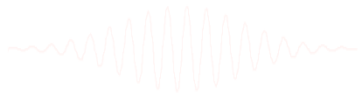


Clustering of ICA components

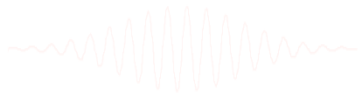
Arnaud Delorme

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



Steps of clustering

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



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 >

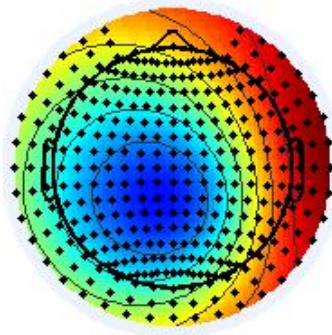
☐ Dataset info (condition, group, ...) differs from study info. [set] = Overwrite dataset info.

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

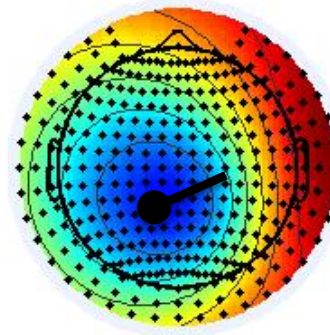
Help Cancel Ok

Computing residual variance (%)

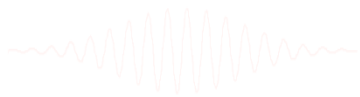
Actual



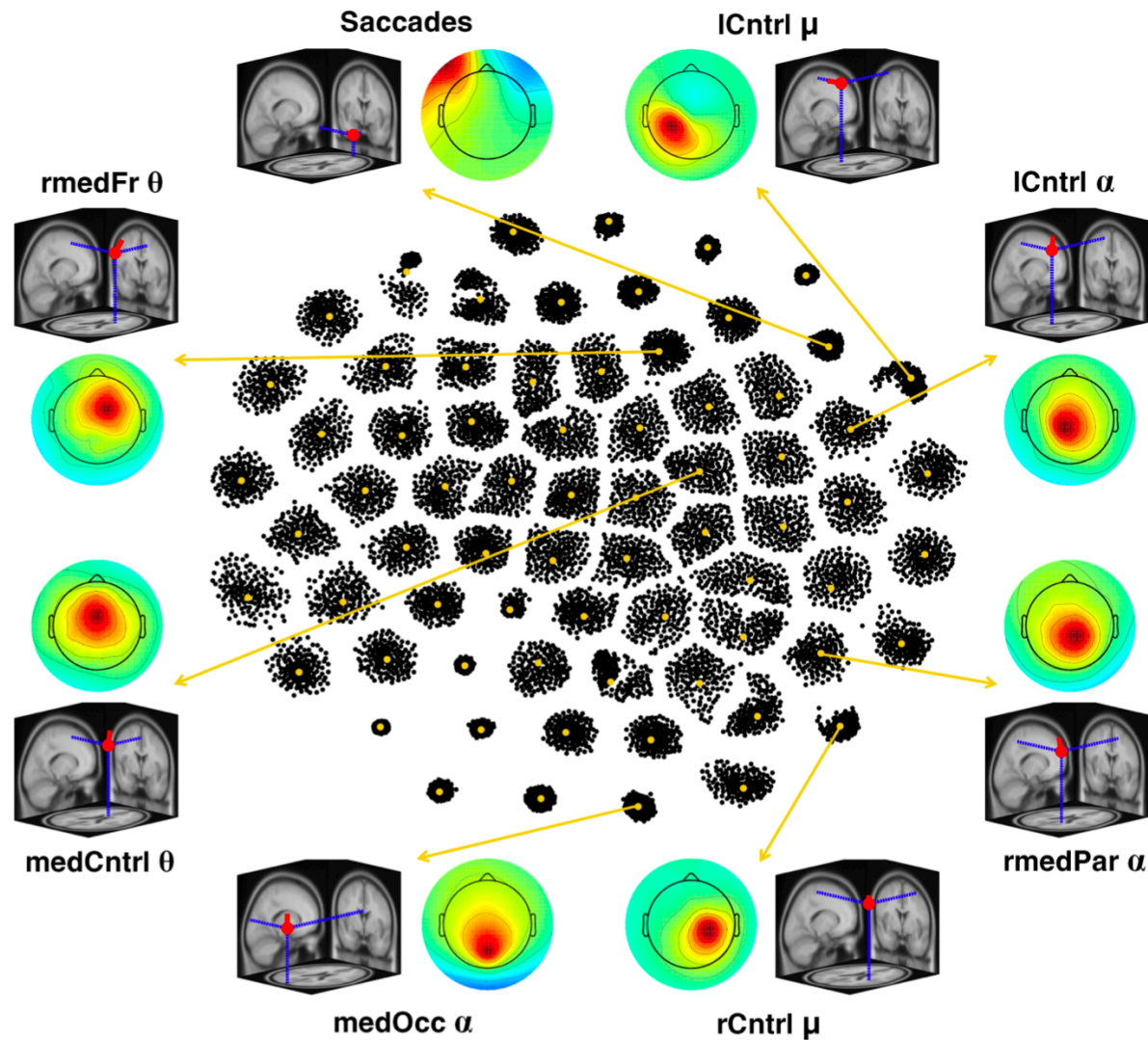
Dipole projection



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



Clustering results example

