Decentralized Temporal Independent Component Analysis: Leveraging fMRI Data in Collaborative Settings[☆]

Bradley T. Baker[⋆†] Anees Abrol[⋆†] Rogers F. Silva[†] Eswar Damaraju[⋆†] Anand D. Sarwate[‡] Vince D. Calhoun[⋆†] Sergey M. Plis[⋆†]

[⋆] *University of New Mexico* [†] *Mind Research Network* [‡] *Rutgers, The State University of New Jersey*

Abstract

The field of neuroimaging has recently witnessed a strong shift towards data sharing; however, current collaborative research projects may be unable to leverage institutional architectures that collect and store data in local, centralized data centers. Additionally, though research groups are willing to grant access for collaborations, they often wish to maintain control of their data locally. These concerns may stem from research culture as well as privacy and accountability concerns. In order to leverage the potential of these aggregated larger data sets, we require tools that perform joint analyses without transmitting the data. Ideally, these tools would have similar performance and ease of use as their current centralized counterparts. In this paper, we propose and evaluate a new algorithm, decentralized joint independent component analysis (djICA), which meets these technical requirements. djICA shares only intermediate statistics about the data, plausibly retaining privacy of the raw information to local sites, thus making it amenable to further privacy protections, for example via differential privacy. We validate our method on real functional magnetic resonance imaging (fMRI) data and show that it enables collaborative large-scale temporal ICA of fMRI, a rich vein of analysis as of yet largely unexplored, and which can benefit from the larger-N studies enabled by a decentralized approach. We show that djICA is robust to different distributions of data over sites, and that the temporal components estimated with djICA show activations similar to the temporal functional modes analyzed in previous work, thus solidifying djICA as a new, decentralized method oriented toward the frontiers of temporal independent component analysis.

Keywords: Independent Component Analysis, temporal Independent Component Analysis, Decentralization, Collaborative Analysis, fMRI

1. Introduction

The benefits of collaborative analysis on fMRI data are deep and far-reaching. Research groups studying complex phenomena (such as mental disorders) often gather data with the intent of performing specific kinds of analyses. However, researchers can often leverage the data gathered to investigate questions beyond the scope of the original study. For example, a study focusing on the role of functional connectivity in mental health patients may collect a brain scan using magnetic resonance imaging (MRI) from all enrolled subjects, but may only examine one particular aspect of the data. The scans gathered for the study, however, are often saved to form a data set associated with that study—they therefore remain available for use in future research. This phenomenon often results in the accumulation of vast amounts of data, distributed in a decentralized fashion across many research sites. In addition, since technological advances have dramatically increased the complexity of data per measurement while lowering their cost, researchers hope to leverage data across multiple research groups to achieve sufficiently large sample sizes that may uncover important, relevant, and interpretable features that characterize the underlying complex phenomenon.

The standard industry solution to data sharing involves each group uploading data to a shared-use data center, such as a cloud-based service like the OpenfMRI data repository [1] or the more-recently proposed OpenNeuro service [2]. Despite the prevalence of such frameworks, centralized solutions may not be feasible for many research applications. For example, since neuroimaging uses data taken from human subjects, data sharing may be limited or prohibited due to issues such as (i) local administrative rules, (ii) local desire to retain control over the data until a specific project has reached completion, (iii) a desire to pool together a large external dataset with a local dataset without the computational and storage cost of downloading all the data, or (iv) ethical concerns of data re-identification. The last point is particularly acute in scenarios involving genetic information, patient groups with rare diseases, and other identity-sensitive applications. Even if steps are taken to assure patient privacy in centralized repositories, the repository maintainers are often forced to deal with monumental tasks of centralized management and standardization. This can require many hours of additional processing, occasionally reducing the richness of

[☆]This work was supported by grants from the NIH grant numbers R01DA040487, P20GM103472, and R01EB020407 as well as NSF grants 1539067 and 1631838. The author(s) declare that there was no other financial support or compensation that could be perceived as constituting a potential conflict of interest.

some of the contributed data [3].

In lieu of centralized sharing techniques, a number of practical decentralization approaches have recently been proposed by researchers looking to perform privatized analyses. For example, the "enhancing neuroimaging genetics through meta analysis" (ENIGMA) consortium [4] allows groups to share local summary statistics rather than gathering all the original imaging data at a single site for a centralized analysis. This method has proven very successful when using both mega- and meta-analysis approaches [4, 5, 6, 7]. Particularly, the meta-analysis at work in ENIGMA has been used for large-scale genetic association studies, with each site performing the same analysis, the same brain measure extraction, or the same regressions, and then aggregating local results globally. Meta-analyses can summarize findings from tens of thousands of individuals, so the summaries of aggregated local data need not be subject to institutional firewalls or even require additional consent from subjects [7, 8]. This approach represents one proven, widely used method for enabling analyses on otherwise inaccessible data.

Although ENIGMA has spurred innovation through massive international collaborations, there are some challenges which complicate the approach. Firstly, the meta-analyses at work in ENIGMA are effectively executed manually: a very time-consuming process. For each experiment, researchers have to write analysis scripts, coordinate with personnel at all participating sites to make sure these scripts are implemented there, adapt and debug scripts at each site, and then gather the results through the use of proprietary software. In addition, an analysis using the ENIGMA approach described above is typically "single-shot," i.e., it does not iterate among sites to compute results holistically, as informed by the global data. From a statistical and machine learning perspective, single-shot model averaging has asymptotic performance with respect to the number of subjects for some types of analysis [9, 10]. However, simple model averaging does not account for variability between sites driven by small sample sizes and cannot leverage multivariate dependence structures that might exist across sites. Furthermore, the ability to iterate over local site computations allows not only continuous refinement of the solution at the global level but also greater algorithmic complexity, enabling multivariate approaches like group ICA [11] and support vector machines [12], and increased efficiency due to parallelism, facilitating the processing of images containing thousands of voxels.

These, together with the significant amount of manual labor required for single-shot approaches to decentralization, motivates decentralized analyses which favor more frequent communication. For example, sites running a global optimization algorithm can communicate following each iteration or after a number of iterations. In this paper, we further previous work in this direction [13] to develop iterative algorithms for collaborative, decentralized feature learning. Namely, we implement a real-data application of a successful algorithm for decentralized independent component analysis (ICA), a widely-used method in neuroimaging applications. Specifically, we show that our decentralized implementation can help further advance the as-of-yet mostly unexplored domain of temporal ICA of functional magnetic resonance imaging (fMRI) data. The resulting method is a ready fit for decentralized collaboration frameworks, such as the COINSTAC neuro-imaging analysis platform [12], which promises innovation in privacy-sensitive decentralized analysis.

Decentralized approaches such as ENIGMA allow research sites to maintain control over data access, thus providing plausible privacy protection at the cost of additional labor in implementing and updating a distributed architecture. For many applications, keeping data stored on sites without transfer of entire data samples may provide substantial privacy. These decentralized methods, however, are amenable to quantifiable measures of privacy, such as differential privacy [14]. In this work, we leave the addition of differential privacy aside, and focus on the presentation of djICA as a separate algorithm first, with plausible privacy; however, we have pursued the addition of differential privacy to djICA elsewhere [15].

One widespread analysis which stands to benefit from decentralization is temporal independent component analysis (tICA). In resting-state fMRI studies, we can assume that the overall spatial networks remain stable across subjects and experiment duration, while the activation of certain neurological regions varies over time and across subjects. Temporal ICA, first utilized for fMRI by Biswal et al. [16], locates temporally independent components corresponding to independent activations of a subjects' intrinsic common spatial networks [17]. Both spatial and temporal ICA evidently provide reliable estimates of these intrinsic networks from fMRI data [18, 19, 20, 21], but, unlike its spatial counterpart, temporal ICA allows spatial correlation between them (i.e. overlaps in the spatial maps) [22]. Spatial and temporal ICA can result in similar estimated networks [23, 18, 24, 19], while temporal ICA provides estimates not otherwise available to spatial ICA [21, 25], specifically for task-related data. Temporal ICA has also proven particularly useful for extracting information from high-resolution fMRI scans with overlapping spatial activations, a feature not available to spatial ICA [26]. Beyond estimation of novel temporal components, temporal ICA can also aid in isolating and removing noise from fMRI signals [27, 28].

While useful, the existing literature for temporal ICA is limited. This can be partially attributed to computational complexity and dependence on statistical sample size, since temporal ICA requires more data points in the time dimension than the typical fMRI time series can offer [18, 19]. Specifically, the ratio of the spatial to the temporal dimension often requires the temporal dimension to be at least similar to the voxel dimension. This often motivates the temporal aggregation of datasets composed of many temporally concatenated subjects. This temporal aggregation is also a key feature of the well-established group spatial ICA in the fMRI literature [29, 30, 31]. Beyond accumulation of subjects, other studies implementing temporal ICA for fMRI utilize higher-resolution scans to perform temporal ICA with fewer subjects [26]. Further methods reduce the spatial dimension to make a temporal ICA tractable: Seifritz et al. [32] use an initial spatial ICA to reduce spatial dimensional by locating a region of interest on which to perform temporal ICA, and Van et al. restrict the temporal analysis to a prede-

termined region of voxels deemed relevant to their particular problem of speech pattern monitoring [33].

Although temporal ICA would benefit tremendously from increasing the temporal frequency of scanners, or analyzing a large number of subjects at a central location, as mentioned above, this is not always feasible. To overcome the challenges of centralized temporal ICA, we present a novel method, decentralized joint Independent Component Analysis (djICA), which allows for the computation of aggregate spatial maps and local independent time courses across decentralized data stored at different servers belonging to independent labs. Our approach combines individual computations performed locally with global processes to obtain both local and global results. The resulting method for temporal ICA produces results with similar performance to the pooled-data case and provides estimated components in line with previous literature, demonstrating the effectiveness of decentralized collaborative algorithms for this difficult task.

In sum, the contributions of this paper are as follows:

- In Section 2, we present decentralized joint independent component analysis (algorithm 1, Section 2.2), which is closely related to Infomax ICA (Section 2.1) with decentralized PCA preprocessing (Section 2.3).
- In Section 3 we include experiments and evaluation of djICA over different subject and site distributions for simulated data sets, including simulated fMRI data, thus providing a baseline result and proper motivation for real-data experiments.
- In Section 4, we perform experiments which evaluate djICA on a real set of fMRI data in a simulated decentralized environment, using a novel pseudo-ground-truth evaluation scheme to compare our results with the pooled case.
- Finally, in Section 5, we discuss the performance of djICA as a novel method for performing temporal ICA in decentralized settings, comparing our results with previously estimated results from the pooled temporal ICA literature.

2. Materials and Methods

In this section, we provide the details of our method for decentralized joint independent component analysis and provide a basis for its evaluation. We first review Independent Component Analysis for the pooled case (where all samples are located on a single site) in Section 2.1, which provides basis for our presentation of the djICA algorithm in section 2.2. In section 2.3 we discuss performing PCA preprocessing in a decentralized setting, and finally, in section 2.4, we discuss our methods for evaluating the djICA algorithm. The code used for evaluation is available on GitHub[1], and its inclusion in the COINSTAC decentralized analysis framework is currently ongoing.

[1] https://github.com/MRN-Code/djica_paper_code_release

2.1. Independent Component Analysis

ICA is a popular blind source separation (BSS) method which attempts to decompose mixed signals into independent components (ICs), or sources, without prior knowledge of the structure of those sources. Empirically, ICA applied to brain imaging data produces robust features which are physiologically interpretable and markedly reproducible across studies [31, 34, 35, 16]. Indeed, while justification for successful ICA of fMRI results had been previously attributed to sparsity alone [36], it has been shown that statistical independence between the underlying sources is in fact a key driving mechanism of ICA algorithms [37], with additional benefits possible by trading off between the two [38].

