Novel Unsupervised Signal Separation Methods for Complex High-dimensional fMRI Data Decomposition

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

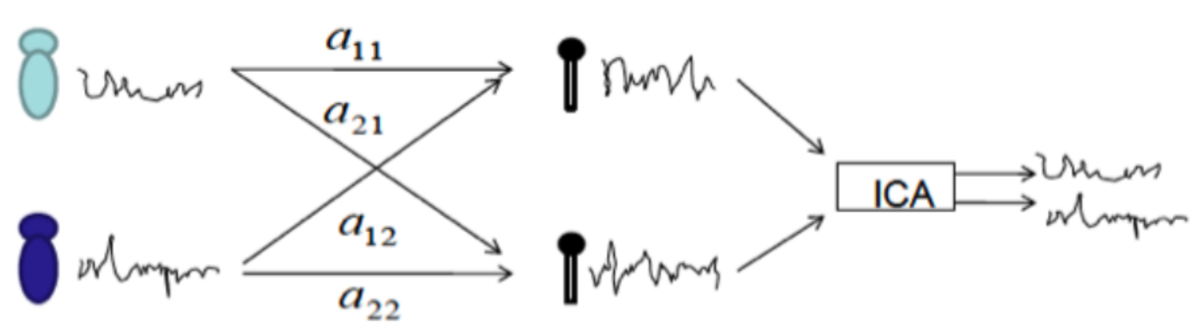
Yikai Wang, Ying Guo

Department of Biostatistics and Bioinformatics, Emory University

INTRODUCTION

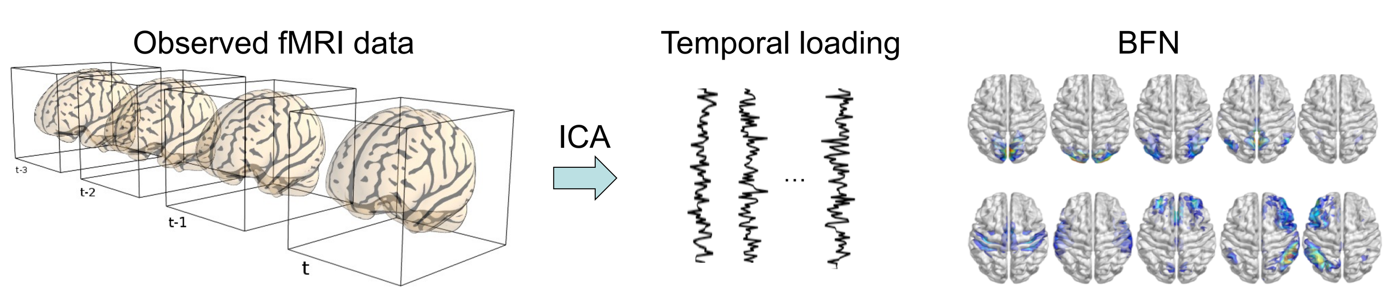
Blind signal separation (BSS), also known as **blind source separation**, is the separation of a set of source signals from a set of mixed signals, without the aid of information about the source signals or the mixing process.

Example: Cocktail party problem



Independent component analysis (ICA) is a signal processing method for separating a multivariate signal into additive subcomponents with the assumption that the subcomponents are non-Gaussian and are statistically independent.

Specifically, ICA is the most commonly used method in brain imaging field for identifying the latent brain functional networks (BFN) based on fMRI data:



$$\boldsymbol{Y}(v) = \boldsymbol{A}\,\boldsymbol{S}(v) + \boldsymbol{e}(v).$$

How to extend ICA to complex group-level study?

- Single ICA + IC Matching
- Temporal-concatenation ICA (TC-GICA)

Hierarchical ICA Framework

Assume we have K fMRI scans, i.e. $\boldsymbol{Y}_k(v)$, and covariates $\boldsymbol{x}_k$:

Level 1: $\boldsymbol{Y}_k(v) = \boldsymbol{A}_k\boldsymbol{S}_k(v) + \boldsymbol{e}_k(v)$,

Level 2: $\boldsymbol{S}_k(v) = \boldsymbol{s}_0(v) + \boldsymbol{b}_{g(k)}(v) + \boldsymbol{\beta}_{c(k)}(v)\boldsymbol{x}_k + \boldsymbol{\varepsilon}_k(v)$.