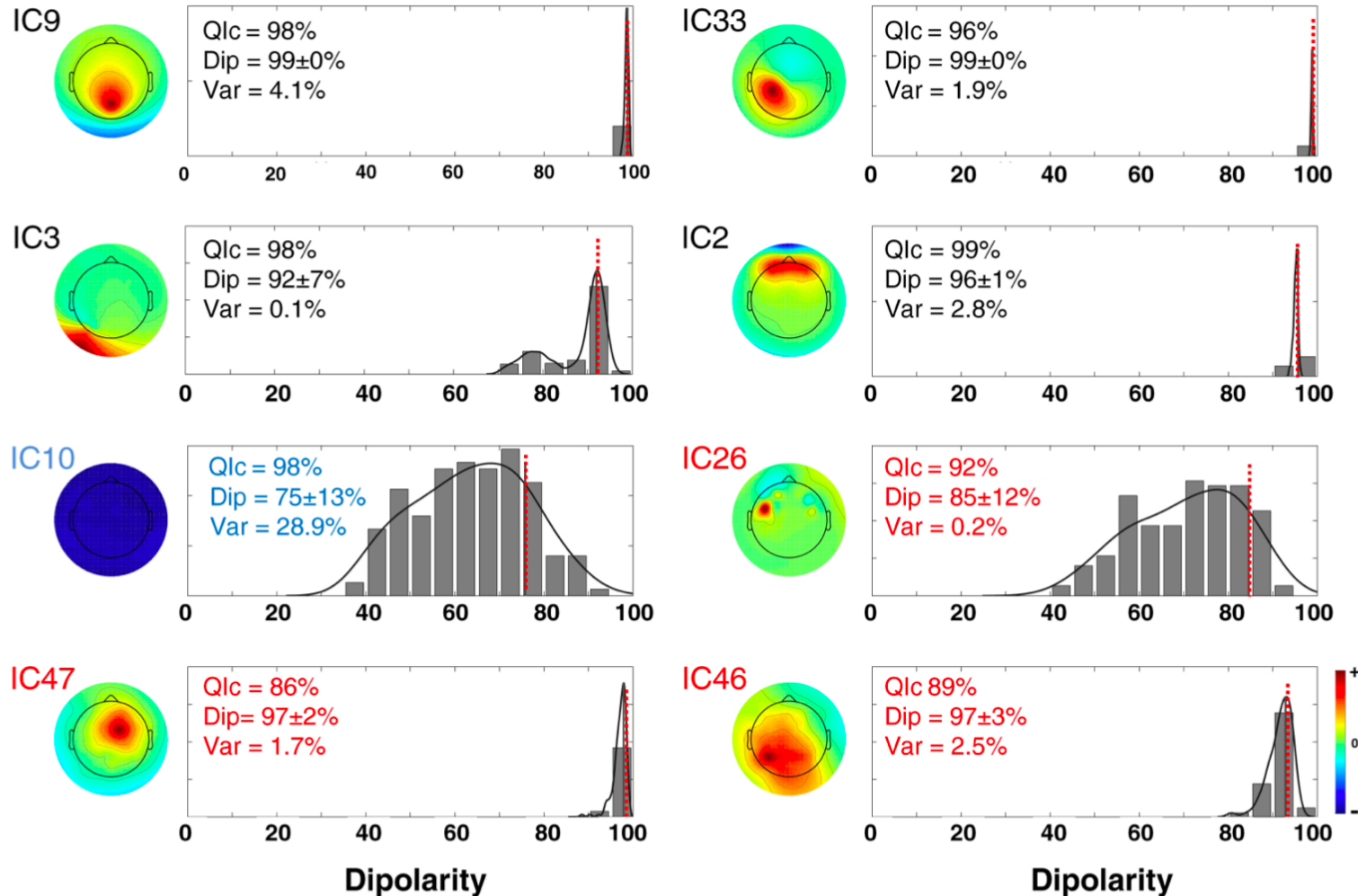


RELICA: A method for estimating the reliability of independent components

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

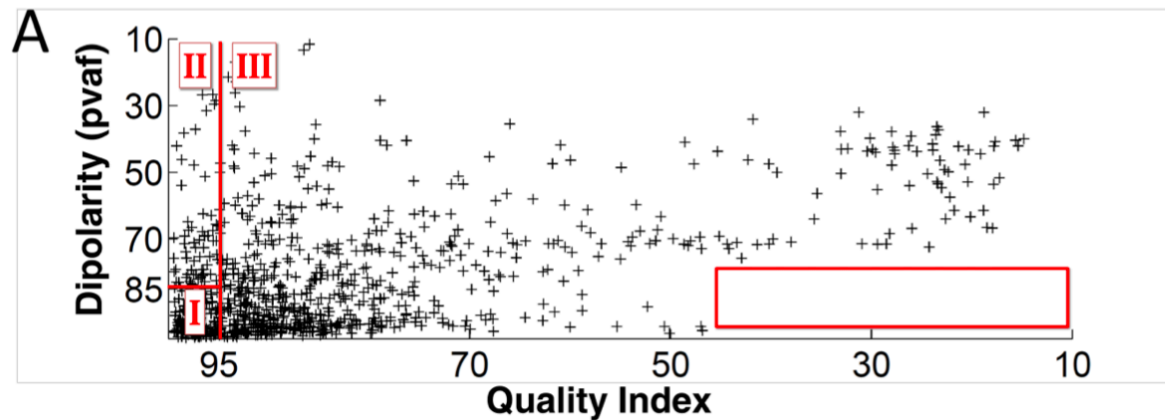
Within-cluster reliability

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

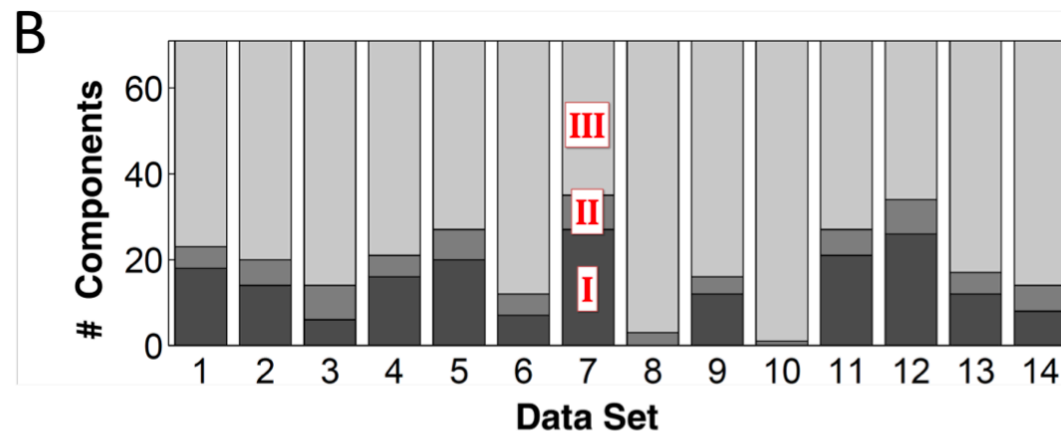
CLASS III

Quality Index below retention threshold

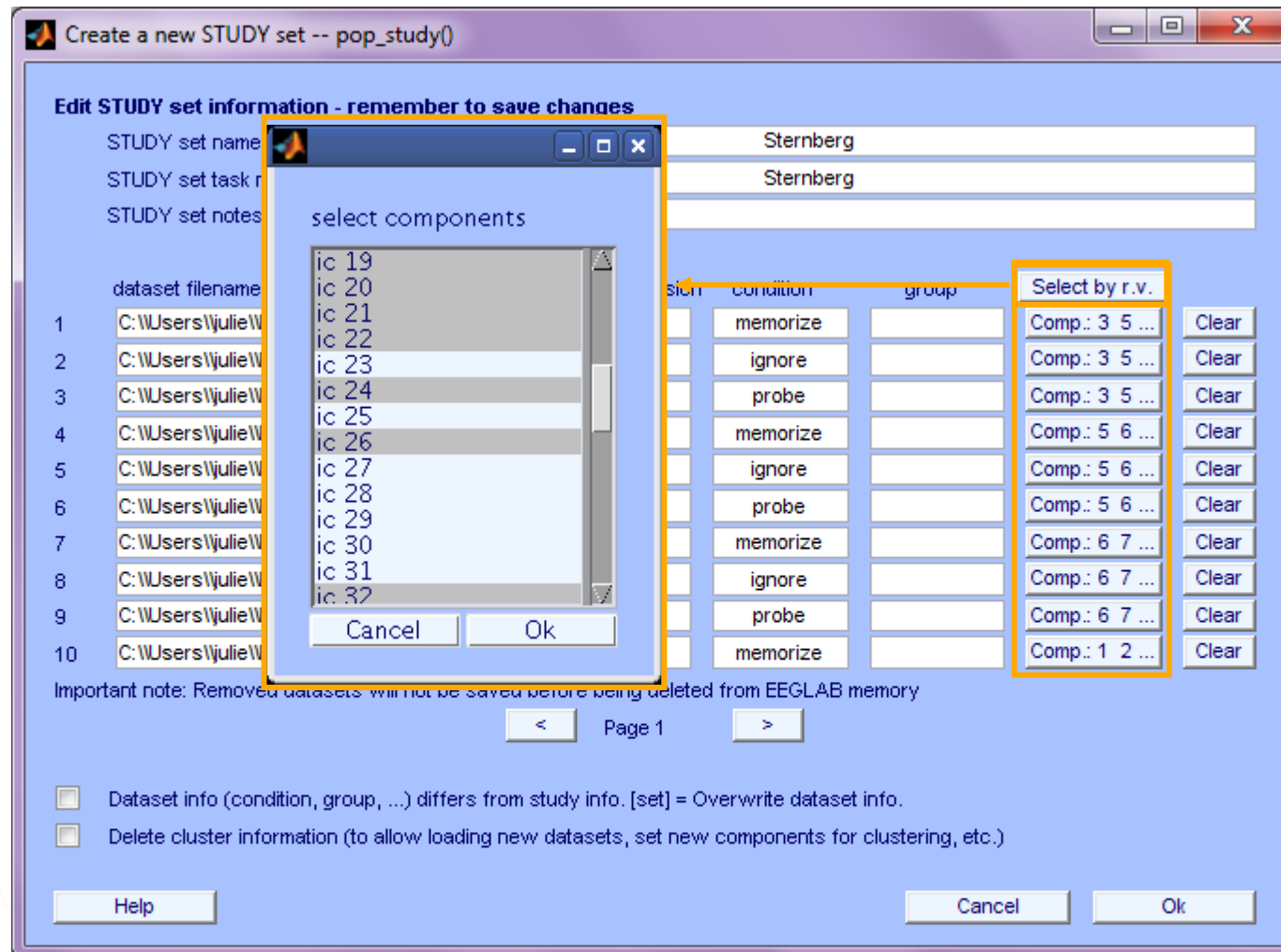
$$dip \pm std > th$$

$$dip \pm std < th$$

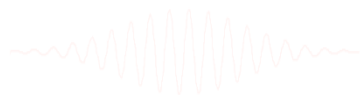
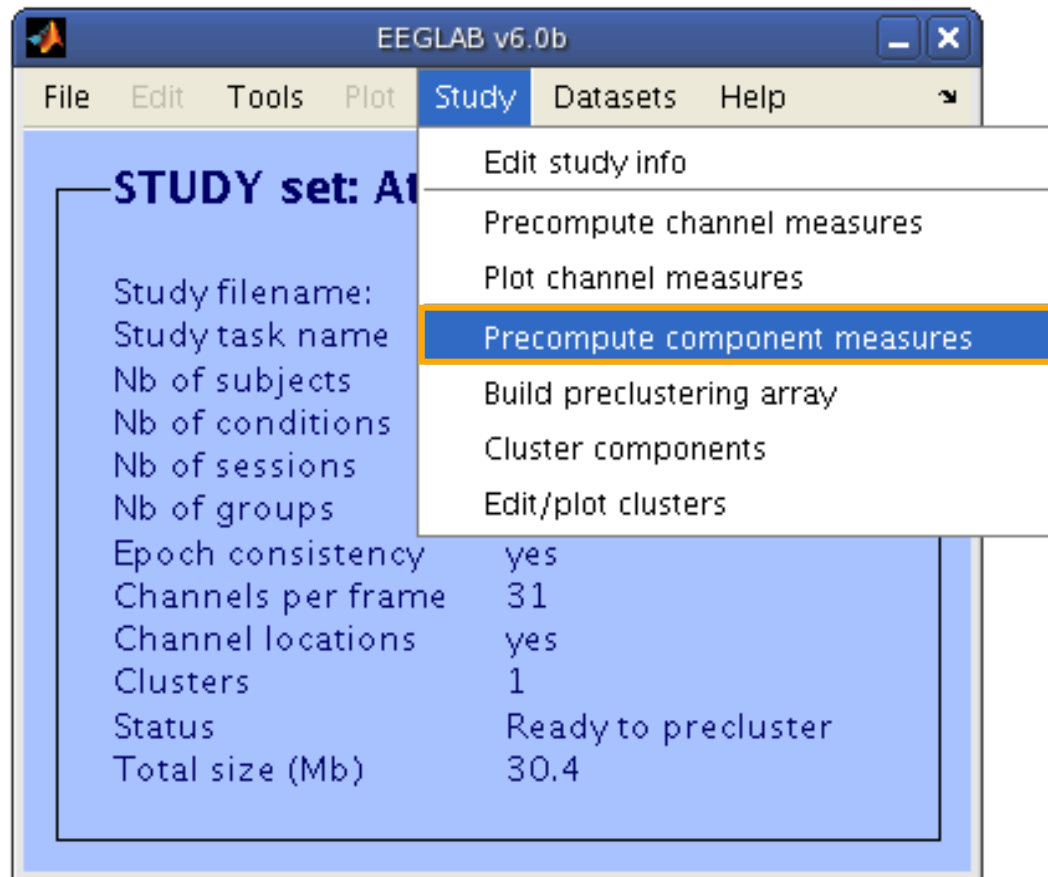
Probable Inseparable **Discard**
noise: variance
explained useful or
**multiple subject
confirmation**



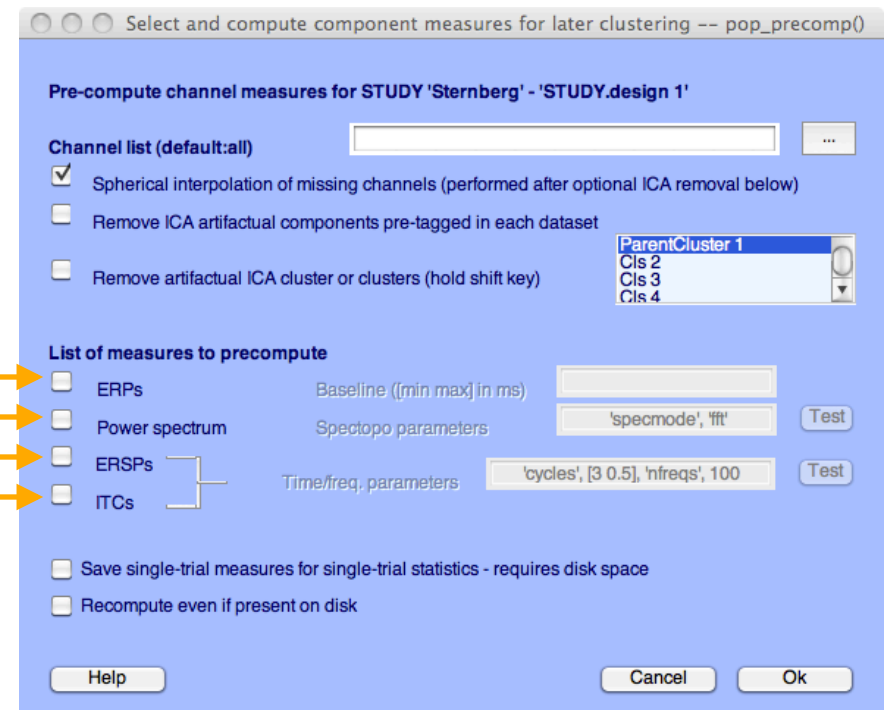
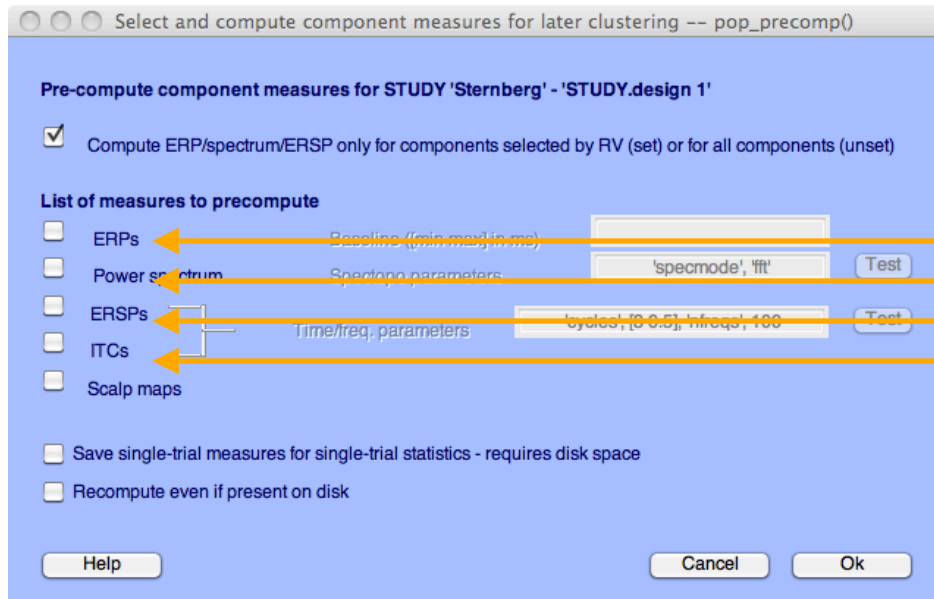
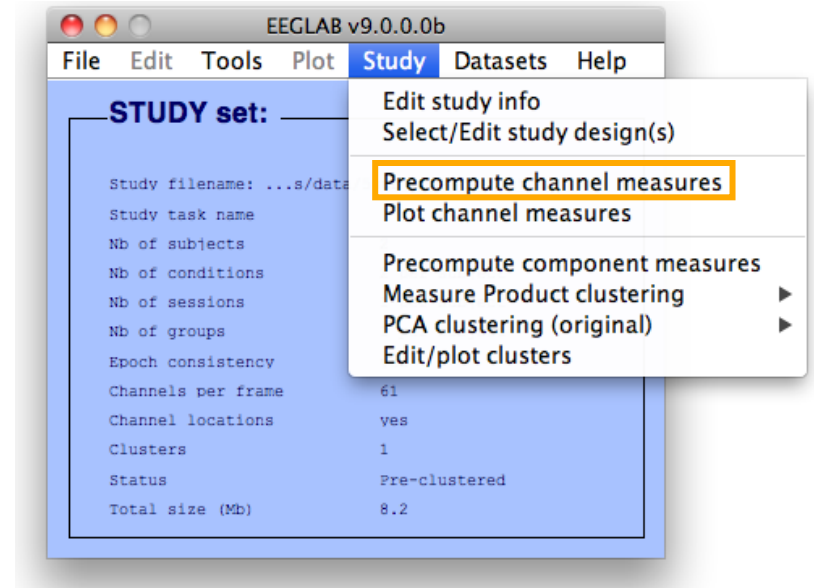
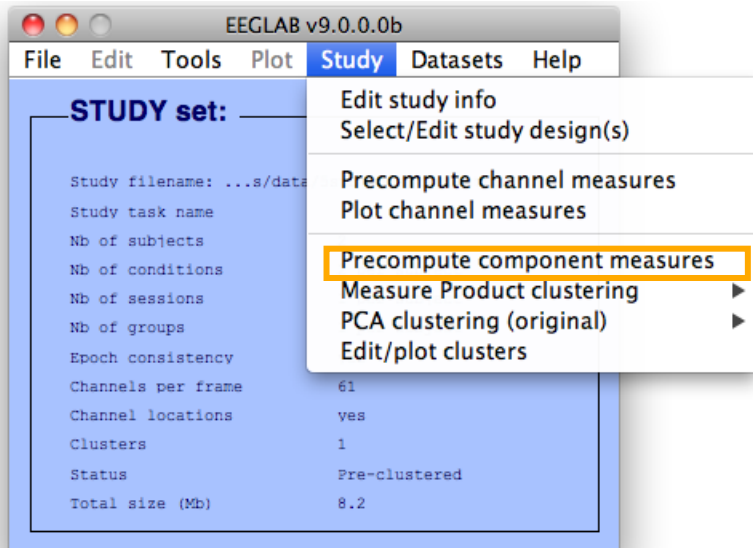
ICs to cluster



