

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Zixi Yang

Date

The impact of initialization in optimization of independent components in functional
magnetic resonance imaging

By

Zixi Yang
Master of Science in Public Health

Department of Biostatistics and Bioinformatics

Benjamin Risk
Thesis Advisor

Ying Guo
Thesis Reader

The impact of initialization in optimization of independent components in functional magnetic resonance imaging

By

Zixi Yang

Bachelor of Science
Central South University
2014

Thesis Committee Chair: Benjamin Risk, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Biostatistics and Bioinformatics
2019

Abstract

The impact of initialization in optimization of independent components in functional magnetic resonance imaging

By Zixi Yang

MELODIC software is widely used to conduct independent component analysis (ICA) of functional magnetic resonance imaging (fMRI), in which the components correspond to resting-state networks and image artifacts. However, the objective function in ICA is non-convex and typically contains local optima. Consequently, multiple initial values improve the ability to find the global optima. The current version of MELODIC software does not allow an evaluation of the impact of multiple initializations. The goal of this paper is to examine the impacts of initialization in MELODIC. To clarify the effects, we applied MELODIC with multiple seeds to two datasets: the tutorial data set available with MELODIC software and a subject from the Autism Brain Imaging Data Exchange. We examined the variability between components estimated from 100 seeds. In both datasets, there were a number of components that exhibited high variability between seeds, especially with components classified as unknowns. These components tended to have lower kurtosis, which may be related to estimating too many components. Some components that were sensitive to initialization contained spatial features indicative of signal or artifacts. We conclude that independent components estimated with MELODIC are sensitive to initialization.

The impact of initialization in optimization of independent components in functional magnetic resonance imaging

By

Zixi Yang

Bachelor of Science
Central South University
2014

Thesis Committee Chair: Benjamin Risk, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Biostatistics and Bioinformatics
2019

Contents

1 Introduction	1
2 Method	2
2.1 Data set description	2
2.1.1 Tutorial Dataset	2
2.1.2 ABIDE Data	2
2.2 ICA Method	3
2.3 Optimization	4
2.4 Visualization	4
3 Results	5
3.1 Multiple initializations for the tutorial data set	5
3.1.1 Local optima with 100 seeds	5
3.1.2 Correlation matrix for differences with 100 seeds	6
3.1.3 Differences among 159 components	6
3.1.4 Orthographic views of signals and artifacts	7
3.1.5 Time courses and spectral density of signals and artifacts	8
3.1.6 Correlation matrix for signals, artifacts and unknowns	10
3.2 Multiple initializations for the ABIDE data set	11
3.2.1 Local optima with 100 seeds	11
3.2.2 Correlation Matrix for differences with 100 seeds	12
3.2.3 Differences among 80 components	12
3.2.4 Orthographic views of signals and artifacts	14
3.2.5 Time courses and spectral density of signals and artifacts	15
3.2.6 Correlation matrix for signals, artifacts and unknowns	16
4 Discussion	17

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) measures the Blood Oxygenation Level Dependent (BOLD) signal in the brain to examine patterns in neural activity. Resting state fMRI (rsfMRI) is used for characterizing spontaneous correlations in BOLD activity when an awake subject is performing no activity. Alterations in resting state correlations have been associated with mental and neurological disorders. There are various sources of artifacts that will interfere with the diagnosis of the effects of true BOLD signals and may lead to misinterpretation of fMRI [Griffanti et al., 2015]. Independent Components Analysis (ICA) is the most widely used technique for identifying different kinds of artifacts in fMRI data. ICA can be used to decompose fMRI data into “signal” components and structure “noise,” such as motion effects, cardiac pulsation, and other non-neuronal physiology [Salimi-Khorshidi et al., 2014]. Cleaned fMRI data without structured “noise” components may be beneficial to diagnoses of brain activity of patients. FMRIB’s ICA-based X-noiseifier (ICA-FIX) is used in the Human Connectome Project processing pipeline, where it may increase statistical sensitivity [Smith et al., 2013].

Even though ICA has been widely applied in analyzing brain imaging data, it requires non-convex optimization, which can be sensitive to initialization [Risk et al., 2014, Himberg and Hyvarinen, 2003]. The corresponding spurious components caused by local optima might cause inaccurate estimations of artifacts and resting-state networks. The probability of finding the global optimum will be increased by running the algorithm several times with different initializations. Re-estimating components with several initializations may improve the ability to identify “signal” components (including resting-state networks) versus “noise” components (artifacts) [Tohka et al., 2008].

MELODIC (Multivariate Exploratory Linear Optimized Decomposition of Independent Components) is a popular artifact detection software in fMRI and used in the ICA-FIX artifact-removal pipeline. However, it estimates independent components (ICs) from a single initialization. In particular, running MELODIC twice on the same dataset may produce different results. Consequently, studies using MELODIC could result in mischaracterized signal and artifacts. Another popular ICA software is Group

ICA of fMRI Toolbox (GIFT), which enables group inferences [Calhoun et al., 2001]. While GIFT can be done with multiple initializations, it is designed for group analysis [Rachakonda et al., 2007]. MELODIC applies kurtosis-based ICA to the 4D pre-processed data set. Components can then be classified as signals and artifacts for the purpose of removing artifacts from subsequent modeling of brain activity, and this approach is part of ICA-FIX [Griffanti et al., 2014]. As ICA-FIX is based on single-session MELODIC output, we focus on its implementation [Salimi-Khorshidi et al., 2014].

In order to illustrate the influence of multiple initializations, we apply MELODIC with 100 initializations on the multiband 6 tutorial data set provided in the ICA-FIX FSL practical. We also examine resting-state data from a subject from the Autism Brain Imaging Data Exchange (ABIDE) of International Neuroimaging Data-Sharing Initiative (INDI). We calculate the objective functions and match the ordered ICs through the modified Hungarian algorithm [Risk et al., 2014]. We compare the mixing matrices based on multidimensional scaling [Torgerson, 1952]. We then compared the similarity between components estimated with multiple initial values.

2 Method

2.1 Data set description

2.1.1 Tutorial Dataset

The multi-band accelerated resting-state fMRI data set (MB6) used in this analysis is provided by the ICA tutorial for using MELODIC software ([ICA link]). This dataset uses a multiband factor of 6. The data dimensions are $106 \times 106 \times 66 \times 790$. MELODIC aims to decompose 4D data sets into spatial and temporal components by using ICA. We prepared the original tutorial data sets as 4D NIfTI data and converted them to multiple 3D images. By default, MELODIC commonly normalizes the variance of time courses for brain image data sets and estimates the number of components of the data automatically. The number of components was estimated by MELODIC as 159. We inputted the original tutorial 4D NIfTI data and estimated ICs by MELODIC with initializing 100 different seeds for further analysis.