- $\boldsymbol{Y}_k(v)$: (T x 1) fMRI time series at location v for scan k;
- $\boldsymbol{A}_k$: (T x q) scan-specific temporal loading matrix;
- $\boldsymbol{S}_k(v)$: (q x 1) scan-specific latent component at location v ;
- $\boldsymbol{s}_0(v)$: (q x 1) population-level spatial map at location v ;
- $\boldsymbol{b}_{g(k)}(v)$: (q x1) group-level random effects at location v ;
- $\boldsymbol{\beta}_{c(k)}(v)$: (q x 1) covariate effects at location v .

Specifically, we set mixture of Gaussian prior on $\boldsymbol{s}_0(v)$. With latent state variable $\boldsymbol{z}(v)$, we have $\boldsymbol{s}_0(v) = \boldsymbol{\mu}_{\boldsymbol{z}(v)} + \boldsymbol{\psi}_{\boldsymbol{z}(v)}$, $\boldsymbol{\psi}_{\boldsymbol{z}(v)} \sim N\left(0, \boldsymbol{\Sigma}_{\boldsymbol{z}(v)}\right)$, where $\boldsymbol{z}(v)$ represents which Gaussian component in MoG that voxel v belongs to.

Hc-ICA (special case for cross-sectional study)

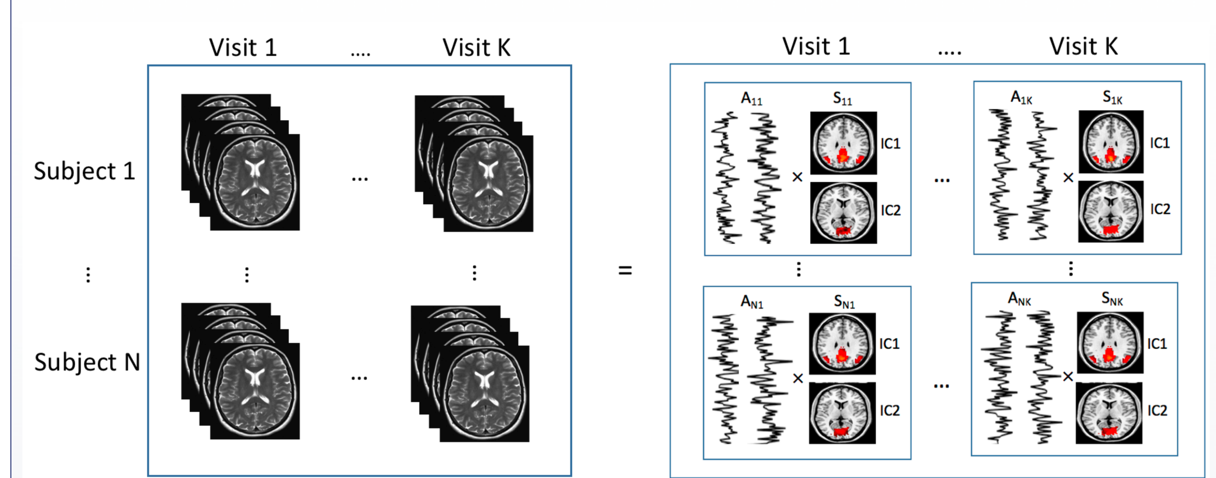
Level 1: $\boldsymbol{Y}_i(v) = \boldsymbol{A}_i\boldsymbol{S}_i(v) + \boldsymbol{e}_i(v)$,

