Clustering of ICA components

Romain Grandchamp

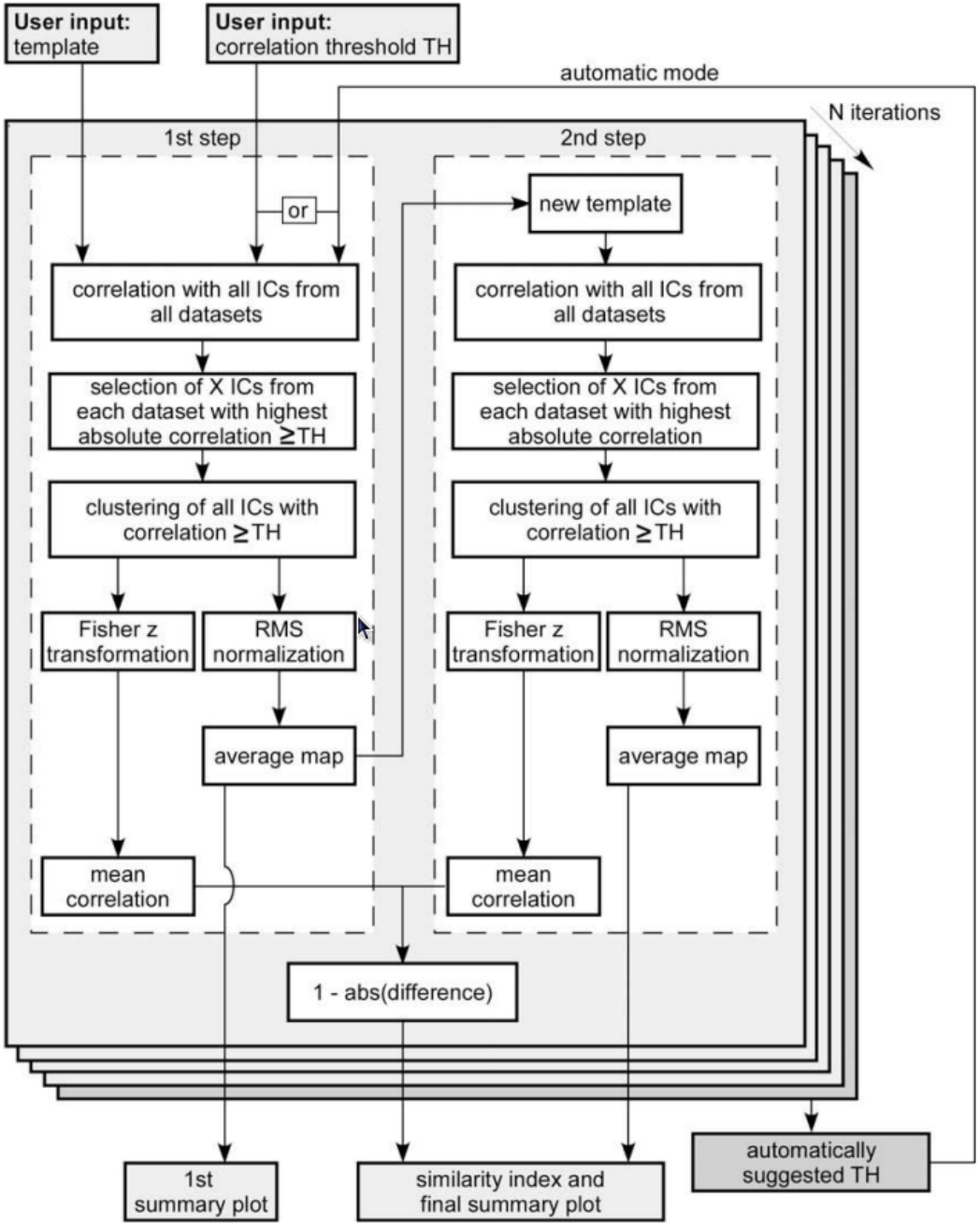
(with Arnaud Delorme, Julie Onton, Nima Bigdely Shamlo, Scott Makeig)

Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice

ICA decomposition of multiple data sets from the same individuals

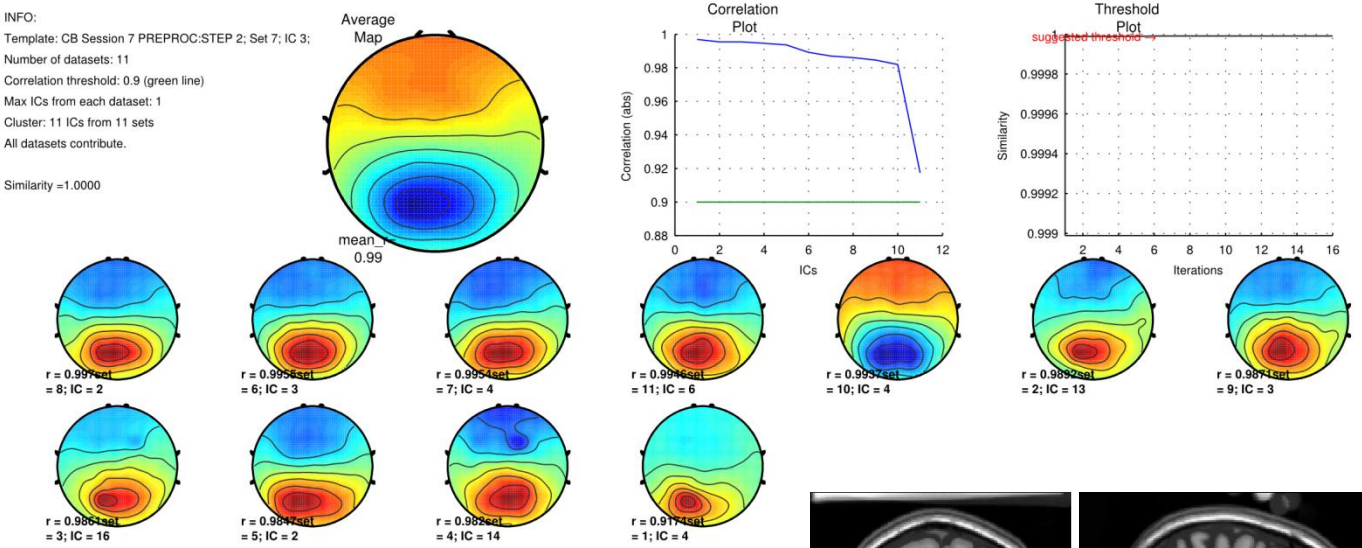
- Experimental protocol
  - Mind wandering experiment
  - 2 subjects
  - 11 x 30 min. sessions
  - 2 sessions per week
  - EEG from Biosemi 64 channels
  - Fs=1024 Hz

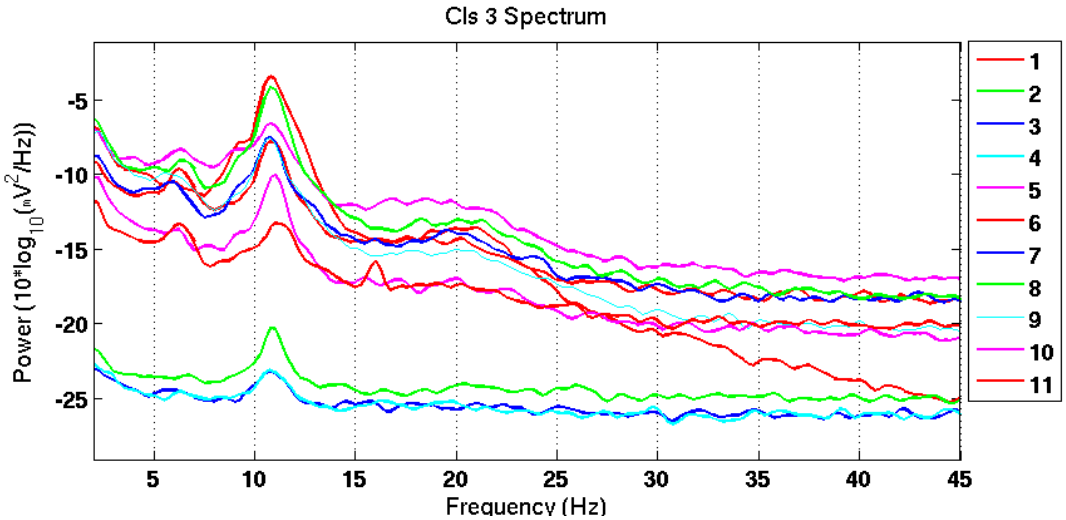


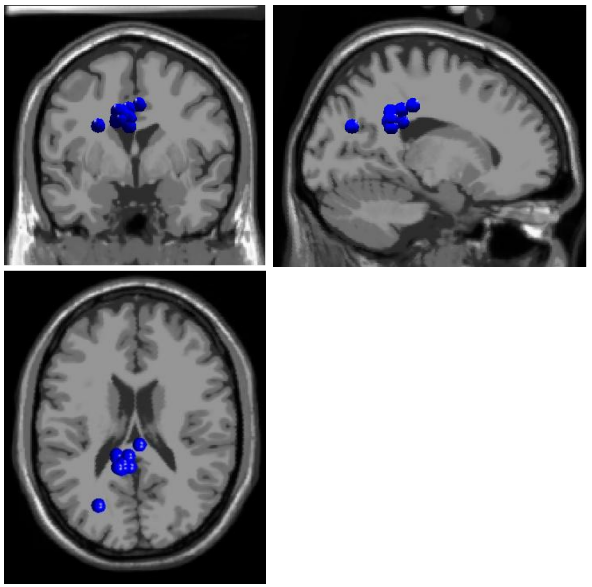
F. Campos Viola et al., “Semi-automatic identification of independent components representing EEG artifact,” Clinical Neurophysiology 120, no. 5 (2009): 868–877.

Results (Cluster 1)

100 % Sessions contribute

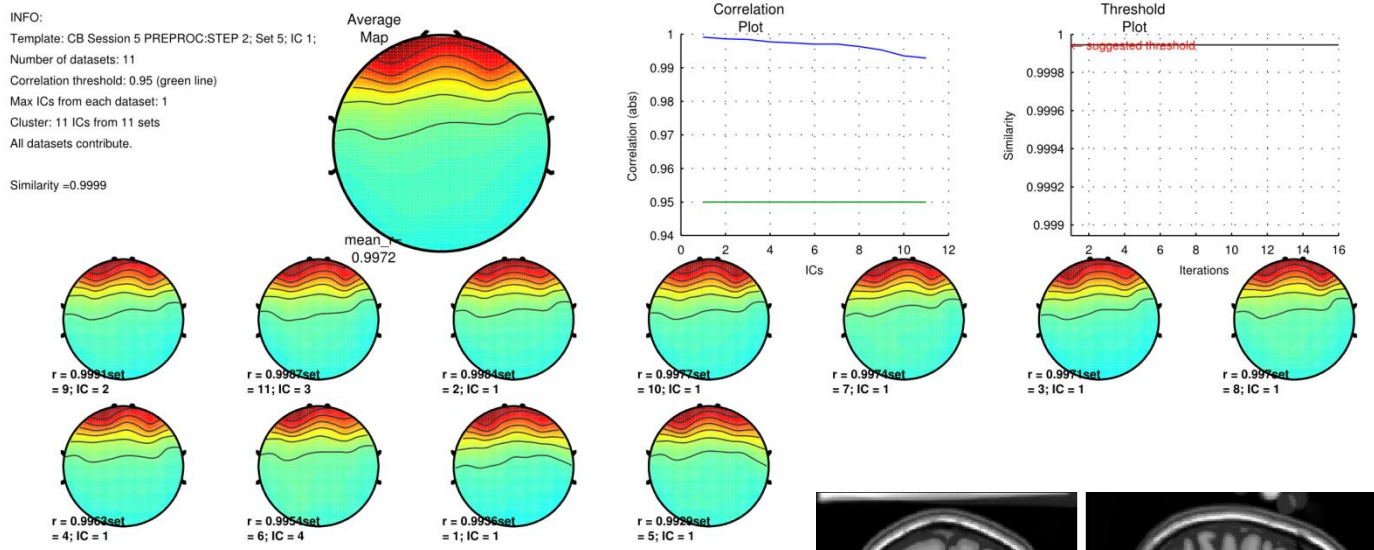


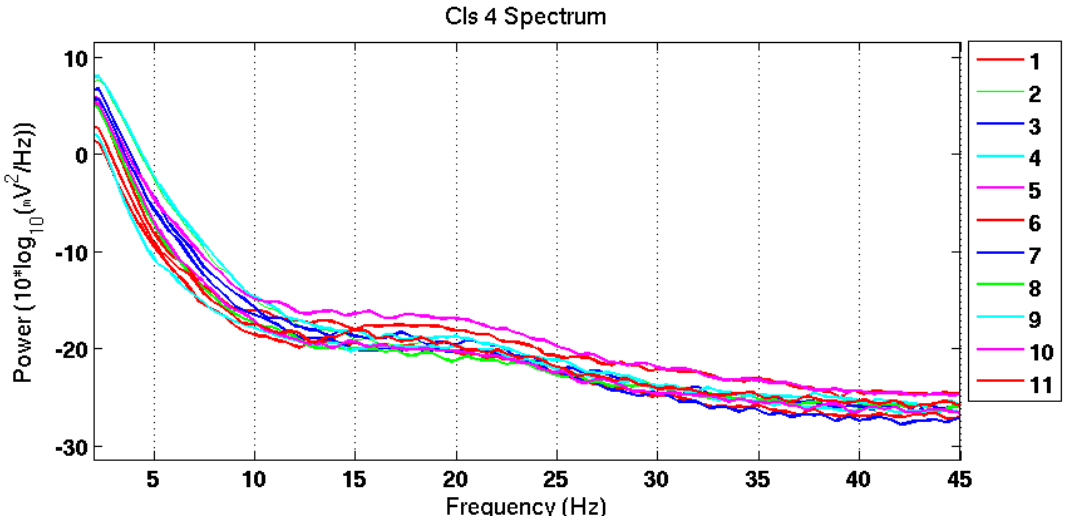


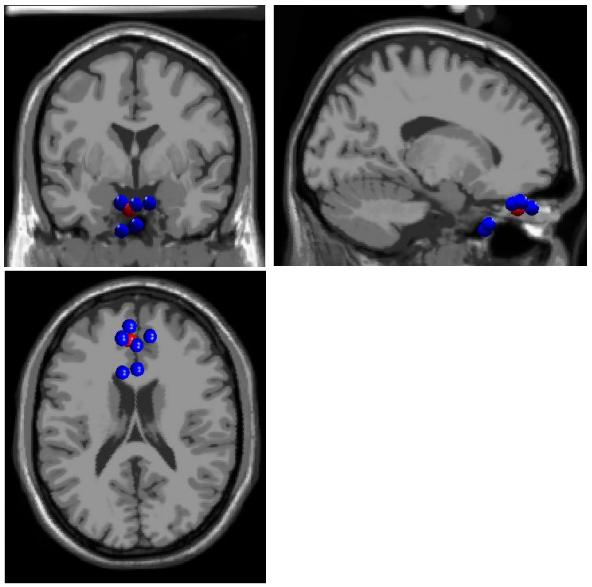


Results (Cluster 2)

100 % Sessions contribute

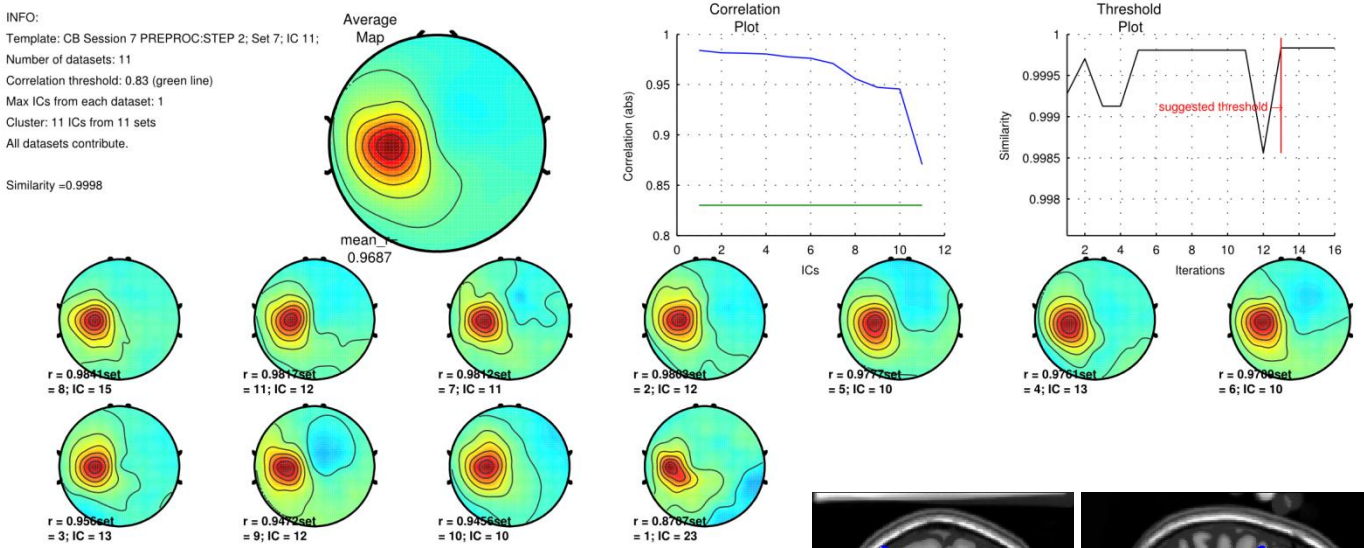


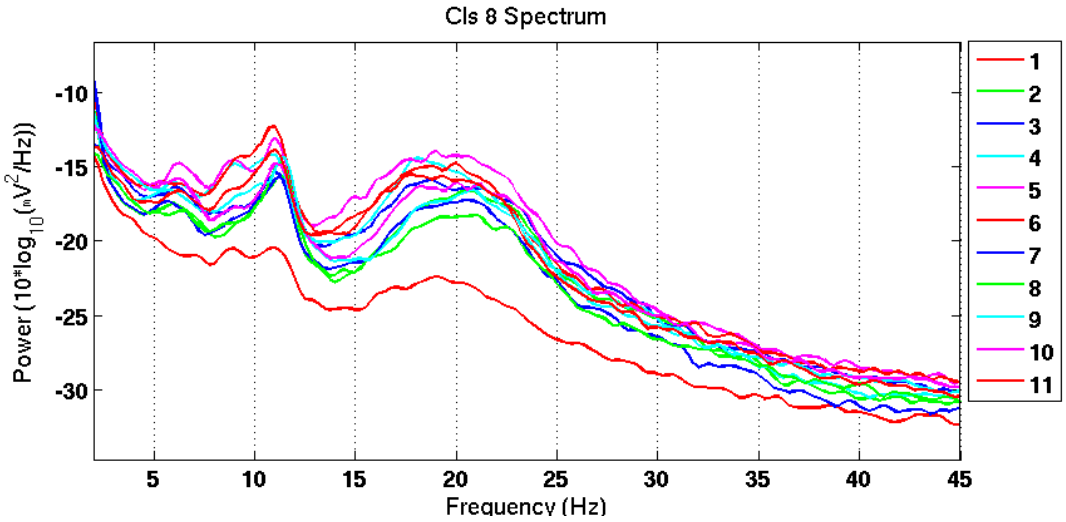


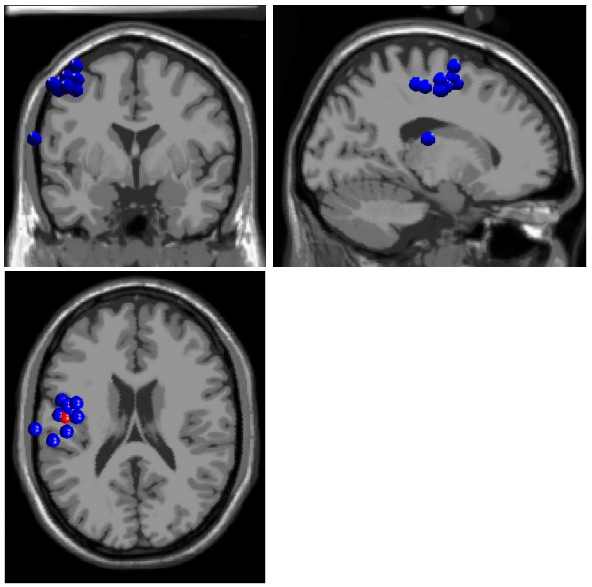


Results (Cluster 8)