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

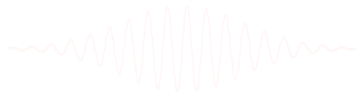
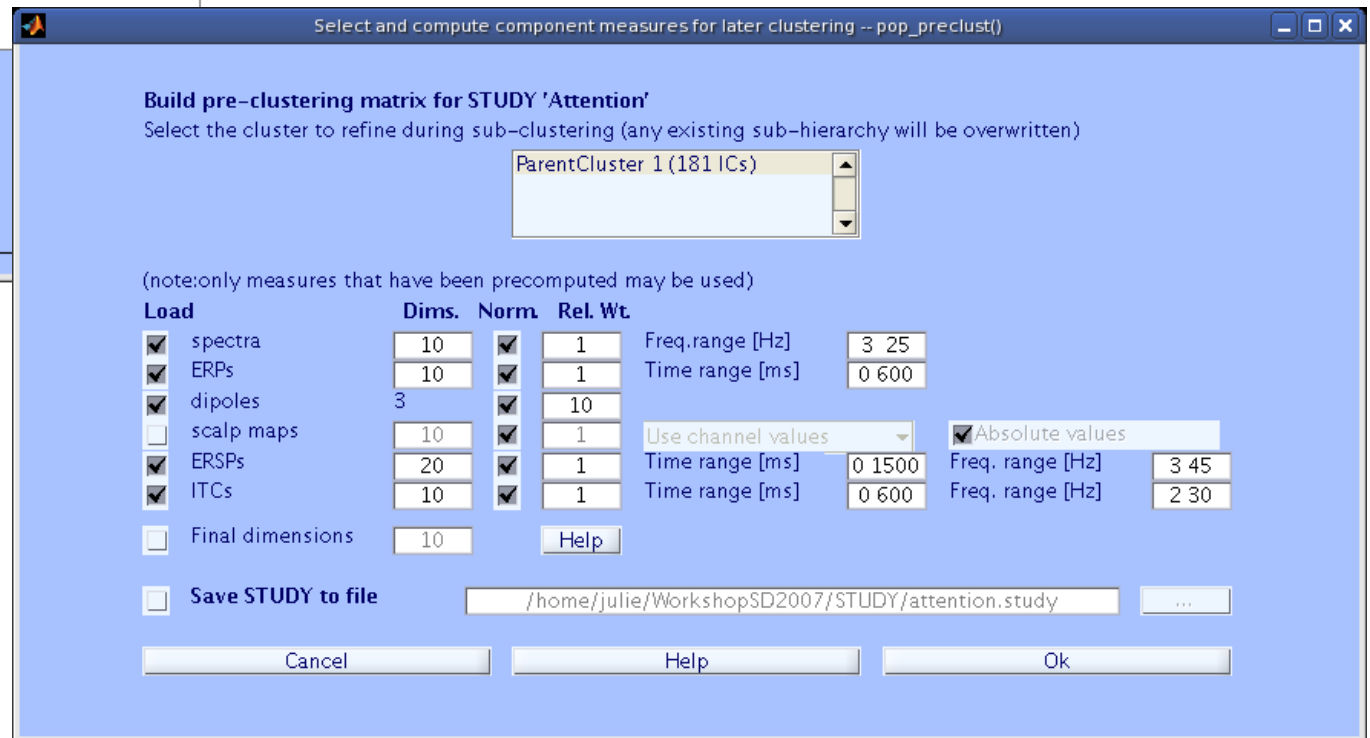
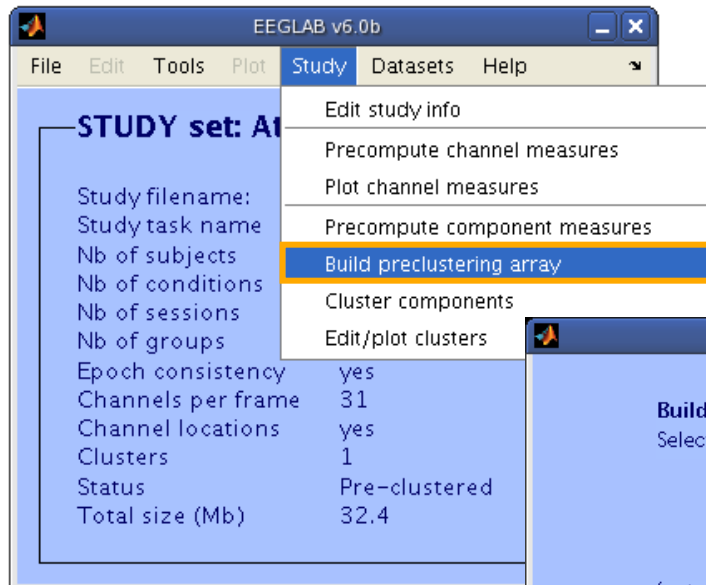
<input checked="" type="checkbox"/> ERPs	Baseline ([min max] in ms)	<input type="text" value="[-200 0]"/>	
<input checked="" type="checkbox"/> Power spectrum	Spectopo parameters	<input type="text" value=""/>	<input type="button" value="Test"/>
<input checked="" type="checkbox"/> ERSPs	Time/freq. parameters	<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

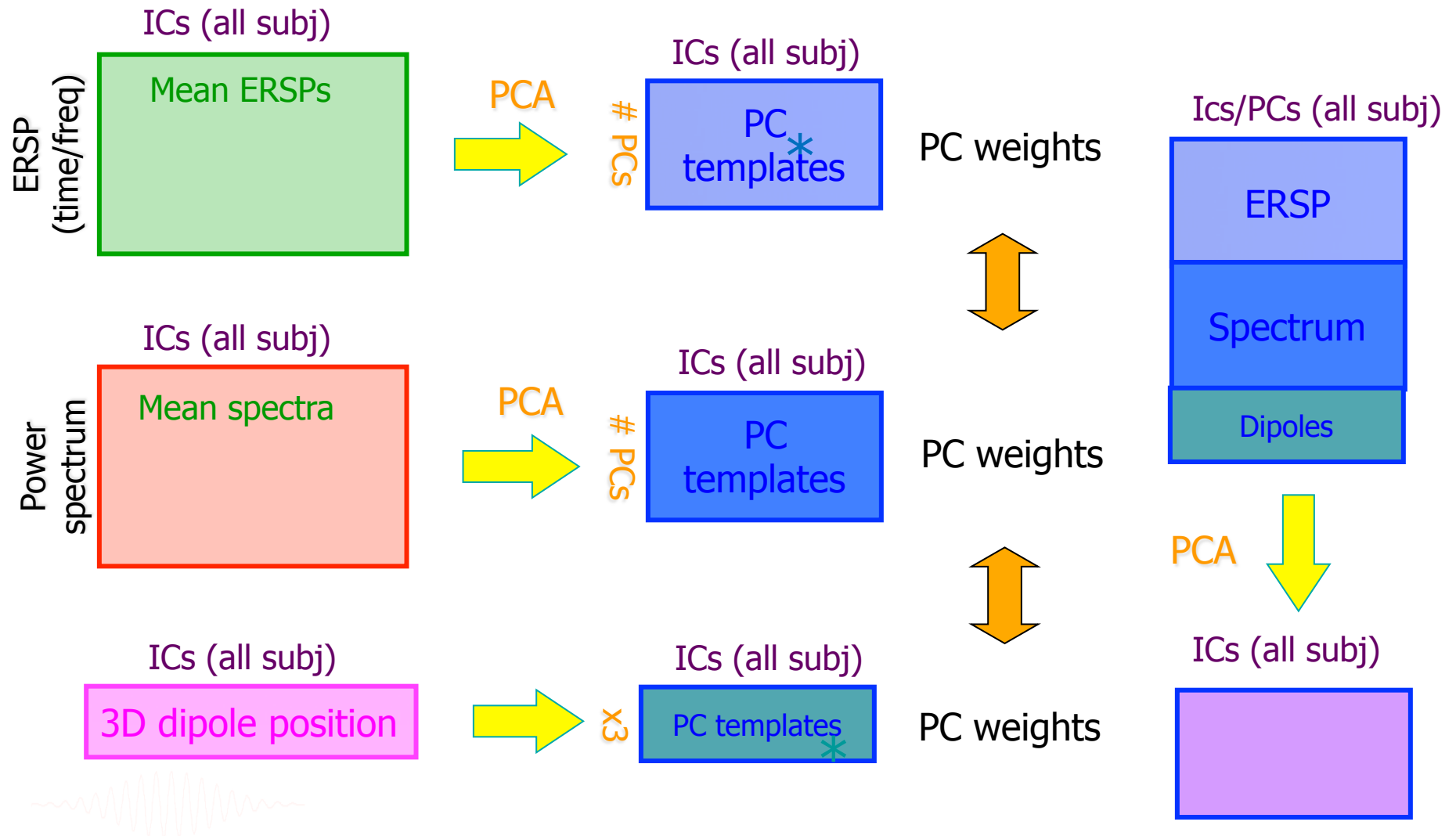
Time-frequency options



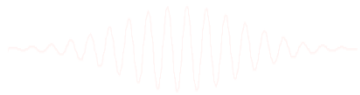
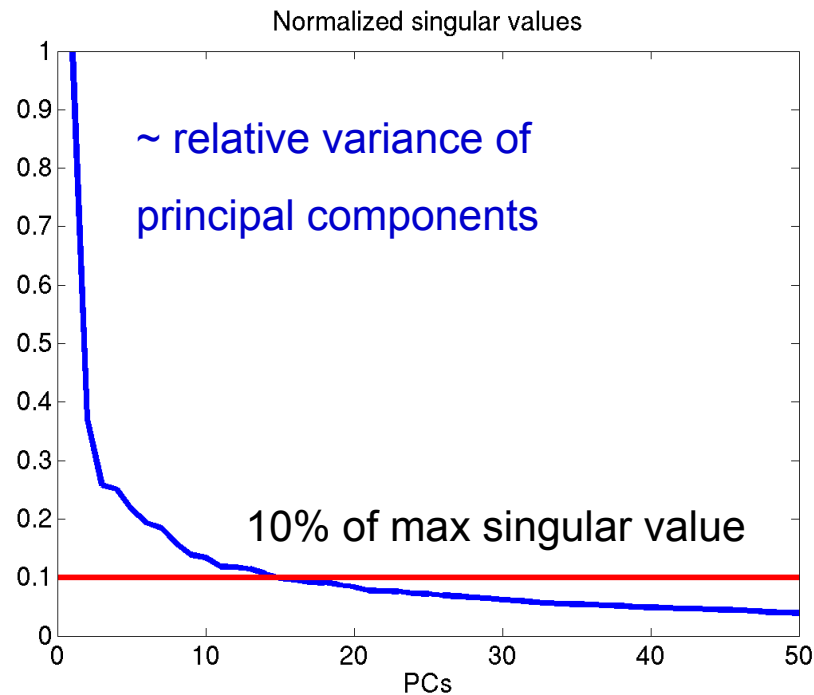
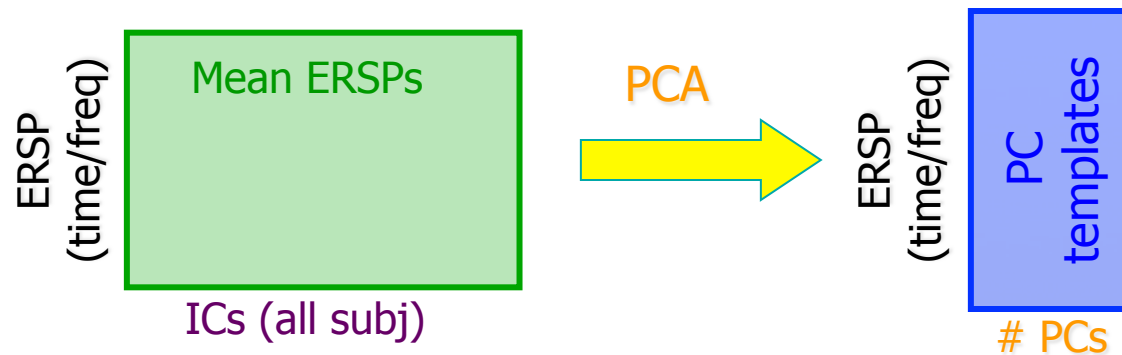
Cluster components



Precluster schematic



Precluster: Use singular values from PCA



Precluster schematic

ICs (all subj)

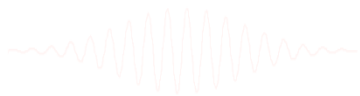
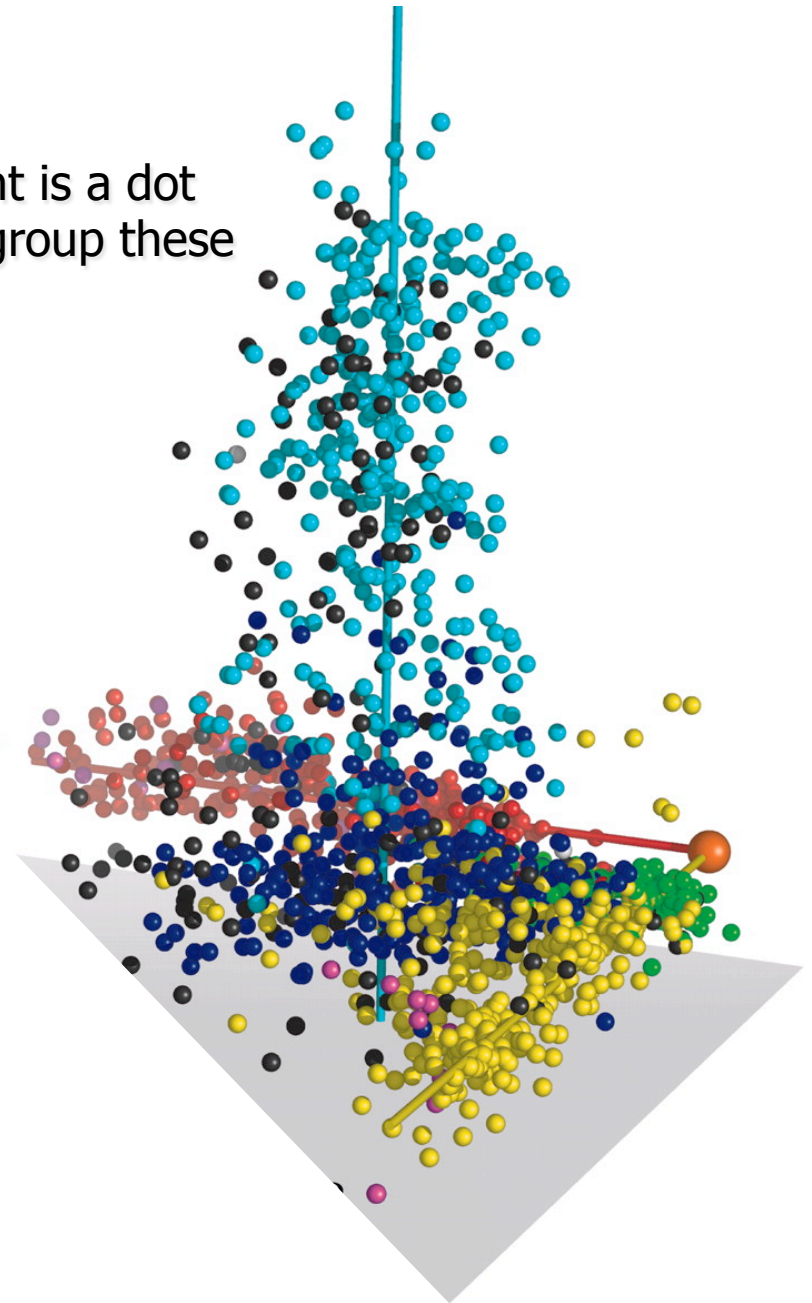