In linear ICA, we model a data matrix $\mathbf{X} \in \mathbb{R}^{d \times N}$ as a product $\mathbf{X} \approx \mathbf{AS}$, where $\mathbf{S} \in \mathbb{R}^{r \times N}$ is composed of N observations from r *statistically independent* components, each representing an underlying signal source. Thus, we can interpret ICA in terms of this generative model, with independent sources $\mathbf{S}$ submitted to a linear mixing process described by a mixing matrix $\mathbf{A} \in \mathbb{R}^{d \times r}$, forming the observed data $\mathbf{X}$. Most ICA algorithms seek to recover the "unmixing matrix" $\mathbf{W} = \mathbf{A}^{-1}$ (or in the case where $\mathbf{A}$ is not square, the pseudo-inverse, $\mathbf{A}^{+}$), by maximizing independence between rows of the product $\mathbf{WX}$, assuming the matrix $\mathbf{A}$ is invertible.

Maximal information transfer (Infomax) [39] is a popular heuristic for estimating $\mathbf{W}$ by maximizing an entropy functional related to $\mathbf{WX}$. This can alternatively be interpreted as a Bayesian estimator with a super-Gaussian prior on the density of the sources. More precisely, with some abuse of notation, let

$$g(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

be the sigmoid function with $g(\mathbf{Z})$ being the result of elementwise application of $g(\cdot)$ on the entries of a matrix or vector $\mathbf{Z}$. The differential entropy of a random vector Z with joint density p is

$$h(Z) = -\int p(Z) \log p(Z) dZ. \tag{2}$$

The objective of Infomax ICA then becomes

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}}\, h(g(\mathbf{WX})). \tag{3}$$

Another class of algorithms includes the famous family of fixed-point methods such as Fast ICA [40, 41, 42]. These locally optimize a "contrast" function such as kurtosis or negentropy.

ICA, along with other methods for BSS, has found wide application. In particular, functional magnetic resonance imaging (fMRI) and other biomedical imaging data use ICA models to interpret subject imaging data [35]. For fMRI, many models assume that functionally connected regions in the brain are systematically nonoverlapping. ICA has been used in applications ranging from interpreting physiology to analyzing task-related signals in both the spatial and temporal domains.

$\mathbf{X}$	$\mathbf{S}$	$\mathbf{A}$	$\mathbf{U\Sigma V}$
Data Matrix	Source Matrix	Mixing Matrix	SVD results
$\mathbf{X}_i$	$\mathbf{X}_{i,red}$	U_i	$\mathbf{G}_i(j)$
Data Site i	Reduced Data Site i	Eigenvectors Site i	Gradient Site i
$\Delta_{\mathbf{W}}(j)$	$\mathbf{W}(j)$	$\mathbf{b}(j)$	(j)
Weight Update	Weight Matrix	Bias	iter j
s	r	d	N
# sites	rank (# ICs)	# rows	# cols
ρ	w_{max}	θ_{max}	α
learning rate	max weight	max angle	anneal rate

Table 1: A summary of important notation used throughout this paper, especially in Algorithms 1, 2, and 3.

Additionally, a number of extensions of ICA exist for the purpose of jointly analyzing multiple data sets to perform a simultaneous decomposition across a large number number of subjects and different modalities [43, 44, 45]. Group spatial ICA (GICA) stands out as the leading approach for multi-subject analysis of task- and resting-state fMRI data [46], building on the assumption that the spatial map components ($\mathbf{S}$) are common (or at least similar) across subjects. Another approach, called joint ICA (jICA) [47], is popular in the field of multimodal data fusion and assumes instead that the mixing process ($\mathbf{A}$) over a group of subjects is common between a pair of data modalities.

A largely unexplored area of fMRI research is group temporal ICA, which, like spatial ICA, assumes common spatial maps but with statistically independent timecourses. Group temporal ICA has been most commonly applied to EEG data [48] but less frequently to fMRI data. Consequently, like jICA, in the fMRI case, the common spatial maps from temporal ICA describe a common mixing process ($\mathbf{A}$) among subjects. However, temporal ICA of fMRI is not typically investigated because the small number of time points in each data set can lead to unreliable estimates. Our decentralized jICA (djICA) approach overcomes that limitation by leveraging information from data sets distributed over multiple sites. This is an important extension of single-subject temporal ICA and a further example of methods which can benefit from leveraging data in collaborative settings.

2.2. Decentralized Joint ICA

Our goal in this paper is to show that the decentralized joint ICA algorithm can be applied to decentralized fMRI data and produce meaningful results for temporal ICA. We present djICA in detail here and provide notation in table.

For an integer n let $[n] = \{1, 2, \ldots, n\}$. Suppose that we have s total sites indexed by $[s]$; each site $i \in [s]$ has a data matrix $\mathbf{X}_i \in \mathbb{R}^{d \times N_i}$ consisting of a total time course of length N_i time points over d voxels. Let $N = \sum_{i=1}^{s} N_i$ be the total length. We model the data at each site as coming from a common (global) mixing matrix $\mathbf{A} \in \mathbb{R}^{d \times r}$ applied to local data sources $\mathbf{S}_i \in \mathbb{R}^{r \times N_i}$. Thus, the total model can be written as

$$\mathbf{X} = [\mathbf{AS}_1\ \mathbf{AS}_2\ \cdots\ \mathbf{AS}_s] \in \mathbb{R}^{d \times N}. \tag{4}$$

Our algorithm, decentralized joint ICA (djICA), uses locally computed gradients to estimate a common, global unmixing matrix $\mathbf{W} \in \mathbb{R}^{r \times d}$ corresponding to the Moore-Penrose pseudoinverse of $\mathbf{A}$ in (4), denoted $\mathbf{A}^+$.

Figure 1 summarizes the overall algorithm in the context of temporal ICA for fMRI data. Each site i has data matrices $\mathbf{X}_{i,m} \in \mathbb{R}^{d \times n_i}$ corresponding to subjects $m \in [M_i]$ with d voxels and n_i time samples. Sites concatenate their local data matrices temporally to form a $d \times n_i M_i$ data matrix $\mathbf{X}_i$, so the total time course length at site i is $N_i = n_i M_i$, and the total number of subjects is $M = \sum_{i=1}^{s} M_i$. Each site performs local PCA (Algorithm 2) using the singular value decomposition (SVD), with matrices $\mathbf{U}_i \in \mathbb{R}^{d \times k}$ and $\mathbf{\Sigma}_i \in \mathbb{R}^{k \times k}$ corresponding to the top k singular vectors and values, respectively. Then, in a decentralized principal component analysis (dPCA) framework, the sites approximate a global PCA (Algorithm 3) to form a common r-dimensional projection matrix $\mathbf{U} \in \mathbb{R}^{d \times r}$. This approach is an adaptation of the sub-sampled time PCA (STP) method [49] to the case of decentralized data, offering an accurate bandwidth-efficient alternative to other dPCA algorithms [50] which can compute the global $\mathbf{U}$ directly (without local PCA) but at the expense of communicating a large $d \times d$ matrix between sites. Finally, all sites project their data onto the subspace corresponding to $\mathbf{U}$ to obtain reduced local datasets $\mathbf{X}_{i,\text{red}} \in \mathbb{R}^{r \times N_i}$.

The projected data is the input to the iterative djICA algorithm that estimates the unmixing matrix $\mathbf{W} \in \mathbb{R}^{r \times r}$, as described in Algorithm 1. The full mixing matrix for the global data is modeled as $\mathbf{A} \approx (\mathbf{WU}^\top)^+ \in \mathbb{R}^{d \times r}$. After initializing $\mathbf{W}$ (for example, as the identity matrix), the djICA algorithm iteratively updates $\mathbf{W}$ using a distributed natural gradient descent procedure [51]. At each iteration j the sites update locally. In lines 5 and 6, the sites adjust the local source estimates $\mathbf{Z}_i = \mathbf{WX}_{i,\text{red}}$ by their bias estimates $\mathbf{b}(j-1)\mathbf{1}^\top \in \mathbb{R}^{r \times N_i}$, followed by the sigmoid transformation $g(\cdot)$; then, local gradients are computed with respect to $\mathbf{W}_i$ and $\mathbf{b}_i$ in lines 7 and 8. Here, $\mathbf{y}_{l,i}(j)$ is the l-th column of $\mathbf{Y}_i(j)$. The sites then send their local gradient estimates $\mathbf{G}_i(j)$ and $\mathbf{h}_i(j)$ to an aggregator site, which aggregates them according to lines 11-13. After updating $\mathbf{W}(j)$ and $\mathbf{b}(j)$, the aggregator checks if any values in $\mathbf{W}(j)$ increased above an upper bound of $w_{\max} = 10^9$ in absolute value. If so, the aggregator resets the global unmixing matrix, sets the current iteration to $j = 0$, and anneals the learning rate by $\rho = 0.9\rho$. Otherwise, before continuing, if the angle between $\Delta_{\mathbf{W}}(j)$ and $\Delta_{\mathbf{W}}(j-1)$ is above $\theta_{\max} = 60°$, the aggregator anneals the learning rate by $\rho = 0.9\rho$, preventing $\mathbf{W}$ from changing too quickly without learning the structure of data. The aggregator sends the updated $\mathbf{W}(j)$ and $\mathbf{b}(j)$ back to the sites. Finally, the algorithm stops when $\|\Delta_{\mathbf{W}}(j)\|_2^2 < t$, and each site recovers the statistically independent source estimates $\mathbf{S}_i$ by

$$\mathbf{S}_i \approx \mathbf{WX}_{i,\text{red}}. \tag{5}$$

For the pooled-data case, Amari et al. [52] demonstrate theoretically that Infomax ICA meets with the conditions that guarantee convergence of $\mathbf{W}$ to an asymptotically stable solution as long as A^{-1} is also asymptotically stable. In other words, the natural gradient provides convergence to an equilibrium point corresponding to a local minimum; however, in the general case for Infomax ICA, it is unfortunately not possible to assure con-