100 % Sessions contribute

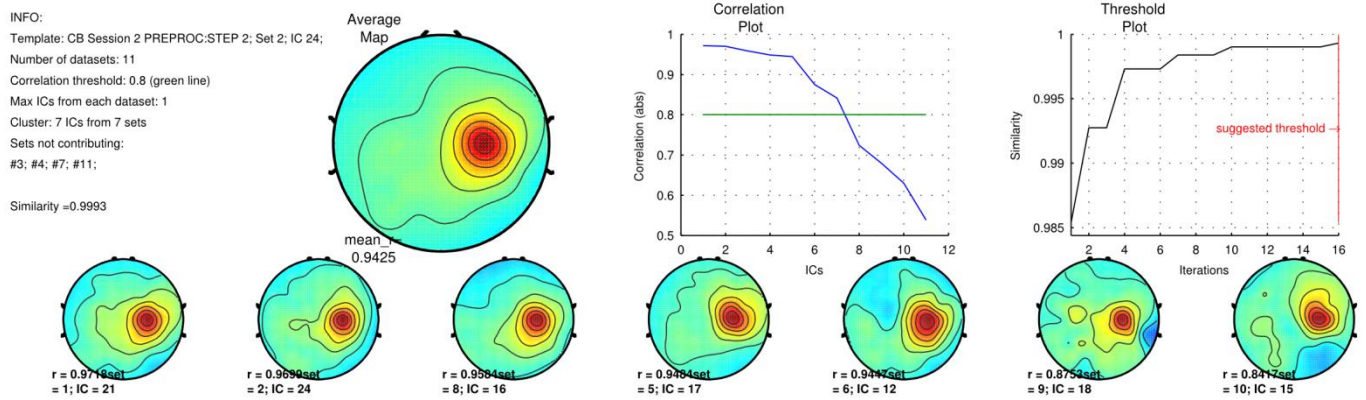


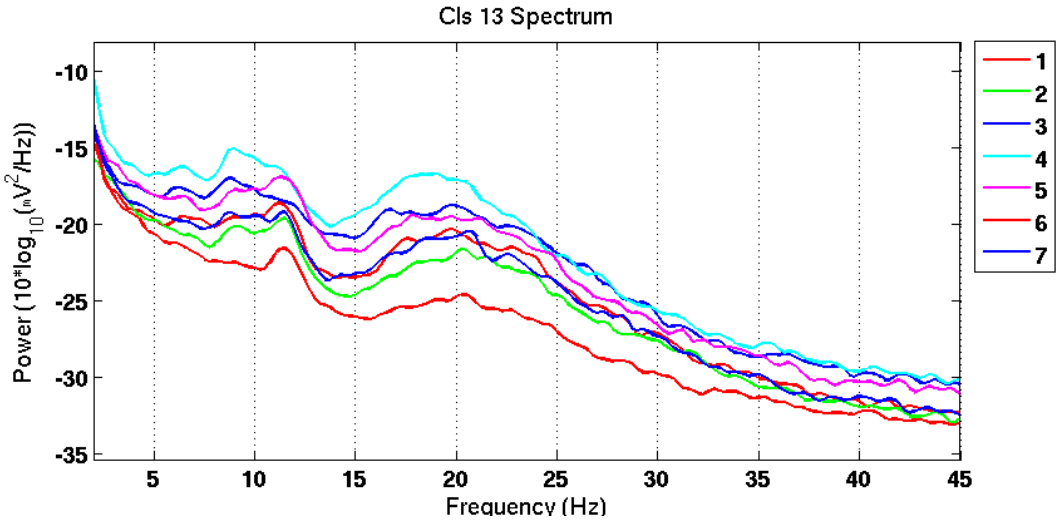


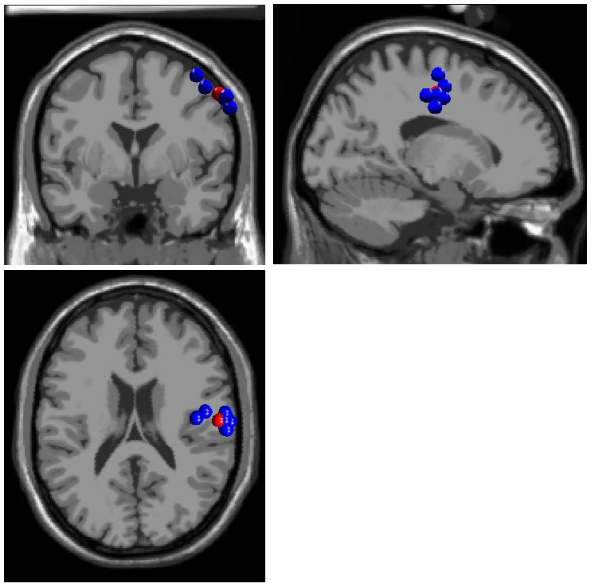


Results (Cluster 13)

63.64% Sessions contribute

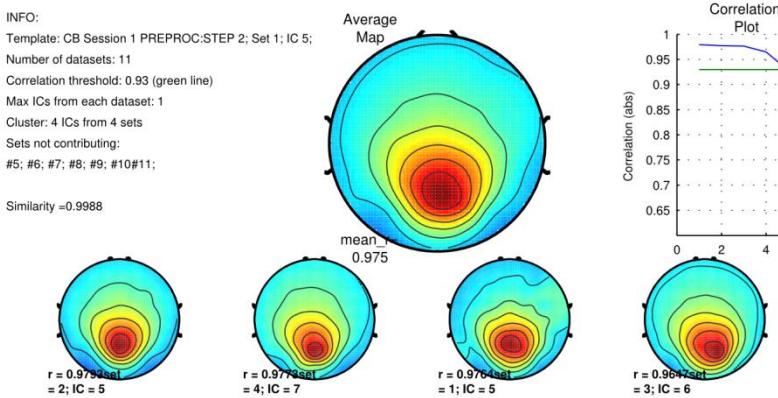


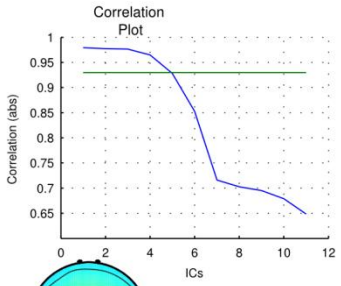


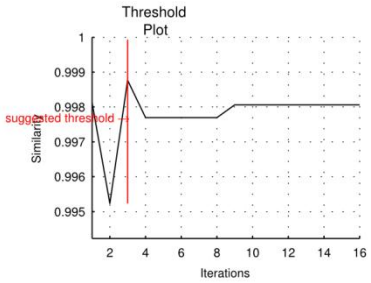


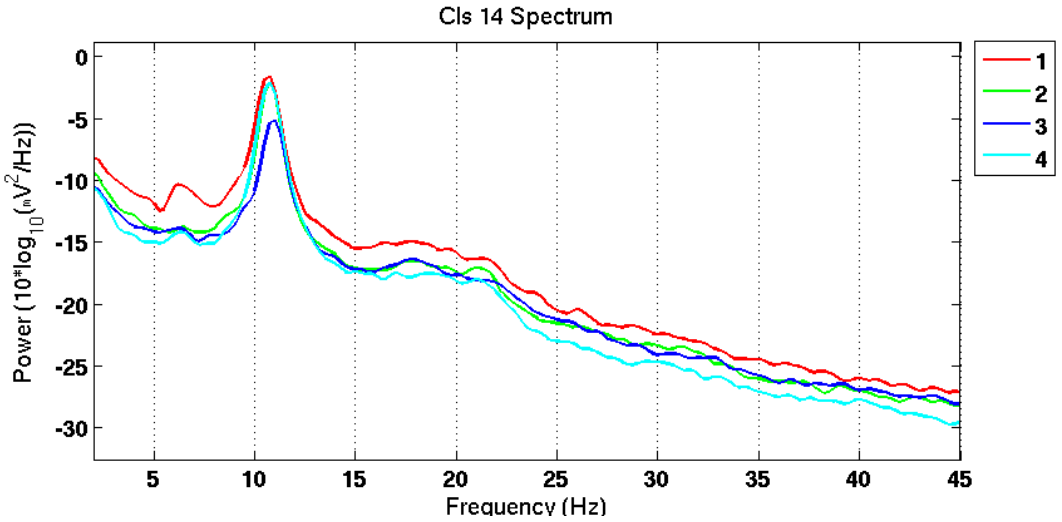
Results (Cluster 14)

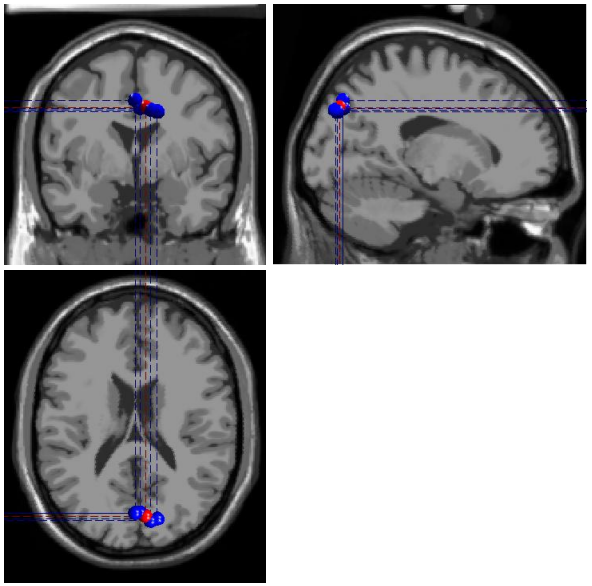
36.36% Sessions contribute











Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice

Validation of the ICA algorithm for EEG

Data

- 13 subjects performing a memory task
- 71 electrodes including EOGs
- more than 300,000 data points/subject

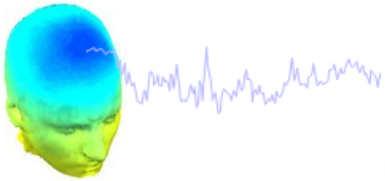
Decomposition

- 23 ICA algorithms plus PCA and Promax

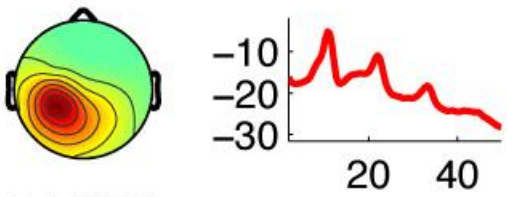
Analysis

- Localization of all components with a single dipole (4-shell spherical model)

