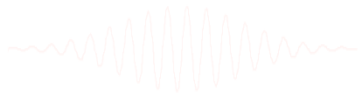
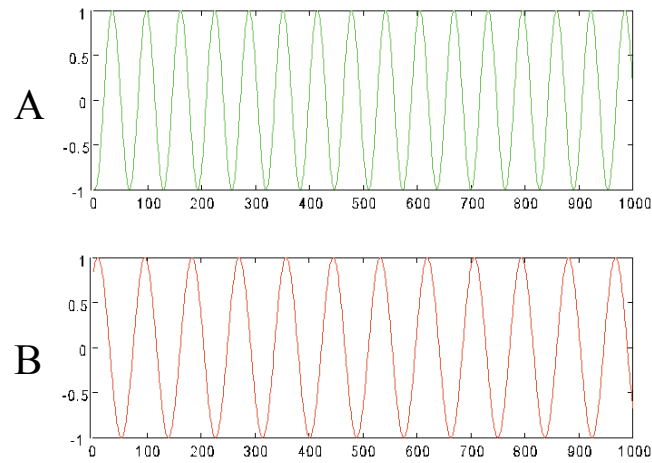


Clustering of ICA components

Arnaud Delorme

(with Julie Onton, Romain Grandchamp, Nima Bigdely Shamlo, Scott Makeig)

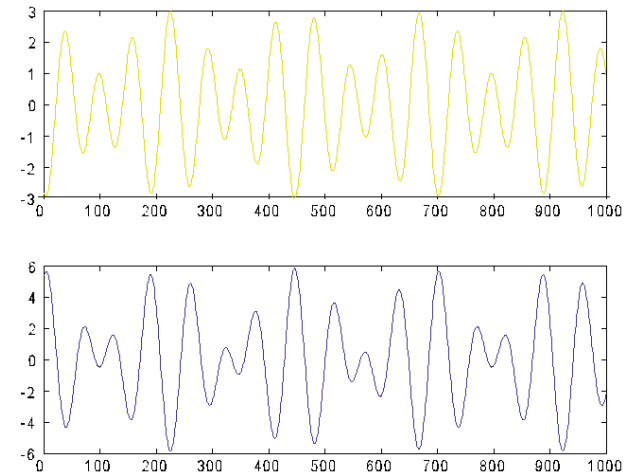
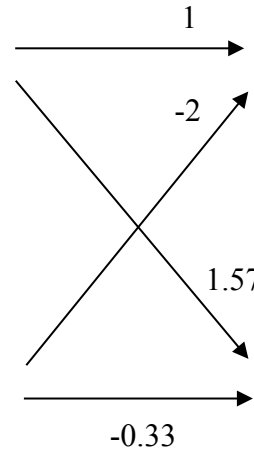




$$Y=[A;B]$$

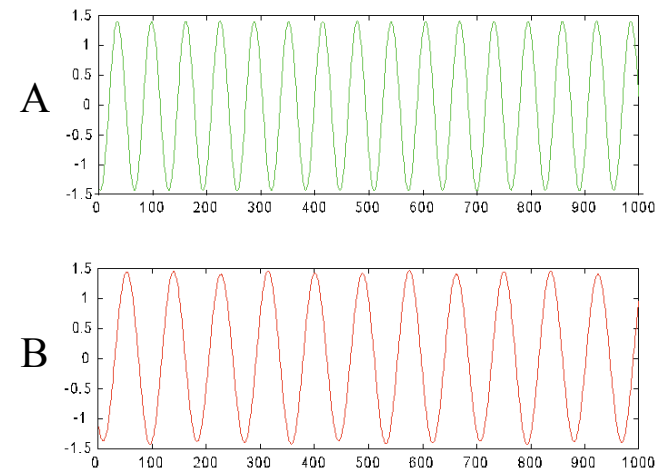
Linear Combination

$$X=YW$$

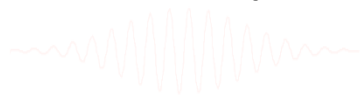


ICA

$$\tilde{Y}=W^{-1}\tilde{X}$$



ICA is a method to recover a version,
of the original sources by multiplying
the data by a unmixing matrix



$$\text{ICA activity } \mathbf{U} = \mathbf{W} \mathbf{X} \text{ Data}$$

Data X

$$\begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 & \dots \\ 0 & -2 & -5 & -1 & 1 & -1 & \dots \\ -1 & 2 & 0 & 1 & 0 & -3 & \dots \end{bmatrix} \begin{matrix} \leftarrow \text{Channel 1} \\ \leftarrow \text{Channel 2} \\ \leftarrow \text{Channel 3} \end{matrix}$$

$$\begin{bmatrix} 5 & 3 & -2 \\ 1 & 2 & 4 \\ 0 & -1 & 3 \end{bmatrix} \begin{matrix} * \\ * \\ * \end{matrix} \begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 & \dots \\ 0 & -2 & -5 & -1 & 1 & -1 & \dots \\ -1 & 2 & 0 & 1 & 0 & -3 & \dots \end{bmatrix} \rightarrow \begin{bmatrix} 3*5+0*3-1*(-2) & 2*5+(-2)*3+2*(-2) & \dots \\ 3*1+0*2-1*4 & 2*1+(-2)*2+2*4 & \dots \\ 5*1-5*2+0*4 & 5*1-5*2+0*4 & \dots \end{bmatrix} \begin{matrix} \leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \end{matrix}$$

Weight matrix W

ICA activity U

$$\text{Data} \rightarrow \mathbf{X} = \mathbf{W}^{-1} \mathbf{U} \rightarrow \text{ICA activity } \mathbf{U}$$

$$\begin{array}{c}
 \begin{bmatrix} 5 & 3 & -2 \\ 1 & 2 & 4 \\ 0 & -1 & 3 \end{bmatrix} \xrightarrow{\begin{matrix} * \\ * \\ * \end{matrix}} \begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 \\ 0 & -2 & -5 & -1 & 1 & -1 & \dots \\ -1 & 2 & 0 & 1 & 0 & -3 \end{bmatrix} \begin{array}{l} \leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \end{array} \\
 \downarrow \\
 \begin{bmatrix} 3*5 + 0*3 - 1*(-2) & 2*5 + (-2)*3 + 2*(-2) \\ 3*1 + 0*2 - 1*4 & 2*1 + (-2)*2 + 2*4 & \dots \\ 5*1 - 5*2 + 0*4 & 5*1 - 5*2 + 0*4 \end{bmatrix} \begin{array}{l} \leftarrow \text{Chan 1} \\ \leftarrow \text{Chan 2} \\ \leftarrow \text{Chan 3} \end{array} \\
 \text{Inverse weight matrix } \mathbf{W}^{-1} \qquad \text{Data } \mathbf{X}
 \end{array}$$



HISTORICAL REMARKS

- Herault & Jutten ("Space or time adaptive signal processing by neural network models", *Neural Nets for Computing Meeting*, Snowbird, Utah, 1986): **Seminal paper, neural network**
- Bell & Sejnowski (1995): Information Maximization
- Amari et al. (1996): Natural Gradient Learning
- Cardoso (1996): JADE
- **Applications of ICA to biomedical signals**
 - EEG/ERP analysis (Makeig, Bell, Jung & Sejnowski, 1996).
 - fMRI analysis (McKeown et al. 1998)



ICA Theory – Cost Functions

Family of BSS algorithms

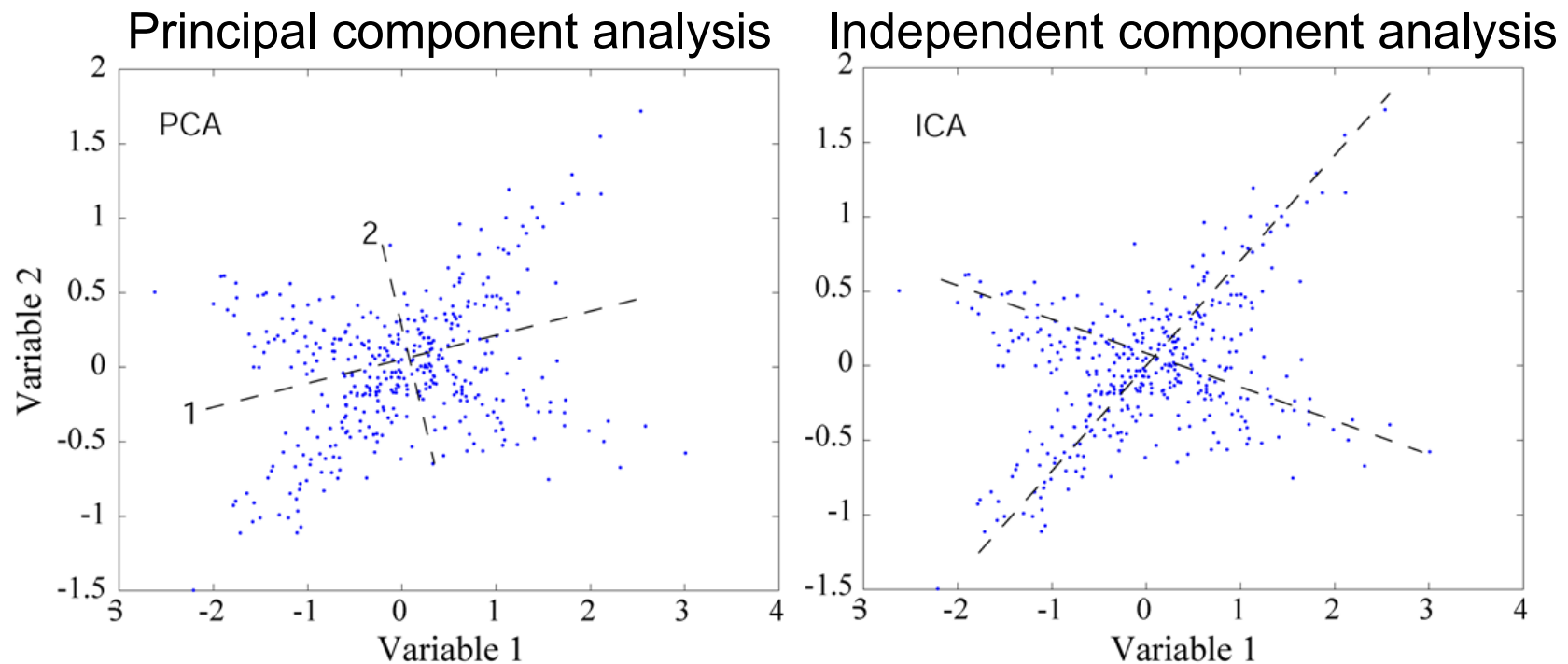
- Information theory (Infomax)
- Bayesian probability theory (Maximum likelihood estimation)
- Negentropy maximization
- Nonlinear PCA
- Statistical signal processing (cumulant maximization, JADE)

A unifying Information-theoretic framework for ICA

- Pearlmutter & Parra showed that InfoMax, ML estimation are equivalent.
- Lee et al. (1999) showed negentropy has the equivalent property to InfoMax.
- Girolami & Fyfe showed nonlinear PCA can be viewed from

ICA and PCA

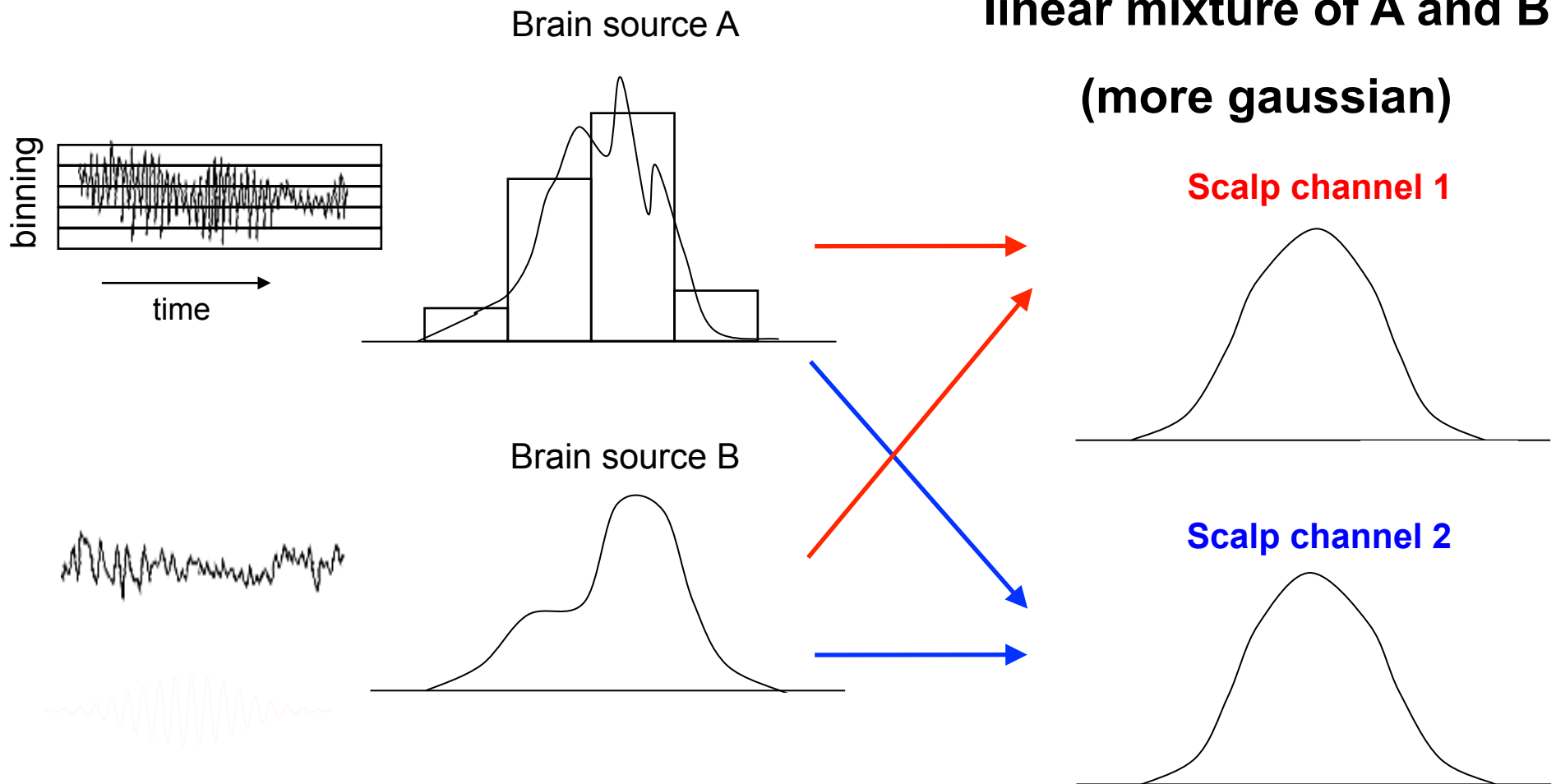
ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,



While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs **statistically independent**, while placing no constraints on the mixing matrix.

Central limit theorem

**Scalp channels =
linear mixture of A and B
(more gaussian)**

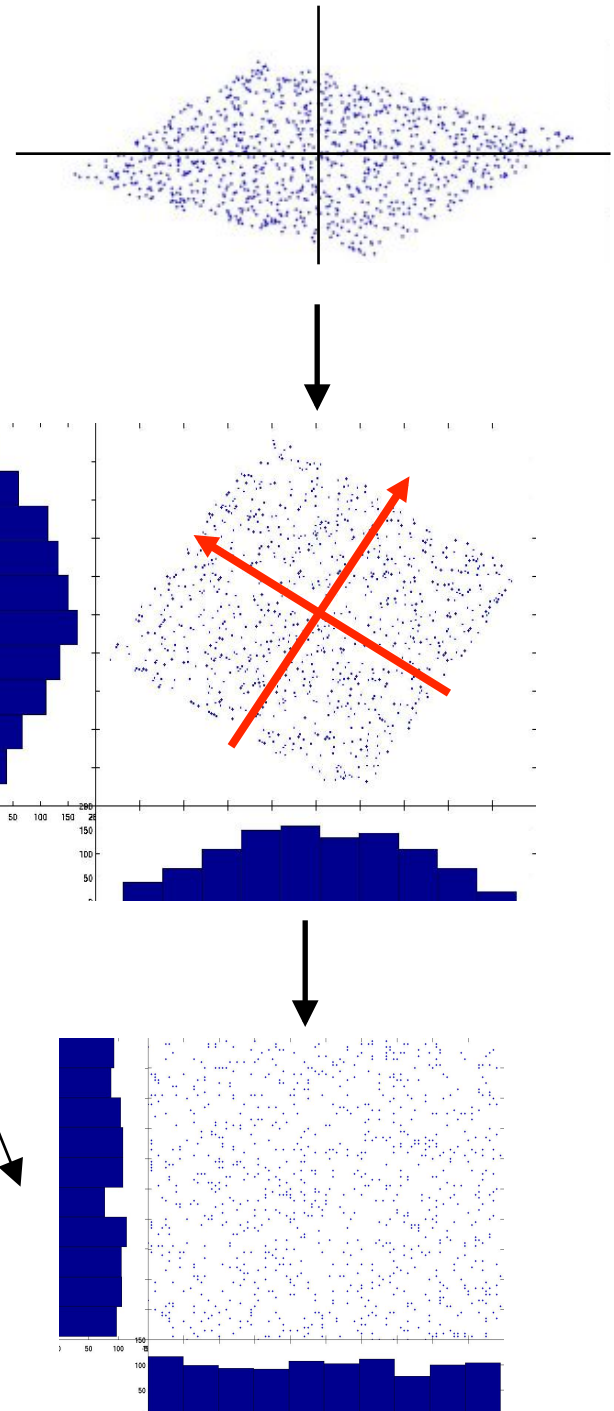


ICA Training Process

Central limit theorem

- Remove the mean
 $\mathbf{x} = \mathbf{x} - \langle \mathbf{x} \rangle$
- 'Sphere' the data by diagonalizing its covariance matrix,
 $\mathbf{x} = \langle \mathbf{x} \mathbf{x}^T \rangle^{-1/2} (\mathbf{x} - \langle \mathbf{x} \rangle)$.
- Update \mathbf{W} according to

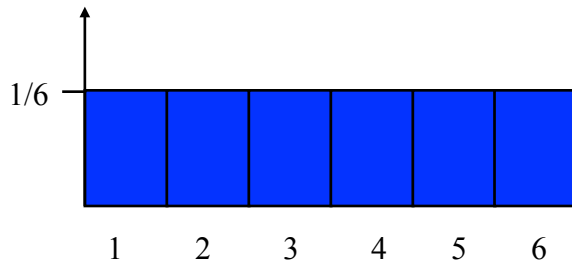
$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}$$



Entropy

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

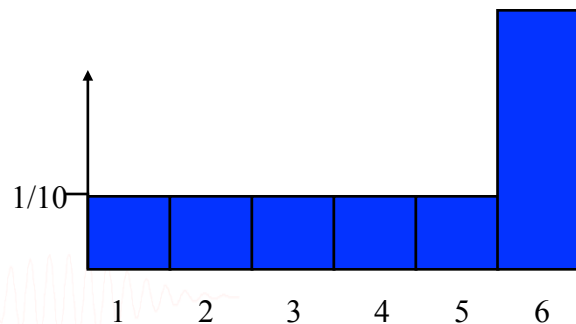
Dice: 1/6



$$H = 6 \left(-\frac{1}{6} \log_2 \left(\frac{1}{6} \right) \right) = 2.58$$

Fake dice (make a 6 half of the time): entropy 2.16 (base 2)

less random



$$H = 5 \left(-\frac{1}{10} \log_2 \left(\frac{1}{10} \right) \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) = 2.16$$

Entropy

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

Joint entropy

$$H(X, Y) = - \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_b p(x, y).$$

Mutual Information

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$



Shannon in his landmark 1948 paper "A Mathematical Theory of Communication."