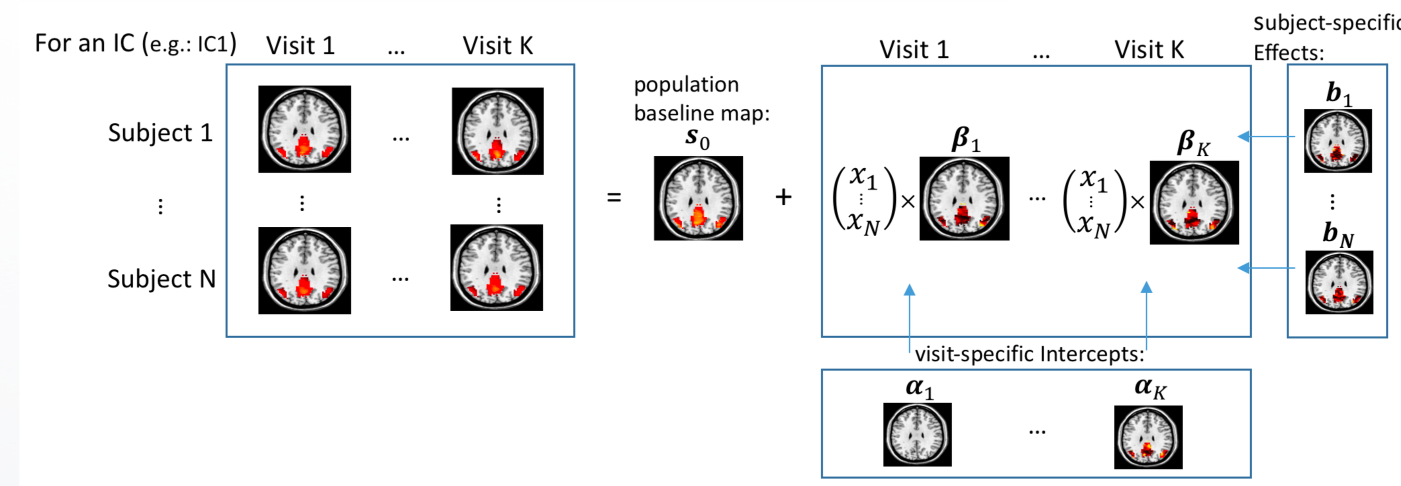
Level 2: $\boldsymbol{S}_i(v) = \boldsymbol{s}_0(v) + \boldsymbol{b}_i(v) + \boldsymbol{\beta}(v)\boldsymbol{x}_i$

Longitudinal ICA (special case for longitudinal study)

Level 1: $\boldsymbol{Y}_{ij}(v) = \boldsymbol{A}_{ij}\boldsymbol{S}_{ij}(v) + \boldsymbol{e}_i(v)$

Level 2: $\boldsymbol{S}_{ij}(v) = \boldsymbol{s}_0(v) + \boldsymbol{b}_i(v) + \boldsymbol{\alpha}_j(v) + \boldsymbol{\beta}_j(v)\boldsymbol{x}_j + \boldsymbol{\varepsilon}_{ij}(v)$.





EM algorithm

Denote $\boldsymbol{L}(v)$ to contain all latent variables except for $\boldsymbol{z}(v)$. Conditioned on $\boldsymbol{z}(v)$, we can estimate the conditional expectation of $\boldsymbol{L}(v)$:

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

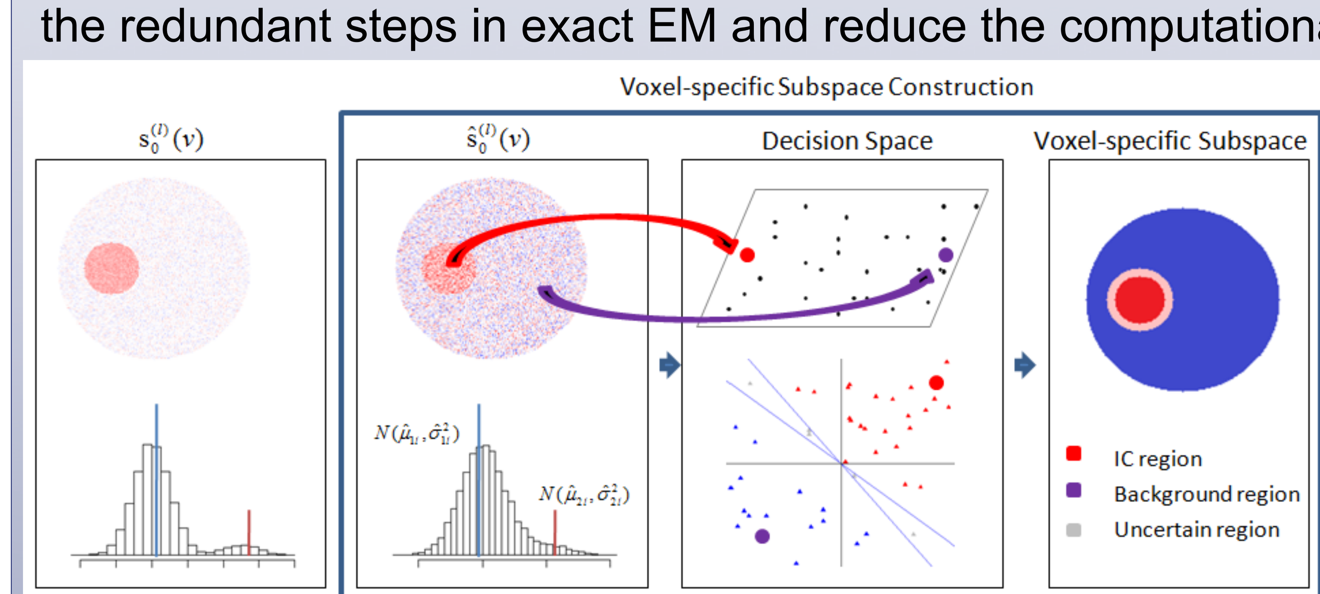
where $\boldsymbol{R}$ represents the set of all possible values of $\boldsymbol{z}(v)$, i.e. $\boldsymbol{R} = \{\boldsymbol{z}^r\}_1^{m^q}$.

Limitation:

This exact EM requires $O(m^q)$ for each voxel for learn the latent structure of $\boldsymbol{z}(v)$, which increases exponentially with the number of ICs.

Stochastic EM

We proposed a stochastic EM to adaptively learn the latent structure of $\boldsymbol{z}(v)$ driven by the data to eliminate the redundant steps in exact EM and reduce the computational complexity.



- Based on current estimation, i.e. $\hat{\boldsymbol{s}}_0(v)$, we classified all voxels for each IC into three classes: IC region, background region and uncertain region through a pre-specified decision rule $\mathcal{F}$.
- Finally we can construct a subspace $\mathcal{R}_v$ for each voxel based on $\mathcal{F}(\hat{\boldsymbol{s}}_0(v))$ by eliminating the original space $\mathcal{R}$.

For example, map each IC element into a m dimensional decision space by:

$$\mathcal{Z}\left(\hat{s}_0^{(l)}(v); \boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2\right) = \left((\hat{s}_0^{(l)}(v) - \mu_{l,1})/\sigma_{l,1}, .., (\hat{s}_0^{(l)}(v) - \mu_{l,m})/\sigma_{l,m}\right)',$$

The decision rule with 2 terms:

$$\mathcal{F}(z_1, z_2; \varepsilon) = \begin{cases} 1, \text{ if } z_1 \le 0, z_2 < 0 \text{ or } z_1 > 0, z_2 < 0, |z_1/z_2| \le 1-\varepsilon, \\ 2, \text{ if } z_1 > 0, z_2 \ge 0 \text{ or } z_1 > 0, z_2 < 0, |z_1/z_2| \ge 1+\varepsilon, \\ 0, \text{ otherwise }, \end{cases}$$

where 1: background, 2: IC region, 0: uncertainty region, $\varepsilon \in (0,1)$.

Approximate Inference

- Stack fMRI data across all visits to have the subject-specific fMRI data, and the time-stacking non-hierarchical L-ICA becomes:

$$\boldsymbol{A}_i'\boldsymbol{y}_i(v) = \boldsymbol{U}\boldsymbol{\mu}_{\boldsymbol{z}(v)} + \alpha(v) + \boldsymbol{X}_i\beta(v) + \boldsymbol{U}\psi_{\boldsymbol{z}(v)} + \boldsymbol{U}\boldsymbol{b}_i(v) + \gamma_i(v) + \boldsymbol{A}_i'\boldsymbol{e}_i(v),$$

