Clustering of ICA components

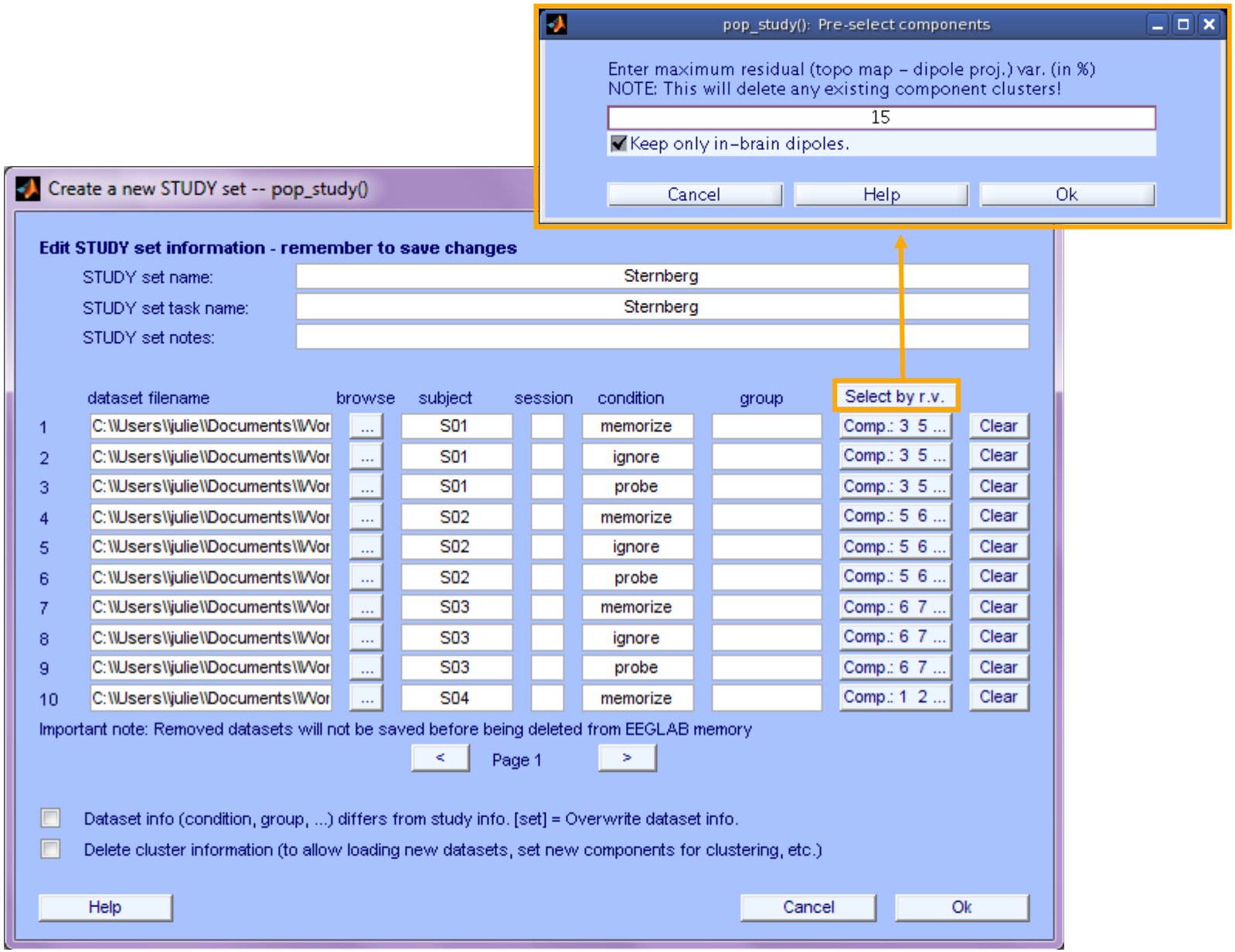
Arnaud Delorme

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

Steps of clustering

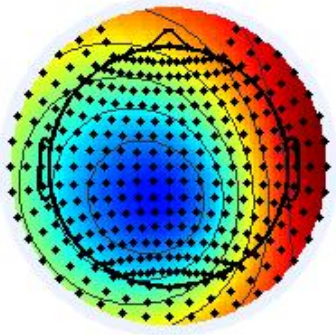
- Select ICA components for clustering
- Precompute measures of interest
- Cluster measures
- Plot clusters and edit them if necessary

Edit dataset info

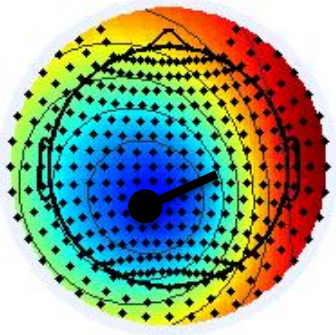


Computing residual variance (%)