From <http://planetmath.org/encyclopedia/ShannonsTheoremEntropy.html>

Contingency table for stress and emotionality

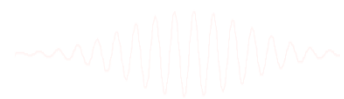
	STRE						
	1	2	3	4	5	6	Total
EMOT= 1	19	4					23
2	11	63	64	3	1		142
3	2	16	18	20	2	2	60
4	1	4	1	9	6	2	23
5			1	2	4	3	10
6				1	1	1	3
Total	33	87	84	35	13	8	

From <http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/node149.html>

Contingency frequencies for stress and emotionality

	STRE					
	1	2	3	4	5	6
EMOT= 1	0.07	0.02				
2	0.04	0.24	0.25	0.01		
3	0.01	0.06	0.07	0.08	0.01	0.01
4		0.02		0.03	0.02	0.01
5				0.01	0.02	0.01
6						

Joint entropy 3.46; exercise: compute mutual information



$$H(X, Y) = - \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_b p(x, y)$$

ICA learning rule

How to make the outputs statistical independent?

Minimize their redundancy or mutual information.

Consider the joint entropy of two components,

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$

Maximizing $H(y_1, y_2) \implies$ minimizing $I(y_1, y_2)$.

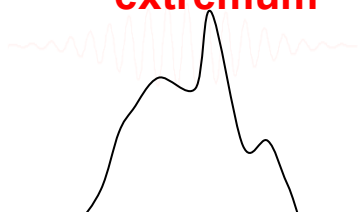
↓
The learning rule:

↓
=0 if the two variables
are independent

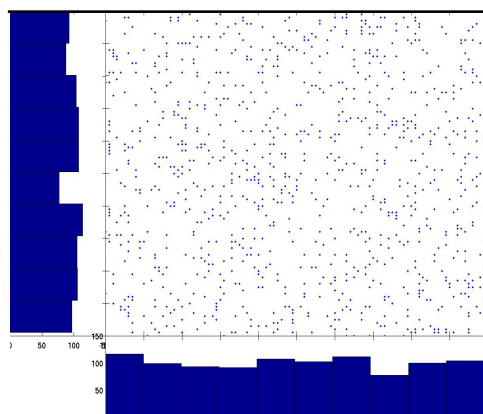
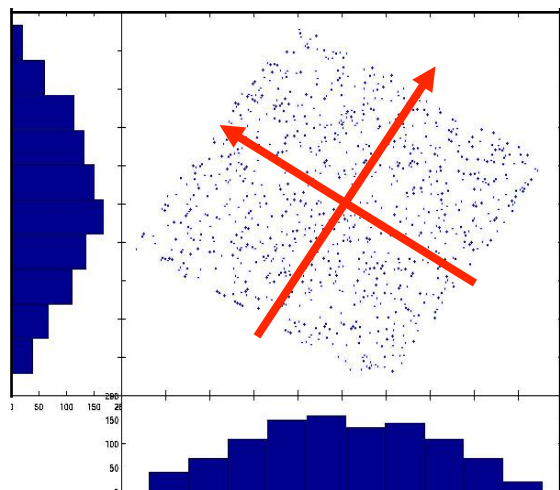
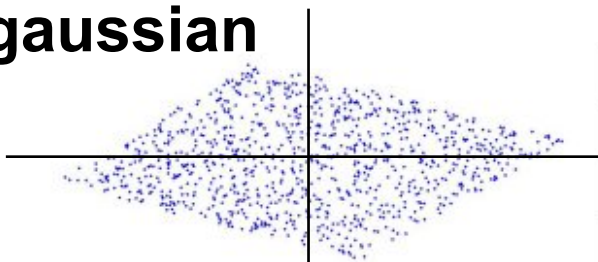
$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \underbrace{\mathbf{W}^T \mathbf{W}}$$

Natural gradient (Amari)

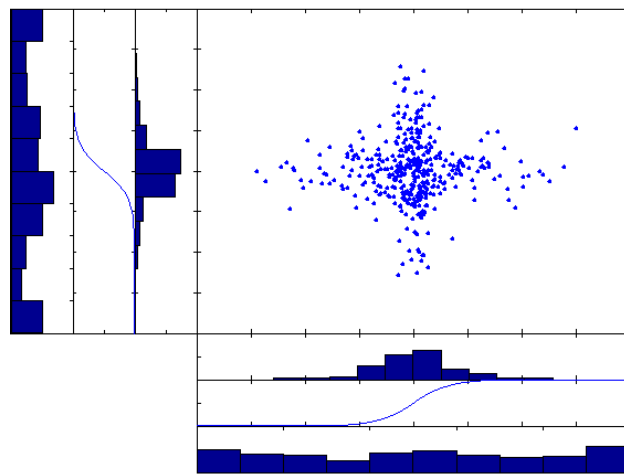
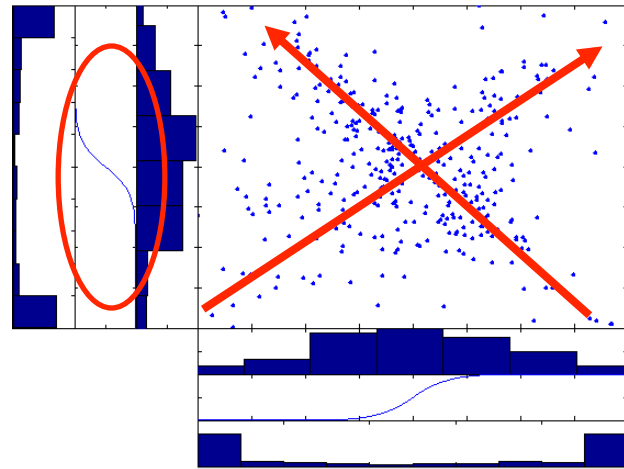
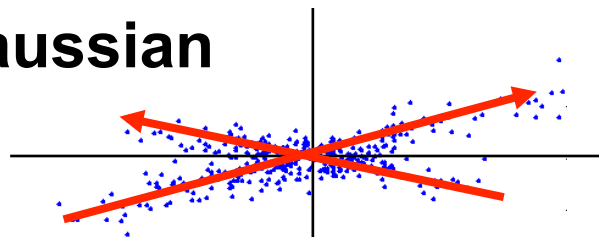
Entropy
extremum



Sub-gaussian



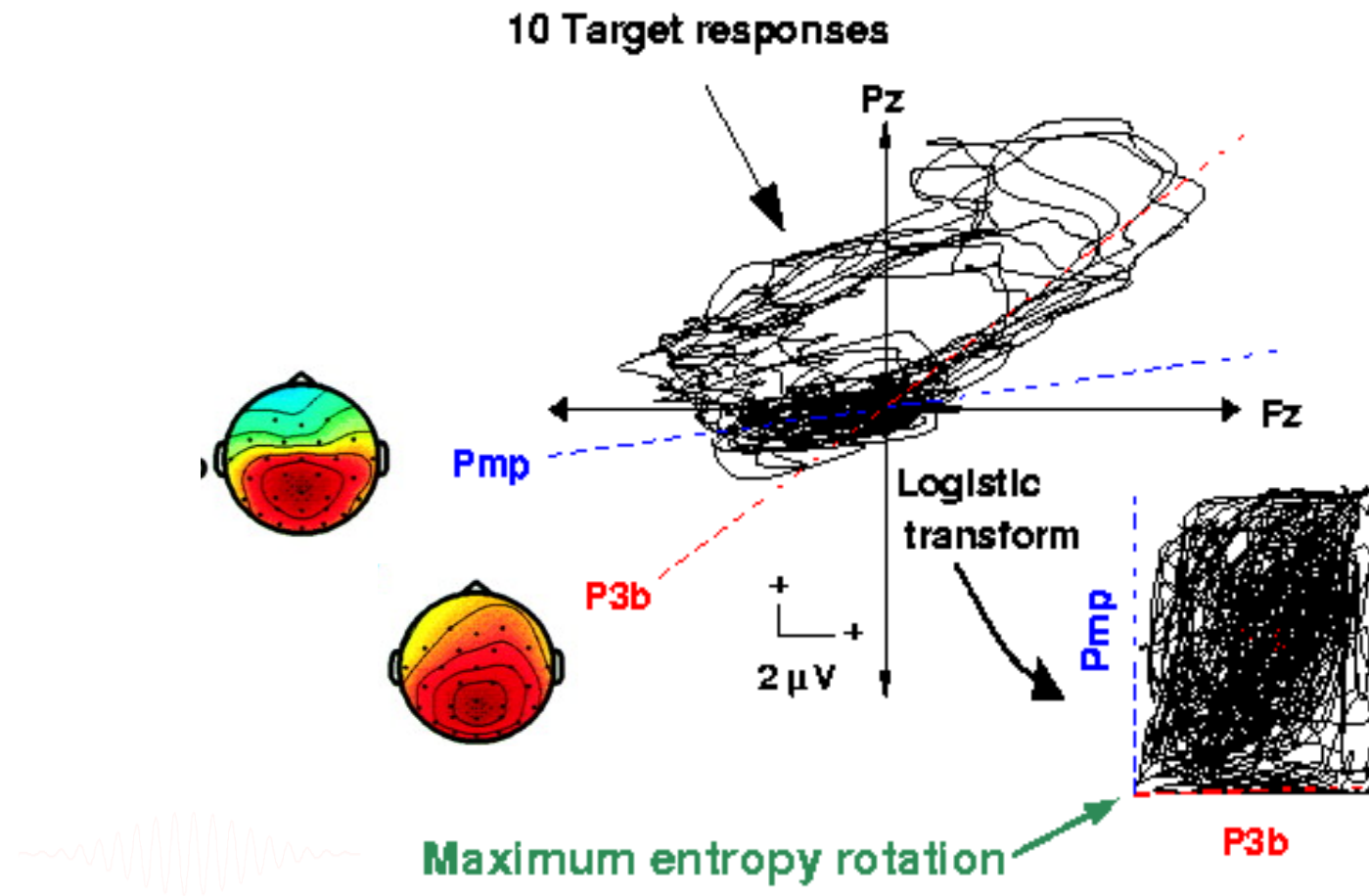
Super-gaussian



Sphering

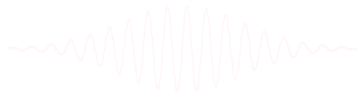
ICA

Independent components of EEG/ERF



Steps of clustering

- Select ICA components for clustering
- Precompute measures of interest
- Cluster measures
- Plot clusters and edit them if necessary



Edit dataset info

Create a new STUDY set -- pop_study()

Edit STUDY set information - remember to save changes

STUDY set name:

STUDY set task name:

STUDY set notes:

	dataset filename	browse	subject	session	condition	group	Select by r.v.	
1	C:\Users\julie\Documents\Wor	...	S01		memorize		Comp.: 3 5 ...	Clear
2	C:\Users\julie\Documents\Wor	...	S01		ignore		Comp.: 3 5 ...	Clear
3	C:\Users\julie\Documents\Wor	...	S01		probe		Comp.: 3 5 ...	Clear
4	C:\Users\julie\Documents\Wor	...	S02		memorize		Comp.: 5 6 ...	Clear
5	C:\Users\julie\Documents\Wor	...	S02		ignore		Comp.: 5 6 ...	Clear
6	C:\Users\julie\Documents\Wor	...	S02		probe		Comp.: 5 6 ...	Clear
7	C:\Users\julie\Documents\Wor	...	S03		memorize		Comp.: 6 7 ...	Clear
8	C:\Users\julie\Documents\Wor	...	S03		ignore		Comp.: 6 7 ...	Clear
9	C:\Users\julie\Documents\Wor	...	S03		probe		Comp.: 6 7 ...	Clear
10	C:\Users\julie\Documents\Wor	...	S04		memorize		Comp.: 1 2 ...	Clear

Important note: Removed datasets will not be saved before being deleted from EEGLAB memory

< Page 1 >

☐ Dataset info (condition, group, ...) differs from study info. [set] = Overwrite dataset info.

☐ Delete cluster information (to allow loading new datasets, set new components for clustering, etc.)

Help Cancel Ok

pop_study(): Pre-select components

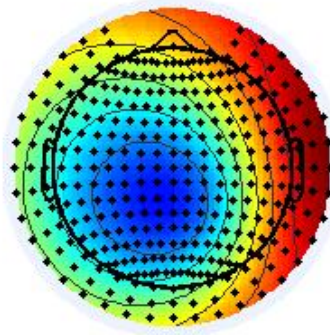
Enter maximum residual (topo map - dipole proj.) var. (in %)
NOTE: This will delete any existing component clusters!

☒ Keep only in-brain dipoles.

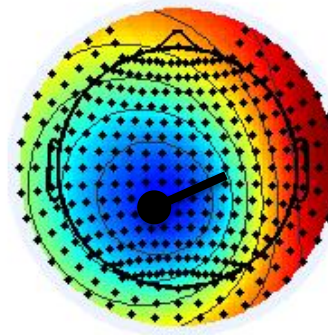
Cancel Help Ok

Computing residual variance (%)

