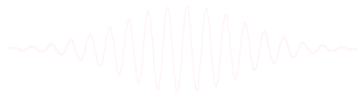# Clustering of ICA components
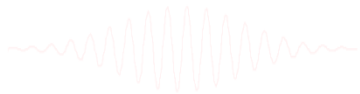
Arnaud Delorme

(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

ICs to cluster
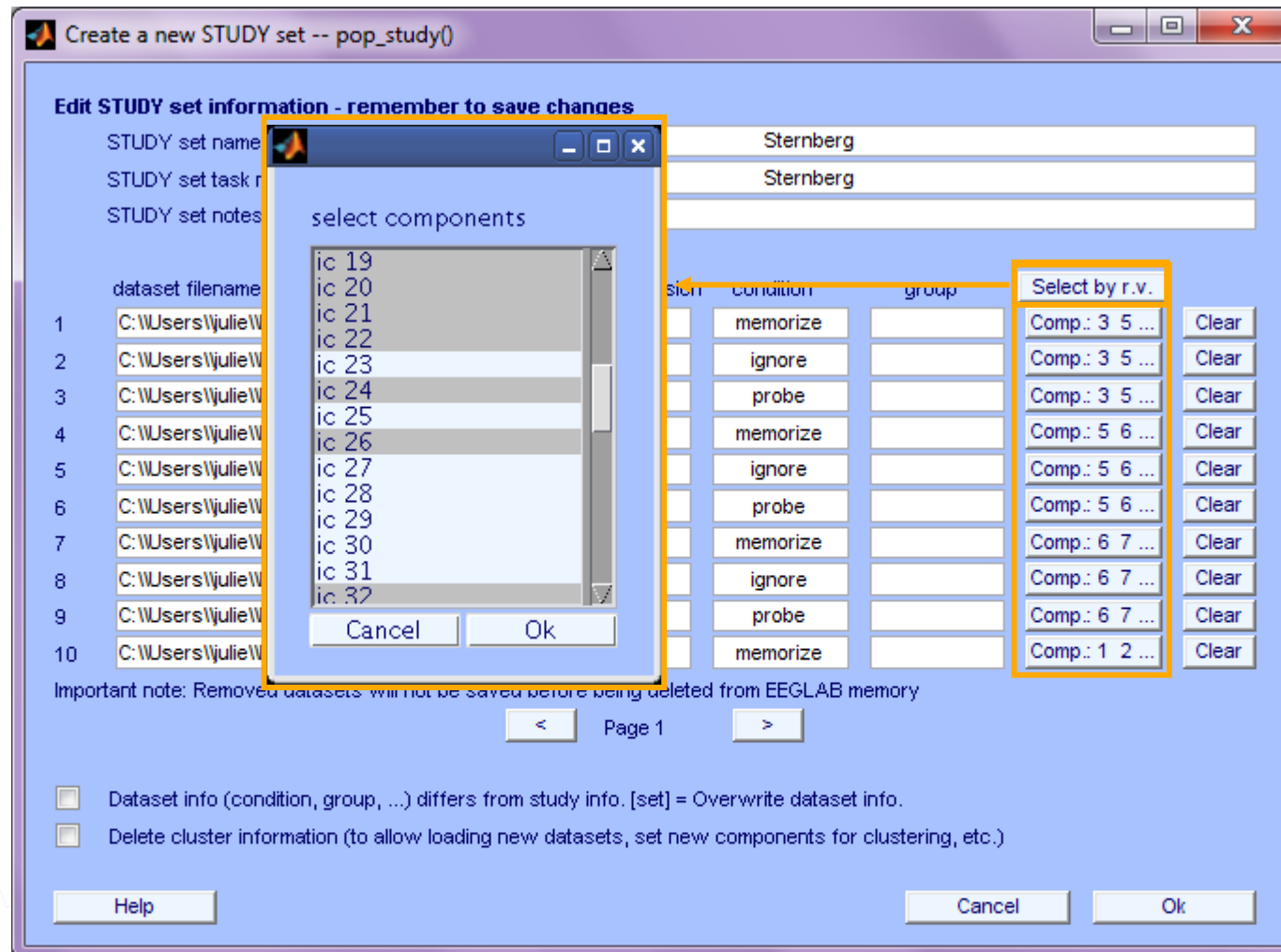
Precompute data measures
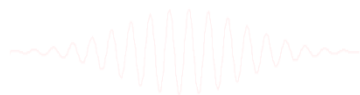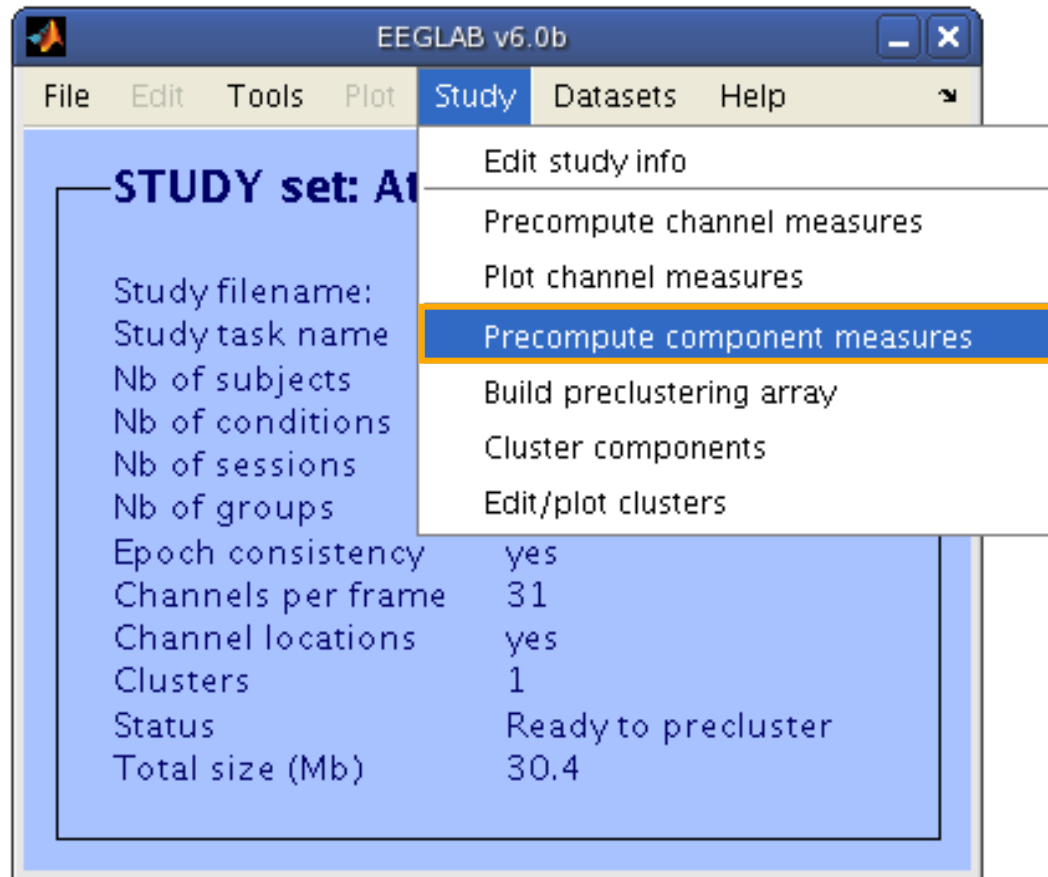
Precompute data measures

