambda}_{i,q}$  contain the first  $q$  eigenvectors and eigenvalues based on the singular value decomposition of  $\tilde{\mathbf{Y}}_i$ . The residual variance,  $\tilde{\sigma}_{i,q}^2$ , is the average of the smallest  $T - q$  eigenvalues that are not included in  $\mathbf{\Lambda}_{i,q}$  representing the variability in  $\tilde{\mathbf{Y}}_i$  that is not accounted by the first  $q$  components. The parameter  $q$ , which is the number of ICs, can be determined using the Laplace approximation method [Minka (2000)]. Throughout the rest of our paper, we will present the model and methodologies based on the preprocessed data  $\mathbf{Y}_i = [\mathbf{y}_i(1), \dots, \mathbf{y}_i(V)]$  ( $i = 1, \dots, N$ ), which are  $q \times V$  matrices.

*2.2. A hierarchical covariate-adjusted ICA model (hc-ICA).* In this section, we present a hierarchical covariate-adjusted ICA (hc-ICA) model for evaluating covariate effects on brain functional networks using multi-subject fMRI data. The first-level model of hc-ICA decomposes a subject's observed fMRI signals into a product of subject-specific spatial source signals and a temporal mixing matrix to capture between-subject variabilities in the spatio-temporal processes in the functional networks. We include a noise term in this ICA model to account for residual variabilities in the fMRI data that are not explained by the extracted ICs, which is

known as probabilistic ICA [Beckmann and Smith (2004)]. To be specific, the first level of hc-ICA is defined as

$$(2) \quad \mathbf{y}_i(v) = \mathbf{A}_i \mathbf{s}_i(v) + \mathbf{e}_i(v),$$

where  $\mathbf{s}_i(v) = [s_{i1}(v), \dots, s_{iq}(v)]'$  is a  $q \times 1$  vector with  $s_{i\ell}(v)$  representing the spatial source signal of the  $\ell$ th IC (i.e., brain functional network) at voxel  $v$  for subject  $i$ . The  $q$  elements of  $\mathbf{s}_i(v)$  are assumed to be independent and non-Gaussian.  $\mathbf{A}_i$  is the  $q \times q$  mixing matrix for subject  $i$  which mixes  $\mathbf{s}_i(v)$  to generate the observed (preprocessed) fMRI data. Since  $\mathbf{Y}_i$  is whitened, the mixing matrix,  $\mathbf{A}_i$ , should be orthogonal [Hyvärinen and Oja (2000)].  $\mathbf{e}_i(v)$  is a  $q \times 1$  vector that represents the noise in the subject's data and  $\mathbf{e}_i(v) \sim N(\mathbf{0}, \mathbf{E}_v)$  for  $v = 1, \dots, V$ . The noise term is assumed to be independent across voxels because the spatial correlation across voxels is modeled by the spatial source signals [Beckmann and Smith (2004), Guo (2011), Hyvärinen, Karhunen and Oja (2001)]. Prior to ICA, preliminary analysis such as pre-whitening [Bullmore et al. (1996)] can be performed to remove temporal correlations in the noise term and to standardize the variability across voxels. Therefore, following previous work [Beckmann and Smith (2004, 2005), Guo (2011), Guo and Pagnoni (2008)], we assume that the covariance for the noise term is the same across voxels and isotropic, that is  $\mathbf{E}_v = v_0^2 \mathbf{I}_q$ . The ICA decomposition in the first-level model is a spatial ICA model since statistical independence is assumed for the spatial maps of brain functional networks. For fMRI data, spatial ICA has become dominant because the spatial independence assumption is well suited to the spatial patterns of most cognitive activation paradigms [McKeown et al. (1998)].

At the second level of hc-ICA, we further model subject-specific spatial source signals  $\mathbf{s}_i(v)$  as a combination of the population-level source signals, the covariate effects and additional between-subject random variabilities:

$$(3) \quad \mathbf{s}_i(v) = \mathbf{s}_0(v) + \boldsymbol{\beta}(v)' \mathbf{x}_i + \boldsymbol{\gamma}_i(v),$$

where  $\mathbf{s}_0(v) = [s_{01}(v), \dots, s_{0q}(v)]'$  is the population-level spatial source signals of the  $q$  statistically independent and non-Gaussian ICs;  $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]'$  is the  $p \times 1$  covariate vector containing subject-specific characteristics such as the treatment or disease group, demographic variables and biological traits;  $\boldsymbol{\beta}(v)$  is a  $p \times q$  matrix where the element  $\beta_{k\ell}(v)$  ( $k = 1, \dots, p, \ell = 1, \dots, q$ ) in  $\boldsymbol{\beta}(v)$  captures the effect of the  $k$ th covariate on the  $\ell$ th functional network at voxel  $v$ ;  $\boldsymbol{\gamma}_i(v)$  is a  $q \times 1$  vector reflecting the random variabilities among subjects after adjusting for covariate effects. We assume  $\boldsymbol{\gamma}_i(v) \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{D})$ , where  $\mathbf{D} = \text{diag}(v_1^2, \dots, v_q^2)$ . IC-specific variances specified in  $\mathbf{D}$  allow us to accommodate different levels of between-subject random variability.

2.3. *Source signal distribution assumptions.* Following Guo (2011) and Guo and Tang (2013), we choose mixtures of Gaussians (MoG) as our source distribution model for the population-level spatial source signals,  $\mathbf{s}_0(v)$ , in (3). MoG have several desirable properties for modeling fMRI signals. Within each BFN, only a small percentage of locations in the brain are activated or deactivated, whereas most brain areas exhibit background fluctuations [Biswal and Ulmer (1999)]. MoG are well suited to model such mixed patterns. Furthermore, MoG can capture various types of non-Gaussian signals [Kostantinos (2000), Xu et al. (1997)] and also offer tractable likelihood-based estimation [McLachlan and Peel (2004)].

Specifically, for  $\ell = 1, \dots, q$  we assume that

$$(4) \quad s_{0\ell}(v) \sim \text{MoG}(\boldsymbol{\pi}_\ell, \boldsymbol{\mu}_\ell, \boldsymbol{\sigma}_\ell^2), \quad v = 1, \dots, V,$$

where  $\boldsymbol{\pi}_\ell = [\pi_{\ell,1}, \dots, \pi_{\ell,m}]'$  with  $\sum_{j=1}^m \pi_{\ell,j} = 1$ ,  $\boldsymbol{\mu}_\ell = [\mu_{\ell,1}, \dots, \mu_{\ell,m}]'$  and  $\boldsymbol{\sigma}_\ell^2 = [\sigma_{\ell,1}^2, \dots, \sigma_{\ell,m}^2]'$ ;  $m$  is the number of Gaussian components in MoG. The probability density of MoG( $\boldsymbol{\pi}_\ell, \boldsymbol{\mu}_\ell, \boldsymbol{\sigma}_\ell^2$ ) is  $\sum_{j=1}^m \pi_{\ell,j} g(s_{0\ell}(v); \mu_{\ell,j}, \sigma_{\ell,j}^2)$ , where  $g(\cdot)$  is the p.d.f. of the (multivariate) Gaussian distribution. In fMRI applications, mixtures of two to three Gaussian components are sufficient to capture the distribution of fMRI spatial signals, with the different Gaussian components representing the background fluctuation and the negative or positive fMRI BOLD effects, respectively [Beckmann and Smith (2004), Guo and Pagnoni (2008)]. Without loss of generality, we denote by  $j = 1$  the background fluctuation state throughout the rest of the paper.

To facilitate derivations in models involving MoG, latent state variables are often used [McLachlan and Peel (2004)]. Here we define latent states  $\mathbf{z}(v) = [z_1(v), \dots, z_q(v)]'$  at voxel  $v$  as follows. For  $\ell = 1, \dots, q$ ,  $z_\ell(v)$  takes a value in  $\{1, \dots, m\}$  with probability  $p[z_\ell(v) = j] = \pi_{\ell,j}$  for  $j = 1, \dots, m$ . Conditional on  $\mathbf{z}(v)$ , we can rewrite our source distribution model as

$$(5) \quad \mathbf{s}_0(v) = \boldsymbol{\mu}_{\mathbf{z}(v)} + \boldsymbol{\psi}_{\mathbf{z}(v)},$$

where  $\boldsymbol{\mu}_{\mathbf{z}(v)} = [\mu_{1,z_1(v)}, \dots, \mu_{q,z_q(v)}]'$  and  $\boldsymbol{\psi}_{\mathbf{z}(v)} = [\psi_{1,z_1(v)}, \dots, \psi_{q,z_q(v)}]'$ ;  $\boldsymbol{\psi}_{\mathbf{z}(v)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{z}(v)})$  with  $\boldsymbol{\Sigma}_{\mathbf{z}(v)} = \text{diag}(\sigma_{1,z_1(v)}^2, \dots, \sigma_{q,z_q(v)}^2)$ .