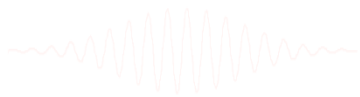
Actual



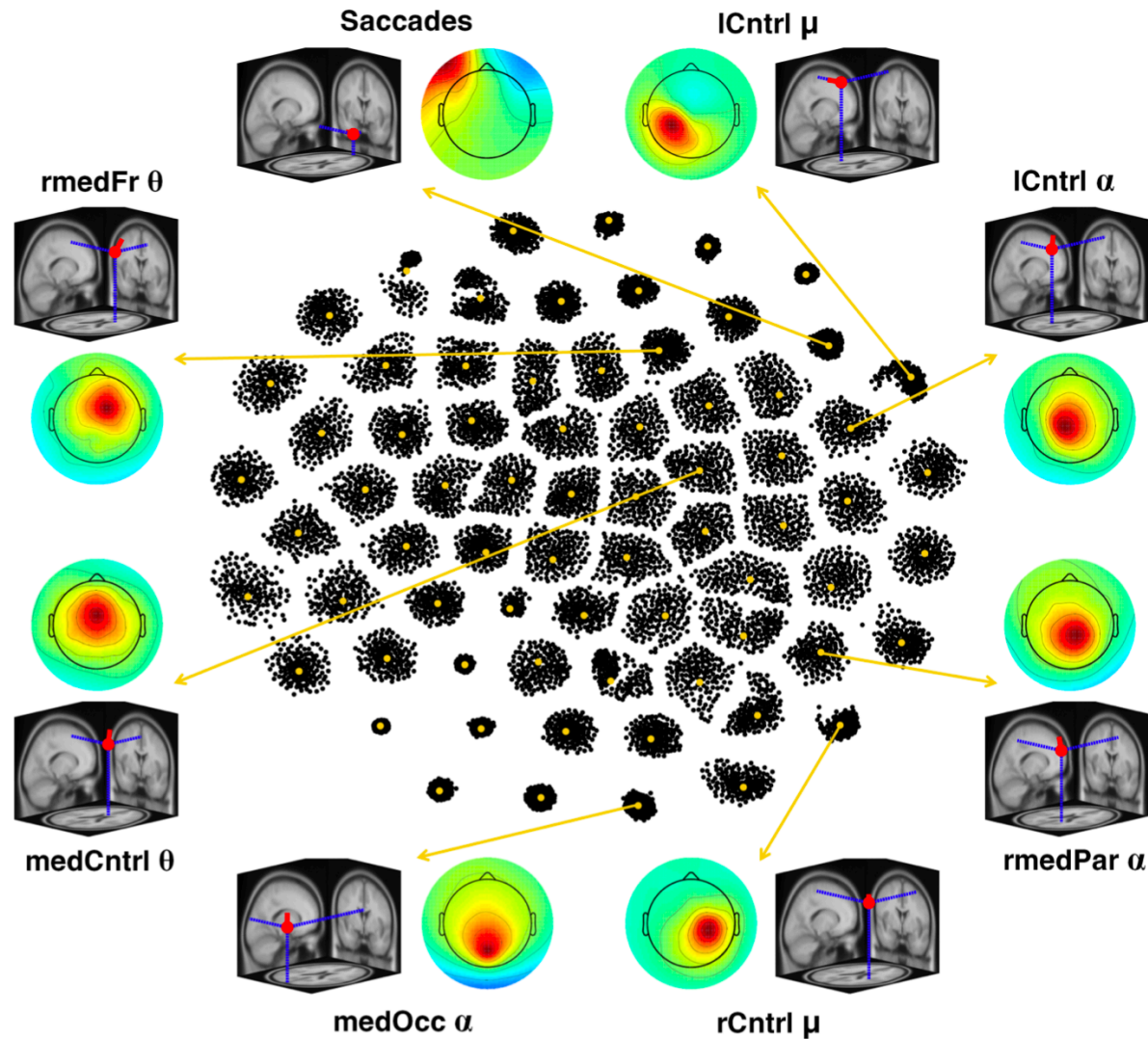
Dipole projection



$$r = \frac{\sum (x_i - \tilde{x}_i)^2}{\sum x_i^2}$$



Clustering results example

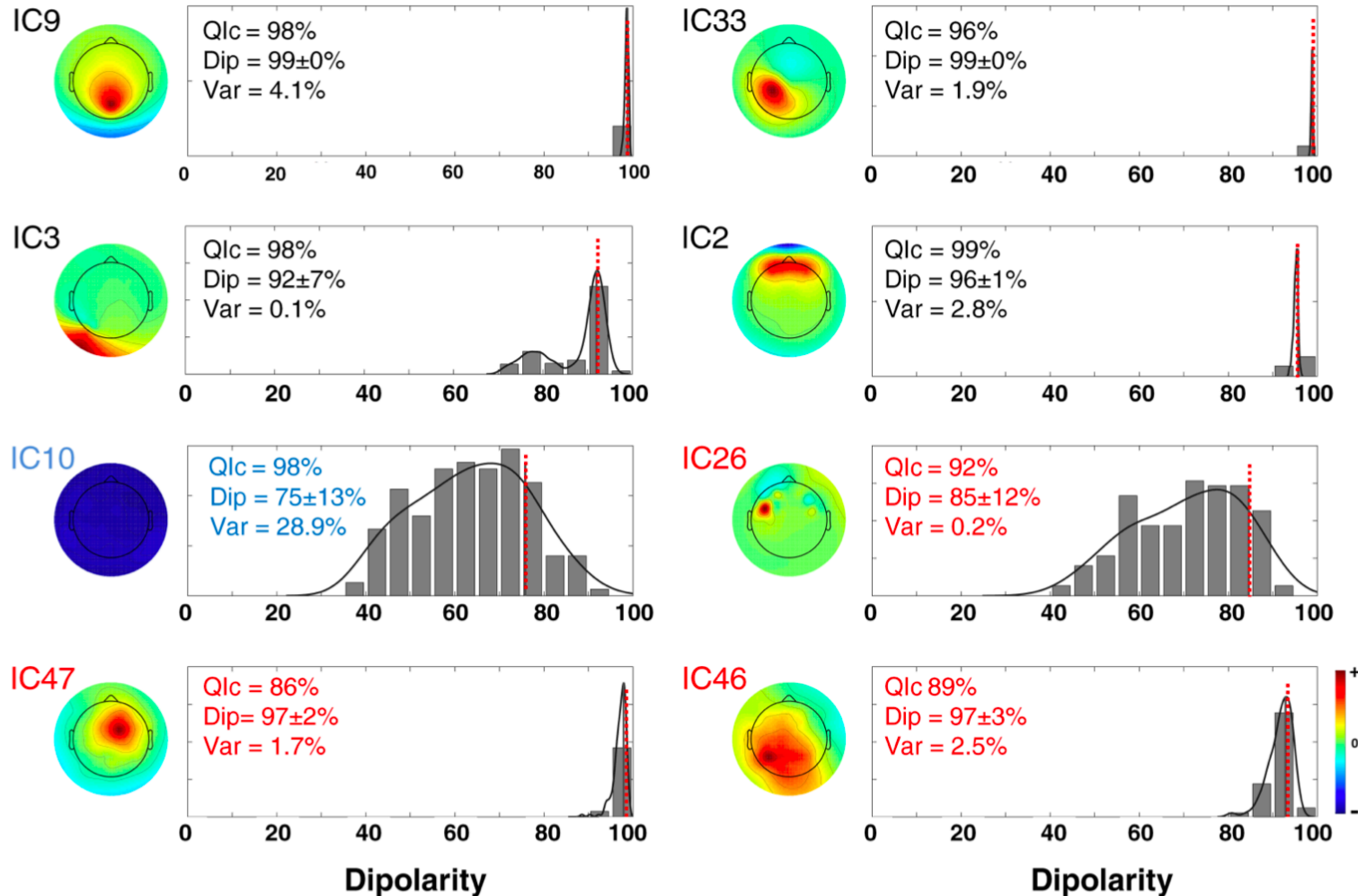


RELICA: A method for estimating the reliability of independent components

Fiorenzo Artoni ^{a,*}, Danilo Menicucci ^b, Arnaud Delorme ^{c,e,f}, Scott Makeig ^c, Silvestro Micera ^{a,d}

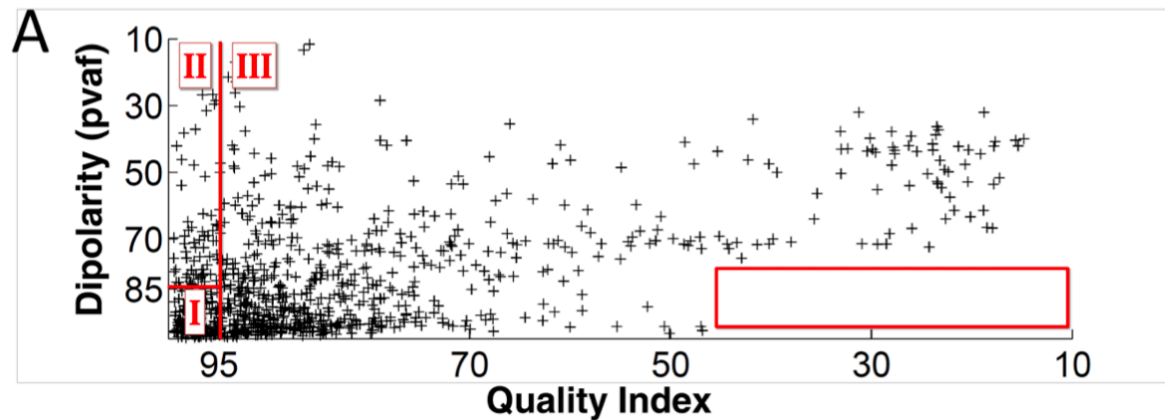
Within-cluster reliability

The distribution of dipolarity within the cluster helps assessing the **quality** and characteristics of Independent Components



Reliability criteria and the $rv < 15\%$

First justification why we should select an $rv < 15\%$ for components to include in further analyses: there is a forbidden region underlined in red, that indicates the absence of



CLASS I

Quality Index and Dipolarity above Retention threshold: **Good**

CLASS II

Quality Index above threshold, dipolarity below: **artifact** or mixing of multiple processes

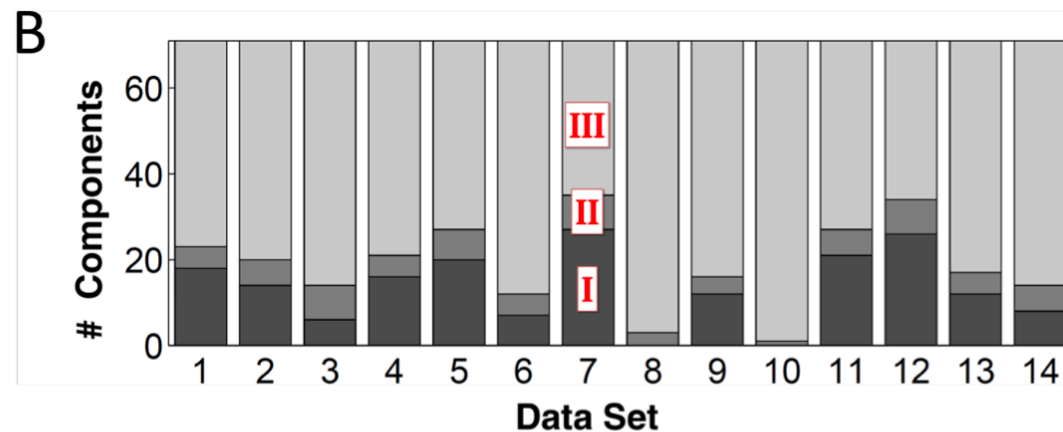
CLASS III

Quality Index below retention threshold

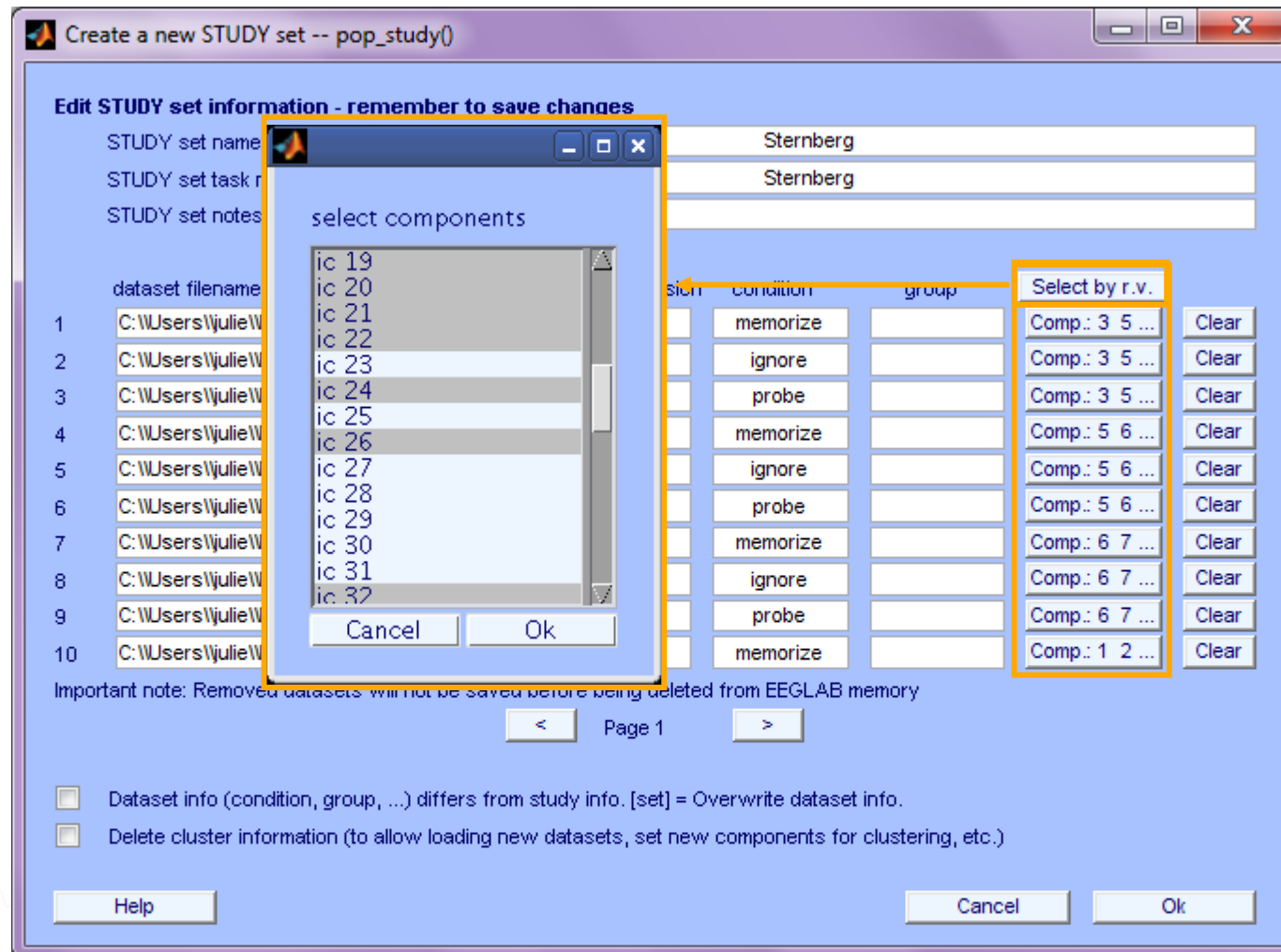
$$dip \pm std > th$$

$$dip \pm std < th$$

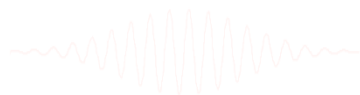
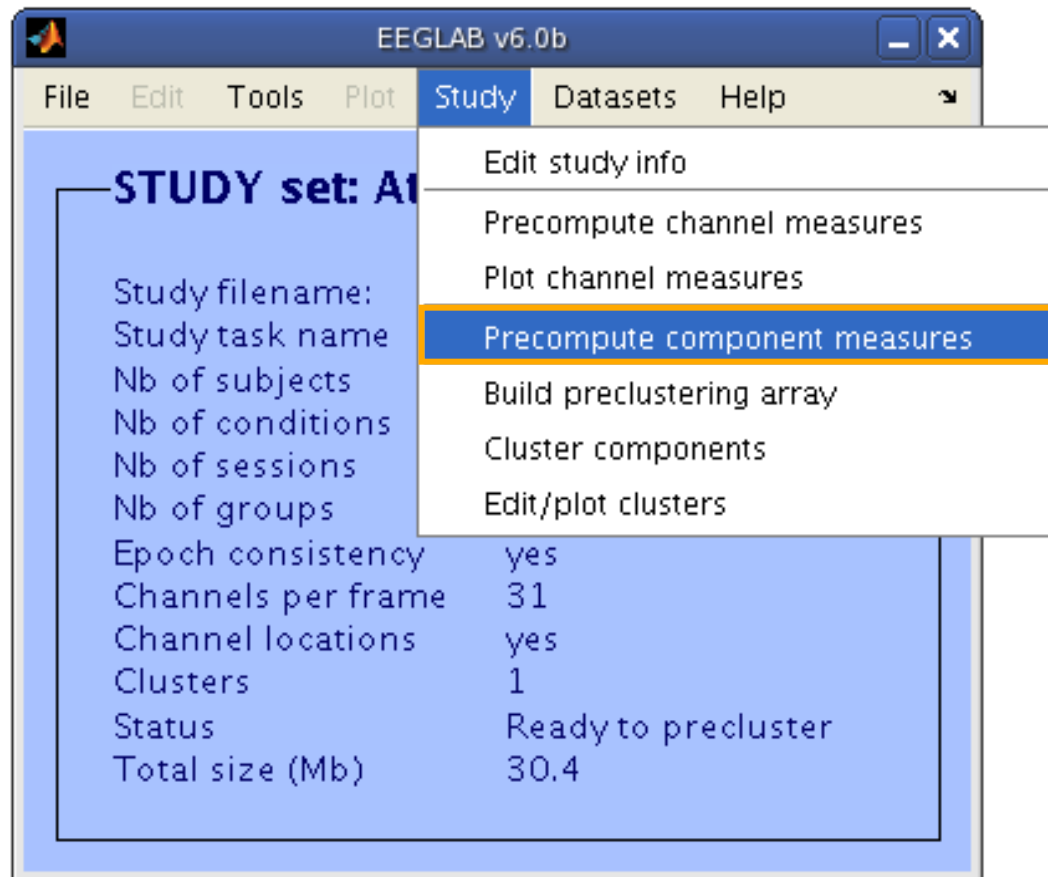
Probable Inseparable **Discard**
noise: variance
explained useful or
**multiple subject
confirmation**



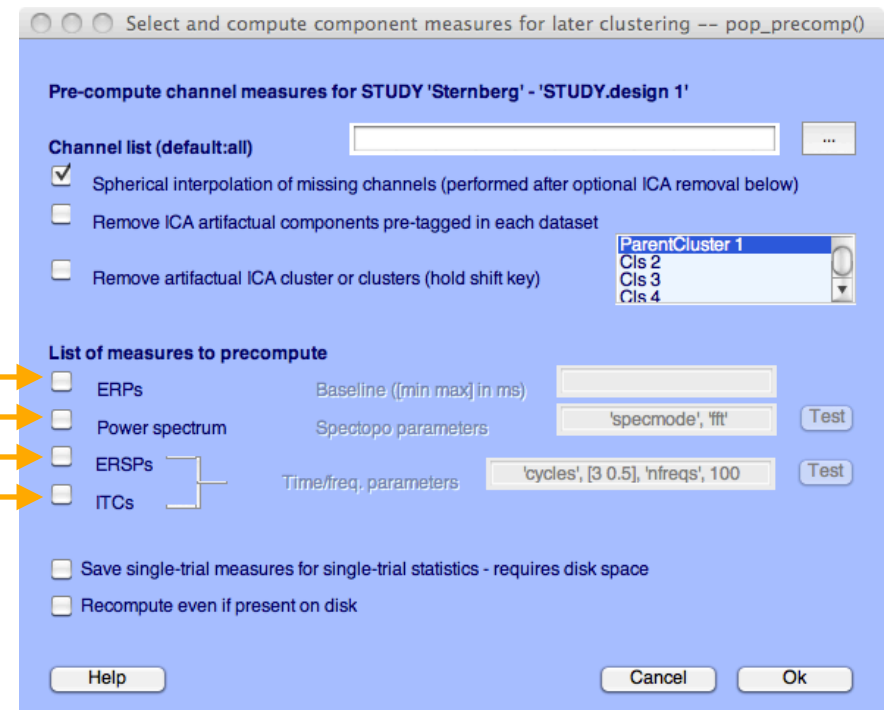
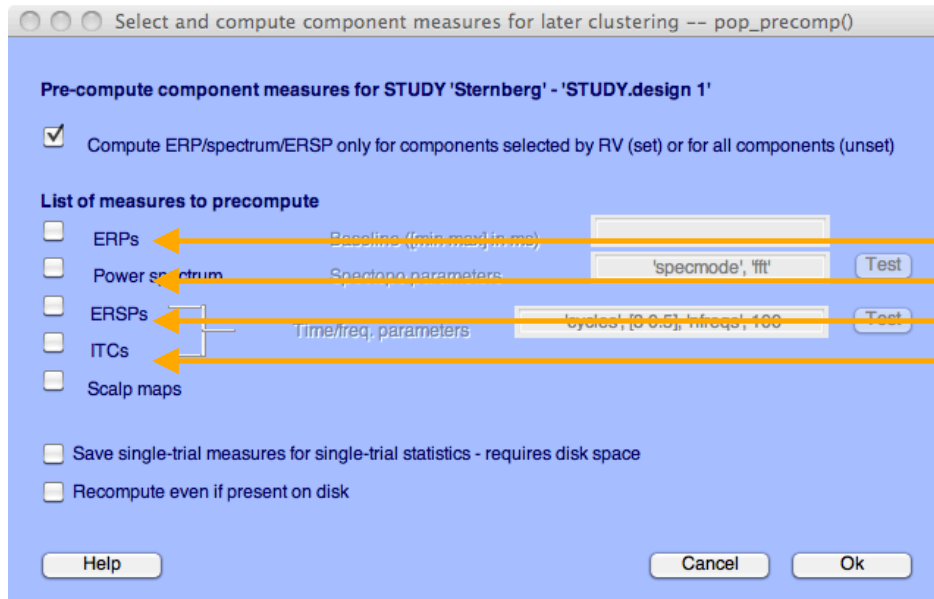
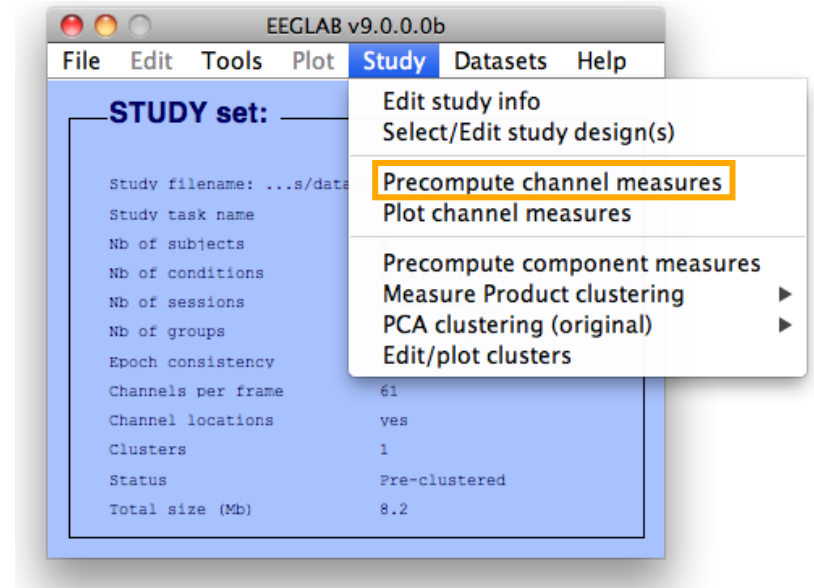
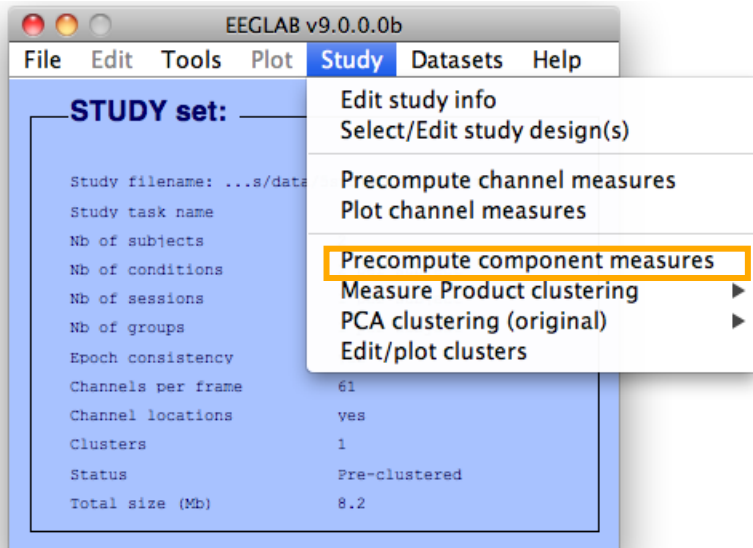
ICs to cluster