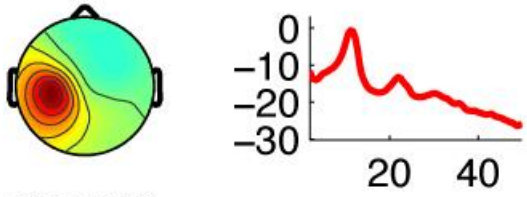
Left μ cluster



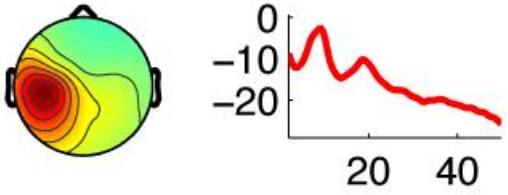
S2 IC47



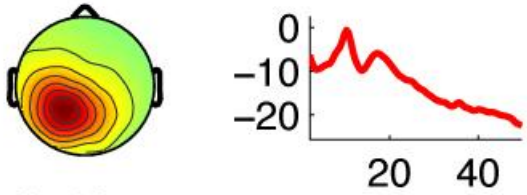
S3 IC47



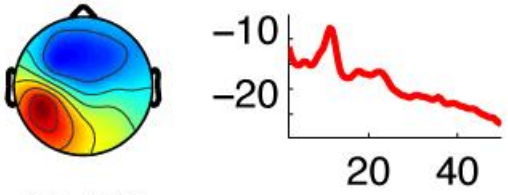
S4 IC37



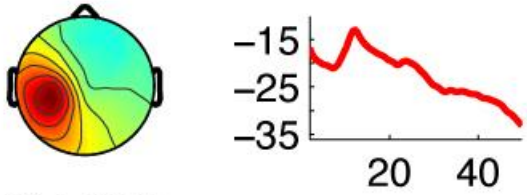
S5 IC48



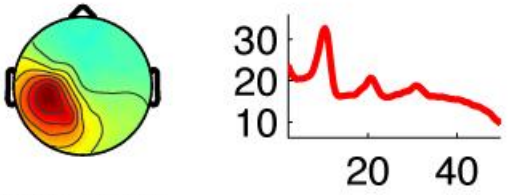
S6 IC46



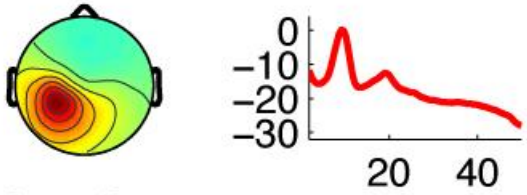
S7 IC35



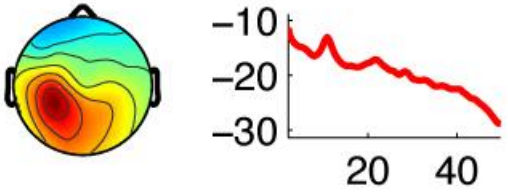
S9 IC7



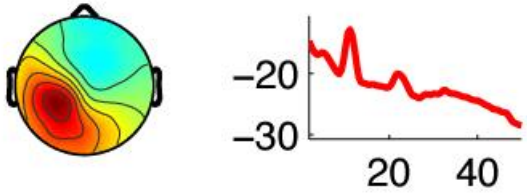
S11 IC45



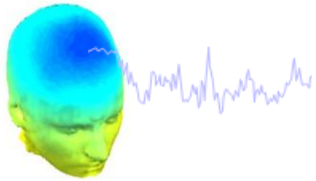
S12 IC45

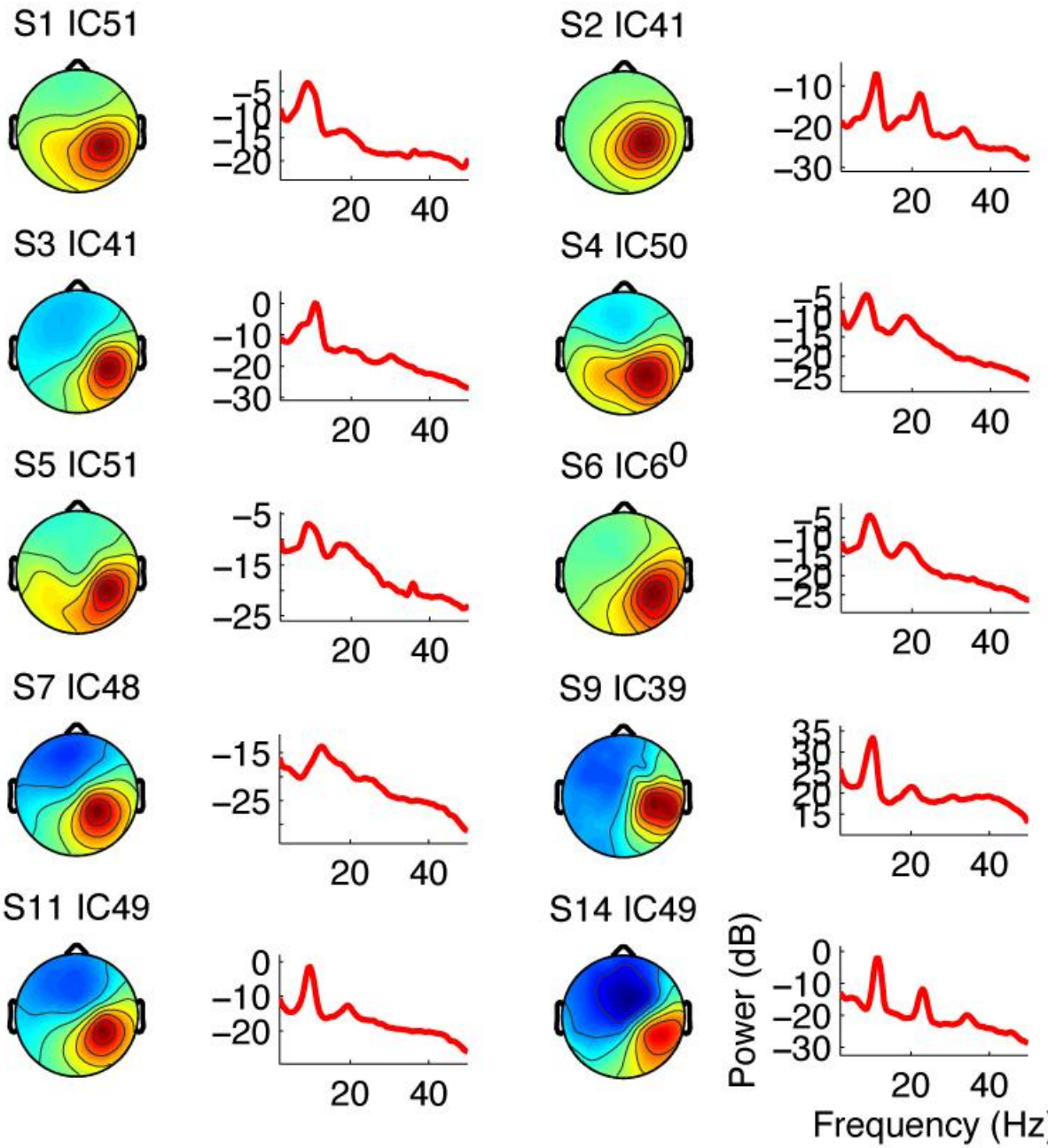


S14 IC45



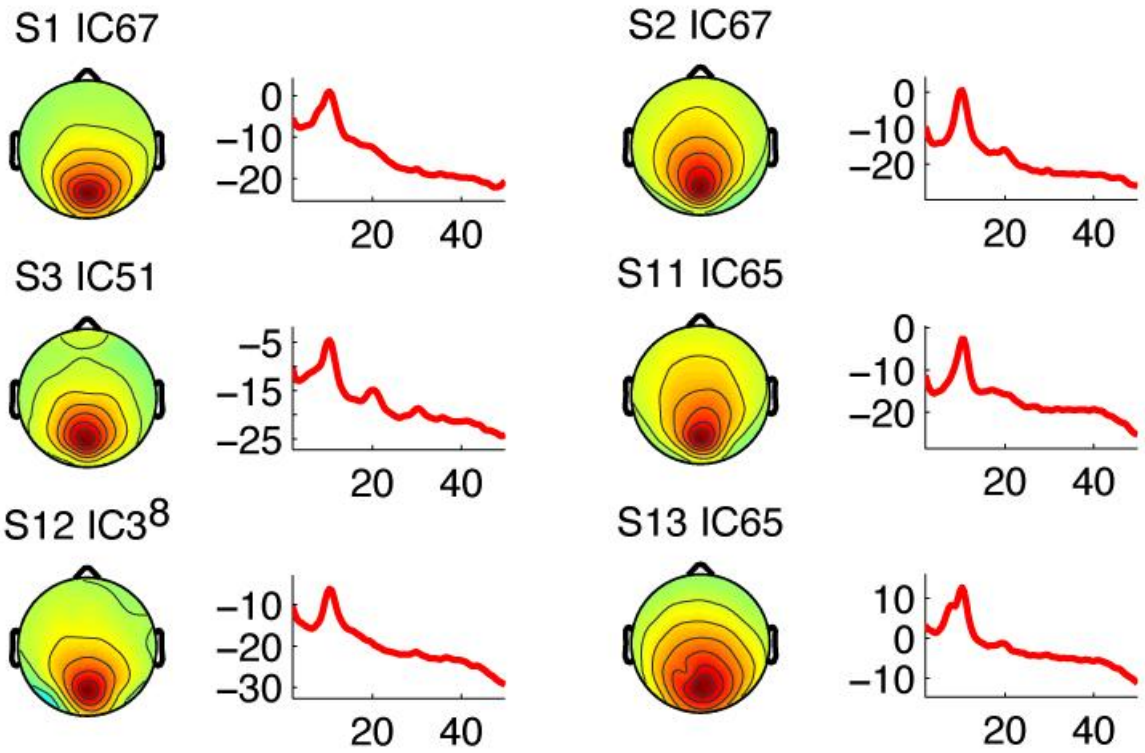
Right μ cluster



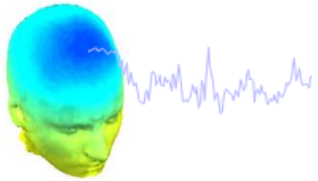


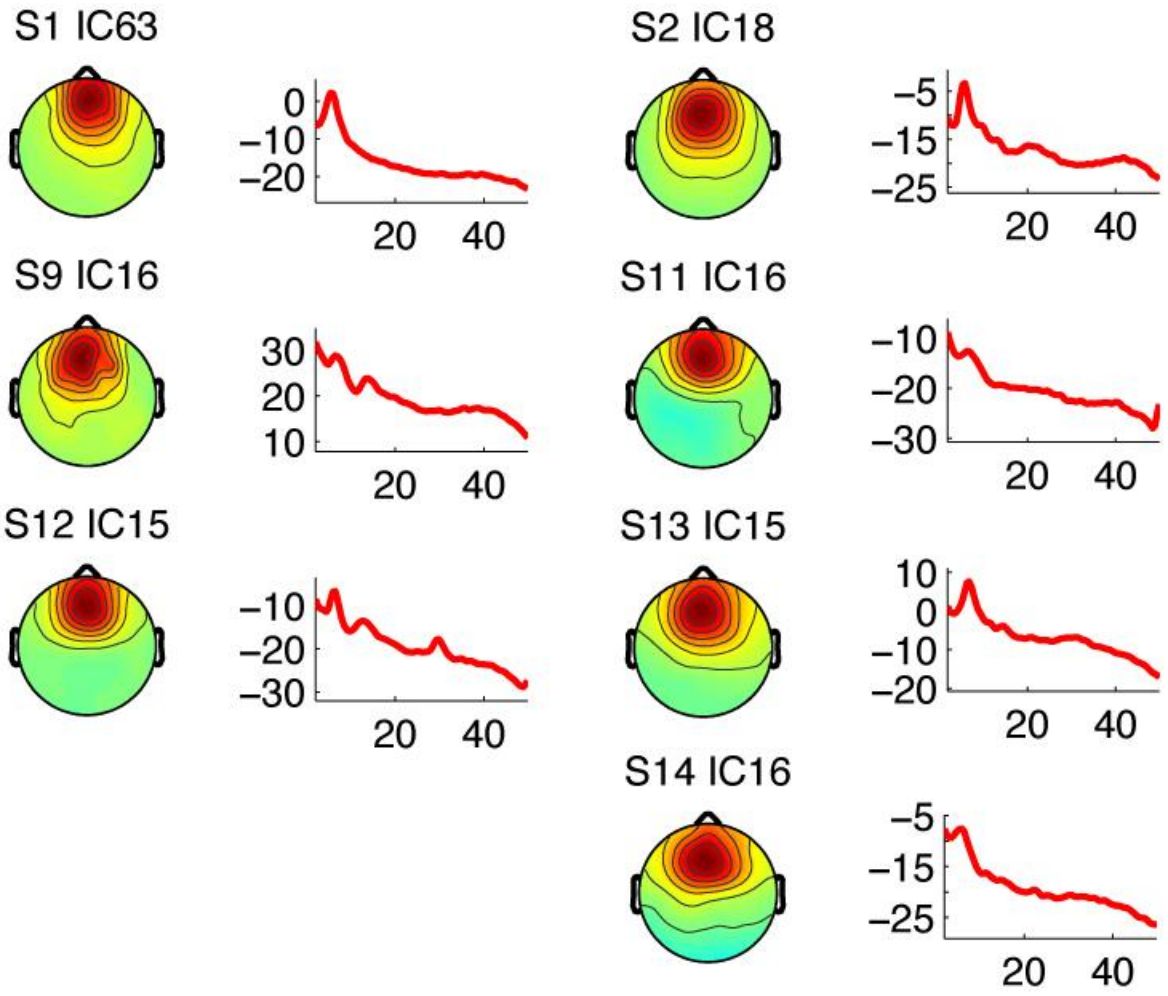
Swartz Center for Computational Neuroscience

Occipital α cluster



Frontal Midline θ cluster





Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice

Edit dataset info

pop_study(): Pre-select components

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

15

Keep only in-brain dipoles.

Cancel | Help | Ok

Create a new STUDY set -- pop_study()

Edit STUDY set information - remember to save changes

STUDY set name: Sternberg

STUDY set task name: Sternberg

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 >

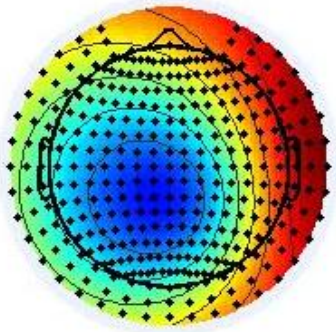
Dataset info (condition, group, ...) differs from study info. [set] = Overwrite dataset info.

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

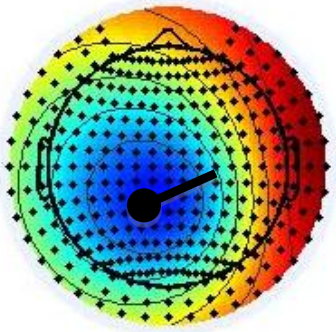
Help | Cancel | Ok

Computing residual variance (%)