- It can be further expressed as:

$$\boldsymbol{y}_i^*(v) = \boldsymbol{X}_i^*\boldsymbol{C}^*(v) + \zeta_i(v),$$

- Similar to Shi and Guo (2016), this model can be viewed as a multivariate linear model at each voxel, and we proposed a variance estimator as,

$$\mathrm{Var}\left\{\hat{\boldsymbol{C}}^*(v)\right\} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{*\prime}\boldsymbol{W}_i(v)^{-1}\boldsymbol{X}_i^*\right)^{-1}$$

- Based on the EM algorithm, the unknown parameters in $\boldsymbol{W}_i(v)$ can be estimated simultaneously,

$$\hat{\boldsymbol{W}}_i(v) = \boldsymbol{U}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}_{mode}(v)} + \hat{\boldsymbol{D}})\boldsymbol{U}' + (\hat{\sigma}^2 + \hat{\tau}^2\boldsymbol{A}_i\boldsymbol{A}_i')\boldsymbol{I}_{qK},$$

- After we plug-in this, the final variance estimator for $\hat{\boldsymbol{C}}^*(v)$ is $\hat{\mathrm{Var}}\left\{\hat{\boldsymbol{C}}^*(v)\right\} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{*\prime}\hat{\boldsymbol{W}}_i(v)^{-1}\boldsymbol{X}_i^*\right)^{-1}$.

- Hypothesis testing on any linear combination of the covariate effects can be performed.
- Standard multiple testing methods can be applied to control family wise error rate or false discovery rate.

Simulation Study

Table 1. Performance of L-ICA and TC-GICA

Subj-Visit Var	Population-level spatial maps Corr.(SD)		Subject/Vist-specific spatial maps Corr.(SD)	
	L-ICA	TC-GICA	L-ICA	TC-GICA
Low				
N=10	0.929 (0.021)	0.853 (0.116)	0.979 (0.016)	0.942 (0.095)
N=20	0.959 (0.015)	0.889 (0.113)	0.981 (0.012)	0.937 (0.093)
N=60	0.984 (0.008)	0.940 (0.109)	0.999 (0.007)	0.951 (0.085)
High				
N=10	0.886 (0.053)	0.621 (0.213)	0.960 (0.044)	0.845 (0.152)
N=20	0.899 (0.042)	0.691 (0.187)	0.962 (0.034)	0.854 (0.141)
N=60	0.958 (0.011)	0.856 (0.162)	0.991 (0.019)	0.900 (0.099)

Subj-Visit Var.	Subject/Vist-specific time courses Corr.(SD)		Covariate Effects MSE(SD)	
	L-ICA	TC-GICA	L-ICA	TC-GICA
Low				
N=10	0.997 (0.004)	0.941 (0.076)	0.152 (0.009)	0.159 (0.068)
N=20	0.998 (0.003)	0.942 (0.075)	0.093 (0.006)	0.153 (0.063)
N=60	1.000 (0.001)	0.957 (0.063)	0.040 (0.000)	0.128 (0.039)
High				
N=10	0.987 (0.019)	0.884 (0.092)	0.253 (0.015)	0.273 (0.101)
N=20	0.990 (0.014)	0.885 (0.093)	0.187 (0.011)	0.239 (0.086)
N=60	0.992 (0.007)	0.910 (0.077)	0.098 (0.004)	0.192 (0.083)

Table 2. Performance of Stochastic EM

Method	Computational time (SD)	Baseline population-level spatial maps Corr.(SD)	Stopping criteria Corr.
$q = 3$			
vss-EM	19.01 (1.09)	0.962 (0.002)	0.990
s-EM	55.26 (0.85)	0.962 (0.001)	0.990
exact EM	98.77 (2.53)	0.963 (0.001)	0.990
$q = 5$			
vss-EM	25.64 (2.53)	0.962 (0.005)	0.990
s-EM	165.89 (5.75)	0.961 (0.004)	0.990
exact EM	656.73 (6.71)	0.962 (0.005)	0.990
$q = 10$			
vss-EM	450.44 (7.21)	0.910 (0.010)	0.900
vss-EM	828.23 (8.11)	0.951 (0.011)	0.950
s-EM	4210.44 (11.21)	0.907 (0.009)	0.900
exact EM	39252.87 (12.01)	0.913 (0.010)	0.900