2.1.2 ABIDE Data

We also analyzed resting-state fMRI data from a subject from the Autism Brain Imaging Data Exchange (ABIDE). ABIDE combines functional and structural brain imaging collected from researchers around the world focusing on the brain connectome in Autism Spectrum Disorder (ASD). Resting State fMRI was acquired during 5 min 20 s-long scan on a 3.0 T Philips scanner using a single-shot, partially parallel,

gradient-recalled echo planar sequence with sensitivity encoding (repetition time [TR]/echo time [TE] = 2500/30ms, flip angle = 75 degrees, sensitivity encoding acceleration factor of 2, 3-mm axial slices with no slice gap, in-plane resolution of 3.053.15 mm [8481 voxels]). Participants were instructed to relax, fixate on a cross-hair and remain as still as possible. rs-fMRI scans were slice-time adjusted using the slice acquired in the middle of the TR as a reference, and rigid body realignment parameters were estimated to adjust for motion. We used subject 29298. The dimensions of the data are $96 \times 96 \times 47 \times 128$. Note the fMRI data is single-band, which generally results in a lower estimate of the number of components than multi-band.

2.2 ICA Method

Suppose the observed data \mathbf{Y}_v is a linear combination of non-Gaussian independent components. Let $\mathbf{Y}_v \in \mathbb{R}^Q$ be a random vector such that $\mathbb{E}\mathbf{Y}_v = 0$ and $\mathbb{E}\mathbf{Y}_v\mathbf{Y}_v' = \mathbf{I}$, where \mathbf{I} is the $Q \times Q$ identity matrix. (In practice, this assumption is met using pre-whitening.) Let

$$\mathbf{Y}_v = \mathbf{S}_v\mathbf{A}. \quad (1)$$

where $\mathbf{S}_v \in \mathbb{R}^Q$ is a random vector with independent components with $\mathbb{E}\mathbf{S}_v = 0$ and $\mathbb{E}\mathbf{S}_v\mathbf{S}_v' = \mathbf{I}$. \mathbf{A} is a $Q \times Q$ mixing matrix of full rank. Since the data are pre-whitened, \mathbf{A} is restricted to the class of orthogonal matrices. Consequently, the components are orthogonal, which is a necessary condition for independence. Let $\mathbf{W} = \mathbf{A}^{-1} = \mathbf{A}'$. The goal of ICA is to find the un-mixing matrix \mathbf{W} to maximize non-gaussianity of \mathbf{S} .

$$\hat{\mathbf{S}}_v = \mathbf{W}\mathbf{Y}_v \quad (2)$$

The matrix decomposition of ICA is identifiable up to signed changes and permutations, as described below. Let \mathbf{P} be a $Q \times Q$ permutation matrix, and we can equivalently represent \mathbf{Y} as

$$\mathbf{Y}_v = \mathbf{S}_v\mathbf{P}\mathbf{P}^{-1}\mathbf{A}. \quad (3)$$

Additionally, we can apply a sign-change matrix \mathbf{D} , which is a diagonal matrix with 1 or -1 on the diagonals. Then we can decompose \mathbf{Y} as follows:

$$\mathbf{Y}_v = \mathbf{S}_v\mathbf{P}\mathbf{D}\mathbf{D}^{-1}\mathbf{P}^{-1}\mathbf{A}. \quad (4)$$

Since each method is non-convex, multiple initializations can lead to different estimations.

2.3 Optimization

We use kurtosis to find the rotation \mathbf{O} that maximizes non-Gaussianity, as it is the default method in MELODIC. For whitened data, kurtosis is equal to $\frac{1}{V} \sum_{v=1}^V (\mathbf{O}'_q Y_v)^4$. Then the estimator is

$$\hat{\mathbf{W}} = \operatorname{argmax}_{\mathbf{O} \in \mathcal{O}} \sum_{v=1}^V \sum_{q=1}^Q (o'_q Y_v)^4. \quad (5)$$

Since \mathcal{O} is a non-convex set, this is a non-convex optimization problem. Consequently, there may be local optima. For each data set, we estimated independent components from 100 initializations. Then we identified the seed with the max objective function value among all 100 seeds, and treated the associated ICs as the template for matching other ICs later. For simplicity, we call the seed identified with the max objective function value the arg max, since it represents our best estimate of the arg max.

2.4 Visualization

First, we used `matchICA()` from the R package `steadyICA` (Risk et al., 2015) to match each IC with the arg max ICs. Second, we calculated their Frobenius distance with $\mathcal{D}(\mathbf{S}_i, \mathbf{S}_j) = \sum_{i=1}^{100} \sum_{j=1}^{100} (\mathbf{S}_i - \mathbf{S}_j)^2$, $i = 1, \dots, 100, j = 1, \dots, 100$, where \mathbf{S}_i denotes the ICs estimated from the i^{th} initialization. Then classical multidimensional scaling (MDS) was used on the Frobenius distance to visualize the dissimilarities of the un-mixing matrix of matched ICs, which can facilitate the identification of clusters corresponding to local optima. Third, we calculated the correlation of components between matched ICs of 100 seeds and formed a $100 \times 100 \times Q$ data array (recall Q is the number of components). We then obtained the median, 0.025 quantile, and 0.975 quantile of those correlations across components, leading to a 100×100 matrix. Then for selected components, we plotted their correlations, which corresponds to a 100×100 matrix from the $100 \times 100 \times Q$ array. We also plotted the kurtosis of the arg max components versus the 0.025, 0.5, 0.975 quantile correlations between the arg max component estimate and the other seeds, which provides insight into whether the issues with convexity are more problematic for components that are close to Gaussianity. If too many components are estimated, some components may be close to Gaussian and not biologically meaningful, in which case the local optima are of lesser concern. Finally, we calculated the component time courses. Define $\hat{\mathbf{M}} = \mathbf{S}' \times \mathbf{X}$, where \mathbf{X} is the centered masked data. Then each row of $\hat{\mathbf{M}}$ corresponds to a component time course. We generated time series plots and spectral density plots of selected components for the argmax and two other initial values. This allowed

a visualization of the impact of local optima on estimated signals and artifacts. We also classified the components as signals, artifacts, or unknowns based on a visual inspection of the orthographic views of the components. (Note a more detailed classification would also use the component time courses and spectral densities, but that was beyond the scope of the current work.) We used software wb-view to visualize the independent components.

3 Results

3.1 Multiple initializations for the tutorial data set

3.1.1 Local optima with 100 seeds

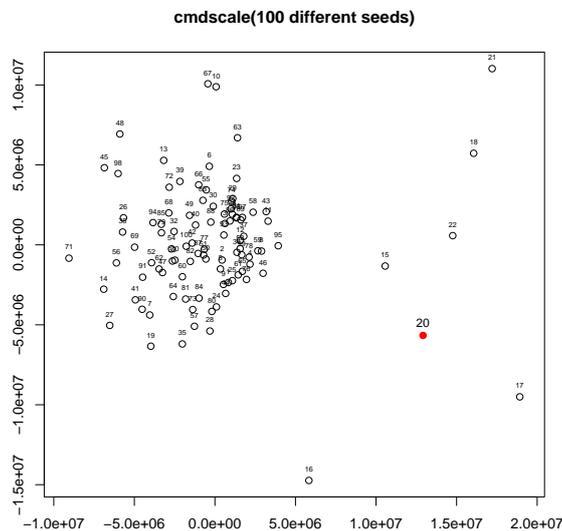


Figure 1: Multidimensional scaling with two axes using the Frobenius distance between matched independent components from 100 initial values from single-subject ICA of the MB6 FSL tutorial data. The estimate with maximum kurtosis is depicted by a red circle (seed 20), and estimates from other seeds were sign-changed and permuted to best match the seed 20.

