



# A Parcellation Based Nonparametric Algorithm for Independent Component Analysis with Application to fMRI Data

Shanshan Li<sup>1\*</sup>, Shaojie Chen<sup>2</sup>, Chen Yue<sup>2</sup> and Brian Caffo<sup>2</sup>

<sup>1</sup> Department of Biostatistics, Indiana University Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA,

<sup>2</sup> Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

## OPEN ACCESS

### Edited by:

Bertrand Thirion,  
Institut National de Recherche en  
Informatique et Automatique, France

### Reviewed by:

Xi-Nian Zuo,  
Chinese Academy of Sciences, China  
Alexis Roche,  
Siemens Healthcare/Centre  
Hospitalier Universitaire Vaudois,  
Switzerland  
Ronald Phlypo,  
Grenoble Images Parole Signal  
Automatique-Lab, France

### \*Correspondence:

Shanshan Li  
sl50@iu.edu

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 14 May 2015

**Accepted:** 11 January 2016

**Published:** 29 January 2016

### Citation:

Li S, Chen S, Yue C and Caffo B  
(2016) A Parcellation Based  
Nonparametric Algorithm for  
Independent Component Analysis  
with Application to fMRI Data.  
*Front. Neurosci.* 10:15.  
doi: 10.3389/fnins.2016.00015

Independent Component analysis (ICA) is a widely used technique for separating signals that have been mixed together. In this manuscript, we propose a novel ICA algorithm using density estimation and maximum likelihood, where the densities of the signals are estimated via p-spline based histogram smoothing and the mixing matrix is simultaneously estimated using an optimization algorithm. The algorithm is exceedingly simple, easy to implement and blind to the underlying distributions of the source signals. To relax the identically distributed assumption in the density function, a modified algorithm is proposed to allow for different density functions on different regions. The performance of the proposed algorithm is evaluated in different simulation settings. For illustration, the algorithm is applied to a research investigation with a large collection of resting state fMRI datasets. The results show that the algorithm successfully recovers the established brain networks.

**Keywords:** blind source separation, density estimation, functional MRI, p-spline bases, signal processing

## 1. INTRODUCTION

This manuscript puts forward two innovations. Firstly, we demonstrate a fast, likelihood motivated and straightforward method for applying independent components analysis (ICA). Secondly, we propose a parcellation based adjustment when the source signals distribute differently across regions. Our work is routed in the context of understanding human brain networks, and we use functional magnetic resonance imaging (fMRI) data for illustration in this manuscript.

We approach our study of fMRI by simultaneously analyzing all voxels. This is in contrast to regional or seed-based approaches (Buckner et al., 2005; Wang et al., 2006; Allen et al., 2007) that restrict attention to carefully chosen locations. Such approaches require strong assumptions on the choice of seeds or parcellation used to define region. Hence voxel-wise approaches are important complementary procedures. Given the volume of voxels under study (usually on the order of fifty thousand non-background ones), flexible yet parsimonious models approaches are required. However, even with parsimonious models, whole brain voxel-level techniques are more empirical and exploratory than their more hypothesis driven regional and seed-based counterparts. Thus, exploratory factor-analytic models are common approaches in voxel-level investigations.

Independent components analysis (ICA) is a factor-analytic approach that has been frequently utilized for the analysis of functional neuroimaging data, because of its success in discovering

important brain networks in many applications (McKeown et al., 1998a,b; Calhoun et al., 2001a; Guo and Pagnoni, 2008). Two key benefits of ICA are its exploratory nature and its often considered reasonable underlying generative model. Specifically, it models collected signals,  $X$ , as linear weighted combinations of independent sources,  $S_1, S_2, \dots, S_p$ . Thus, we can write the noise-free ICA model as  $X = AS$ , where  $S = [S_1, S_2, \dots, S_Q]$  and  $A$  is a  $Q \times Q$  full rank matrix, the so-called mixing matrix. The goal of ICA is to recover the underlying signals  $S_1, S_2, \dots, S_Q$  from their observed mixtures  $X_1, X_2, \dots, X_Q$ . Note that, in the context of fMRI, the independent components  $S_1, S_2, \dots, S_Q$  are often interpreted as brain networks and  $A$  is the mixing matrix characterizing the temporal pattern of the corresponding brain networks.

Various algorithms for ICA have been proposed in the literature. See Hyvärinen et al. (2001); Comon and Jutten (2010); Risk et al. (2014) for comprehensive reviews. One common procedure is to postulate a parametric family for the source distributions and then obtain the independent components (ICs) by optimizing a contrast function that measures the distribution property of the output (Samworth and Yuan, 2012). The contrast functions can be selected based on different measures, such as entropy, mutual independence, high-order decorrelations, divergence between the joint distribution of the output and some model, etc. (Cardoso, 1998). These include the popular FastICA algorithm (Hyvärinen and Oja, 2000), the JADE algorithm (Cardoso, 1999), the Pearson ICA algorithm (Karvanen et al., 2000), and a few other algorithms proposed by Comon (1994); Amari and Cardoso (1997); Li and Adali (2010). An alternative procedure is to assume smooth densities for the source distributions and use nonparametric or semiparametric approaches to estimate those density functions. The mixing matrix can then be derived using maximum likelihood method. For example, Bach and Jordan (2003) developed a nonparametric estimation approach based on canonical correlations in a reproducing kernel Hilbert space. Hastie and Tibshirani (2002) expressed the source distribution by an exponentially tilted Gaussian function and used the fixed-point algorithm for estimation of the mixing matrix. Boscolo et al. (2004) used kernel density estimation techniques to model the underlying densities and quasi-Newton method for optimization. Guo and Pagnoni (2008) used Gaussian mixture models for the source distribution and provided an expectation-maximization (EM) framework for estimation, assuming Gaussian noise in the model. Eloyan et al. (2013) estimated the source distribution by using mixture density estimates, and proposed a constrained EM algorithm for estimation.

The benefit of the likelihood-based ICA algorithm is that, as a byproduct of the ICA algorithm, one obtains the fully specified likelihood of the ICA model which can be used for further statistical inference. For example, based on the fully specified likelihood, one can conduct Bayesian analysis or perform likelihood based model selection. However, the existing likelihood-based ICA algorithms are mostly semi-parametric and are usually computationally intensive. In this manuscript, we aim to develop a likelihood-based algorithm that is exceedingly simple and truly blinded to the source distributions.

We propose to estimate the density function of the ICs via histogram smoothing, following a well-known approach in the penalized spline literature. At its core, likelihood-based ICA requires estimation of the mixing matrix and flexible density estimation for the ICs. Our approach, like many other likelihood-based approaches, iteratively estimates these components separately using block maximization. In contrast to other approaches, we use an exceedingly simple density estimation technique via histogram smoothing. Specifically, we assume the bin counts of the frequency histogram follow a Poisson distribution and express the mean counts as sum of B-spline bases via generalized linear model. To smooth the histogram, we follow Eilers and Marx (1996) to construct a penalized likelihood with a difference penalty on coefficients of adjacent B-splines. Apart from its simplicity, a benefit of this approach is speed. Density estimation and evaluation for tens of thousands of voxels is time consuming, and worse, is performed within an iterative algorithm. Using histogram smoothing, the voxel-level calculation reduces to estimating a histogram, a very fast process.

We briefly mention that, in our primary area of application, fMRI, we focus entirely on noise-free group spatial independent component analysis. By assuming noise-free model, noise in the data is absorbed into the estimated ICs and the mixing matrix. By using spatial ICA model, the fMRI data is decomposed into spatial maps multiplied by their respective time courses, where the maps are drawn from spatial distributions that are statistically independent (Calhoun et al., 2001a). The spatial independence assumption is well suited to the sparse nature of the spatial pattern for typical brain activation (McKeown and Sejnowski, 1998; Guo and Pagnoni, 2008). The time courses estimated from spatial ICA describe the temporal characteristics of functional networks, i.e., areas of temporal correlation in the BOLD signal. For multi-subject fMRI data, we assume common spatial maps for all subjects and subject-specific mixing matrices, therefore, we can concatenate all subjects' data in the temporal domain, and apply ICA to the aggregated data matrix. The group mixing matrix is the concatenated time course for all subjects. Individual mixing matrices can be backreconstructed by partitioning the group mixing matrix into submatrices corresponding to each subject.