Actual

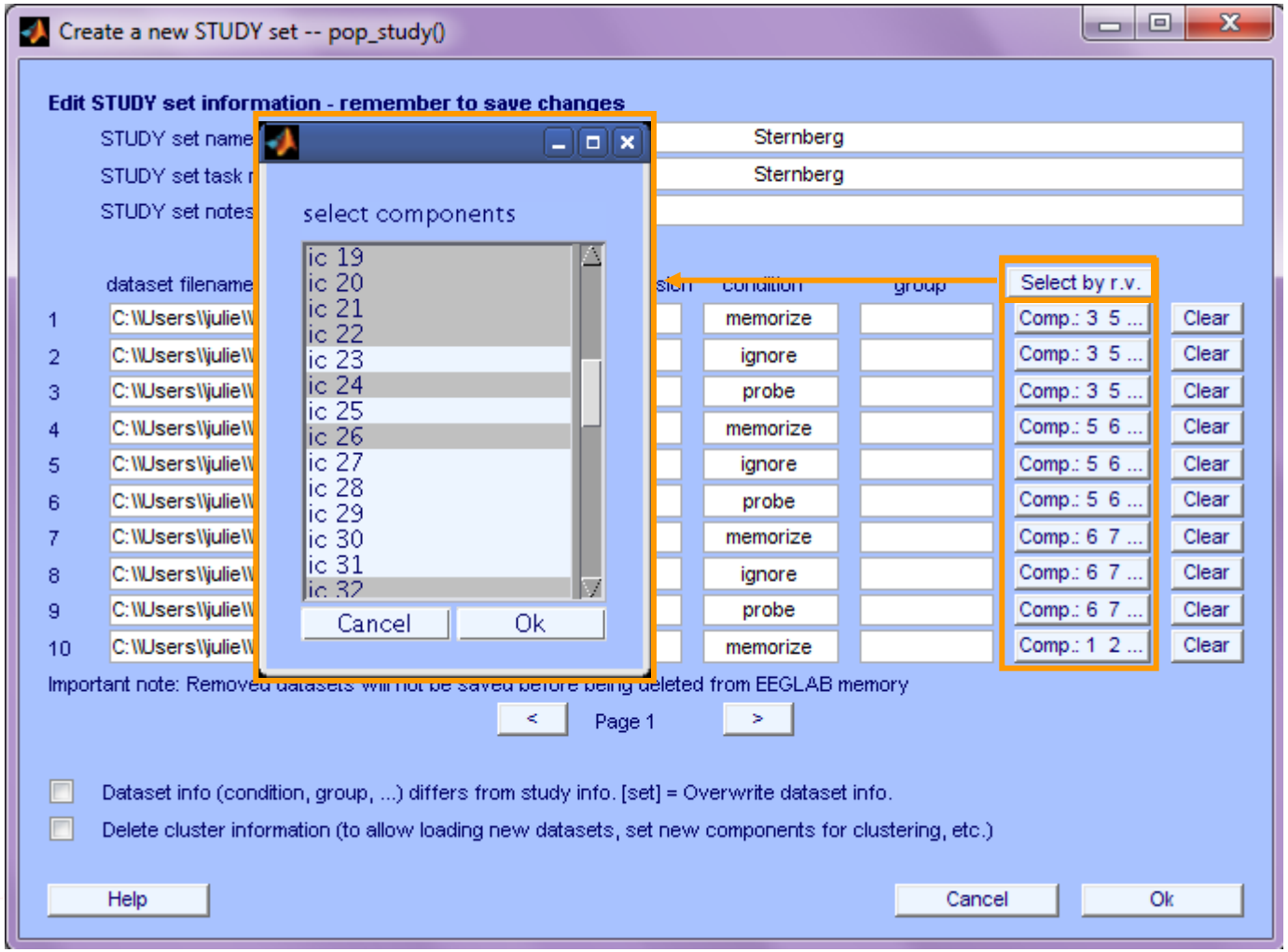


Dipole projection

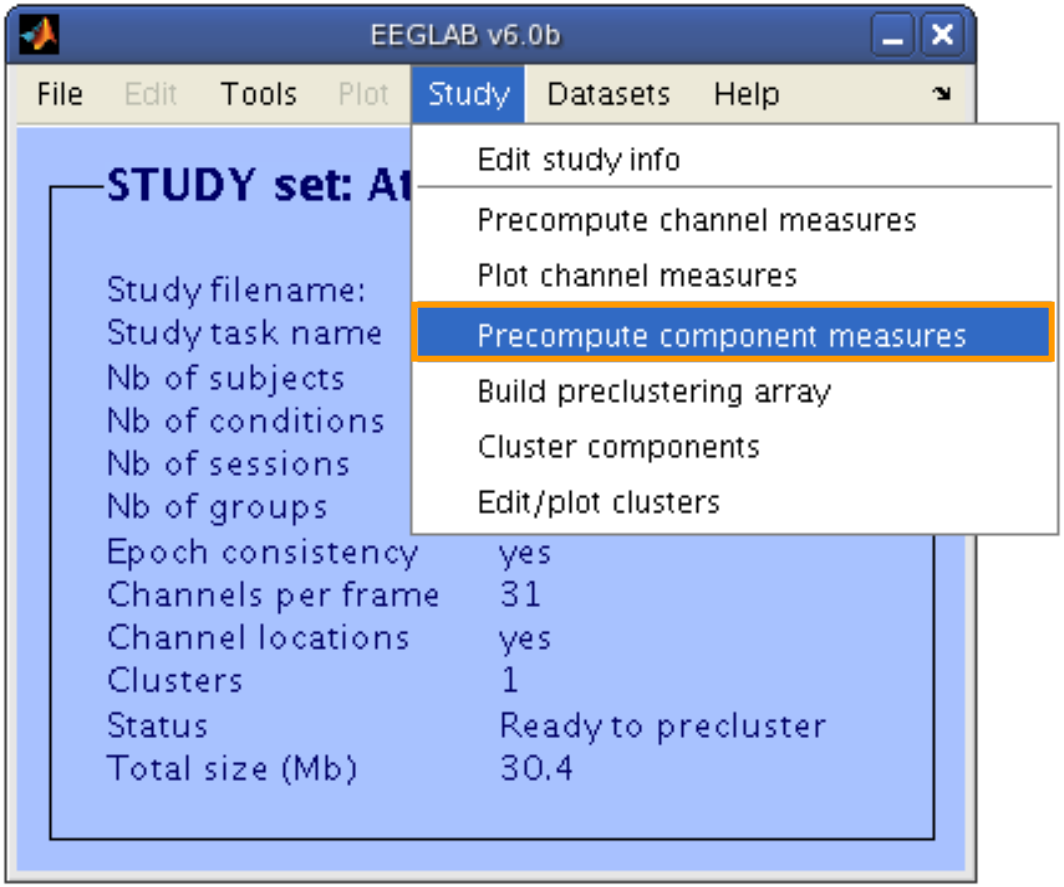


$$r = \Sigma(x_i - \tilde{x}_i)^2 / \Sigma x_i^2$$

ICs to cluster

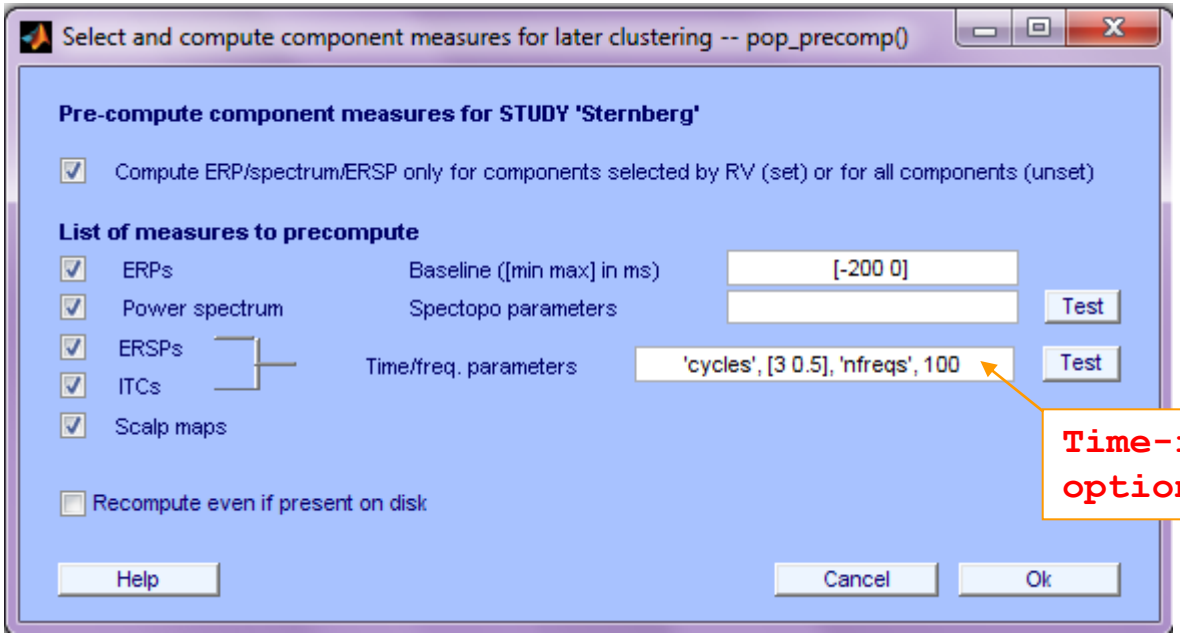


Precompute data measures



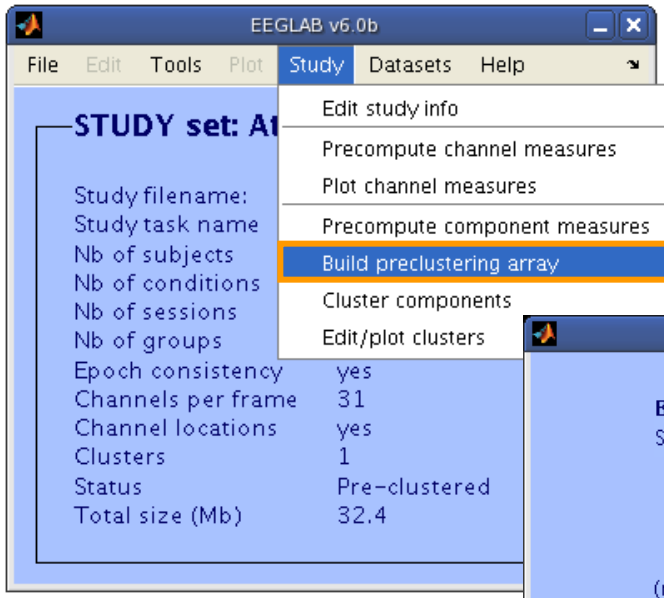
Precompute data measures

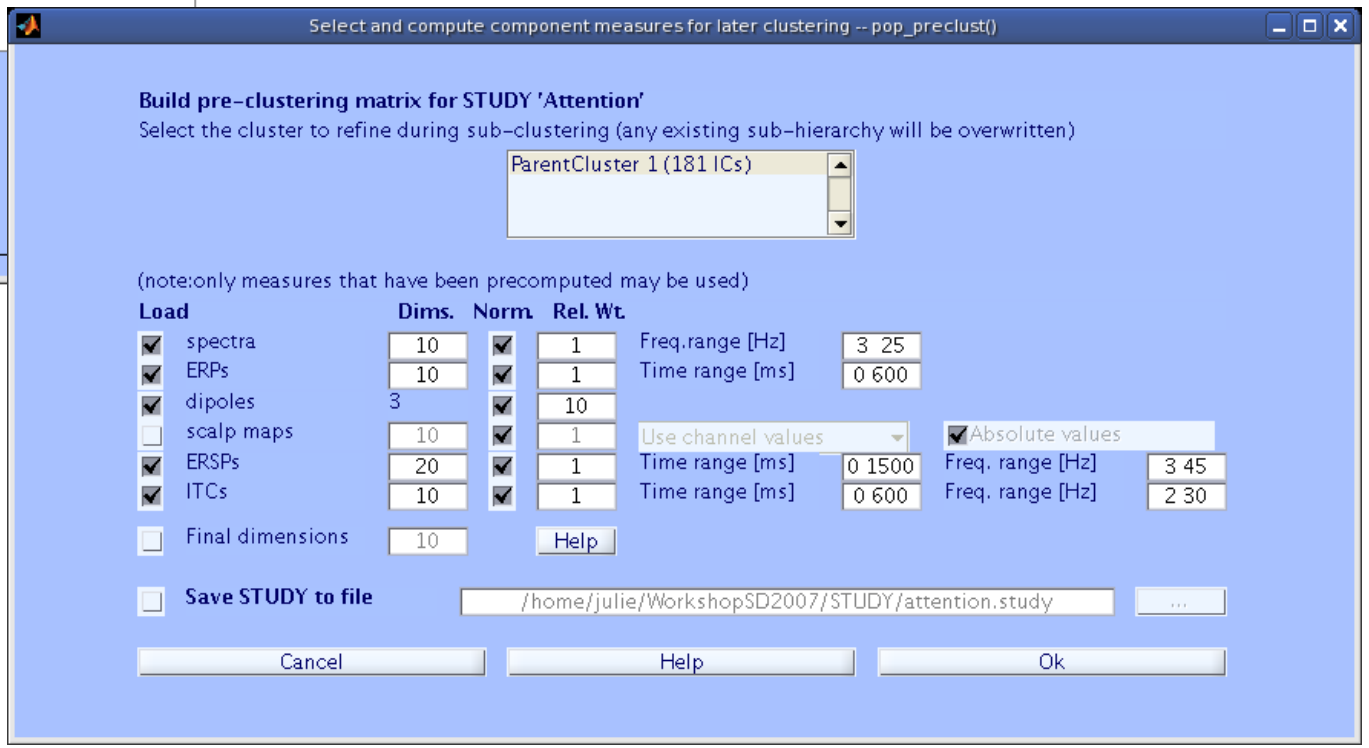
TIP: Compute all measures so you can test different combinations for clustering



Time-frequency options

3. Cluster components





Precluster schematic

Swartz Center for Computational Neuroscience

ERSP (time/freq)

Mean ERSPs

ICs (all subj)

PCA

ICs (all subj)

PC templates

PCs

* PC weights

Power spectrum

Mean spectra

ICs (all subj)

PCA

ICs (all subj)

PC templates

PCs

* PC weights

3D dipole position

ICs (all subj)

PC templates

x3

* PC weights

concatenate

concatenate

ICs (all subj)

ERSP

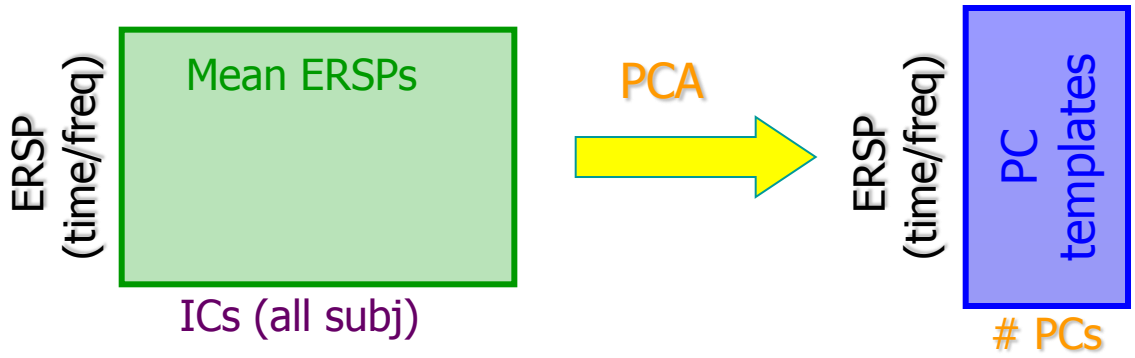
Spectrum

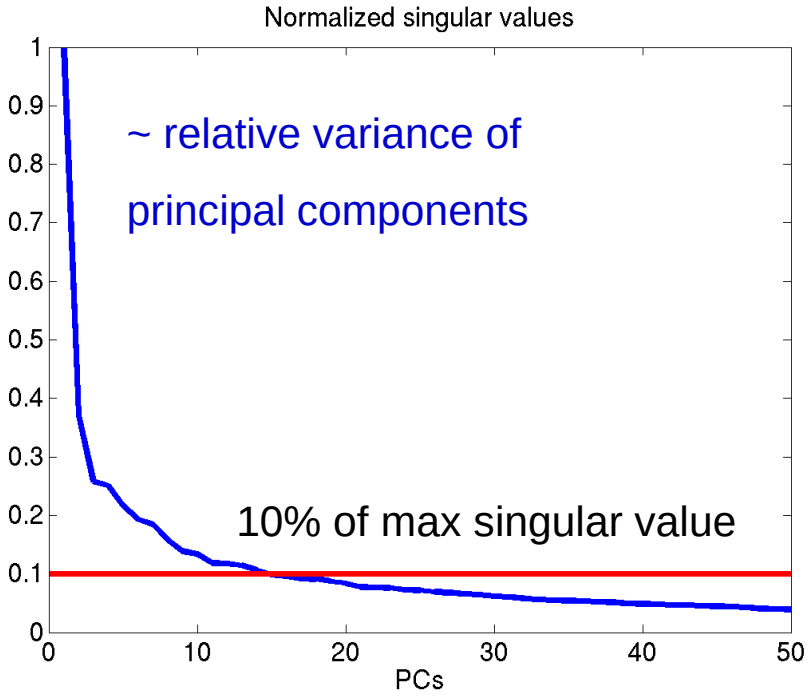
Dipoles

PCA

ICs (all subj)

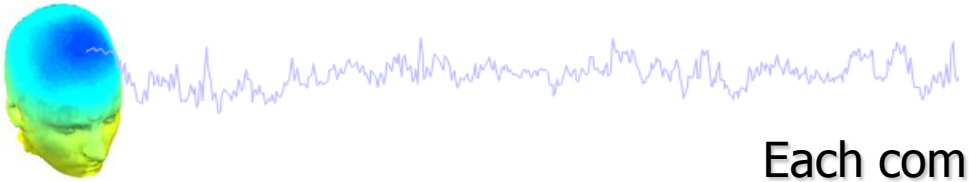
Precluster: Use singular values from PCA



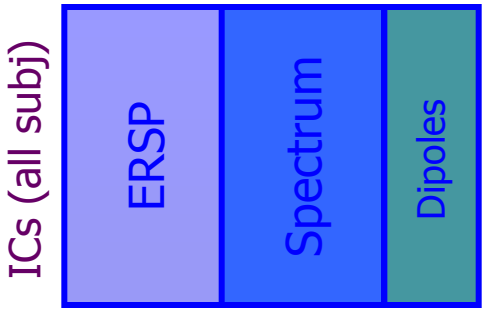


Credit: Julie Onton

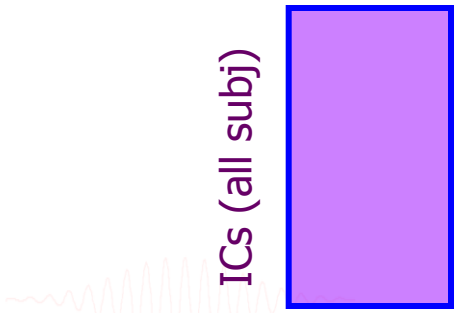
Precluster schematic

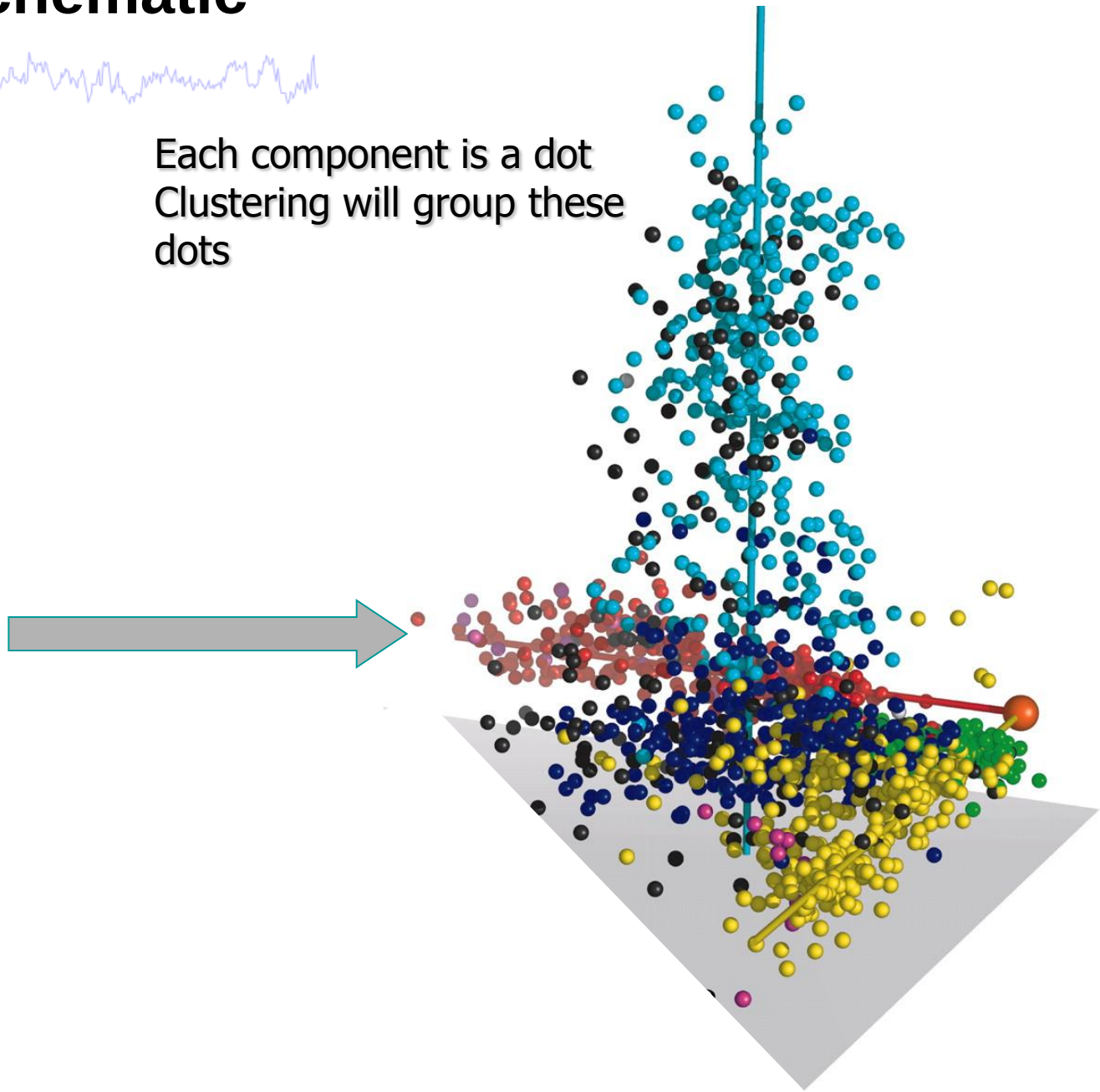


Each component is a dot
Clustering will group these dots

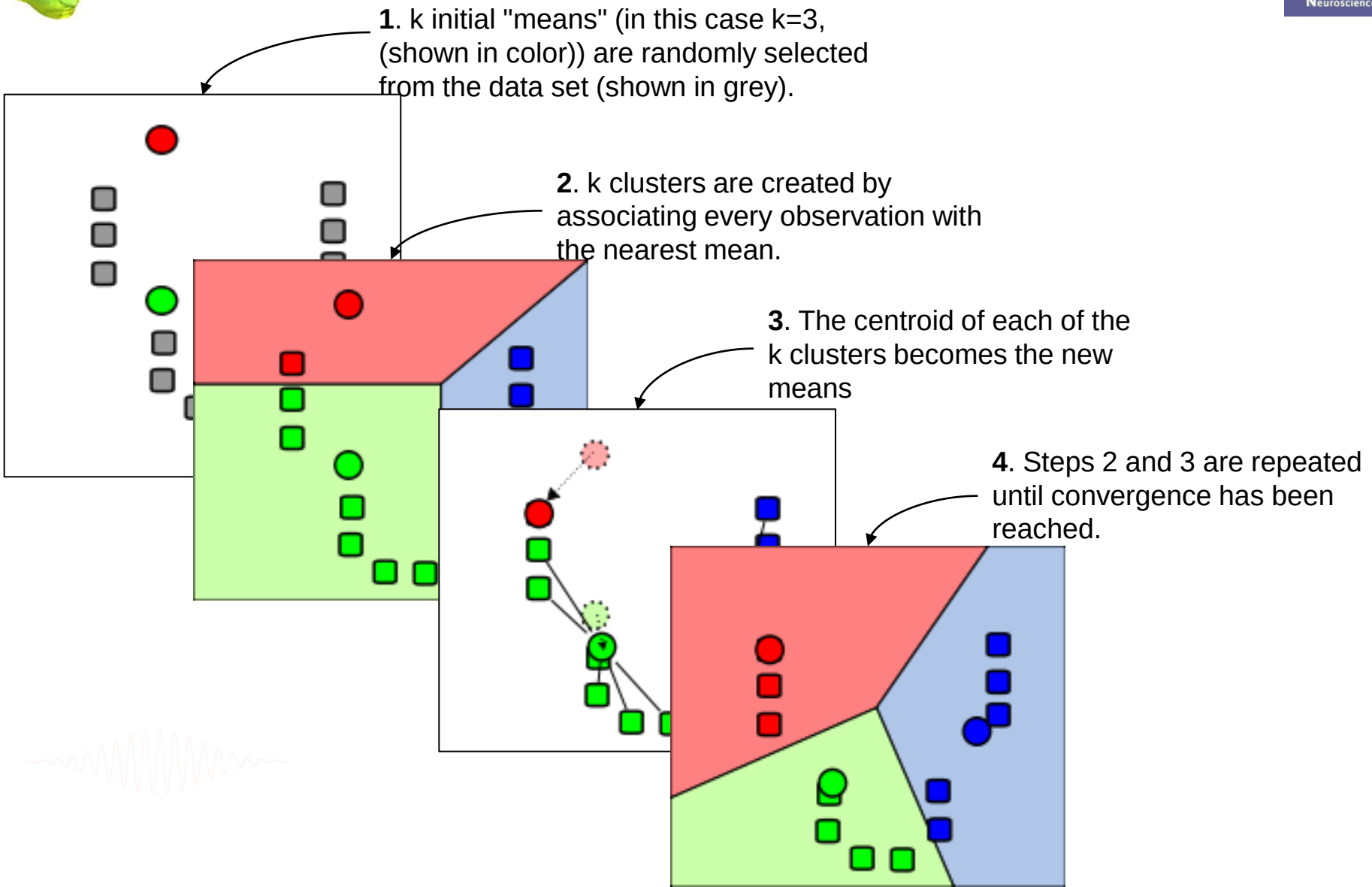


OR





Classical KMean



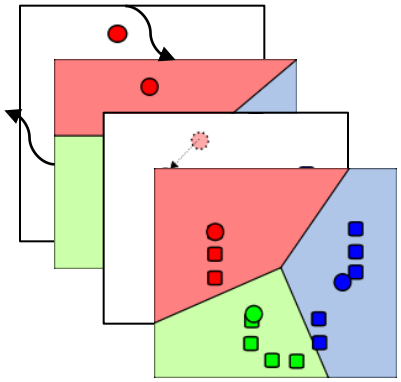
1. k initial "means" (in this case k=3, (shown in color)) are randomly selected from the data set (shown in grey).

2. k clusters are created by associating every observation with the nearest mean.