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

Time-frequency options

Cluster components
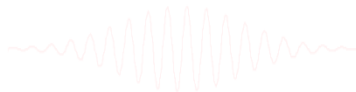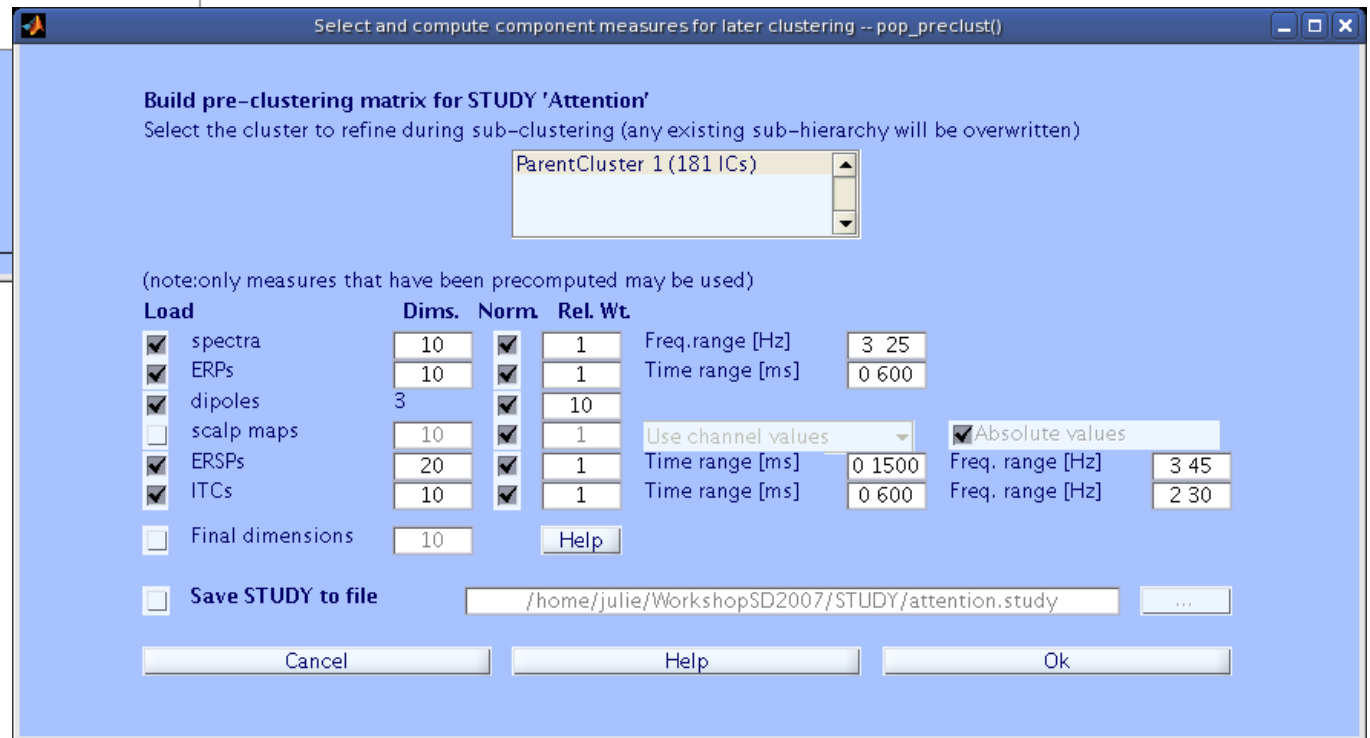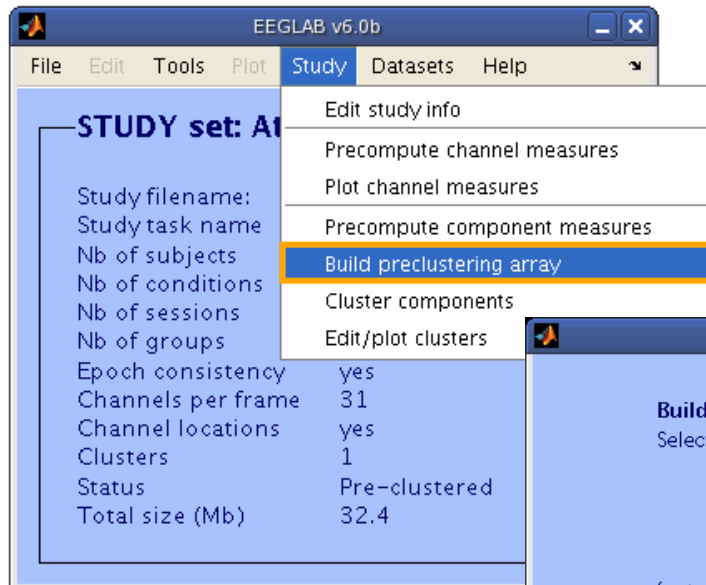
Precluster schematic
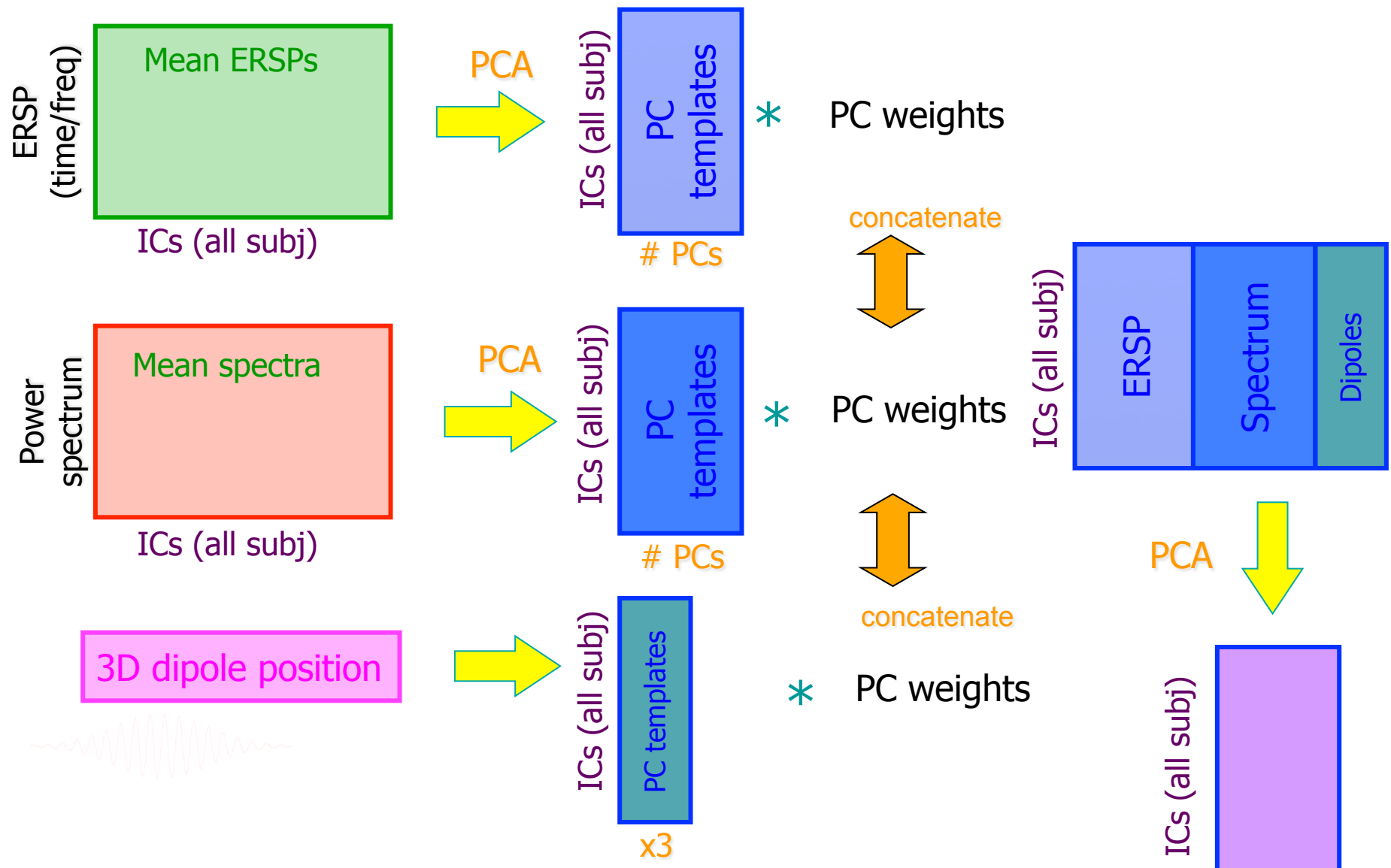
Precluster: Use singular values from PCA

Precluster schematic

ICs (all subj)

| ERSP | Spectrum | Dipoles |

OR

ICs (all subj)
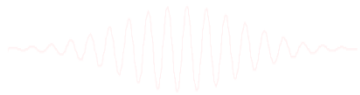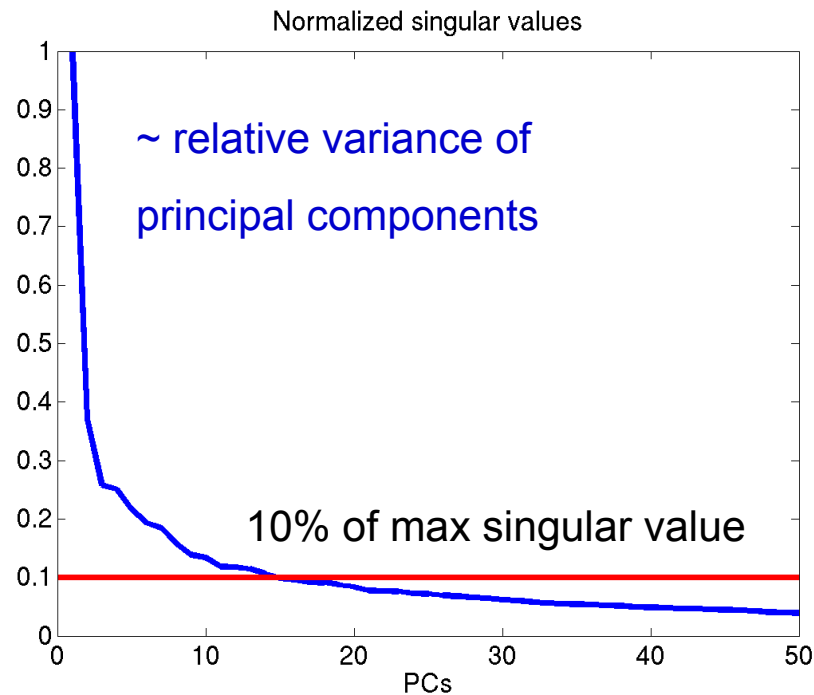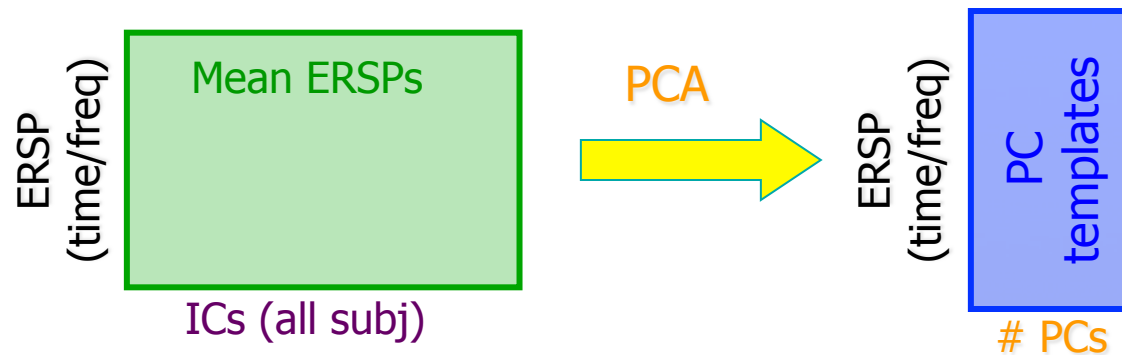
Each component is a dot
Clustering will group these
dots