The remainder of the paper is organized as follows. Section 2 describes the p-spline based ICA algorithm and considers relaxation of the *i.i.d* signal assumption. Section 3 shows the performance of the proposed algorithm in simulation study. Section 4 provides the application of the proposed algorithm to the 1000 Functional Connectome Project ([https://www.nitrc.org/projects/fcon\\_1000/](https://www.nitrc.org/projects/fcon_1000/)), while Section 5 gives a discussion.

## 2. METHODS

### 2.1. Description of ICA Methodology

Independent component analysis models collected signals as linear weighted combinations of independent sources. Notationally, let  $X_i$  be a  $T \times V$  matrix for subject  $i = 1, \dots, I$ . In the context of fMRI,  $T$  indicates scans while  $V$  indicates

voxels. Assume the number of ICs is  $Q$ . The ICA model specifies  $X_i = A_i S$ , where  $A_i$  is a  $T \times Q$  mixing matrix and  $S$  is a  $Q \times V$  matrix of ICs. By assuming common spatial maps across subjects, we can stack the individual matrices in the temporal domain. Let  $X = [X_1^T, X_2^T, \dots, X_I^T]^T$  be the  $TI \times V$  group data matrix, and  $A = [A_1^T, A_2^T, \dots, A_I^T]^T$  be the  $TI \times Q$  group mixing matrix. Spatial group ICA simply specifies the standard model

$$X = AS. \tag{1}$$

We use parentheses to index matrices so that  $X(t, v)$  is element  $(t, v)$  of  $X$  and define  $X(t, \cdot)$  as row  $t$  of  $X$  and  $X(\cdot, v)$  as column  $v$ . Then, model (1) could be rewritten as  $X(t, v) = \sum_{q=1}^Q A(t, q)S(q, v)$  and  $X = \sum_{q=1}^Q A(\cdot, q)S(q, \cdot)$ .

We assume that  $E[X] = \mu_x = 0$  and hence  $E[S] = \mu_s = 0$ . If this assumption were not made, the ICA model would imply  $X - \mu_x = A(S - \mu_s)$ , which is exactly an ICA model with a centered data matrix and the ICs having mean 0. Hence,  $X$  is demeaned prior to analyses and  $\mu_s$  is assumed to be zero. Similarly, since  $A(\cdot, q)S(q, \cdot) = \{A(\cdot, q)/c\} * \{cS(q, \cdot)\}$ , ICs are only identified up to scalar multiplication. Thus, we assume that  $\text{Var}\{S(q, v)\} = 1$  for  $q = 1, \dots, Q$  and  $v = 1, \dots, V$ .

ICA gets its name by assuming that  $S(q, \cdot) \perp\!\!\!\perp S(q', \cdot)$  when  $q \neq q'$ , where  $\perp\!\!\!\perp$  implies statistical independence. However, standard variations of ICA also assumes that  $\{S(q, v)\}_{v=1}^V$  is an *i.i.d* collection, which we also adopt for now. The *i.i.d* assumption will be relaxed later in the next subsection. As a consequence of these assumptions,  $X(\cdot, v) \perp\!\!\!\perp X(\cdot, v')$  when  $v \neq v'$ ; yet note that  $X(t, \cdot)$  is not (necessarily) independent of  $X(t', \cdot)$ .

Typically,  $Q < TI$  and Equation (1) is overdetermined. A two-stage dimension reduction is often performed to reduce the computational load and avoid overfitting (Calhoun et al., 2001a; Beckmann and Smith, 2005; Guo and Pagnoni, 2008; Eloyan et al., 2013; Risk et al., 2014). Specifically, in the first stage, an SVD is performed in the temporal domain within subject, where the first  $R$  eigenvectors are retained. The dimension for the group data matrix then becomes  $RI \times V$ . In the second stage, an SVD is performed on the group data matrix obtained from the first stage and the first  $Q$  eigenvectors are retained to force a determined linear system for the group ICA model. This discards information in the data. However, one hopes that by selecting the first  $Q$  singular vectors, the most relevant features of the data will be retained. The choice of  $R$  and  $Q$  could be based on various criteria, including variance explained, information-theoretic criteria, and practical considerations. This is not a major concern in this article.

## 2.2. ICA Through Fast Nonparametric Density Estimation

ICA estimates  $S$  by seeking an unmixing matrix, say  $\hat{B}$ , such that  $\hat{B}X$  is a good approximation to the original sources  $S$ . Let  $B = A^{-1}$  be the estimand of interest. Notationally following Hyvärinen et al. (2001), if  $f_q$  is the density for  $S(q, v)$  for  $v = 1, \dots, V$ , and  $f = (f_1, \dots, f_Q)$ , then standard multivariate random variable transformation results imply that the joint density of

$X(\cdot, v)$  is

$$\begin{aligned} g\{X(\cdot, v)\} &= |\det(B)| \prod_{q=1}^Q f_q\{S(q, v)\} \\ &= |\det(B)| \prod_{q=1}^Q f_q\{B(q, \cdot)X(\cdot, v)\}, \end{aligned}$$

therefore the joint log-likelihood including all contributions for  $v = 1, \dots, V$  is

$$\mathcal{L}(B, f) = \sum_{v=1}^V \sum_{q=1}^Q \log[f_q\{B(q, \cdot)X(\cdot, v)\}] + V \log|\det(B)|.$$

It is generally not possible to solve the joint likelihood for the parameters in  $f_q$  and  $B$  simultaneously. Instead, an iterative optimization is often performed. Specifically, given the current estimate of  $B$  at iteration  $k$ , say  $\hat{B}^{(k)}$ , one can get an estimate for  $S$  via  $\hat{S}^{(k)} = \hat{B}^{(k)}X$ . Given  $\hat{S}^{(k)}(q, \cdot)$ , density estimation techniques can be used to obtain  $\hat{f}_q^{(k)}$ , the estimate of  $f_q$  at iteration  $k$ .

We suggest the use of histogram smoothing as the density estimation technique, where the bin counts of the frequency histogram are assumed to follow a Poisson distribution and a penalized likelihood is constructed to produce a smooth density estimate. The details of histogram smoothing can be found in Eilers and Marx (1996), and we provide a sketch below. (Readers not familiar with statistical smoothing may skip the rest of this paragraph). Notationally, let  $c^{(k)}(q, 0) < c^{(k)}(q, 1) < \dots < c^{(k)}(q, J)$  be equidistant histogram cutpoints, where  $c^{(k)}(q, 0) = -\epsilon + \min \hat{S}^{(k)}(q, \cdot)$  and  $c^{(k)}(q, J) = \epsilon + \max \hat{S}^{(k)}(q, \cdot)$ . The number  $\epsilon$  is added to avoid numerical boundary effects. Let  $n^{(k)}(q, j) = \sum_{v=1}^V I\{c^{(k)}(j-1) < \hat{S}^{(k)}(q, \cdot) \leq c^{(k)}(j)\}$ , for  $j = 1, \dots, J$ , be the count of values between cutpoints  $j-1$  and  $j$  for row  $q$  of  $\hat{S}^{(k)}$ . Define the midpoints of intervals  $[c^{(k)}(q, j-1), c^{(k)}(q, j)]$  by  $m^{(k)}(q, j)$  for  $j = 1, \dots, J$ . We obtain a density estimate via the log-linear model  $n^{(k)}(q, j) \sim \text{Poisson}\{\lambda^{(k)}(q, j)\}$ , where  $\log\{\lambda^{(k)}(q, \cdot)\} = \sum_{l=1}^L D^{(k)}\{m^{(k)}(q, \cdot), l\}\beta^{(k)}(q, l)$ . Here the log function is presumed to act component-wise on vectors,  $D^{(k)}$  is a B-spline basis design matrix,  $L$  is the number of knots for B-splines, and  $\beta^{(k)}(q, \cdot)$  is a vector of coefficients. To avoid overfitting the B-spline model, and to avoid sensitivity to the degrees of freedom, we choose a large value for the degrees of freedom and put a squared penalty on the coefficients. Let  $\mu^{(k)}(q, j)$  denote the expectation of  $n^{(k)}(q, j)$ , then the penalized log likelihood takes the form (Eilers and Marx, 1996)