3. The centroid of each of the k clusters becomes the new means

4. Steps 2 and 3 are repeated until convergence has been reached.

Customized KMean
(no more than 1 session per cluster)



1. A first KMean solution is computed for N clusters

2. Select the cluster with minimum residual distance to centroid

3. Keeps at most one component per session

(min dist. to centroid)

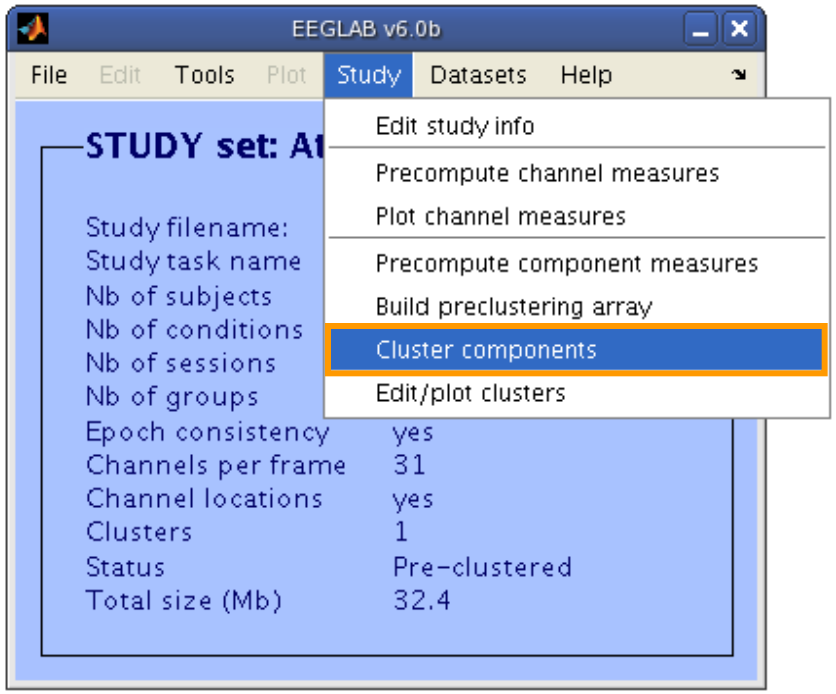
4. Store the resulting cluster

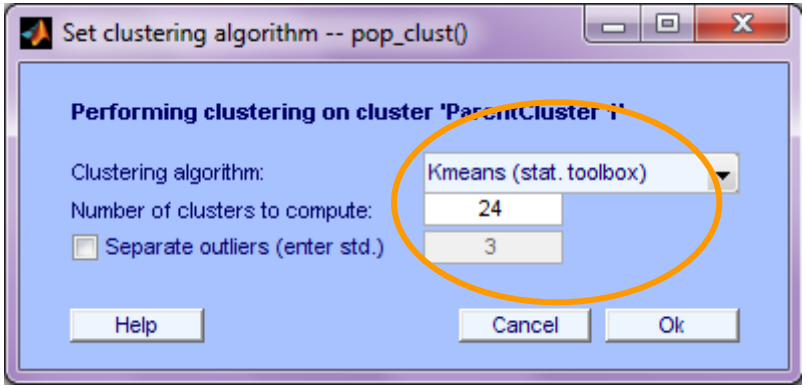
5. Remove the cluster's ICs from the pool of all ICs

6. Compute a new KMean solution for N-1 clusters on the new pool

7. Loop until the desired number of selected clusters is reached

Cluster components





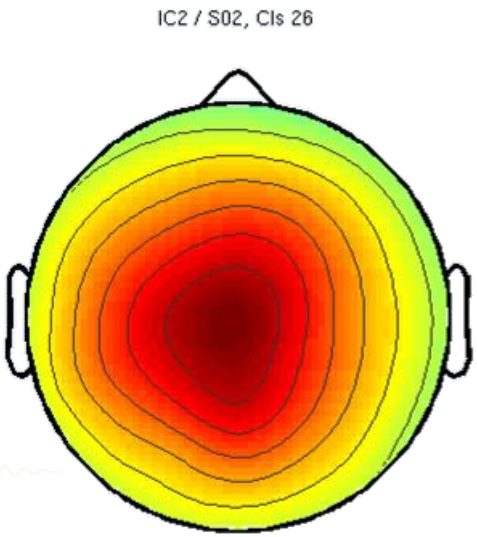
Choosing data measures

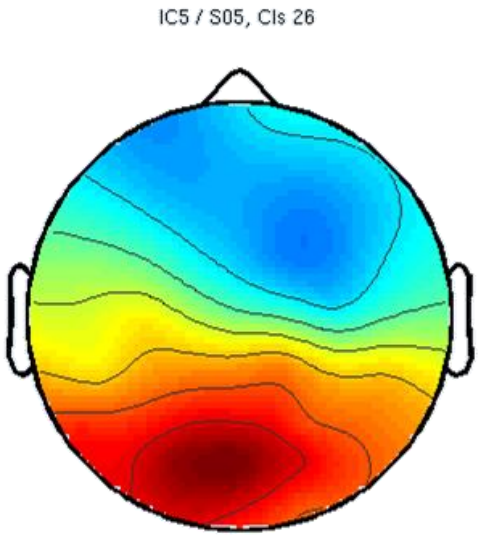
What measure(s) should you use?

- It depends on your final cluster criteria…
  - If for example, your priority is dipole location, then cluster only based on dipole location…

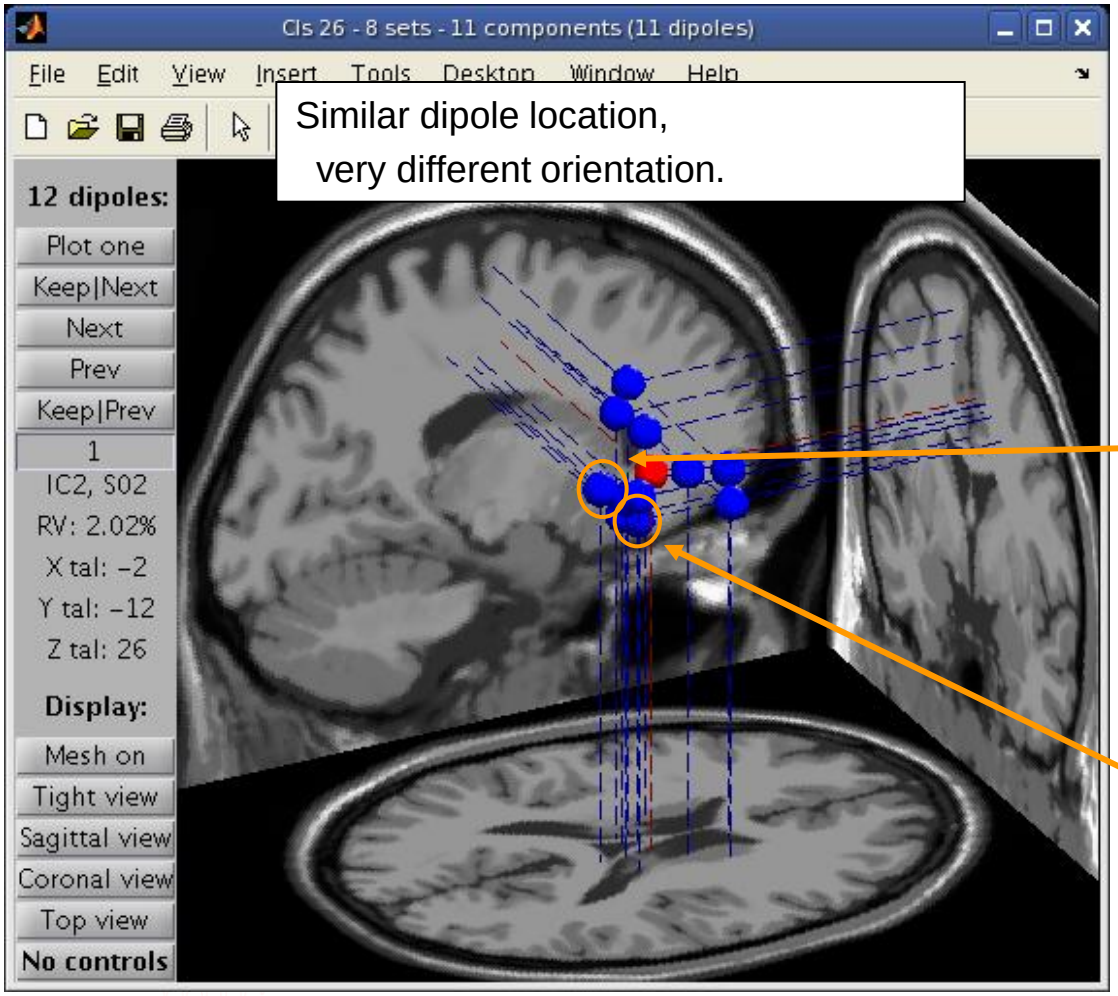
But consider:

- What is the difference between these two components?



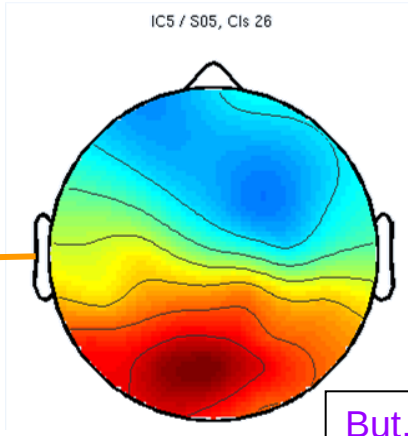


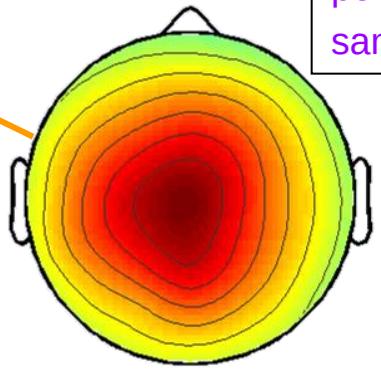
Choosing data measures



Similar dipole location,
very different orientation.

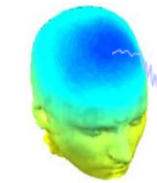
Obvious dramatic effect on scalp map topography:

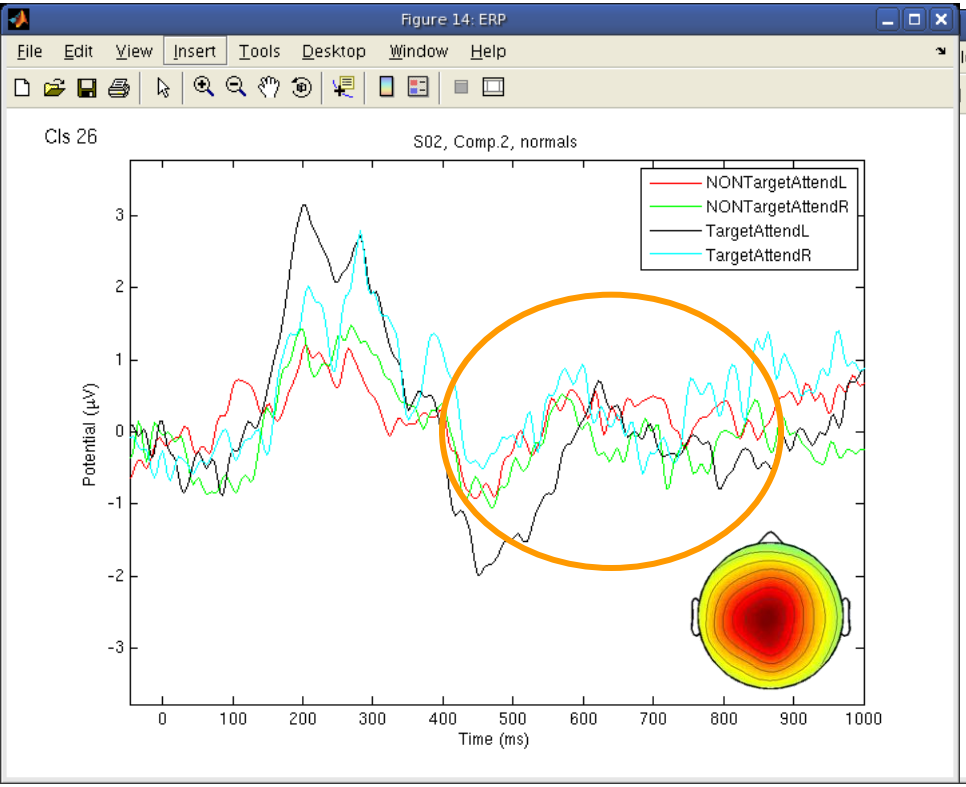


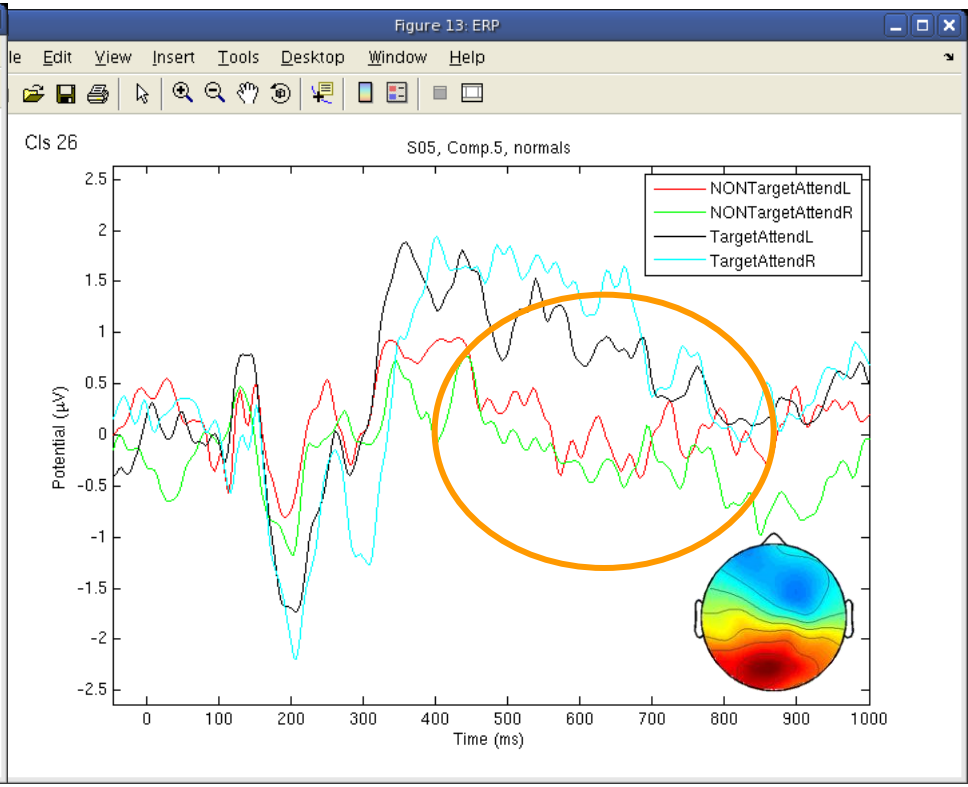


But, do they perform the same functions?