Actual

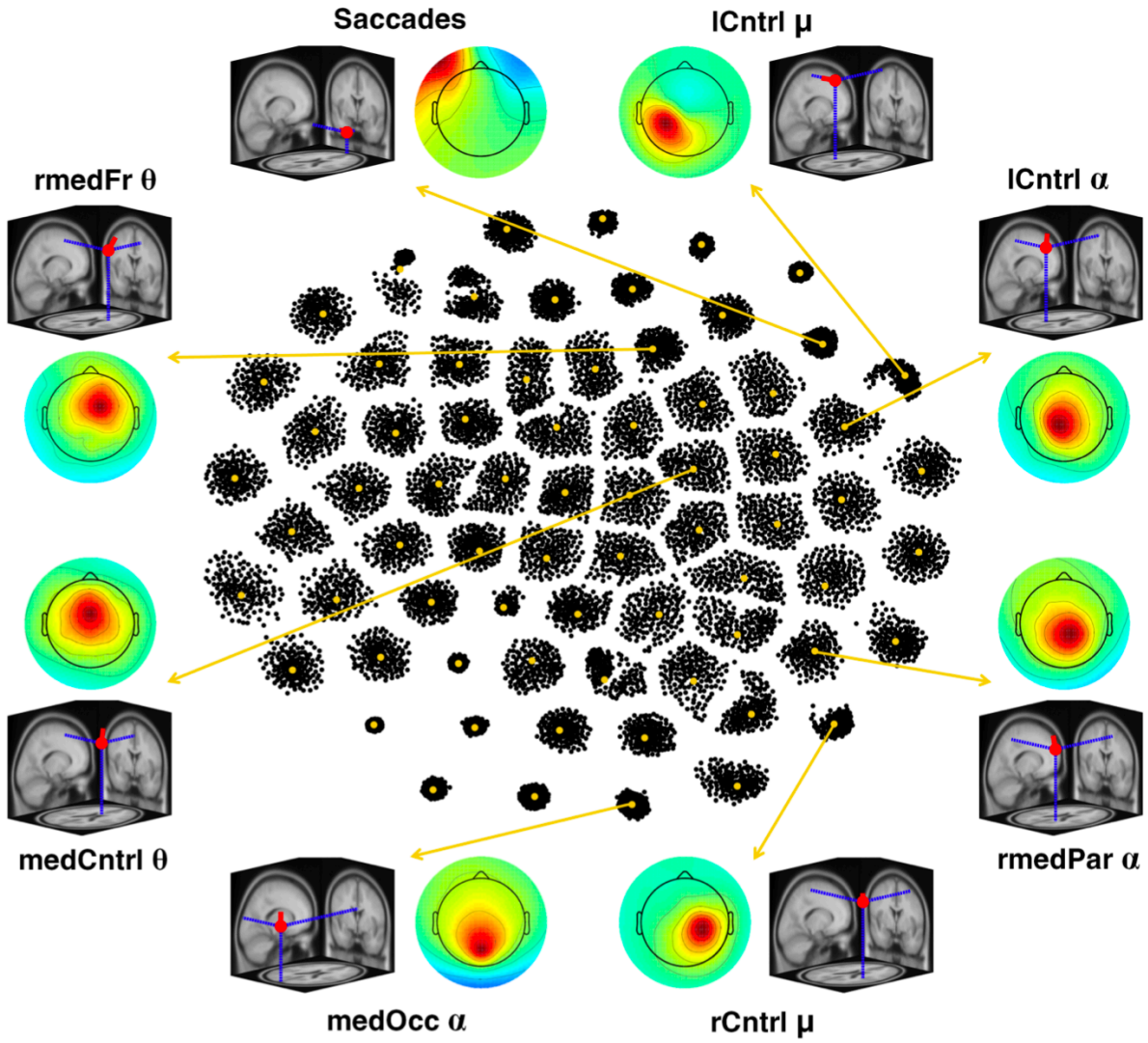


Dipole projection



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

Clustering results example

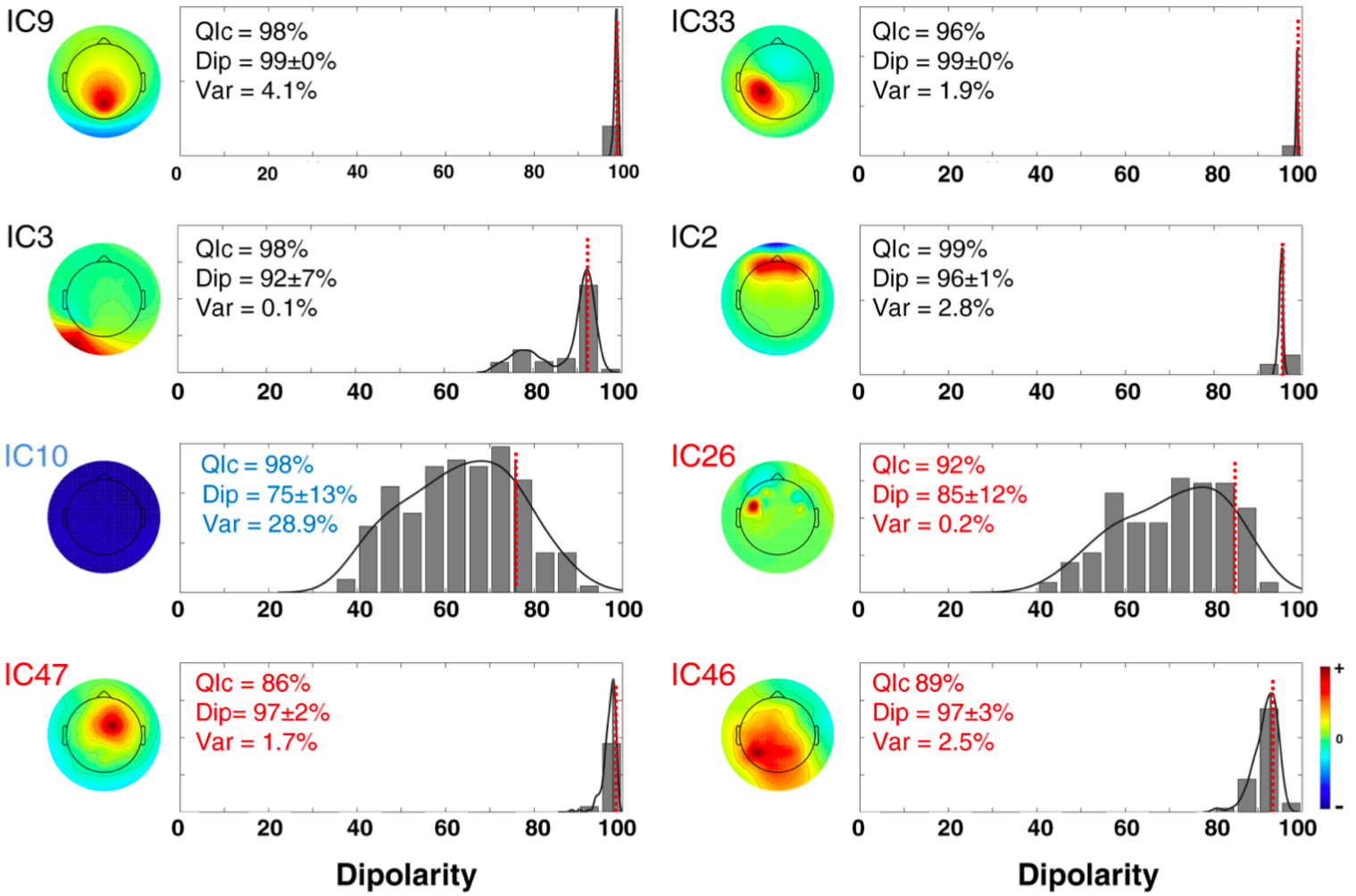


RELICA: A method for estimating the reliability of independent components