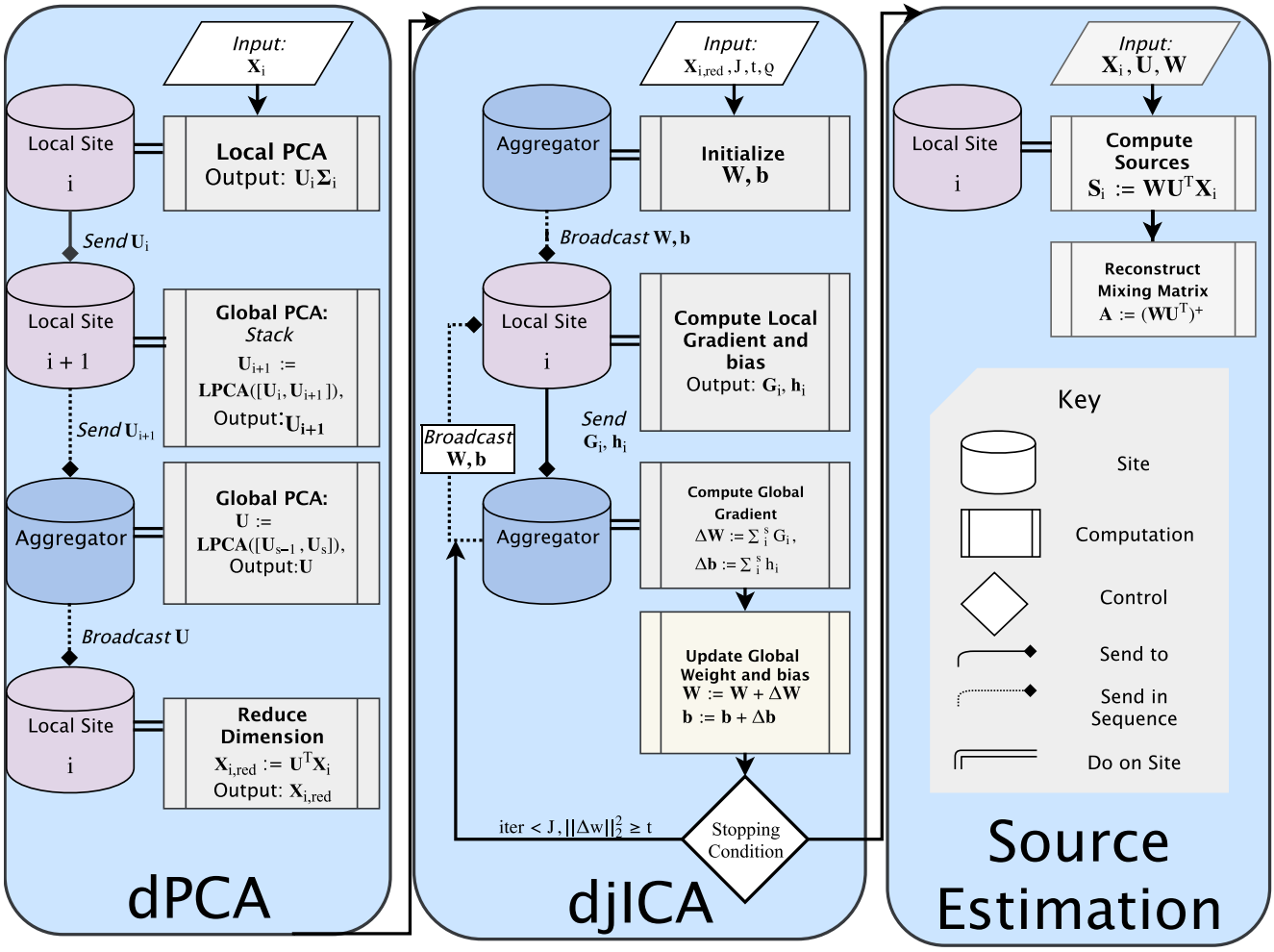

Figure 1: An overview of the djICA pipeline. Each panel in the flowchart represents one stage in the pipeline and provides an overview of the processes done on local sites and on the aggregator site, as well as communication between nodes. The dPCA panel corresponds to Algorithms 2 and 3, the djICA panel corresponds to Algorithm 1, and the Source Estimation panel corresponds to the procedure for computing local sources given in equation (5). On each panel, local site i is an arbitrary site in the decentralized network, and local site $i+1$ represents the next site in a given ordering over the decentralized network. *Broadcast* communication sends data to all sites, and *Send* communication sends data to one site. Lines with an arrowhead indicate procedural flow. Lines with diamond endpoints indicate communication flow. Dotted lines with diamond endpoints indicate that the sending process occurs iteratively to neighbors until the aggregator is reached, or in the case of broadcasting, indicates that all nodes receive the latest update. Double-lines indicate site-specific computations.

vergence to a global minimum, i.e. complete separation of the source signals.

In the decentralized-data case, djICA converges to the solution of the pooled-data case: the assumption of a common mixing matrix across subjects assures that the global gradient sum is identical to the pooled-data gradient on average, likewise moving the global weight matrix towards convergence.

Indeed, since the global iterates of djICA are taken as the average of the individually computed, on-site gradients, djICA run on a full-batch case (where each site has access to the full batch of data) is equivalent to the pooled version of infomax ICA. We show this empirically in section 1 of the supplementary material included with this work.

For our purposes, we chose the hyper-parameter values as specified in the "Required" parameters for Algorithm 1, and we utilized the stochastic version of the algorithm with block size $b = \left\lfloor \sqrt{\frac{\min(N_i)}{20}} \right\rfloor$, where $\min(N_i)$ is the minimum number of concatenated time-points across all sites. We summarize these parameters in Table 2.

param.	t	J	ρ	w_{max}	θ_{max}	α
value	10^{-6}	1024	0.015/ln(r)	10^9	60°	0.9

Table 2: A summary of the hyper-parameters used in all experiments, for both simulations, and real-data scenarios.

2.3. PCA preprocessing

Here, we describe the decentralized principal component analysis (dPCA) algorithms used for dimension reduction and whitening in the djICA pipeline. The dPCA algorithm is a preprocessing step that standardizes the data prior to djICA and should also be decentralized so that the benefits of using a decentralized joint ICA are not made moot by dependence on a previous pooled step. There are many approaches to approximating the global PCA with a distributed algorithm [53].

Algorithm 1 decentralized joint ICA (djICA)

Require: data $\{\mathbf{X}_{i,\text{red}} \in \mathbb{R}^{r\times N_i} : i \in [s]\}$, where r is the same across sites, tolerance level $t = 10^{-6}$, $j = 0$, maximum iterations $J = 1024$, initial learning rate $\rho = 0.015/\ln(r)$, maximum weight entry $w_{\max} = 10^9$, maximum angle $\theta_{\max} = 60°$, annealing rate $\alpha = 0.9$

Initialize $\mathbf{W}(0) \in \mathbb{R}^{r\times r}$ ▷ for example, $\mathbf{W}(0) = \mathbf{I}$
for $j < J$ **and** $\|\Delta_{\mathbf{W}}(j)\|_2^2 \geq t$ **do**
  $j = j + 1$
  for all sites $i = 1, 2, \ldots, s$ **do**
    $\mathbf{Z}_i(j) = \mathbf{W}(j-1)\mathbf{X}_{i,\text{red}} + \mathbf{b}(j-1)\mathbf{1}^\top$
    $\mathbf{Y}_i(j) = g(\mathbf{Z}_i(j))$
    $\mathbf{G}_i(j) = \rho\,(\mathbf{I} + (\mathbf{1} - 2\mathbf{Y}_i(j))\mathbf{Z}_i(j)^\top)\mathbf{W}(j-1)$
    $\mathbf{h}_i(j) = \rho \sum_{l=1}^{N_i}(\mathbf{1} - 2\mathbf{y}_{l,i}(j))$
    Send $\mathbf{G}_i(j)$ and $\mathbf{h}_i(j)$ to the aggregator site.
  end for
  At the aggregator site, update global variables
  $\Delta_{\mathbf{W}}(j) = \sum_{i=1}^{s} \mathbf{G}_i(j)$
  $\mathbf{W}(j) = \mathbf{W}(j-1) + \Delta_{\mathbf{W}}(j)$
  $\mathbf{b}(j) = \mathbf{b}(j-1) + \sum_{i=1}^{s} \mathbf{h}_i(j)$
  Check Upper-Bound Conditions
  if $w_{i,j} \in \mathbf{W}, |w_{i,j}| > w_{\max}$ **then** ▷ $\mathbf{W}$ has blown-up
    Re-Initialize $\mathbf{W}(0)$, $j = 0$, $\rho = \alpha\rho$
  else if $\angle(\Delta_{\mathbf{W}}(j), \Delta_{\mathbf{W}}(j-1)) > \theta_{\max}$ **then**
    $\rho = \alpha\rho$ ▷ Prevent $\mathbf{W}$ from changing too quickly
  end if
  Broadcast global $\mathbf{W}(j)$ and $\mathbf{b}(j)$ to all sites.
end for

Algorithm 2 Local PCA algorithm (LocalPCA)

Require: data $\mathbf{X} \in \mathbb{R}^{d\times N}$ and intended rank k

Compute the SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$.
Let $\mathbf{\Sigma}^{(k)} \in \mathbb{R}^{k\times k}$ contain the largest k singular values and $\mathbf{U}^{(k)} \in \mathbb{R}^{d\times k}$ the corresponding singular vectors.
Save $\mathbf{U}^{(k)}$ and $\mathbf{\Sigma}^{(k)}$ locally and return $\mathbf{P} = \mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)}$.

Algorithm 3 Global PCA algorithm (GlobalPCA)

Require: s sites with data $\{\mathbf{X}_i \in \mathbb{R}^{d\times N_i} : i = 1, 2, \ldots, s\}$, intended final rank r, local rank $k \geq r$.

Choose a random order π for the sites.
$\mathbf{P}(1) = \text{LocalPCA}(\mathbf{X}_{\pi(1)}, \min\{k, \text{rank}(\mathbf{X}_{\pi(1)})\})$
for all $j = 2, 3, \ldots, s$ **do**
  Set site index $i = \pi(j)$
  Send $\mathbf{P}(j-1)$ from site $\pi(j-1)$ to site $\pi(j)$
  $k' = \min\{k, \text{rank}(\mathbf{X}_i)\}$
  $\mathbf{P}' = \text{LocalPCA}(\mathbf{X}_i, k')$
  $k' = \max\{k', \text{rank}(\mathbf{P}(j-1))\}$
  $\mathbf{P}(j) = \text{LocalPCA}([\mathbf{P}'\ \mathbf{P}(j-1)], k')$
end for
$r' = \min\{r, \text{rank}(\mathbf{P}(s))\}$
$\mathbf{U} = \text{NormalizeTopColumns}(\mathbf{P}(s), r')$ ▷ At last site
Send $\mathbf{U}$ to sites $\pi(1), \ldots, \pi(s-1)$.
for all sites $i = 1, 2, \ldots, s$ **do**
  $\mathbf{X}_{i,\text{red}} = \mathbf{U}^\top\mathbf{X}_i$ ▷ The locally reduced data
end for

We first chose to examine dPCA from Bai et al. [50]. Their proposed dPCA algorithm bypasses local data reduction, and thus works directly with the full data, which motivates its choice for some of our simulated experiments. One major downside of their approach, however, is that it requires the transfer of a large orthogonal matrix between all sites, thus increasing bandwidth usage significantly. As an alternative to the approach presented by Bai et al., a two-step dPCA approach was considered based on the STP approach [54] recently developed for large PCA of multi-subject fMRI data. One advantage of this approach is that only a small matrix $\mathbf{P} \in \mathbb{R}^{d\times k}$ is transmitted from one site to another, a significant decrease compared to the large $d \times d$ matrix [50]. The downside is that there are no bounds on the accuracy of the final $\mathbf{U}$ and results can vary slightly with the order in which sites and subjects are processed. Nonetheless, our results suggest that the two-step dPCA approach, described in Algorithms 2 and 3, yields a fairly good estimate of $\mathbf{U}$. In principle, any suitable decentralized PCA algorithm could replace the two methods tested here. Thus, we leave room for future improvements of our framework to find the most effective dPCA approach for the djICA pipeline.

Algorithm 3 uses a peer-to-peer scheme to iteratively refine $\mathbf{P}(j)$, with the last site broadcasting the final $\mathbf{U}$ to all sites. $\mathbf{U}$ is the matrix containing the top r' columns of $\mathbf{P}(s)$ with largest L_2-norm, but normalized to unit L_2-norm instead. Following the recommendation in Calhoun et al. [54], we set $r = 20$ and $k = 5 \cdot r$ for our simulations.

2.4. Evaluation Strategy

All of our experiments were run using the MATLAB 2007b parallel computation toolbox, on a Linux Server running Ubuntu 12.04 LTS, with a 9.6GHz processor (four Intel Xeon E7-4870 @ 2.40GHz each), a 120MB L3 cache (30MB L3 cache per processor), and 512GB of RAM. For any one experiment, we only used a maximum of 8 cores, due to a need to share the server with other researchers.

As a performance metric for our experiments we choose the Moreau-Amari [51] inter-symbol interference (ISI):

$$ISI(Q) = \frac{1}{2r(r-1)}\Bigg[\sum_{i=1}^{r}\Big(\sum_{j=1}^{r}\frac{|Q_{ij}|}{max_k|Q_{ik}|} - 1\Big) + \sum_{j=1}^{r}\Big(\sum_{i=1}^{r}\frac{|Q_{ij}|}{max_k|Q_{kj}|} - 1\Big)\Bigg]. \tag{6}$$

This is a function of the square matrix $\mathbf{Q} = \hat{\mathbf{W}}\mathbf{A}$, where $\hat{\mathbf{W}} = \mathbf{W}\mathbf{U}^\top$, $\mathbf{W}$ is the estimated unmixing matrix from Algorithm 1, $\mathbf{U}$ is the orthonormal projection matrix retrieved from dPCA, and $r = \text{rank}(\mathbf{Q})$, i.e. the number of sources. In particular, a lower ISI measure indicates a better estimation of a set of ground-truth components.