A) Type 1 Error Analysis

B) Power Analysis

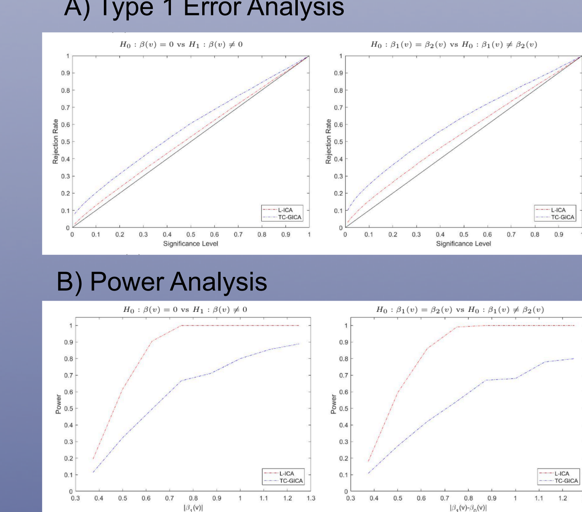


Figure 1. Performance of Approximate Inference

- L-ICA provides more accurate and robust estimation than TC-GICA.
- L-ICA has better statistical power and smaller type 1 error than TC-GICA.
- Proposed stochastic EM (vss-EM) is much more efficient than exact EM and common subspace EM.