2.4. *Maximum likelihood estimation.* We develop a unified maximum likelihood estimation method via the EM algorithm that simultaneously estimates all parameters in the hc-ICA model. Based on (2), (3) and (5), the complete data log-likelihood for the hc-ICA model is

$$(6) \quad l(\Theta; \mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{Z}) = \sum_{v=1}^V l_v(\Theta; \mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{Z}),$$

where  $\mathcal{Y} = \{\mathbf{y}_i(v) : i = 1, \dots, N; v = 1, \dots, V\}$ ,  $\mathcal{X} = \{\mathbf{x}_i : i = 1, \dots, N\}$ ,  $\mathcal{S} = \{\mathbf{s}_i(v) : i = 0, \dots, N, v = 1, \dots, V\}$  and  $\mathcal{Z} = \{\mathbf{z}(v) : v = 1, \dots, V\}$ ; the parameters

are  $\Theta = \{\{\boldsymbol{\beta}(v)\}, \{\mathbf{A}_i\}, \mathbf{E}, \mathbf{D}, \{\boldsymbol{\pi}_\ell\}, \{\boldsymbol{\mu}_\ell\}, \{\boldsymbol{\sigma}_\ell^2\} : i = 1, \dots, N, v = 1, \dots, V, \ell = 1, \dots, m\}$ . The detailed expression for the complete data log-likelihood function at each voxel  $v$  is

$$\begin{aligned}
 l_v(\Theta; \mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{Z}) &= \sum_{i=1}^N [\log g(\mathbf{y}_i(v); \mathbf{A}_i \mathbf{s}_i(v), \mathbf{E}) \\
 (7) \qquad \qquad \qquad &+ \log g(\mathbf{s}_i(v); \mathbf{s}_0(v) + \boldsymbol{\beta}(v)' \mathbf{x}_i, \mathbf{D})] \\
 &+ \log g(\mathbf{s}_0(v); \boldsymbol{\mu}_{\mathbf{z}(v)}, \boldsymbol{\Sigma}_{\mathbf{z}(v)}) + \sum_{\ell=1}^q \log \pi_{l, \mathbf{z}_\ell(v)}.
 \end{aligned}$$

2.4.1. *The exact EM algorithm.* We first present an exact EM which has an explicit E-step and M-step to obtain ML estimates for the parameters in hc-ICA.

*E-step:* In the E-step, given the parameter estimates  $\hat{\Theta}^{(k)}$  from the last step, we derive the conditional expectation of the complete data log-likelihood given the observed data as follows:

$$(8) \qquad Q(\Theta \mid \hat{\Theta}^{(k)}) = \sum_{v=1}^V E_{\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v)} [l_v(\Theta; \mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{Z})],$$

where  $\mathbf{y}(v) = [\mathbf{y}_1(v)', \dots, \mathbf{y}_N(v)']'$  represents the group data vector from the  $N$  subjects at voxel  $v$ ,  $\mathbf{s}(v) = [\mathbf{s}_1(v)', \dots, \mathbf{s}_N(v)', \mathbf{s}_0(v)']'$  is the vector containing latent source signals on both the population and individual level. The detailed definition of  $Q(\Theta \mid \hat{\Theta}^{(k)})$  is available in Section 1 of the web Supplementary Materials. The evaluation of  $Q(\Theta \mid \hat{\Theta}^{(k)})$  relies on obtaining  $p[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  as well as its marginal distributions, which consist of the following three steps. First, we determine  $p[\mathbf{s}(v) \mid \mathbf{z}(v), \mathbf{y}(v); \hat{\Theta}^{(k)}]$ , which is a multivariate Gaussian distribution. Second, we evaluate the probability mass functions  $p[\mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  through an application of Bayes' theorem. We finally obtain  $p[\mathbf{s}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  by convolving the distributions derived in the previous two steps. More details can be found in Section 2 of the Supplementary Material.

Given these probability distributions, we can derive the analytical forms for the conditional expectation in (8). For illustration purposes, two main quantities of interest in (8) are given as follows:

$$\begin{aligned}
 E[\mathbf{s}(v) \mid \mathbf{y}(v); \Theta] &= \sum_{\mathbf{z}(v) \in \mathcal{R}} p[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta] E[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \Theta], \\
 E[\mathbf{s}(v)^{\otimes 2} \mid \mathbf{y}(v); \Theta] &= \sum_{\mathbf{z}(v) \in \mathcal{R}} p[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta] E[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \Theta]^{\otimes 2} \\
 &+ \sum_{\mathbf{z}(v) \in \mathcal{R}} p[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta] \text{Var}[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \Theta],
 \end{aligned}$$

where  $\mathcal{R}$  represents the set of all possible values of  $\mathbf{z}(v)$ , that is,  $\mathcal{R} = \{\mathbf{z}^r\}_{r=1}^{m^q}$  where  $\mathbf{z}^r = [z_1^r, \dots, z_q^r]'$  and  $z_\ell^r \in \{1, \dots, m\}$  for  $\ell = 1, \dots, q$ ; the notation  $\mathbf{a}^{\otimes 2}$  for vector  $\mathbf{a}$  stands for  $\mathbf{a}\mathbf{a}'$ .

Based on the results presented above, our E-step is fully tractable without the need for iterative numerical integrations.

*M-step:* In the M-step, we update the current parameter estimates  $\hat{\Theta}^{(k)}$  to

$$(9) \quad \hat{\Theta}^{(k+1)} = \arg \max_{\Theta} Q(\Theta \mid \hat{\Theta}^{(k)}).$$

We have derived explicit formulas for all parameter updates. The updating rules are provided in Section 3 of our Supplementary Material.

The estimation procedure for the exact EM algorithm is summarized in Algorithm 1. See Sections 1–3 of the Supplementary Material for more details. After obtaining  $\hat{\Theta}$ , we can estimate the population- and individual-level source signals as well as their variability based on the mean and variance of their conditional distributions, that is,  $[\mathbf{s}_0(v) \mid \mathbf{y}(v); \hat{\Theta}]$  and  $[\mathbf{s}_i(v) \mid \mathbf{y}(v); \hat{\Theta}]$ . These conditional moments are directly obtainable from the E-step of our algorithm upon convergence and no separate post-ICA steps are required. As a referee pointed out, one could also estimate the source signals using the MAP estimator. As a major difference from TC-GICA, our subject-specific ICs  $\{\mathbf{s}_i\}$  are estimated simultaneously with the population-level IC  $\mathbf{s}_0$  instead of being reconstructed via post-ICA *ad hoc* approaches. Therefore, all the subject ICs are aligned to the population ICs in our model specification and estimation, which eliminates the need to match ICs across

**Algorithm 1** The Exact EM Algorithm

**Initial values:** Start with initial values  $\hat{\Theta}^{(0)}$  which can be obtained based on estimates from existing group ICA software.

**repeat**

**E-step:**

1. Determine  $p[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  and its marginals using the proposed three-step approach:

1.a Evaluate the multivariate Gaussian  $p[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \hat{\Theta}^{(k)}]$ ;

1.b Evaluate  $p[\mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ ;

1.c  $p[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}] = p[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \hat{\Theta}^{(k)}] \times p[\mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ ;

$$p[\mathbf{s}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}] = \sum_{\mathbf{z}(v) \in \mathcal{R}} p[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}];$$

2. Evaluate conditional expectations in  $Q(\Theta \mid \hat{\Theta}^{(k)})$ .

**M-step:**

Update  $\beta(v), \mathbf{A}_i, \pi_{\ell, j}, \mu_{\ell, j}, \sigma_{\ell, j}^2$ ;

Update the variance parameters  $\mathbf{D}, \mathbf{E}$ .

**until**  $\frac{\|\hat{\Theta}^{(k+1)} - \hat{\Theta}^{(k)}\|}{\|\hat{\Theta}^{(k)}\|} < \varepsilon$

difference subjects. This is an advantage of our approach over single-subject ICA-based analysis.

In fMRI analysis, researchers are often interested in thresholded IC maps to identify “significantly activated” voxels in each BFN. Following previous work [Guo (2011)], we propose a thresholding method based on the mixture distributions for this purpose (Section 6 of the Supplementary Material).

*2.4.2. The subspace-based approximate EM algorithm.* One major limitation of the exact EM algorithm is that its complexity increases exponentially with regard to the number of ICs. Specifically,  $\mathcal{O}(m^q)$  operations are required for the exact EM algorithm to complete. The main reason is that, at each voxel, the exact EM evaluates and sums the conditional distributions across the whole sample space  $\mathcal{R}$  of the latent state variables  $\mathbf{z}(v)$ , which has a cardinality of  $m^q$ . A standard way to alleviate this issue is through mean field variational approximation. This method has been used by Attias (1999, 2000) for single subject ICA and by Guo (2011) for TC-GICA. However, the variational method cannot be easily generalized to other models such as hierarchical ICA because the derivation of the variational approximate distributions depends heavily on the model specifications. In most cases, the estimates for the variational parameters do not have analytically tractable expressions and require extra numerical iterations, which sometimes causes convergence problems.

In this section, we propose a new approximate EM algorithm for solving MoG-based ICA models in fMRI studies. Compared with the exact EM that needs  $\mathcal{O}(m^q)$  operations, this new EM algorithm only requires  $\mathcal{O}(mq)$  operations. The key idea behind the approximate algorithm is that instead of considering the whole sample space  $\mathcal{R}$  of the latent state vector  $\mathbf{z}(v)$ , we only focus on a small subspace of  $\mathcal{R}$  in the algorithm. Theorem 1 provides the definition for the subspace and shows that, under certain conditions, the distribution of the latent state vectors is concentrated to the proposed subspace.

**THEOREM 1.** Define  $\mathcal{R} = \{\mathbf{z}^r = [z_1^r, \dots, z_q^r] : z_\ell^r = j \text{ with } j \in \{1, \dots, m\}, \ell = 1, \dots, q\}$  for  $r = 1, \dots, m^q$ , which is the domain of  $\mathbf{z}(v)$ . For all  $\mathbf{z}(v) \in \mathcal{R}$ , suppose that  $p[z_\ell(v) = j] = \pi_{\ell,j}$  and that  $p[\mathbf{z}(v) = \mathbf{z}^r] = \prod_{\ell=1}^q \pi_{\ell,z_\ell^r}$  [i.e.,  $\mathbf{z}(v)$  has independent elements]. Define  $\tilde{\mathcal{R}}$  as  $\tilde{\mathcal{R}} = \mathcal{R}_0 \cup \mathcal{R}_1$ , where  $\mathcal{R}_0 = \{\mathbf{z}^r \in \mathcal{R} : z_\ell^r = 1, \ell = 1, \dots, q\}$  and  $\mathcal{R}_1 = \{\mathbf{z}^r \in \mathcal{R} : \exists \text{ one and only one } \ell, \text{ s.t., } z_\ell^r \neq 1\}$ . Then, for any  $0 < \varepsilon < 1$ , if  $\pi_{\ell,1} > \frac{q}{q+\sqrt{\varepsilon}}$  for all  $\ell = 1, \dots, q$ , we have  $p[\mathbf{z}(v) \in \tilde{\mathcal{R}}] > 1 - \varepsilon$ .