From the multiple initializations, we found that the IC estimates when the seed equaled 20 had the largest objective function value compared to the estimates from the other 99 seeds. From the 2 dimensional plot produced with multidimensional scaling, there is one big cluster containing most of the estimates. Interestingly, the arg max did not cluster with any other estimates.

3.1.2 Correlation matrix for differences with 100 seeds

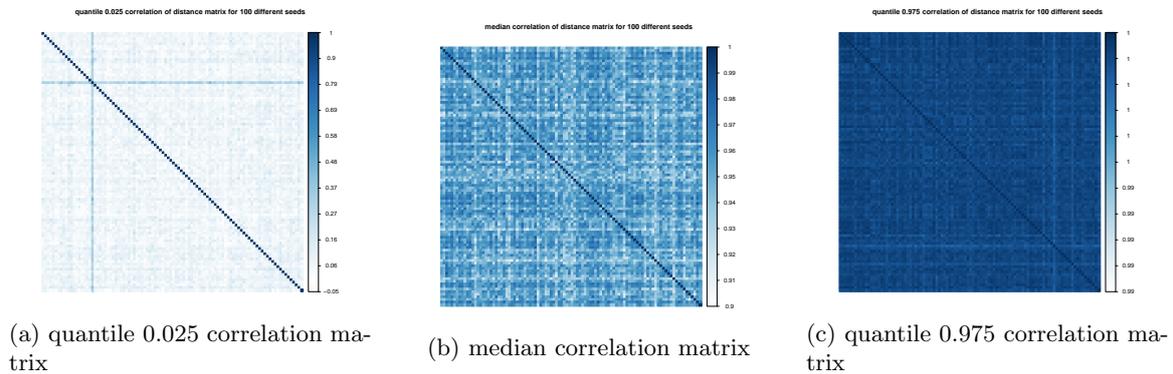


Figure 2: Correlation Matrix for the tutorial dataset.

From the plots in Figure 2, darker blue corresponds to higher correlations, which indicates greater consistency across initializations. The second plot is for the median correlation and the third plot is for 0.975 quantile correlation. Both of these two plots have dark blue in general, which means most of the components across initializations are highly consistent. We observed most of the correlations as being above 0.9 in plot b of Figure 2, and the range of plot c of Figure 2 is from 0.99 to 1. However, in the 0.025 quantile, the correlations are relatively low. The range of correlations for this plot is from -0.05 to 1. This indicates that some seeds produce estimates substantially different from the arg max; in other words, some of the components are not estimated consistently. From Figure 2, we know there are at least four components that are estimated poorly, since the 0.025 quantile of 159 is four.

3.1.3 Differences among 159 components

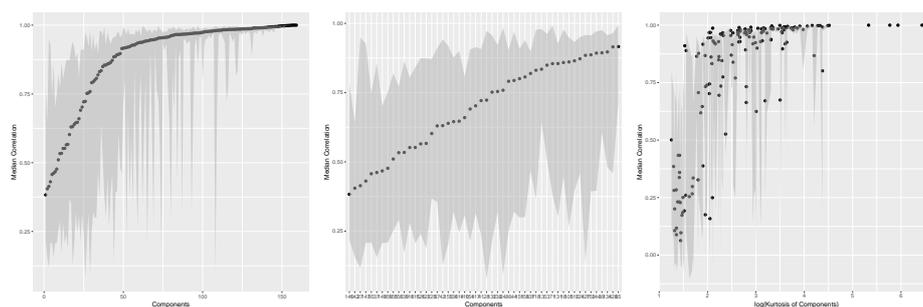


Figure 3: Median correlation for components with the arg max seed along with quantile 0.975 and 0.025 bands.

The left figure is for the overall view, and the middle figure is an enlarged version for components with relatively low correlations. The right figure depicts the correlation between components and the arg max versus the log kurtosis for arg max components.

We made a 100×159 correlation matrix among 159 components between the arg max and all the other 99 seeds. Then we calculated quantile 0.025, median and quantile 0.975 correlations of those 99

different seeds for each component. After sorting those correlations by median correlation, we plotted the median correlation for all 159 components first, along with quantile 0.975 correlation and quantile 0.025 correlation as upper and lower bands in the plot, which is the left figure above. Then we selected the components with comparably low median correlations (less than 0.9), which resulted in 48 components. Figure 3 b plots the median, 0.025, and 0.975 quantile correlations for these components. The proportion of components with low correlation in this tutorial data set is 0.302. In Figure 3 c (right plot), we plotted the log kurtosis for the 159 components as x axis, and the 0.025, 0.5, and 0.975 quantiles of the correlation between matched components and the arg max as the y axis. Components with high kurtosis tended to have a higher median correlation with the arg max, indicating fewer issues with local optima. However, some components with high kurtosis exhibited variation across seeds, as represented by low 0.025 quantiles.

3.1.4 Orthographic views of signals and artifacts

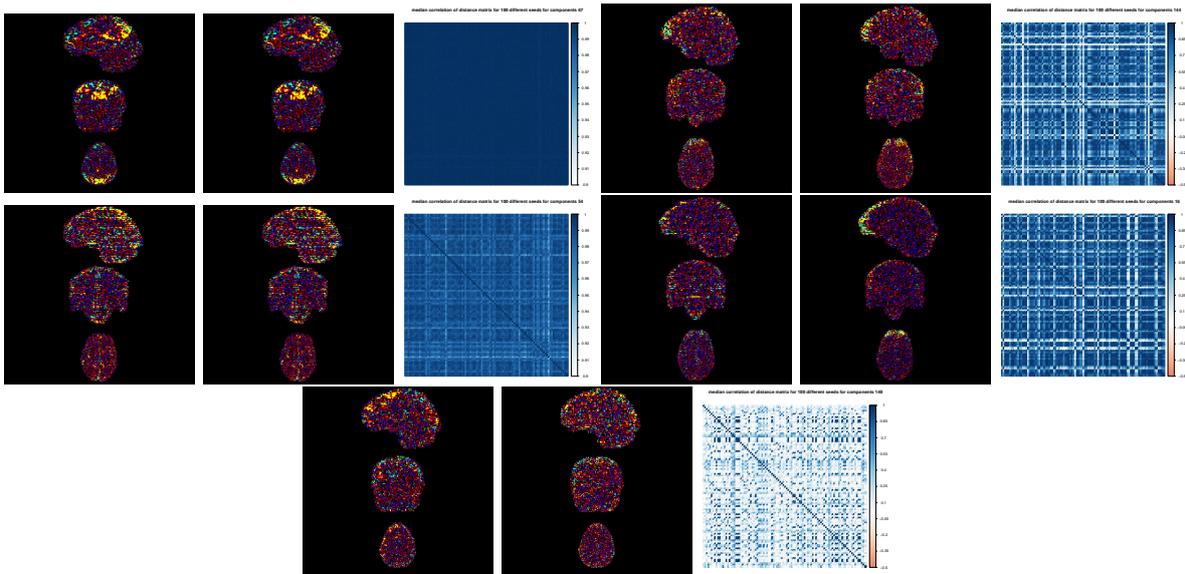


Figure 4: Orthographic views of subject's components and related correlation matrix. The first row depicts components that appear to be signals (component 67, 144), the second row depicts components that appear to be artifacts (component 54, 16), and the third row is the component with lowest median correlation with the argmax (component 140).