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.
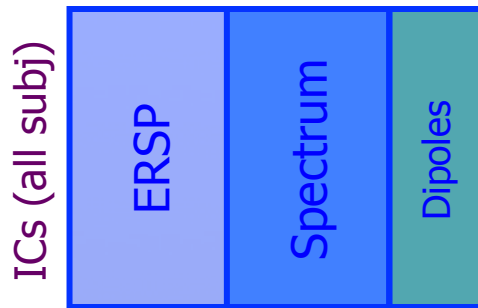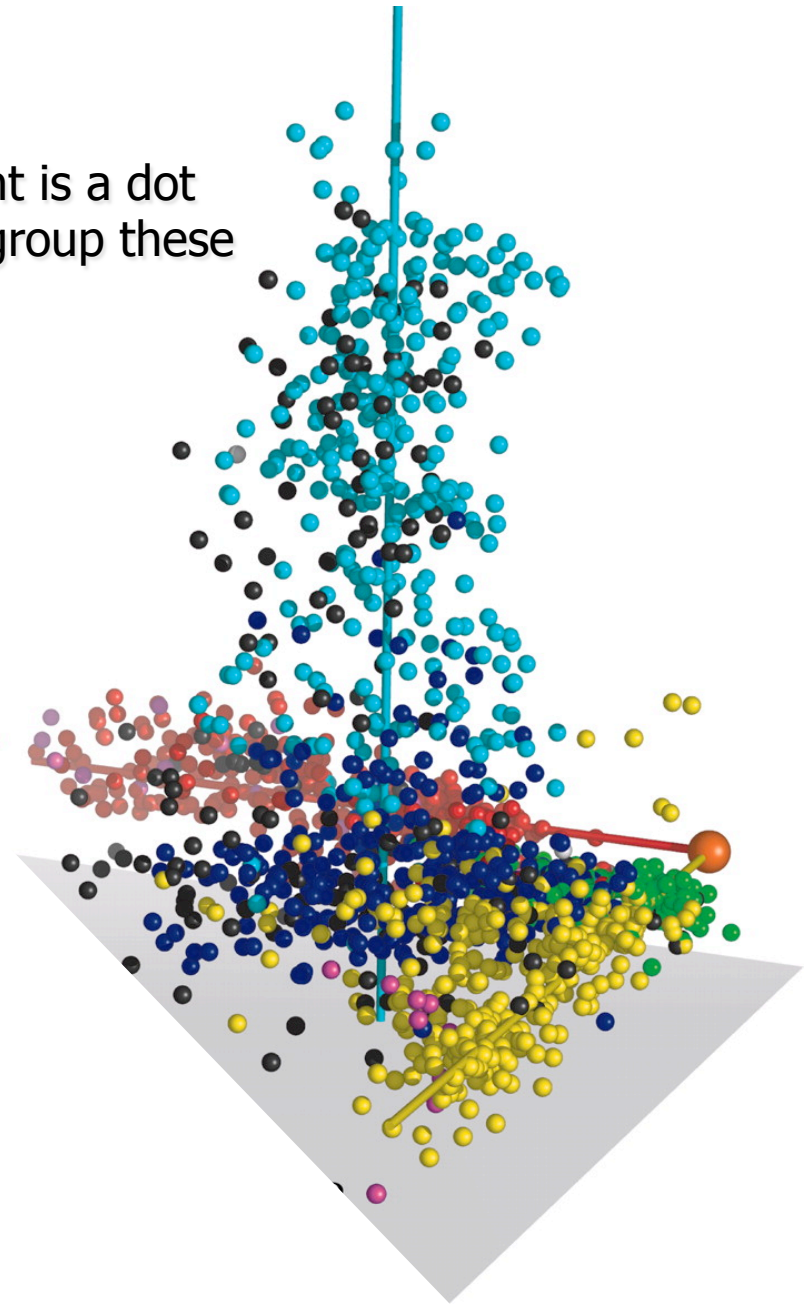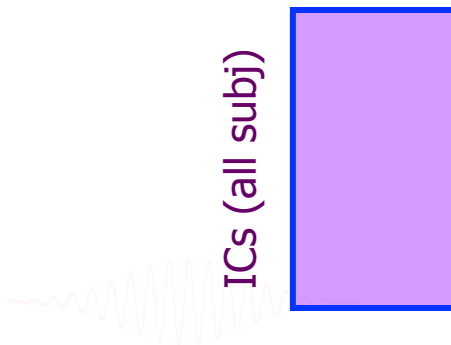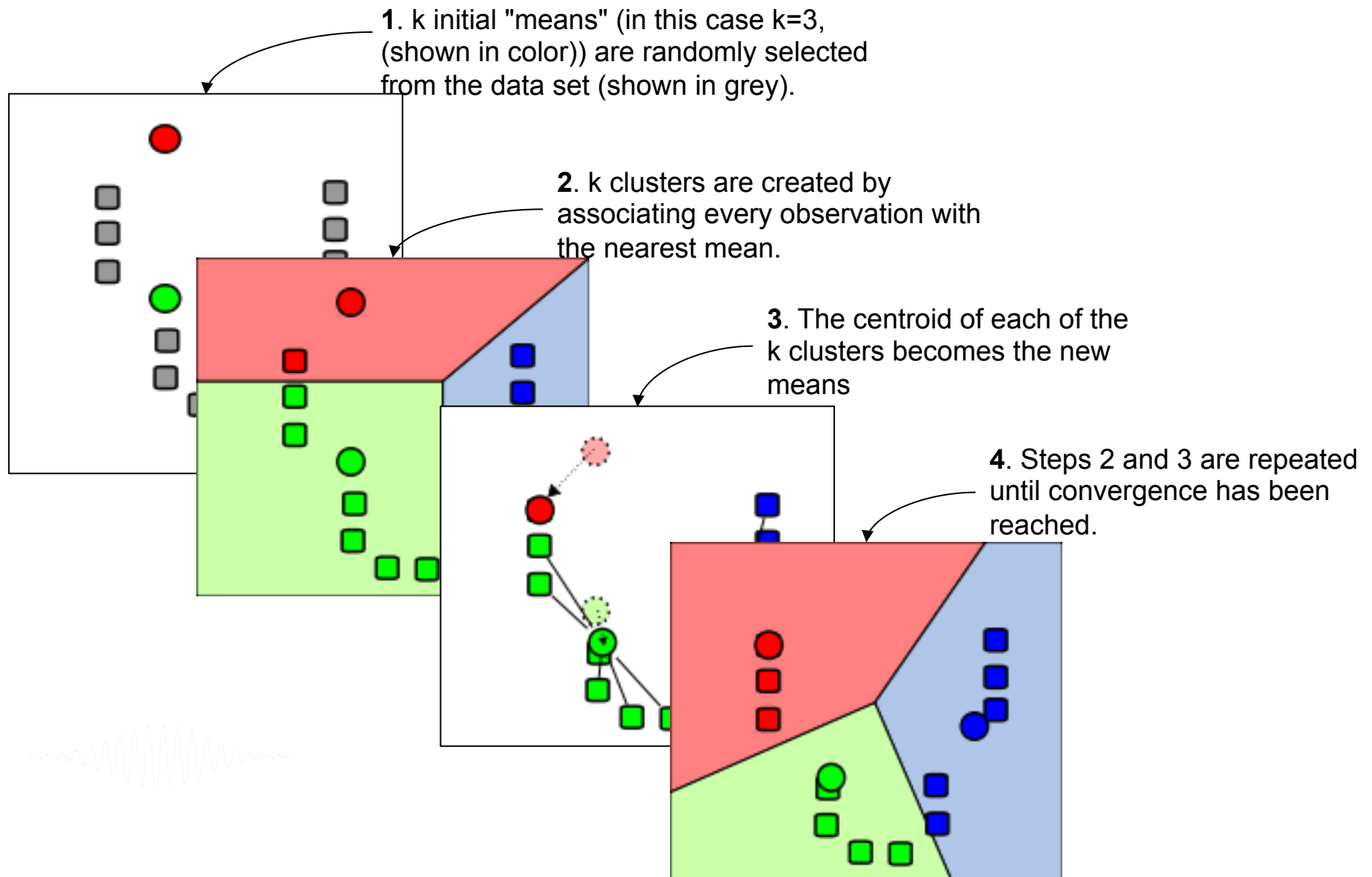
Cluster components
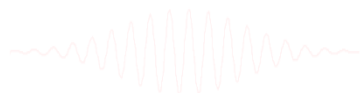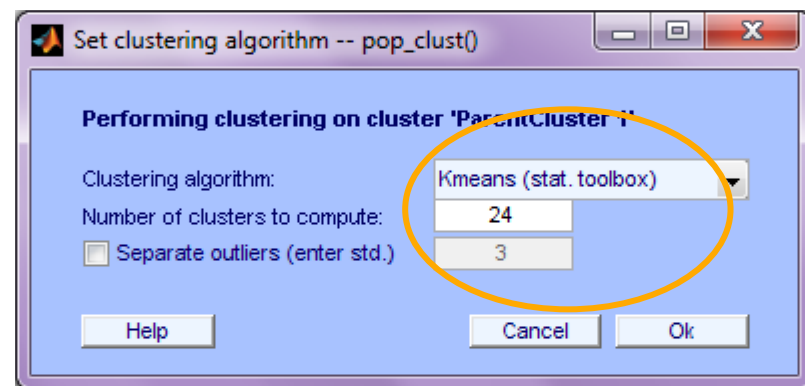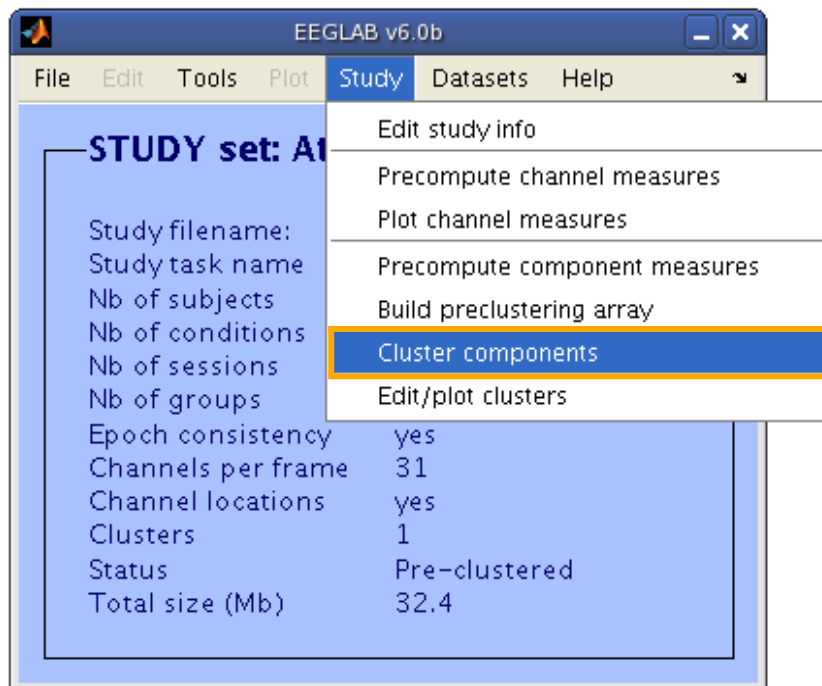
Choosing data measures