3. Experiments with Simulated Data

First, we test djICA in a simulated environment where we can manufacture a known ground-truth and use djICA to reconstruct this ground-truth under different mixing and site configurations. For this simulated case, we explicitly construct the signal matrices, $\mathbf{S}$, and the mixing matrix $\mathbf{A}$ (using the methods described in section 3), such that the source matrices are statistically independent and provide, thus providing the assurance that a solution to underyling BSS problem exists. If djICA performs well in this simulated case, where a solution is given, we can thus justify further experiments with real data, where a solution to the underlying BSS problem is not readiily available. To this end, we evaluated 5 different scenarios for synthesis and analysis of synthetic data, as summarized in Table 3. Based on what we have learned from these various scenarios, which include different PCA preprocessing strategies, we can construct a promising pipeline for djICAwhich can then be translated to the real data case.

Two kinds of mixing matrices $\mathbf{A}$ were used for experimentation:

1. Lower dimensional square mixing matrices were generated using MATLAB's `randn` function [55], which generates matrices whose elements are selected from an i.i.d. Gaussian distribution.

2. Higher dimensional mixing matrices were generated using the MIALab's fMRI simulation toolbox (simTB) [56]. The simTB spatial maps are intended to simulate spatial components of the brain which contribute to the generation of the simulated time course. Higher dimensional mixtures were masked using a simple circular mask which drops empty voxels outside of the generated spatial map.

For the first two scenarios indicated in Table 3, we generated i.i.d. Gaussian mixing matrices $\mathbf{A} \in \mathbb{R}^{r\times r}$. For the higher-dimensional problems (scenarios 3-5), we used the simTB spatial maps [56] to generate different $\mathbf{A} \in \mathbb{R}^{d\times r}$ mixing matrices.

The independent signals $\mathbf{S}_m$ were simulated using a generalized autoregressive (AR) conditional heteroscedastic (GARCH) model [57, 58], which has been shown to be useful in models of causal source separation [59] and time-series analyses of data from neuroscience experiments [59, 60], especially resting-state fMRI time courses [61, 62]. We simulated fMRI time courses using a GARCH model by generating an AR process (no moving average terms) randomly such that the AR series converges. We chose a random order between 1 and 10 and random AR coefficients $\{\alpha[\ell]\}$ such that $\alpha[0] \in [0.55, 0.8]$ and $\alpha[\ell] \in [-0.35, 0.35]$ for $\ell > 0$. For the error terms $\delta_t = \sigma_t \epsilon_t$, we used an ARMA model driven by ϵ_t from a generalized normal distribution with shape parameter 100 (so it was approximately uniform on $[-1, 1]$) and $\sigma_t^2 = 0.1 + 0.1y[t-1]^2 + 0.75\sigma[t-1]^2$. For each of 1024 simulated subjects, we generated 20 time courses with 250 time points, each after a "burn-in" period of 20000 samples, checking that all pair-wise correlations between the 20 time courses stayed below 0.35. We generated a total of 1024 mixed datasets for each experiment by computing $\mathbf{X}_m = \mathbf{A}\mathbf{S}_m$.

In summary, we considered the following combinations of algorithm, preprocessing, and mixing matrix: 1) pooled (centralized) temporal ICA with no preprocessing (no data reduction) and a square i.i.d Gaussian mixing-matrix, 2) djICA with no preprocessing and a square i.i.d. Gaussian mixing-matrix, 3) pooled temporal ICA with LocalPCA preprocessing (Algorithm 2) and a simTB mixing matrix, 4) djICA with dPCA from Bai et al. [50] and a simTB mixing matrix, and 5) djICA with GlobalPCA (Algorithm 3) and a simTB mixing matrix.

scenario	algorithm	preprocessing	mixing matrix A
1	ICA (pooled)	none	i.i.d. Gaussian
2	djICA	none	i.i.d. Gaussian
3	ICA (pooled)	LocalPCA	simTB map
4	djICA	One-Step dPCA [50]	simTB map
5	djICA	GlobalPCA	simTB map

Table 3: Five scenarios considered for synthesis and analysis of simulated data experiments.

3.1. Simulation Results

In this section, the results for simulated experiments are presented. We are particularly interested in understanding how the proposed algorithm performs with different kinds of preprocessing, and how the results improve as a function of the global number of subjects, the global number of sites, or how the subjects are distributed over sites.

To test how the algorithms compare as we increase the data at a fixed number of sites, we fixed $s = 2$ sites and evaluated all five scenarios in Table 3, splitting the data evenly per site in the non-pooled cases. Figure 2a shows ISI versus the total data set size. As the data set increases all algorithms improve and, more importantly, the distributed versions perform nearly as well as the pooled-data counterparts. Results are averaged over 10 randomly generated mixing matrices.

To test how the algorithms compare as we increase the number of sites s, we fix $M_i = 32$ subjects per site. Figure 2b demonstrates the convergence of the ISI curve with an increasing amount of combined data, with results averaged over 10 randomly generated mixing matrices. Again, we see that the performance of djICA is very close to the centralized pooled performance, even for such a small number of subjects per site.

To test how splitting the data sets across more sites affects performance, we fixed the total of 1024 subjects and investigated the effect of splitting them over a growing number of sites s. Thus, the concentration of data per site M_i decreased with increasing number of sites such that for small s each site had more data sets and for large s each site had fewer data sets. Figure 2c shows that the performance of djICA is very close to that of the pooled-data ICA, even with more and more sites holding fewer and fewer data points. This implies that we can support largely decentralized data with little loss in performance.

4. Experiments with Real Data

The simulated experiments illustrate the clear benefit djICA provides by enabling the joint analysis of large decen-

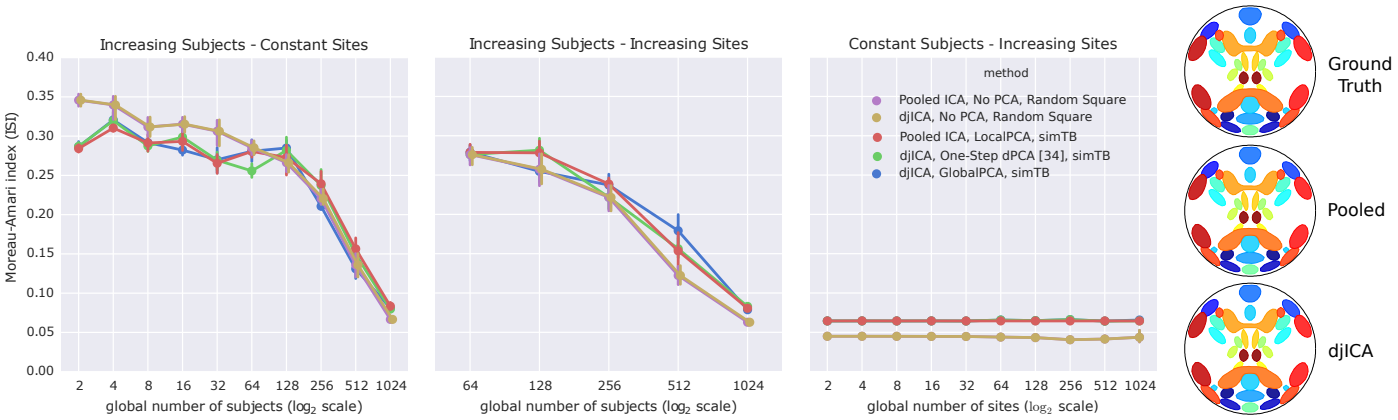

Figure 2: The ISI for pooled and decentralized algorithms for different distributions of subjects over sites under the five simulated scenarios indicated in Table 3. Panel 2a illustrates an increasing number of subjects over two, fixed sites. Panel 2b illustrates an increasing number of sites, with the number of subjects per site staying constant at 32 subjects per site, with the number of sites starting at 2 and increasing by a factor of two. Panel 2c illustrates 1024 total subjects distributed over an increasing number of sites. Panel 2d shows the 20 Ground-Truth spatial-maps, along with the estimated spatial-maps from Pooled ICA and djICA with 1024 subjects on 2 sites. In the cases with no PCA (panels 2a-**??**), the pooled and decentralized algorithms perform identically.

tralized data sets. In this section, we describe the methods utilized for real-data experiments with resting-state fMRI datasets. These experiments are intended to illustrate the effectiveness of djICA (Algorithm 1) in the particular domain of exploratory analysis of fMRI data. As mentioned earlier, the benefits of using this algorithm for fMRI analysis are numerous, and the experiments here aim to both highlight those benefits and illustrate the robustness of the algorithm when compared to pooled analyses.

4.1. Data Description

In this section, we describe the data sets used for real data analysis. The purpose here is to describe the preprocessing steps specific to the data utilized here. Experiments used data gathered on-site, according to the protocol in [46]. The data were collected using a 3-Tesla Siemens Trio scanner with a 12-channel radio frequency coil. T2*-weighted functional images were acquired using a gradient-echo EPI sequence with TE = 29 ms, TR = 2 s, flip angle = 75°, slice thickness = 3.5 mm, slice gap = 1.05 mm, field of view 240 mm, matrix size = 64×64, voxel size = 3.75 mm × 3.75 mm × 4.55 mm. In terms of duration, resting-state scans were a minimum of 2 min 8 s (64 volumes) long, on average 5 min 16 s (158 volumes) long, and at maximum 10 min 2 s (301 volumnes) long (see Table 4). In contrast to [46], subjects with greater number of time-points were retained in order to illustrate the general robustness of djICA to variation in the time-course length.

In terms of preprocessing, the data underwent rigid body alignment for head motion, slice-timing correction, spatial normalization to MNI space (using SPM5), regression of 6 motion parameters and their derivatives in addition to any trends (up to cubic or quintic), and spatial smoothing using a 10 mm^3 full-width at half-maximum (FWHM) Gaussian kernel.

We also used the minimum description length (MDL) criterion [63] to estimate the number of independent components for each individual subject with the algorithm available in the

min	mean	mode	median	max	range	std
64	158	158	158	301	237	9

Table 4: Statistics on the number of timepoints in the data set.

MIALab's Group ICA of fMRI toolbox (GIFT) [64, 65, 11]. experiment. The median number of components over 2038 subjects was 50, and the mean was 49.4636. In all experiments, we thus elected to estimate $r = 50$ real components from the data.

4.2. Real Data with "Real" Ground-Truth

Our ultimate goal is to show that djICA can provide reasonable decentralized estimates for real fMRI components which are comparable to the pooled case. Thus, we first perform a pooled analysis in order to establish a "pseudo" ground-truth that can be used to evaluate djICA's performance on real component estimation. We estimated $r = 50$ real independent components from $M = 2038$ subjects by running a pooled instance of temporal ICA on a single site. The performance of djICA was assessed by matching the estimated decentralized components to the pooled components via the Hungarian algorithm [66] and then computing the ISI between the two sets of components. For PCA preprocessing, to avoid high communication costs, we elect to test only the GlobalPCA method given in algorithm 3.

Using the pooled estimations as our basis for comparison, we then tested djICA in four distinct scenarios, varying the distribution of subjects across a network as follows:

1. when the global number of subjects in the network increases, but the number of sites in the network stays constant,
2. when the number of subjects per site stays constant, and the number of sites in the network increases,
3. when the global number of subjects in the network stays constant, but the number of sites in the network increases (subjects distributed evenly across sites), and

4. when the global number of subjects in the network stays constant, and subjects are randomly distributed across sites.

For the first three scenarios, we cap the maximum number of subjects in the network at 1024 so that we can achieve an even distribution of subjects in terms of powers of 2, and so that the figures compare more directly with the simulated experiments. Furthermore, in order to get a more detailed picture of the effects of small numbers of subjects per site, we also evaluate djICA in the third scenario with a higher global number of subjects ($M = 2000$), and closely examine the results in Figure 4. For each of these scenarios, we performed 10 repeated estimations of djICA components, where each run randomly assigned subjects to different sites (without duplication of subjects).