Based on Figure 1, the estimates from seed 17 and the arg max are different. We used the software `wb-view` to create orthographic views of the brain for certain components with chosen seeds. To the right of those plots, we plotted correlation matrices across all seeds for the corresponding component to illustrate overall variability in estimates of that component. From Figure 4 above, for the first row, we chose two components that appear to be related to neural activity (signal components). Component 67 (top left) did not vary across initializations, represented by high correlations. It appears to contain portions

of the default mode network, and thus captures neural activity. In contrast, component 144 differed across initializations (top right). Note the spectral density of this component's time course indicates activity in the 0.01-0.08 Hz band, which suggests neural activity; see below. The second row depicts examples of components that appear to correspond to artifacts, components 54 and 16. Component 54 has clear multi-band artifacts in both initializations. In contrast, component 16 appears to have different regions of activation, with large positive and negative values in the frontal lobe. The correlation matrix for component 54 is high compared to component 16, indicating the estimation of component 54 is consistent across initializations, while component 16 is less consistent. Finally for the last row, the component 140 was chosen based on its lowest correlation among all the components. The left column of orthographic views is for estimated ICs with seed 17, and the right column of views is for ICs with arg max seed 20. Overall, the correlation matrix of component 140 has lowest correlation. The orthographic views of seed 17 and 20 for component 140 have significantly different signals. In this component, the seed associated with the arg max appears to have less structure than other seeds, which is contrary to our expectations.

3.1.5 Time courses and spectral density of signals and artifacts

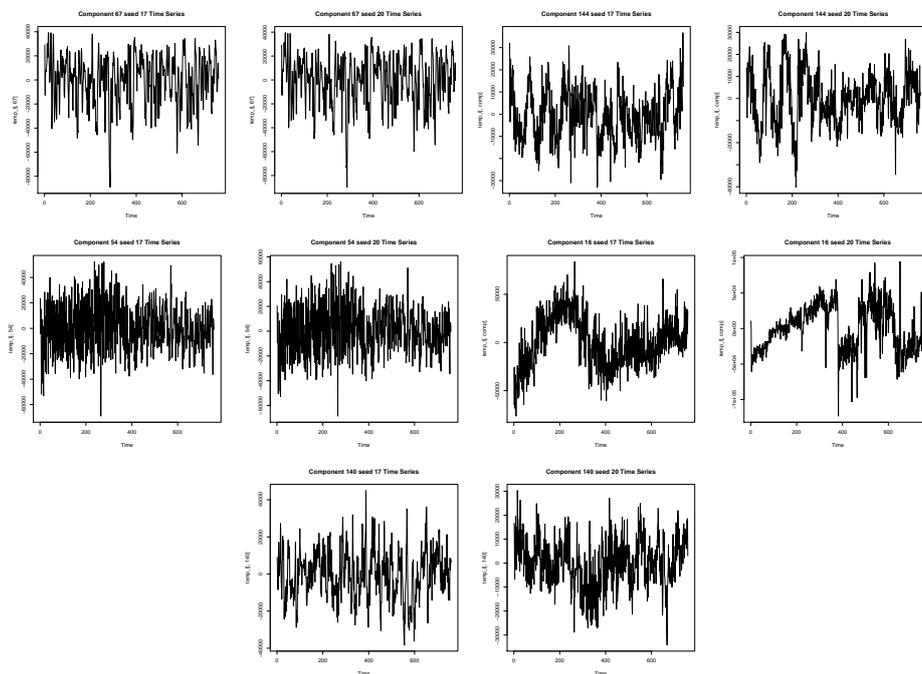


Figure 5: Time Courses. The first row is for signal components 67 and 144, the second row is for artifact components 54, 16, and the third row is for component 140 with lowest correlation. For the first and second row, the first column is for seed 17, the second column is for the arg max seed 20, the third column is for seed 17, the fourth column is for seed 20. For the third row, the first column is for seed 17 and second column is for seed 20.

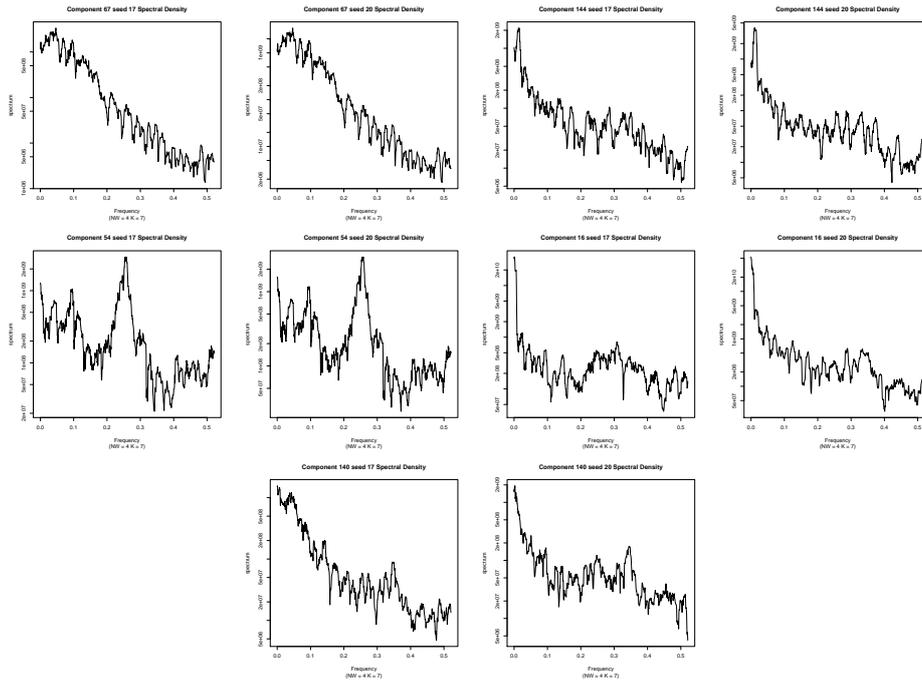


Figure 6: Spectral Densities. The first row is for components 67 and 144 (classified as signals), the second row is for components 54 and 16 (classified as artifacts), and the third row is for component 140 with lowest correlation. For the first and second row, the first column is for seed 17, the second column is for the arg max seed 20, the third column is for seed 17, the fourth column is for seed 20. For the third row, the first column is for seed 17 and second column is for seed 20.

We generated time series plots and spectral density plots for seeds 17 and 20, which provides further insight into the classification of the components as signal versus artifact, and provides insight into the variability between components. In Figure 5, the time series for components 144, 16 and 140 differ greatly between the estimates from seed 17 and 20. These time courses had low correlations in the corresponding components (Figure 4). Note that the time course in component 16 contains jumps for the arg max (seed 20) but not for component 16, suggesting that the arg max estimate may better capture a motion-related artifact. In Figure 6, component 67 and 144 have greater power at frequencies between 0.01 - 0.10 Hz, suggesting signals related to neural activity. The component 54 in the second row has large power around 0.25 Hz, which suggests an artifact. The component 16 has high power near 0 Hz but not greater than 0.02 Hz, suggesting an artifact. The third row for component 140 had high power in the neural-related frequencies in seed 17 but to a lesser extent in seed 20, and its classification is uncertain.