Fiorenzo Artoni [a,*], Danilo Menicucci [b], Arnaud Delorme [c,e,f], Scott Makeig [c], Silvestro Micera [a,d]

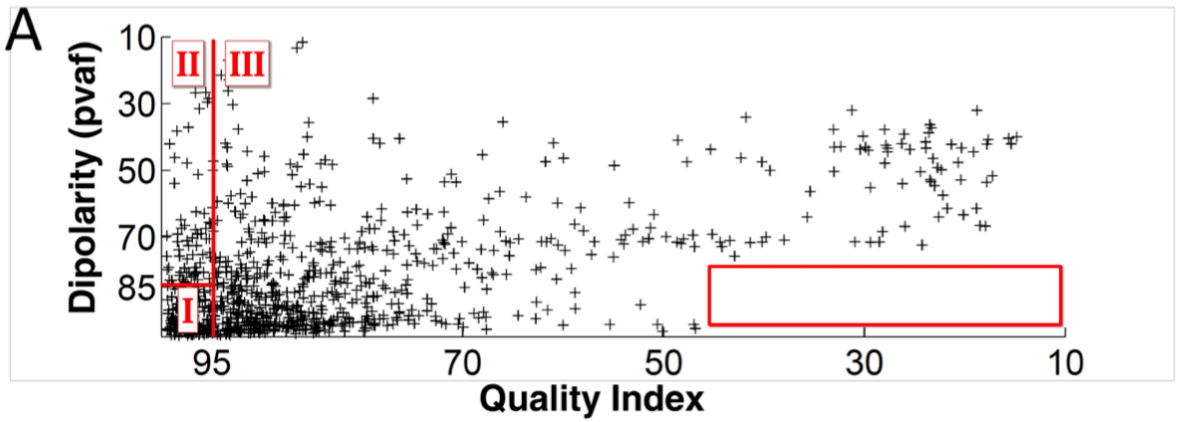
Within-cluster reliability

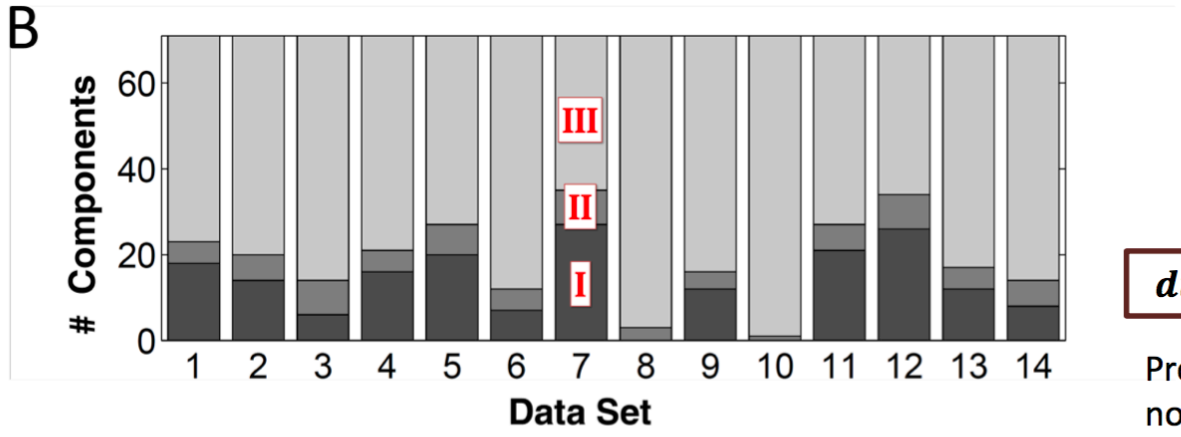
The distribution of dipolarity within the cluster helps assessing the **quality** and characteristics of Independent Components