$$\begin{aligned} \mathcal{L} &= \sum_{j=1}^J n^{(k)}(q, j) \ln \mu^{(k)}(q, j) - \sum_{j=1}^J \mu^{(k)}(q, j) \\ &\quad - \delta \sum_{l=3}^L \frac{\{\Delta^2 \beta^{(k)}(q, l)\}^2}{2}, \end{aligned}$$

where  $\delta$  is a parameter controlling the smoothness of the fit,  $\Delta$  denotes the difference operator,  $\Delta^2 \beta(\cdot, l) = \beta(\cdot, l) - 2\beta(\cdot, l-1) + \beta(\cdot, l-2)$ . The resulting model is then a generalized linear

mixed model on the counts. The B-spline basis is evaluated at the midpoint of the cutpoint interval. However, via interpolation, the smoother gives an estimate for all values, thus yielding a continuous function, say  $\hat{f}_q^{(k)}(s)$ , which is the density estimate.

Using generalized linear mixed models to penalize smoothing has become standard practice and is well described in Ruppert et al. (2003). Histogram smoothing as a density estimate appears to be less commonly used. However, we note that this p-spline based density smoother has very attractive properties (Eilers and Marx, 1996). First, it results in a proper density. Secondly, it elegantly handles boundary issues, unlike other density estimators (such as kernel density estimator). Thirdly, the estimated density conserves the first few empirical moments (means and variances) of the histogram, depending on the order of the B-splines. More details regarding these properties can be found in Eilers and Marx (1996). Note that, conservation of moments is an important property that guarantees the identifiability of the ICA model. We choose a cubic B-spline which then conserves the first two moments of the histogram.

Furthermore, due to the convenient differentiation properties of B-spline bases and the simple exponential (Poisson) model, the first and second derivatives of  $\hat{f}_q^{(k)}$  are immediately available, where  $\hat{d}_q^{(k)} = \exp\{\hat{f}_q^{(k)}\} \beta^{(k)}(q, \cdot) dD^{(k)}$ . Thus, derivatives of  $\mathcal{L}(B)$  are available in closed form, making gradient- and Hessian-based optimization algorithms easy to implement. This is useful for the stage of the algorithm for obtaining the next iterate of  $B$ . Accordingly, we use a Newton-Raphson method to update the mixing matrix. Specifically, let  $\mathcal{L}'$  and  $\mathcal{L}''$  denote the first and second derivatives of the log likelihood. At the  $k$ th iteration, we update  $B$  by

$$B^{(k+1)} = B^{(k)} - \mathcal{L}''(B^{(k)})^{-1} \mathcal{L}'(B^{(k)}). \quad (2)$$

The starting values of  $B$  should satisfy the condition that the underlying ICs are the same for all subjects. Following Eloyan et al. (2013), we decompose the full matrix  $X$  using the population value decomposition  $X = U\Sigma V^T$  (Crainiceanu et al., 2011), and the starting values of the  $B_i$  are chosen as the  $i$ th block of the rows of  $U\Sigma$ . Thus, given a starting value for  $B$ , histogram smoothing is used to obtain  $f_q$ , then given the update for  $f_q$ , the natural gradient algorithm is used to obtain  $B$  and these steps are iterated until convergence. Let  $P$  denote  $B^{(k)}(B^{(k+1)})^{-1}$ . We use the Amari metric between  $B^{(k+1)}$  and  $B^{(k)}$  as our convergence criterion (Amari, 1998), where the metric is defined as

$$d\{B^{(k)}, B^{(k+1)}\} = \frac{1}{2Q} \sum_{i=1}^Q \left( \sum_{j=1}^Q \frac{|P_{ij}|}{\max_j |P_{ij}|} - 1 \right) + \frac{1}{2Q} \sum_{j=1}^Q \left( \sum_{i=1}^Q \frac{|P_{ij}|}{\max_i |P_{ij}|} - 1 \right).$$

The Amari metric is useful, as it is invariant to permutation of the ordering of the ICs, a necessary condition for a convergence metric to be useful.

## 2.3. ICA Based on Parcellation

Most ICA algorithms (such as the commonly used fastICA, JADE, etc.) assume that  $\{S(q, v)\}_{v=1}^V$  is an *i.i.d* collection for all  $q = 1, \dots, Q$ . Intrinsically, this is to assume that the values of the ICs are independent draws from a density. The *i.i.d* assumption is made for simplicity, but it may not hold for fMRI data. Calhoun et al. (2001b) considered possible violations of the independence assumption for task-based fMRI data. They found that the ICA algorithm was successful when the correlation in the signal was small, but it may fail when the signals are highly dependent. However, for most task-based fMRI and resting-state fMRI data, the correlation between voxels is negligible. Therefore, we do not pursue the approach to deal with violation of the independence assumption here. Instead, we consider relaxation of the identically distributed assumption.

Specifically, we propose to account for the difference in the activity across the brain by allowing different density distribution in different regions. To this end, we adopt the functional parcellation of the brain activity map proposed by Yeo et al. (2011). The parcellation includes 17 functional networks in the cerebral cortex, that is,  $I = 18$  ROIs for the whole brain. We assume the signals are *i.i.d* within region but could be differently distributed across region. Under this assumption, the density function  $f_q$  can be written as the sum of the region-specific density function, that is,

$$f_q(s) = \sum_{i=1}^I I(s \in R_i) f_{iq}(s),$$

where  $R_i$  denotes the  $i$ th ROI,  $f_{iq}$  is the density function on  $R_i$ . Thus,  $f_{iq}$  takes positive values on the  $i$ th region and zero elsewhere. The density estimate of  $f_{iq}$  can be obtained using the same procedure as proposed in Section 2.2, confined to the  $i$ th region. The estimate for  $f_q$  can be constructed by taking the sum of  $\hat{f}_{iq}$ . The rest of the ICA algorithm follows the proposed procedure in Section 2.2.

The proposed ICA algorithm can be summarized as follows:

1. Choose an initial value of for the mixing matrix  $B$ .
2. Alternate until convergence of  $B$  using the Amari metric.
  - a. Let  $S = BX$ .
  - b. For each IC  $q$ , calculate the density function  $f_{iq}(s)$  on the  $i$ th ROI,  $i = 1, 2, \dots, I$ , using the p-spline based density estimation algorithm.
  - c. Get  $f_q(s) = \sum_{i=1}^I I(s \in R_i) f_{iq}(s)$ .
  - d. Update the mixing matrix  $B$  using the Newton-Raphson method, see Equation (2).

Note that, in the special case that  $f_{1q} = f_{2q} = \dots = f_{iq}$ , the above algorithm reduces to the algorithm proposed in Section 2.2 assuming *i.i.d* signals across the entire brain.