3.1.6 Correlation matrix for signals, artifacts and unknowns

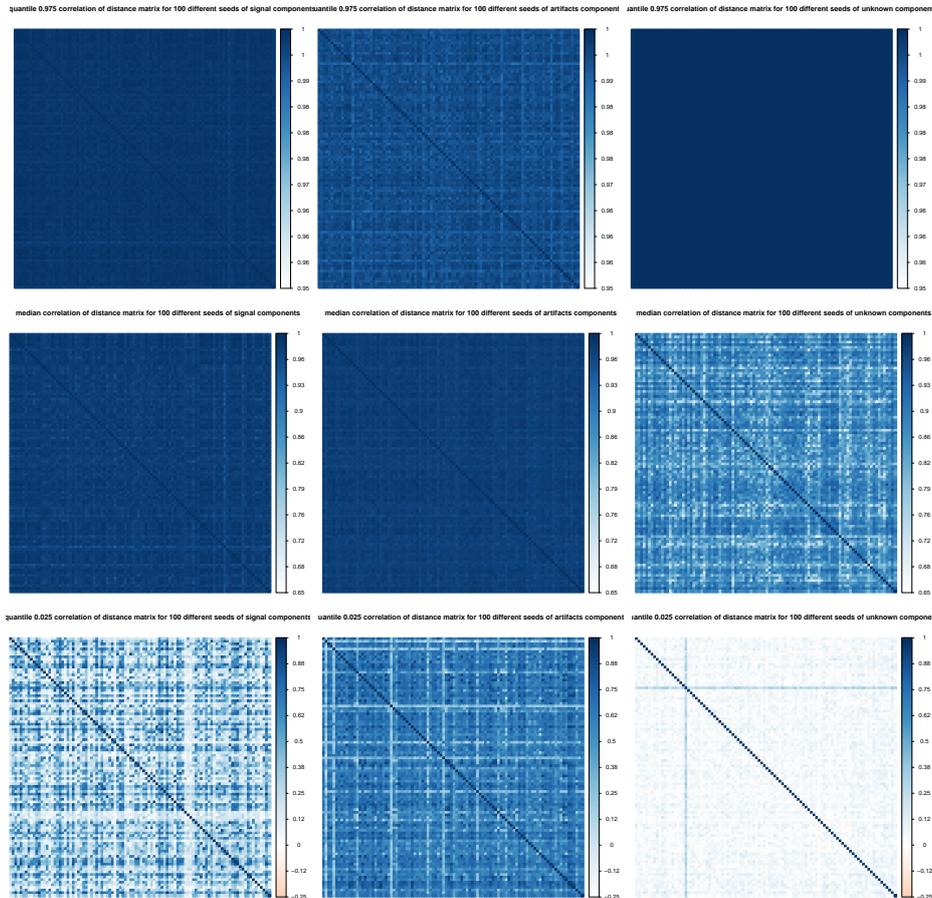


Figure 7: correlation matrix for signals, artifacts and unknowns. First row is for quantile 0.975 correlations, second row is for median correlations, the third row is for quantile 0.025 correlations. First column is for signals, the second column is for artifacts, and the third column is for unknowns.

We examined the consistency across initializations when grouping components as signals, artifacts and unknowns. We calculated the 100 by 100 correlation matrix for the three groups of components. We separated components into signals, artifacts and unknowns and calculated their quantile 0.025, median, quantile 0.975 correlation matrix. From Figure 7, quantile 0.975 for signals, artifacts and unknowns are high. The median correlations for signals and artifacts are relatively high, while the median correlation for unknowns is more variable, with some values around 0.65. The signals components have lower 0.025 quantile correlations than artifacts components, and unknowns components have the lowest correlations in the third row. The reason might be related to the fact many of the components classified as unknown contain little spatial structure and appear to be closer to Gaussianity. The unknowns may reflect the estimation of too many components, where the number of components was estimated using MELODIC, such that these components are in fact Gaussian noise. Additionally, it may reflect issues that may arise when non-Gaussian components are closer to Gaussianity, which may result in more local optima. Next,

we examine differences in estimates from multiple initializations applied to a subject from the ABIDE data set.

3.2 Multiple initializations for the ABIDE data set

3.2.1 Local optima with 100 seeds

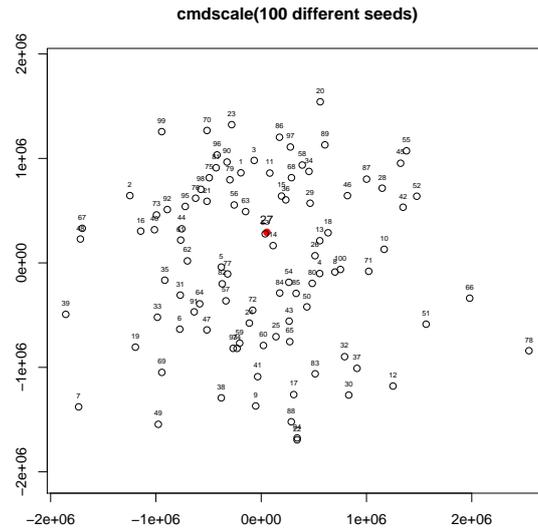


Figure 8: Multidimensional scaling with two dimensions using the Frobenius distance between matched independent components from 100 initial values from single-subject ICA of a subject from the ABIDE dataset. The estimate with maximum kurtosis is depicted by a red circle (seed 27), and estimates from other seeds were sign-changed and permuted to best match the components from seed 27.

We repeated the previous analysis for subject 29298 from the ABIDE data set. For optimizing the ABIDE data set with subject 29298, ICs estimated with seed 27 had the largest sum objective function compared to ICs estimated with the other 99 seeds. From the two dimensional representation with multidimensional scaling, it seems like there is one cluster with ICs with all 100 seeds. The arg max is located in the center of the initial values.

3.2.2 Correlation Matrix for differences with 100 seeds

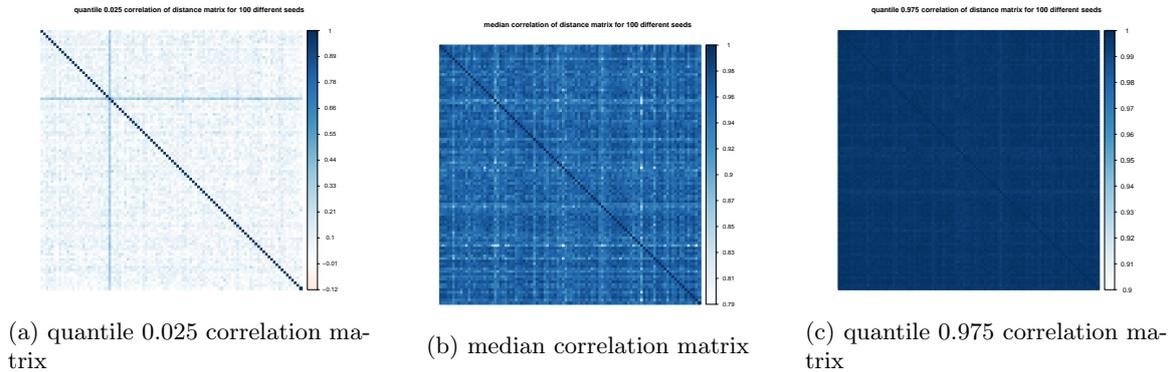


Figure 9: Correlation matrix for the ABIDE data set.