Precompute data measures



Pre-compute measures



Precompute data measures

TIP: Compute all measures so you can test different combinations for clustering

Select and compute component measures for later clustering -- pop_precomp()

Pre-compute component measures for STUDY 'Sternberg'

☒ Compute ERP/spectrum/ERSP only for components selected by RV (set) or for all components (unset)

List of measures to precompute

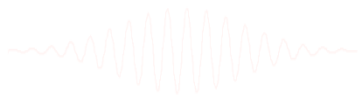
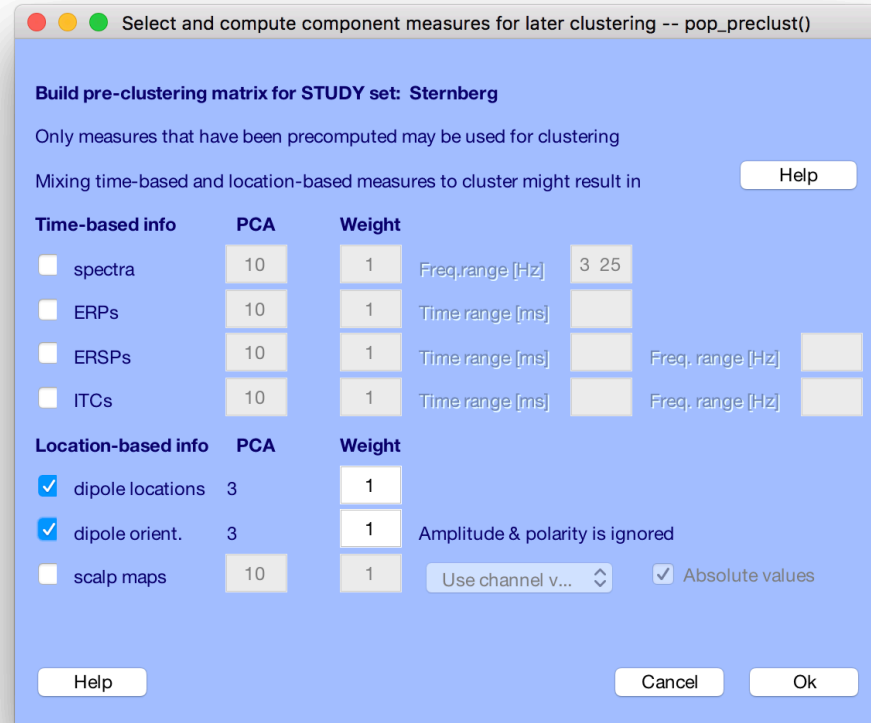
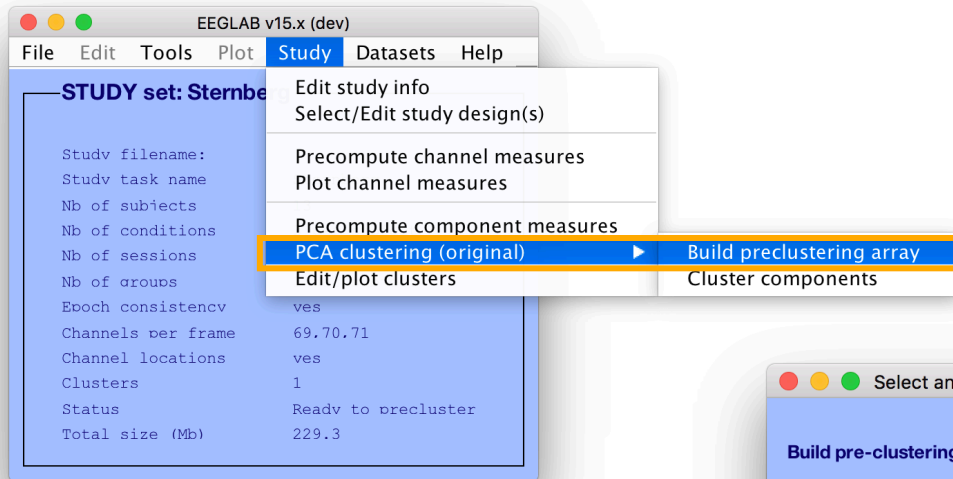
<input checked="" type="checkbox"/> ERPs	Baseline ([min max] in ms)	<input type="text" value="[-200 0]"/>	
<input checked="" type="checkbox"/> Power spectrum	Spectopo parameters	<input type="text" value=""/>	<input type="button" value="Test"/>
<input checked="" type="checkbox"/> ERSPs	Time/freq. parameters	<input type="text" value="'cycles', [3 0.5], 'nfreqs', 100"/>	<input type="button" value="Test"/>
<input checked="" type="checkbox"/> ITCs			
<input checked="" type="checkbox"/> Scalp maps			

☐ Recompute even if present on disk

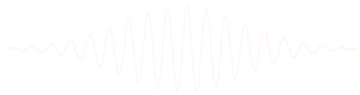
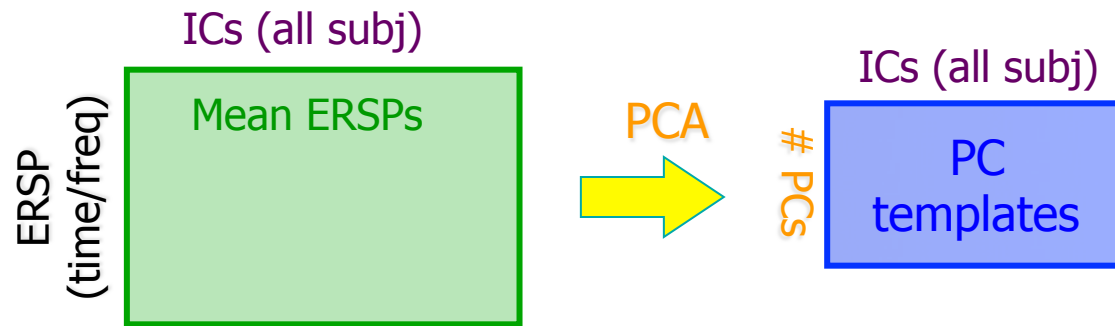
Time-frequency options



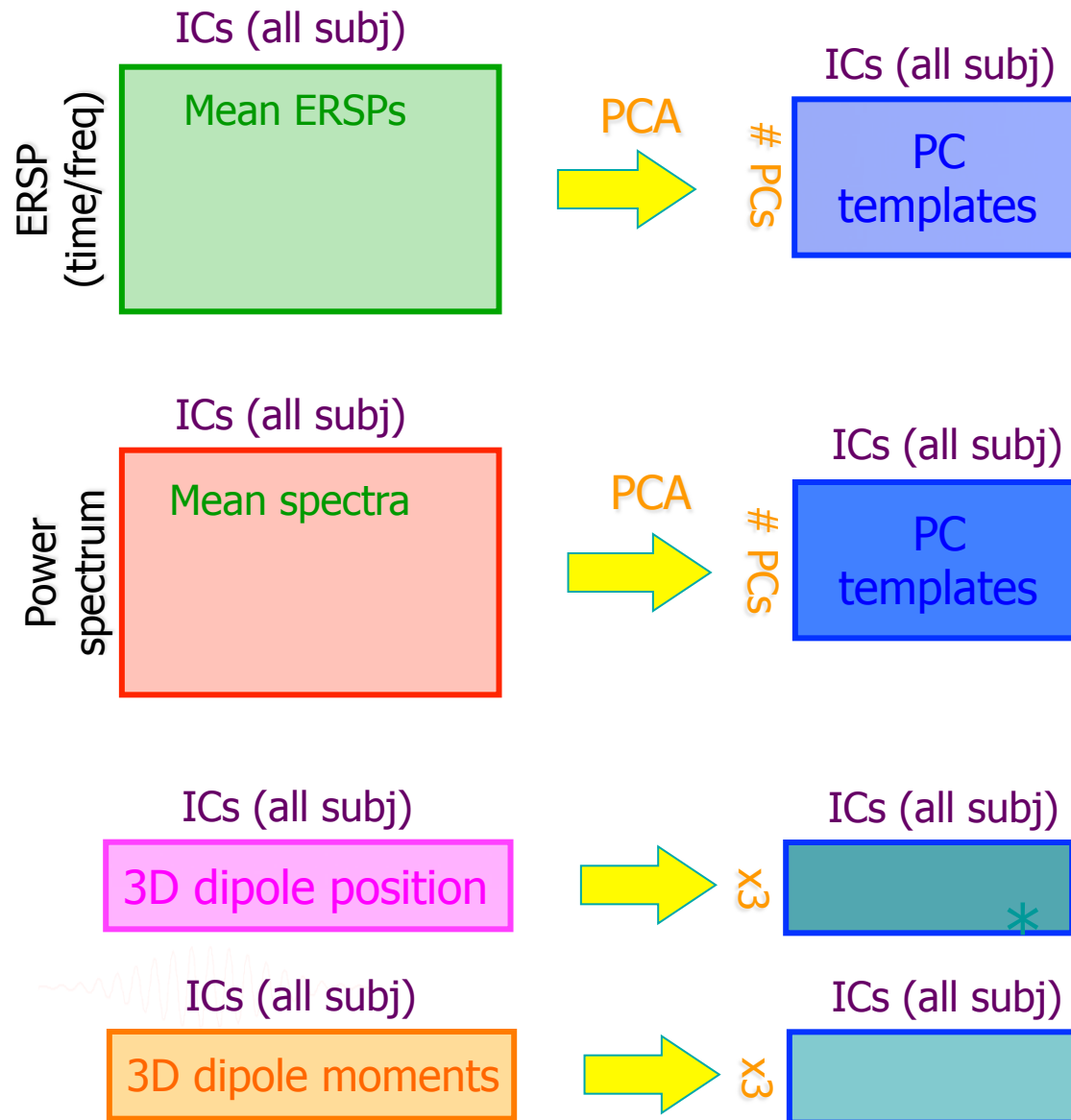
Cluster components



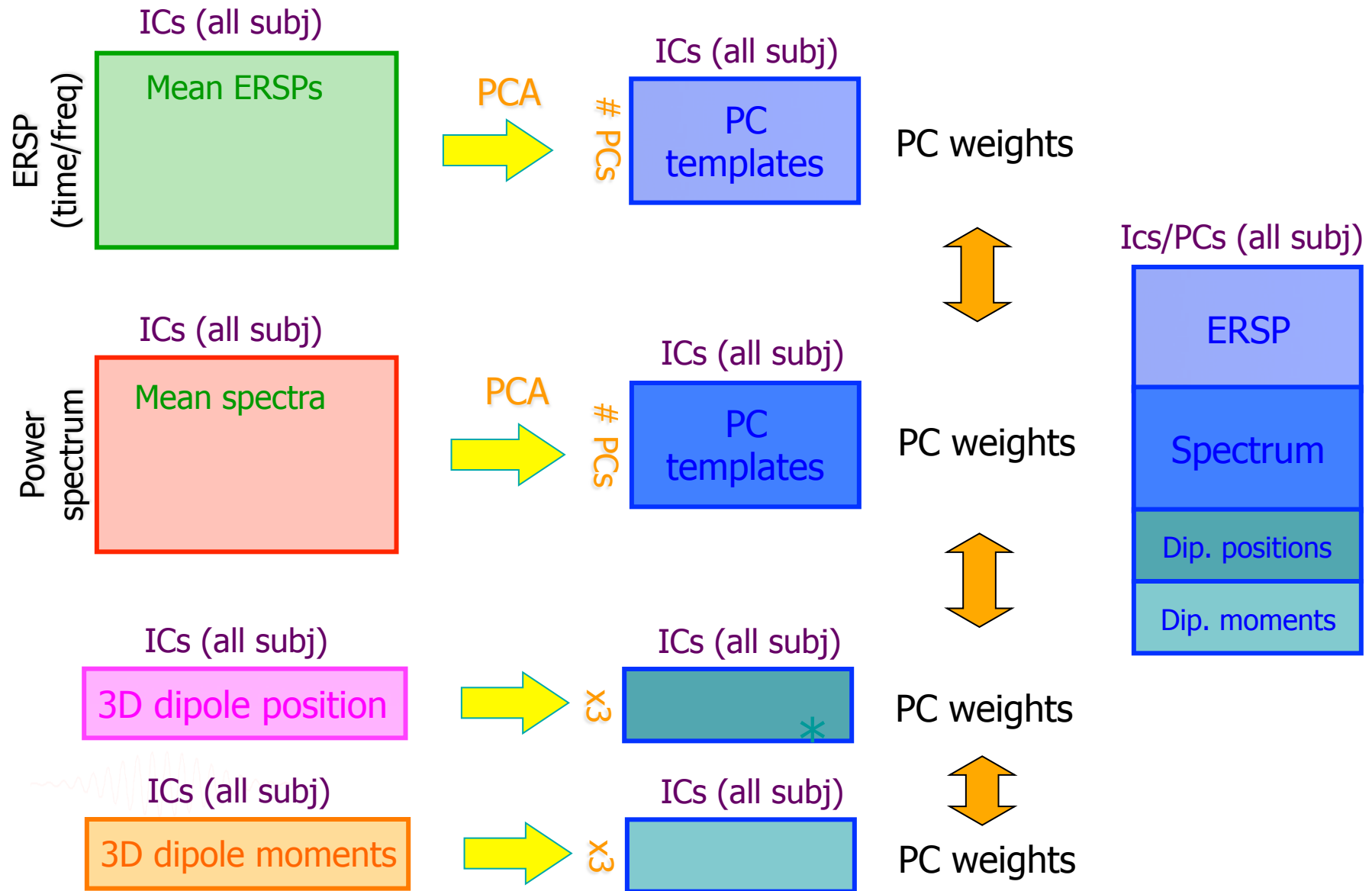
Precluster schematic



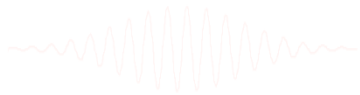
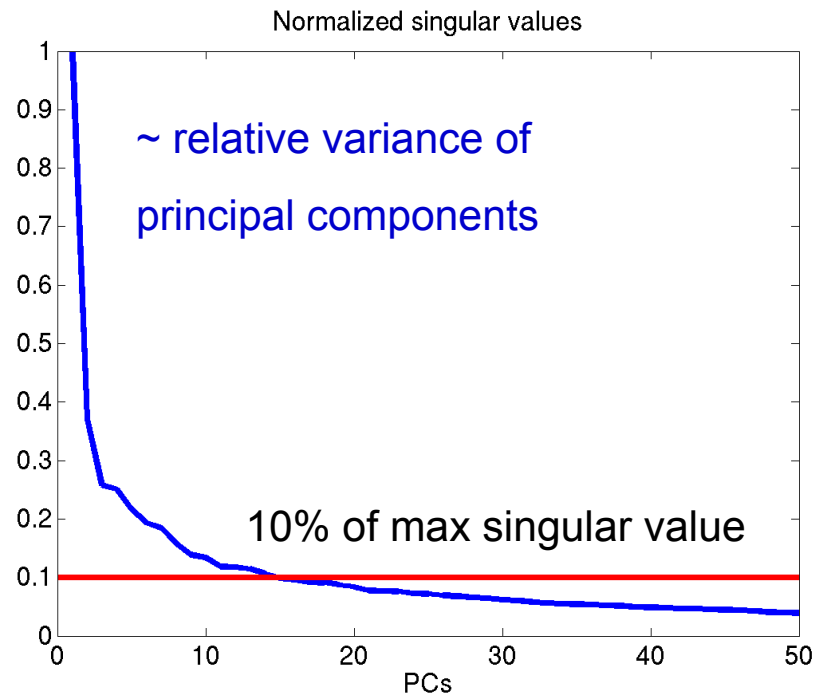
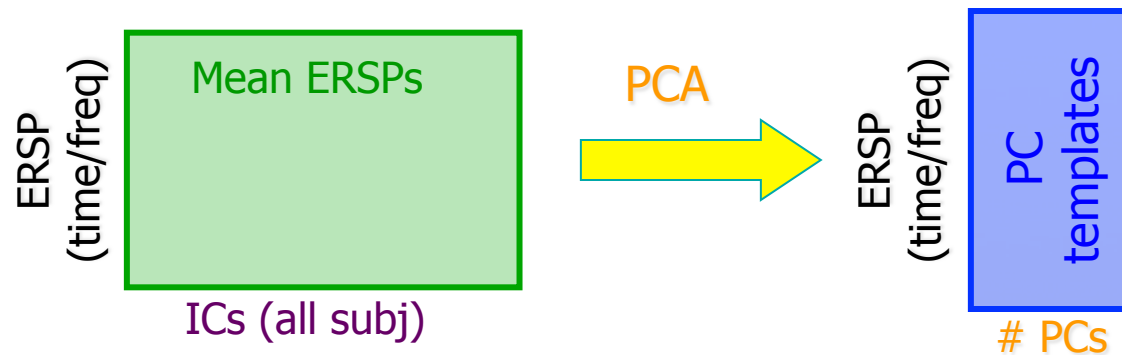
Precluster schematic