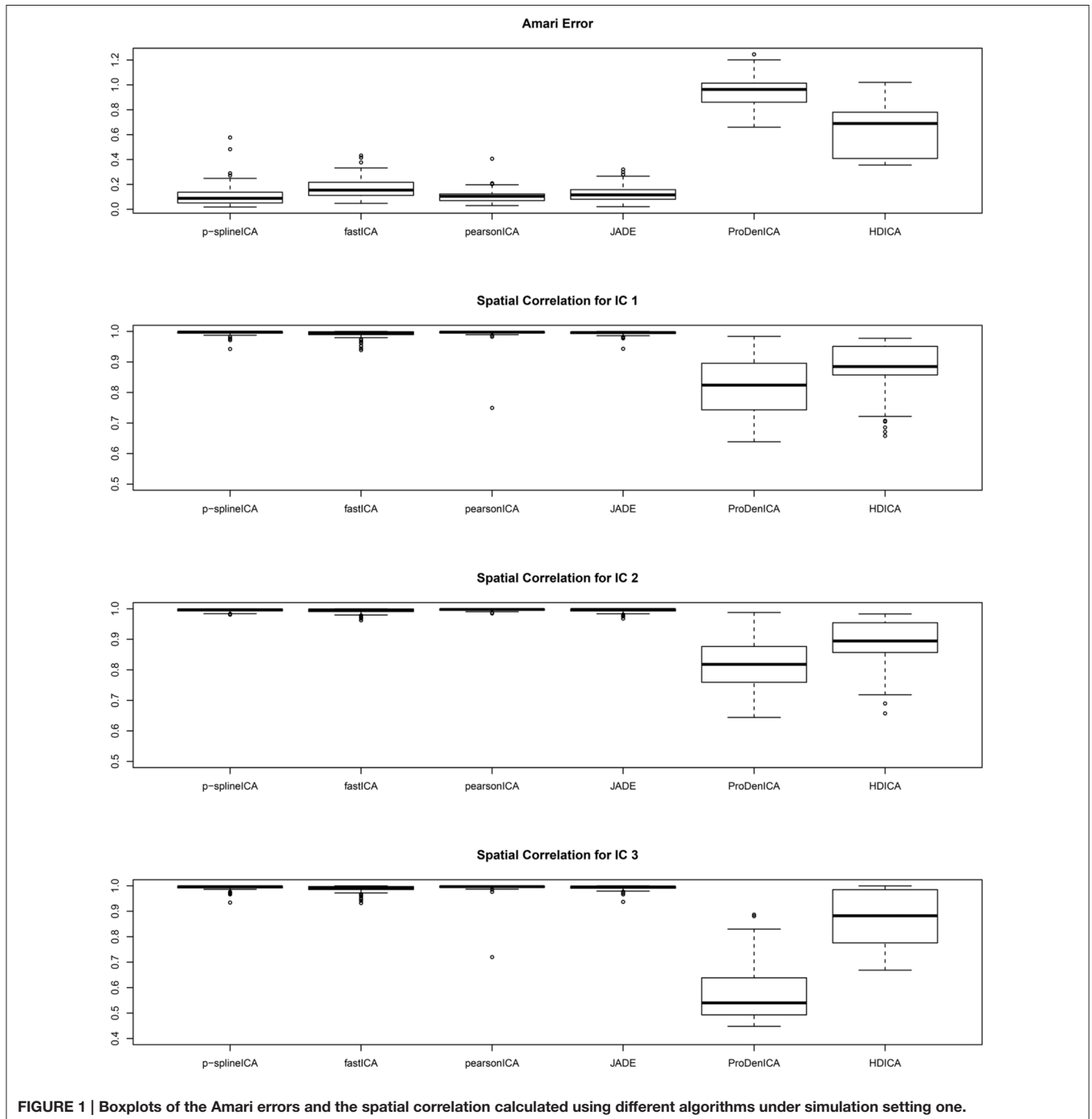
## 3. SIMULATION

We conduct simulation studies to evaluate the performance of the proposed ICA algorithm. We consider four settings where

**TABLE 1 | The average computation time (in seconds) per simulation replication using different algorithms in the simulation study.**

Simulation	p-spline ICA	fastICA	Pearson ICA	JADE	ProDenICA	HDICA
Scenario 1	6.21	< 0.01	0.01	0.01	3.12	308.34
Scenario 2	5.89	< 0.01	0.01	< 0.01	1.67	76.59
Scenario 3	7.27	0.01	0.01	0.01	1.74	123.84

data are generated using different distributions. We compare the results of the proposed algorithm with fastICA (Hyvärinen et al., 2001), JADE (Cardoso, 1999), Pearson ICA (Karvanen et al., 2000), ProDenICA (Hastie and Tibshirani, 2002), and HDICA (Eloyan et al., 2013). We implement the algorithms fastICA, ProDenICA, JADE, PearsonICA using the R packages “fastICA” (Marchini et al., 2013), “ProDenICA” (Hastie and Tibshirani, 2010), “JADE” (Nordhausen et al., 2014), and “PearsonICA”



(Karvanen, 2006). The proposed p-spline based ICA algorithm and the HDICA (Eloyan et al., 2013) are also implemented in R.

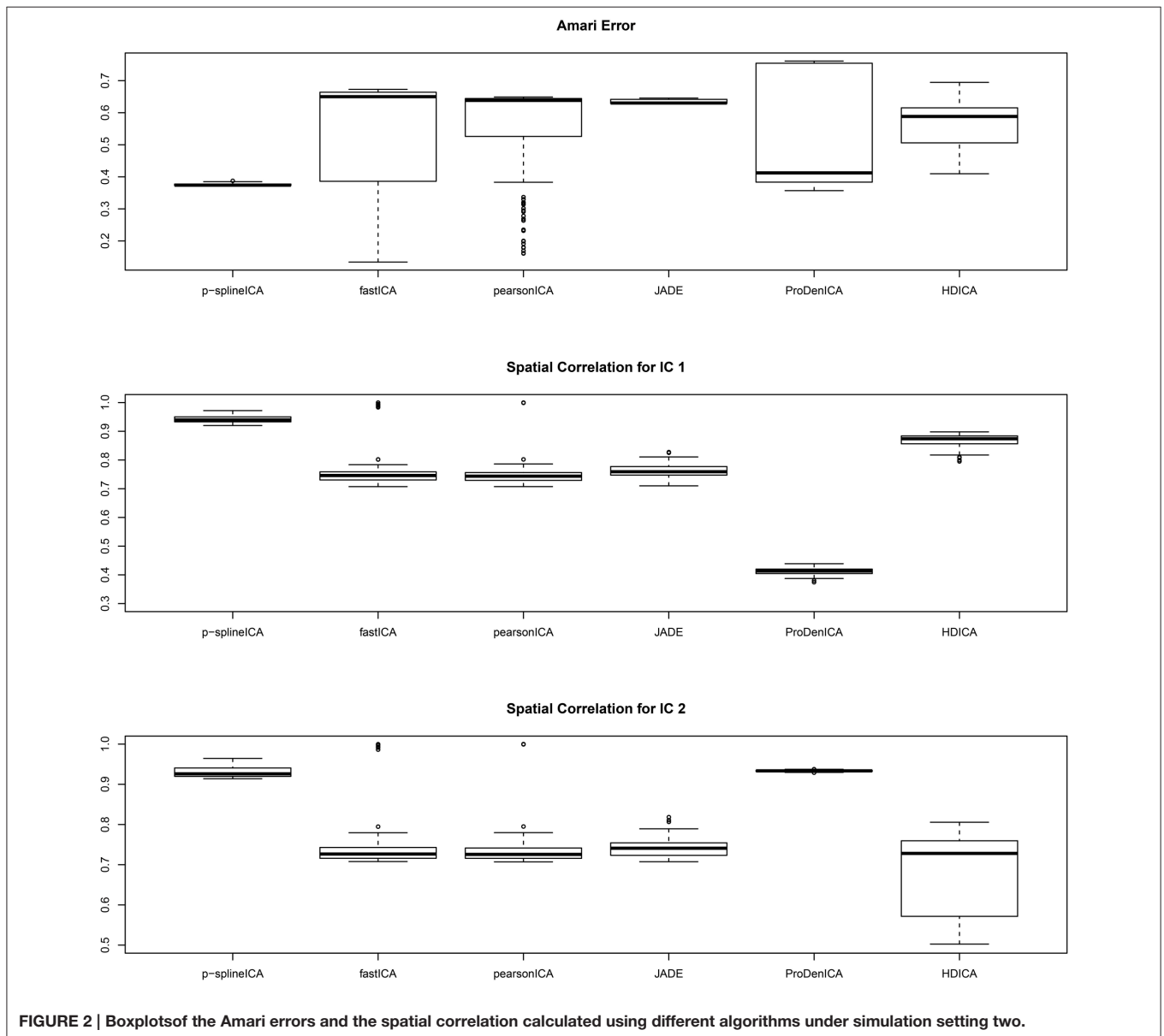
The computation environment is a multi-core Linux cluster with more than 680 cores running in the average of 2.5 GHz speed and 4.4 TB of memory. On average, the contrast-function based algorithms (fastICA, PearsonICA, JADE) perform much faster than the likelihood-based algorithms (p-spline ICA, ProDenICA, HDICA). (See **Table 1** for a summary of the computation time using different algorithms.) However, since those are essentially two different sets of algorithms, we restrict the comparison of the computational intensity within the category of likelihood-based algorithm.