Reliability criteria and the rv<15%

First justification why we should select an **r.v <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

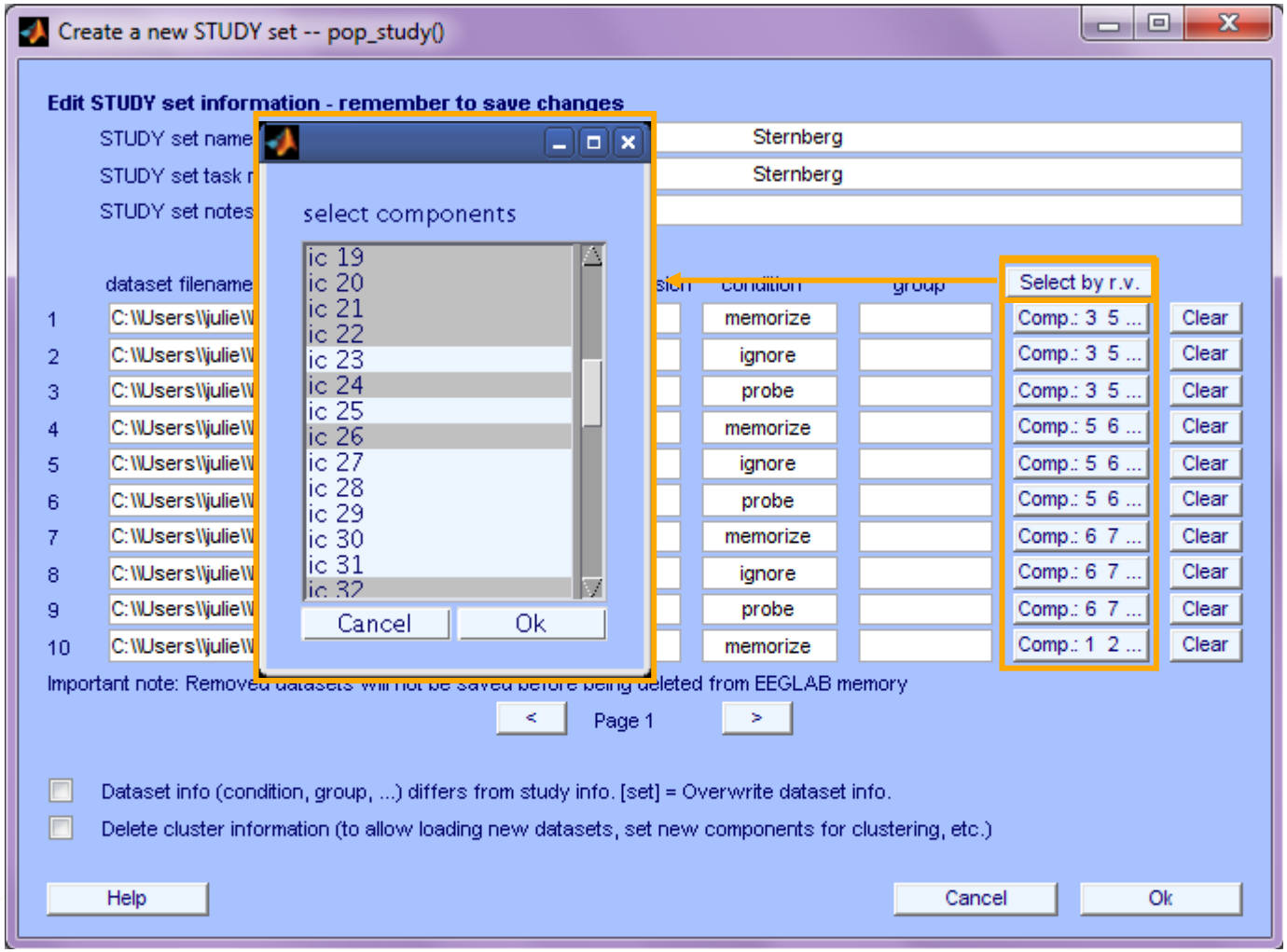