In Figure 9, the color shade of blue represents correlations of estimated ICs with 100 different seeds. Darker blue is related to higher correlations, which means greater consistency across initializations. The middle plot is for the median correlation, the right plot is 0.975 quantile correlation and the left plot is 0.025 quantile correlation. Clearly, 0.025 quantile correlation matrix has relatively lighter blue, and note the range is from -0.12 to 1. Most of the correlations in plot b and plot c of Figure 9 is above 0.9. Thus, most of the components were consistent across different initial values. However, the estimated components from certain seeds are different from the arg max, and some of them are not estimated consistently. From Figure 9, there are at least two components that differ between initializations, since the 0.025 quantile of 80 is two.

3.2.3 Differences among 80 components

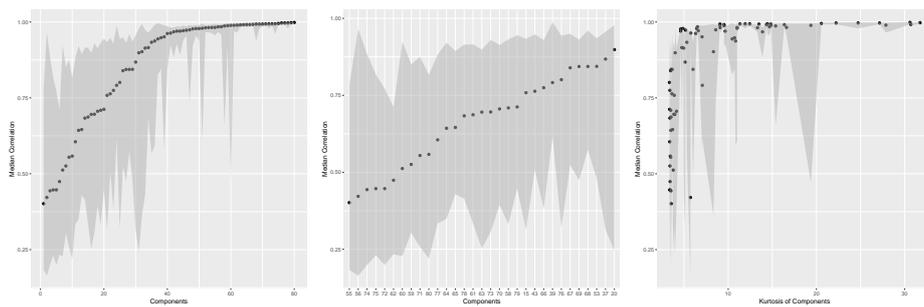


Figure 10: Median correlation for components with the arg max seed along with quantile 0.975 and 0.025 bands

The left figure is for the overall view, and the middle figure is an enlarged version for components with relatively low correlations. The right figure depicts the correlation between matched components and arg max versus kurtosis for arg max components.

We made a 100×80 correlation matrix among 80 components between the arg max seed 27 and all the other 99 different seeds for components of subject 29298 in the ABIDE data set. We calculated the sorted

quantile 0.025, median and quantile 0.975 correlations of those 100 different seeds for each component, and then filled the left plot with median correlation for all 80 components with the corresponding quantile 0.975 correlation and quantile 0.025 correlation as upper and lower bounds (Figure 10, left plot). We enlarged the part for components with low median correlations (less than 0.9) in the middle figure, whose number in total is 32. The proportion of components with low correlation in subject 29298 ABIDE data is $32/80$ which equals to 0.4. We also plotted the 0.025, 0.5, and 0.975 quantiles of the correlation between kurtosis of matched components and the best estimate of the arg max. From the right plot of Figure 10, components with higher kurtosis generally have higher median correlation with the arg max, which means fewer issues with local optima. But, as shown in the figure, there are low 0.025 quantiles for certain components with intermediate levels of kurtosis, indicating variation existed across seeds. The proportion of components with low correlations for the ABIDE data and tutorial data is similar to each other.

3.2.4 Orthographic views of signals and artifacts

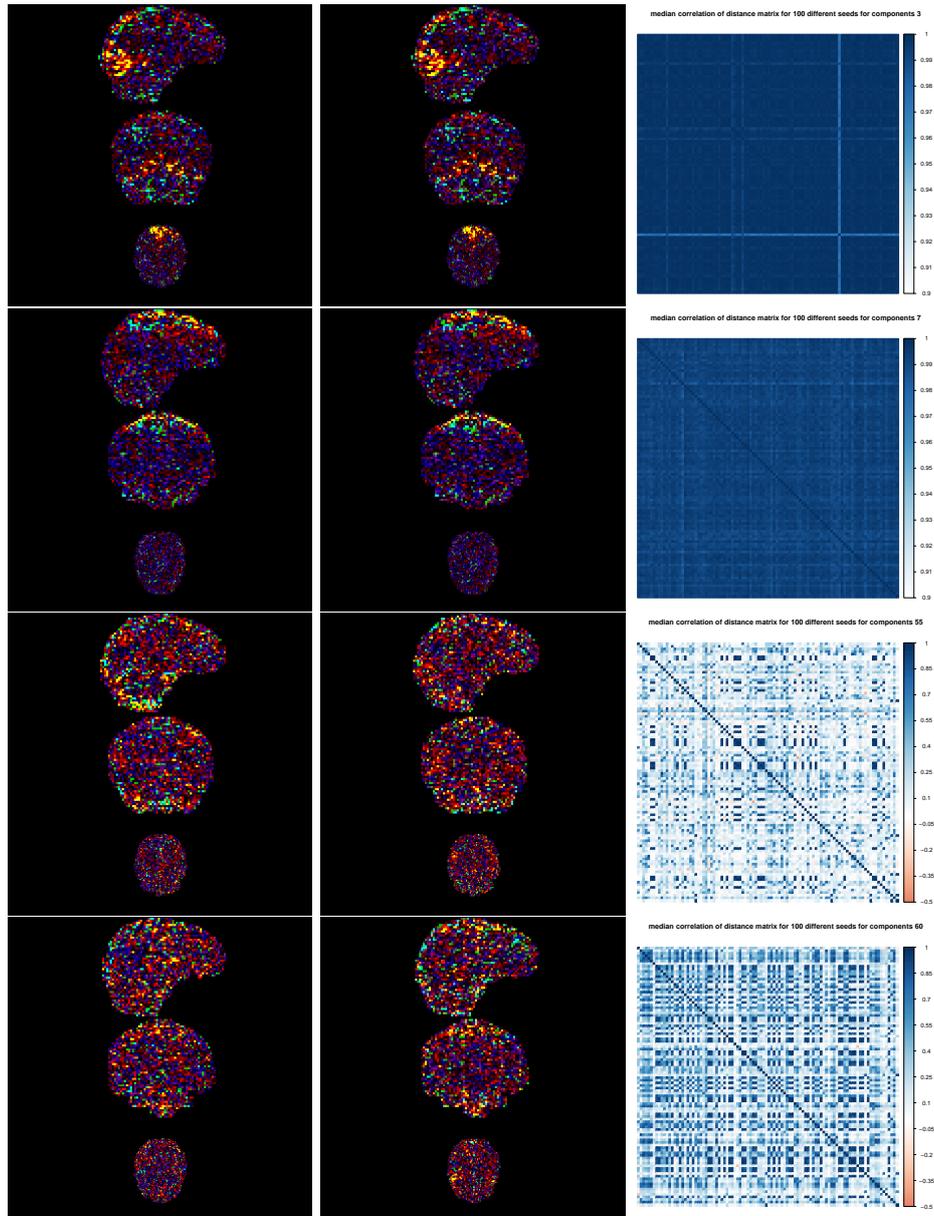


Figure 11: Orthographic views of subject's components and correlation matrix. First row is a component with signals (component 3), second row is a component with artifacts (component 7), the third row is the component with lowest correlation (component 55), the fourth row is a component with low correlation that shows some signs of artifacts (component 60).