Choosing data measures

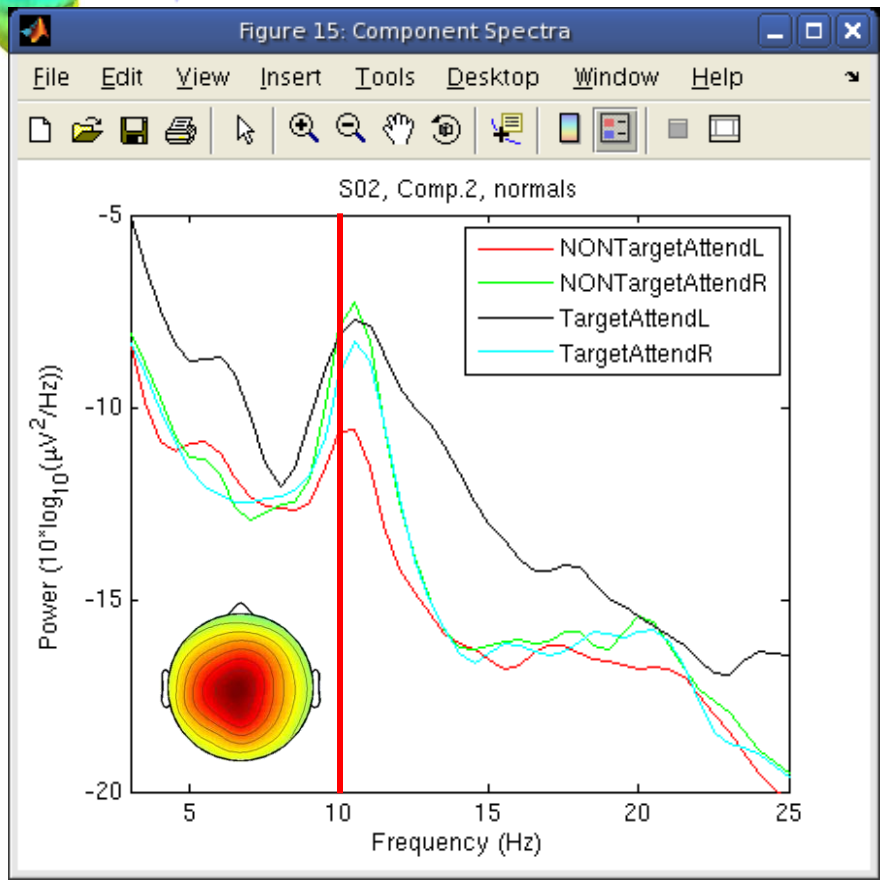


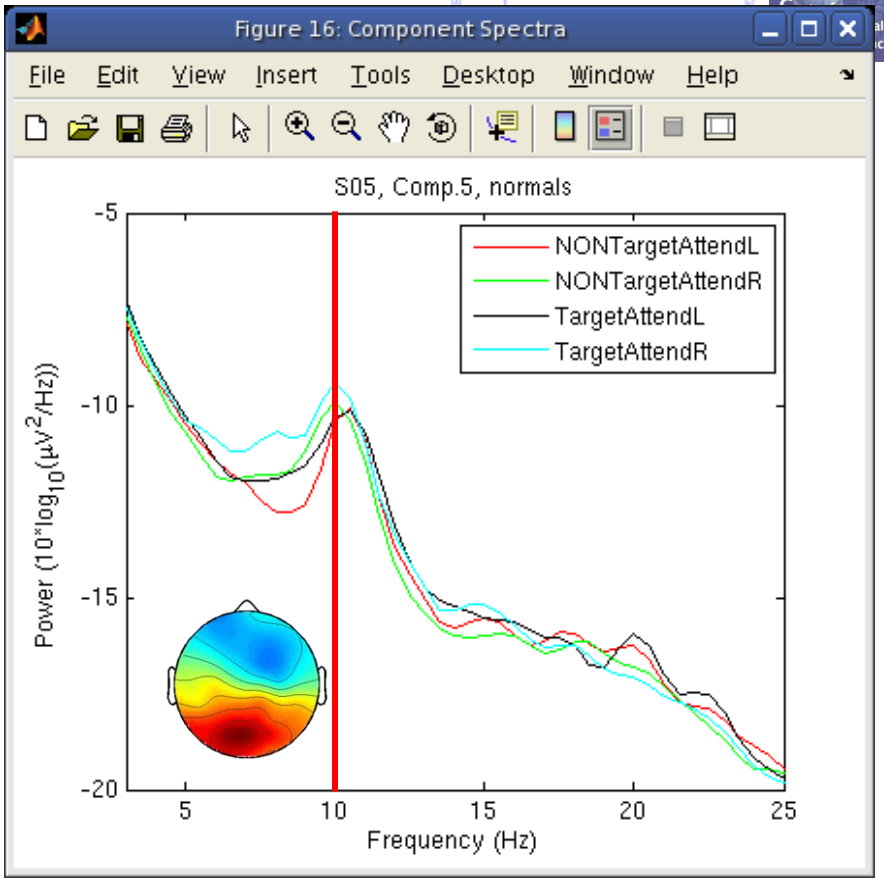




ERPs seem different…

Choosing data measures

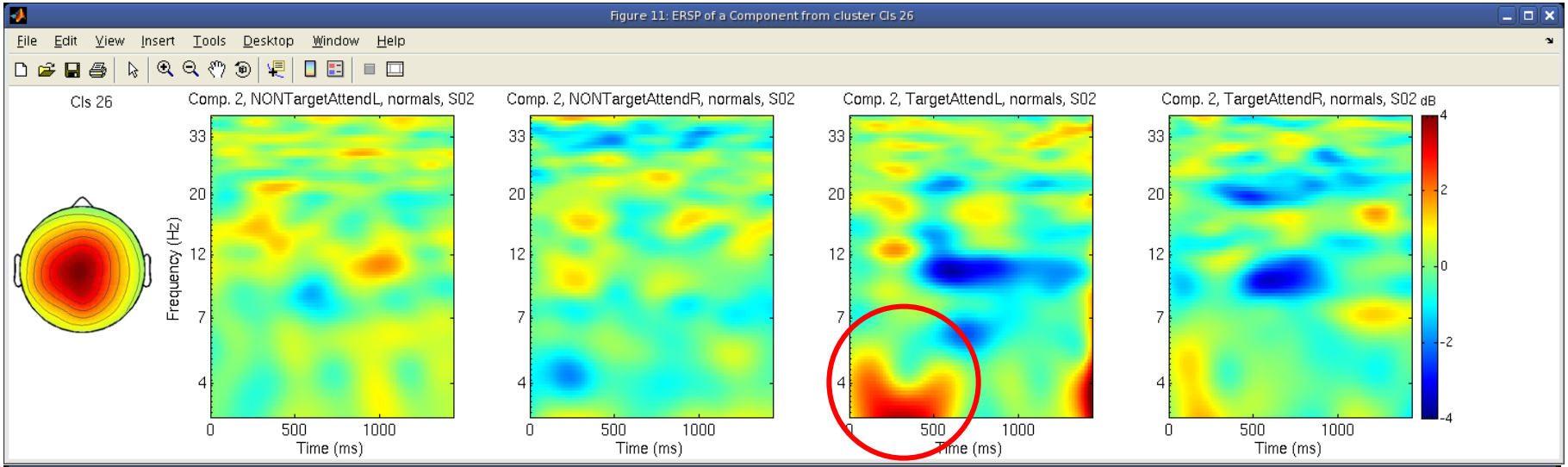


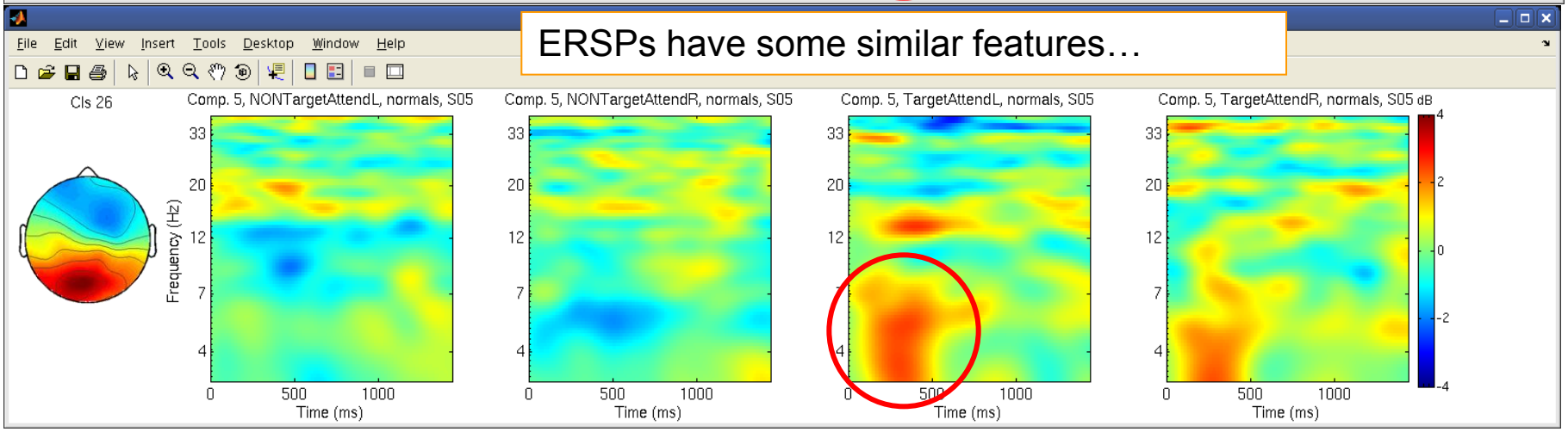


Spectra are similar, but they have

variable responses to different conditions…

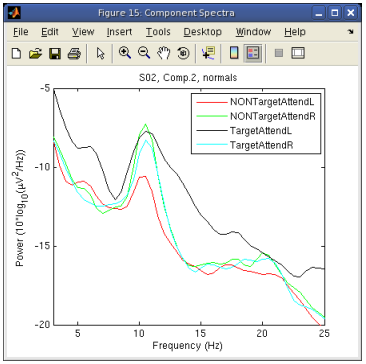
Choosing data measures

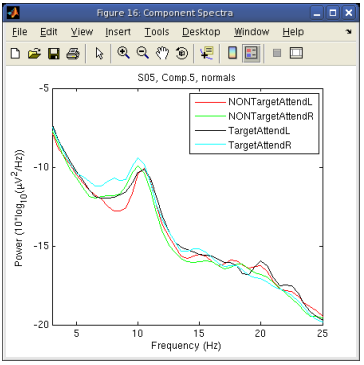




ERSPs have some similar features…

Choosing data measures

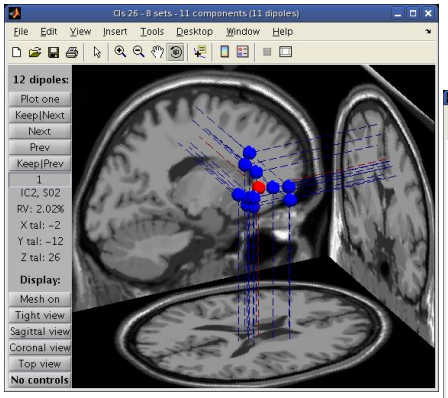


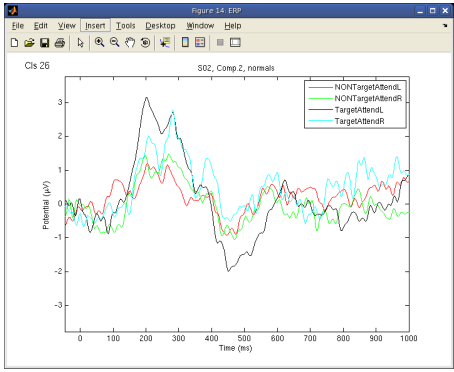


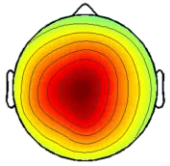
What data measures should you use?

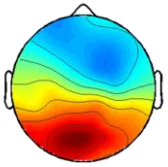
It depends…

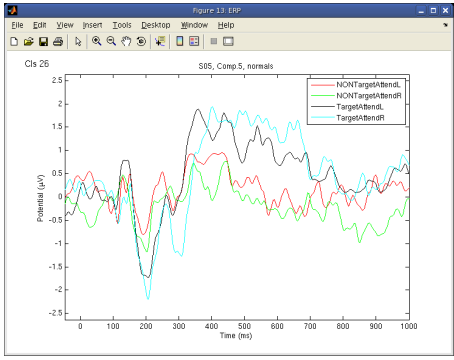
- broadly-matched ICs: use many/all of the measures.
- specifically-matched ICs: use one/few of the measures.