The proof of the theorem is relegated to Section 4 of the Supplementary Material. Based on the above theorem, when  $\varepsilon \approx 0$ , that is,  $p[z_\ell(v) = 1] \approx 1$ , we have  $p[\mathbf{z}(v) \in \tilde{\mathcal{R}}] \approx 1$ . For fMRI data, the latent state  $j = 1$  in MOG model (4) corresponds to background fluctuation. Therefore, Theorem 1 implies that, for each IC, if latent states at most voxels are background fluctuation, the probability distribution of the latent state vector  $\mathbf{z}(v)$  in our hc-ICA will be mostly restricted to the

subspace  $\tilde{\mathcal{R}}$ . The condition in Theorem 1, that is,  $p[z_\ell(v) = 1] \approx 1$ , is supported by fMRI data because previous literature maintains that the fMRI spatial source signals are sparse across the brain [Daubechies et al. (2009), Mckeown et al. (1998)]; that is, within a specific BFN, that is, IC, most of the voxels exhibit background fluctuations with only a very small proportion of voxels being activated (or deactivated). The restriction of the latent states vector to the subspace  $\tilde{\mathcal{R}}$  implies that there is little chance for the same voxel to be activated in more than one IC. Biologically, this means that there is little overlapping in the activated regions across different BFNs, which has been supported by findings in the existing neuroimaging literature.

Based on this result, we propose a subspace-based approximate EM for our ICA model. The approximate EM follows similar steps as the exact EM. The main difference is that we restrict the conditional distribution of the latent state vector  $\mathbf{z}(v)$  to the subspace  $\tilde{\mathcal{R}}$  in the E-step and M-step; that is, the conditional expectations in the E-step are evaluated with a subspace-based approximate distribution  $\tilde{p}[\mathbf{z}(v) = \mathbf{z}^r \mid \mathbf{y}(v); \hat{\Theta}^{(k)}] = p[\mathbf{z}(v) = \mathbf{z}^r \mid \mathbf{y}(v); \hat{\Theta}^{(k)}] / \sum_{r \in \tilde{\mathcal{R}}} p[\mathbf{z}(v) = \mathbf{z}^r \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  where  $\mathbf{z}^r \in \tilde{\mathcal{R}}$  (see Section 5 of the Supplementary Material for a detailed treatment). Since the subspace  $\tilde{\mathcal{R}}$  has a cardinality of  $(m - 1)q + 1$ , the approximate EM only requires  $\mathcal{O}(mq)$  operations to complete. The concentration of measures to the subspace leads to the simplification in evaluating the conditional expectations in the E-step. For example,

$$(10) \quad \tilde{E}[\mathbf{s}(v) \mid \mathbf{y}(v); \Theta] = \sum_{\mathbf{z}(v) \in \tilde{\mathcal{R}}} \tilde{p}[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta] E[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \Theta],$$

which implies that, instead of summing over  $m^q$  latent states in  $\mathcal{R}$ , we only need to perform  $(m - 1)q + 1$  summations across the subspace of  $\tilde{\mathcal{R}}$ . The subspace-based EM also reduces computation time in the M-step. Specifically, when updating the parameters for the MoG source distribution model, we now use approximate conditional marginal moments. For example, as compared with the exact results, we use the following approximate moment when updating parameters for the Gaussian mixtures:

$$(11) \quad \begin{aligned} &\tilde{E}[s_{0\ell}(v) \mid z_\ell(v) = j, \mathbf{y}(v); \Theta] \\ &= \frac{\sum_{\mathbf{z}(v) \in \tilde{\mathcal{R}}^{(\ell,j)}} \tilde{p}[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta] E[s_{0\ell}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \Theta]}{\sum_{\mathbf{z}(v) \in \tilde{\mathcal{R}}^{(\ell,j)}} \tilde{p}[\mathbf{z}(v) \mid \mathbf{y}(v); \Theta]}, \end{aligned}$$

where  $\tilde{\mathcal{R}}^{(\ell,j)} = \{\mathbf{z}^r \in \tilde{\mathcal{R}} : z_\ell^r = j\}$ , whose cardinality equals  $(m - 1)(q - 1) + 1$  if  $j = 1$  and 1 if  $j \neq 1$ . Comparing to its exact counterpart,  $\mathcal{R}^{(\ell,j)} = \{\mathbf{z}^r \in \mathcal{R} : z_\ell^r = j\}$ , which has a cardinality of  $m^{q-1}$ ; this can dramatically simplify the updating of  $\pi_{\ell,j}$ ,  $\mu_{\ell,j}$  and  $\sigma_{\ell,j}^2$  in the M-step. We summarize the approximate EM algorithm as Algorithm 2.

**Algorithm 2** The Subspace-based Approximate EM Algorithm

**Initial values:** Start with initial values  $\hat{\Theta}^{(0)}$ .

**repeat**

**E-step:**

1. Determine  $\tilde{p}[\mathbf{s}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  and its marginals as follows:
  - 1.a Evaluate the multivariate Gaussian  $p[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \hat{\Theta}^{(k)}]$ ;
  - 1.b Evaluate  $\tilde{p}[\mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$  on the subset  $\tilde{\mathcal{R}}$ ;
  - 1.c  $\tilde{p}[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}] = p[\mathbf{s}(v) \mid \mathbf{y}(v), \mathbf{z}(v); \hat{\Theta}^{(k)}] \times \tilde{p}[\mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ ;
$$p[\mathbf{s}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}] = \sum_{\mathbf{z}(v) \in \tilde{\mathcal{R}}} \tilde{p}[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v), \hat{\Theta}^{(k)}];$$
2. Evaluate conditional expectations in  $Q(\Theta \mid \hat{\Theta}^{(k)})$  with regard to  $\tilde{p}[\mathbf{s}(v), \mathbf{z}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ .

**M-step:**

Update  $\boldsymbol{\beta}(v)$ ,  $\mathbf{A}_i$ ,  $\pi_{\ell,j}$ ,  $\mu_{\ell,j}$ ;  $\sigma_{\ell,j}^2$  with the modification of replacing the exact conditional moments with their counterparts based on  $\tilde{p}[\mathbf{s}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ .  
Update  $\mathbf{D}$ ,  $\mathbf{E}$  with similar modifications of replacing the exact conditional moments with those based on  $\tilde{p}[\mathbf{s}(v) \mid \mathbf{y}(v); \hat{\Theta}^{(k)}]$ .

**until**  $\frac{\|\hat{\Theta}^{(k+1)} - \hat{\Theta}^{(k)}\|}{\|\hat{\Theta}^{(k)}\|} < \varepsilon$

*2.5. Inference for covariate effects in hc-ICA model.* Typically, statistical inference in maximum likelihood estimation is based on the inverse of the information matrix which is used to estimate the asymptotic variance-covariance matrix of the MLEs. Since Standard EM algorithms only provide parameter estimates, extensions to the EM algorithm have been developed to estimate the information matrix [Louis (1982), Meilijson (1989), Meng and Rubin (1991)]. However, these methods are computationally expensive for the proposed hc-ICA model due to the following reasons. First, the dimension of the information matrix for our model is huge due to the large number of parameters. Second, the ML estimates,  $\hat{\boldsymbol{\beta}}(v)$ ,  $v = 1, \dots, V$ , are not independent across voxels because they rely on the estimates of the same set of parameters such as the mixing matrices. Consequently, the information matrix of the hc-ICA model is ultra-high-dimensional and is not sparse, which makes it extremely challenging to invert.

In this section, we present a statistical inference procedure for covariate effects in the hc-ICA model. The proposed method is developed based on the connection between the hc-ICA and standard linear models. Our method aims to provide an efficient approach to estimate the asymptotic standard errors of the covariate effects at each voxel, that is,  $\hat{\boldsymbol{\beta}}(v)$  ( $v = 1, \dots, V$ ), by directly using the output from our EM algorithms. Specifically, we first rewrite the hc-ICA model in a nonhierarchical form by collapsing the two-level models in (2) and (3), and then multiplying

the orthogonal mixing matrix  $\mathbf{A}_i$  on both sides:

$$(12) \quad \mathbf{A}'_i \mathbf{y}_i(v) = \mathbf{s}_0(v) + \mathbf{X}_i \text{vec}[\boldsymbol{\beta}(v)'] + \boldsymbol{\gamma}_i(v) + \mathbf{A}'_i \mathbf{e}_i(v),$$

where  $\mathbf{X}_i = \mathbf{x}'_i \otimes \mathbf{I}_q$ . (12) can be re-expressed as follows:

$$(13) \quad \mathbf{y}^*_i(v) = \mathbf{X}_i \text{vec}[\boldsymbol{\beta}(v)'] + \boldsymbol{\zeta}_i(v),$$

where  $\mathbf{y}^*_i(v) = \mathbf{A}'_i \mathbf{y}_i(v) - \mathbf{s}_0(v)$ , and  $\boldsymbol{\zeta}_i(v) = \boldsymbol{\gamma}_i(v) + \mathbf{A}'_i \mathbf{e}_i(v)$  is a multivariate zero-mean Gaussian noise term. The model in (13) can be viewed as a general multivariate linear model at each voxel. The major distinction of (13) from the standard linear model is that the dependent variable  $\mathbf{y}^*(v)$  not only depends on the observed data  $\mathbf{y}(v)$  but also involves unknown parameters  $\mathbf{A}_i$  and latent variables  $\mathbf{s}_0(v)$ . Given the similarity between the hc-ICA and the standard linear models, we propose a variance estimator for  $\text{vec}[\hat{\boldsymbol{\beta}}(v)']$  following the linear model theory.