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
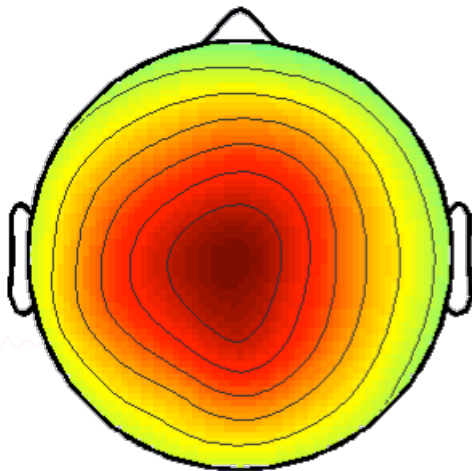
IC5 / S05, Cls 26

Choosing data measures

Similar dipole location,
very different orientation.
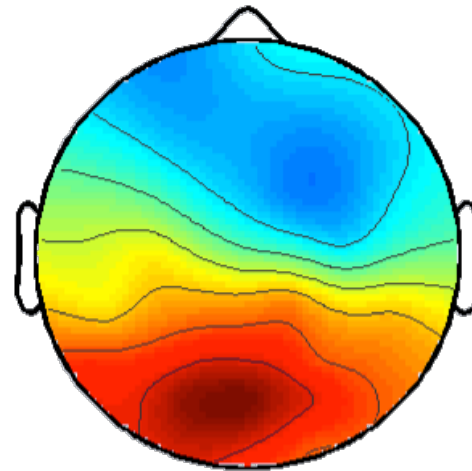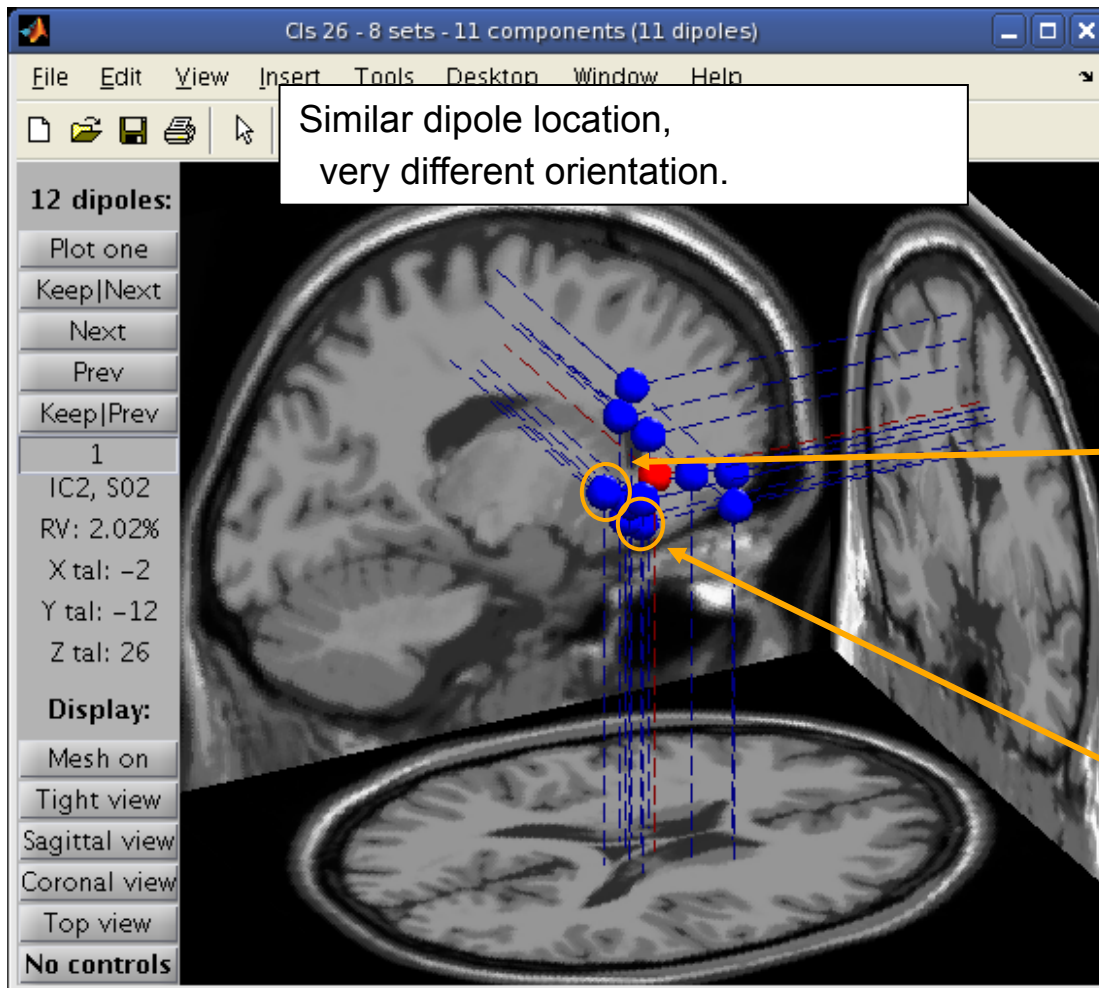
Obvious dramatic effect on
scalp map topography:

IC5 / S05, Cls 26
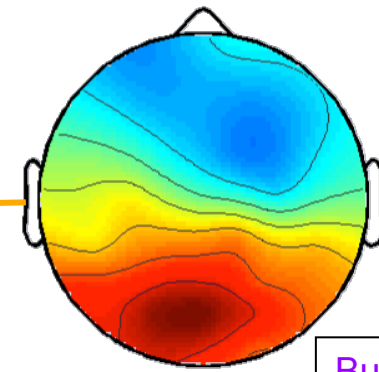
IC2 / S02, Cls 26
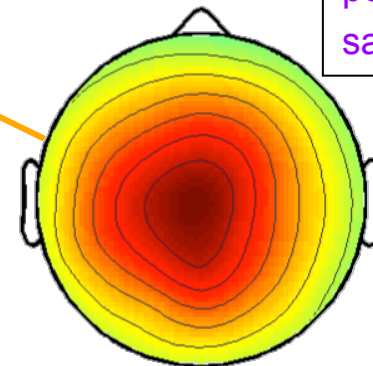
But, do they
perform the
same functions?