For the fourth and last scenario, 2000 subjects were randomly distributed across sites. Firstly, we selected a parametric probability distribution $P(\Theta)$ with parameters Θ. We then sampled 100 different values from P, where each value corresponds with a potential number of subjects on the i-th site (M_i). We discarded any values below 4, so that each site has a minimum of 4 subjects per site, and took the ceiling of each real value so that site distributions are given as natural numbers. We then selected the first $s-1$ values, with s being the number of sites, such that $\sum_{i=1}^{s-1} M_i < M$, where M is the global number of subjects in the network. For the final site, we set $M_s = M - \sum_{i=1}^{s-1} M_i$ so that the total number of subjects in the network will remain constant at M. This process results in a varying number of sites between successive samplings, which we found more appealing for testing as opposed to a randomization method that would distribute a fixed number of subjects across a fixed number of sites. We also considered the effect of different distribution parameter values (Θ) to assess the performance of djICA.

4.3. "Real" Ground-Truth Results

In this section, we present the results of djICA on the four experiments described above. In all cases, djICA is compared with a pooled case involving $M = 2038$ subjects, comparing across conditions using the Moreau-Amari ISI index as we did in the simulated experiments, now treating the pooled case as a our real-data "ground-truth".

4.3.1. How do the estimated components compare as we increase the data, with a fixed number of sites?

In Figure 3a, we evaluate the ISI index for djICA using real-data in a scenario where the global number of subjects increases, but the number of sites in the network is fixed. This figure illustrates that as the number of subjects increases, the estimated djICA components converge towards the components computed in the pooled case.

4.3.2. How do the estimated components compare as we increase the number of sites, with a fixed amount of data sets per site?

In Figure 3b, we evaluate the ISI index for djICA using real-data in a scenario where the number of subjects on each site is held constant, while the number of sites in the network increases. This figure further illustrates that as the global number of subjects increases, the estimated djICA components converge towards the components computed in the pooled case. Indeed, 1024 global subjects was sufficient for good performance across the smaller network.

4.3.3. How does spreading the data sets across more sites affect performance?

In Figure 3c, we evaluate the ISI index for djICA using real data in a scenario where the global number of subjects across the entire network is held constant, while the number of sites in the network increases. This figure illustrates that it is the global number of subjects included in the analysis, rather than the number of subjects per site, that mostly affects the performance of djICA. The concentration of subjects per site only begins to affect the performance of djICA when it is very low. At four subjects per site (256 sites in panel 3c), the performance is slightly worse than in previous runs. Thus, in Figure 4, we provide more detailed results for the particular scenario using 2000 subjects in order to illustrate the effects of low number of subjects per site.

The three cases A, B, and C in Figure 4 illustrate the performance of the algorithm for the minimum, median and maximum inter-symbol interference (ISI) scenarios respectively on an increasing number of sites, from 10 repeated runs for each subject-site distribution. Each run randomly placed different subjects on different sites. For all three cases, the plot to the left illustrates the correlations of each pooled ICA component with their corresponding match in each decomposition. It is evident from these plots that the correlation values for the majority of the 50 components clustered tightly above the mean correlation values (black horizontal bar) for the entire range of number of sites, thus suggesting that the mean correlation values for all cases were driven to a lower value by a few outliers (poorly replicated components). In fact, the lowest mean for the worst ISI case was 0.83 ($s = 1000$). However, in general, the mean value of the component correlations for any given case decreased with an increase in number of sites.

A one-way analysis of variance (ANOVA) and multiple comparison of means tests were used to determine the specific cases in which the mean Fisher-transformed correlation estimates (z-space) were significantly different (at a corrected significance level of 5% using Tukey's honest significant difference (HSD) method). For the minimum ISI case, significant differences were observed in the first five decompositions ($s = 4$, 10, 25, 50, and 125) as compared to the last decomposition ($s = 1000$), the first two decompositions ($s = 4$ and 10) as compared to the second-to-last ($s = 500$), and the first decomposition ($s = 4$) as compared to the fourth, fifth and sixth ($s = 50$, 125 and 250). The median ISI case showed an almost identical set of differences, except for one less significant difference between the first *four* decompositions ($s = 4$, 10, 25 and 50) as compared to the last decomposition ($s = 1000$). Finally, for the maximum ISI case, the first four decompositions ($s = 4$, 10, 25 and 50) had mean correlation estimates significantly different than the last two ($s = 500$ and 1000), and the first decomposition ($s =$

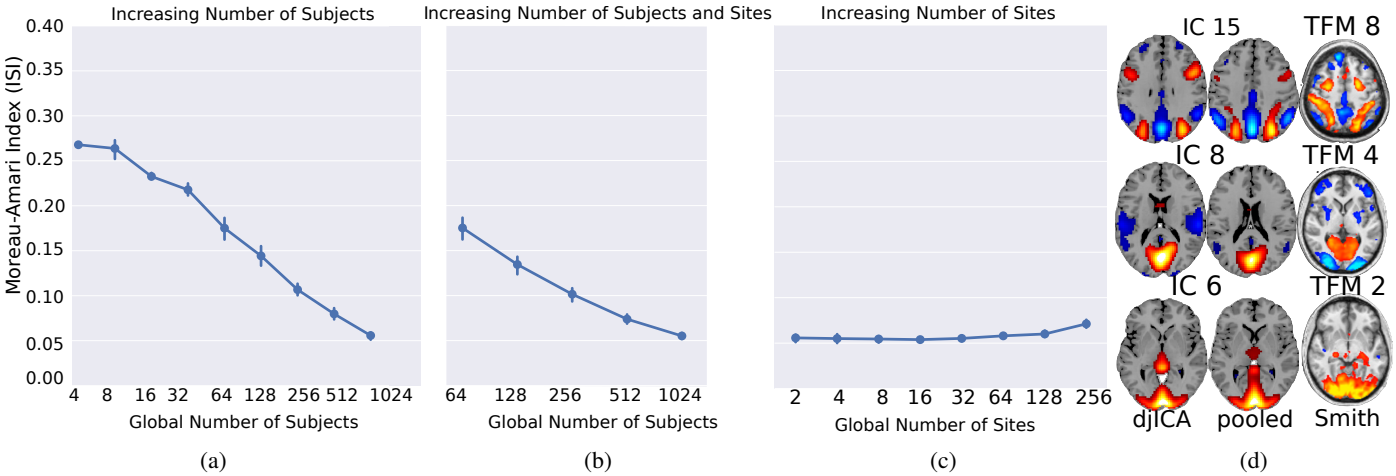

Figure 3: The estimated ISI for real-data djICA over different distributions of subjects over sites. Panel 3a illustrates an increasing number of subjects over two fixed sites. Panel 3b illustrates an increasing number of sites, with the number of subjects per site staying constant. Panel 3c illustrates 1024 global subjects distributed over an increasing number of sites. Panel 3d shows three of the spatial maps from djICA with over 2016 subjects evenly split over 16 sites, the pooled temporal ICA "pseudo ground-truth" with 2038 subjects, and the corresponding temporal fluctuation modes (TFM) from Smith et al. [21].

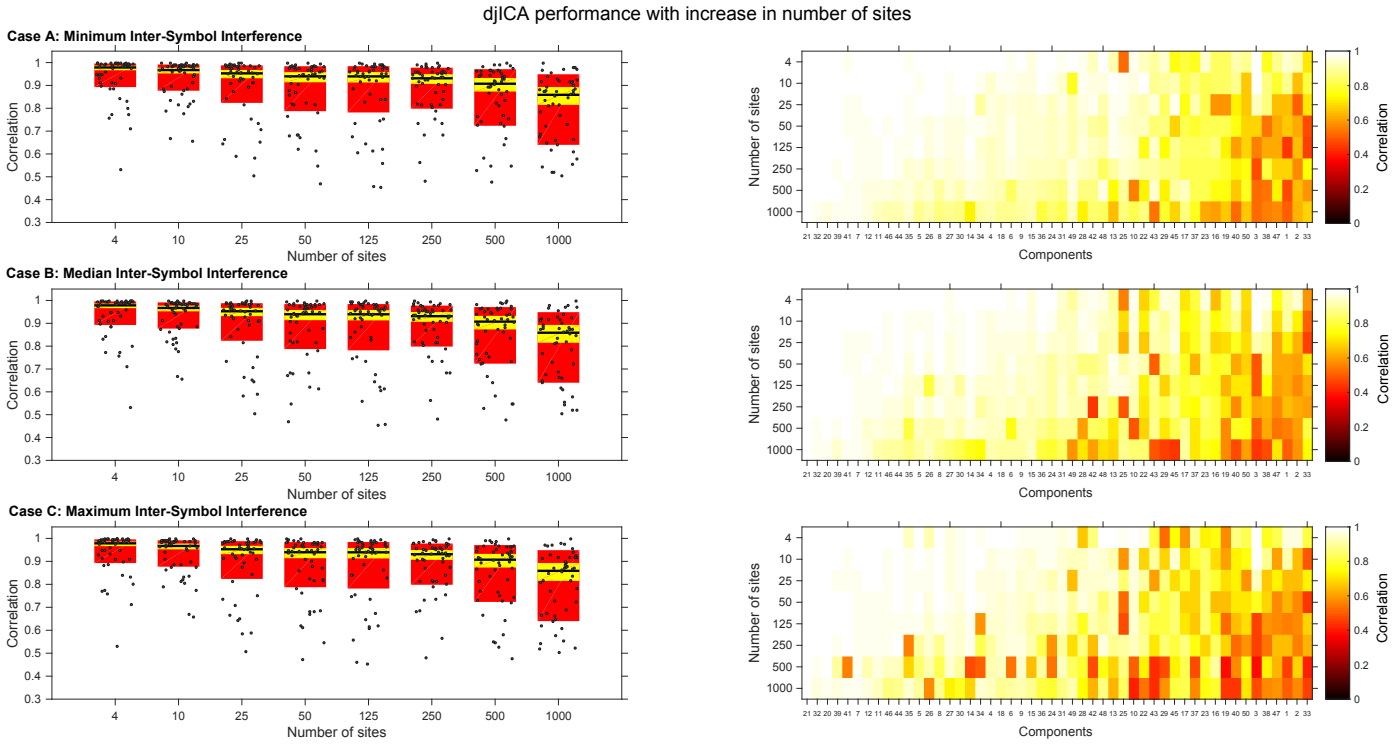

Figure 4: Keeping the number of subjects fixed at 2000 and increasing the number of sites, we examine the correlations of the estimated components from djICA with the corresponding best match component from the pooled ICA case. The plots to the left illustrate the correlations between pooled ICA components and their best matched djICA components, from runs with the minimum (Case A, best), median (Case B), and maximum (Case C, worst) ISI selected out of 10 total runs. On the box-plots, the black horizontal bar represents the mean value of the Fisher-transformed correlations (z-space) for a specific decomposition transformed back to correlation space (r-space), the yellow shaded areas give the 95% confidence intervals of the Fisher-transformed correlations (z-space) transformed back to r-space, and the red box boundaries show the sample standard deviation over the Fisher-transformed correlations (z-space) transformed back to r-space. The panels to the right are a component-specific depiction of the similarity between the estimated djICA components and their corresponding pooled ICA component. Lighter colors indicate that the estimated component highly resembled the pooled ICA component estimated from 2038 subjects. The components (columns) are arranged in descending order of correlations for the minimum ISI case, and this sorting order was retained for the median and maximum ISI cases.

4) showed additional significant differences as compared to the fifth and sixth (s = 125 and 250). This overall pattern clearly indicates: (1) deterioration of performance with lower number of subjects per site; (2) significantly lower mean correlations for very low number of subjects per site (8 at s = 250, 4 at s = 500, and 2 at s = 1000).