Precluster schematic



Precluster: Use singular values from PCA

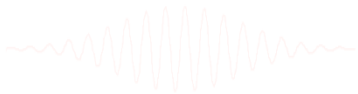
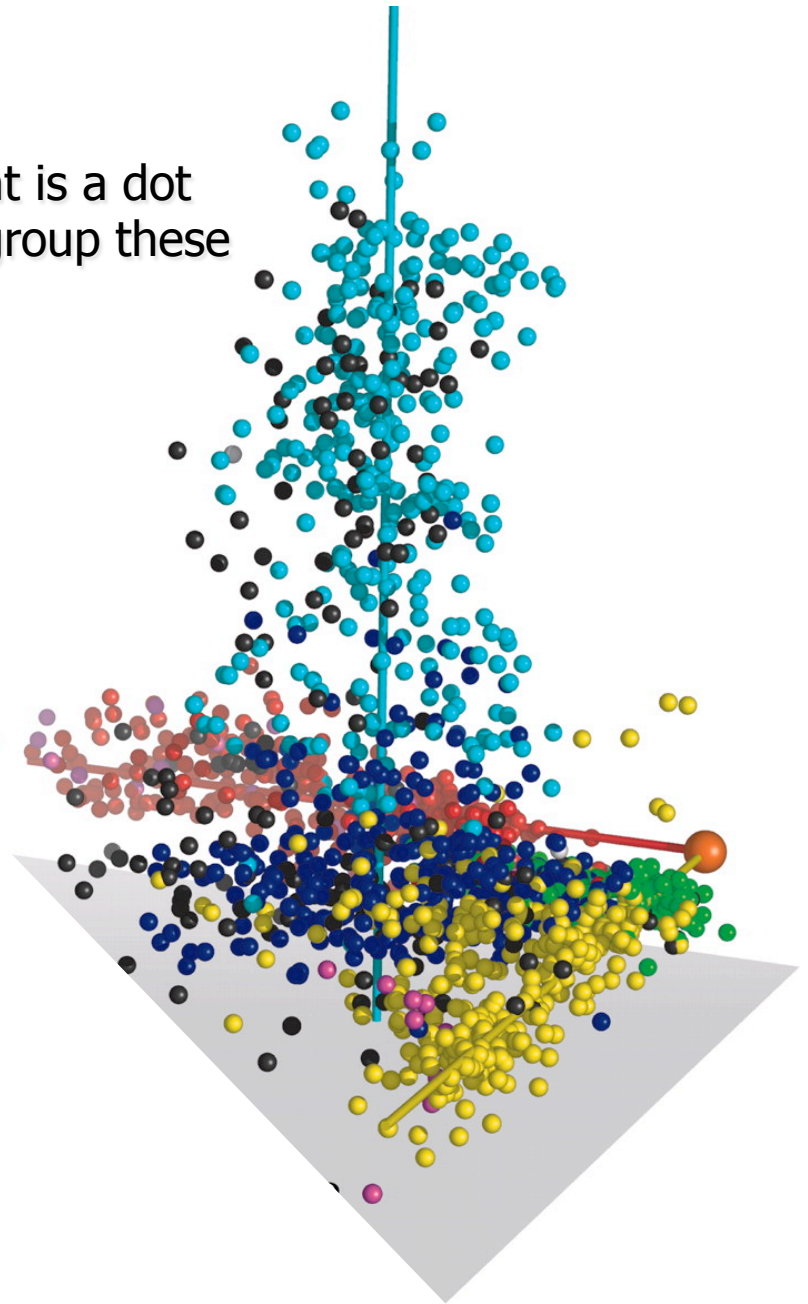


Precluster schematic

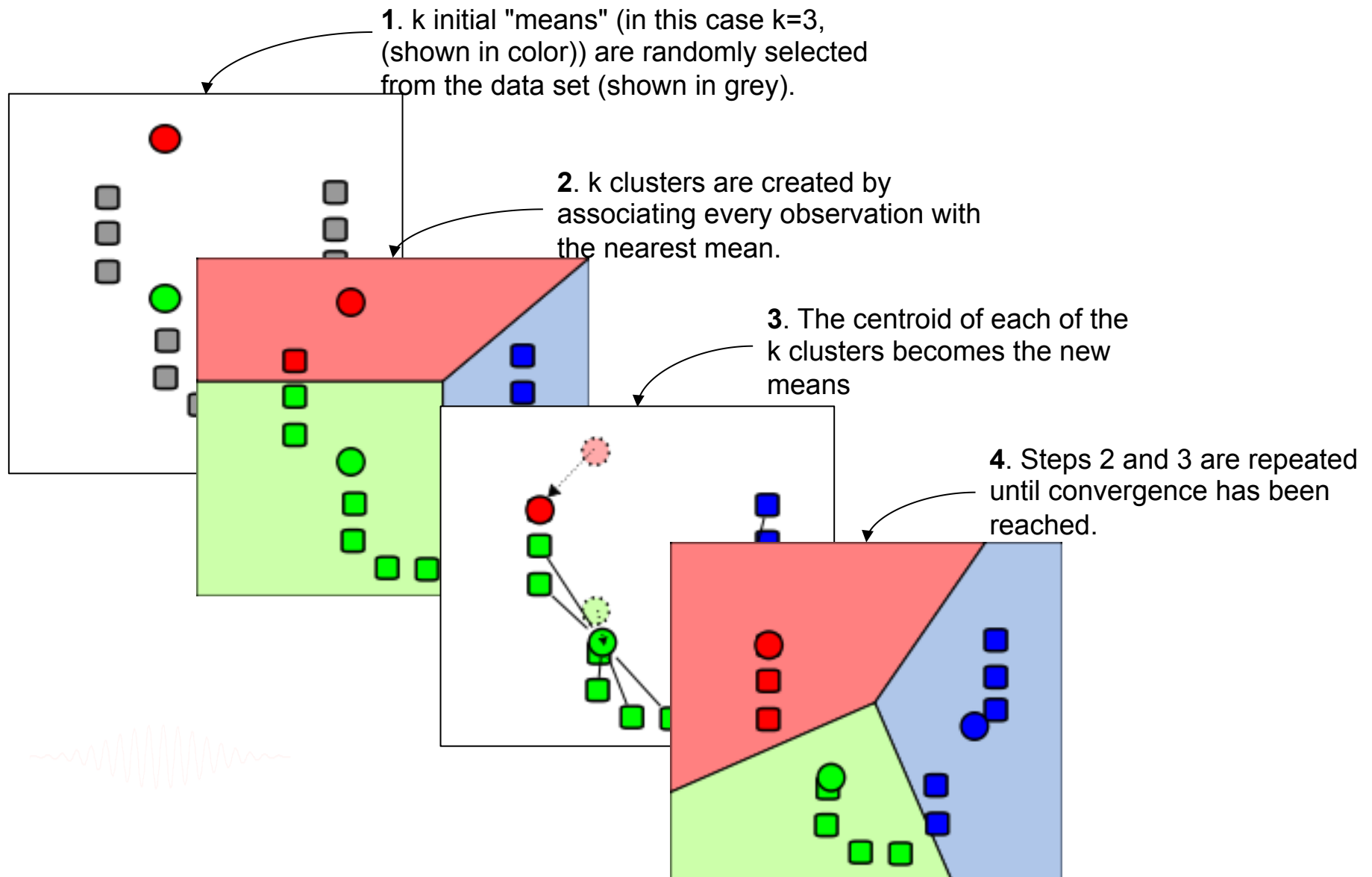
ICs (all subj)



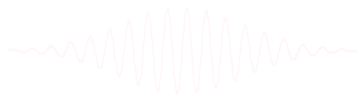
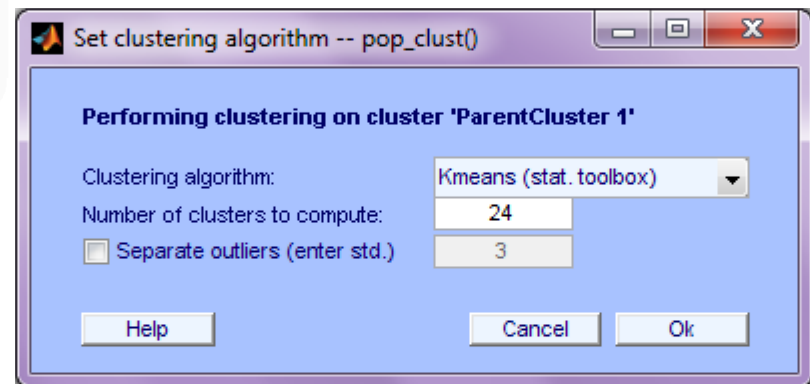
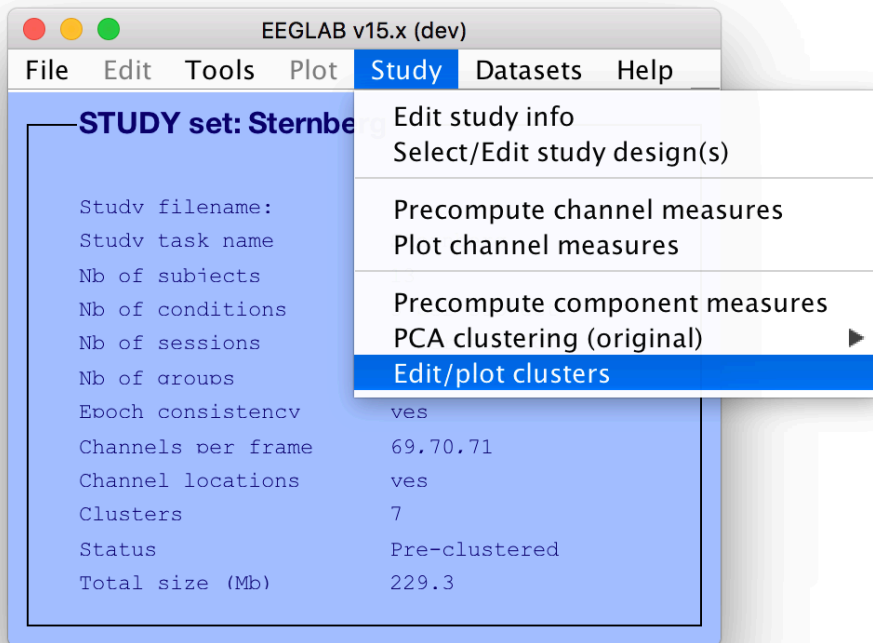
Each component is a dot
Clustering will group these
dots



Classical KMean



Cluster components



Choosing data measures

What measure(s) should you use?

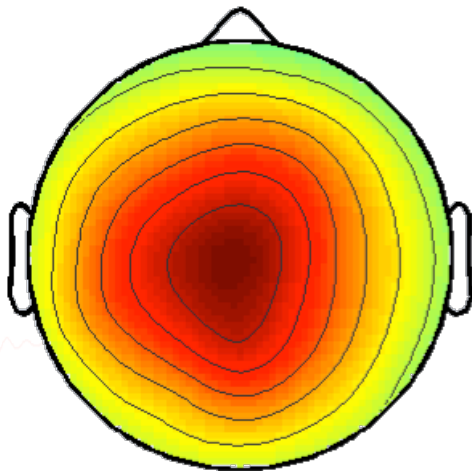
It depends on your final cluster criteria...

- If for example, your priority is dipole location, then cluster only based on dipole location...

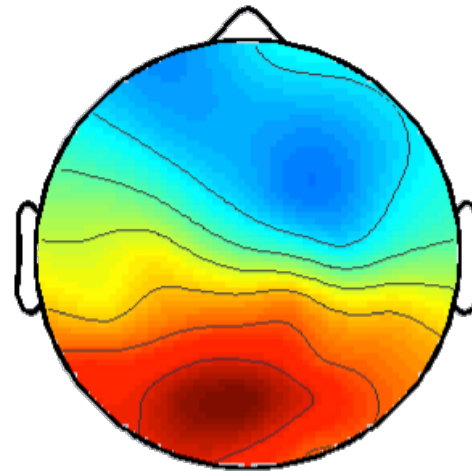
But consider:

- What is the difference between these two components?

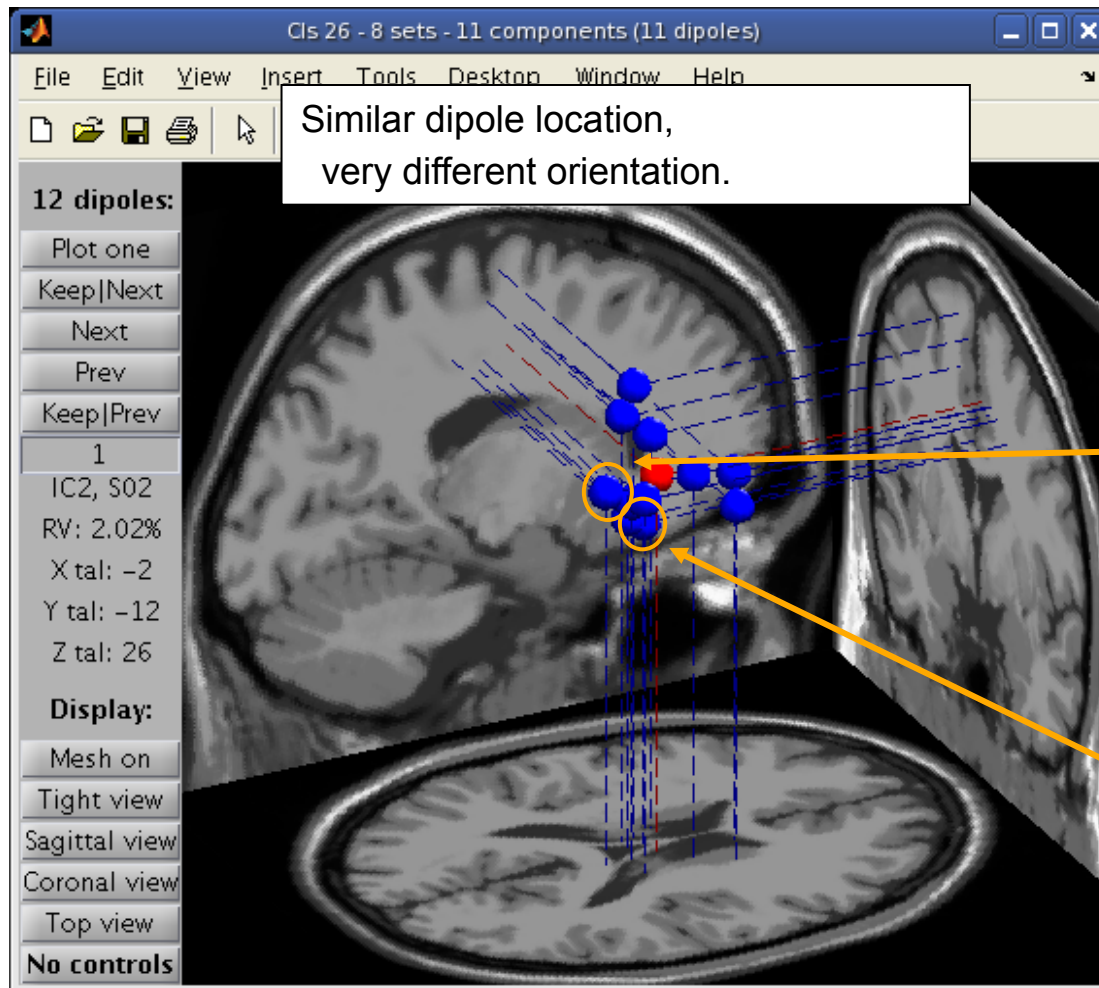
IC2 / S02, Cls 26



IC5 / S05, Cls 26

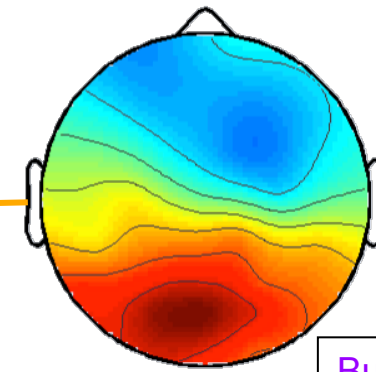


Choosing data measures

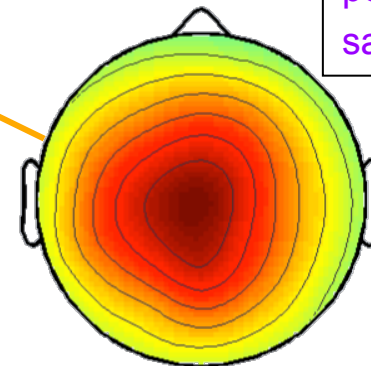


Obvious dramatic effect on scalp map topography:

IC5 / S05, Cls 26

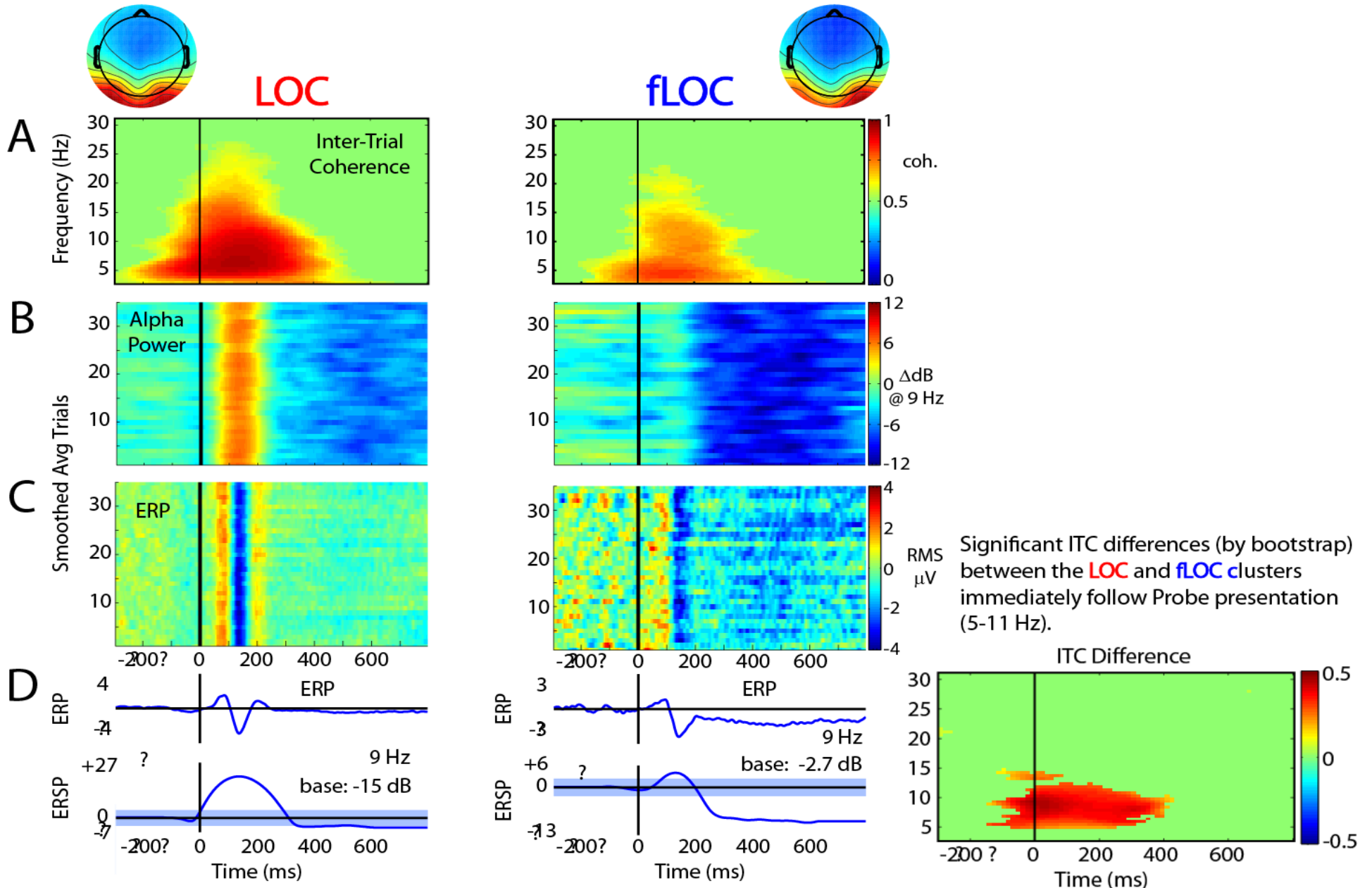


IC2 / S02, Cls 26

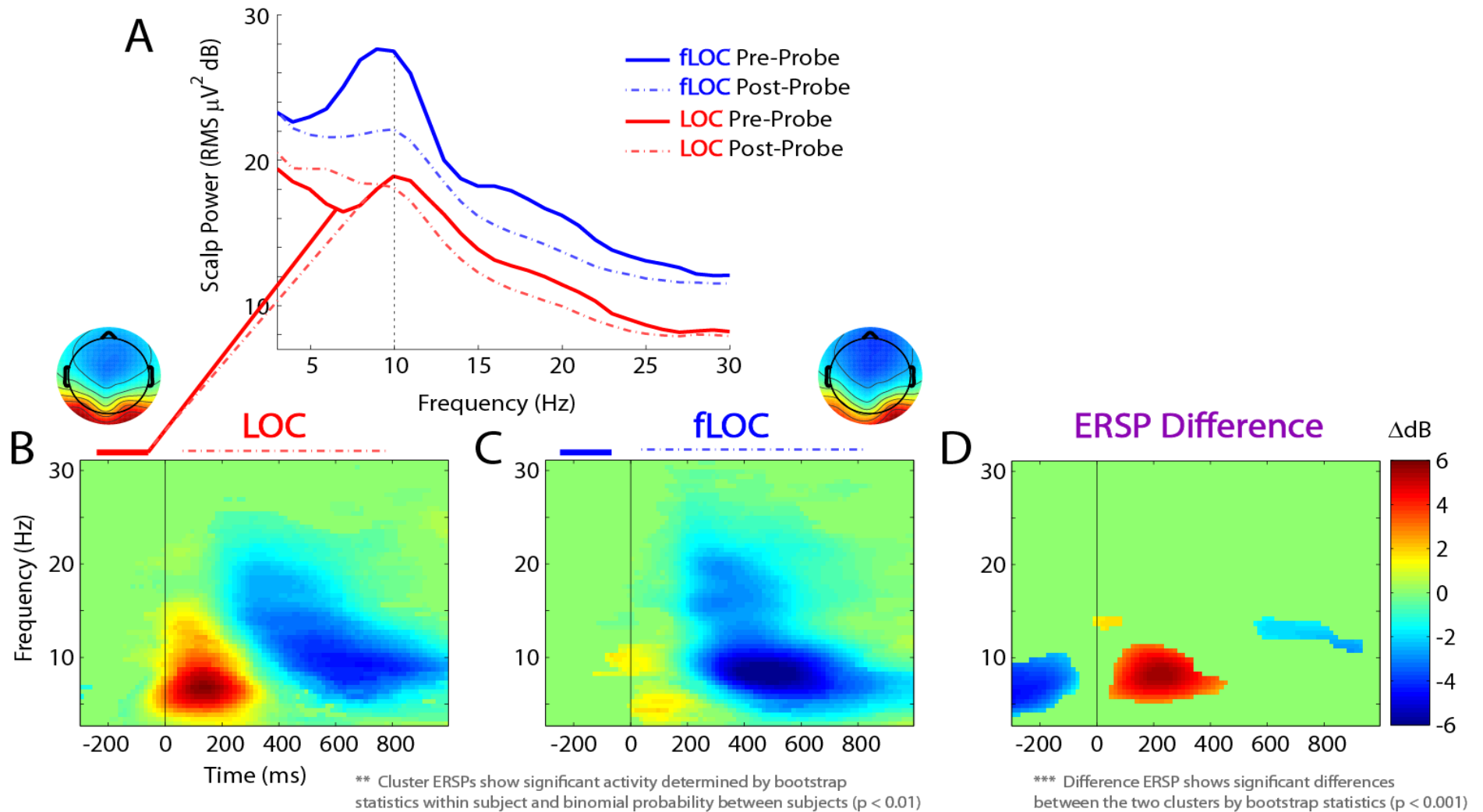


But, do they perform the same functions?

Subject differences?



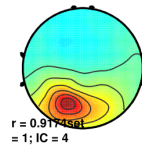
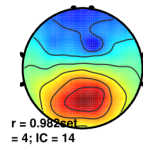
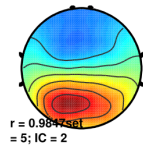
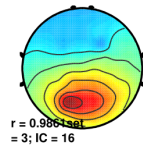
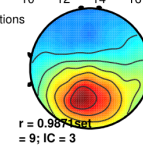
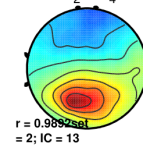
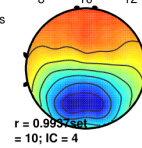
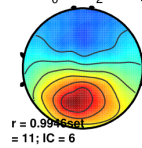
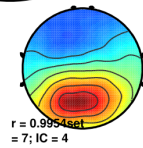
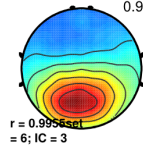
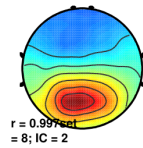
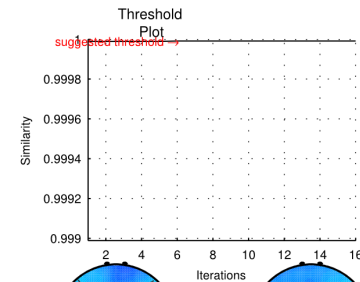
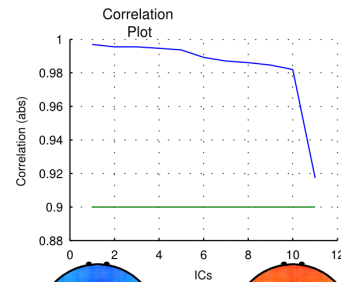
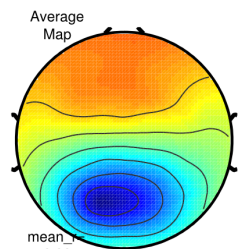
Subject differences?



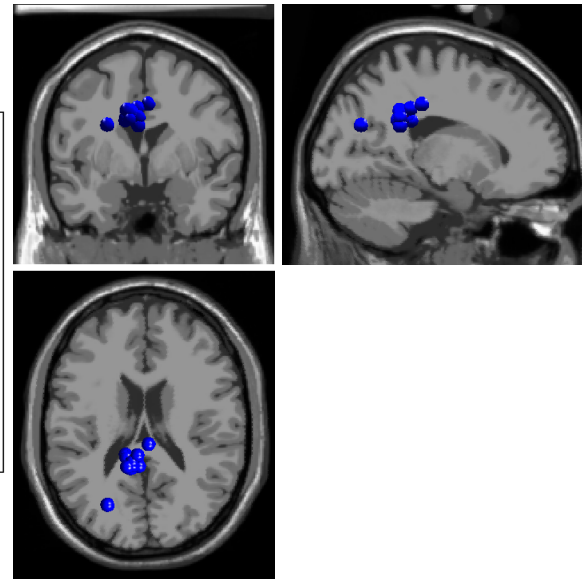
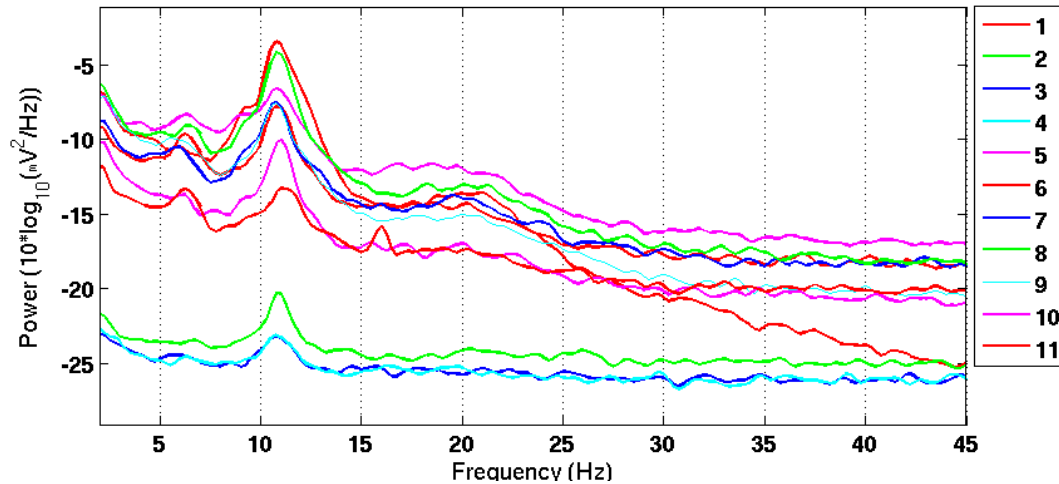
Results (Cluster 1)

100 % Sessions contribute

INFO:
Template: CB Session 7 PREPROC:STEP 2; Set 7; IC 3;
Number of datasets: 11
Correlation threshold: 0.9 (green line)
Max ICs from each dataset: 1
Cluster: 11 ICs from 11 sets
All datasets contribute.
Similarity = 1.0000



Cls 3 Spectrum



Results (Cluster 2)

100 % Sessions contribute

INFO:

Template: CB Session 5 PREPROC:STEP 2; Set 5; IC 1;

Number of datasets: 11

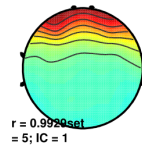
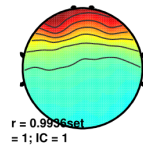
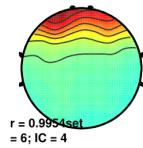
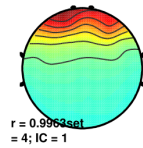
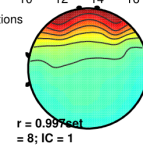
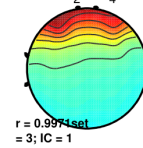
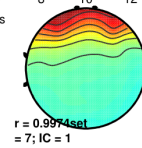
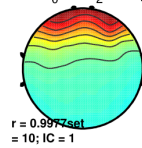
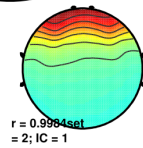
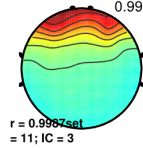
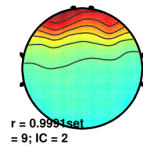
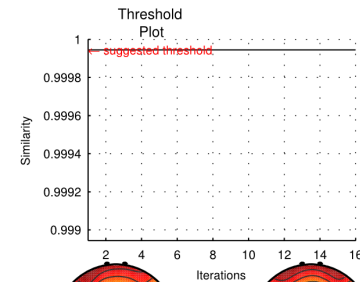
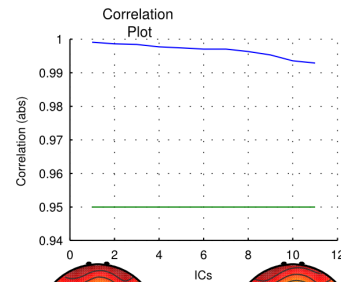
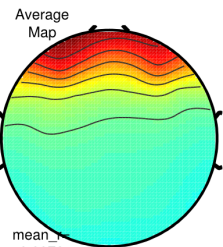
Correlation threshold: 0.95 (green line)

Max ICs from each dataset: 1

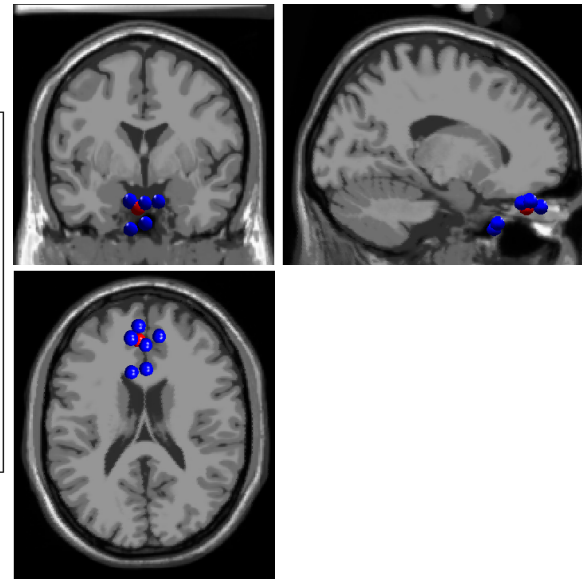
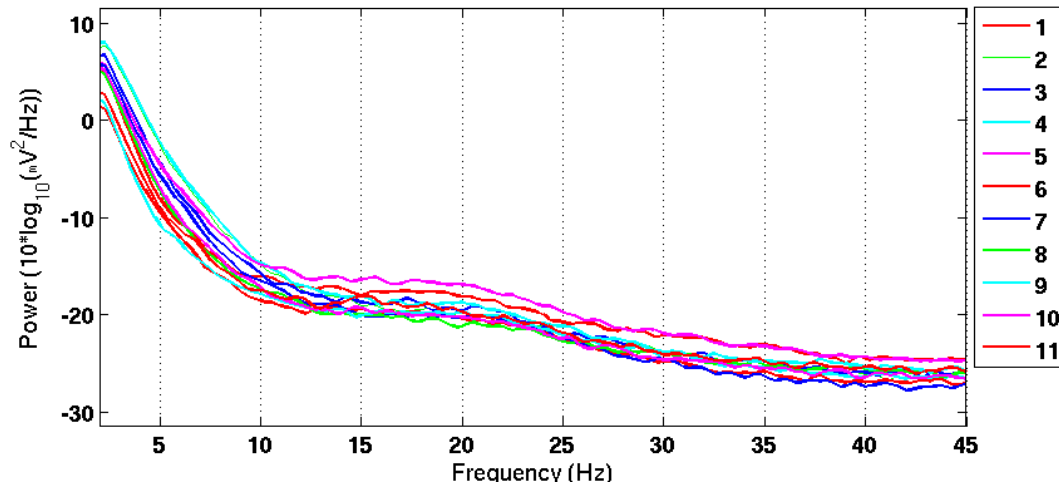
Cluster: 11 ICs from 11 sets

All datasets contribute.

Similarity = 0.9999



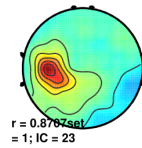
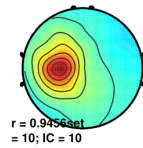
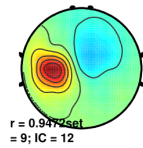
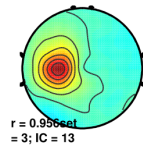
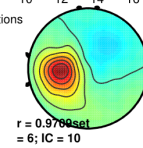
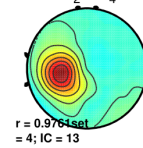
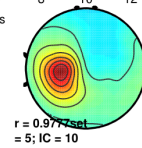
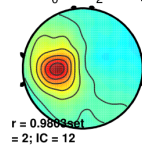
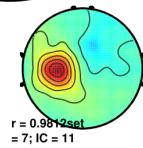
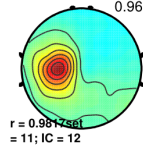
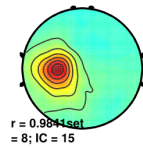
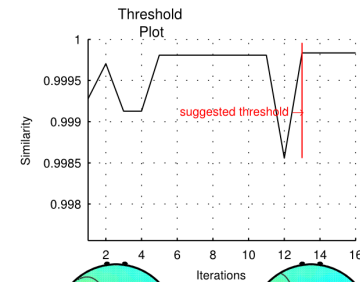
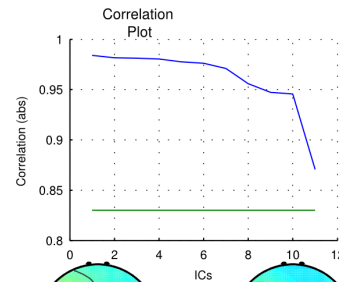
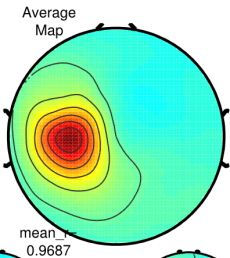
Cls 4 Spectrum



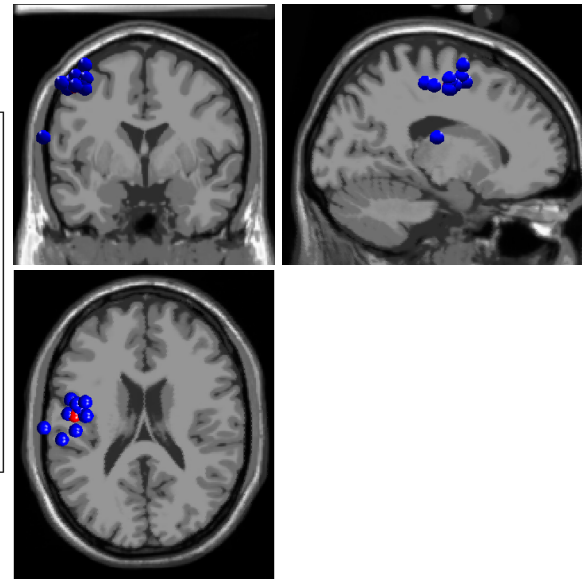
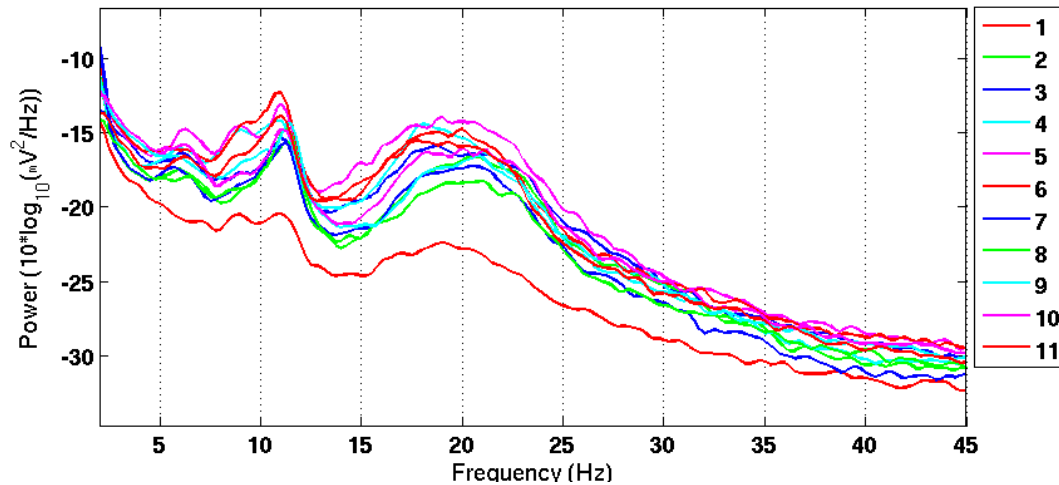
Results (Cluster 8)