OR

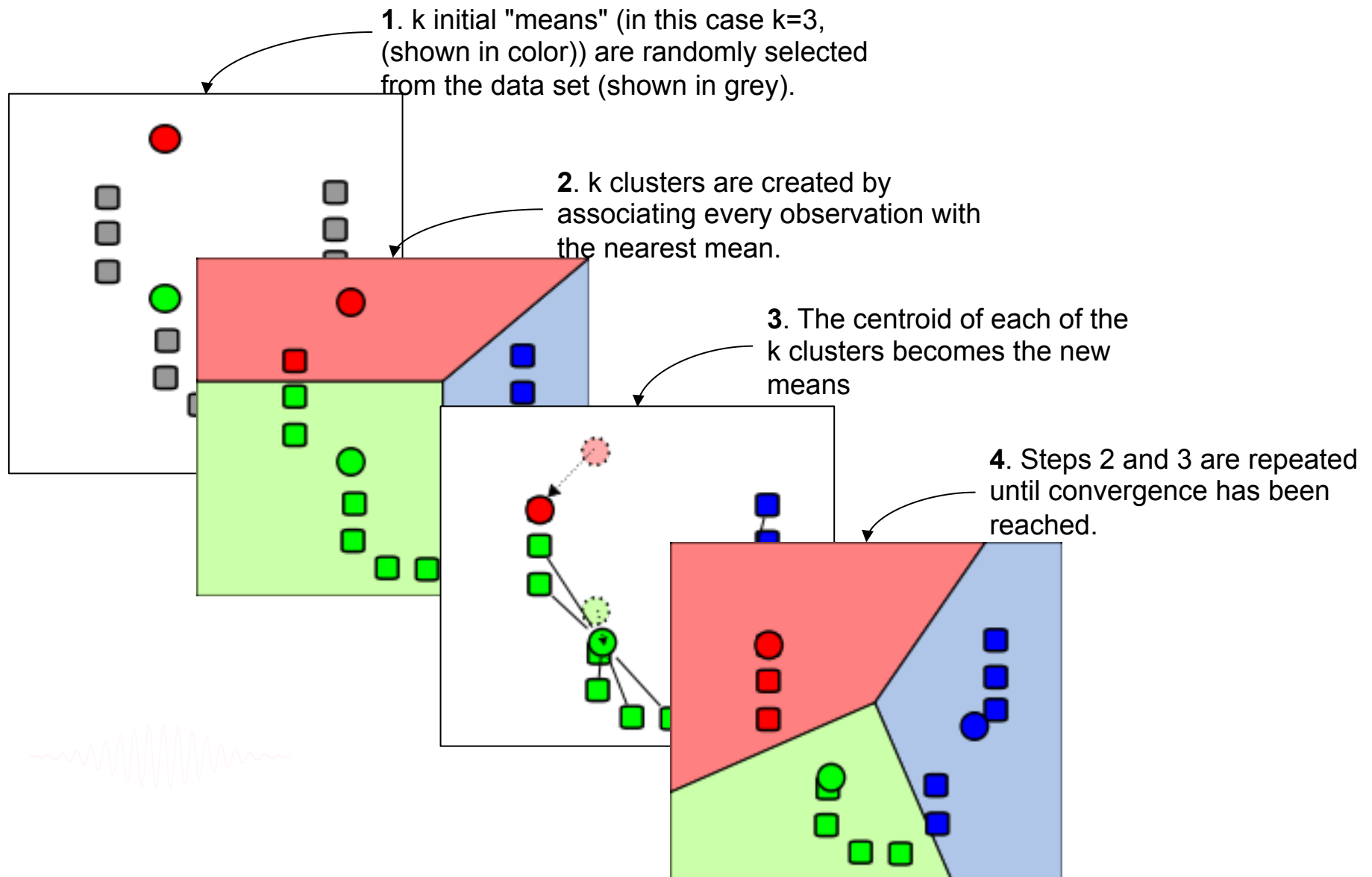
ICs (all subj)



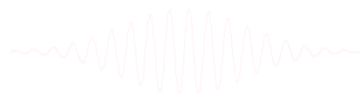
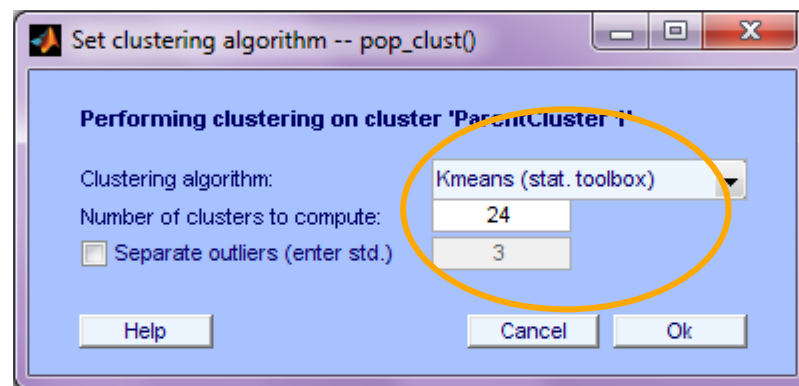
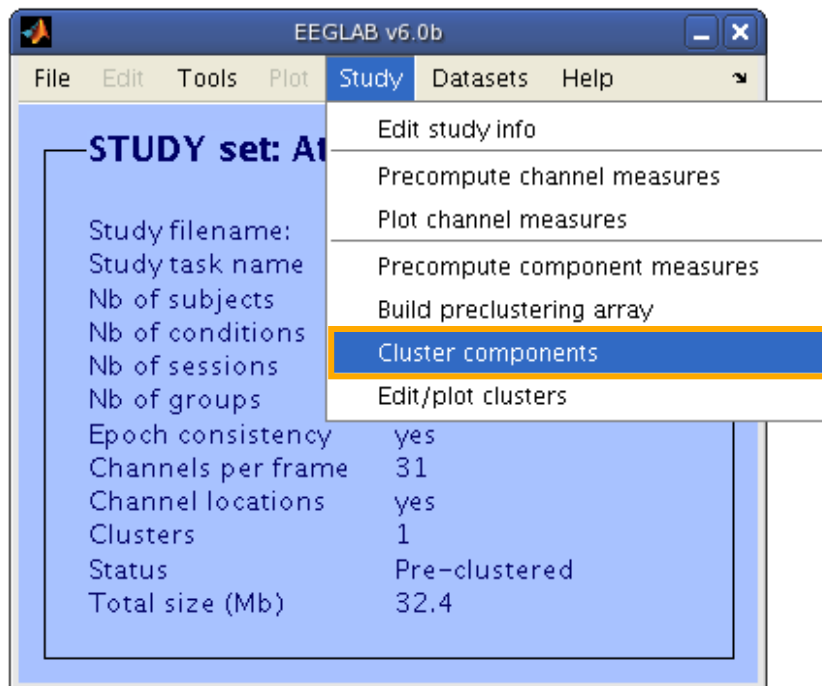
Each component is a dot
Clustering will group these
dots



Classical KMean



Cluster components



Choosing data measures

What measure(s) should you use?

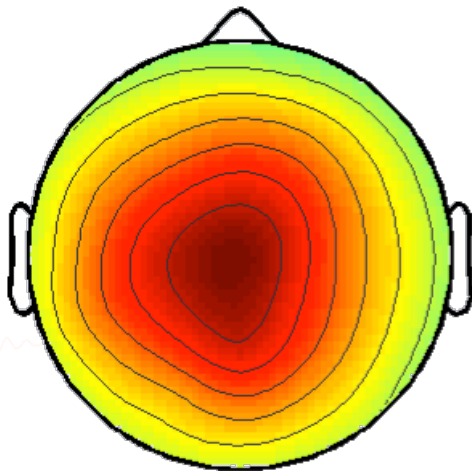
It depends on your final cluster criteria...

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

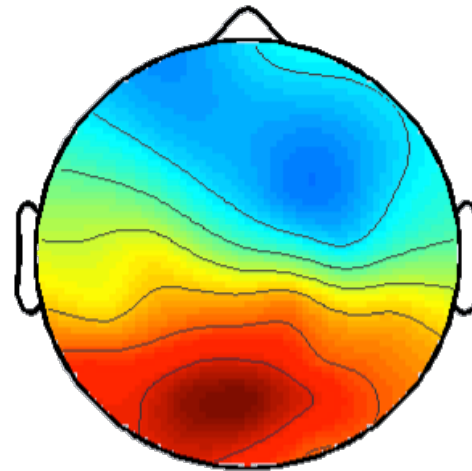
But consider:

- What is the difference between these two components?

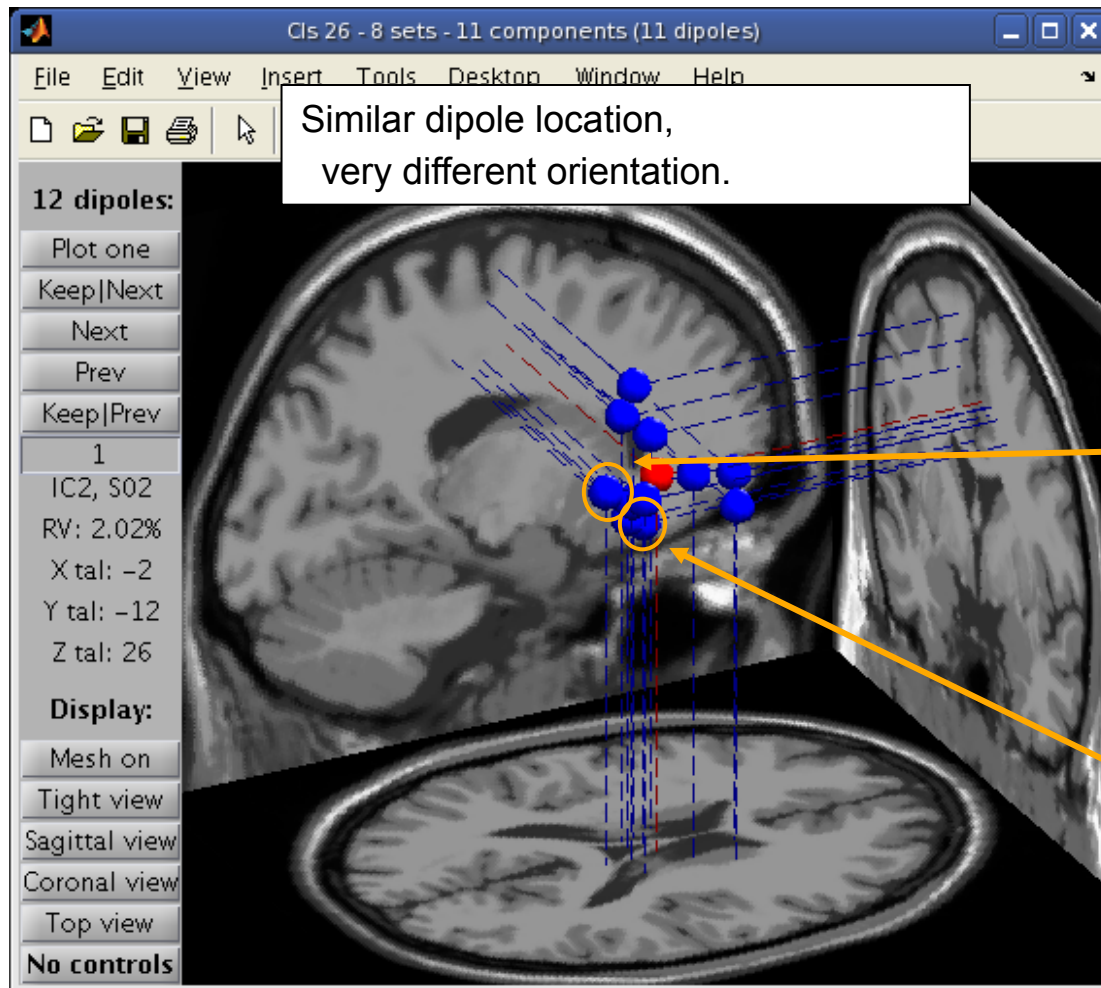
IC2 / S02, Cls 26



IC5 / S05, Cls 26

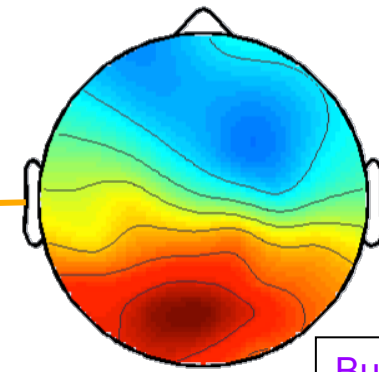


Choosing data measures

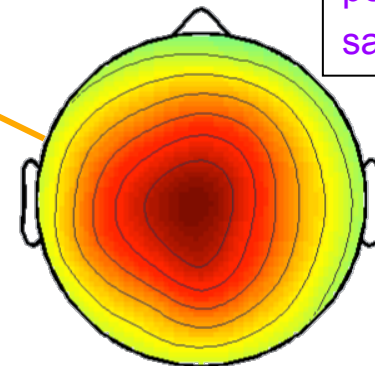


Obvious dramatic effect on scalp map topography:

IC5 / S05, Cls 26

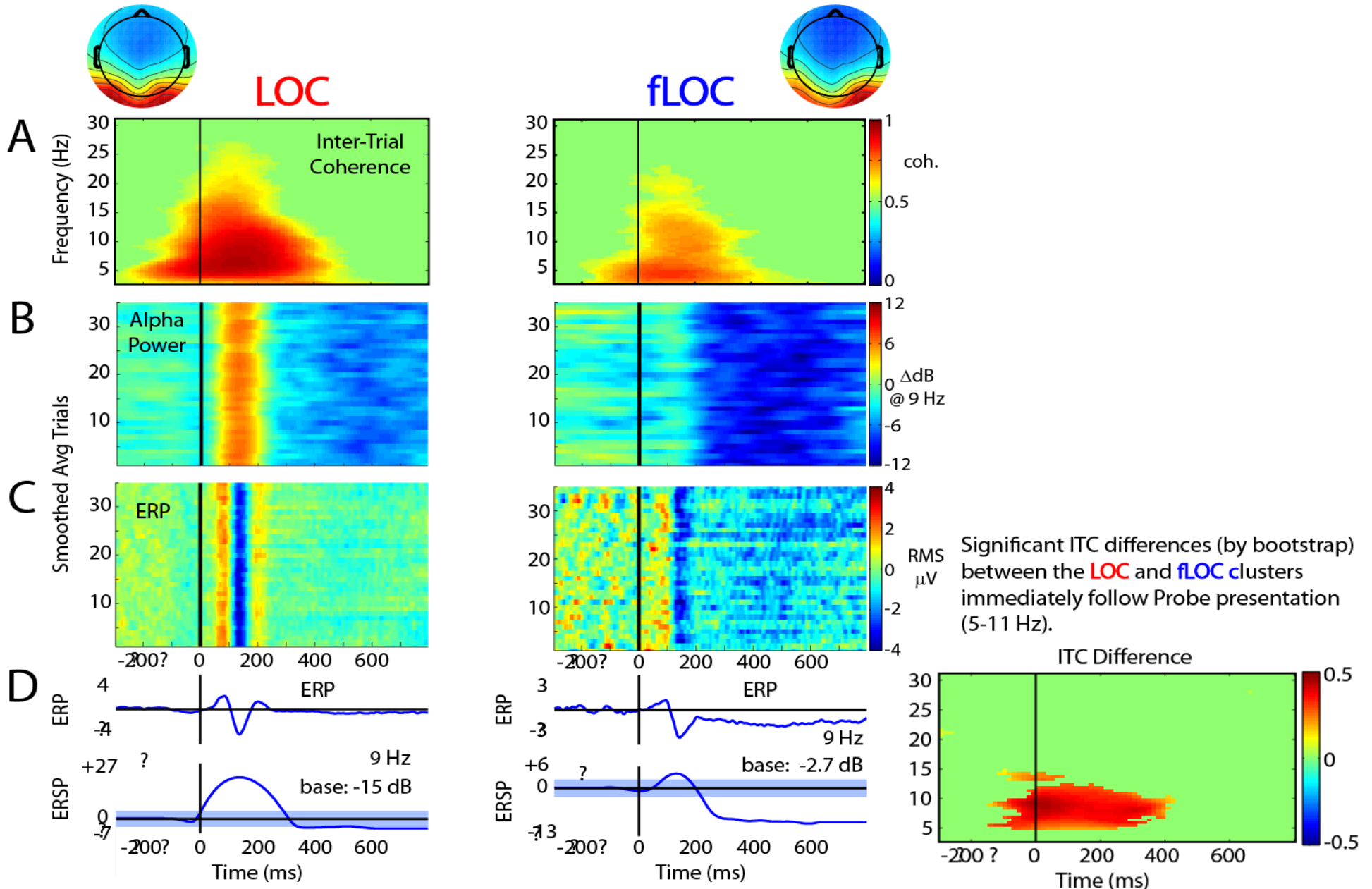


IC2 / S02, Cls 26

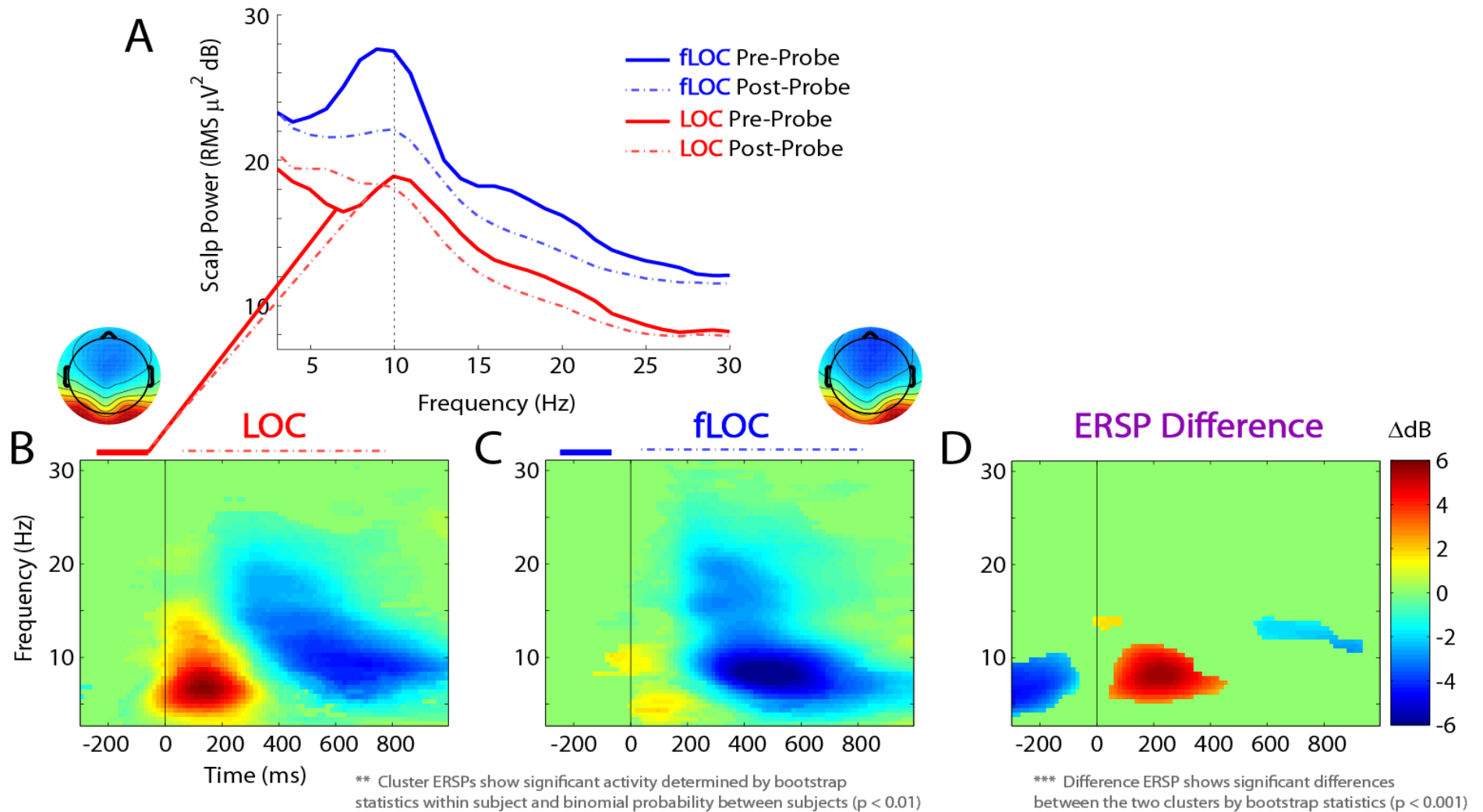


But, do they perform the same functions?

Subject differences?



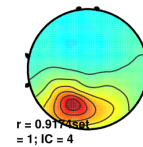
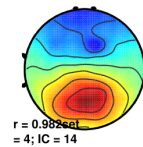
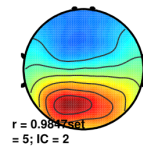
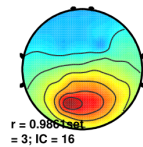
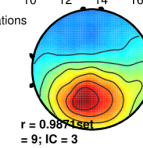
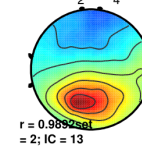
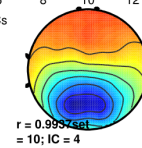
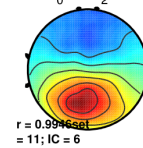
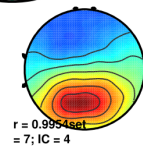
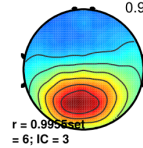
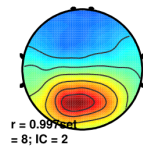
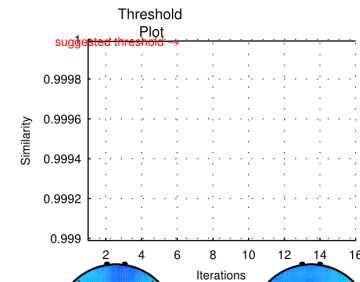
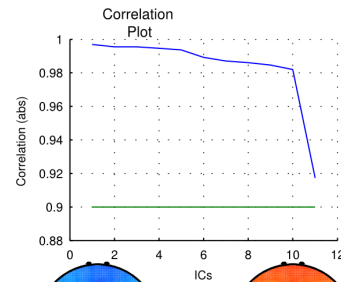
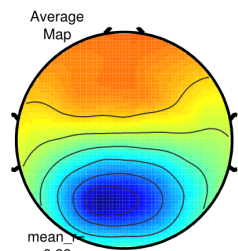
Subject differences?



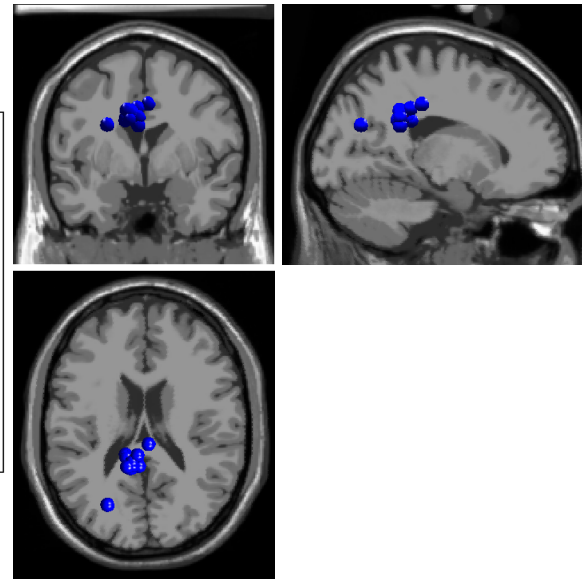
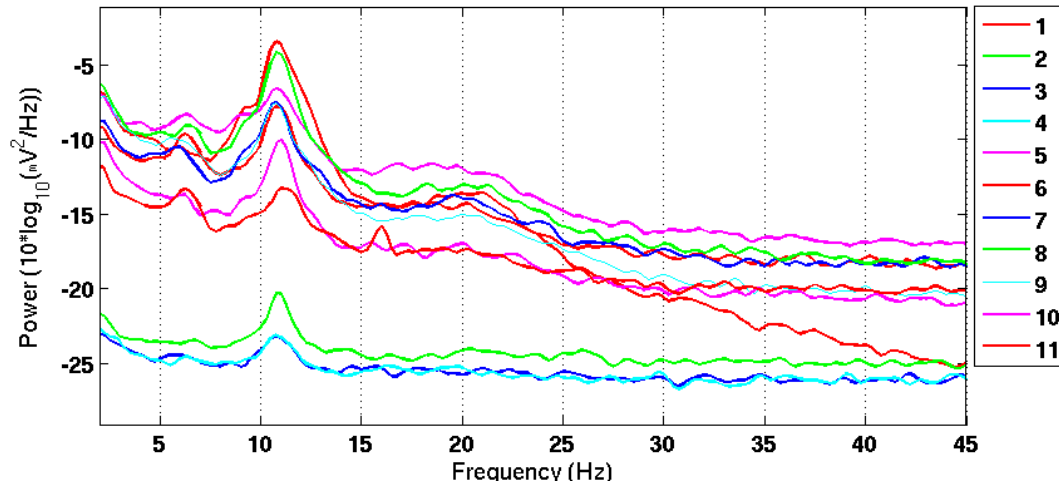
Results (Cluster 1)

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

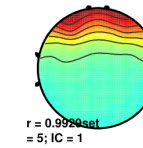
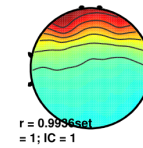
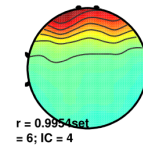
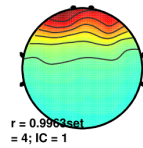
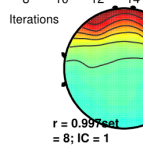
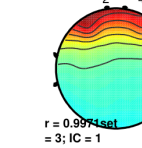
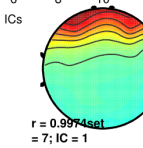
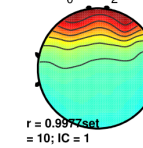
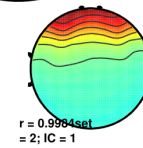
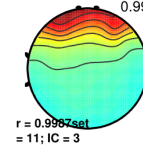
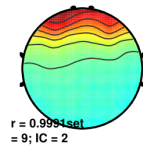
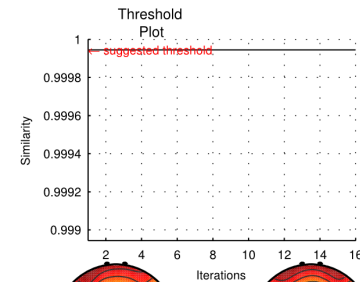
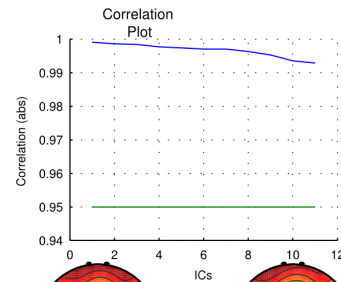
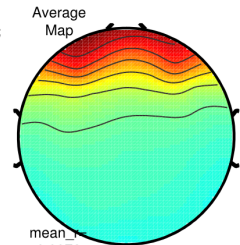


Results (Cluster 2)

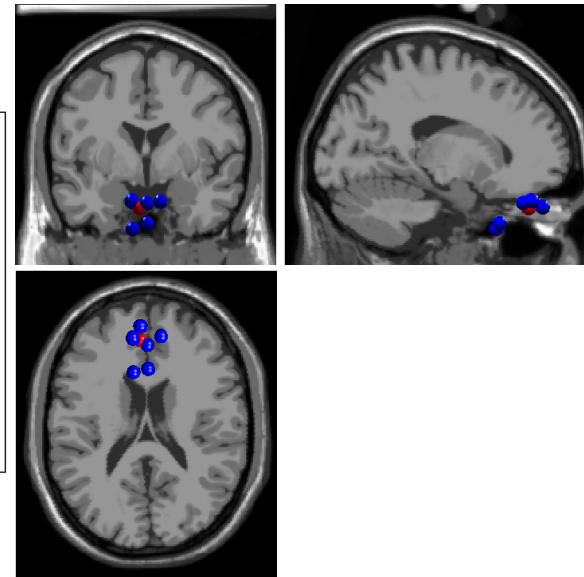
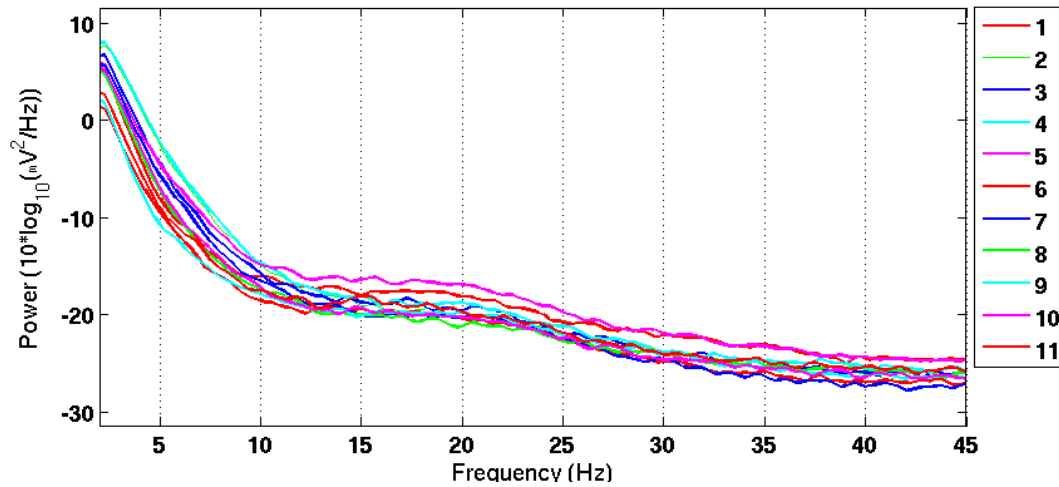
100 % Sessions contribute