Note that, for a standard linear model, the asymptotic variance for  $\text{vec}[\hat{\boldsymbol{\beta}}(v)']$  can be obtained by

$$(14) \quad \text{Var}\{\text{vec}[\hat{\boldsymbol{\beta}}(v)']\} = \frac{1}{N} \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{W}(v)^{-1} \mathbf{X}_i \right)^{-1},$$

where  $\mathbf{W}(v)$  is the variance of the Gaussian noise in the linear model. Then the variance of  $\text{vec}[\hat{\boldsymbol{\beta}}(v)']$  can be estimated by plugging in an estimator for  $\mathbf{W}(v)$  in (14). Following this result, we consider a variance estimator for  $\text{vec}[\hat{\boldsymbol{\beta}}(v)']$  based on (14) by plugging in the empirical variance estimator  $\widehat{\mathbf{W}}(v) = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}^*_i(v) - \mathbf{X}_i \text{vec}[\hat{\boldsymbol{\beta}}(v)'])^{\otimes 2}$  [Seber and Lee (2012)]. Because the dependent variable  $\mathbf{y}^*(v)$  in (13) is not directly observable, we estimate  $\mathbf{y}^*_i(v)$  using the ML estimates from our EM algorithm as  $\widehat{\mathbf{y}}^*_i(v) = \widehat{\mathbf{A}}'_i \mathbf{y}_i(v) - \widehat{\mathbf{s}}_0(v)$ , where  $\widehat{\mathbf{s}}_0(v) = E[\mathbf{s}_0(v) \mid \mathbf{y}(v), \hat{\Theta}]$ ; that is, we modify the empirical variance estimator  $\widehat{\mathbf{W}}(v)$  as follows:

$$(15) \quad \widetilde{\mathbf{W}}(v) = \frac{1}{N} \sum_{i=1}^N (\widehat{\mathbf{A}}'_i \mathbf{y}_i(v) - E[\mathbf{s}_0(v) \mid \mathbf{y}(v), \hat{\Theta}] - \mathbf{X}_i \text{vec}[\hat{\boldsymbol{\beta}}(v)'])^{\otimes 2}.$$

Thus, our final variance estimator is  $\widehat{\text{Var}}\{\text{vec}[\hat{\boldsymbol{\beta}}(v)']\} = \frac{1}{N} (\sum_{i=1}^N \mathbf{X}'_i \widetilde{\mathbf{W}}(v)^{-1} \mathbf{X}_i)^{-1}$ .

Hypothesis testing on the covariate effects at each voxel can be performed by calculating the Z-statistics based on the proposed variance estimator and determining the corresponding  $p$ -values. Our method can test whether a certain covariate has significant effects on each of the BFNs at the voxel level. Based on the parametric Z-statistic maps, one can also apply standard multiple testing methods to control the family-wise error rate (FWER) or the false discovery rate (FDR) when testing the covariate effects within a BFN.

We note that our variance estimator may underestimate the variabilities in  $\hat{\boldsymbol{\beta}}(v)$  because it does not account for variabilities in estimating  $\mathbf{A}_i$  and  $\mathbf{s}_0(v)$ . As a result, when performing hypothesis testing, the actual type-I errors of the proposed test statistic can be relatively higher than the nominal level. We evaluate via simulation studies the performance of this proposed inference procedure in Section 4.

**3. Application to fMRI data from Grady PTSD study.** We applied the proposed method to the fMRI data collected from the Grady PTSD study. In this study, 92 African American women were recruited as part of a larger study conducted by the Grady Health System in Atlanta, GA. The Structured Clinical Interview for DSM-IV (SCID) [First (1995)] was administered to all subjects and was used for diagnosis of PTSD. In addition, participants completed the Beck Depression Inventory (BDI) [Beck, Steer and Carbin (1988), Beck et al. (1996)] for depression assessment. Out of the 92 subjects, 39 met a diagnosis of PTSD (PTSD+) and 53 did not meet the criteria for PTSD (PTSD-). The ages of these women at the time of study ranged from 20 to 62 (Mean  $\pm$  SD:  $35 \pm 12$  for PTSD+ group;  $39 \pm 12$  for PTSD- group; between-group test  $p = 0.1096$ ). The BDI depression scores were significantly higher in subjects with PTSD diagnosis (Mean  $\pm$  SD:  $16.6 \pm 9.0$  for PTSD+,  $8.3 \pm 7.8$  for PTSD-,  $p < 10^{-5}$ ).

*3.1. fMRI experimental design, image acquisition and preprocessing.* MRI scans were obtained in a 3.0T Siemens scanner. Participants received task stimuli through a flexible mirror attached to the radio frequency coil of the scanner. The mirror reflected a computer screen placed at the end of the MRI aperture. During all experiments, a white cross appeared on a black background for 500 msec; it was replaced by an X or an O “Go” signal for 1000 msec and followed by 750 msec of blank screen. On a response pad, the subjects pressed 1 for X and 2 for O. The subjects were instructed to respond to each trial as fast as they could unless the “NoGo” signal appeared (i.e., the background changed to red), in which case they should not press either button. The task comprised four runs separated by three 20’s rest periods. Each run contained 26 “Go”, 13 “NoGo” and 14 blank trials distributed randomly.

A T1-weighted high-resolution anatomical image was first acquired (176 sagittal slices, voxel size:  $1 \times 1 \times 1$  mm). During task administration, a series of T2-weighted functional images (echo-planar, 26 axial slices, voxel size:  $3.75 \times 3.75 \times 4$  mm, TR = 2.53 s, TE = 30 ms) were acquired. The fMRI data were converted and preprocessed using Statistical Parametric Mapping, version 5 (SPM5, Wellcome Trust Centre for Neuroimaging, London, UK: <http://www.fil.ion.ucl.ac.uk/spm/>). Functional volumes were corrected for slice acquisition timing differences and subject movement. The anatomical image was registered to the mean of the corrected functional images and subsequently spatially normalized to the MNI standard brain space. These normalization parameters from the MNI space were used for the functional images, which were smoothed with an 8 mm FWHM Gaussian kernel. Prior to ICA analysis, we performed additional preprocessing steps, including centering, dimension reduction and whitening as described in Section 2.1, on the fMRI data.

*3.2. Analysis and findings.* The preprocessed data from the 92 subjects were decomposed using the proposed hc-ICA model into 16 ICs (the number is chosen

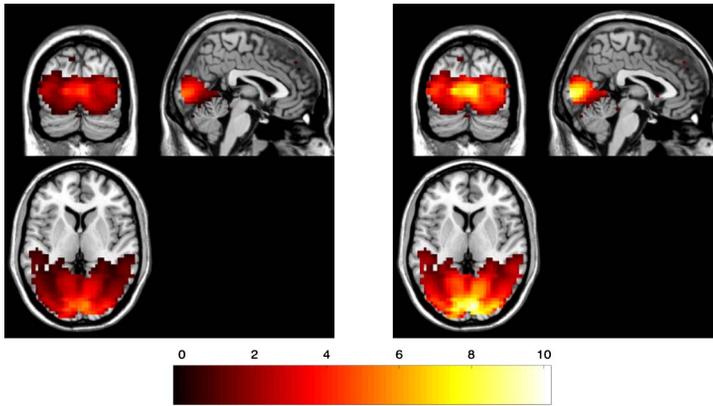
from the GIFT package: <http://mialab.mrn.org/software/gift/index.html>). To compare the networks between the two PTSD groups, we included PTSD diagnosis as the primary covariate of interest in the hc-ICA (PTSD $-$  = 0, PTSD $+$  = 1). We also included subject's age and BDI score as covariates to control for potential confounding effects. We estimated the parameters in the hc-ICA model using the subspace-based EM algorithm implemented in MATLAB, which is available at the authors' website. To ensure the validity of the results, we initialized the EM algorithm with 50 different starting values. The resulting estimates of the parameters were mostly close to each other. In this analysis, we reported the estimates corresponding to the highest observed data likelihood. More details about this robustness check are included in the Supplementary Materials.

Among the extracted ICs, we identified two components of particular interest. The first network had the highest positive temporal correlation with the task time series, which were the task series convolved with the hemodynamic response function (HRF). The spatial pattern of this network features the visual cortex, which responded to the visual stimuli presented in the Go/NoGo task. In Figure 1(A), we present the hc-ICA model-based estimates of the visual network for both the PTSD $-$  and PTSD $+$  groups. The two subpopulation maps were estimated at the median age (36 year old) and the median BDI scores (BDI = 10) to control for confounding effects. They were all thresholded based on the conditional probability of activation (Section 6 of the Supplementary Material). According to Figure 1(A), the PTSD $+$  group demonstrated stronger spatial source signals in the visual network as compared to the PTSD $-$  group with the same ages and BDI scores. It is worth noting that the existing group ICA methods cannot provide such model-based estimates of the brain networks for subpopulations defined by specific covariate patterns.

The second network of interest mainly includes the posterior cingulate cortex (PCC), the medial prefrontal cortex (mPFC) and the lateral parietal cortex (LPC). This network is known as the "default mode network", which shows increased activities during resting states and decreased activities during cognitive tasks [Raichle et al. (2001)]. Its temporal responses have the largest negative correlation with the task time series. Figure 1(B) presents the hc-ICA model-based estimates of this network for the two PTSD subpopulations (also adjusted at the median age and the median BDI score). Based on Figure 1(B), the default mode network of the PTSD $+$  women demonstrated stronger functional connectivity during the Go/NoGo tasks.