100 % Sessions contribute

INFO:
Template: CB Session 7 PREPROC:STEP 2; Set 7; IC 11;
Number of datasets: 11
Correlation threshold: 0.83 (green line)
Max ICs from each dataset: 1
Cluster: 11 ICs from 11 sets
All datasets contribute.
Similarity = 0.9998

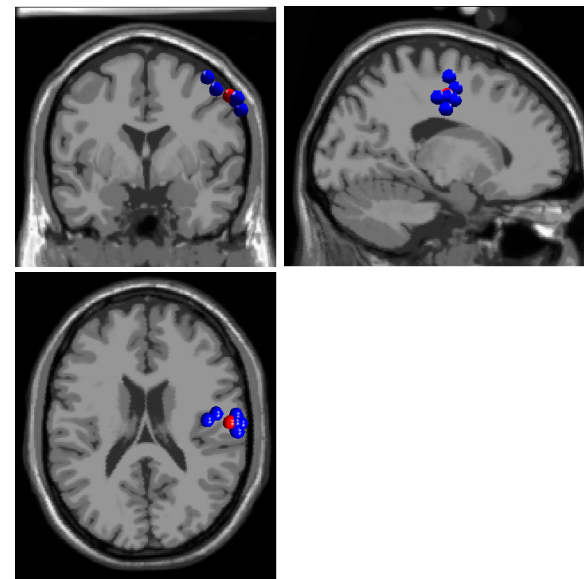
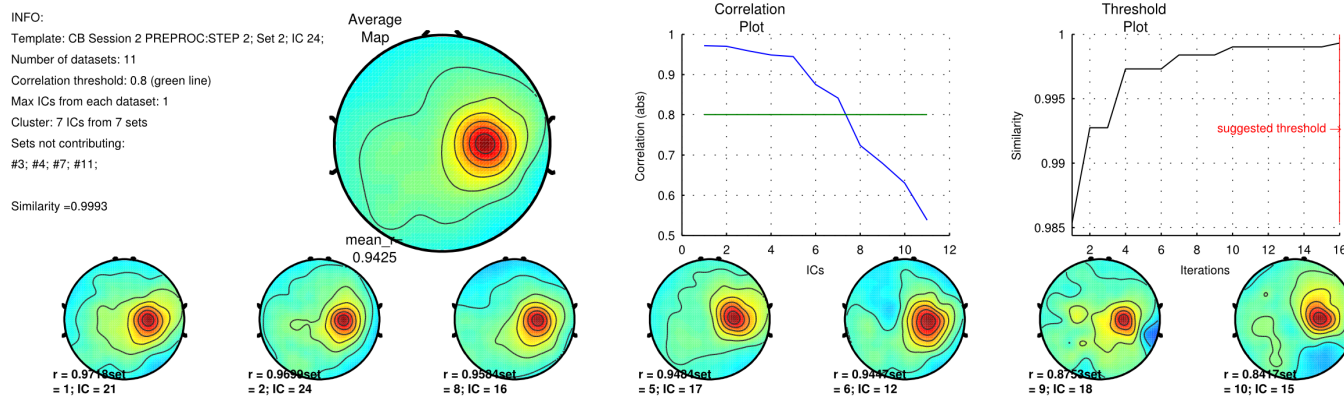


Cls 8 Spectrum



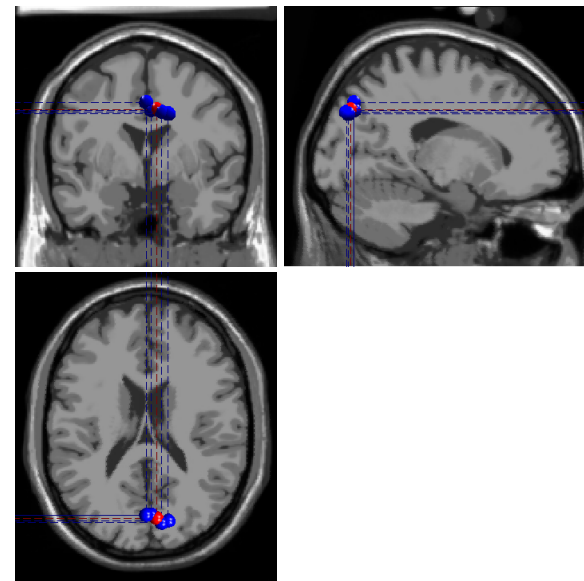
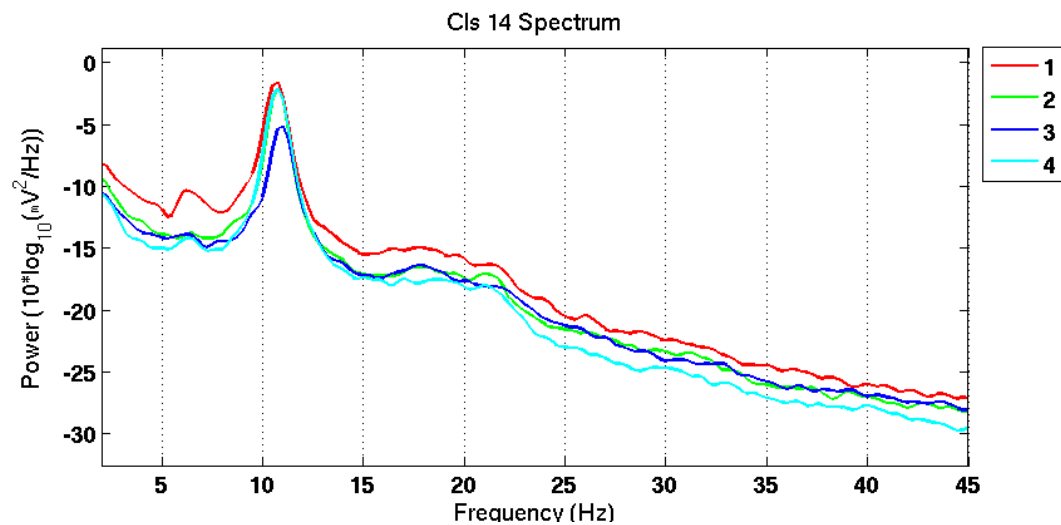
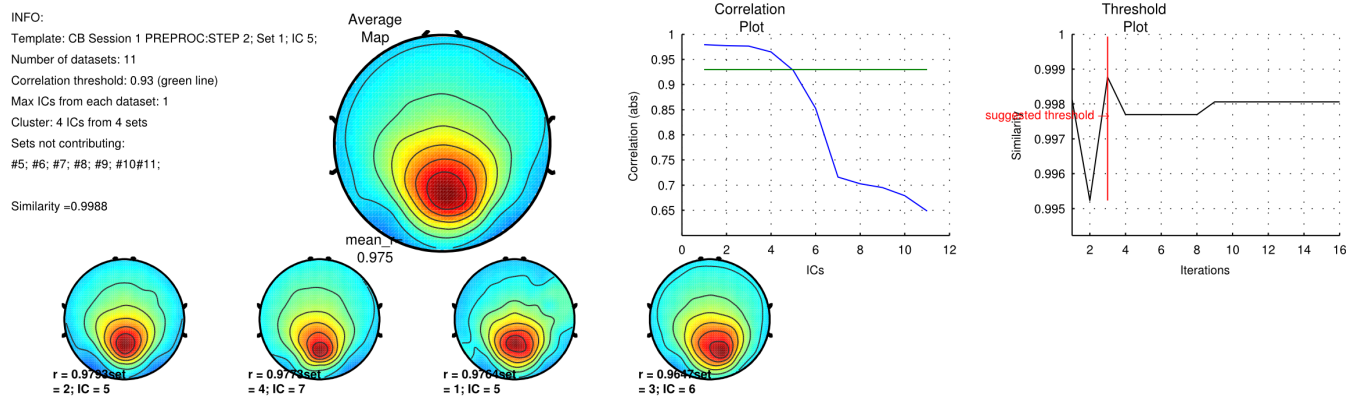
Results (Cluster 13)

63.64% Sessions contribute

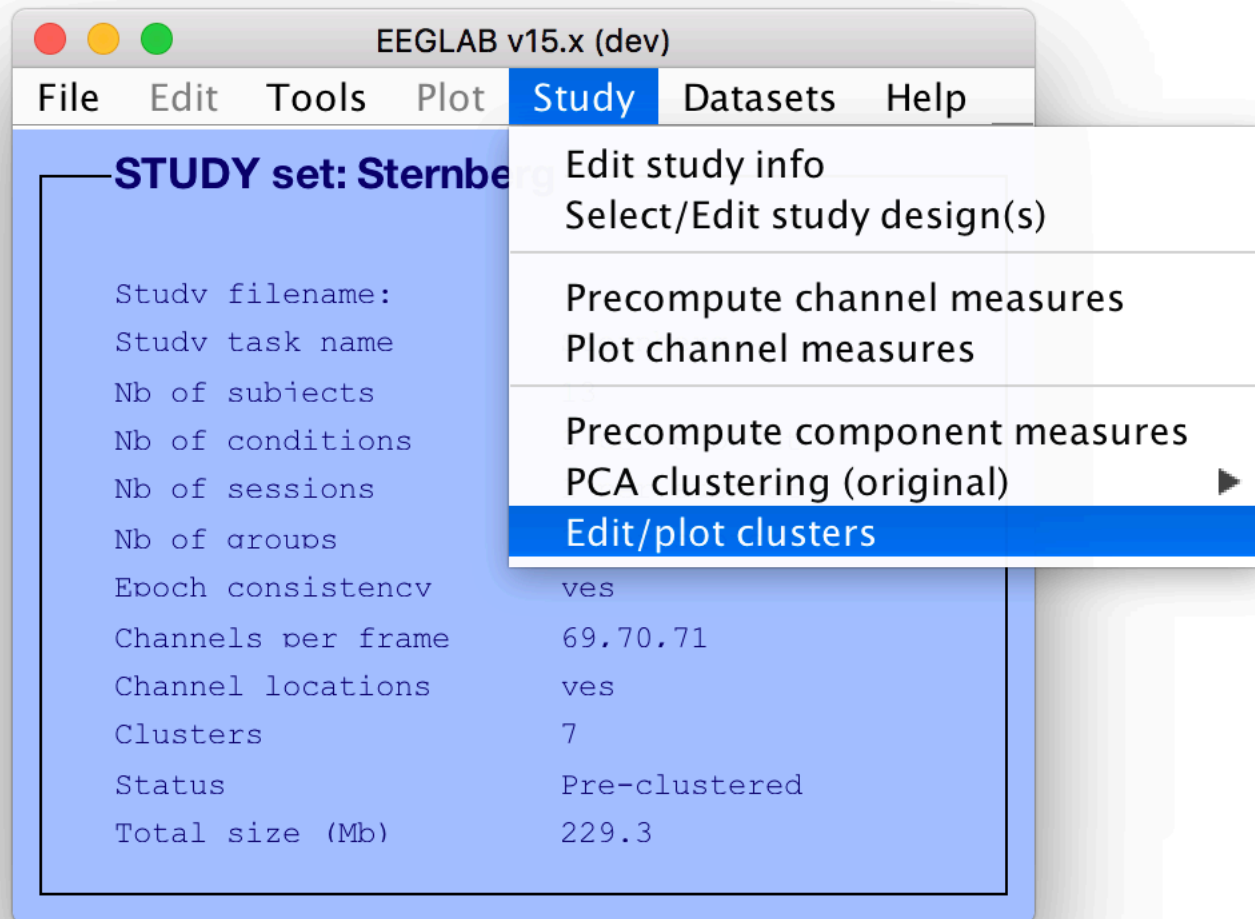


Results (Cluster 14)