Choosing data measures
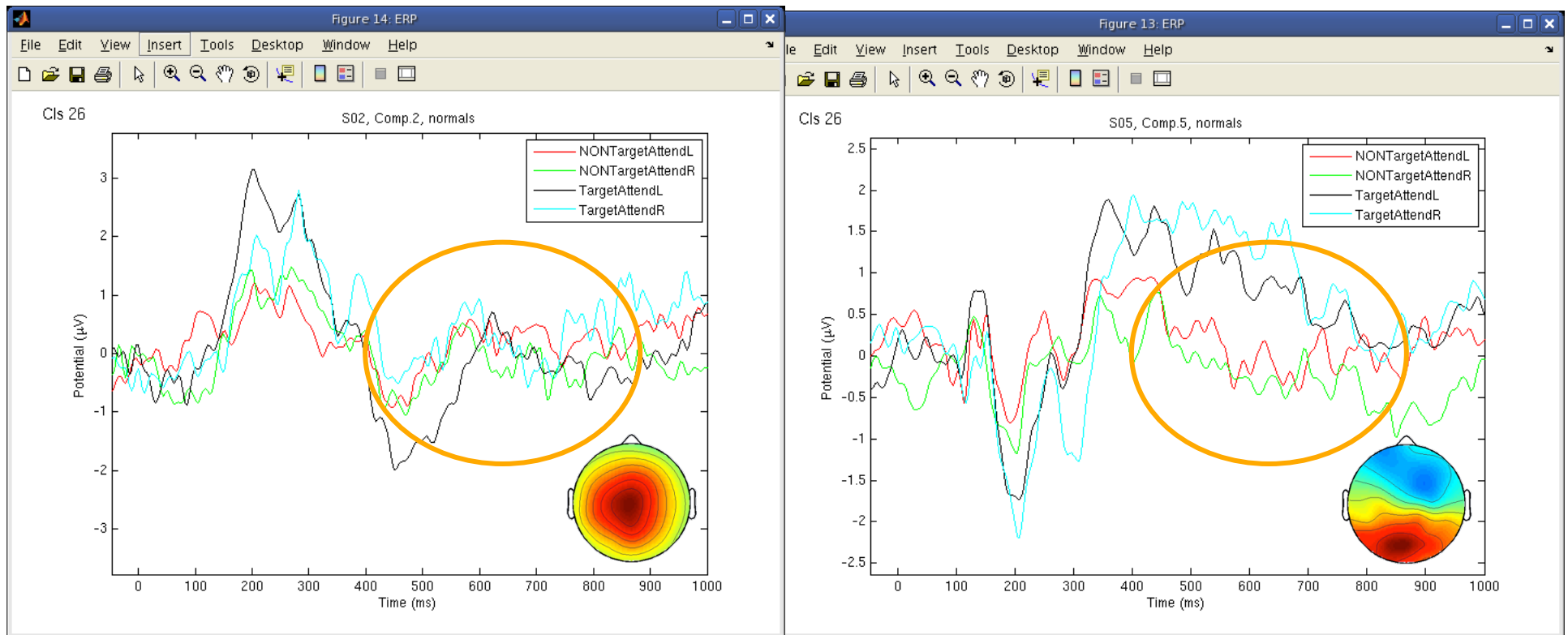
ERPs seem different…
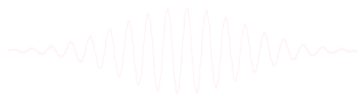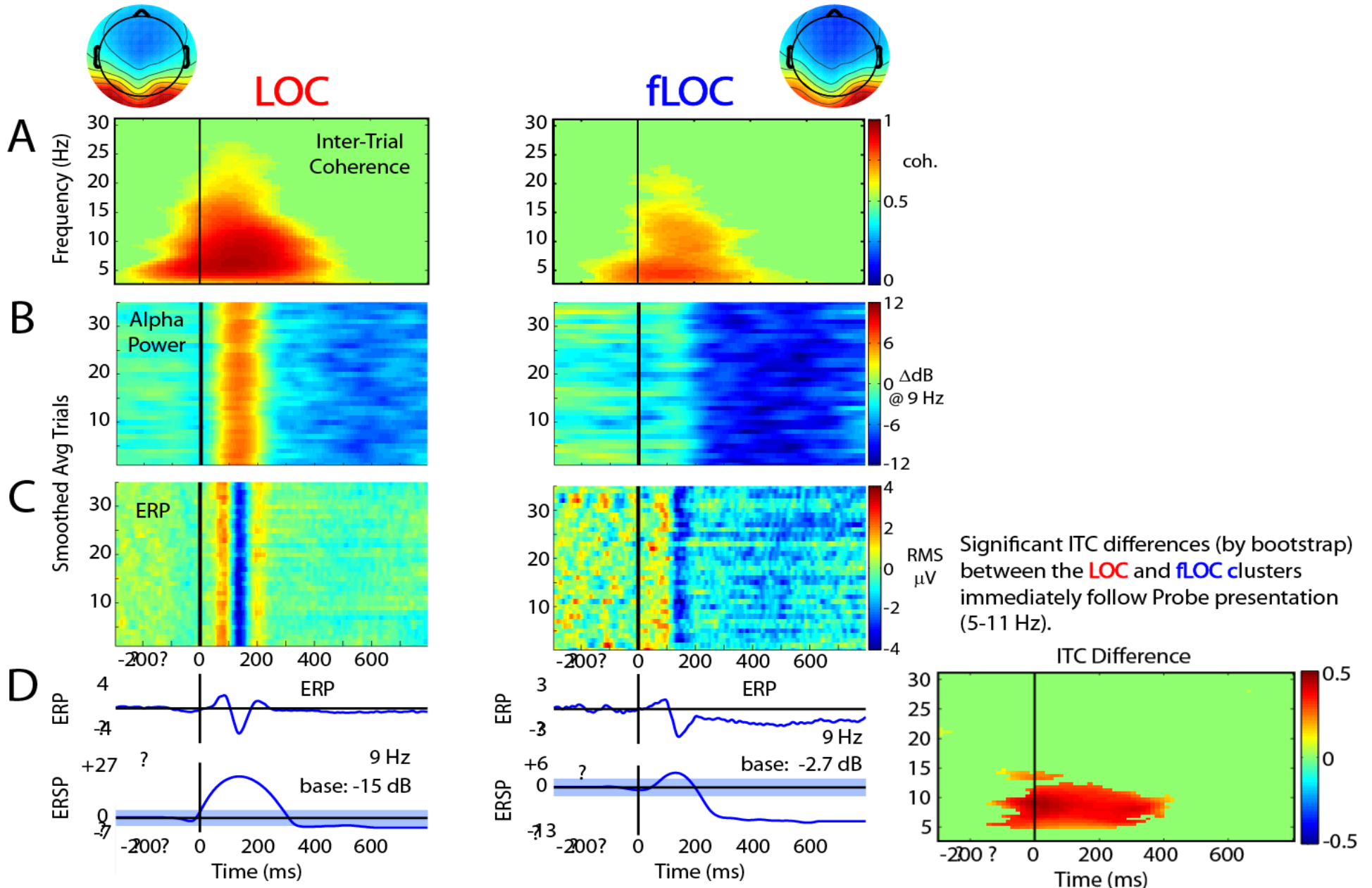
Subject differences?

Subject differences?

A — Scalp Power (RMS µV² dB) vs Frequency (Hz)
- fLOC Pre-Probe
- fLOC Post-Probe
- LOC Pre-Probe
- LOC Post-Probe

B — LOC

C — fLOC

** Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects (p < 0.01)

D — ERSP Difference   ΔdB

*** Difference ERSP shows significant differences between the two clusters by bootstrap statistics (p < 0.001)

Subject differences?