We then applied the proposed inference procedure to formally test the PTSD group differences in these two networks while controlling for the potential confounding effects from age and depression status. We also applied the method in Genovese, Lazar and Nichols (2002) to calculate the FDR corrected  $p$ -values for the between-group tests. For comparison, we also used a TC-GICA-based method, dual-regression ICA [Beckmann et al. (2009), Filippini et al. (2009)], to examine the group differences. Dual-regression ICA is one of the most commonly

(A): the task network featuring the visual cortex  
PTSD PTSD+



(B): the default mode network  
PTSD PTSD+

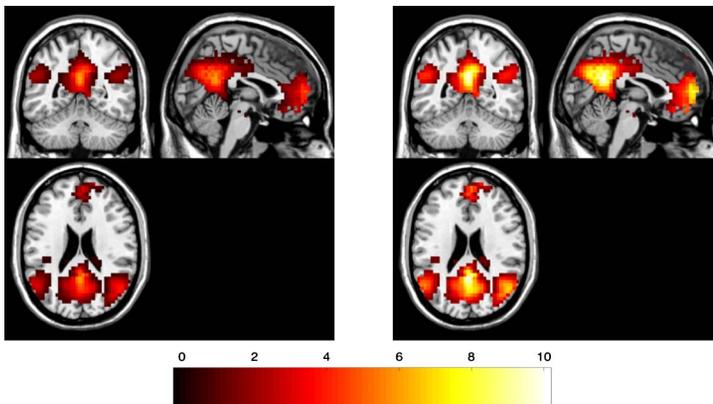


FIG. 1. The estimated subpopulational maps for the PTSD– and PTSD+ women at the median age (36 years old) and the median depression score ( $BDI = 10$ ): Panel (A) shows the estimates for the network featuring the visual cortex, which has the highest positive correlation with the task time series. Panel (B) shows the estimates for the default mode network, which has the largest negative correlation with the task time series. All IC maps are thresholded at the posterior probability of activation above 0.9. PTSD+ women show stronger IC signals in both networks.

used methods in the neuroimaging community for estimating subject-specific IC maps and performing between-group comparisons; see Reineberg et al. (2015), Smith et al. (2014) for some examples of its application. It is also adopted as a standard analytical tool by the well-known Human Connectome Project (<http://www.humanconnectomeproject.org/>). The dual-regression procedure typically tests group differences via permutation tests which cannot adjust for any confounding factors. To provide a fair comparison between hc-ICA and dual-re-

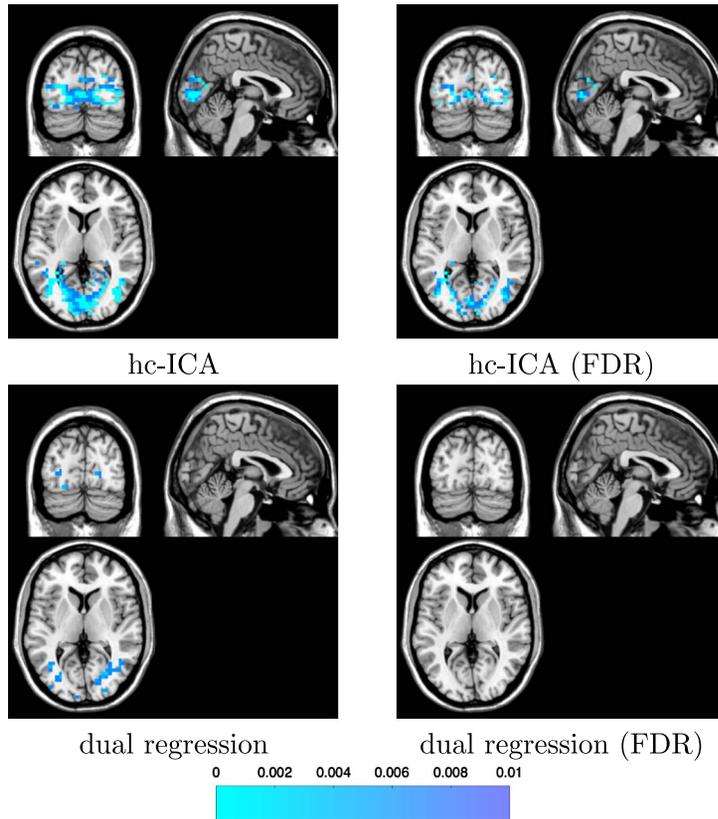


FIG. 2.  $p$ -values, thresholded below 0.01, for comparing the adjusted PTSD group differences ( $PTSD- < PTSD+$ ) in the task-related network: hc-ICA found increased spatial source signals at the central part of the visual cortex among  $PTSD+$  women, which remained significant after FDR control; dual regression found much less group differences in the network, all of which became insignificant with the FDR control.

gression, we performed an additional regression analysis on the reconstructed subject IC maps from dual regression using the same set of covariates as in hc-ICA and then tested PTSD group differences with adjustment for age and BDI.

The  $p$ -values for testing group differences in the task network, which features the visual cortex, are presented in Figure 2. Based on Figure 2, hc-ICA detected that  $PTSD+$  women showed significantly stronger spatial signals than  $PTSD-$  women in major parts of the visual network. This finding still held after FDR correction. These enhanced activities in visual cortex among PTSD subjects were previously reported in other fMRI studies involving visual stimuli [Hendler et al. (2003)]. The dual-regression analysis, however, found only a few differences in this network between the two groups, and all of them became insignificant after FDR control.

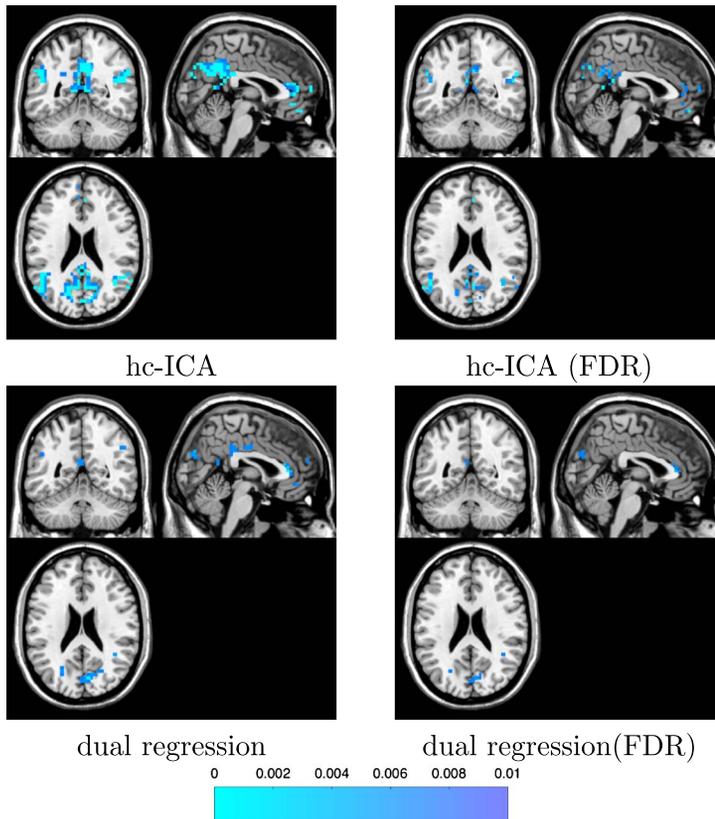


FIG. 3.  $p$ -values, thresholded below 0.01, for comparing the adjusted PTSD group differences (PTSD- < PTSD+) in the default mode network: hc-ICA finds stronger network activities across all the major regions of this network for PTSD+ women. Many of these identified voxels still appear after FDR control; dual-regression findings only discover a few differences in the PCC and mPFC regions.

Figure 3 shows the  $p$ -values regarding the group differences on the default mode network. Compared with PTSD- women, our method showed that the default mode network of PTSD+ women had significantly stronger source signals in all regions of the network as compared to the PTSD- women. This implies that functional connectivities among the brain regions within this network were stronger for the PTSD+ women, after controlling for subjects' age and depression status. Our results are consistent with recent findings in neuroscience literature that report abnormally high functional connectivity within the default mode network during both resting states and tasks for patients with mental disorders such as schizophrenia, depression and PTSD [Daniels et al. (2011), Greicius et al. (2007), Whitfield-Gabrieli et al. (2009)]. In comparison, dual regression only identified a few distinctions between the two groups in the PCC and mPFC regions but did not

detect any differences in the LPC part of the default mode network. After FDR correction, none of the findings based on dual regression remained significant.

**4. Simulation study.** We conducted three sets of simulation studies to (1) evaluate the performance of the proposed hc-ICA model as compared with the existing TC-GICA model, (2) to compare the accuracy of the subspace-based approximate EM algorithm vs. the exact EM22 and (3) to evaluate the performance of the proposed inference method for testing covariate effects based on hc-ICA.