For all cases A, B, and C, the component-specific performance of the algorithm can be traced in the correlation intensity

images to the right. It is evident that all three cases feature high correlations for most of the components, especially in decompositions with lower number of sites. However, there are a few components that exhibit poor (or outlying) performance. For example, correlations for the last few (rightmost) components degrade significantly in decompositions with higher number of sites, while in the minimum ISI case the correlation for component 25 is unexpectedly low for the first decomposition ($s = 4$) and higher for the remaining decompositions. Finally, it can be concluded from these correlation images that although the algorithm performance degrades with increasing number sites, the same set of components tends to be consistently well replicated.

4.3.4. How does randomly splitting the data sets across more sites affect performance?

In real-world scenarios, fMRI data is not evenly distributed across research sites, thus motivating an investigation into the effect of randomly distributing the number of subjects per site on the effectiveness of djICA . In Figure 5, we compare the estimated ISI of djICA using real-data in a scenario where 2000 subjects are randomly assigned to sites by sampling the number of subjects on each site from a given distribution $P(\Theta)$. We tested three different distributions for site assignment: normal, exponential, and uniform, setting the mean and standard deviation both at 128. The sampling process generates nodes (research sites) with different dataset sizes, and we then run djICA and compute the ISI as given above. We ran djICA five times for each distribution, resampling the number of subjects per site each time, and then plotted the ISI for each run. Each panel in figure 5 also illustrates a graph of a network where each site is represented as a node. The size of each node in the network corresponds to the number of subjects on a given site, which is sampled from the given distribution.

As the figure shows, uniformly distributing subjects across the network reduces variance in computations when compared to normally distributed subjects; however, all of the given runs do not vary more than 0.02 with respect to the ISI, and all fall below 0.1 ISI, indicating favorable performance.

4.3.5. How do the estimated maps compare with previous results?

In Figure 3d we provide the spatial maps from three of the highest-correlated components estimated using djICA and pooled ICA for comparison to their corresponding temporal fluctuation modes (TFM) from Smith et al. [21], which also investigated temporal ICA of fMRI. We discuss this comparison in the following section.

5. Discussion

In contrast to systems optimized for processing large amounts of data by making computation more efficient (Apache Spark, H2O and others), we focus on a different setting common in research collaborations: data are expensive to collect, are spread across multiple sites, and possibly not shareable directly. To that end, we proposed a *distributed data* joint ICA algorithm that, in synthetic experiments, finds underlying sources in decentralized data nearly as accurately as its centralized counterpart. This shows that algorithms like djICA may enable collaborative processing of decentralized data by combining local computation and communication of local summaries. djICA represents an important iteration towards toolboxes for computing on data distributed across private sites with an emphasis on collaboration. While other distributed methods for decentralized fMRI analysis have been recently proposed [67, 68], djICA in particular is able to benefit from the unique opportunity of globally accumulated multi-subject data.

To further validate our method we have evaluated it in experiments on real fMRI data. Our use of djICA to perform temporal ICA of fMRI produces results which compare well to the pooled version of the algorithm and produces estimated components which compare well with other work on temporal fMRI analysis [21] that uses much more elaborate multi-step analyses techniques. Additionally, djICA is robust to random allocation of subjects to sites, generally performing well with a high number of globally accumulated subjects, and insensitive to how these subjects are distributed across the sites. We have discovered one edge-case for real-data djICA in which having less than four subjects per site across all sites in the network leads to a slight decrease in global performance. While further investigation using a robust hyper-parameter search (which we did not pursue in this paper) may mitigate this performance reduction, the scenario where all or many sites in a collaborative analysis would each have fewer than four subjects is highly unlikely. Other decentralized approaches to fMRI analysis, such as approaches which use the ENIGMA consortium [4], do not explore this edge-case. Indeed, the lowest number of subjects on a site within the ENIGMA consortium was 36, with the majority of other sites in the consortium possessing over 200 subjects [7].

Our decentralized djICA algorithm is a good fit for decentralized collaborative frameworks, such as the COINSTAC collaboration platform, and is amenable to the privacy guarantees including in those platforms. The inclusion of djICA in a system like COINSTAC would allow for shared analysis between members of pre-arranged consortia without the exchange of raw data. This alone provides a level of plausible privacy to djICA which is not available to centralized ICA approaches. As we have explored elsewhere, djICA can be easily extended to include quantifiable notions of privacy, such as differential privacy [15]. Further investigation is required, however, to investigate the robustness of both plausible and differential privacy to scenarios involving malicious participants in the consortium. For example, it has been shown that malicious participants in a collaborative classification task using a decentralized Deep Neural Network can reconstruct data samples by utilizing Generative Adversarial Networks to leverage shared gradient information [69]. It is currently unclear whether or not methods such as djICA suffer from this information leakage issue; however, the issue demands future attention.

Privacy aside, real-world networks can suffer from a number of additional implementation issues: individual sites may have different computing hardware and messages may be dropped

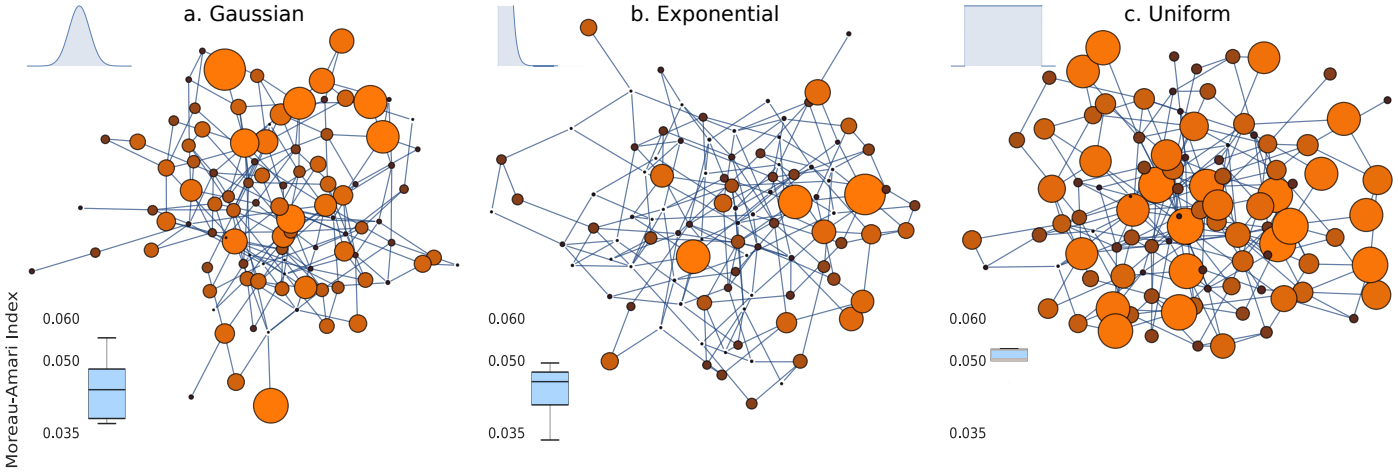

Figure 5: An illustration of the effect of randomly distributing subjects across a decentralized data network. Each panel contains an example graph of connected nodes in the network, where each node represents a site in the network. The size of each node in the network corresponds to the number of subjects located on that site. The estimated ISI is computed after running djICA over 5 repeated runs, where each distinct run utilized a different network sampled from the same distribution. Panel (a) illustrates a network where the number of subjects on each site was sampled from a gaussian distribution, panel (b) illustrates a network of subjects where the number of subjects on each site was sampled from an exponential distribution, and panel (c) illustrates a network where the number of subjects on each site was sampled from a uniform distribution. In the bottom-left the corner of each panel, we plot the ISI after performing djICA for 5 different runs, where each run resampled the number of subjects on each site from the given distribution.

due to network latency or slow processing. While it is likely that issues such as hardware variance will not significantly influence the analysis in the decentralized case, other practical considerations should be handled by the overall software framework in which djICA would be included. The djICA algorithm can easily be made more robust to by including features such as timeouts, automated resets in response to errors or dropout, thresholds for minimum sufficient participation from each site, and so on.

Additionally, a number of decentralization-friendly heuristic choices can be made to improve runtime or performance beyond that of the default settings in djICA. For example, a stochastic gradient for weight updates can be computed over blocks (or mini-batches) of data in order to improve runtime. Thus, the block size b can be chosen as a heuristic or evaluated as a hyper-parameter in order to examine the tradeoff between algorithm runtime and performance. Other hyper-parameters worth investigating are the tolerance level t, initial learning rate ρ, maximum iterations J, and the number of components chosen for local PCA.

In a pooled environment with a known ground-truth, it makes sense to find optimal values for these hyper-parameters using a grid search, or other hyper parameter selection method. In many real-life collaborative environments, however, a thorough hyper parameter search across sites may be impractical, and as far as we have found, no established method exists for hyper parameter optimization across decentralized sites. Finally, real-data problems often lack a reliable ground-truth, which makes it even more difficult and time-consuming to verify the effectiveness of multiple hyper-parameters. Nonetheless, in situations where a reliable ground-truth *is* available, such as in realistic simulations, one simple solution would be to aggregate locally searched hyper-parameters; however, this method is unlikely to yield good performance if the number of subjects varies widely between sites, or if many of the sites contain only a small amount of data. Another potential solution would be to have each site participate in a global search using a randomly sampled subset of the local data. This may prove effective provided that enough data can be made available from each site, but would come at the expense of additional computation, and additional release of information from each site. In ICA for fMRI, certain auxiliary measures, such as the cross-correlation between components or the kurtosis of estimated independent components, could be used to assess performance empirically, starting with an initial heuristic choice of parameters and making adjustments if the auxiliary measures (or other indirect validation surrogates) indicate it would be helpful to do so.

Due to the lack of sufficient data problem that our method solves, temporal ICA networks from resting state neuroimaging data are rarely reported in the literature. A straightforward comparison of our observed networks with typical ones is not possible. However, our maps should be comparable, to some extent, to temporal fluctuation modes (TFMs) reported in Smith et al. [21], which performed temporal ICA on denoised spatial ICA component time courses. A qualitative comparison of the observed ground-truth maps in our work to the TFM maps reported in their work suggests similarities in certain spatial map activation patterns between the two. Component 15 resembles TFM 8 with task positive regions (dorsal visual regions and frontal eye fields) anti-correlated to the default mode (posterior cingulate, angular gyri, and medial prefrontal cortex). Component 8, with anti-correlated foveal and high-eccentricity visual areas corresponding to surround suppression observed in task studies, shows a good resemblance to TFM 4 in that work. As observed in TFM 2, component 6 shows coactivation patterns of lateral visual areas and parts of thalamus. Component 17 from our work, shows a good correspondence to TFM 13 in that work, with DMN regions anti-correlated with bilateral supra-

marginal gyri and language regions, albeit without strong lateralization reported in that work. TFM 1 and component 14 in this work, demonstrate anti-correlated somatosensory regions to DMN regions of the brain. A couple other TFMs, 12 and 15, show moderate correspondence to components 11 and 9, respectively.

The differences between the networks we observed and the TFMs reported in Smith et al. [21] may stem from methodological differences and choice of number of independent components. In that work, instead of performing direct temporal ICA on preprocessed data to identify fluctuation modes, Smith et al. [21] use a two-step approach: firstly, performing a high model order spatial ICA, identifying artifactual components, and regressing out their variance from the time courses of seemingly non-artifactual components, and secondly, performing a temporal ICA on these denoised time courses. In contrast, we perform direct temporal ICA, leveraging the large number of samples available in large collaborative studies and directly getting to dynamics of fMRI. Therefore, the amount of variance captured during the PCA step in both methods differs. We identify 19 non-artifactual spatial modes, out of our 50 estimated components; all with spatial map activation patterns localized to gray matter regions and corresponding power spectra of independent time courses showing higher low frequency amplitude, as observed for intrinsic connectivity networks from spatial ICA analyses. These maps are included in Figure 6. Finally, the data utilized in that work was from 36 ten minute-runs from 5 subjects, roughly sampled at TR=0.8s, which yielded 24000 concatenated timepoints, in contrast to roughly 300000 concatenated timepoints from 2000 subjects in this study, which is arguably a more general result.