$dip \pm std > th$

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

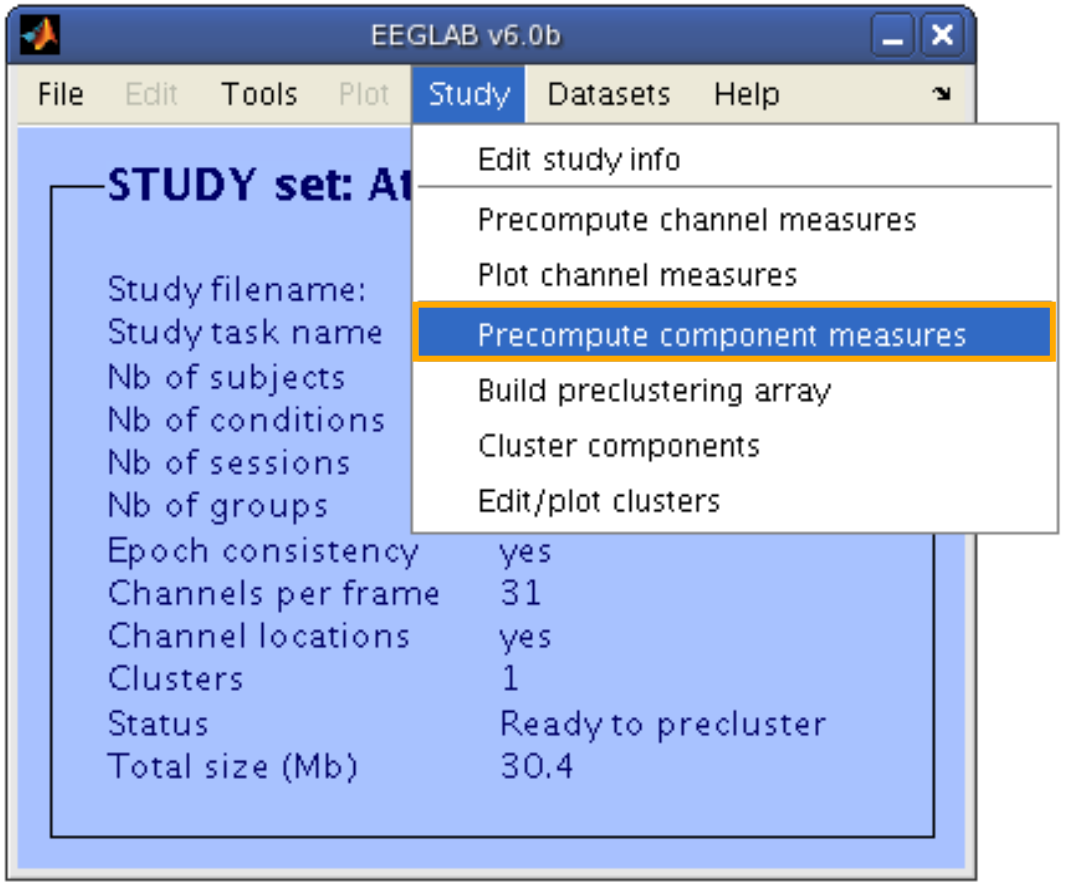
$dip \pm std < th$

Discard