In the first set of simulation studies, we assume there are  $Q = 3$  independent components, and they are generated by  $S(1, \cdot) \sim \text{Weibull}(1, 1)$ ,  $S(2, \cdot) \sim \text{Gamma}(1, 1)$ , and  $S(3, \cdot) \sim$

$\text{Gamma}(2, 2)$ , respectively. Standard Gaussian noises are added to the generated ICs. The mixing matrix is given by

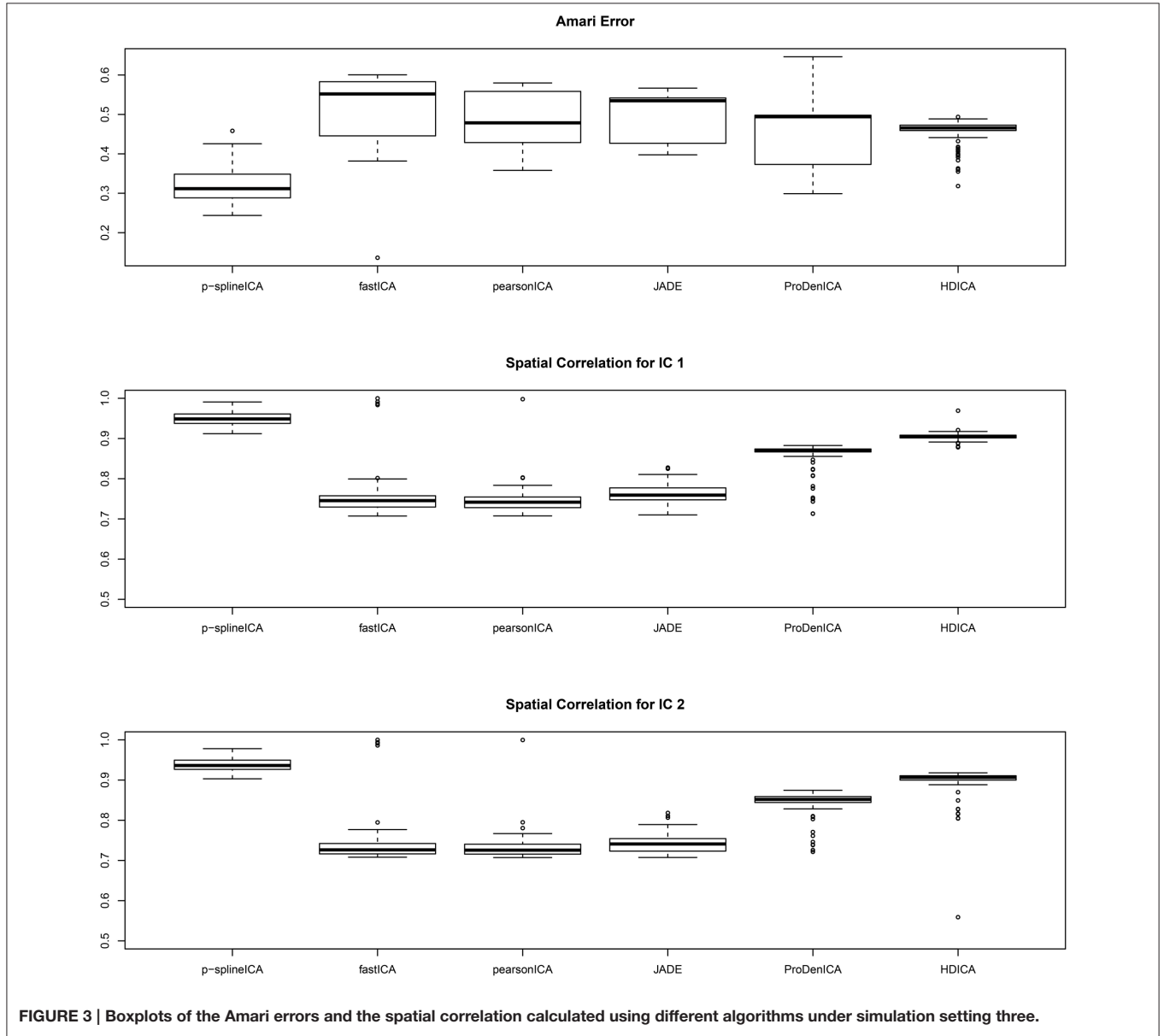
$$A = \begin{pmatrix} 2 & 1 & 2 \\ 3 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}.$$

**Figure 1** summarizes the simulation results based on 200 replications. The spatial correlation is the absolute correlation between the estimated spatial map and the true spatial map without noise. The Amari error is computed to evaluate the accuracy of the estimated mixing matrix (Amari, 1998). It is seen from the boxplots of the spatial correlation and the Amari errors that the proposed ICA algorithm performs equally well as fastICA, JADE, and PearsonICA, and all

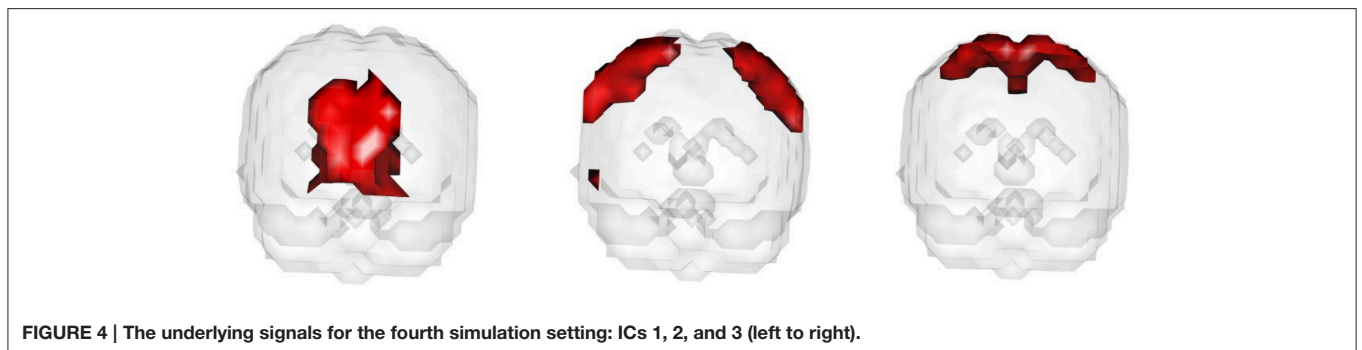


these algorithms perform substantially better than ProDenICA algorithm. ProDenICA fails probably due to the extreme values introduced by the noise (See more discussion in Risk et al.,

2014). This shows that the ProDenICA is sensitive to extreme values, while our algorithm is robust to extreme values. The average computation time per replication is 6.21 s using the



**FIGURE 3 |** Boxplots of the Amari errors and the spatial correlation calculated using different algorithms under simulation setting three.



**FIGURE 4 |** The underlying signals for the fourth simulation setting: ICs 1, 2, and 3 (left to right).

p-spline ICA, 3.12 s using ProDenICA, and 308.34 s using HDICA.