*4.1. Simulation study I: Performance of the hc-ICA vs. TC-GICA.* In the first simulation study, we evaluate the performance of the proposed hc-ICA model compared with dual-regression ICA. We simulated fMRI data from three underlying source signals, that is,  $q = 3$ , and considered three sample sizes with the number of subjects of  $N = 10, 20, 40$ . For each source, we generated a 3D spatial map with the dimension of  $25 \times 25 \times 4$ , and the activated signals in each source are displayed in Figure 4(A). For spatial source signals, we first generated population-level spatial maps, that is,  $\{s_0(v)\}$ , as the activated signals plus Gaussian random variability of a variance of 0.5. We then generated two covariates for each subject with one being categorical [ $x_1 \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.5)$ ] and the other being continuous [ $x_2 \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(-1, 1)$ ]. The covariate effects maps, that is,  $\{\beta(v)\}$ , are presented in Figures 4(B1)–(B2) where the covariate effect parameters at each voxel took values from  $\{0, 1.5, 1.8, 2.5, 3.0\}$ . Additionally, we generated Gaussian subject-specific random effects, that is,  $\gamma_i(v)$ , and considered three levels of between-subject variability: low [ $\mathbf{D} = \text{diag}(0.1, 0.3, 0.5)$ ], medium [ $\mathbf{D} = \text{diag}(1.0, 1.2, 1.4)$ ] and high [ $\mathbf{D} = \text{diag}(1.8, 2.0, 2.5)$ ]. The subject-specific spatial source signals were then simulated as the linear combination of the population-level signals, covariate effects and subject-specific random effects. For temporal responses, each source signal had a time series of length of  $T = 200$  that was generated based on time courses from real fMRI data, and hence represented realistic fMRI temporal dynamics. We generated subject-specific time sources that had similar frequency features but different phase patterns [Guo (2011)], which represented temporal dynamics in resting-state fMRI signals. After simulating the spatial maps and time courses for the source signals, Gaussian background noise with a standard deviation of 1 ( $\mathbf{E} = \mathbf{I}_q$ ) was added to generate observed fMRI data.

We applied both hc-ICA and dual-regression ICA to the simulated data. The computational time was about 10 min ( $N = 10$ ), 16 min ( $N = 20$ ) and 25 min ( $N = 40$ ) for hc-ICA using the exact EM and around 45 s (approximately the same among all  $N$ 's) for dual regression for each simulated dataset using a desktop PC with an Intel i7 3.6 GHz quad core CPU. Following previous work [Beckmann and Smith (2005), Guo (2011)], we evaluate the performance of each method based on the correlations between the activation signals and estimated signals in both temporal and spatial domains. To compare the performance in estimating

## (A) Population-level IC maps

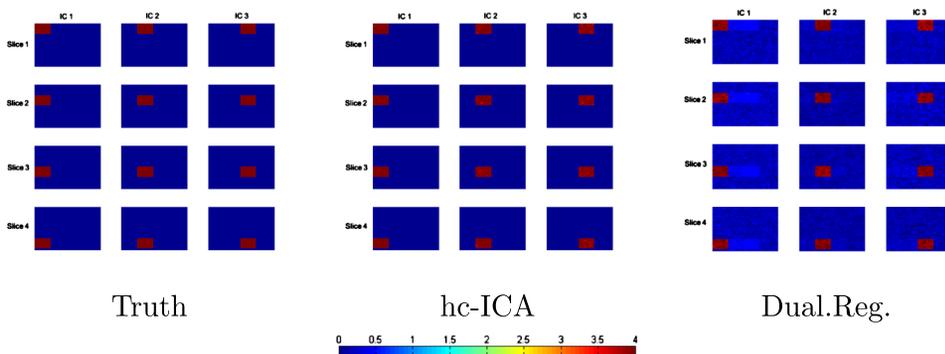
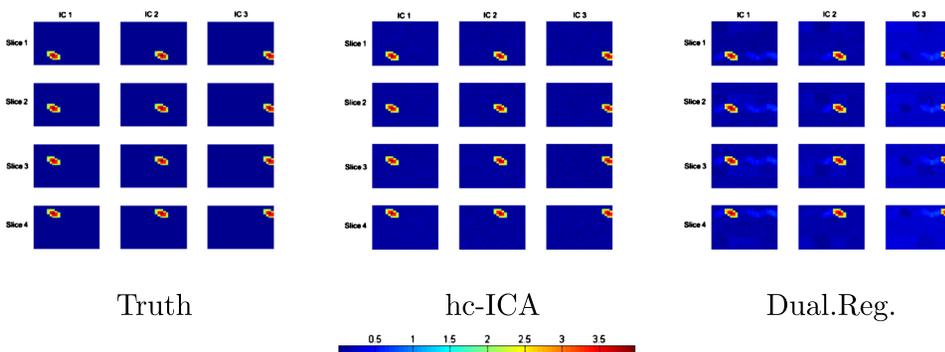
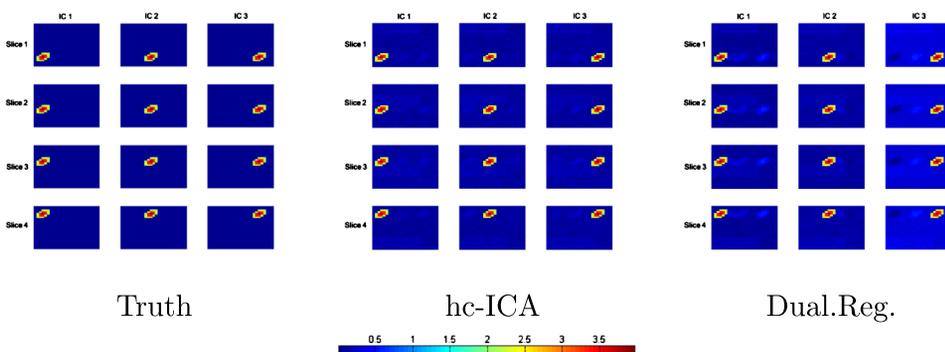
(B1) Covariate effects of the binary covariate,  $\boldsymbol{x}_1$ (B2) Covariate effects of the continuous covariate,  $\boldsymbol{x}_2$ 

FIG. 4. Comparison between our method and dual-regression ICA: truth, estimates from our model and estimates from dual regression ( $N = 10$ , between-subject variabilities are medium) are displayed based on 100 runs. All the images displayed are averaged across the 100 Monte Carlo datasets. Population-level spatial maps are shown in Figure 4(A). The results of dual-regression ICA are contaminated by the covariate effects. The results from our method are more accurate. Covariate effect estimates are shown in Figure 4(B1) and Figure 4(B2), respectively. The results of dual regression show clear mismatching, while our method provides accurate estimates.

the covariate effects, we report the mean squared errors (MSEs) of  $\hat{\boldsymbol{\beta}}(v)$  defined by  $\frac{1}{V} \sum_{v=1}^V \|\hat{\boldsymbol{\beta}}(v) - \boldsymbol{\beta}(v)\|_{\mathcal{F}}^2$  averaged across simulation runs. Here  $\|\cdot\|_{\mathcal{F}}$  is the Frobenius norm for a matrix. Since ICA recovery is permutation invariant, each estimated IC was matched with the original source with which it had the highest spatial correlation. We present the simulation results in Table 1. The results

TABLE 1

*Simulation results for comparing our hc-ICA method against dual-regression ICA based on 100 runs. Values presented are the mean and standard deviation of correlations between the true and estimated: subject-specific spatial maps, population-level spatial maps and subject-specific time courses. The mean and standard deviation of the MSE of the covariate estimates are also provided*

Btw-subj Var	Population-level spatial maps Corr. (SD)		Subject-specific spatial maps Corr. (SD)	
	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.
Low				
$N = 10$	0.982 (0.003)	0.956 (0.018)	0.984 (0.004)	0.945 (0.023)
$N = 20$	0.990 (0.002)	0.968 (0.014)	0.996 (0.002)	0.949 (0.008)
$N = 40$	0.992 (0.002)	0.976 (0.005)	0.996 (0.001)	0.956 (0.002)
Medium				
$N = 10$	0.942 (0.017)	0.914 (0.048)	0.943 (0.011)	0.882 (0.030)
$N = 20$	0.954 (0.002)	0.938 (0.034)	0.959 (0.004)	0.890 (0.016)
$N = 40$	0.961 (0.002)	0.949 (0.020)	0.968 (0.003)	0.893 (0.009)
High				
$N = 10$	0.833 (0.146)	0.740 (0.164)	0.894 (0.108)	0.689 (0.303)
$N = 20$	0.850 (0.129)	0.795 (0.143)	0.909 (0.084)	0.695 (0.281)
$N = 40$	0.871 (0.055)	0.809 (0.102)	0.928 (0.035)	0.705 (0.259)
Btw-subj Var.	Subject-specific time courses Corr. (SD)		Covariate effects Corr. (SD)	
	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.
Low				
$N = 10$	0.998 (0.001)	0.987 (0.010)	0.048 (0.019)	0.154 (0.055)
$N = 20$	0.998 (0.001)	0.995 (0.004)	0.021 (0.003)	0.127 (0.044)
$N = 40$	0.998 (0.001)	0.994 (0.004)	0.012 (0.001)	0.111 (0.030)
Medium				
$N = 10$	0.993 (0.010)	0.970 (0.028)	0.273 (0.088)	0.485 (0.151)
$N = 20$	0.998 (0.003)	0.976 (0.016)	0.117 (0.015)	0.285 (0.076)
$N = 40$	0.998 (0.002)	0.991 (0.008)	0.064 (0.005)	0.187 (0.041)
High				
$N = 10$	0.948 (0.021)	0.903 (0.045)	0.387 (0.157)	0.783 (0.325)
$N = 20$	0.978 (0.018)	0.925 (0.029)	0.224 (0.075)	0.532 (0.271)
$N = 40$	0.990 (0.015)	0.934 (0.022)	0.131 (0.056)	0.389 (0.198)

show that hc-ICA provides more accurate estimates for the source signals on both the population and subject levels. It leads to smaller mean squared errors in estimating the covariate effects. We also display the estimated population-level IC maps and the covariate effects maps from both methods in Figure 4. The hc-ICA shows, in Figure 4, much better performance in correctly detecting the true activation patterns and covariate effects for each IC. In comparison, the estimates of the population-level IC maps from dual regression show clear “cross-talk” between the ICs. Furthermore, the estimated covariate effects maps based on dual regression are noisier plus some mismatches across the ICs.