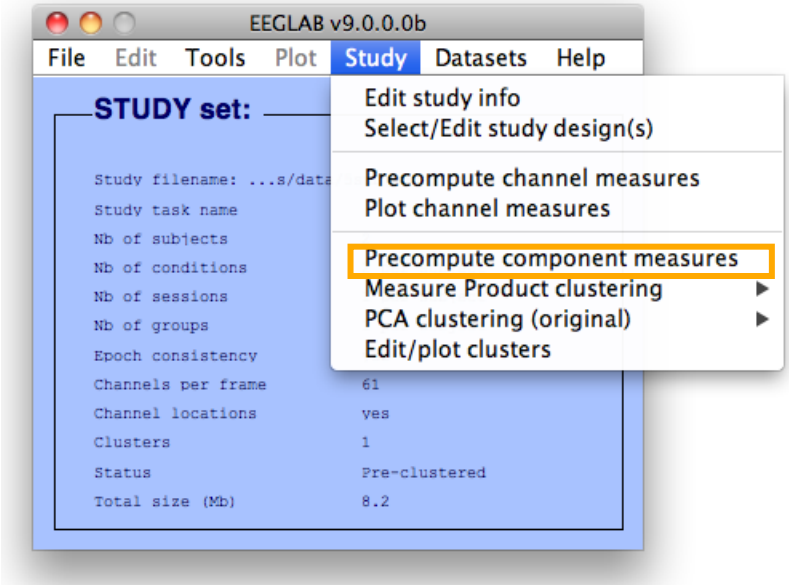
ICs to cluster

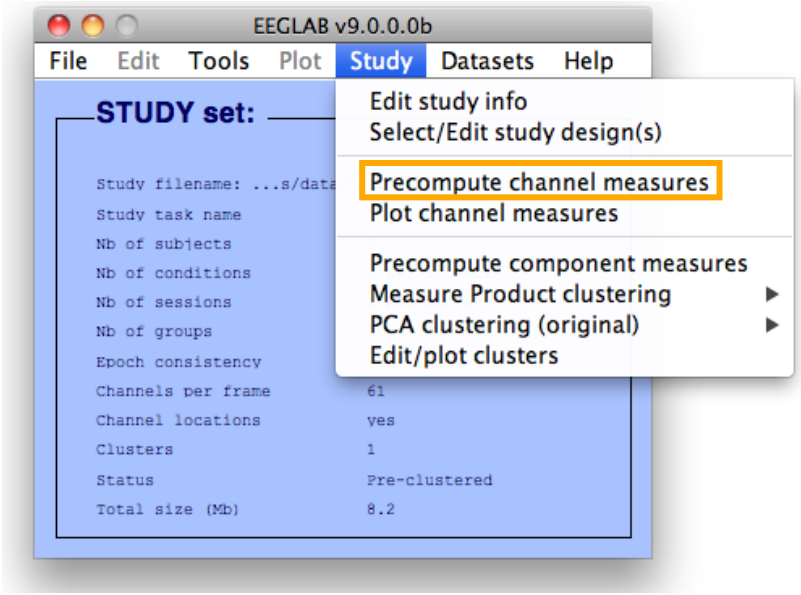


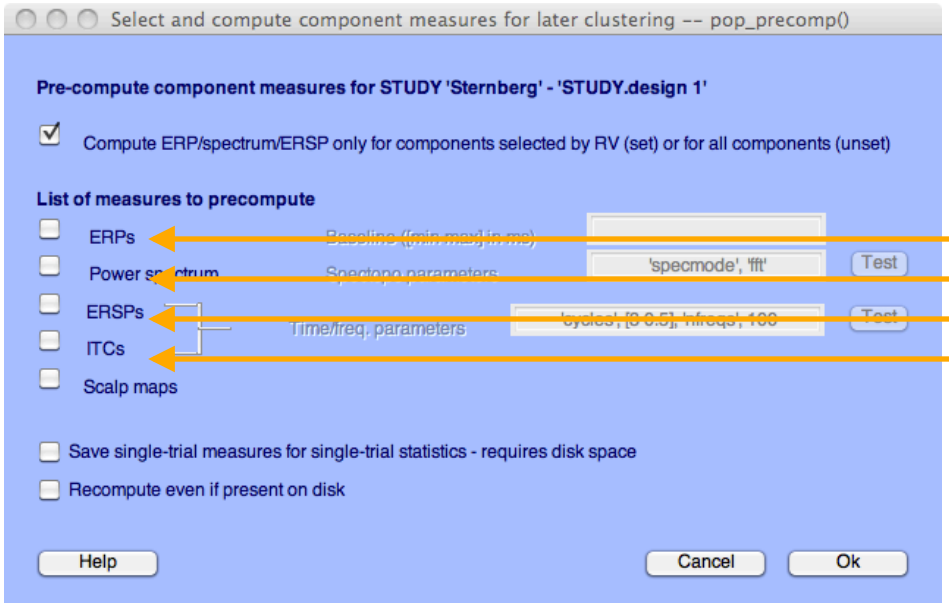
Precompute data measures