In the second setting, we assume the number of source signals  $Q = 2$ , and we generate the signals based on parcellation. Specifically, we partition the real line into 10 intervals, with cutoffs at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of the normal distribution. For the first IC, the density function is uniformly distributed within each interval, but the overall shape is approximately normal. For the second IC, the density function follows Laplace distribution within each interval, and the overall shape is approximately normal. The mixing matrix is given by

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}.$$

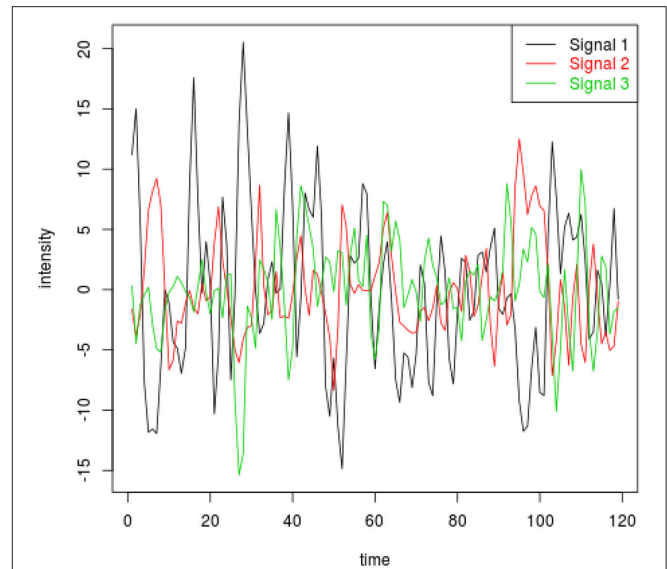
The boxplots of the spatial correlation and the Amari errors based on 200 replications are summarized in **Figure 2**. Under the second scenario, the underlying signals have region specific densities, and the overall density functions for both components are approximately normal. All the competing algorithms considered in the simulation studies show substantial bias. These algorithms fail to recover the true signals because they heavily depend on the non-gaussianity assumption. On the contrary, the proposed algorithm accounts for the effect of parcellation and recovers the true signals with relatively high accuracy. The proposed algorithm substantially outperforms all the competing algorithms under the second setting. The average computation time per replication is 5.89 s using p-spline ICA, 1.67 s using ProDenICA, and 76.59 s using HDICA.

In the third setting, we generate multi-subject data with number of subject  $I = 3$ . The source signals are the same as those in the second setting, and the mixing matrices for the three subjects are given by

$$A_1 = \begin{pmatrix} 1 & 0.25 \\ 0.5 & -0.5 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 2 \\ 0.5 & -0.5 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}.$$

The simulation results are summarized in **Figure 3**, where in each simulation replication, the Amari error is calculated as the average of the Amari errors for all three subjects. The results show that, for multi-subject data, the proposed algorithm successfully recovers both the common spatial signals and the individual mixing matrices. In addition, for similar reasons as in the second setting, the proposed algorithm substantially outperforms all the competing algorithms.

In the fourth setting, we generate the ICs and mixing matrices by mimicking signals from real fMRI data. Specifically, we run fastICA on 10 subjects from the NITRC 1000 Connectome dataset to get twenty ICs (networks). Three of the twenty networks are chosen as the true signals, and they are shown in **Figure 4**. The time courses are also signals from real data, obtained in a similar way as in Calhoun et al. (2009). They are shown in **Figure 5**. We first apply a two-stage dimension reduction using the method as described in Section 2.1. Then we apply the proposed atlas-based ICA algorithm using the brain parcellation proposed by Yeo et al. (2011). The correlation matrix



**FIGURE 5 |** Time courses corresponding to ICs 1, 2, and 3 for the fourth simulation setting.

between the true signals and the estimated signals using the proposed algorithm is

$$\begin{pmatrix} -0.999 & -0.025 & 0.007 \\ 0.028 & 0.999 & -0.006 \\ -0.004 & -0.003 & 0.999 \end{pmatrix}.$$

The results indicate that our proposed p-spline based ICA algorithm is successful in recovering signals from real fMRI data.

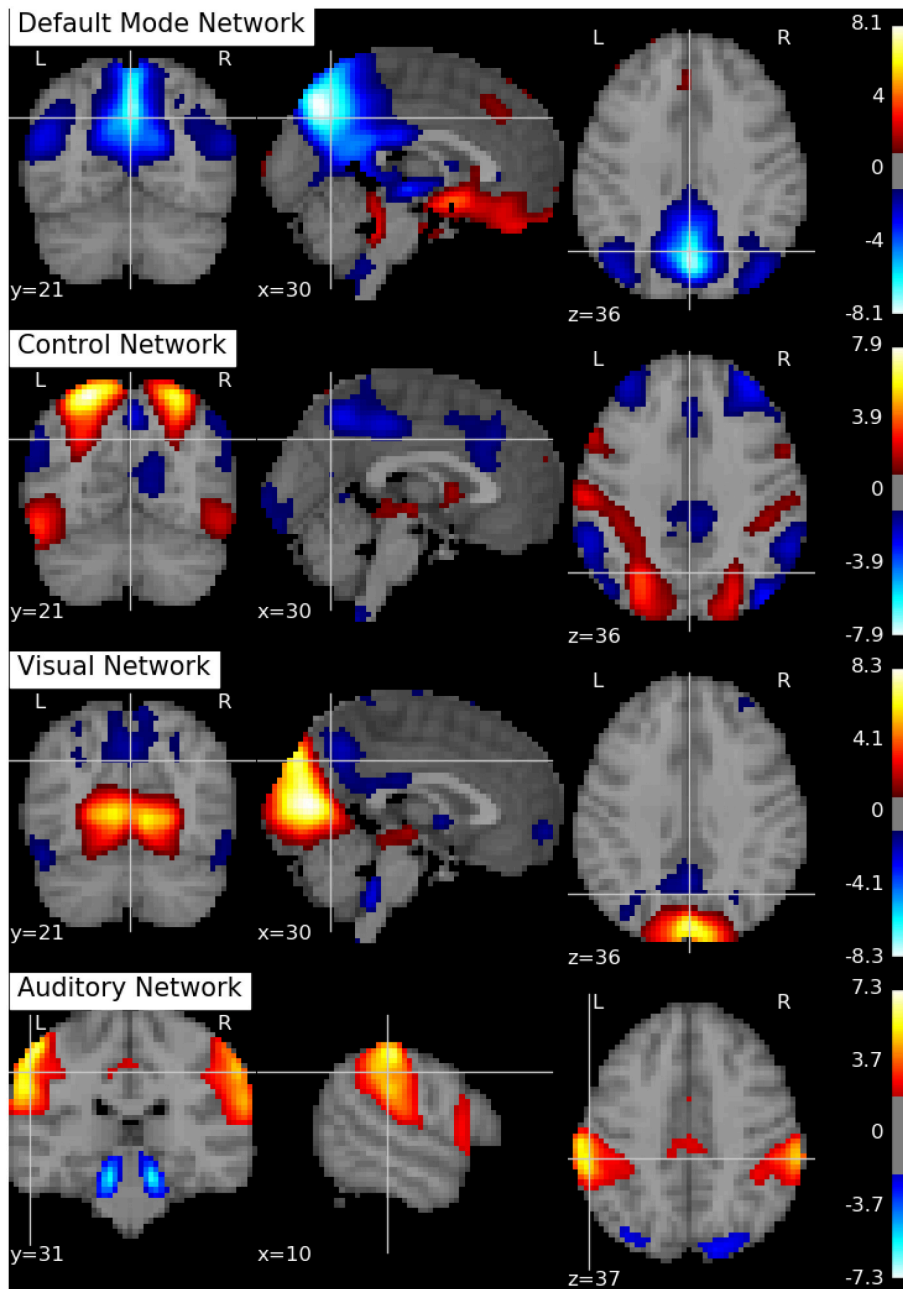
## 4. APPLICATION

We apply our proposed algorithm to the 1000 Functional Connectomes Project dataset, which consists of thousands of resting state scans combined across multiple sites with the goal of facilitating discovery and analysis of brain networks (Biswal et al., 2010). It is one of the largest freely available fMRI datasets. The fMRI scans were collected when the subjects stayed in the scanner for 2.2–20 min in resting state. Scanning parameters used to acquire the data from each site are detailed elsewhere (for complete information see [https://www.nitrc.org/projects/fcon\\_1000/](https://www.nitrc.org/projects/fcon_1000/)).