First column is for seed 77, the second column is for arg max seed 27. The third column is the correlation matrix among 100 different seeds.

In Figure 11, for the first row, we chose component 3, which represents neural activity (signal components); for the second row, the component 7 is chosen as an example that may correspond to an artifact; for the third row, we chose the component with lowest correlation, 55; and for the fourth row, the component 60 may represent movement (artifact components) with relatively low correlation. The

left column of orthographic views is for estimation with seed 77, and the right column of the views is for the arg max seed 27. From Figure 11, orthographic views of components 3 and 7 for seed 77, 27 are very similar, and these components had high correlations with the arg max across initializations. In contrast, component 55 and 60 had low correlations, as shown in the correlation matrices (range from -0.5 to 1), and the orthographic views are different between initializations. The component 55 with the arg max (seed 27) has less structure than seed 77. For component 60, seed 77 has less structure than the arg max seed.

3.2.5 Time courses and spectral density of signals and artifacts

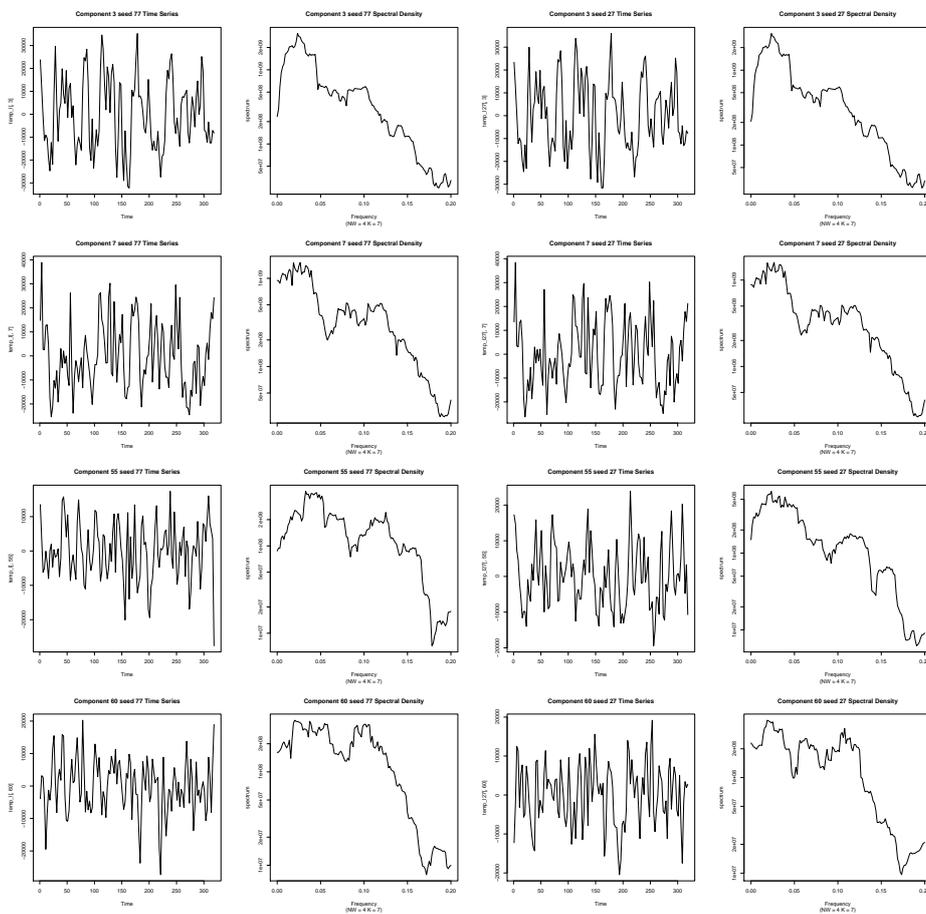


Figure 12: Time Courses along with Spectral Density. The first row is for signal component 3, the second row is for artifacts component 7, the third row is for component 55 with lowest correlation, the fourth row is for component 60 with low correlation. The first column is for seed 77, the second column is for seed 27 (the arg max).

We generated time series plots and spectral density plots for seed 77 and the arg max seed 27. The time series plot and spectral density plot for components 3 and 7, which had high correlations, had very similar patterns between seed 77 and 27, but trends for component 55 and 60 with seed 77 were different

than seed 27. Overall, these plots seem less informative for the ABIDE dataset than the tutorial MB 6 data, where the temporal resolution is lower in the ABIDE data versus the multiband data.

3.2.6 Correlation matrix for signals, artifacts and unknowns

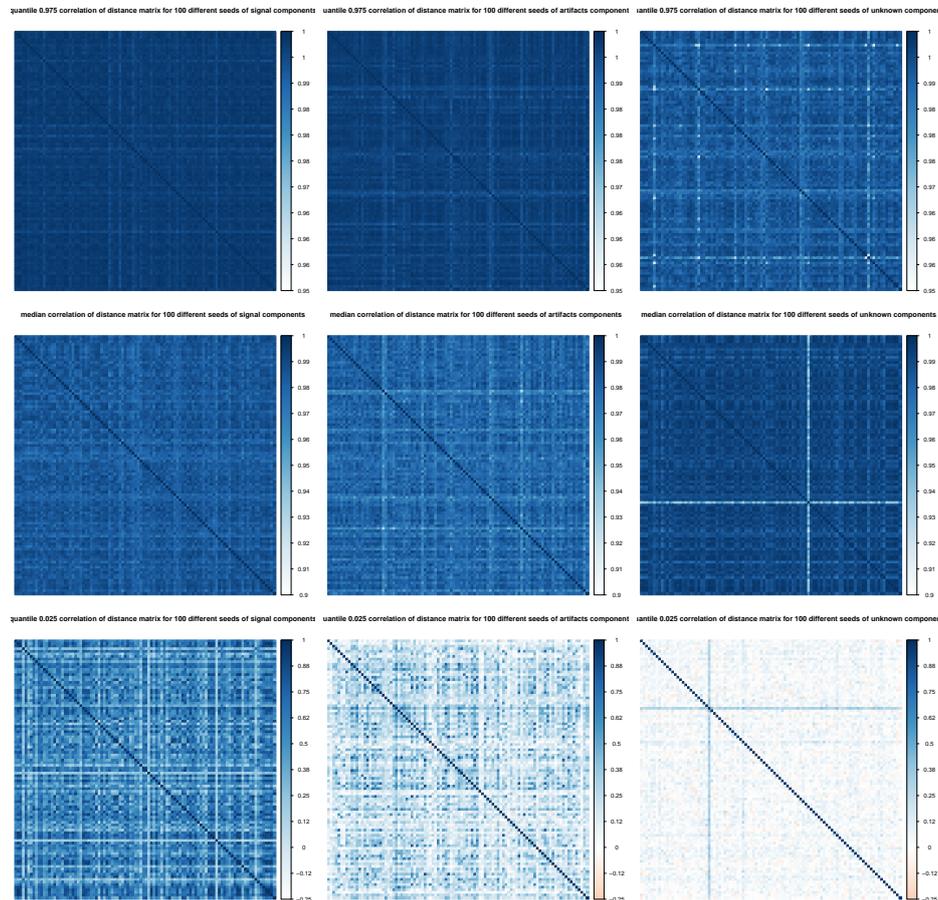


Figure 13: correlation matrix for signals, artifacts and unknowns First row is for the quantile 0.975 correlations, second row is for median correlations, the third row is for the quantile 0.025 correlations. First column is for signals, the second column is for artifacts, and the third column is for unknowns.