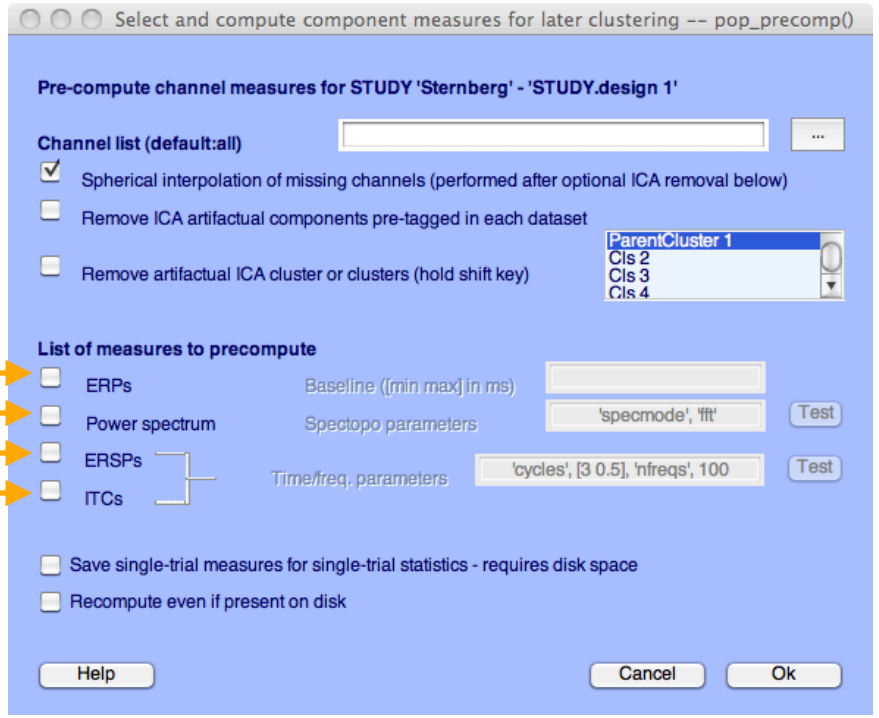


Pre-compute measures



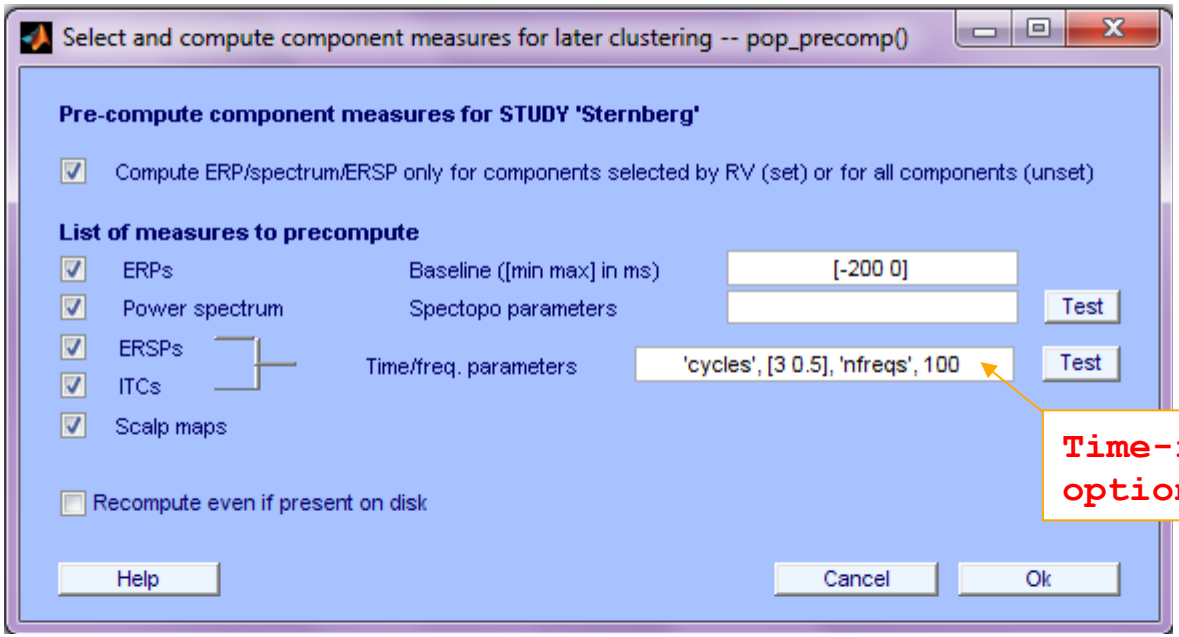






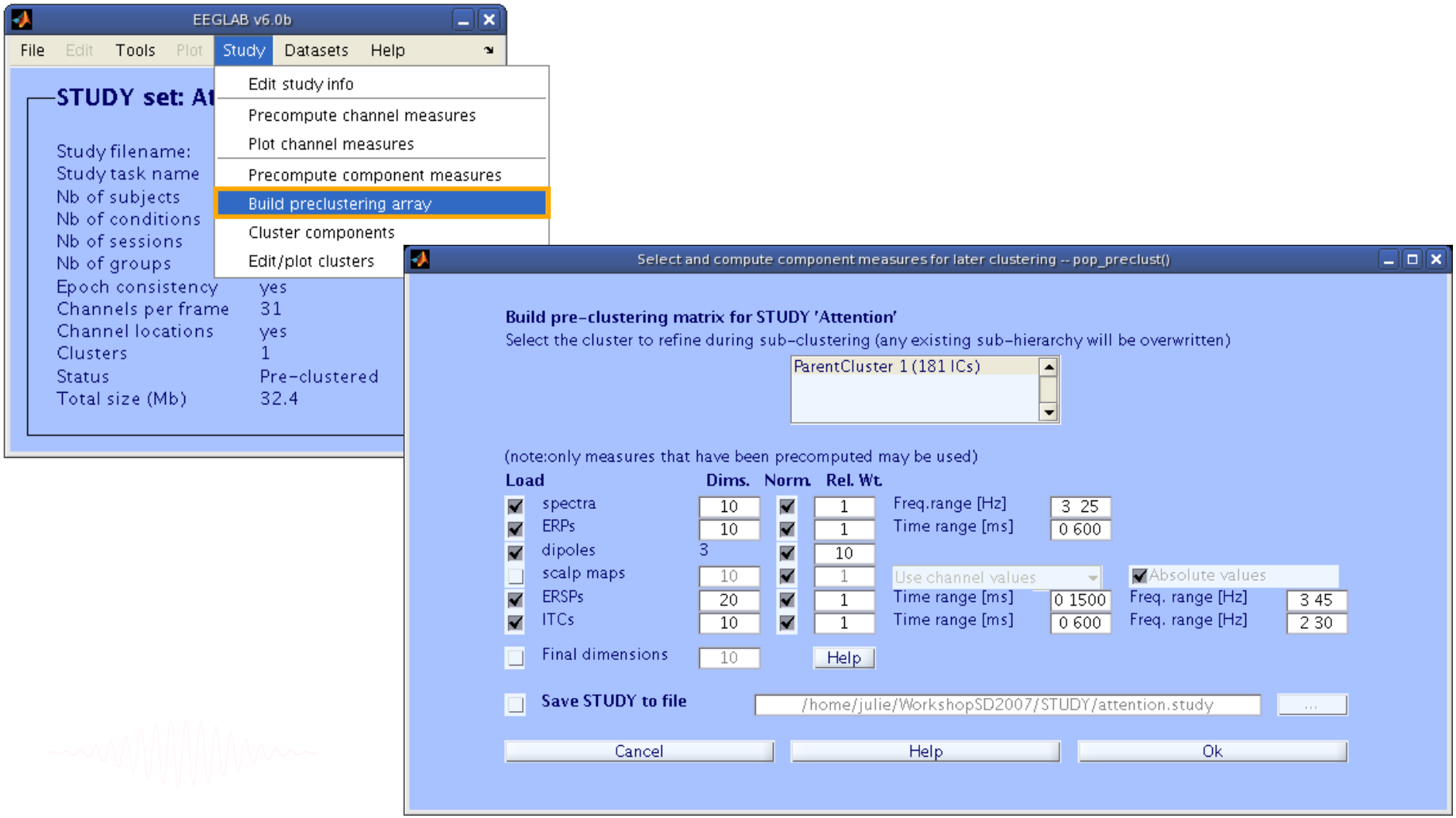
Precompute data measures