6. Conclusions & Future Work

We have presented djICA, a novel method for decentralized temporal Independent Component Analysis, which represents a step toward facilitating large, collaborative analyses of data in a decentralized fashion. We evaluated djICA on simulated and real fMRI data, with both experiments illustrating the benefits of djICA, namely the increased availability of a larger, otherwise inaccessible, subject pool shared across multiple sites. Additionally, since djICA does not communicate subject data across sites but only gradients, it is amenable to privatization via approaches like differential privacy [14], thus further opening the potential for collaboration between sites where direct sharing of data is not possible. Indeed, the increased availability of data provided by decentralized methods like djICA enables data-intensive, and thus underutilized, analyses like temporal Independent Component Analysis. Our comparison to the results from Smith et al. [21] confirms that djICA produces comparable temporal components. Finally, djICA and other methods like it foster further research on previously unexplored temporal dynamics in fMRI, such as the effects on temporal ICA of common confounds often found in datasets consisting of multi-site data.

Additional extensions to the methods provided here include reducing the bandwidth of the method and designing privacy-preserving variants, possibly, with differential privacy guarantees, which we have previously investigated for simulated cases [15]. In such cases, reducing the iteration complexity will help guarantee more privacy and hence incentivize larger research collaborations. If we were to return to the simulated data case, additional explorations into robust hyper-parameter searches and the deliberate corruption by noise may prove interesting for discovering and ameliorating further edge-cases for djICA. Beyond temporal ICA, decentralized spatial ICA is also worth investigation, and could be paired with decentralized clustering to evaluate decentralized dynamic functional network connectivity. Finally, Infomax ICA represents only one optimization approach to perform ICA, and while it is amenable to decentralization, other algorithms for ICA, such as fastICA [41] or the flexible entropy bound minimization (EBM) [70] approach, may provide other benefits beyond ease of decentralization.

7. Bibliography

[1] R. A. Poldrack, D. M. Barch, J. P. Mitchell, T. D. Wager, A. D. Wagner, J. T. Devlin, C. Cumba, O. Koyejo, M. P. Milham, Toward open sharing of task-based fMRI data: the OpenfMRI project, Frontiers in Neuroinformatics 7. doi:10.3389/fninf.2013.00012.

[2] K. Gorgolewski, O. Esteban, G. Schaefer, B. Wandell, R. Poldrack, Openneuro—a free online platform for sharing and analysis of neuroimaging data, in: Organization for Human Brain Mapping (OHBM), Vancouver, Canada, 2017, p. 1677.

[3] R. A. Poldrack, K. J. Gorgolewski, Making big data open: data sharing in neuroimaging, Nature Neuroscience 17 (11) (2014) 1510–1517. doi:10.1038/nn.3818.

[4] P. M. Thompson, J. L. Stein, S. E. Medland, D. P. Hibar, A. A. Vasquez, M. E. Renteria, R. Toro, N. Jahanshad, G. Schumann, B. Franke, et al., The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data, Brain Imaging and Behavior 8 (2) (2014) 153–182. doi:10.1007/s11682-013-9269-5.

[5] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L Whitwell, C. Ward, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, Journal of Magnetic Resonance Imaging 27 (4) (2008) 685–691. doi:10.1002/jmri.21049.

[6] P. M. Thompson, O. A. Andreassen, A. Arias-Vasquez, C. E. Bearden, P. S. Boedhoe, R. M. Brouwer, R. L. Buckner, J. K. Buitelaar, K. B. Bulayeva, D. M. Cannon, et al., ENIGMA and the individual: predicting factors that affect the brain in 35 countries worldwide, NeuroImage 145 (2017) 389–408. doi:10.1016/j.neuroimage.2015.11.057.

[7] T. G. van Erp, D. P. Hibar, J. M. Rasmussen, D. C. Glahn, G. D. Pearlson, O. A. Andreassen, I. Agartz, L. T. Westlye, U. K. Haukvik, A. M. Dale, et al., Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium, Molecular Psychiatry 21 (4) (2016) 547. doi:10.1038/mp.2015.63.

[8] D. P. Hibar, J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivières, N. Jahanshad, R. Toro, K. Wittfeld, L. Abramovic, M. Andersson, et al., Common genetic variants influence human subcortical brain structures, Nature 520 (7546) (2015) 224. doi:10.1038/nature14101.

[9] R. Mcdonald, M. Mohri, N. Silberman, D. Walker, G. S. Mann, Efficient large-scale distributed training of conditional maximum entropy models, in: Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, A. Culotta (Eds.), Advances in Neural Information Processing Systems 22 (NIPS 2009), Curran Associates, Inc., 2009, pp. 1231–1239.

[10] M. Zinkevich, M. Weimer, L. Li, A. J. Smola, Parallelized stochastic gradient descent, in: J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, A. Culotta (Eds.), Advances in Neural Information Processing Systems 23 (NIPS 2010), Curran Associates, Inc., 2010, pp. 2595–2603.

[11] V. D. Calhoun, T. Adalı, Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnos-

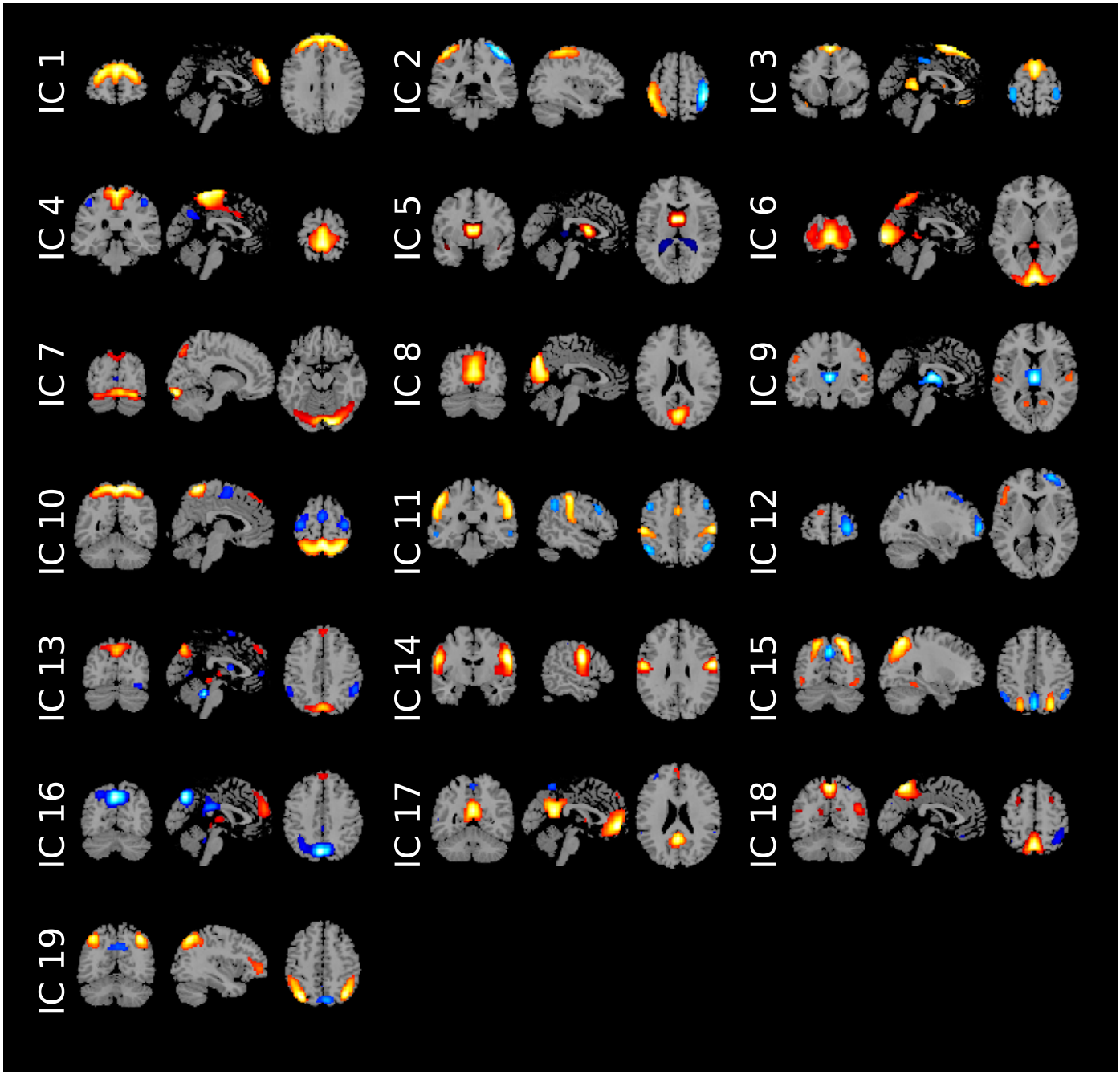

Figure 6: The 19 identified non-artifactual spatial modes, with spatial map activation patterns localized to gray matter regions. Component 15 resembles TFM 8 from Smith et al.. [21], with task positive regions (dorsal visual regions and frontal eye fields) anti-correlated to the default mode (posterior cingulate, angular gyri, and medial prefrontal cortex). Component 8 shows anti-correlated foveal and high-eccentricity visual areas corresponding to surround suppression observed in task studies, and resembles to TFM 4 in Smith et al.. Component 6 shows coactivation patterns of lateral visual areas and parts of thalamus. Component 17 shows a good correspondence to TFM 13 in Smith et al.. with DMN regions anti-correlated with bilateral supramarginal gyri and language regions, but without strong lateralization reported in that Smith et al.. Component 14 demonstrates anti-correlated somatosensory regions to DMN regions of the brain. A couple other TFMs from Smith. et al.. 12 and 15, show moderate correspondence to components 11 and 9 from our estimation.

tic discovery, IEEE Reviews in Biomedical Engineering 5 (2012) 60–73. doi:10.1109/RBME.2012.2211076.

[12] S. Plis, A. D. Sarwate, D. Wood, C. Dieringer, D. Landis, C. Reed, S. R. Panta, J. A. Turner, J. M. Shoemaker, K. W. Carter, P. Thompson, K. Hutchison, V. D. Calhoun, COINSTAC: A privacy enabled model and prototype for leveraging and processing decentralized brain imaging data, Frontiers in Neuroscience 10 (365). doi:10.3389/fnins.2016.00365.

[13] B. T. Baker, R. F. Silva, V. D. Calhoun, A. D. Sarwate, S. M. Plis, Large scale collaboration with autonomy: Decentralized data ICA, in: Proceedings of the IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, 2015. doi:10.1109/MLSP.2015.7324344.

[14] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in: S. Halevi, T. Rabin (Eds.), Theory of Cryptography, Vol. 3876 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 265–284. doi:10.1007/11681878_14.

[15] H. Imtiaz, R. Silva, B. Baker, S. M. Plis, A. D. Sarwate, V. D. Calhoun, Privacy-preserving source separation for distributed data using independent component analysis, in: Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 2016, pp. 123–127. doi:10.1109/CISS.2016.7460488.

[16] B. B. Biswal, J. L. Ulmer, Blind source separation of multiple signal sources of fMRI data sets using independent component analysis, Journal of Computer Assisted Tomography 23 (2) (1999) 265–271.

[17] J. Stone, J. Porrill, C. Buchel, K. Friston, Spatial, temporal, and spatiotemporal independent component analysis of fMRI data, in: Spatio-temporal Modelling and its applications, Citeseer, Department of Statistics, University of Leeds, 1999, pp. 7–9.

[18] V. Calhoun, T. Adalı, G. Pearlson, J. Pekar, Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms, Human Brain Mapping 13 (1) (2001) 43–53. doi:10.1002/hbm.1024.