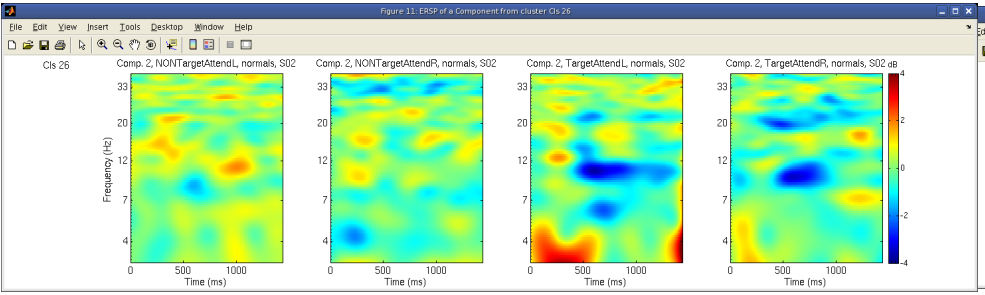


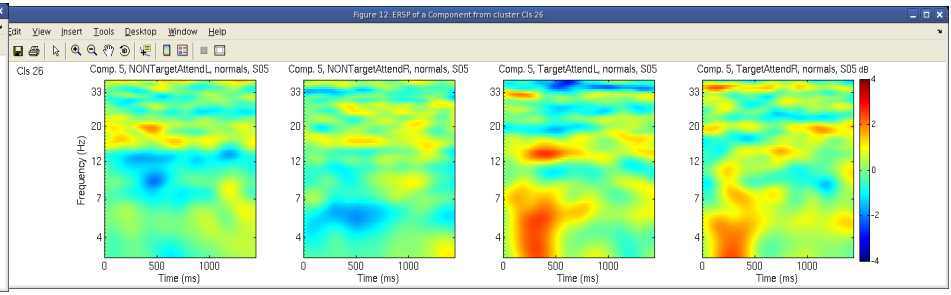




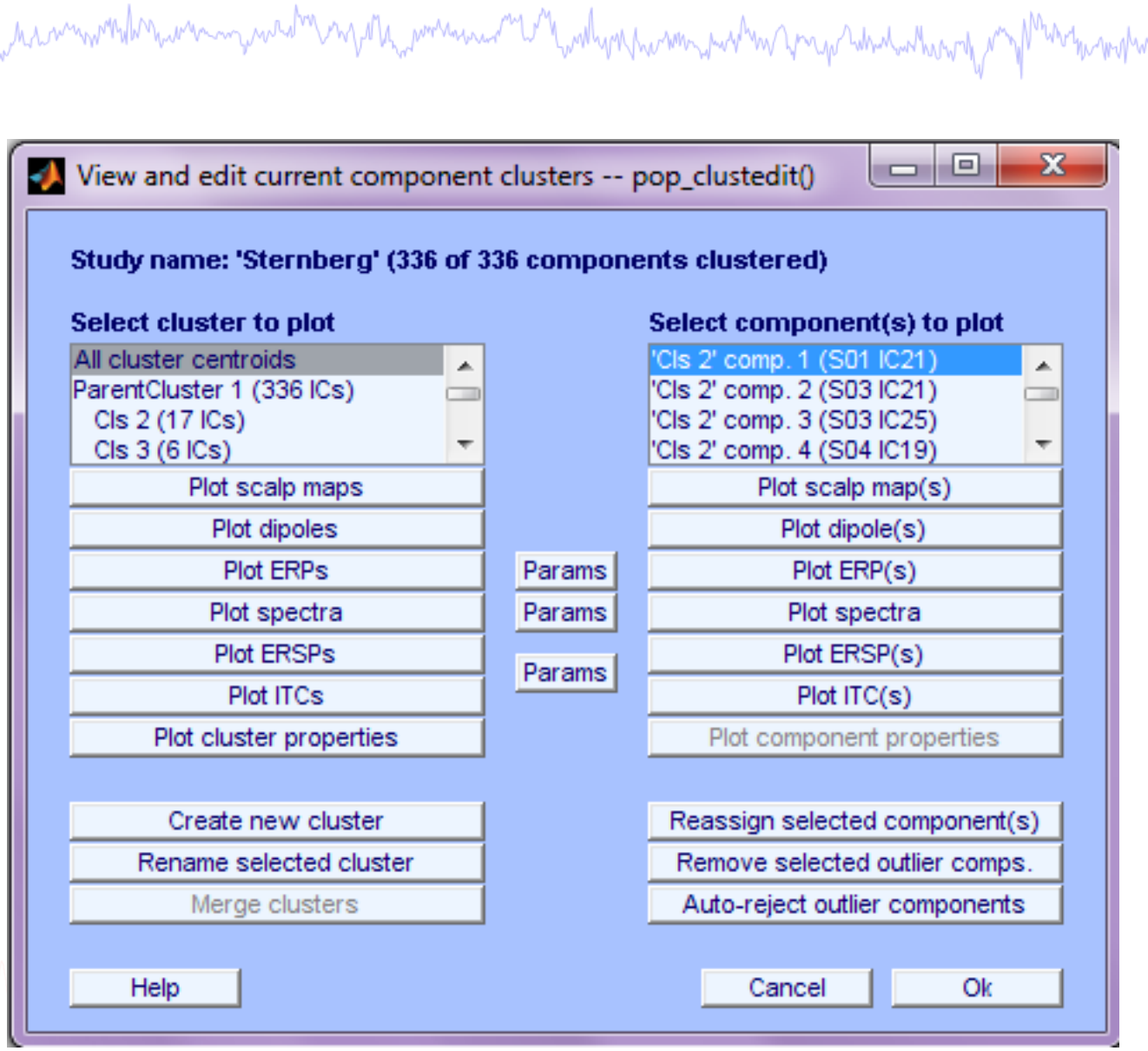




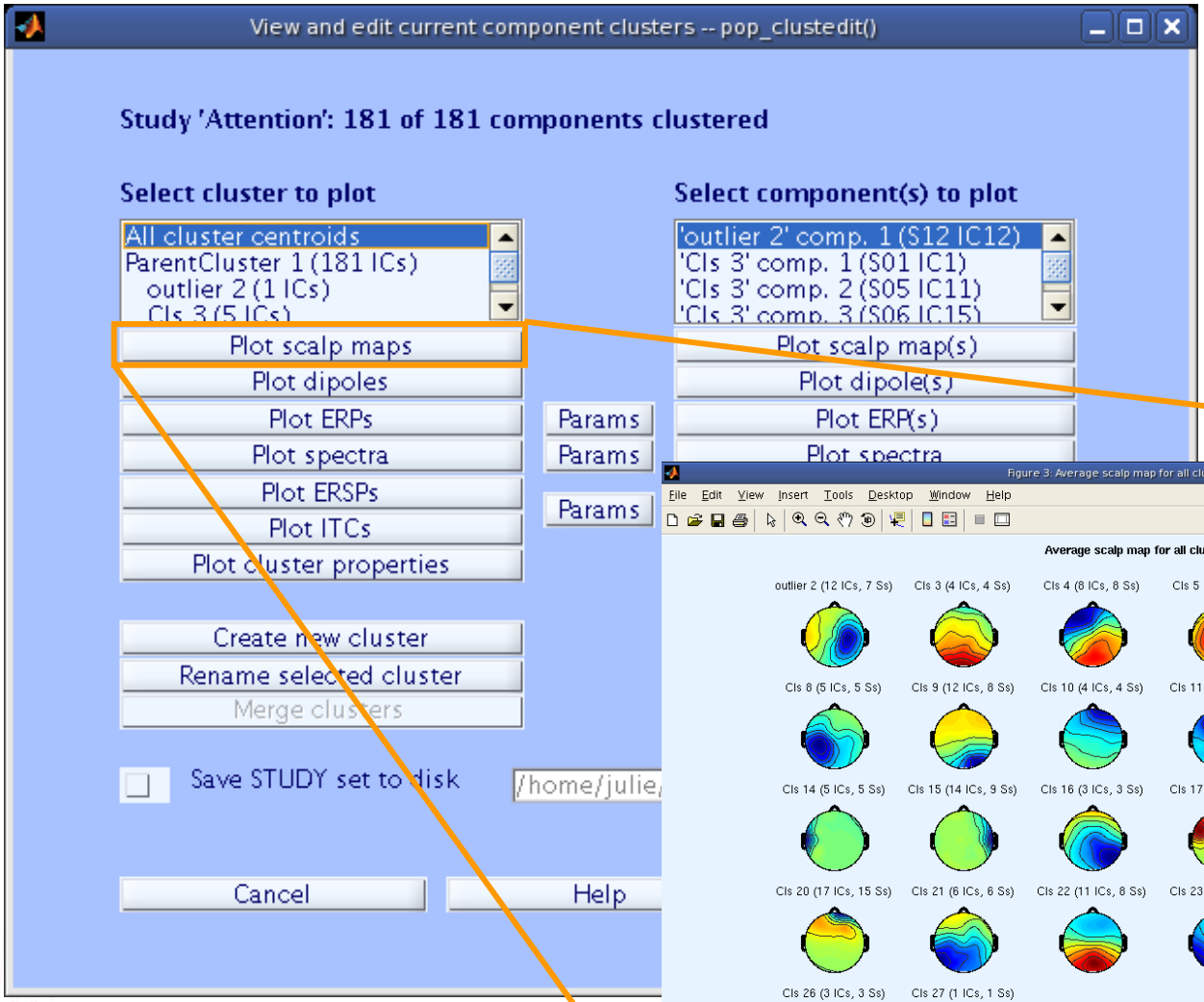


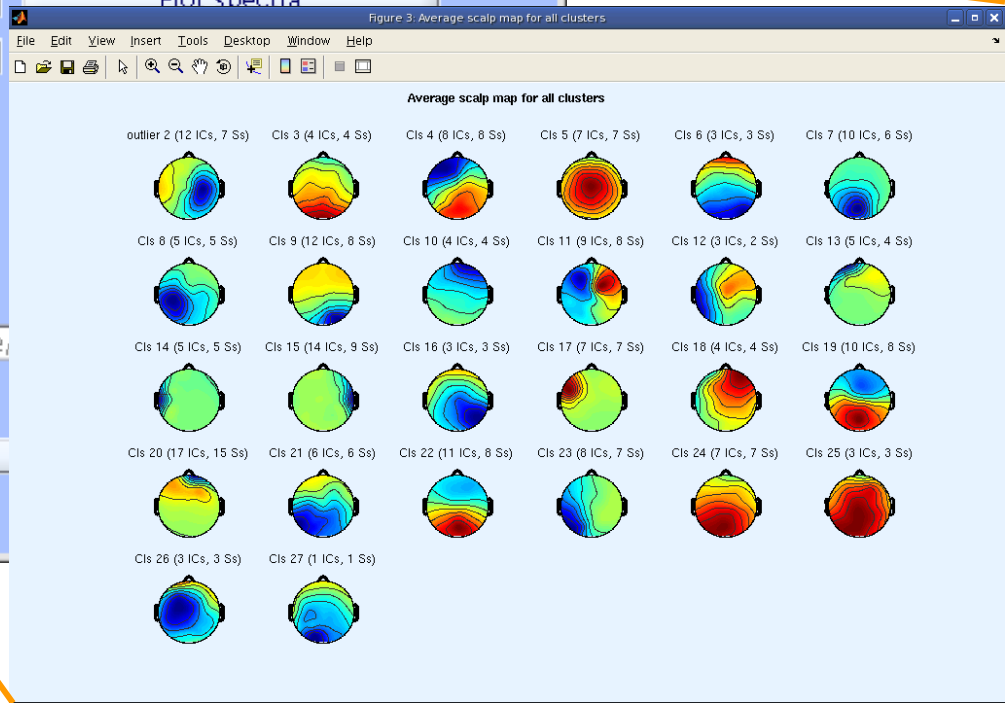


Plot/edit clusters



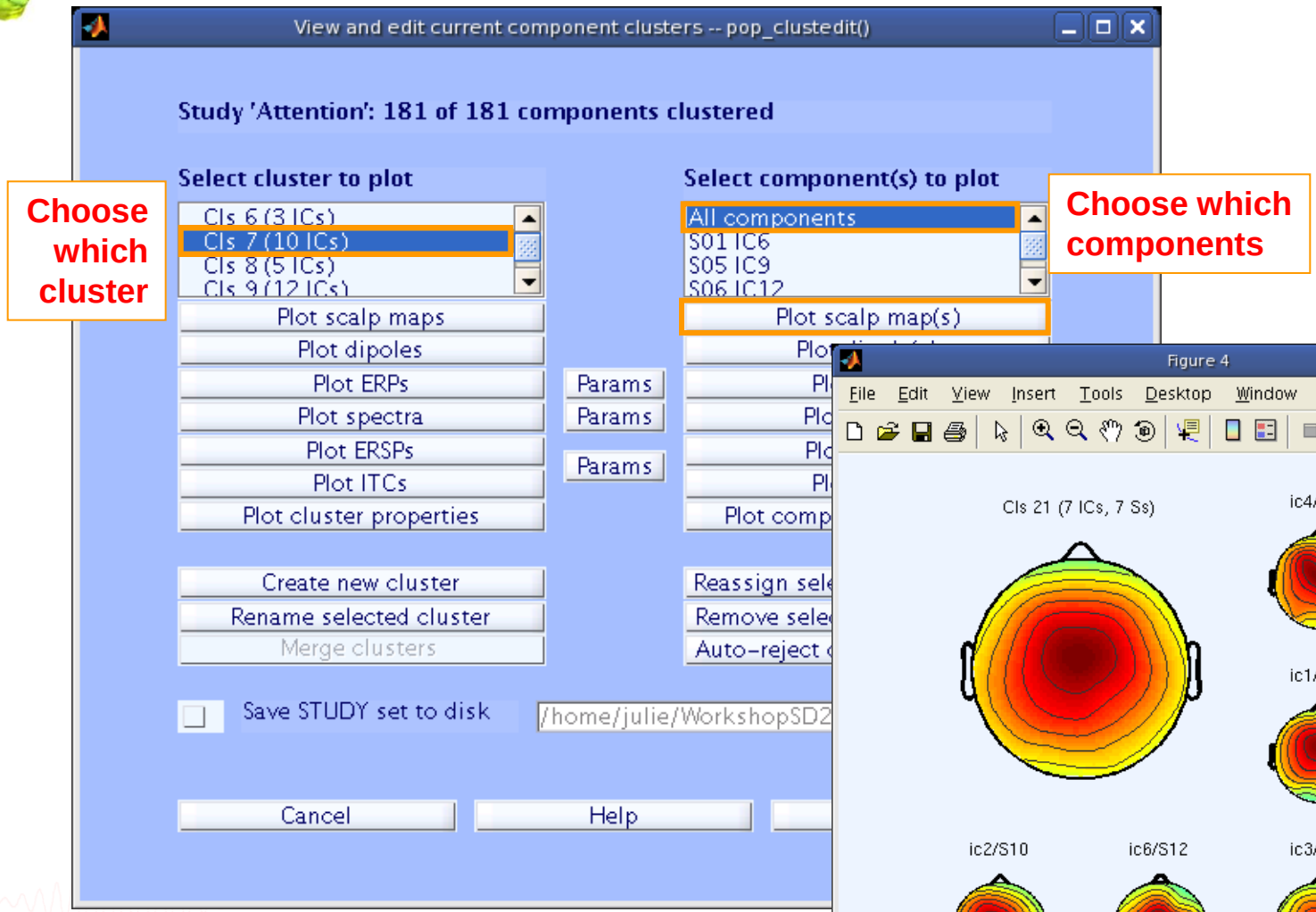
Plot cluster data



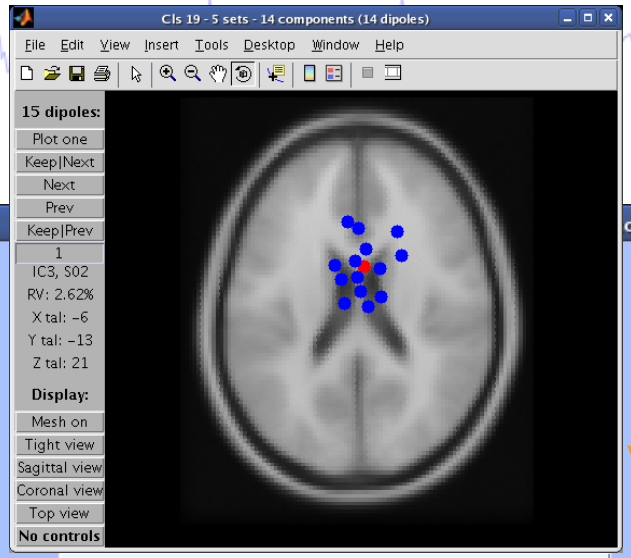


Plot mean scalp maps for easy reference

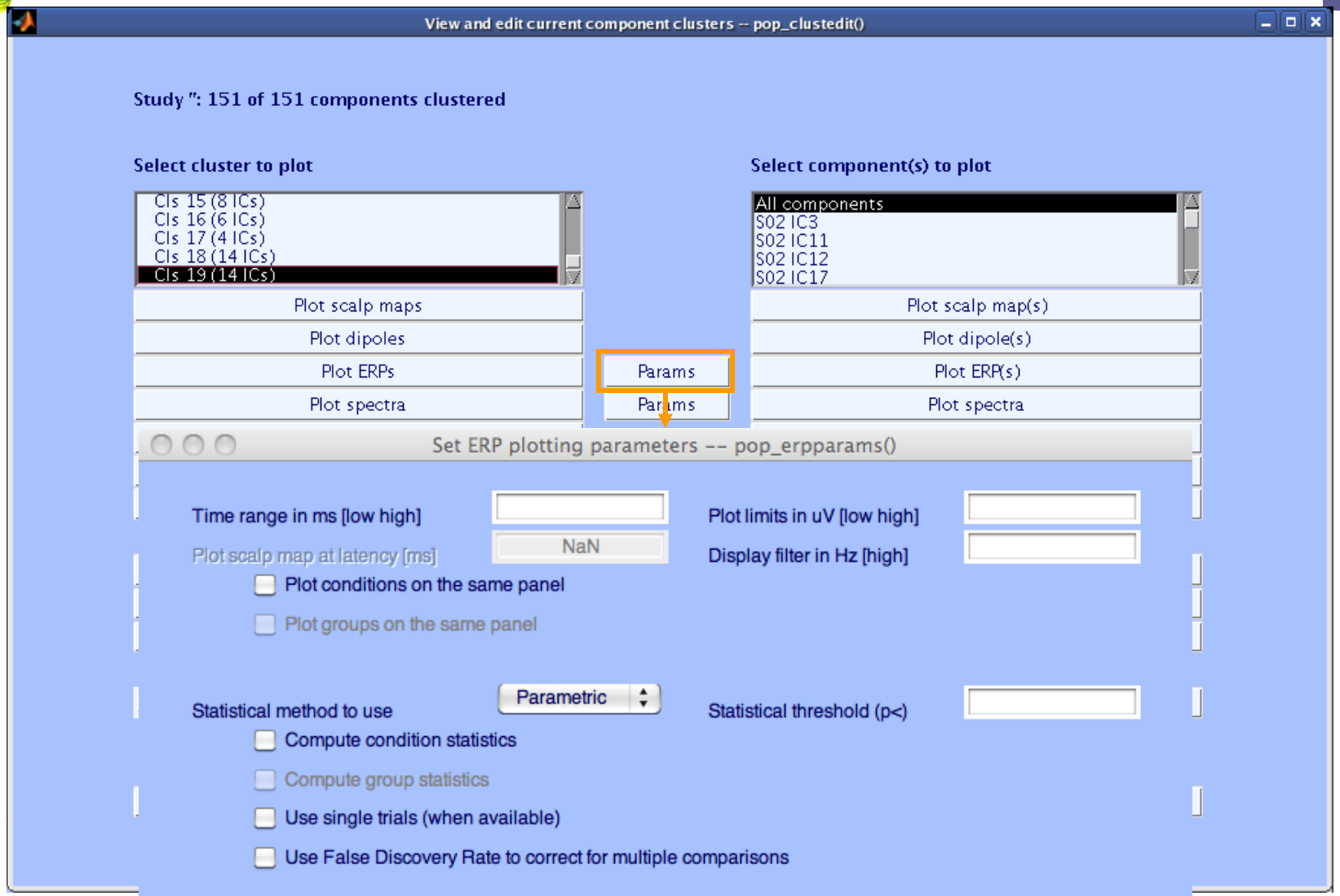
Plot cluster data



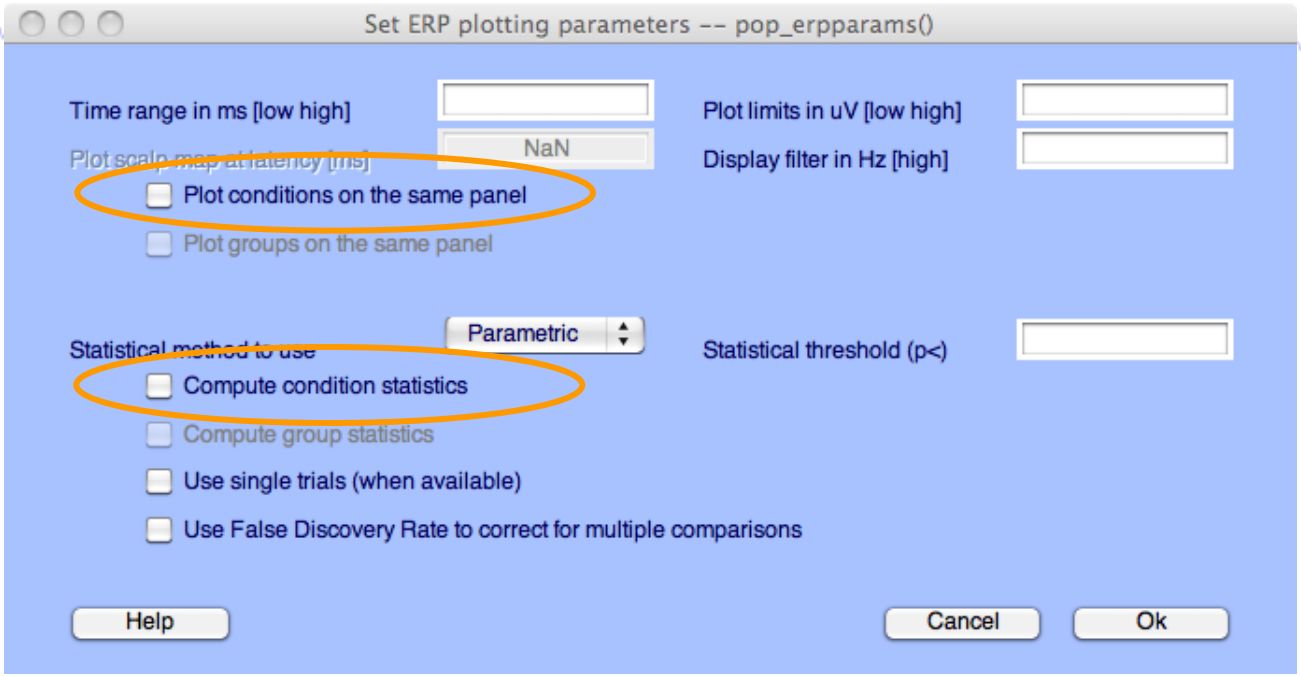
Plot cluster data

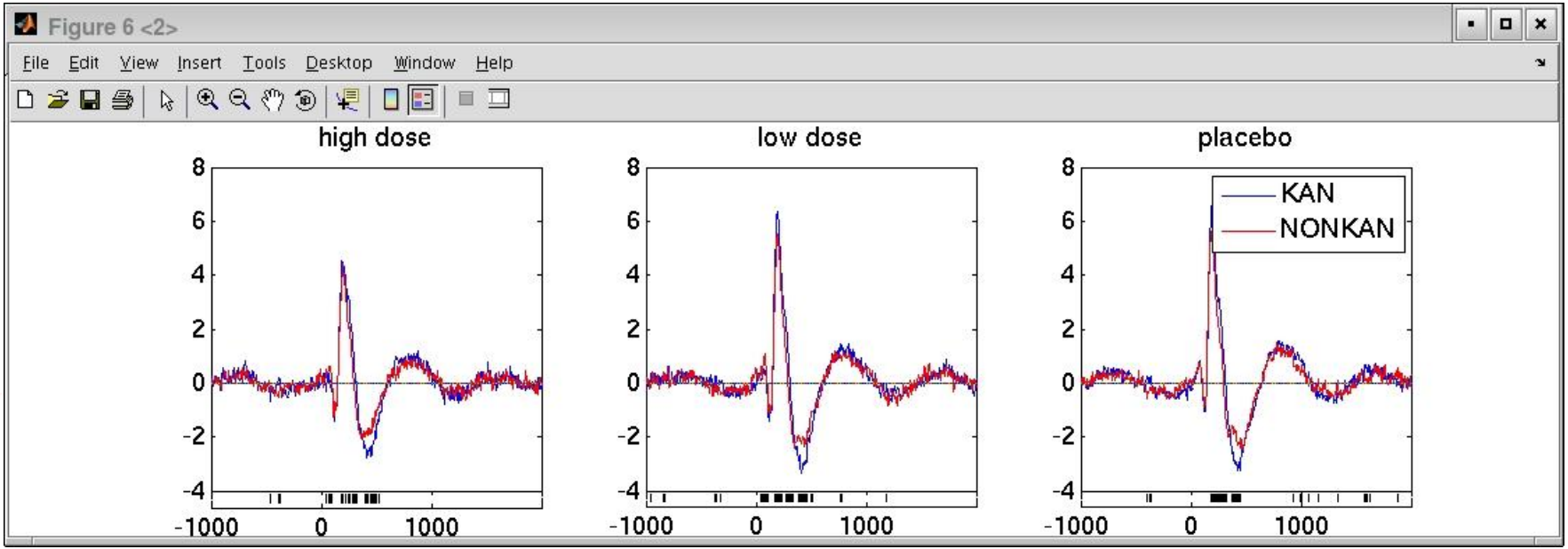


Plot cluster data

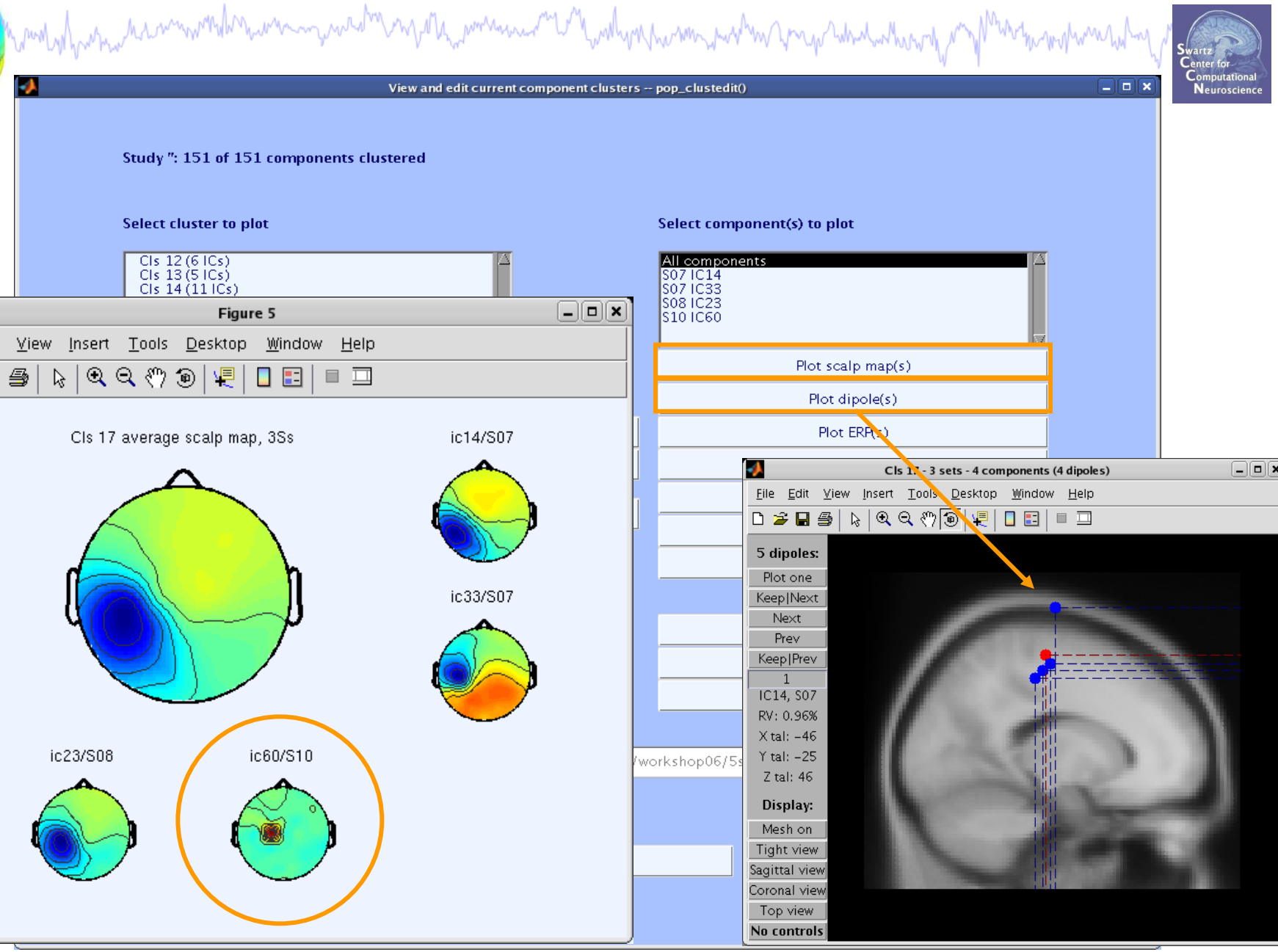


Other plotting options…

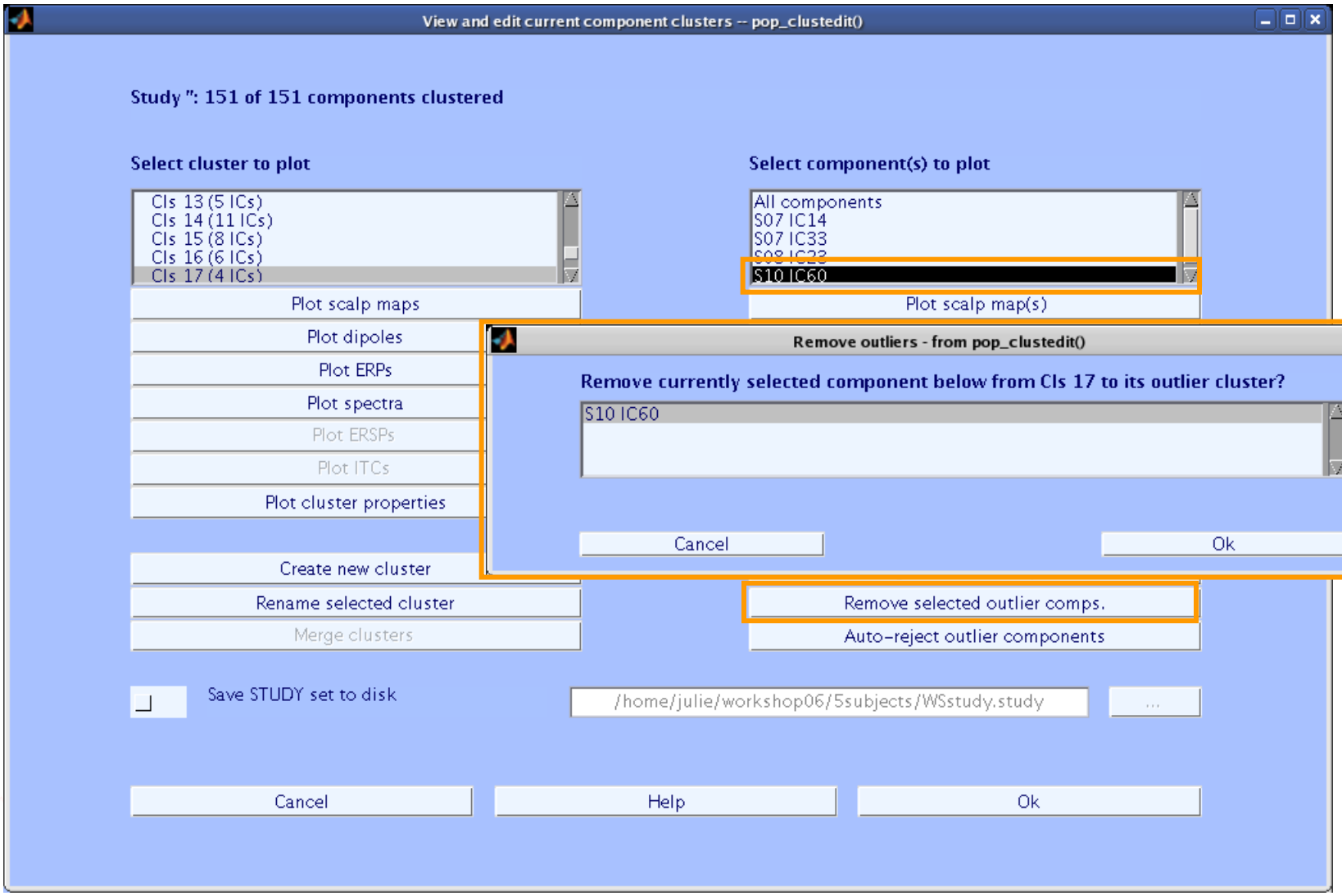




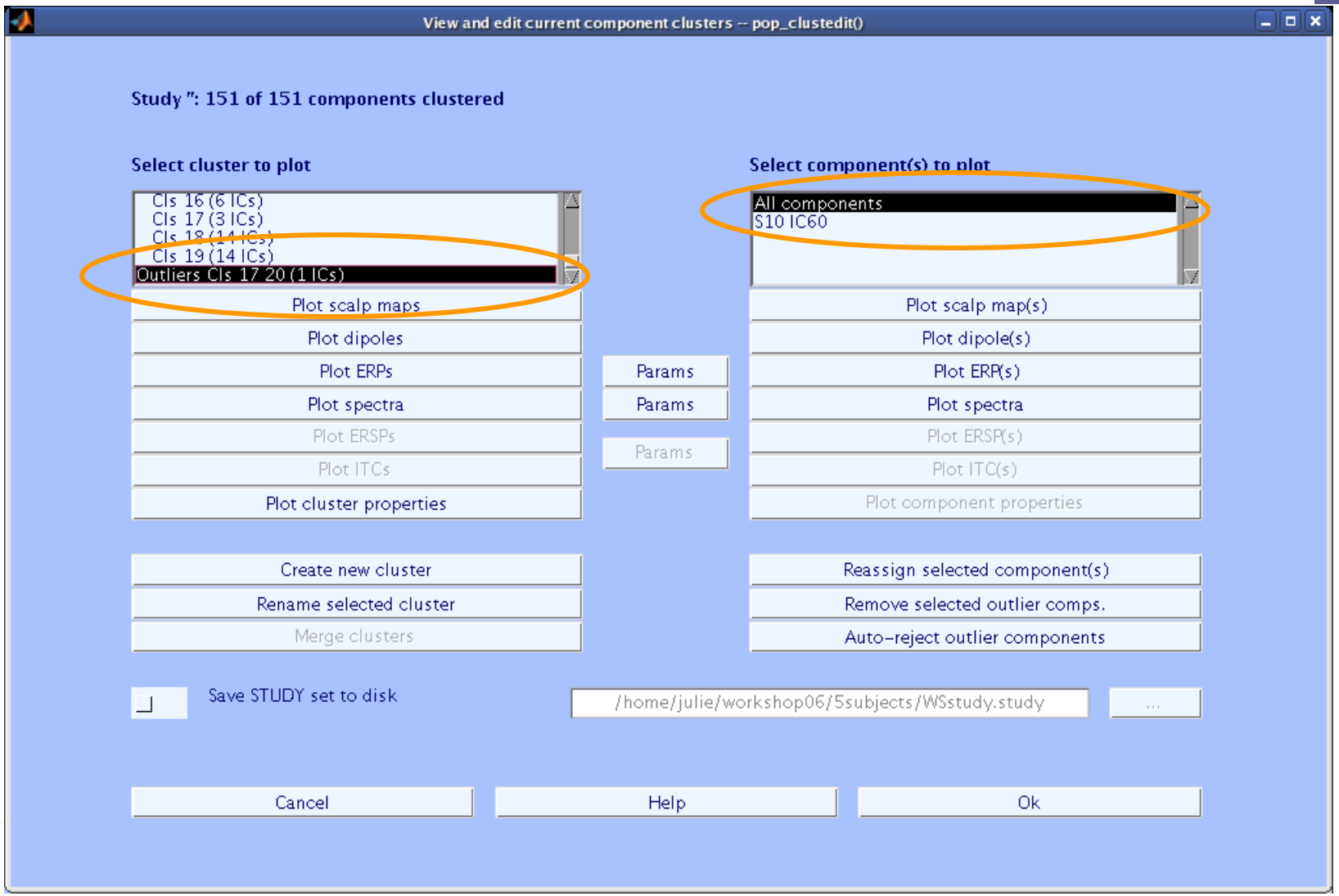
Reassigning components



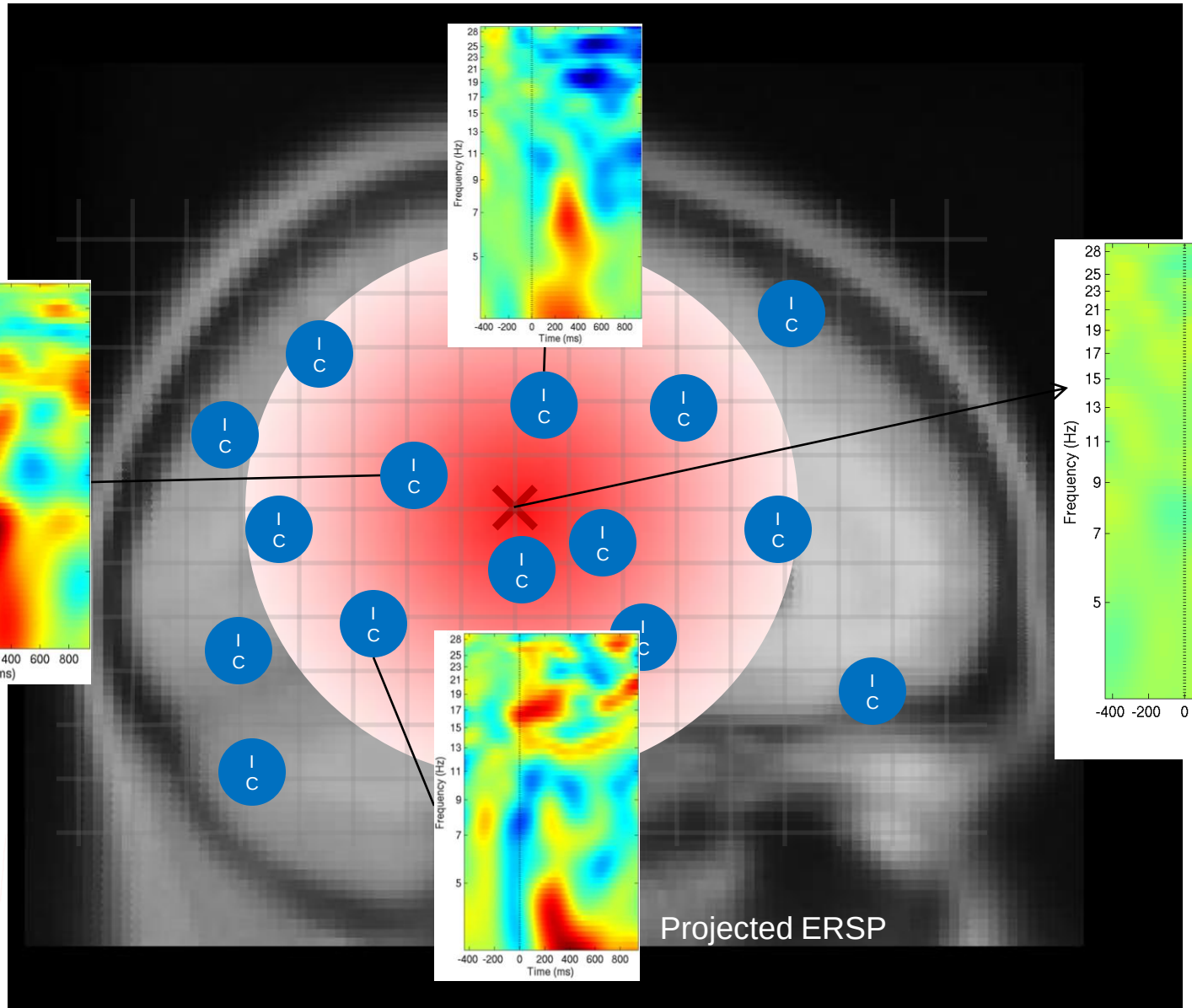
Reassigning components