Plot/edit clusters
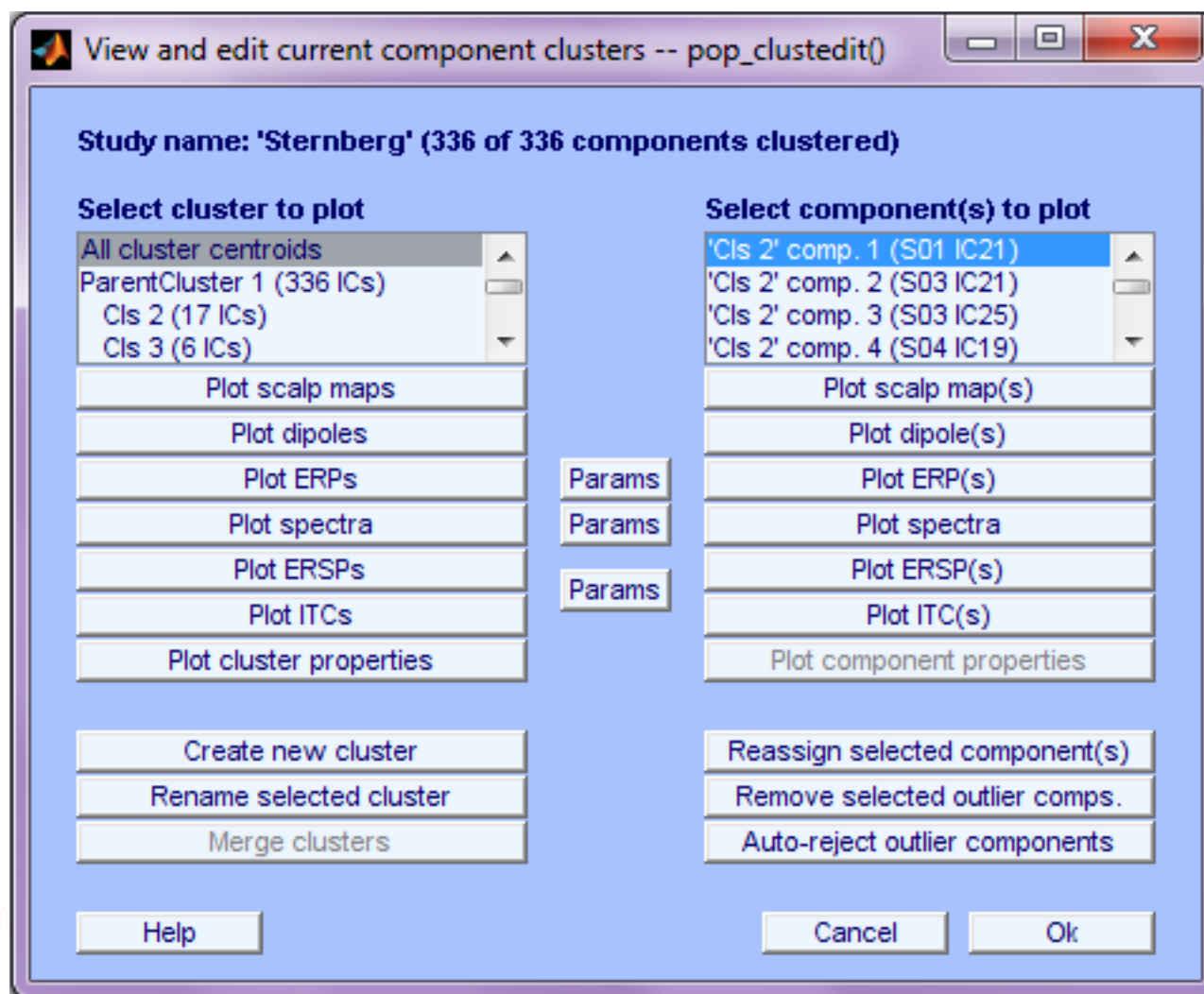
Plot cluster data
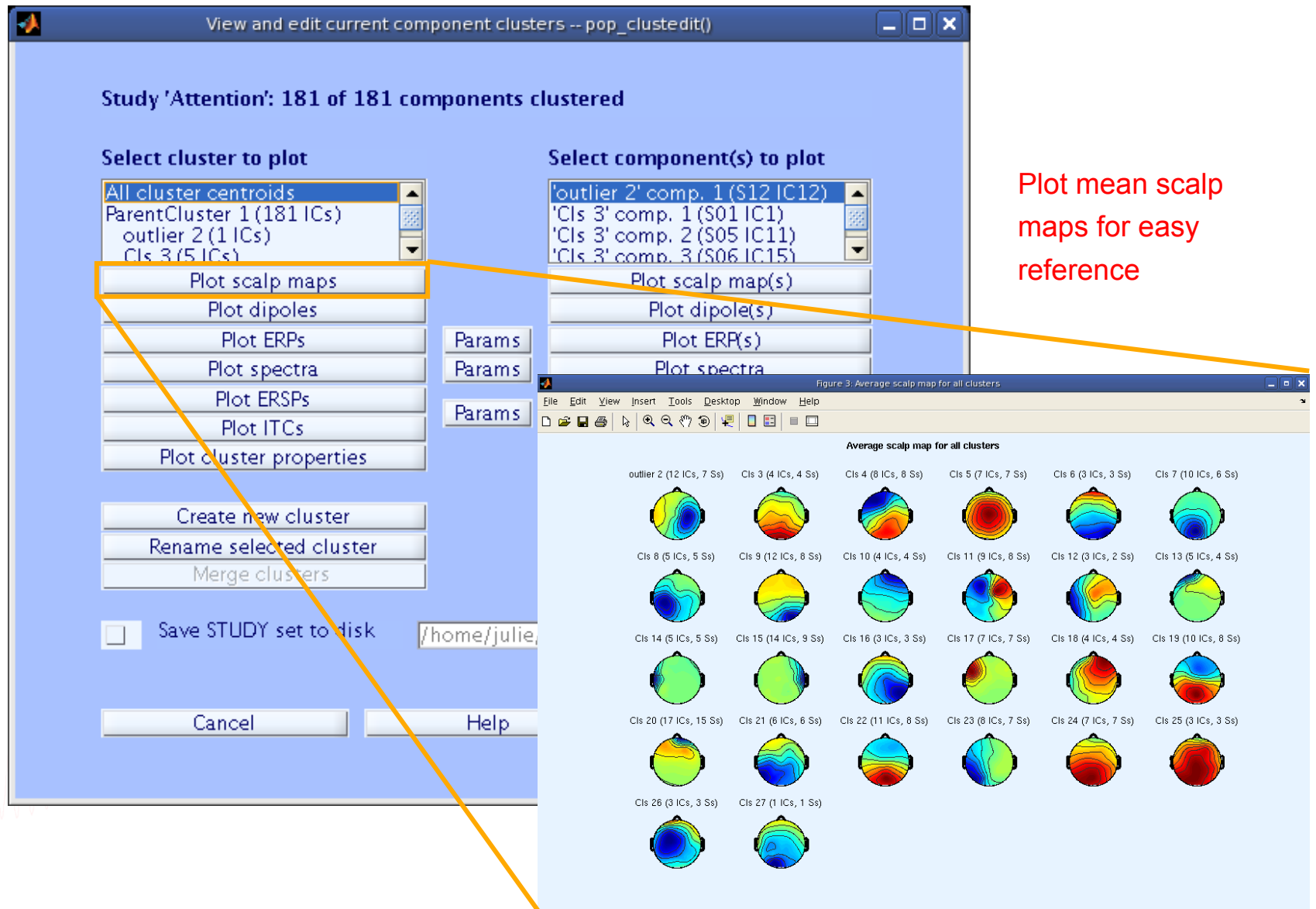
Plot mean scalp maps for easy reference