As the quality and scanning parameters vary across sites, we focus on data from the largest site, Cambridge, which contains  $I = 50$  subjects. For the subjects used in this analysis, the number of time points is  $T = 119$ . We use the MNI template to remove the background noise and to retain voxels that are in the actual brain. For each subject, we have a  $T \times V$  dimensional matrix  $X_i$ . The group data matrix  $X$  is obtained by concatenating  $I$  subjects' data in the temporal domain.

Following Biswal et al. (2010), we assume there are  $Q = 20$  independent components in this application. An SVD is



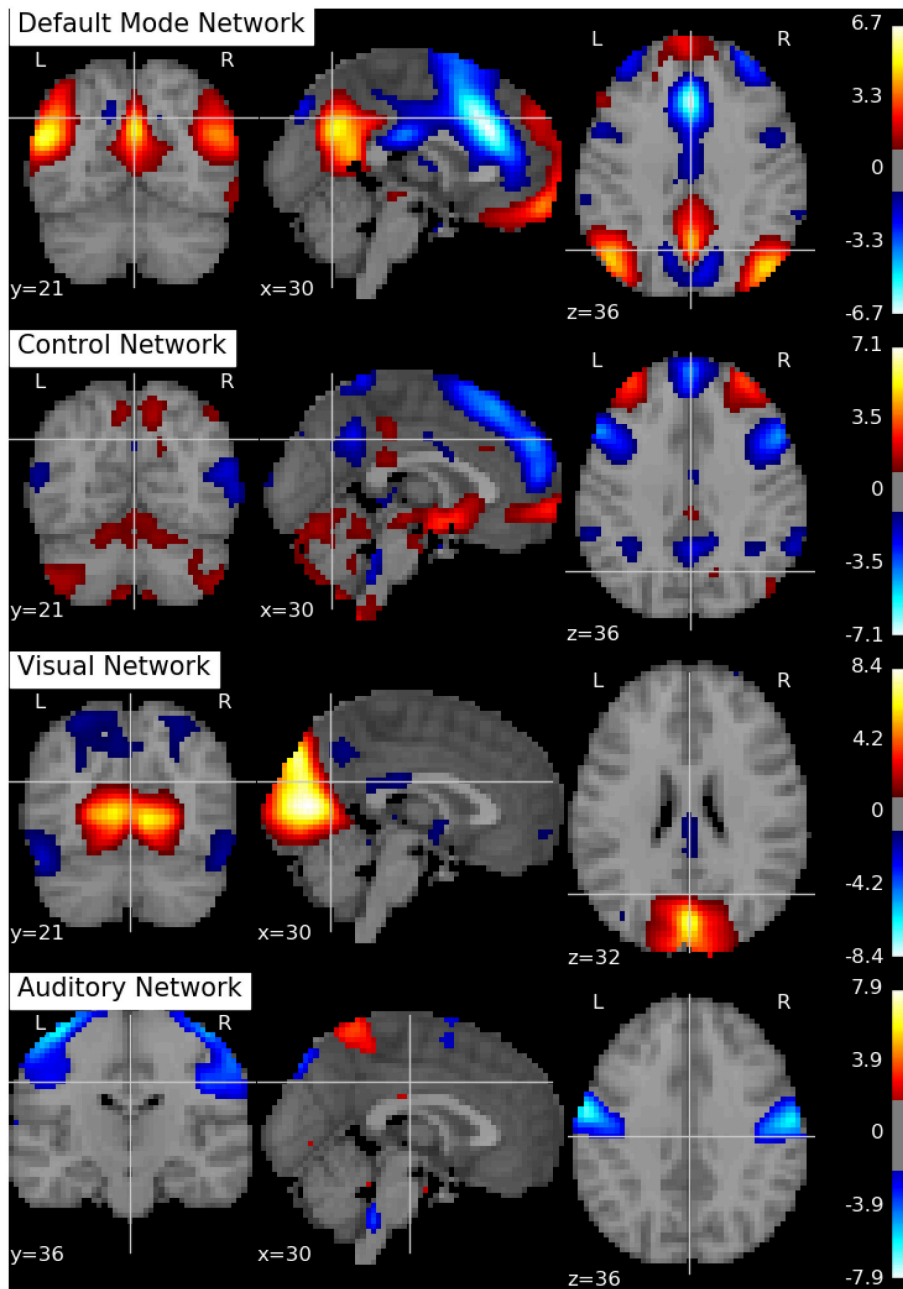


**FIGURE 6 |** Brain networks obtained by the p-spline ICA algorithm using data from the 1000 Functional Connectome Project.

performed to reduce the dimension of the aggregated data matrix to  $Q \times V$ . The ICA algorithms are then applied to the reduced data matrix and the Python toolbox Nilearn (Abraham et al., 2014) is used for visualization of the estimation results. Specifically, the estimated ICs using the proposed p-spline based ICA algorithm are shown in **Figure 6**. Several main brain networks including the default mode network (DMN) and the control network are successfully identified by the proposed algorithm. As a comparison, the results from fastICA are shown in **Figure 7**. The ICs estimated by fastICA and the p-spline ICA are matched by

correlation. Of the 20 pairs, the highest correlation is 0.99, the lowest correlation is 0.52, and the median correlation is 0.93. Specifically, the following is a list of these correlations for the major brain networks: visual network (0.99), auditory network (0.98), DMN (0.96), and control network (0.92).

As suggested by an anonymous reviewer, we investigate the impact of the dimension of the reduced space on the final results. Specifically, we select different values of  $R$  and  $Q$  (the number of eigenvectors in the subject-level and group-level dimension reduction), and rerun the ICA algorithm on the dimension



**FIGURE 7 |** Brain networks obtained by the fastICA algorithm using data from the 1000 Functional Connectome Project.

reduced dataset. We set  $R = 15, 20, 30$  and  $Q = 15, 20, 30$ , respectively. Similarly as in Li et al. (2007), we find that the IC estimates are well separated when  $Q = 15, 20$ . When  $Q = 30$ , the estimation of the major networks shows degradation and a few of the other estimated components seem to be noise. Specifically, the correlations for the major brain networks using  $R = 20, Q = 15$ , and  $R = 20, Q = 20$  are as follows: visual network (0.96), auditory network (0.73), DMN (0.86) and control network (0.84). In addition, the correlations for the major brain networks using  $R = 20, Q = 30$ , and  $R = 20, Q = 20$

are as follows: visual network (0.78), auditory network (0.61), DMN (0.88) and control network (0.69). In summary, we find that, although the estimation results depend on the number of components, the major networks appear to be robust against the choices of number of components.

## 5. DISCUSSION

Independent component analysis is a factor-analytic approach that is commonly used in analyzing fMRI data. In this

manuscript, we present a novel and simple ICA algorithm that is fast, likelihood based and straightforward to program. The algorithm is nonparametric, data-driven, and is blind to the particular distribution of the underlying signals. As a byproduct of the algorithm, we obtain the likelihood function of the ICA model which can be used for further statistical inference. It should be noted that, the likelihood function in our algorithm is a profile likelihood, since we are mainly interested in the mixing matrix estimates and the parameters over the spline basis are nuisance parameters. Indeed, one could also study the coefficients on the spline basis in a full likelihood, but this is not the goal of this manuscript, hence the variance of the estimator of the mixing matrix depends on the variance of the nuisance parameters.