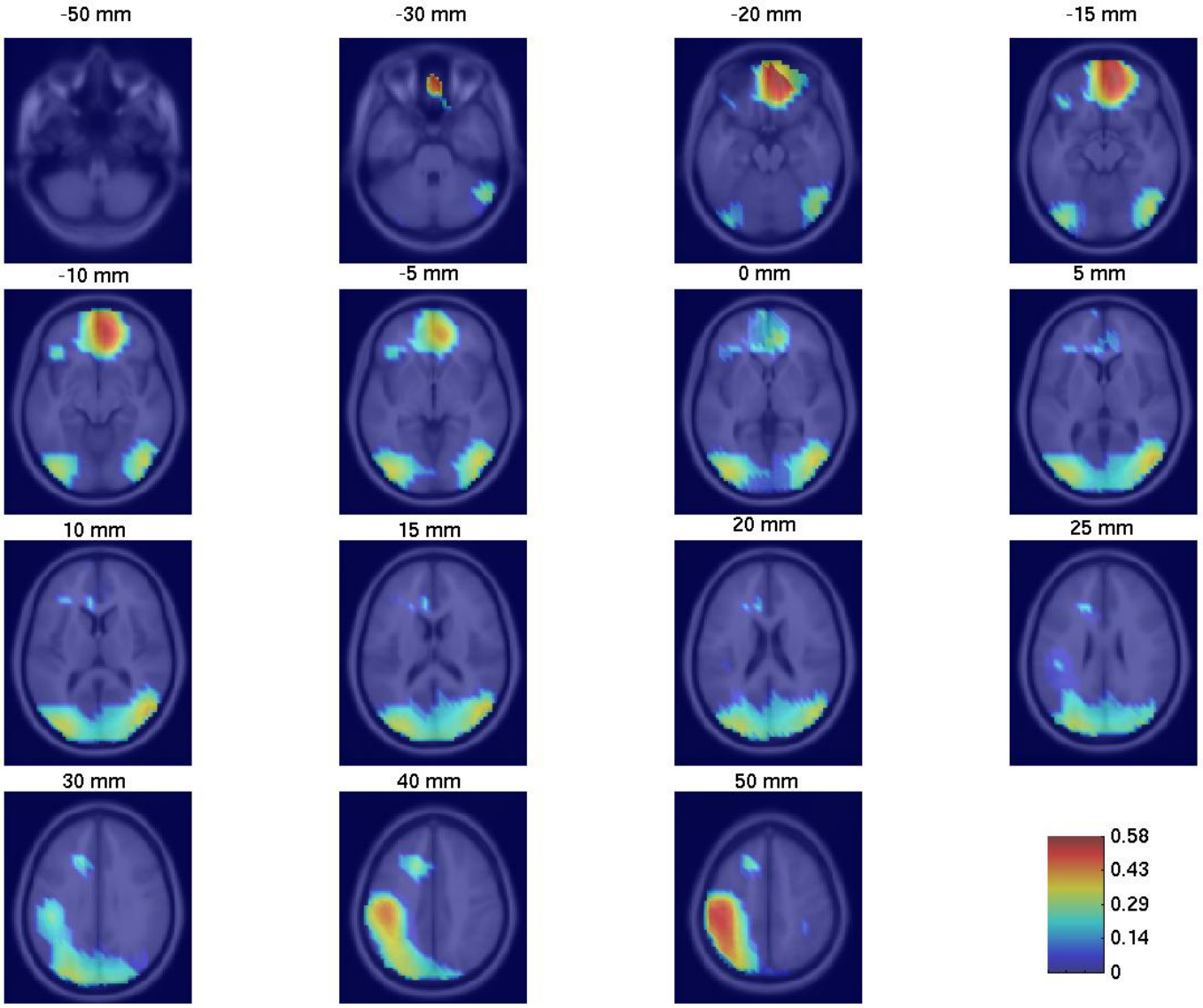
Outlier cluster reassignment



Measure Projection Toolbox



Measure Projection Toolbox



Exercise

Precluster (pre-computation already done) and cluster components using measures of your choice. Experiment with different measures.

Inter iteration Cluster Consistency

Iterations

Clusters	1	2	3	4	5	6	7	8	9	10	Mean
3	100	100	100	100	100	100	100	100	100	100	**100**
4	100	100	100	100	100	100	90	100	100	100	**99**
5	90	40	10	90	90	60	100	10	60	90	**64**
6	60	0	100	60	100	90	60	60	90	60	**68**
7	90	100	90	90	60	90	90	100	90	90	**89**
8	80	80	60	80	40	80	80	80	80	100	**76**
9	60	90	50	60	80	60	0	10	60	50	**52**
10	40	90	10	40	0	50	50	0	50	60	**39**
11	60	20	0	0	10	60	10	90	60	60	**37**
12	100	50	50	100	50	100	100	50	100	50	**75**
13	50	10	20	50	90	50	50	10	50	20	**40**
14	20	10	10	20	20	30	20	20	30	30	**21**