*4.2. Simulation study II: Performance of the approximate EM.* In the second simulation study, we compare the performance of the exact EM algorithm with the approximate EM for the hc-ICA model. We simulated fMRI data for ten subjects and considered three model sizes with the number of source signals of  $q = 3, 6, 10$ . The fMRI data were generated using methods similar to that in Simulation Study I with 10 subjects and low between-subject variabilities. We then fitted the proposed hc-ICA model using both the exact EM and the approximate EM. Results from Table 2 show that the accuracy of the subspace-based EM is comparable with regard to that of the exact EM in both the spatial and temporal domains and on both population and subject level. The major advantage of the subspace-based EM is that it was much faster than the exact EM. This advantage becomes more clear with the increase of the number of ICs. For  $q = 10$ , the subspace-based EM only uses about 2% computation time of the exact EM.

The convergence rates are the same between the two EM algorithms. We note that as  $q$  increased to 10, the convergence rates slightly decrease to 96%, which are lower than the EM algorithm for the TC-GICA model in Guo (2011). The main reason is that compared with the model in Guo (2011), which assumes common spatial maps across subjects, hc-ICA involves a significantly larger number of parameters and latent variables by incorporating subject-specific IC maps and spatially varying covariate effects on each IC. The dramatic increase in the number of parameters for hc-ICA with larger  $q$  leads to the slightly decreased convergence rate. In practice, if the EM algorithm experiences convergence issues due to a large number of ICs, one can consider using existing group ICA software to first identify the uninteresting ICs, linearly remove them from the observed data and then perform hc-ICA on the new data with a smaller number of ICs. This technique has been commonly used in ICA applications to remove artifact-related components [Griffanti et al. (2014), Tohka et al. (2008)].

*4.3. Simulation study III: Performance of the proposed inference procedures for covariate effects.* We examine the performance of our inference procedures for  $\hat{\beta}(v)$  in the third simulation study. We simulated fMRI datasets with two source signals and considered sample sizes of  $N = 20, 40, 80$ . We generated two covariates in the same manner as in Simulation Study I. To facilitate computation, we

TABLE 2

*Simulation results for comparing the subspace-based approximate EM and the exact EM based on 50 runs. The mean and standard deviation of correlations between the true and estimated spatial maps and time courses are presented. The mean and standard deviation of the MSE of the covariate estimates are also provided*

# of IC	Population-level spatial maps Corr. (SD)		Subject-specific spatial maps Corr. (SD)	
	Exact EM	Approx. EM	Exact EM	Approx. EM
$q = 3$	0.981 (0.003)	0.981 (0.001)	0.986 (0.004)	0.981 (0.002)
$q = 6$	0.980 (0.006)	0.980 (0.006)	0.985 (0.012)	0.981 (0.011)
$q = 10$	0.969 (0.022)	0.963 (0.020)	0.972 (0.027)	0.970 (0.022)
# of IC	Subject-specific time courses Corr. (SD)		Covariate effects Corr. (SD)	
	Exact EM	Approx. EM	Exact EM	Approx. EM
$q = 3$	0.998 (0.001)	0.998 (0.000)	0.048 (0.020)	0.048 (0.019)
$q = 6$	0.997 (0.003)	0.995 (0.002)	0.069 (0.024)	0.070 (0.022)
$q = 10$	0.992 (0.016)	0.992 (0.009)	0.105 (0.033)	0.112 (0.028)
# of IC	Time in minutes		Proportions of convergence	
	Exact EM	Approx. EM	Exact EM	Approx. EM
$q = 3$	9.91	5.22	100%	100%
$q = 6$	71.05	9.09	100%	100%
$q = 10$	860.10	19.02	96%	96%

generated images with the dimension of  $20 \times 20$ . The variance of between-subject random variabilities was 0.25 for both spatial source signals, and the within-subject variance was 0.4. We applied our hc-ICA method and dual-regression ICA for the simulated datasets and tested for the covariate effects using both methods. The hypotheses were  $H_0 : \beta_{k\ell}(v) = 0$  versus  $H_1 : \beta_{k\ell}(v) \neq 0$  at each voxel. Specifically, for hc-ICA, hypothesis tests were conducted for  $\beta(v)$  using the test proposed in Section 2.5. In comparison, the dual-regression method tested covariate effects by performing post-ICA regressions of the estimated subject-specific IC maps. We estimated the Type-I error rate with the empirical probabilities of not rejecting  $H_0$  at voxels such that  $\beta_{k\ell}(v) = 0$ . We also estimated the power of the tests with the empirical probabilities of rejecting  $H_0$  at voxels with nonzero values for the covariate effects, that is,  $\beta_{k\ell}(v) \in \{1.5, 1.8, 2.5, 3.0\}$ . We report the average of the Type-I error rates at various significance levels, as well as the powers with regard to different alternative hypothesis, in Table 3. According to Table 3, the type-I error rates from our inference method are always lower than those from dual-regression ICA. We do note that our Type-I error rates are slightly higher

TABLE 3

Simulation results for the inference of  $\beta(v)$  based on 1000 runs. Type-I errors are averaged across all voxels with  $\beta_{k\ell}(v) = 0$ ; powers are averaged across voxels having the same values of  $\beta_{k\ell}(v) \neq 0$

Size	$N = 20$		$N = 40$		$N = 80$	
	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.
<i>Type-I error analysis:</i>						
0.01	0.014	0.029	0.012	0.025	0.012	0.018
0.05	0.062	0.084	0.056	0.076	0.055	0.062
0.10	0.129	0.205	0.118	0.190	0.112	0.149
0.50	0.522	0.580	0.516	0.565	0.514	0.557
0.80	0.835	0.872	0.820	0.856	0.810	0.840
$\beta(v)$	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.	hc-ICA	Dual. Reg.
<i>Power analysis (test size: 0.05):</i>						
1.5	0.144	0.130	0.256	0.203	0.404	0.284
1.8	0.268	0.224	0.474	0.390	0.812	0.548
2.5	0.589	0.475	0.862	0.705	0.963	0.839
3.0	0.907	0.845	1.000	0.922	1.000	1.000

than the nominal level mainly due to the approximation in the inference procedure. From Table 3, we can also see that our method consistently demonstrates higher statistical power than dual-regression ICA. The results indicate that the proposed inference method based on hc-ICA provides more reliable and powerful inference about the covariate effects on the functional networks than the TC-GICA based dual-regression method.

**5. Discussion.** We propose a hierarchical covariate-adjusted ICA (hc-ICA) model to formally quantify and test differences in brain functional networks related to subjects' demographic, clinical and biological characteristics. Our hc-ICA approach can be applied to study brain networks in both task-related and resting state fMRI studies. We develop a maximum likelihood estimation method based on EM algorithms for hc-ICA. We use an efficient approximate procedure to make inferences about covariate effects in our model. Simulation studies show that our methods provide more accurate estimation and inference for covariate effects on brain networks than the widely used dual-regression method. Application of hc-ICA to the Grady PTSD Study reveals important differences in brain functional networks between PTSD+ and PTSD- African American women, after adjusting for their ages and depression scores.

One of the main challenges in statistical modeling of brain imaging is the heavy computation load. In this paper, we develop computationally efficient estimation and inference procedures for the proposed hc-ICA model. In particular, by exploiting sparsity in fMRI source signals, the subspace-based EM algorithm sig-

nificantly reduces the computational time via concentration of probability masses on a subspace of the latent multinomial variables. We show theoretically that the subspace-based approximate method is supported by the characteristics of fMRI signals. We demonstrate empirically that the approximate EM provides highly accurate results. The definition of the subspace implies that there is little overlap in the spatial distributions of fMRI source signals. This is supported by findings in the neuroscience literature which showed that brain functional networks are mostly separate [Beckmann et al. (2005), Smith et al. (2009)]. However, there are a few network hubs in the brain, consisting of a very small proportion of voxels, that may be involved in multiple networks. To investigate the performance of the subspace-based EM in this case, we have conducted additional simulation studies which generated data from overlapping source signals. Results show that the subspace EM still maintains good performance in recovering overlapping spatial signals.

Our hc-ICA model estimation is performed via a formal and unified maximum likelihood estimation which simultaneously estimates all parameters and latent variables in the model. By doing so, we improve the accuracy in estimating the brain networks on both population and individual level significantly; we also achieve higher statistical power in detecting differences in the networks. This holistic estimation approach does lead to heavier computation load compared with TC-GICA two-stage methods. The computation can be accelerated using several strategies. First, based on preliminary analysis of the data, we can identify ICs that are not of strong interest in a study, apply the standard procedure mentioned at the end of Section 4.2 to remove them from the data, and then apply the hc-ICA model to investigate group differences in the remaining ICs. Second, we can also apply standard multi-process/multi-thread computing techniques to reduce computational time at a large scale since most parts of our EM algorithm can be parallelized for each voxel (see the Supplementary Material, Sections 2–3, for details).

One potential extension to hc-ICA is to incorporate spatial dependence on modeling the spatially varying covariate effects  $\beta(v)$ . This can help increase the accuracy in detecting covariate-related network differences when they are spatially correlated. Furthermore, we can accommodate spatial dependence in the residual terms in both the first and the second level of the hc-ICA model, which may help improve the accuracy and efficiency of the proposed hc-ICA framework for investigating differences between functional networks.

**Acknowledgments.** We thank Dr. Tanja Jovanovic and Mr. Tim Ely from Emory University for processing and sharing the data. We also thank the Editor, the Associate Editor and the two referees for their valuable comments and constructive suggestions that have helped improve our manuscript significantly. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## SUPPLEMENTARY MATERIAL

**Supplement to the paper “Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis”** (DOI: [10.1214/16-AOAS946SUPP](https://doi.org/10.1214/16-AOAS946SUPP); .zip). This document presents the following contents: details about our EM algorithm and the approximate EM algorithm; the proof of Theorem 1; the criteria of selecting activating voxels within each brain network; the specification of initial guesses to our algorithm; additional simulation results with  $q = 10$ ; additional simulation results and theoretical results on the robustness of the approximate EM; additional results on the robustness of our method in real data analysis.