For ABIDE data subject 29298, Figure 13 shows evidence of differences between signals, artifacts, and unknowns. Similar to the tutorial data set, quantile 0.975 and median correlations are high, while quantile 0.025 correlation for signals, artifacts and unknowns are relatively low. Notably, the quantile 0.025 correlation matrix for unknowns and artifacts are low. In the tutorial data set, the signals have lower correlations compared to artifacts for the 0.025 quantile correlation matrix, but in the ABIDE data, the artifacts have lower correlations than signals for the 0.025 quantile correlation matrix. Note that the classification of components has some degree of subjectivity, which will influence these patterns.

4 Discussion

In this study, we found that independent components estimated using MELODIC can differ between initializations. Most components did not differ greatly between initializations, as evident by high correlations, but a notable number displayed substantial differences. This suggests that initialization may impact fMRI preprocessing pipelines that use ICA for artifact removal, such as ICA-FIX [Salimi-Khorshidi et al., 2014] [Griffanti et al., 2014], which was used in the Human Connectome Project [Van Essen et al., 2013]. Future research should examine whether the differences between components due to initialization can have significant impacts on the subsequent analyses, for example, whether functional connectivity analyses using ICA-FIX data are impacted by sensitivity to initialization.

From the results in part 3.1, we saw that ICs from the arg max seed did not cluster with the other 99 seeds. This may be unusual for a functional MRI data set. A previous study found more distinct clusters in group ICA [Risk et al., 2014], with multiple initializations more clearly converging on an argmax. From an examination of the components, there were differences among components estimated with different initializations. In our examination of the correlation between 100 different seeds, we found the median and quantile 0.975 have high correlation, while quantile 0.025 has relatively low correlation, which means there are some components which are poorly estimated. In the correlation plot describing median correlation across components with the arg max, the proportion of components with correlation less than 0.9 is around 0.3. Additional research should examine why there was a lack of distinct clusters. For example, the lack of clusters may be due to hundreds of local optima, or it may be related to the convergence criteria used in MELODIC. It may also be related to convergence and/or local optima issues when too many components are estimated (discussed below).

In the ABIDE dataset, the arg max seed 27 is in the center of the other estimates, but again clusters were not very clearly defined. As in the tutorial data set, the quantile 0.025 correlation matrix was low, which indicated that there were some components that differed across initializations. The proportion of components with median correlation less than 0.9 is 0.4, which is larger but still close to the proportion of the tutorial data set. Orthographic views, time series plots and spectral density plots for component 55 with lowest correlation differed between the arg max and other initializations, while for signal component 3 and artifacts component 4, they appear very similar. When we separate components into signals, artifacts and unknowns, the quantile 0.975 and median correlations were still high through out all three kinds of components. However, especially for unknowns, quantile 0.025 correlations tend to be relatively low. Due to the fact that the classification of signals, artifacts and unknowns is based on

visual inspection, there may be misclassifications that impact these findings. Overall, the multiple initial values did not have a significant influence on a majority of components, while the components with low correlations were affected obviously.

In this study, we have used the max among all initializations as the best estimate of the independent components. However, we found examples where the component corresponding to another initialization exhibited more meaningful spatial patterns than the arg max (component 16 in Figure 4, 5, 6, component 55 in Figure 11, 12). When the ICA model is true, the arg max is clearly the best choice. In real data applications, there may be other approaches that could be considered. ICASSO performs cluster analysis on the independent components pooled across multiple initializations, as performed in GIFT ICA tool box [\[Rachakonda et al., 2007\]](#). Additionally, we found that many components did not exhibit strong spatial features. Related, components with lower kurtosis tended to have more variability across initializations (component 16, 144, 140 in Figure 4, 5, 6, component 55, 60 in Figure 11, 12). MELODIC uses a probabilistic PCA to estimate the number of components [\[Minka, 2001\]](#). If too many components are estimated, then some of the components may be closer to Gaussian. When using multidimensional scaling to visualize how results cluster into local basin of attractions, these issues may lead to a lack of clustering. The extension of tests of the number of non-Gaussian components [\[Jin et al., 2019\]](#) is an important avenue for future research.

Future research should examine patterns of local optima in additional subjects. Here, we examined the impacts in a multiband dataset (MB 6 tutorial) and a single-band dataset (a subject from the ABIDE dataset), and future research should examine the generalizability of these findings to other studies.

References

- [Calhoun et al., 2001] Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). Group ica of functional mri data: separability, stationarity, and inference. In *Proc. Int. Conf. on ICA and BSS San Diego, CA. p*, volume 155.
- [Griffanti et al., 2015] Griffanti, L., Dipasquale, O., Laganà, M. M., Nemni, R., Clerici, M., Smith, S. M., Baselli, G., and Baglio, F. (2015). Effective artifact removal in resting state fmri data improves detection of dmn functional connectivity alteration in alzheimer’s disease. *Frontiers in human neuroscience*, 9:449.
- [Griffanti et al., 2014] Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., Zsoldos, E., Ebmeier, K. P., Filippini, N., Mackay, C. E., et al. (2014). Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *Neuroimage*, 95:232–247.
- [Himberg and Hyvarinen, 2003] Himberg, J. and Hyvarinen, A. (2003). Icasto: software for investigating the reliability of ica estimates by clustering and visualization. In *2003 IEEE XIII Workshop on Neural Networks for Signal Processing (IEEE Cat. No.03TH8718)*, pages 259–268.
- [Jin et al., 2019] Jin, Z., Risk, B. B., and Matteson, D. S. (2019). Optimization and testing in linear non-gaussian component analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal*.
- [Minka, 2001] Minka, T. P. (2001). Automatic choice of dimensionality for pca. In *Advances in neural information processing systems*, pages 598–604.
- [Rachakonda et al., 2007] Rachakonda, S., Egolf, E., Correa, N., and Calhoun, V. (2007). Group ica of fmri toolbox (gift) manual. *Dostupné z <http://www.nitrc.org/docman/view.php/55/295/v1>*. 3d-GIFTManual pdf [cit. 2011-11-5].
- [Risk et al., 2015] Risk, B., James, N., and Matteson, D. (2015). steadyica: Ica and tests of independence via multivariate distance covariance. *R package version*, 1.
- [Risk et al., 2014] 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(1):224–236.

- [Salimi-Khorshidi et al., 2014] Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, 90:449–468.
- [Smith et al., 2013] Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., et al. (2013). Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168.
- [Tohka et al., 2008] Tohka, J., Foerde, K., Aron, A. R., Tom, S. M., Toga, A. W., and Poldrack, R. A. (2008). Automatic independent component labeling for artifact removal in fmri. *Neuroimage*, 39(3):1227–1245.
- [Torgerson, 1952] Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419.
- [Van Essen et al., 2013] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79.