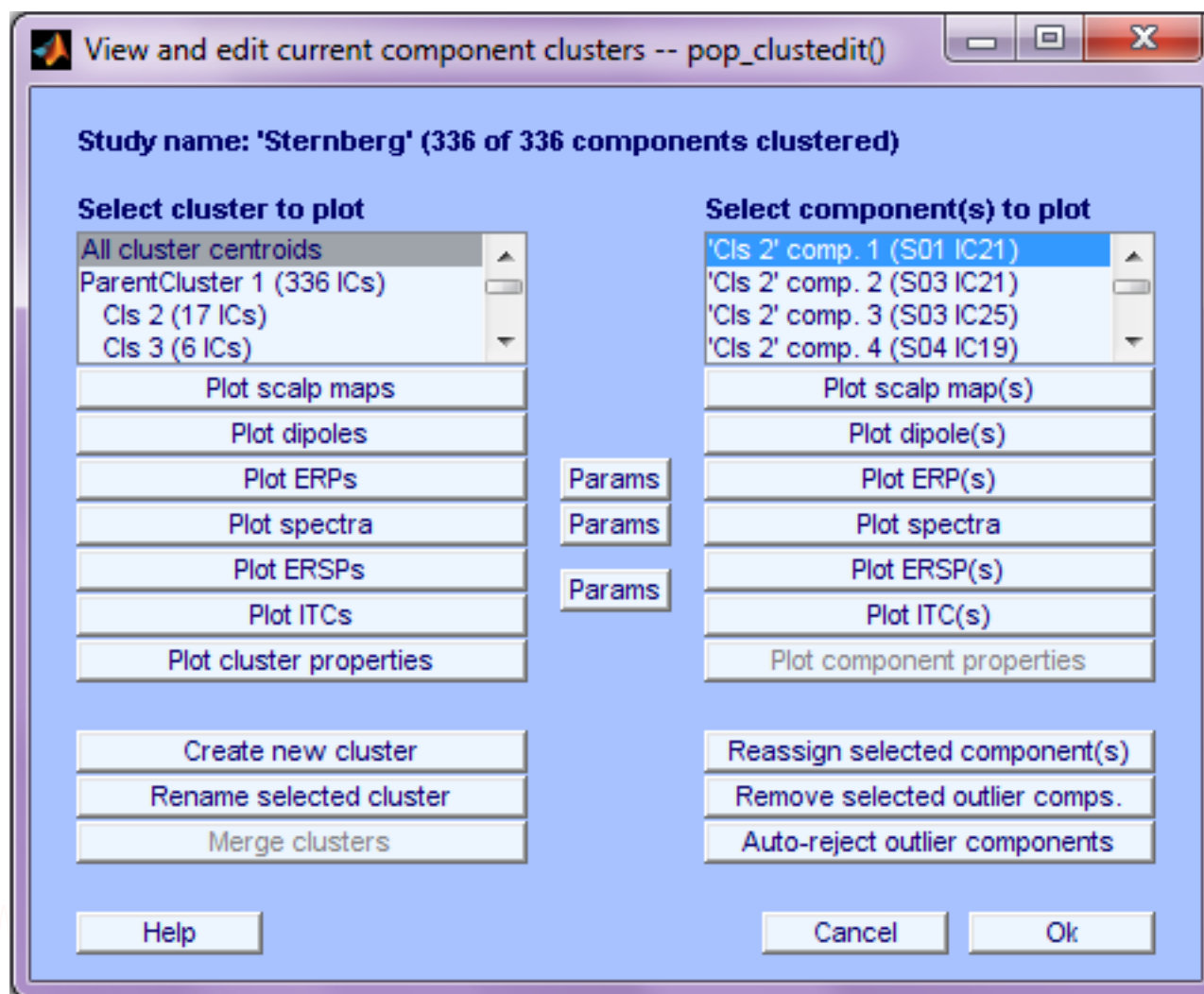
36.36% Sessions contribute



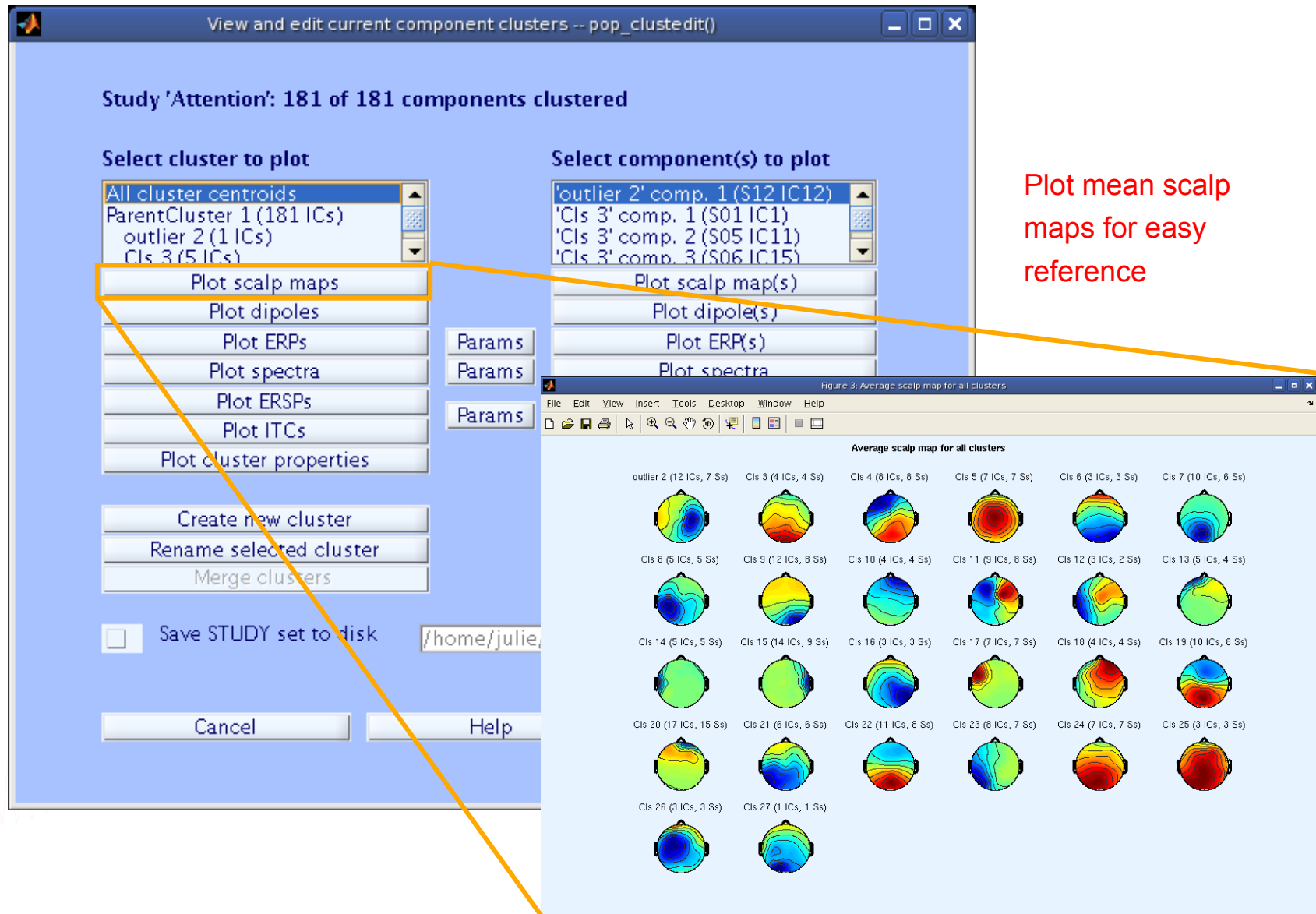
View and edit clusters



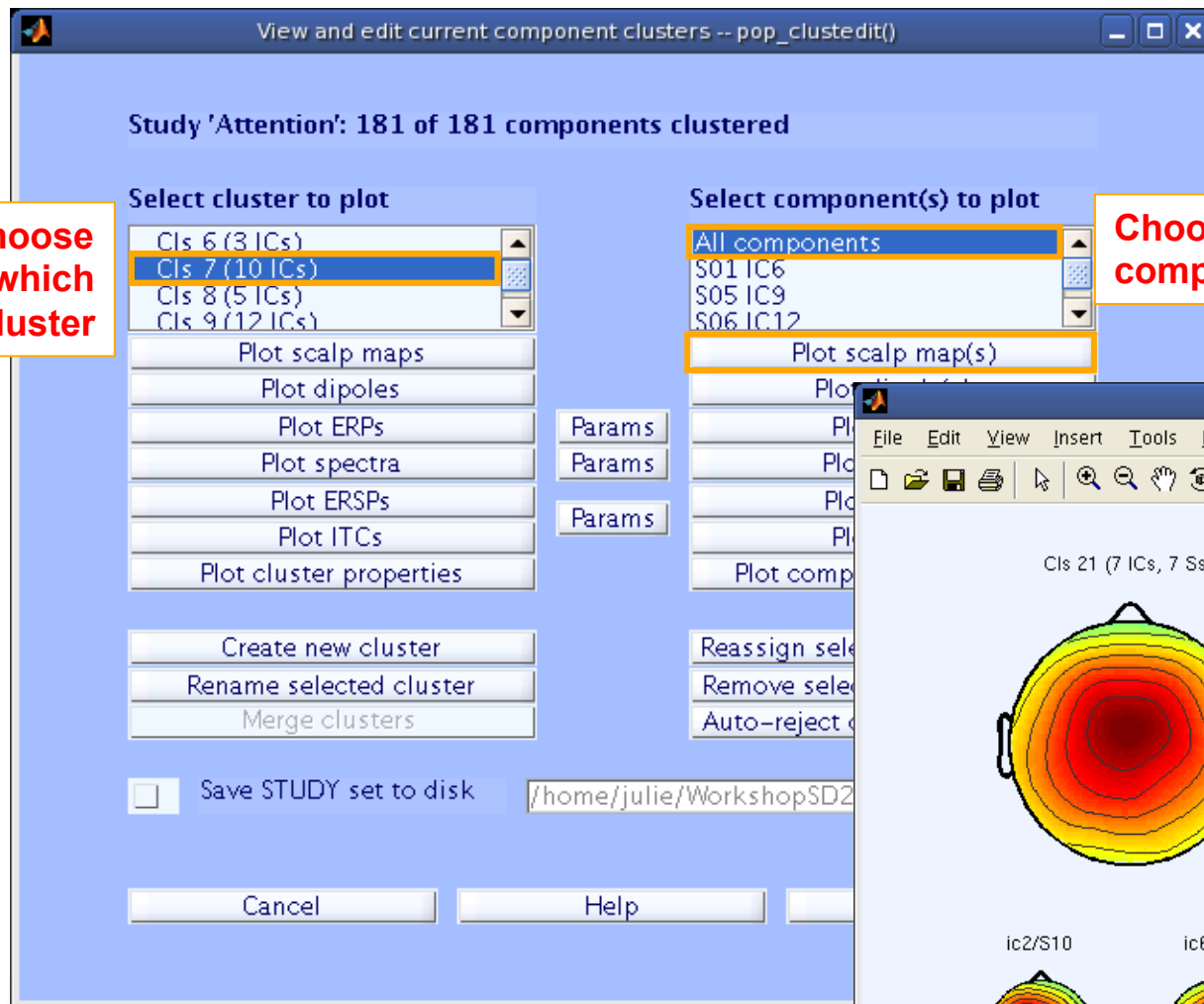
Plot/edit clusters



Plot cluster data

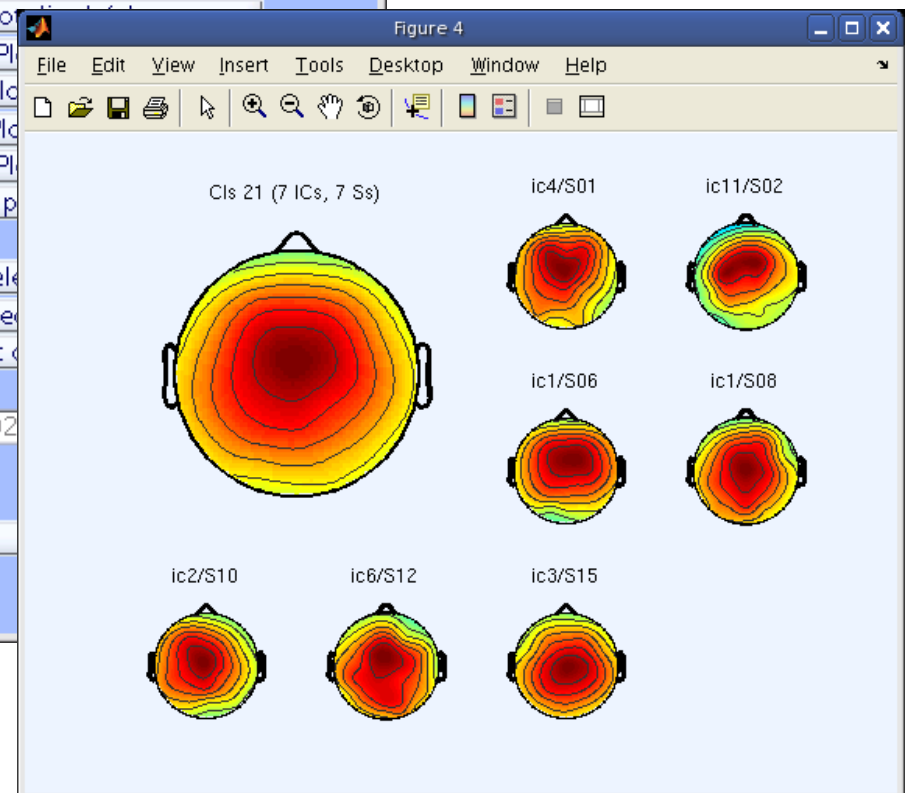


Plot cluster data

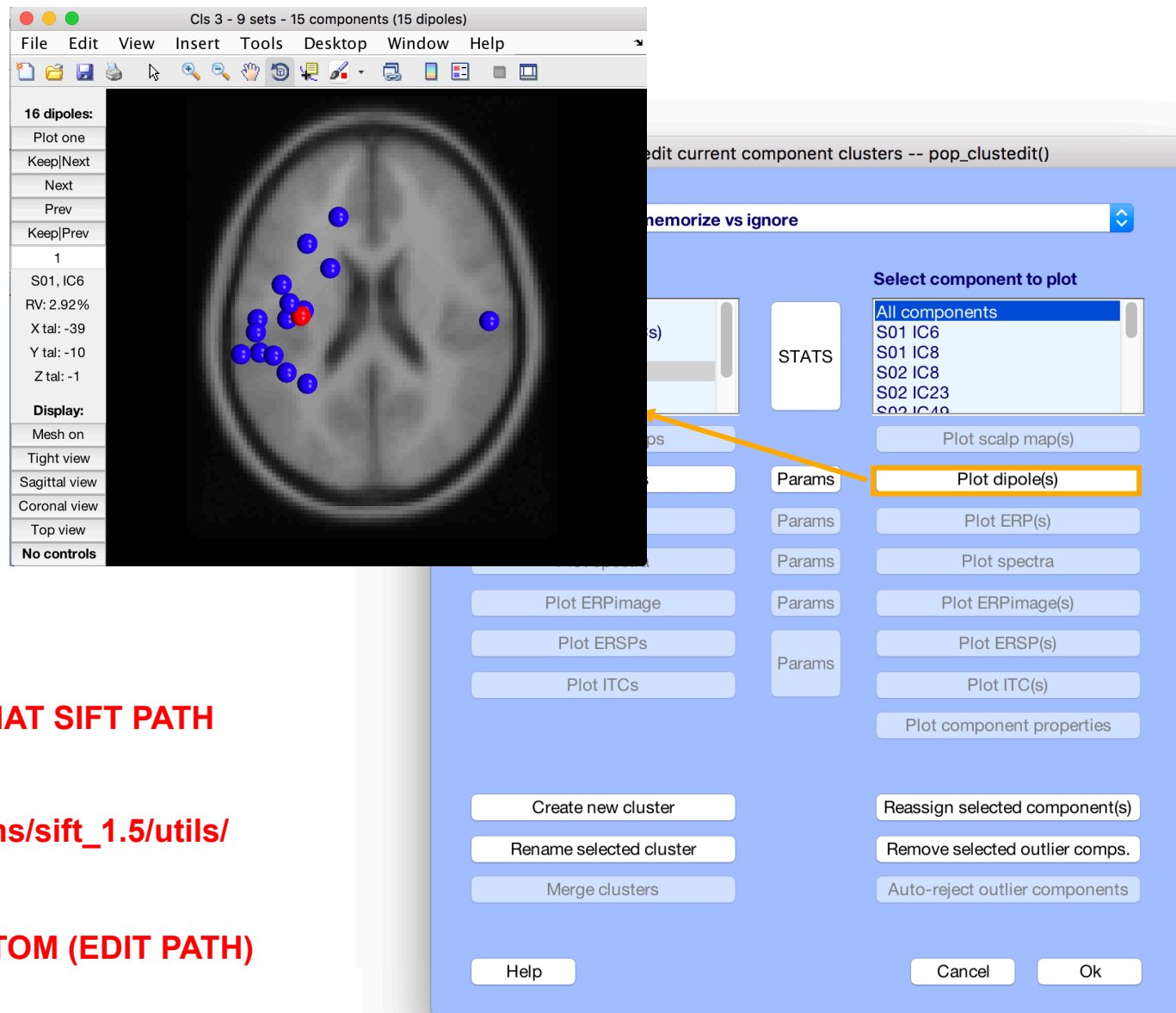


Choose
which
cluster

Choose which
components



Plot cluster data

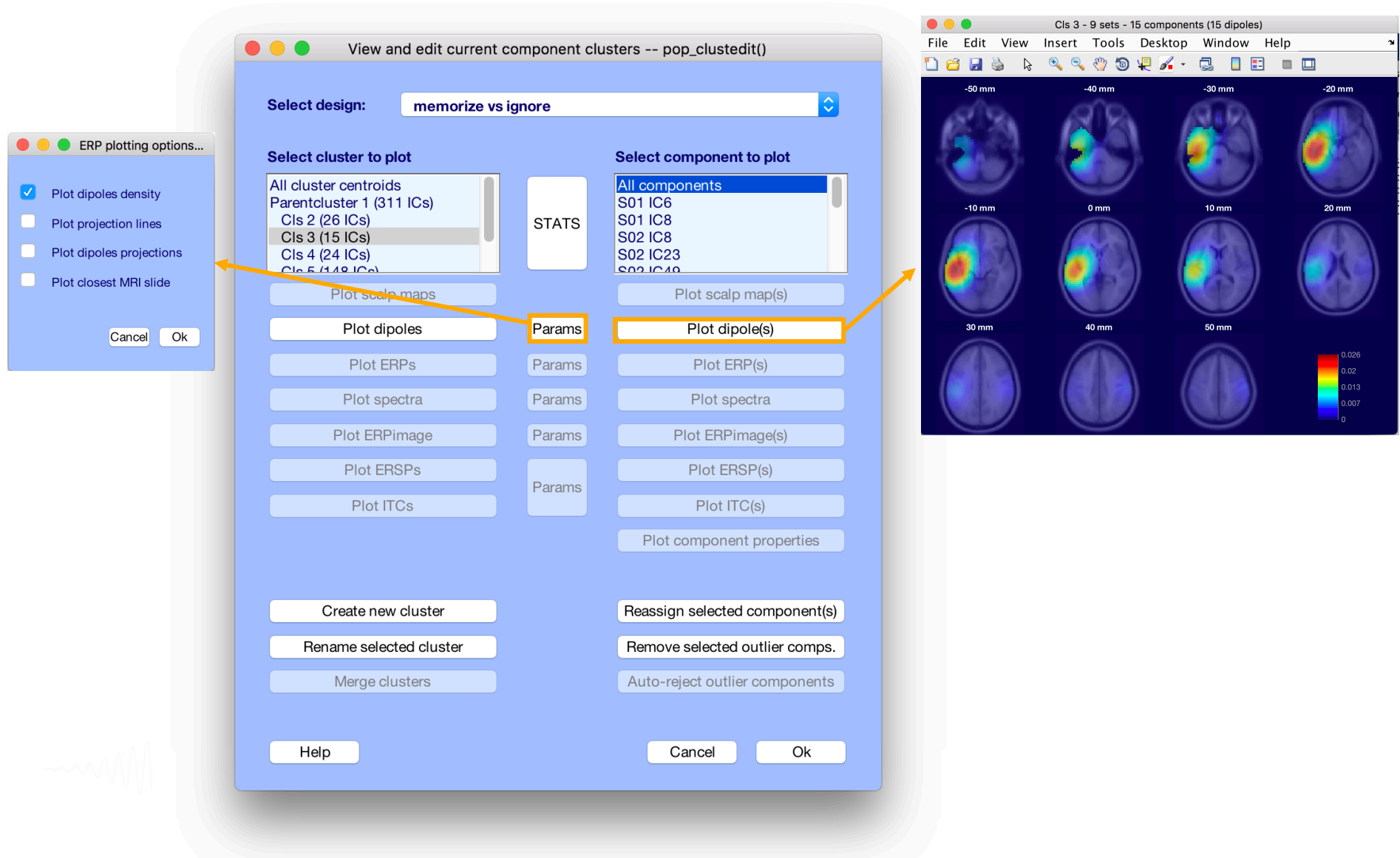


MAKE SURE THAT SIFT PATH

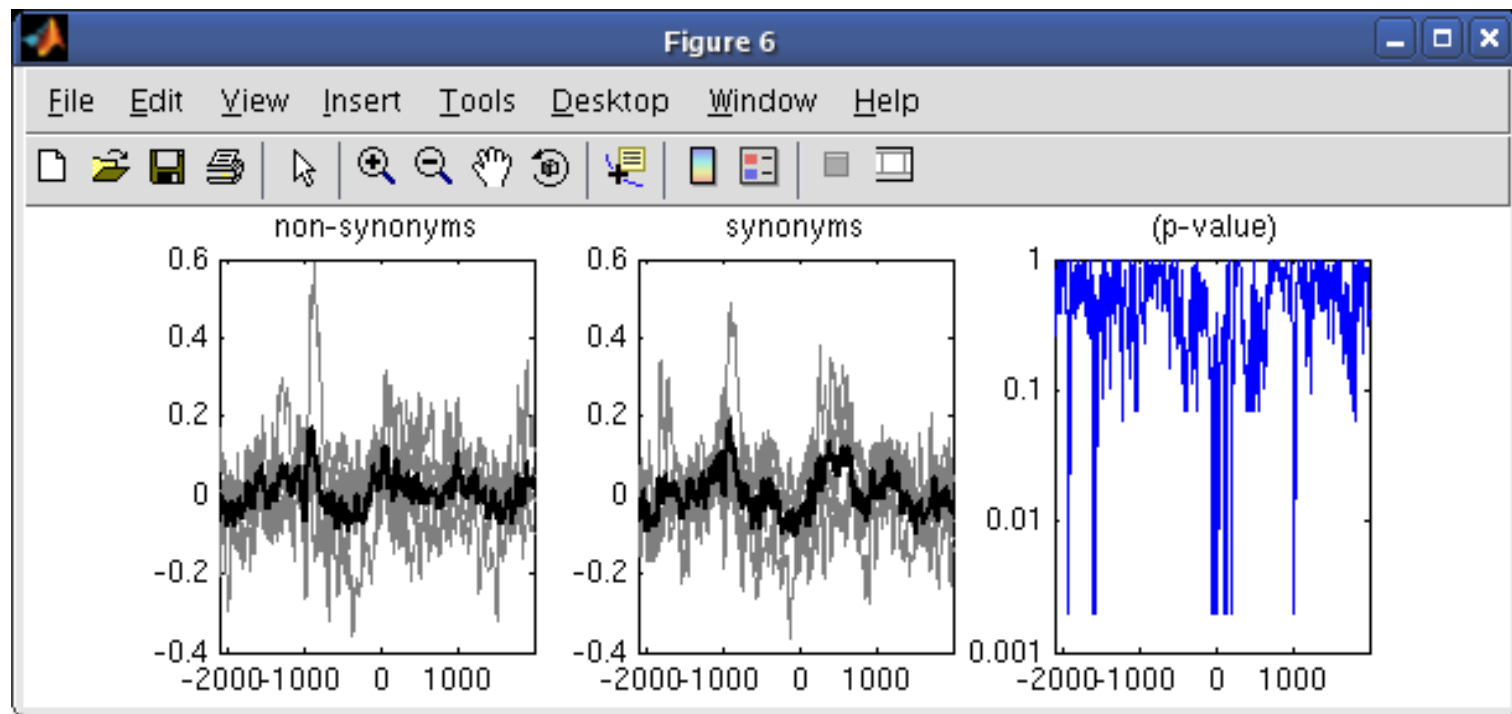
.../eeglab/plugins/sift_1.5/utis/

IS AT THE BOTTOM (EDIT PATH)

Plot cluster data



Plot cluster ERP



Exercise

- Load the STUDY stern.study
- Precompute **spectrum**, **ERP** and **scalp maps** for components
- Precluster and cluster components using **dipole locations** and **dipole moments** (affinity clustering)
- Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
- Remove outliers if any
- Plot significant difference (parametric statistics) for one component cluster spectrum between the two conditions ignore vs memorize