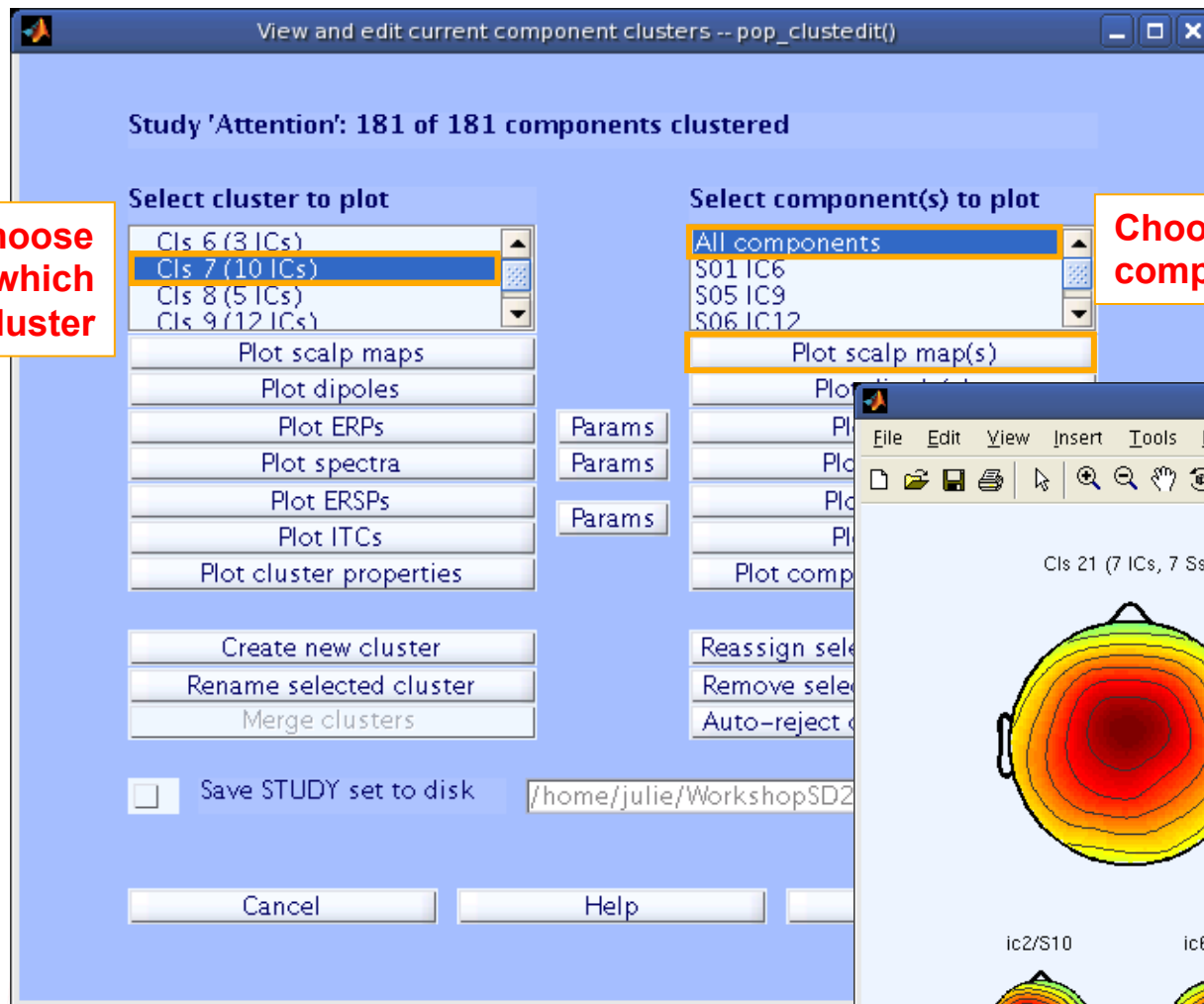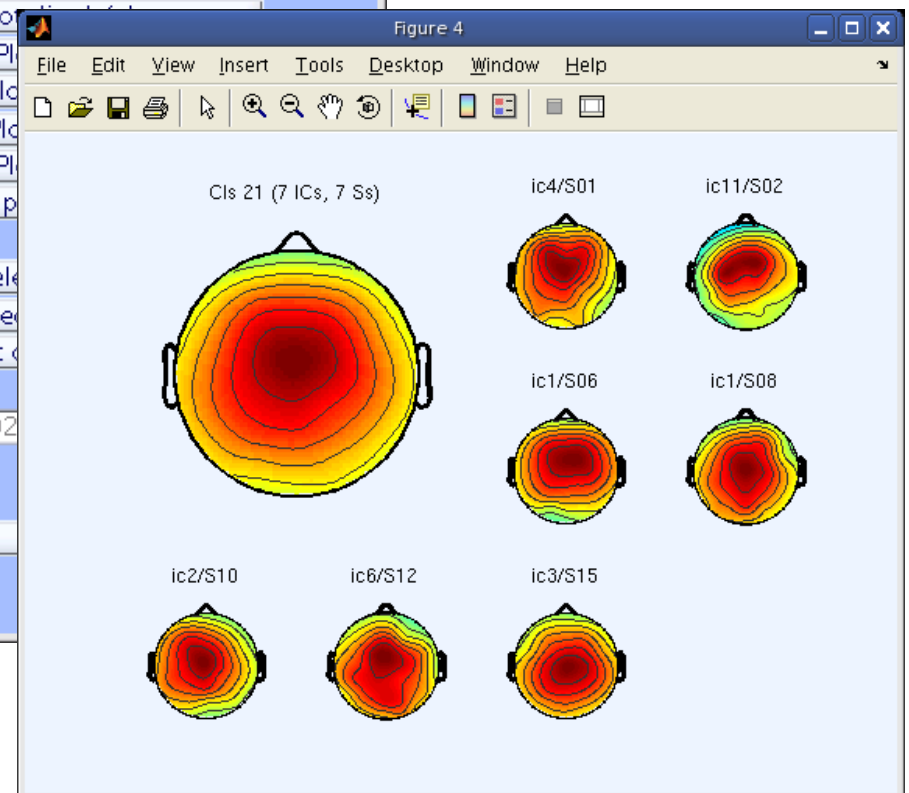
Plot cluster data
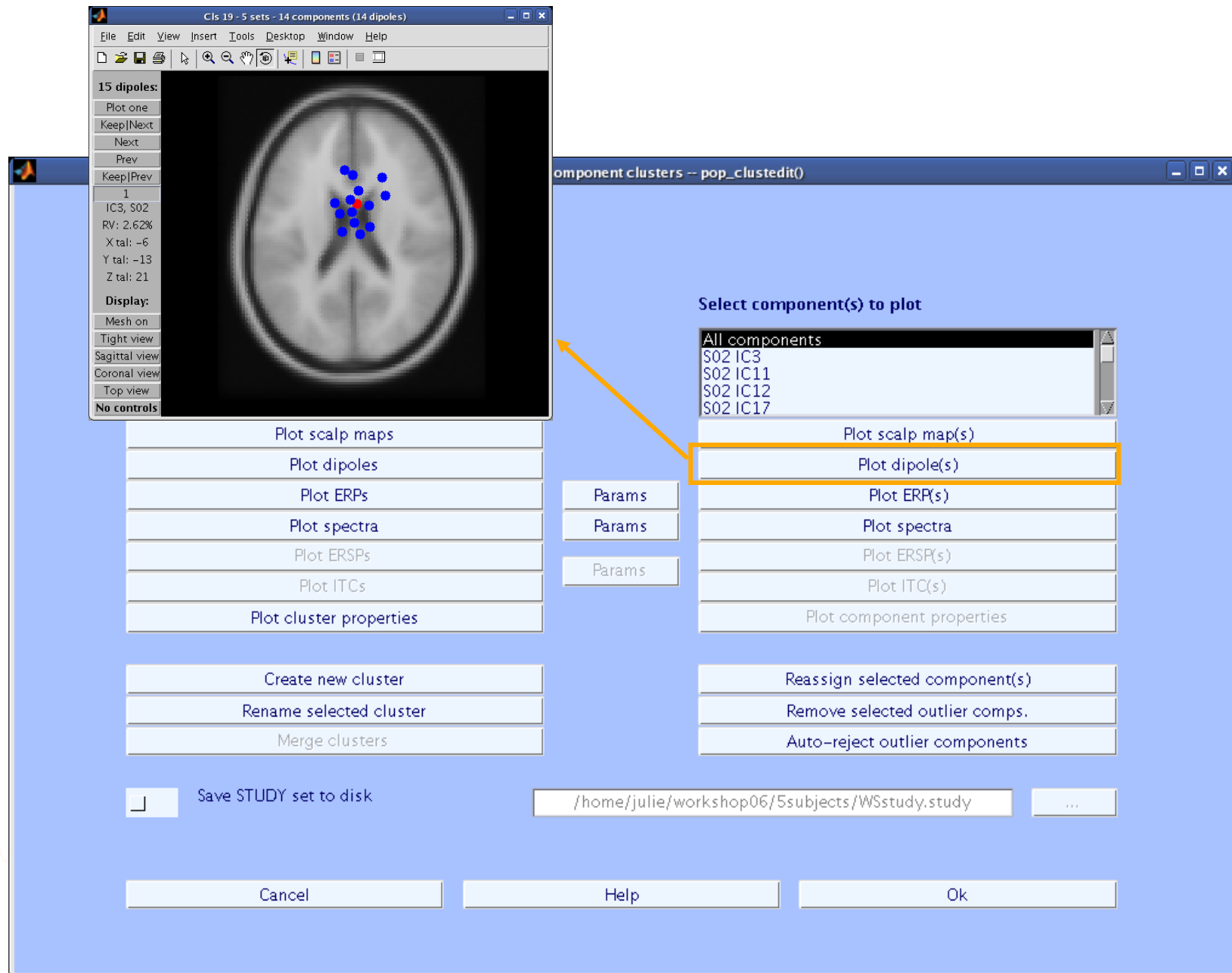
Plot cluster data