ADNI2 Study

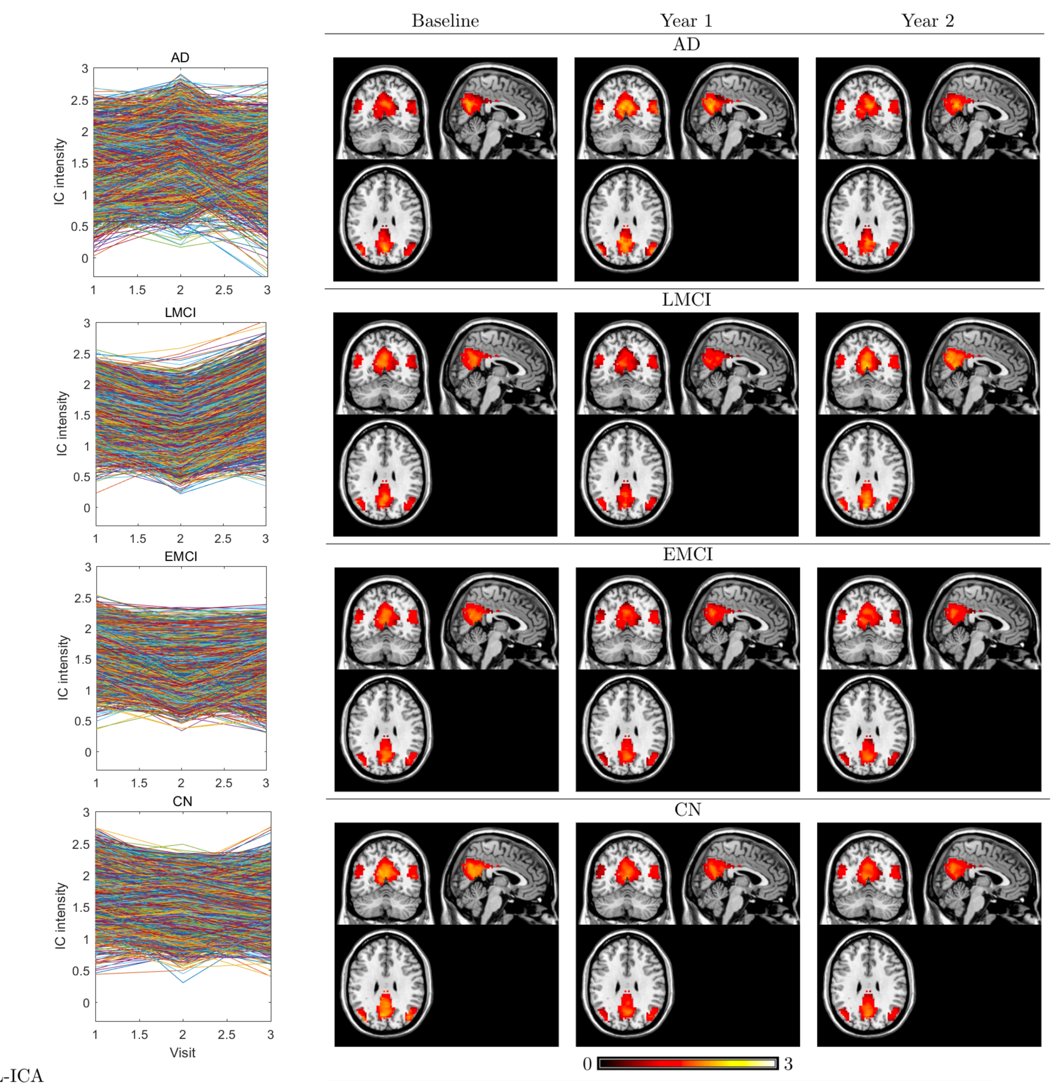


Figure 2. Group-level DMN based on LICA

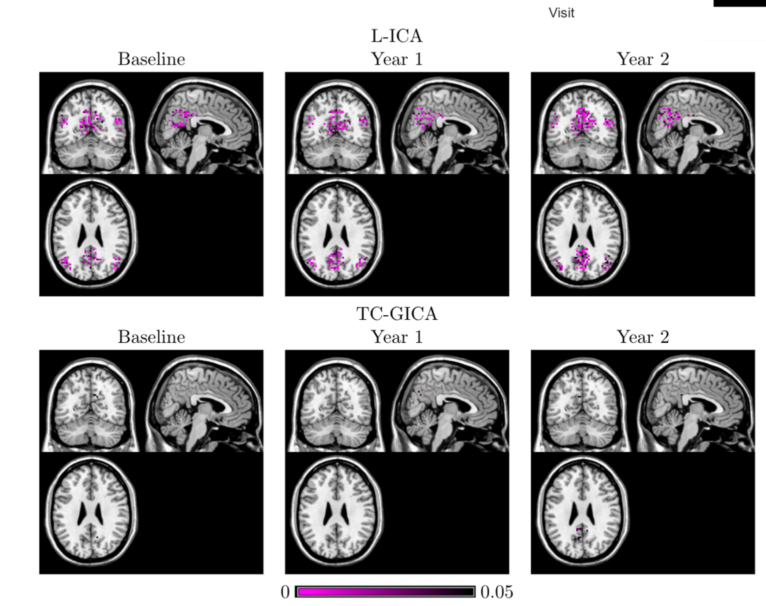


Figure 3. P-values on AD vs CN

ADNI2 is a longitudinal study aiming at examining changes in neuroimaging with the progression of mild cognitive impairment (MCI) and Alzheimer's Disease (AD).

- Longitudinal rs-fMRI images from 51 subjects that were collected at screening, 1 year and 2 year.
- Among 51 subjects, 16 were normal, 17 had EMCI, 12 had LMCI and 6 were diagnosed with AD at baseline

Summary

1: General hierarchical ICA modeling framework with broad applications.

2: Highly efficient stochastic EM algorithm with space encoding.

3: Approximate inference procedure for covariate effects.

Current / Related Works

- Connectivity ICA for network-valued data analysis;
- Discrete ICA for discrete data analysis;
- Template-driven single scan ICA : a robust estimation;
- Multi-site ICA to account for batch effects

References

1. Shi, Ran, and Ying Guo. "Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis." *The annals of applied statistics* 10.4 (2016): 1930.
2. Lukemire, Joshua, et al. "HINT: A Toolbox for Hierarchical Modeling of Neuroimaging Data." *arXiv preprint arXiv:1803.07587* (2018).
3. Wang, Yikai, et al "Longitudinal Independence Component Modeling of fMRI data ", arXiv preprint arXiv(2018).

ACKNOWLEDGEMENTS

- Supported by the NIH grants ROI MH105561.

Contact: Yikai Wang johnzon.wyk@gmail.com