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



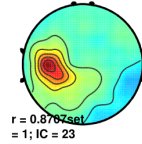
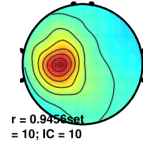
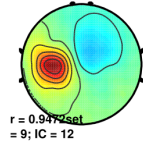
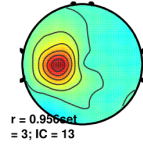
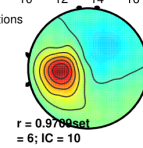
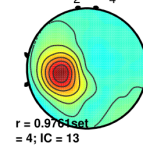
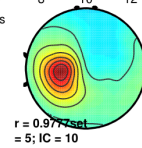
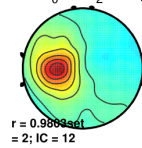
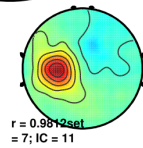
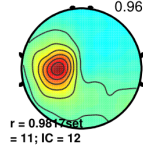
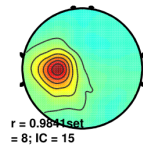
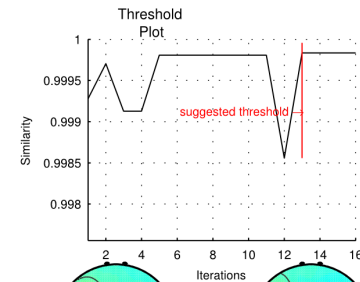
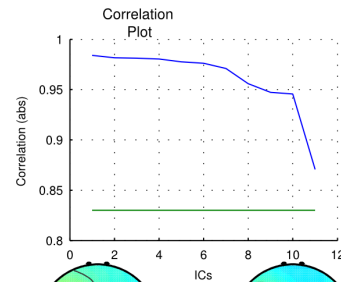
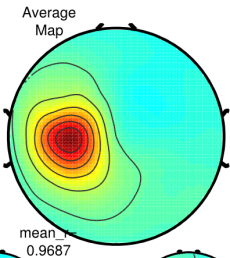
Cls 4 Spectrum



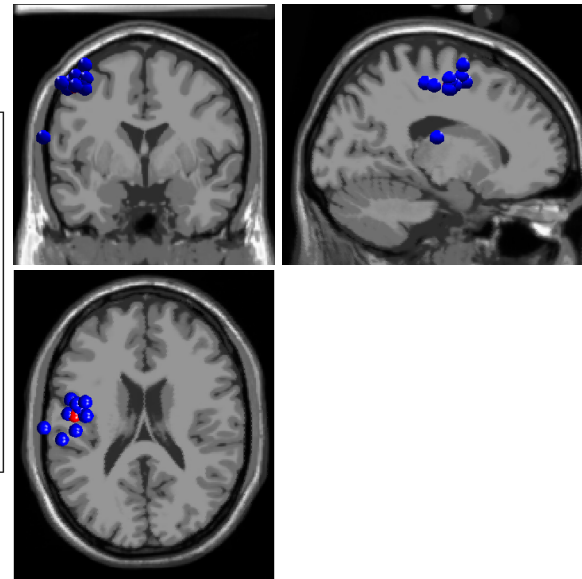
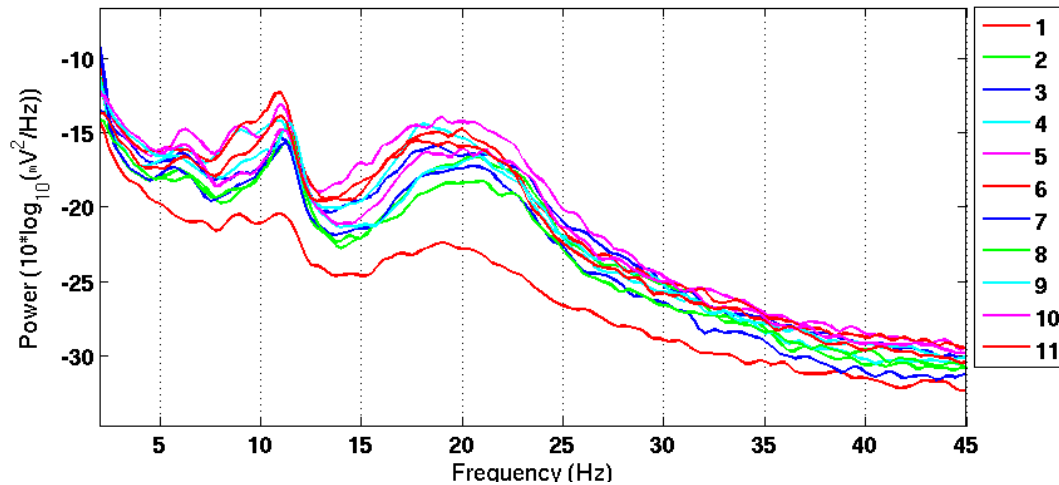
Results (Cluster 8)

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

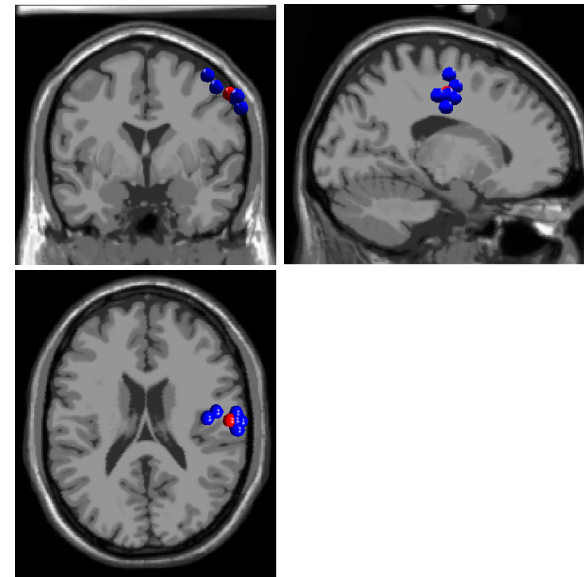
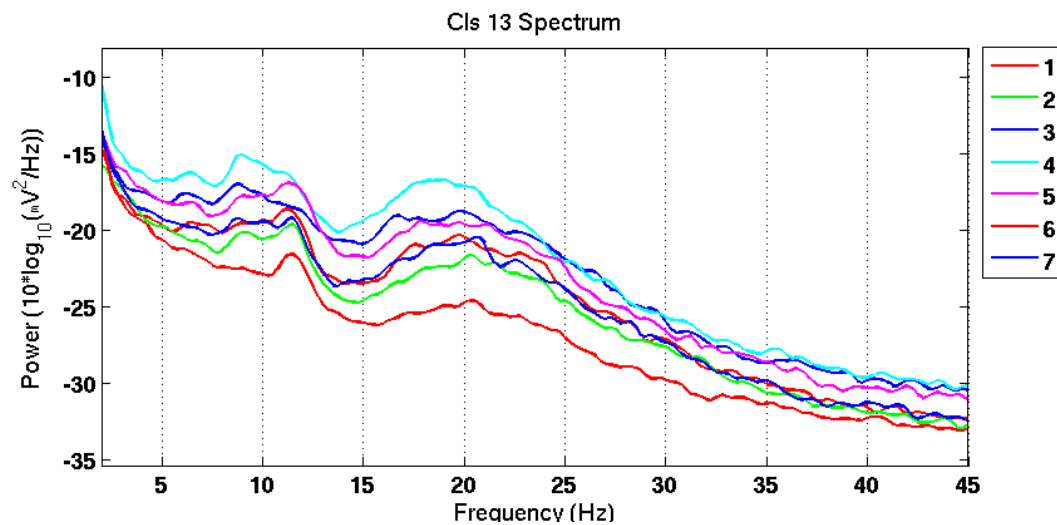
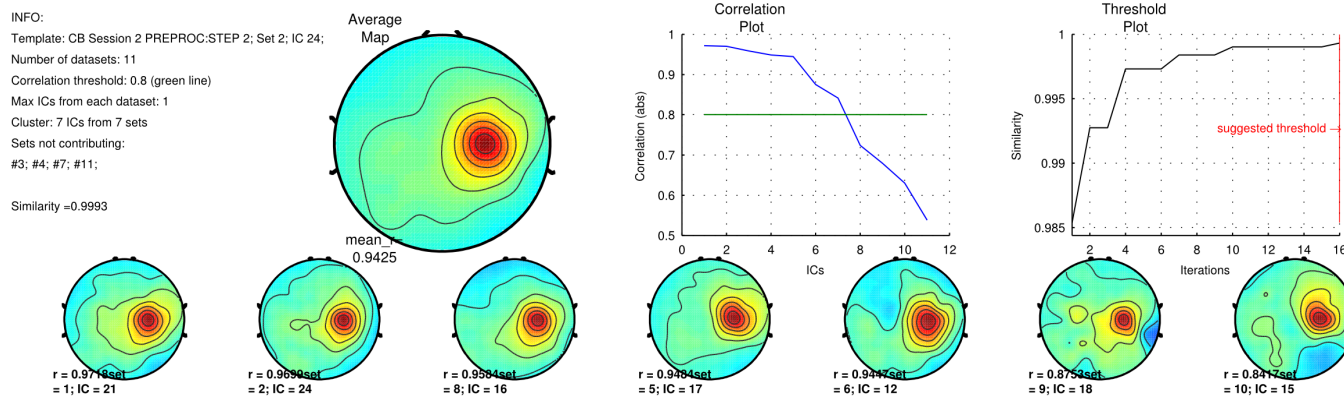


Cls 8 Spectrum



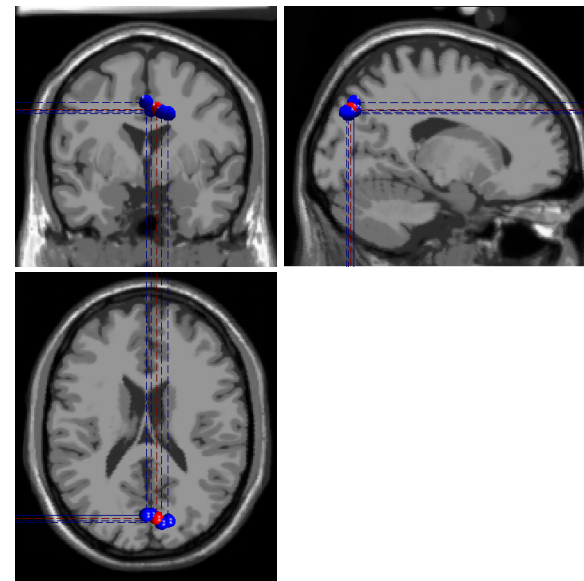
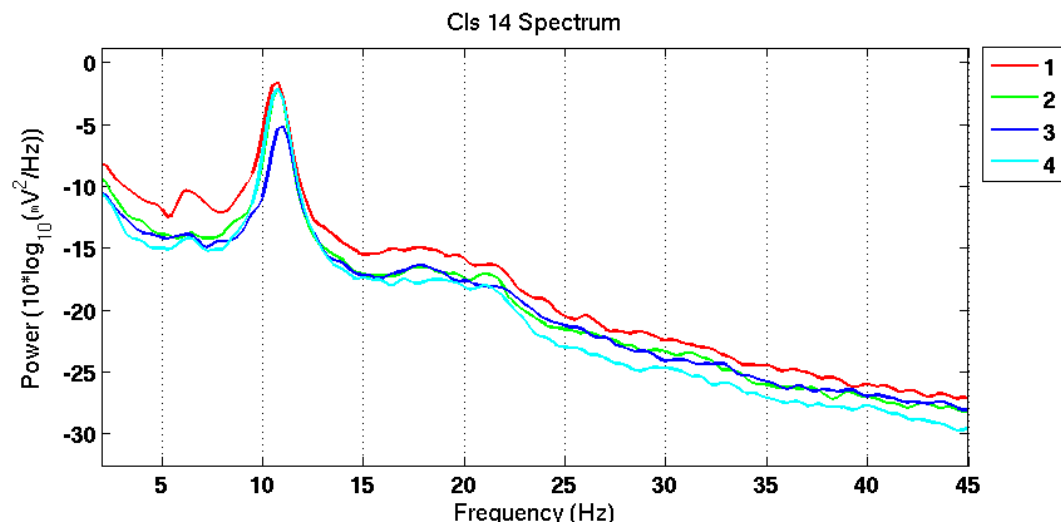
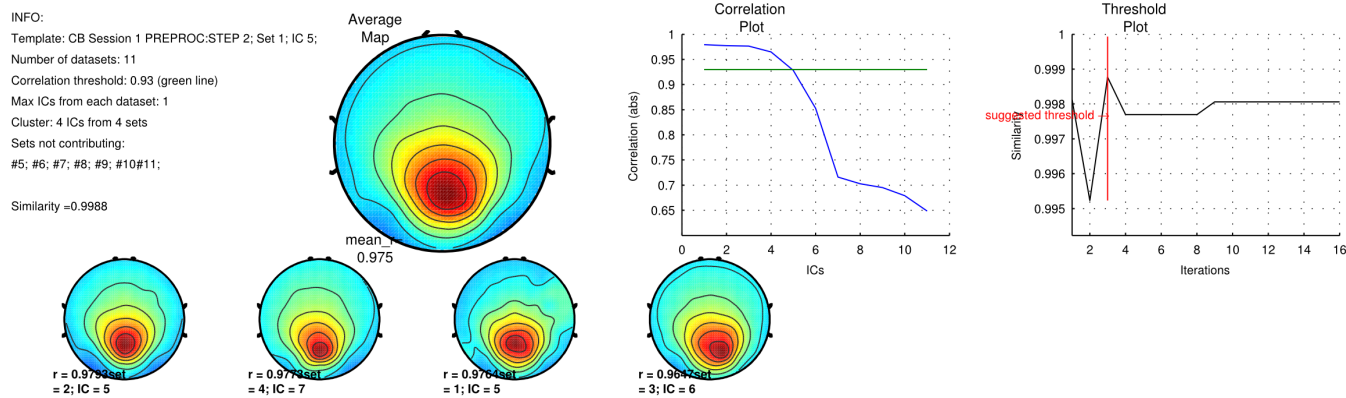
Results (Cluster 13)