Plot cluster ERP
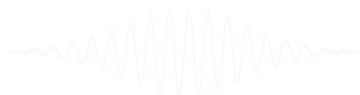
Reassigning components

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

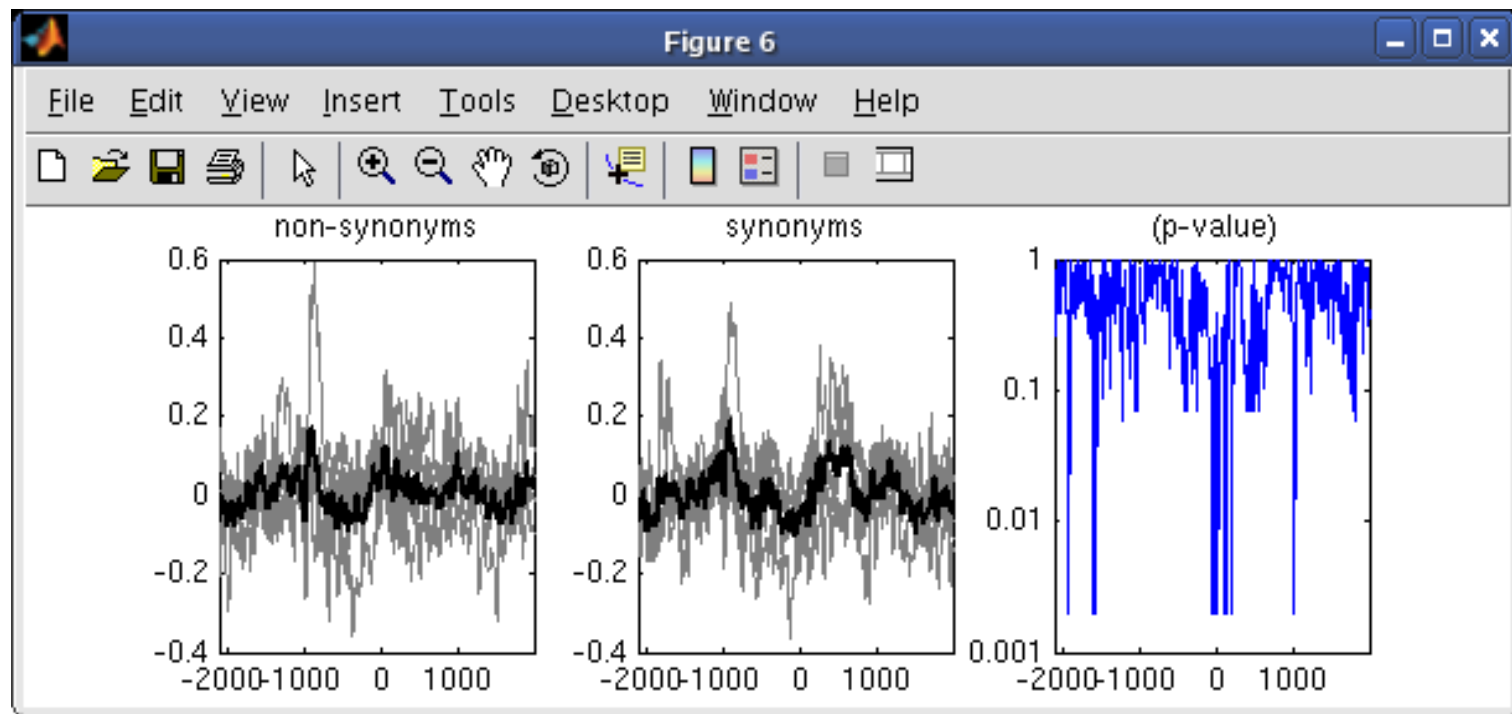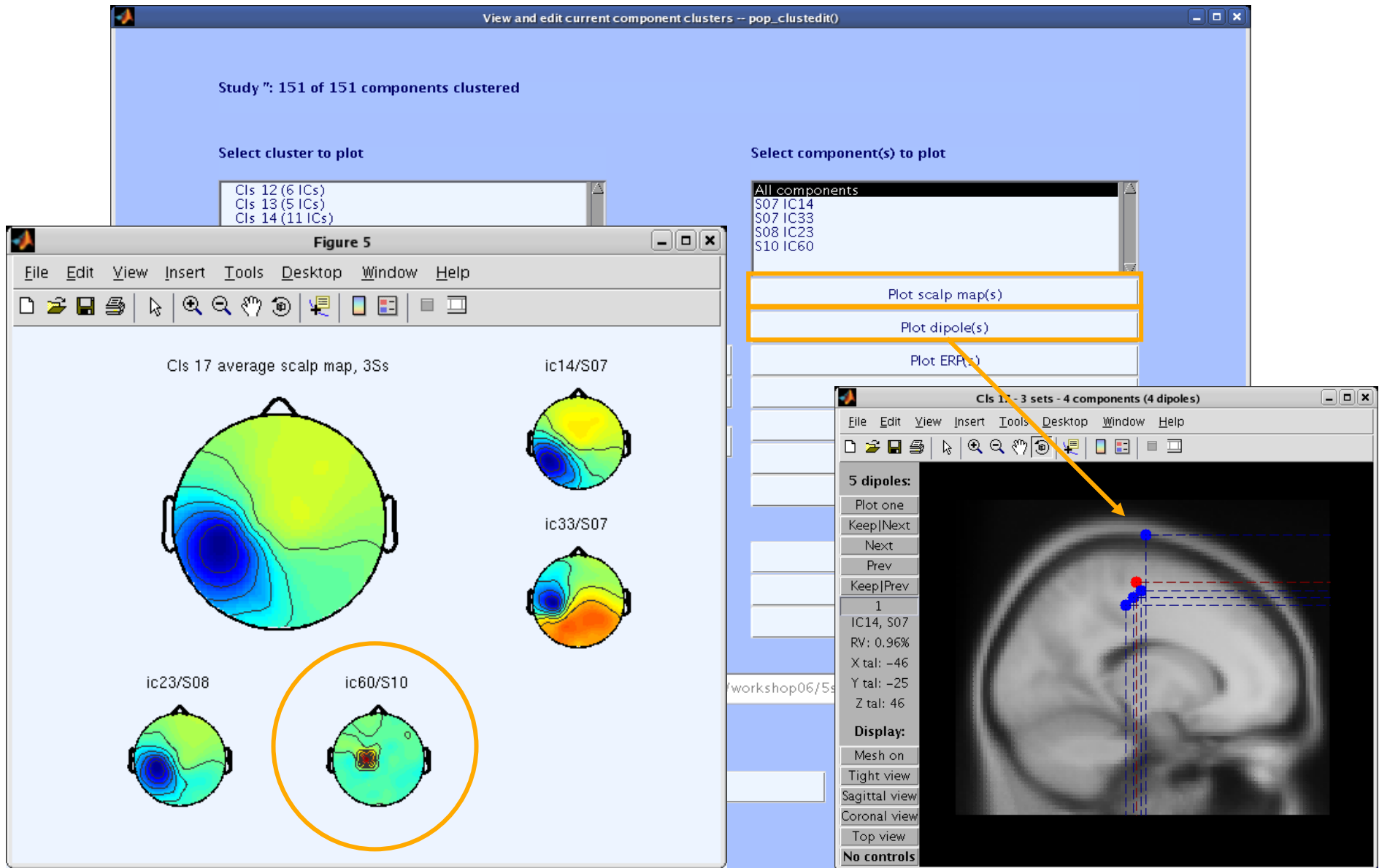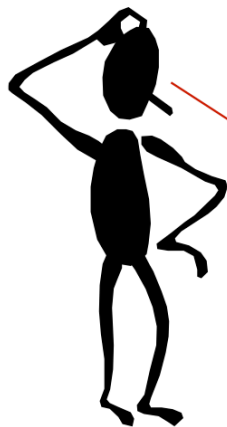Select and compute component measures for later clustering -- pop_preclust()

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.		
spectra	10	1	Frequency range [Hz]	3 25
ERPs	10	1	Latency range in ms [lo hi]	-2100 1995
dipoles	3	10		
scalp maps	10	1	Use channel values	Absolute values
ERSPs	10	1	Time/freq. parameters	e', [3 25], 'cycles', [3 0.5], 'pa
ITCs	10	1	Time/freq parameters	', [3 25], 'cycles', [3 0.5], 'p4
Final dimensions	10	Help		

Save STUDY to file   /data/common4/amo/6subjects/N400preclust.study   ..

Cancel   Help   Ok

Measure projection

(EEGLAB extension by Nima Bigdely Shamlo)
only has one pre-clustering parameter.
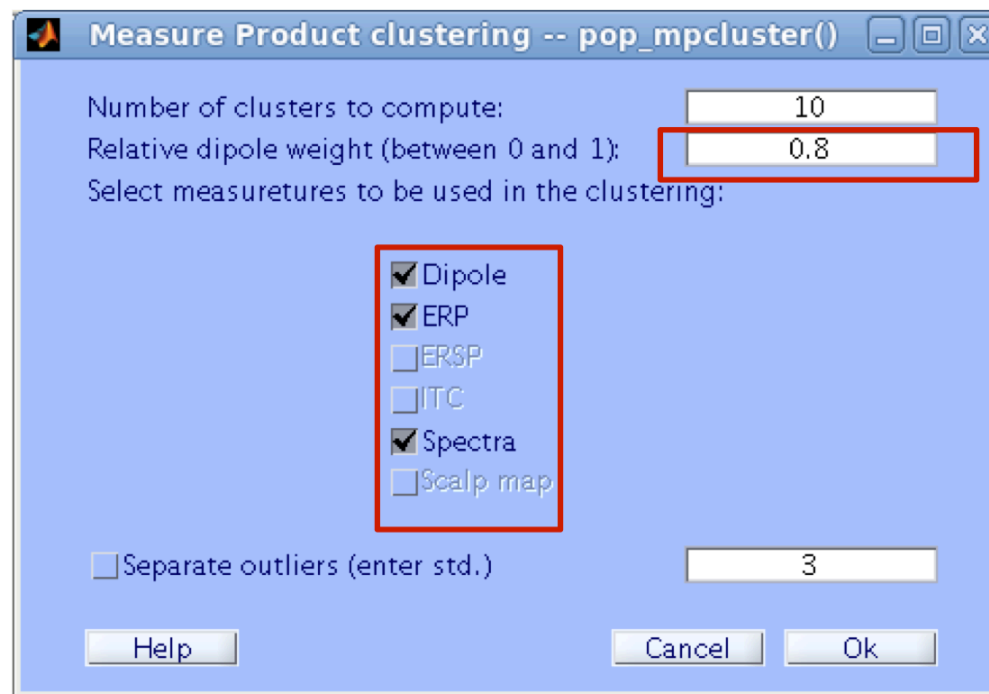
(Affinity clustering by Pernet, Martinez, Delorme)

Exercise

- Load the STUDY
- Precompute ERP, power spectrum and scalp topographies
- Precluster and cluster components using spectrum and dipoles location
- Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
- Plot significant difference for one component cluster spectrum between the two conditions in the default design