[19] X. Gao, T. Zhang, J. Xiong, Comparison between spatial and temporal independent component analysis for blind source separation in fmri data, in: Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on, Vol. 2, IEEE, 2011, pp. 690–692.

[20] M. J. McKeown, L. K. Hansen, T. J. Sejnowsk, Independent component analysis of functional mri: what is signal and what is noise?, Current opinion in neurobiology 13 (5) (2003) 620–629.

[21] S. M. Smith, K. L. Miller, S. Moeller, J. Xu, E. J. Auerbach, M. W. Woolrich, C. F. Beckmann, M. Jenkinson, J. Andersson, M. F. Glasser, et al., Temporally-independent functional modes of spontaneous brain activity, Proceedings of the National Academy of Sciences 109 (8) (2012) 3131–3136. doi:10.1073/pnas.1121329109.

[22] K. J. Friston, Modes or models: a critique on independent component analysis for fmri, Trends in cognitive sciences 2 (10) (1998) 373–375.

[23] S. Dodel, J. M. Herrmann, T. Geisel, Comparison of temporal and spatial ica in fmri data analysis, in: Proceedings of ICA2000, the Second International Conference on Independent Component Analysis and Signal Separation, 2000, pp. 543–547.

[24] K. Petersen, L. K. Hansen, T. Kolenda, On the independent components of functional neuroimages.

[25] V. D. Calhoun, T. Adali, V. McGinty, J. J. Pekar, T. Watson, G. Pearlson, fmri activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis, NeuroImage 14 (5) (2001) 1080–1088.

[26] R. N. Boubela, K. Kalcher, W. Huf, C. Kronnerwetter, P. Filzmoser, E. Moser, Beyond noise: using temporal ICA to extract meaningful information from high-frequency fMRI signal fluctuations during rest, Frontiers in Human Neuroscience 7 (2013) 168. doi:10.3389/fnhum.2013.00168.

[27] M. F. Glasser, T. S. Coalson, J. D. Bijsterbosch, S. J. Harrison, M. P. Harms, A. Anticevic, D. C. Van Essen, S. M. Smith, Using temporal ica to selectively remove global noise while preserving global signal in functional mri data, bioRxiv (2017) 193862.

[28] E. B. Beall, M. J. Lowe, Isolating physiologic noise sources with independently determined spatial measures, Neuroimage 37 (4) (2007) 1286–1300.

[29] V. D. Calhoun, T. Adalı, G. D. Pearlson, J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, Human Brain Mapping 14 (3) (2001) 140–151. doi:10.1002/hbm.1048.

[30] N. Correa, T. Adalı, V. D. Calhoun, Performance of blind source separation algorithms for fMRI analysis using a group ICA method, Magnetic Resonance Imaging 25 (5) (2007) 684–694. doi:10.1016/j.mri.2006.10.017.

[31] V. D. Calhoun, J. Liu, T. Adalı, A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data, NeuroImage 45 (1) (2009) S163–S172. doi:10.1016/j.neuroimage.2008.10.057.

[32] E. Seifritz, F. Esposito, F. Hennel, H. Mustovic, J. G. Neuhoff, D. Bilecen, G. Tedeschi, K. Scheffler, F. Di Salle, Spatiotemporal pattern of neural processing in the human auditory cortex, Science 297 (5587) (2002) 1706–1708. doi:10.1126/science.1074355.

[33] V. van de Ven, F. Esposito, I. K. Christoffels, Neural network of speech monitoring overlaps with overt speech production and comprehension networks: a sequential spatial and temporal ica study, Neuroimage 47 (4) (2009) 1982–1991.

[34] M. Svensén, F. Kruggel, H. Benali, ICA of fMRI group study data, NeuroImage 16 (3) (2002) 551–563. doi:10.1006/nimg.2002.1122.

[35] V. D. Calhoun, T. Adalı, Unmixing fMRI with independent component analysis, IEEE Engineering in Medicine and Biology Magazine 25 (2) (2006) 79–90. doi:10.1109/MEMB.2006.1607672.

[36] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J. Cohen, J. Haxby, Independent component analysis for brain fMRI does not select for independence, Proceedings of the National Academy of Sciences 106 (26) (2009) 10415–10422. doi:10.1073/pnas.0903525106.

[37] V. Calhoun, V. Potluru, R. Phlypo, R. Silva, B. Pearlmutter, A. Caprihan, S. Plis, T. Adalı, Independent component analysis for brain fMRI does indeed select for maximal independence, PloS One 8 (8) (2013) e73309. doi:10.1371/journal.pone.0073309.

[38] Z. Boukouvalas, Y. Levin-Schwartz, V. D. Calhoun, T. Adalı, Sparsity and independence: Balancing two objectives in optimization for source separation with application to fmri analysis, Journal of the Franklin Institute 355 (4) (2017) 1873–1887. doi:10.1016/j.jfranklin.2017.07.003.

[39] A. J. Bell, T. J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Computation 7 (6) (1995) 1129–1159. doi:10.1162/neco.1995.7.6.1129.

[40] A. Hyvärinen, A family of fixed-point algorithms for independent component analysis, in: Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, 1997, pp. 3917–3920. doi:10.1109/ICASSP.1997.604766.

[41] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634. doi:10.1109/72.761722.

[42] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4–5) (2000) 411–430.

[43] R. Silva, S. Plis, J. Sui, M. Pattichis, T. Adalı, V. Calhoun, Blind source separation for unimodal and multimodal brain networks:a unifying framework for subspace modeling, EEE Journal of Selected Topics in Signal Processing 10 (7) (2016) 1134–1149. doi:10.1109/JSTSP.2016.2594945.

[44] J. Sui, T. Adalı, G. Pearlson, V. Calhoun, An ICA-based method for the identification of optimal fMRI features and components using combined group-discriminative techniques, NeuroImage 46 (1) (2009) 73–86. doi:10.1016/j.neuroimage.2009.01.026.

[45] J. Liu, V. Calhoun, Parallel independent component analysis for multimodal analysis: Application to fMRI and EEG data, in: Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2007), 2007, pp. 1028–1031. doi:10.1109/ISBI.2007.357030.

[46] E. Allen, E. Erhardt, E. Damaraju, W. Gruner, J. Segall, R. Silva, M. Havlicek, S. Rachakonda, J. Fries, R. Kalyanam, A. Michael, A. Caprihan, J. Turner, T. Eichele, S. Adelsheim, A. Bryan, J. Bustillo, V. Clark, S. Feldstein Ewing, F. Filbey, C. Ford, K. Hutchison, R. Jung, K. Kiehl, P. Kodituwakku, Y. Komesu, A. Mayer, G. Pearlson, J. Phillips, J. Sadek, M. Stevens, U. Teuscher, R. Thoma, V. Calhoun, A baseline for the multivariate comparison of resting state networks, Frontiers in Systems Neuroscience 5 (2). doi:10.3389/fnsys.2011.00002.

[47] V. Calhoun, T. Adalı, N. Giuliani, J. Pekar, K. Kiehl, G. Pearlson, Method for multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data, Human Brain Mapping 27 (1) (2006) 47–62.

doi:10.1002/hbm.20166.

[48] T. Eichele, S. Rachakonda, B. Brakedal, R. Eikeland, V. D. Calhoun, EEGIFT: A toolbox for group temporal ICA event-related EEG, Computational Intelligence and Neuroscience 2011 (2011) Article ID 129365. doi:10.1155/2011/129365.

[49] S. Rachakonda, R. F. Silva, J. Liu, V. D. Calhoun, Memory efficient PCA methods for large group ICA, Frontiers in Neuroscience 10 (2016) 17. doi:10.3389/fnins.2016.00017.

[50] Z.-J. Bai, R. Chan, F. Luk, Principal component analysis for distributed data sets with updating, in: APPT 2005: Advanced Parallel Processing Technologies, Vol. 3756 of Lecture Notes in Computer Science, Springer, 2005, pp. 471–483. doi:10.1007/11573937_51.

[51] S. i. Amari, A. Cichocki, H. H. Yang, A new learning algorithm for blind signal separation, Advances in NIPS (1996) 757–763.

[52] S.-I. Amari, T.-P. Chen, A. Cichocki, Stability analysis of learning algorithms for blind source separation, Neural Networks 10 (1997) 1345–1351. doi:10.1016/S0893-6080(97)00039-7.

[53] H. Imtiaz, A. D. Sarwate, Differentially private distributed principal component analysis, in: Proceedings of the 43rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018), Calgary, Canada, 2018.

[54] V. D. Calhoun, R. F. Silva, T. Adalı, S. Rachakonda, Comparison of PCA approaches for very large group ICA, NeuroImage 118 (2015) 662–666. doi:10.1016/j.neuroimage.2015.05.047.

[55] MATLAB, rand:Uniformly distributed random numbers, mathworks. `https://www.mathworks.com/help/matlab/ref/rand.html`

[56] E. Erhardt, E. Allen, Y. Wei, T. Eichele, V. Calhoun, SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability, NeuroImage 59 (4) (2012) 4160–4167. doi:10.1016/j.neuroimage.2011.11.088.

[57] R. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50 (4) (1982) 987–1007. doi:10.2307/1912773.

[58] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (3) (1986) 307–327. doi:10.1016/0304-4076(86)90063-1.

[59] K. Zhang, A. Hyvärinen, Source separation and higher-order causal analysis of MEG and EEG, in: 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010), AUAI Press, Catalina Island, California, 2012, pp. 709–716.

[60] T. Ozaki, Time Series Modeling of Neuroscience Data, 1st Edition, CRC Press, Boca Raton, FL, USA, 2012.

[61] Q. Luo, G. Tian, F. Grabenhorst, J. Feng, E. Rolls, Attention-dependent modulation of cortical taste circuits revealed by Granger causality with signal-dependent noise, PLoS Comp. Bio. 9 (10) (2013) e1003265. doi:10.1371/journal.pcbi.1003265.

[62] M. Lindquist, Y. Xu, M. Nebel, B. Caffo, Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach, NeuroImage 101 (2014) 531 – 546. doi:10.1016/j.neuroimage.2014.06.052.

[63] R. Balan, Estimator for number of sources using minimum description length criterion for blind sparse source mixtures, in: D. M. E., J. C. J., A. S. A., P. M. D. (Eds.), ICA 2007: Independent Component Analysis and Signal Separation, Vol. 4666 of Lecture Notes in Computer Science, Springer, 2007, pp. 333–340. doi:10.1007/978-3-540-74494-8_42.

[64] E. Egolf, K. A. Kiehl, V. D. Calhoun, Group ICA of fMRI toolbox (GIFT), in: Proceedings of Human Brain Mapping, Budapest, Hungary, 2004.

[65] M. R. N. Medical Image Analysis Lab, Group ICA Of fMRI Toolbox (GIFT) [cited February 11, 2018]. `http://mialab.mrn.org/software/gift`

[66] H. W. Kuhn, The Hungarian method for the assignment problem, in: M. Jünger, et al. (Eds.), 50 Years of Integer Programming 1958-2008, Springer, Berlin, Heidelberg, 2010, pp. 29–47. doi:10.1007/978-3-540-68279-0_2.

[67] N. P. Wojtalewicz, R. F. Silva, V. D. Calhoun, A. D. Sarwate, S. M. Plis, Decentralized independent vector analysis, in: Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, New Orleans, LA, USA, 2017, pp. 826–830. doi:10.1109/ICASSP.2017.7952271.

[68] N. Lewis, S. Plis, V. Calhoun, Cooperative learning: Decentralized data neural network, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, Anchorage, AK, USA, 2017, pp. 324–331. doi:10.1109/IJCNN.2017.7965872.

[69] B. Hitaj, G. Ateniese, F. Perez-Cruz, Deep models under the gan: information leakage from collaborative deep learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2017, pp. 603–618.

[70] X.-L. Li, T. Adali, Complex independent component analysis by entropy bound minimization, IEEE Transactions on Circuits and Systems I: Regular Papers 57 (7) (2010) 1417–1430. doi:10.1109/TCSI.2010.2046207.