## REFERENCES

- ANAND, A., LI, Y., WANG, Y., WU, J., GAO, S., BUKHARI, L., MATHEWS, V. P., KALNIN, A. and LOWE, M. J. (2005). Activity and connectivity of brain mood regulating circuit in depression: A functional magnetic resonance study. *Biol. Psychiatry* **57** 1079–1088.
- ATTIAS, H. (1999). Independent factor analysis. *Neural Comput.* **11** 803–851.
- ATTIAS, H. (2000). A variational Bayesian framework for graphical models. *Adv. Neural Inf. Process. Syst.* **12** 209–215.
- BECK, A. T., STEER, R. A. and CARBIN, M. G. (1988). Psychometric properties of the beck depression inventory: Twenty-five years of evaluation. *Clin. Psychol. Rev.* **8** 77–100.
- BECK, A. T., STEER, R. A., BROWN, G. K. et al. (1996). Manual for the beck depression inventory-II.
- BECKMANN, C. F. and SMITH, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imag.* **23** 137–152.
- BECKMANN, C. F. and SMITH, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage* **25** 294–311.
- BECKMANN, C. F., DELUCA, M., DEVLIN, J. T. and SMITH, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 1001–1013.
- BECKMANN, C. F., MACKAY, C. E., FILIPPINI, N. and SMITH, S. M. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *NeuroImage* **47** S148.
- BISWAL, B. B. and ULMER, J. L. (1999). Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *J. Comput. Assist. Tomogr.* **23** 265–271.
- BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198.
- BULLMORE, E., BRAMMER, M., WILLIAMS, S. C., RABE-HESKETH, S., JANOT, N., DAVID, A., MELLERS, J., HOWARD, R. and SHAM, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35** 261–277.
- CALHOUN, V. D., ADALI, T., PEARLSON, G. D. and PEKAR, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* **14** 140–151.
- CAMPBELL, D. G., FELKER, B. L., LIU, C.-F., YANO, E. M., KIRCHNER, J. E., CHAN, D., RUBENSTEIN, L. V. and CHANEY, E. F. (2007). Prevalence of depression–PTSD comorbidity: Implications for clinical practice guidelines and primary care-based interventions. *Journal of General Internal Medicine* **22** 711–718.
- CHEN, C.-H., RIDLER, K., SUCKLING, J., WILLIAMS, S., FU, C. H., MERLO-PICH, E. and BULLMORE, E. (2007). Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biol. Psychiatry* **62** 407–414.

- COLE, L. J., FARRELL, M. J., GIBSON, S. J. and EGAN, G. F. (2010). Age-related differences in pain sensitivity and regional brain activity evoked by noxious pressure. *Neurobiol. Aging* **31** 494–503.
- DANIELS, J. K., FREWEN, P., MCKINNON, M. C. and LANIUS, R. A. (2011). Default mode alterations in posttraumatic stress disorder related to early-life trauma: A developmental perspective. *J. Psychiatry Neurosci.* **36** 56–59.
- DAUBECHIES, I., ROUSSOS, E., TAKERKART, S., BENHARROSH, M., GOLDEN, C., D'ARDENNE, K., RICHTER, W., COHEN, J. D. and HAXBY, J. (2009). Independent component analysis for brain fMRI does not select for independence. *Proc. Natl. Acad. Sci. USA* **106** 10415–10422.
- FILIPPINI, N., MACINTOSH, B. J., HOUGH, M. G., GOODWIN, G. M., FRISONI, G. B., SMITH, S. M., MATTHEWS, P. M., BECKMANN, C. F. and MACKAY, C. E. (2009). Distinct patterns of brain activity in young carriers of the APOE- $\epsilon$ 4 allele. *Proc. Natl. Acad. Sci. USA* **106** 7209–7214.
- FIRST, M. B. (1995). Structured clinical interview for the DSM (SCID). In *The Encyclopedia of Clinical Psychology*. American Psychiatric Press, Washington, DC.
- GENOVESE, C. R., LAZAR, N. A. and NICHOLS, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15** 870–878.
- GREICIUS, M. D., FLORES, B. H., MENON, V., GLOVER, G. H., SOLVASON, H. B., KENNA, H., REISS, A. L. and SCHATZBERG, A. F. (2007). Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* **62** 429–437.
- GRIFFANTI, L., SALIMI-KHORSHIDI, G., BECKMANN, C. F., AUERBACH, E. J., DOUAUD, G., SEXTON, C. E., ZSOLDOS, E., EBMEIER, K. P., FILIPPINI, N., MACKAY, C. E. et al. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95** 232–247.
- GUO, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67** 1532–1542. [MR2872404](#)
- GUO, Y. and PAGNONI, G. (2008). A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage* **42** 1078–1093.
- GUO, Y. and TANG, L. (2013). A hierarchical model for probabilistic independent component analysis of multi-subject fMRI studies. *Biometrics* **69** 970–981. [MR3146792](#)
- HENDLER, T., ROTSHTEIN, P., YESHURUN, Y., WEIZMANN, T., KAHN, I., BEN-BASHAT, D., MALACH, R. and BLEICH, A. (2003). Sensing the invisible: Differential sensitivity of visual cortex and amygdala to traumatic context. *NeuroImage* **19** 587–600.
- HIMBERG, J., HYVÄRINEN, A. and ESPOSITO, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage* **22** 1214–1222.
- HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis* **46**. Wiley, New York.
- HYVÄRINEN, A. and OJA, E. (2000). Independent component analysis: Algorithms and applications. *Neural Netw.* **13** 411–430.
- KESSLER, R. C., SONNEGA, A., BROMET, E., HUGHES, M. and NELSON, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Arch. Gen. Psychiatry* **52** 1048–1060.
- KOSTANTINOS, N. (2000). Gaussian mixtures and their applications to signal processing. In *Advanced Signal Processing Handbook*. CRC Press, New York.
- LEE, S., SHEN, H., TRUONG, Y., LEWIS, M. and HUANG, X. (2011). Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging. *J. Amer. Statist. Assoc.* **106** 1009–1024. [MR2894760](#)
- LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233. [MR0676213](#)

- MCKEOWN, M. J., MAKEIG, S., BROWN, G. G., JUNG, T.-P., KINDERMANN, S. S., KINDERMANN, R. S., BELL, A. J. and SEJNOWSKI, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* **6** 160–188.
- MCLACHLAN, G. and PEEL, D. (2004). *Finite Mixture Models*. Wiley, New York.
- MEILIJSON, I. (1989). A fast improvement to the EM algorithm on its own terms. *J. Roy. Statist. Soc. Ser. B* **51** 127–138. [MR0984999](#)
- MENG, X.-L. and RUBIN, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *J. Amer. Statist. Assoc.* **86** 899–909.
- MINKA, T. P. (2000). Automatic choice of dimensionality for PCA. In *NIPS* **13** 598–604. MIT Press, Cambridge, MA.
- QUITON, R. L. and GREENSPAN, J. D. (2007). Sex differences in endogenous pain modulation by distracting and painful conditioning stimulation. *Pain* **132 Suppl 1** S134–S149.
- RAICHLER, M. E., MACLEOD, A. M., SNYDER, A. Z., POWERS, W. J., GUSNARD, D. A. and SHULMAN, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. USA* **98** 676–682.
- REINEBERG, A. E., ANDREWS-HANNA, J. R., DEPUE, B. E., FRIEDMAN, N. P. and BANICH, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage* **104** 69–78.
- SEBER, G. A. and LEE, A. J. (2012). *Linear Regression Analysis* **936**. Wiley, New York.
- SHELIN, Y. I., BARCH, D. M., PRICE, J. L., RUNDLE, M. M., VAISHNAVI, S. N., SNYDER, A. Z., MINTUN, M. A., WANG, S., COALSON, R. S. and RAICHLER, M. E. (2009). The default mode network and self-referential processes in depression. *Proc. Natl. Acad. Sci. USA* **106** 1942–1947.
- SHI, R. and GUO, Y. (2016). Supplement to “Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis.” DOI:10.1214/16-AOAS946SUPP.
- SMITH, S. M., FOX, P. T., MILLER, K. L., GLAHN, D. C., FOX, P. M., MACKAY, C. E., FILIPPINI, N., WATKINS, K. E., TORO, R., LAIRD, A. R. et al. (2009). Correspondence of the brain’s functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* **106** 13040–13045.
- SMITH, D. V., UTEVSKY, A. V., BLAND, A. R., CLEMENT, N., CLITHERO, J. A., HARSCH, A. E., CARTER, R. M. and HUETTEL, S. A. (2014). Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. *NeuroImage* **95** 1–12.
- TOHKA, J., FOERDE, K., ARON, A. R., TOM, S. M., TOGA, A. W. and POLDRACK, R. A. (2008). Automatic independent component labeling for artifact removal in fMRI. *NeuroImage* **39** 1227–1245.
- WHITFIELD-GABRIELI, S., THERMENOS, H. W., MILANOVIC, S., TSUANG, M. T., FARAONE, S. V., MCCARLEY, R. W., SHENTON, M. E., GREEN, A. I., NIETO-CASTANON, A., LAVIOLETTE, P. et al. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc. Natl. Acad. Sci. USA* **106** 1279–1284.
- XU, L., CHEUNG, C., YANG, H. and AMARI, S. (1997). Maximum equalization by entropy maximization and mixture of cumulative distribution functions. In *Proc. of ICNN’97* 1821–1826. Springer, New York.

DEPARTMENT OF BIostatISTICS AND BIOinformatics  
ROLLINS SCHOOL OF PUBLIC HEALTH  
EMORY UNIVERSITY  
1518 CLIFTON RD.  
ATLANTA, GEORGIA 30322  
USA  
E-MAIL: [yguo2@emory.edu](mailto:yguo2@emory.edu)