63.64% Sessions contribute

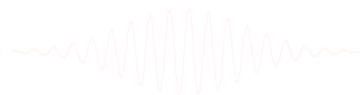
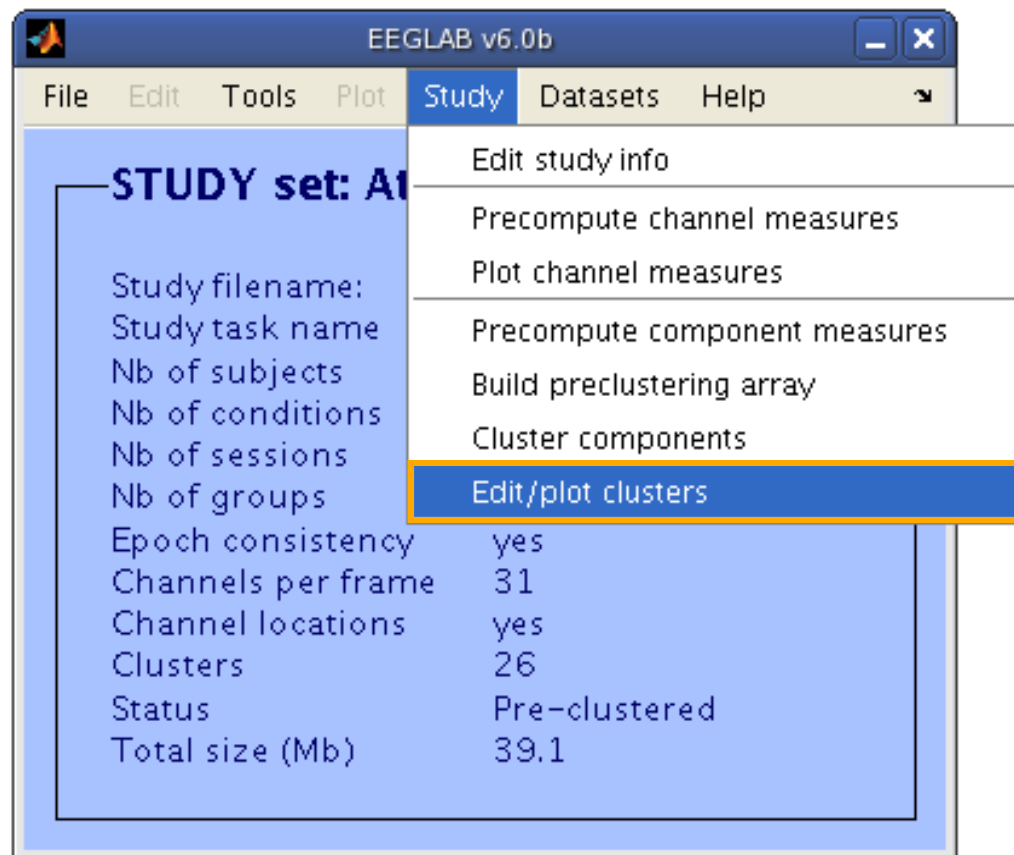


Results (Cluster 14)