The proposed algorithm is extended to allow for region specific IC density functions, on the rationale that most signals of interest are reasonably confined to a subset of the entire anatomical brain space (Guo and Pagnoni, 2008). When the source signals distribute identically across brain, the estimation accuracy of the parcellation-based estimator is similar to that of the full-brain estimator, because it becomes equivalent to the full-brain estimator. However, when the source signals distribute differently across brain, the full-brain estimator may result in substantial bias while the parcellation-based estimator can successfully recover the source signals. It should be noted that the parcellation based adjustment can be applied to other ICA algorithms as well. Indeed, for any gradient-based ICA, one can do the adjustment by taking a weighted sum over the updates of each of the parcellations, where the weights account for the number of samples in the parcellations. This flexibility ensures the generalizability of the proposed parcellation based adjustment.

Simulation studies show that our proposed algorithm works well in both the simple and complex situations, and it

substantially outperforms the existing ICA algorithms when the identically distributed assumption of the source signals is violated. In applying the proposed algorithm to the fMRI data, we choose to account for the difference in brain activities across regions by using the brain parcellation proposed by Yeo et al. (2011). Our data application results show that the proposed algorithm successfully identifies the main brain networks in the 1000 Functional Connectomes Project dataset.

There are a few directions for future research. Firstly, the test-retest reliability of the intrinsic brain networks is an important issue and has been studied extensively in recent years. For example, Zuo et al. (2010) found that a few functionally relevant components (such as the default mode, auditory-motor and executive control) show the highest reliability across all components. It would be interesting to compare different ICA algorithms in identifying and characterizing those functionally relevant components. Secondly, there are a variety of existing brain parcellation schemes, including those proposed by Tzourio-Mazoyer et al. (2002); Fischl et al. (2004); Beckmann et al. (2009); Yeo et al. (2011). It would be interesting to study the optimal choice of parcellation under different scientific scenarios. Thirdly, as pointed out by an anonymous reviewer, pre-whitening, although a standard pre-processing procedure, may result in loss of information and bias in estimation (Cardoso, 1994). It would be interesting to investigate alternative pre-processing procedures to avoid the bias introduced by pre-whitening.

## ACKNOWLEDGMENT

This work was partly supported by a grant from the Simons Foundation (#354180, Shanshan Li).

## REFERENCES

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8:14. doi: 10.3389/fninf.2014.00014
- Allen, G., Barnard, H., McColl, R., Hester, A. L., Fields, J. A., Weiner, M. F., et al. (2007). Reduced hippocampal functional connectivity in Alzheimer disease. *Arch. Neurol.* 64, 1482–1487. doi: 10.1001/archneur.64.10.1482
- Amari, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Comput.* 10, 251–276. doi: 10.1162/089976698300017746
- Amari, S.-I., and Cardoso, J.-F. (1997). Blind source separation-semiparametric statistical approach. *IEEE Trans. Signal Process.* 45, 2692–2700. doi: 10.1109/78.650095
- Bach, F. R., and Jordan, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* 3, 1–48. doi: 10.1109/icassp.2003.1202783
- Beckmann, C. F., and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage* 25, 294–311. doi: 10.1016/j.neuroimage.2004.10.043
- Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190. doi: 10.1523/JNEUROSCI.3328-08.2009
- Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107
- Boscolo, R., Pan, H., and Roychowdhury, V. P. (2004). Independent component analysis based on nonparametric density estimation. *IEEE Trans. Neural Netw.* 15, 55–65. doi: 10.1109/TNN.2003.820667
- Buckner, R. L., Snyder, A. Z., Shannon, B. J., LaRossa, G., Sachs, R., Fotenos, A. F., et al. (2005). Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* 25, 7709–7717. doi: 10.1523/JNEUROSCI.2177-05.2005
- Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001a). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. doi: 10.1002/hbm.1048
- Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001b). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13, 43–53. doi: 10.1002/hbm.1024
- Calhoun, V. D., Liu, J., and Adall, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45, 163–172. doi: 10.1016/j.neuroimage.2008.10.057
- Cardoso, J.-F. (1994). "On the performance of orthogonal source separation algorithms," in *Proceedings of EUSIPCO*, Vol. 94 (Edinburgh), 776–779.
- Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proc. IEEE* 86, 2009–2025. doi: 10.1109/5.720250
- Cardoso, J.-F. (1999). High-order contrasts for independent component analysis. *Neural Comput.* 11, 157–192. doi: 10.1162/089976699300016863

- Comon, P. (1994). Independent component analysis, a new concept? *Signal Process.* 36, 287–314. doi: 10.1016/0165-1684(94)90029-9
- Comon, P., and Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Cambridge, MA: Academic Press.
- Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. M., and Punjabi, N. M. (2011). Population value decomposition, a framework for the analysis of image populations. *J. Am. Stat. Assoc.* 106, 775–790. doi: 10.1198/jasa.2011.ap10089
- Eilers, P., and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Stat. Sci.* 11, 89–121. doi: 10.1214/ss/1038425655
- Eloyan, A., Crainiceanu, C. M., and Caffo, B. S. (2013). Likelihood-based population independent component analysis. *Biostatistics* 14, 514–527. doi: 10.1093/biostatistics/kxs055
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., et al. (2004). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22. doi: 10.1093/cercor/bhg087
- Guo, Y., and Pagnoni, G. (2008). A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage* 42, 1078–1093. doi: 10.1016/j.neuroimage.2008.05.008
- Hastie, T., and Tibshirani, R. (2002). *Independent Component Analysis Through Product Density Estimation*. Technical Report, Department of Statistics, Stanford University.
- Hastie, T., and Tibshirani, R. (2010). *ProDenICA: Product Density Estimation for ICA using tilted Gaussian Density Estimates*. R package version 1.0.
- Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Hoboken, NJ: John Wiley and Sons. doi: 10.1002/0471221317
- Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5
- Karvanen, J. (2006). *Package PearsonICA*. R package version 1.2-4.
- Karvanen, J., Eriksson, J., and Koivunen, V. (2000). “Pearson system based method for blind separation,” in *Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)* (Helsinki), 585–590.
- Li, X.-L., and Adali, T. (2010). Independent component analysis by entropy bound minimization. *IEEE Trans. Signal Process.* 58, 5151–5164. doi: 10.1109/TSP.2010.2055859
- Li, Y.-O., Adali, T., and Calhoun, V. D. (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Human Brain Mapp.* 28, 1251–1266. doi: 10.1002/hbm.20359
- Marchini, J. L., Heaton, C., and Ripley, B. D. (2013). *fastICA: FastICA Algorithms to Perform ICA and Projection Pursuit*. R package version 1.2-0.
- McKeown, M. J., Jung, T. P., Makeig, S., Brown, G., Kindermann, S. S., Lee, T. W., et al. (1998a). Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc. Natl. Acad. Sci. U.S.A.* 95, 803–810. doi: 10.1073/pnas.95.3.803
- McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., et al. (1998b). Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188.
- McKeown, M. J., and Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Hum. Brain Mapp.* 6, 368–372.
- Nordhausen, K., Cardoso, J.-F., Miettinen, J., Oja, H., Ollila, E., and Taskinen, S. (2014). *JADE: JADE and other BSS Methods as Well as Some BSS Performance Criteria*. R package version 1.9-92.
- Risk, B. B., Matteson, D. S., Ruppert, D., Eloyan, A., and Caffo, B. S. (2014). An evaluation of independent component analyses with an application to resting-state fMRI. *Biometrics* 70, 224–236. doi: 10.1111/biom.12111
- Ruppert, D., Wang, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, MA: Cambridge University Press. doi: 10.1017/cbo9780511755453
- Samworth, R. J., and Yuan, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Stat.* 40, 2973–3002. doi: 10.1214/12-AOS1060
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., et al. (2006). Changes in hippocampal connectivity in the early stages of Alzheimer’s disease: evidence from resting state fMRI. *Neuroimage* 31, 496–504. doi: 10.1016/j.neuroimage.2005.12.033
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011
- Zuo, X.-N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Li, Chen, Yue and Caffo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.