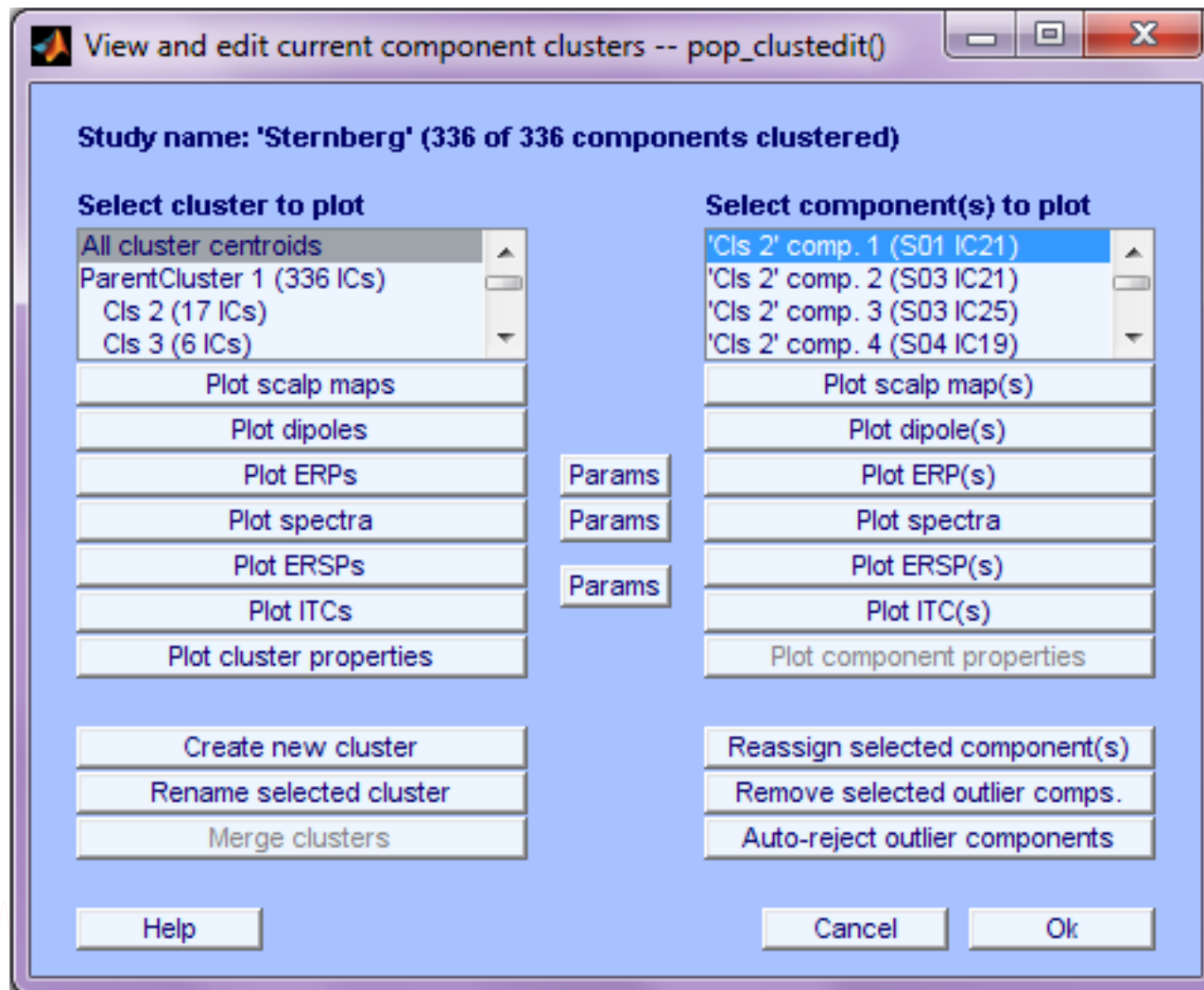
36.36% Sessions contribute



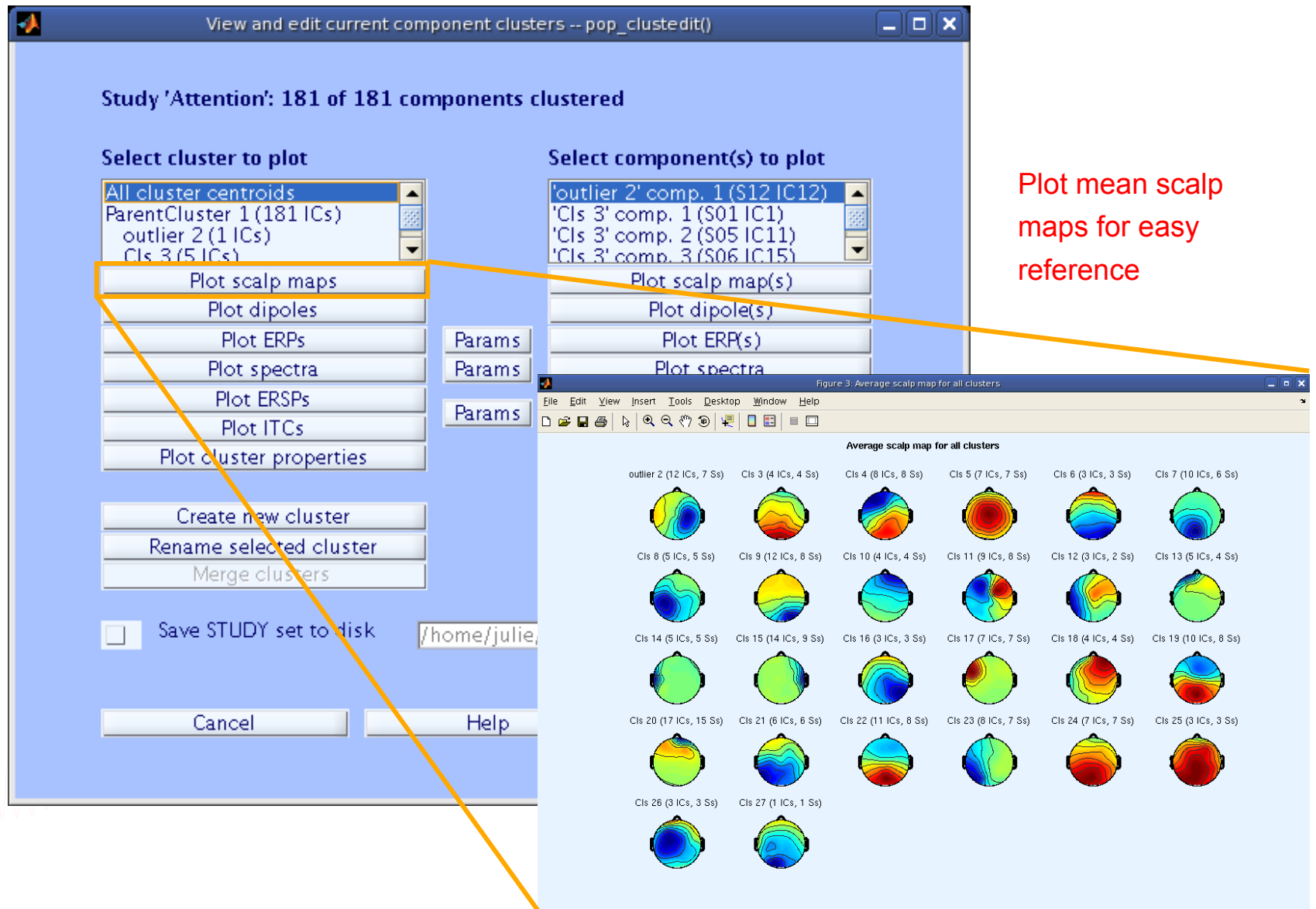
View and edit clusters



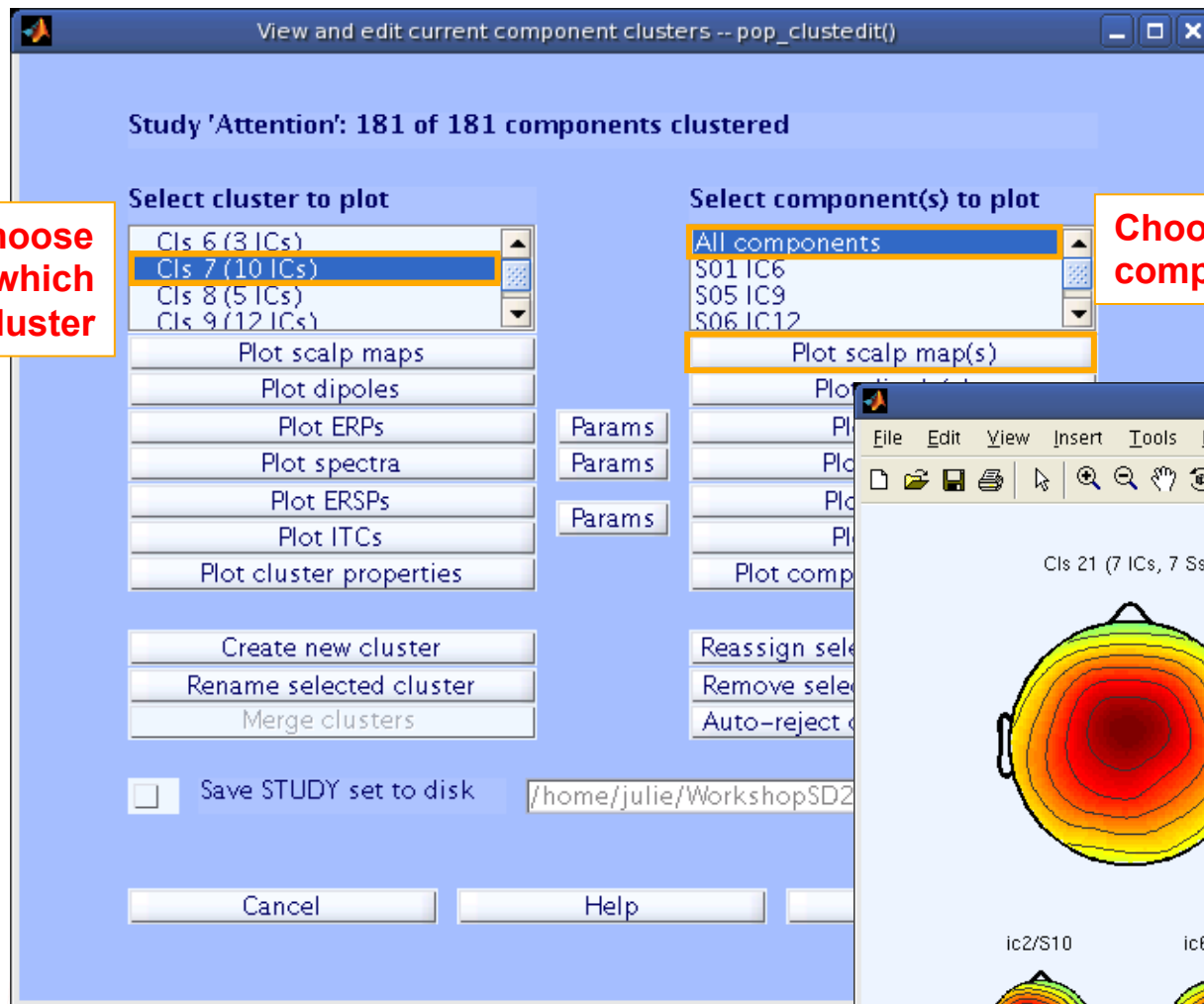
Plot/edit clusters



Plot cluster data

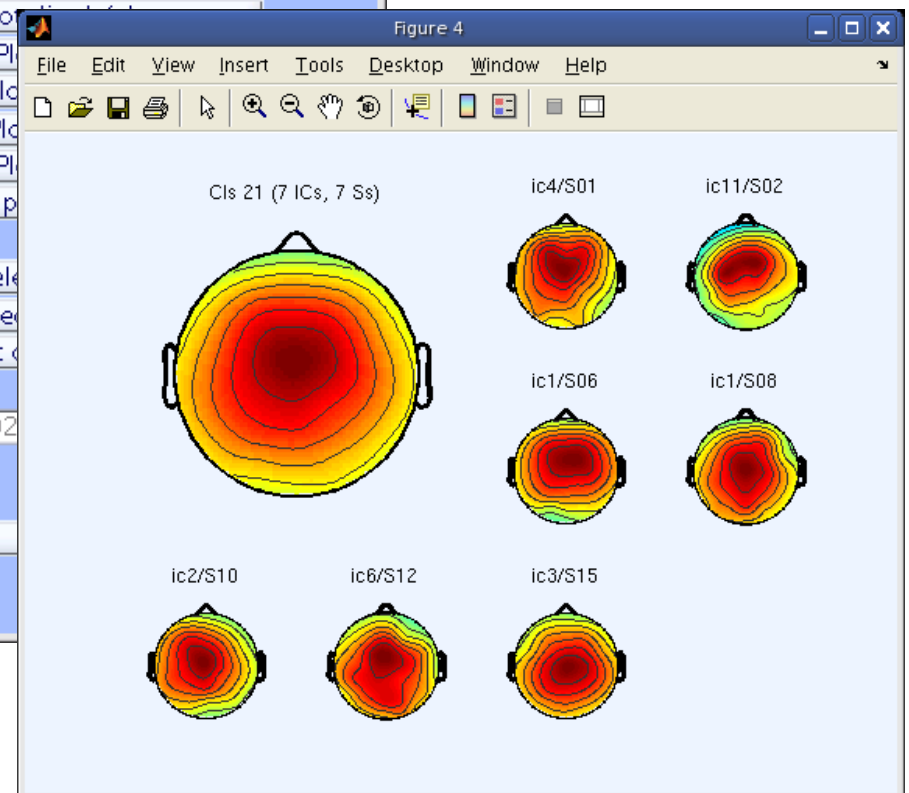


Plot cluster data

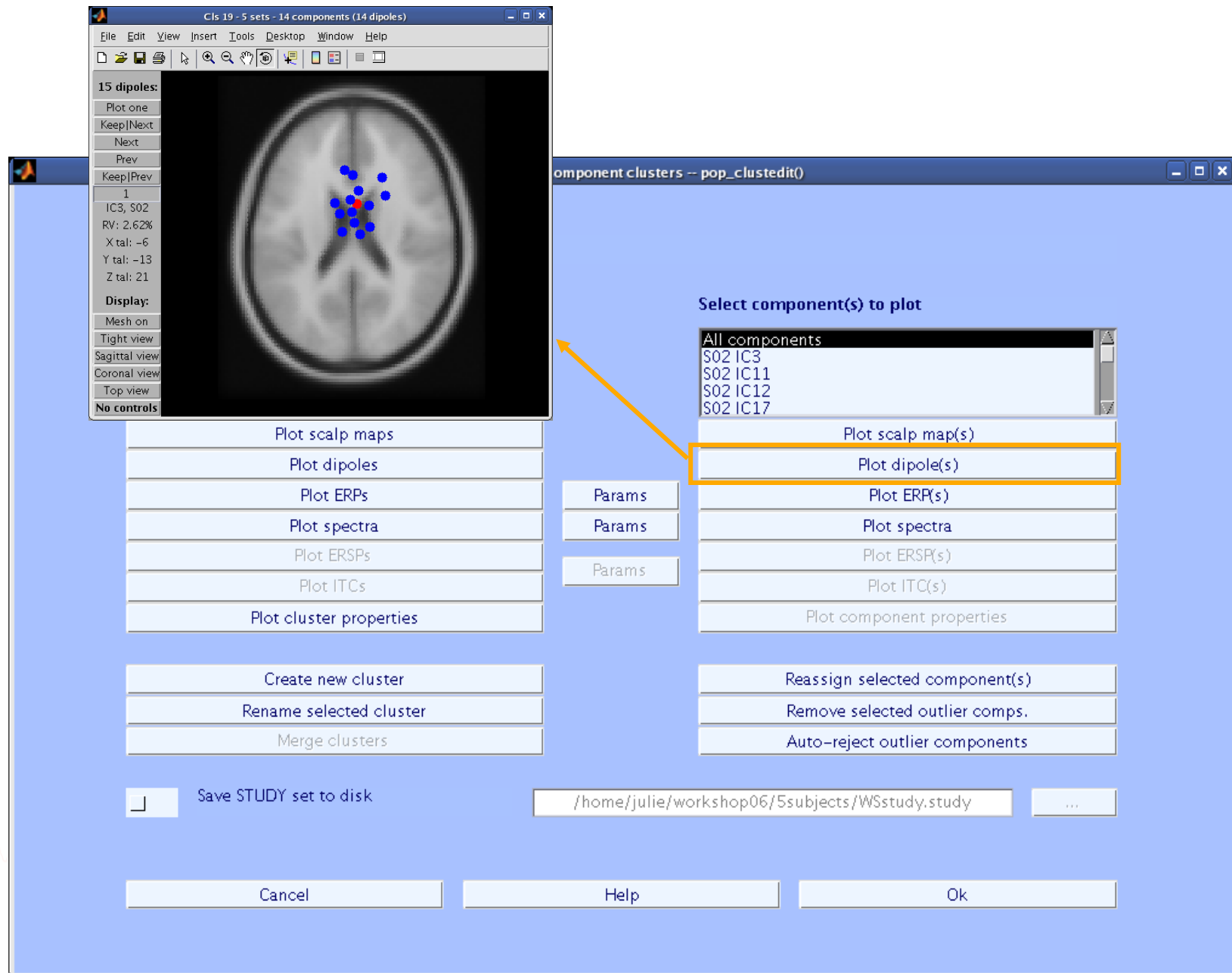


Choose
which
cluster

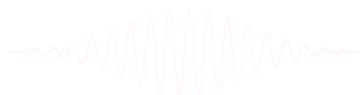
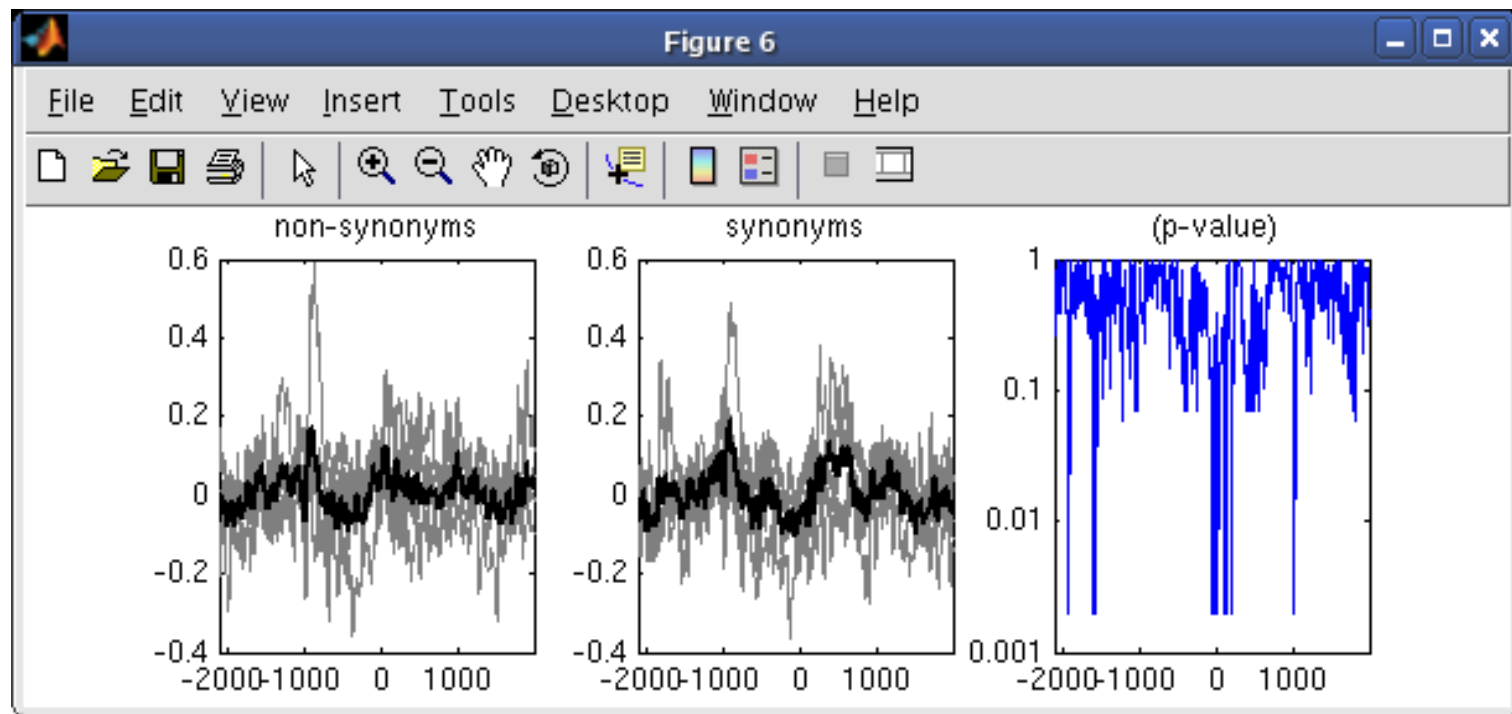
Choose which
components



Plot cluster data



Plot cluster ERP



Issue with standard clustering

Large parameter space problem: many different clustering solutions can be produced by changing parameters and measure subsets. Which one should we choose?

EEGLAB clustering has ~12 parameters



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

Pre-compute measures on which to cluster components from study 'N400STUDY'
Select the cluster to refine during sub-clustering (any existing sub-hierarchy will be overwritten)

ParentCluster 1 (151 ICs)

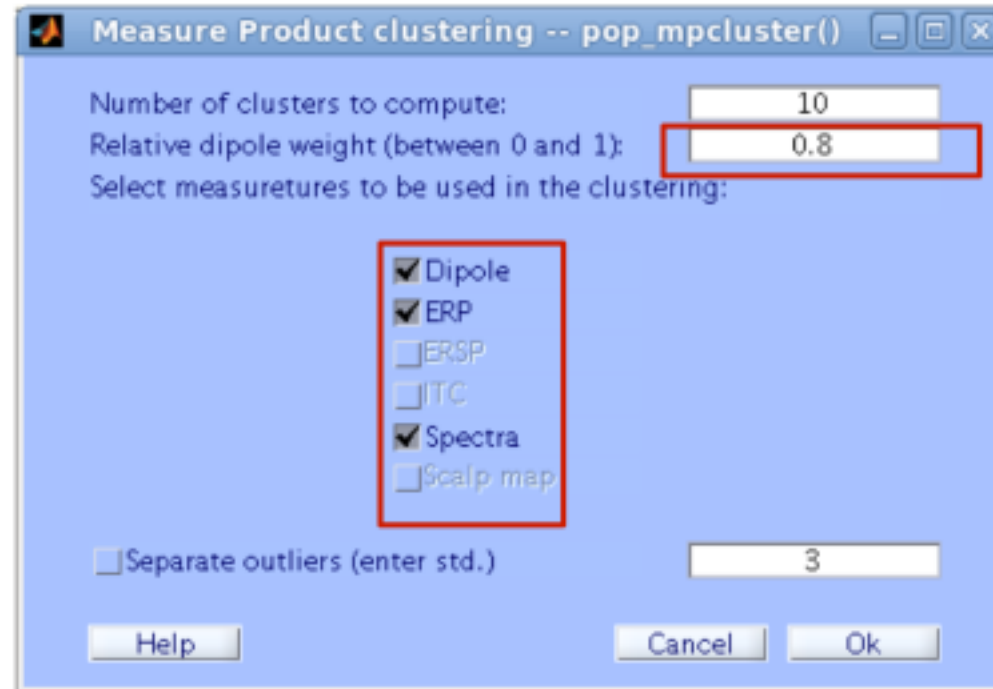
Pre-compute or Load	Dims.	Norm.	Rel. Wt.
<input checked="" type="checkbox"/> spectra	10	<input checked="" type="checkbox"/> 1	Frequency range [Hz]
<input checked="" type="checkbox"/> ERPs	10	<input checked="" type="checkbox"/> 1	Latency range in ms [lo hi]
<input checked="" type="checkbox"/> dipoles	3	<input checked="" type="checkbox"/> 10	
<input checked="" type="checkbox"/> scalp maps	10	<input checked="" type="checkbox"/> 1	Use channel values <input type="checkbox"/>
<input checked="" type="checkbox"/> ERSPs	10	<input checked="" type="checkbox"/> 1	Time/freq. parameters
<input checked="" type="checkbox"/> ITCs	10	<input checked="" type="checkbox"/> 1	Time/freq. parameters
<input checked="" type="checkbox"/> Final dimensions	10	Help	

☐ Save STUDY to file

Cancel Help Ok

Measure projection

(EEGLAB extension by Nima Bigdely Shamlo)
only has one pre-clustering parameter.



(Affinity clustering by Pernet, Martinez, Delorme)

Exercise

- Load the STUDY stern.study
- Precompute **spectrum**, **ERP** and **scalp maps** for components
- Precluster and cluster components using **spectrum and dipoles**
- 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