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

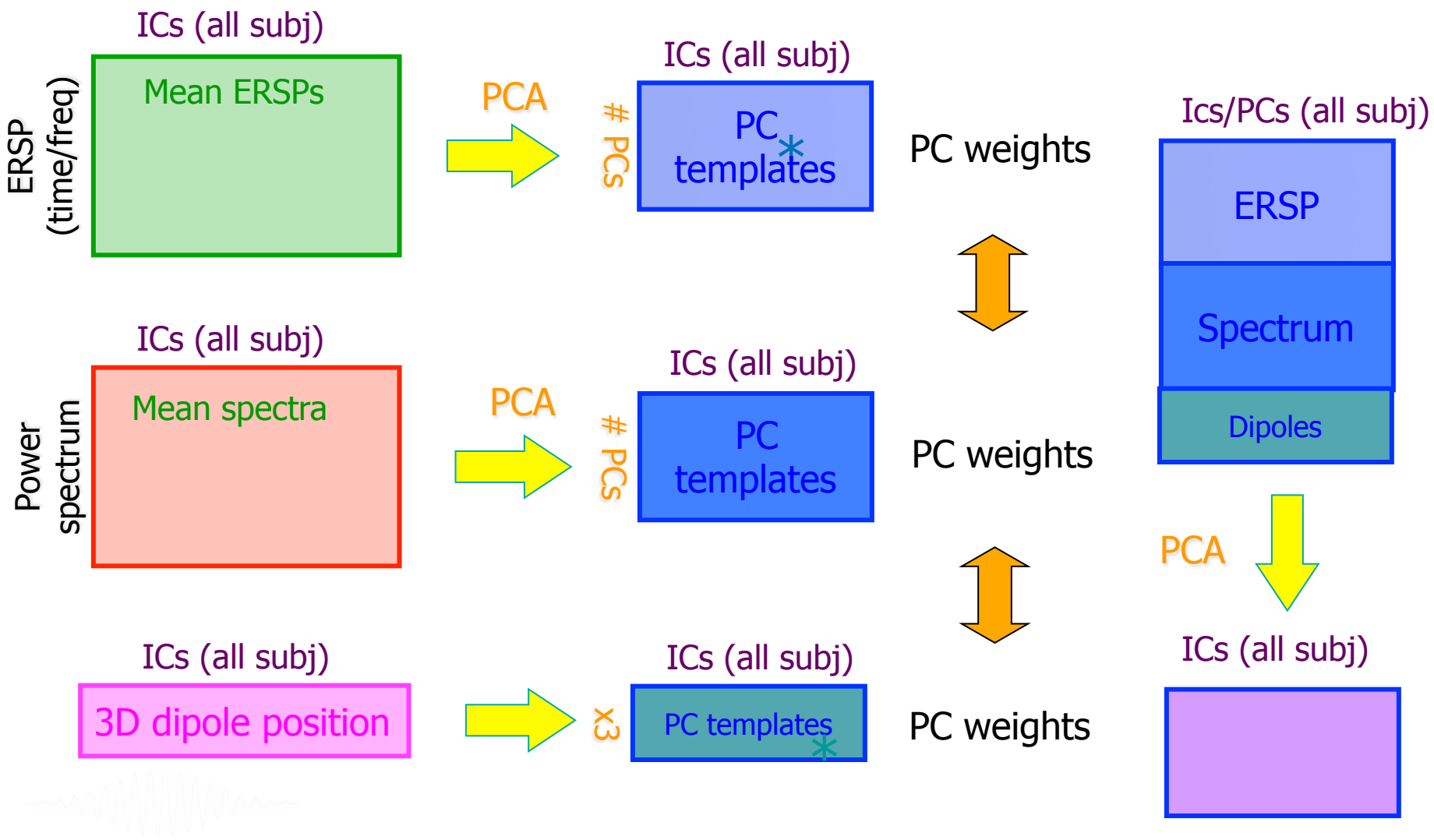


Time-frequency options

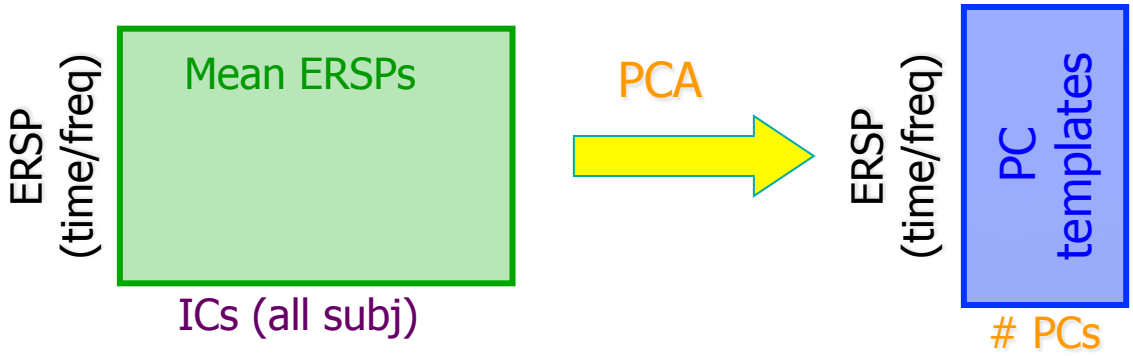
Cluster components

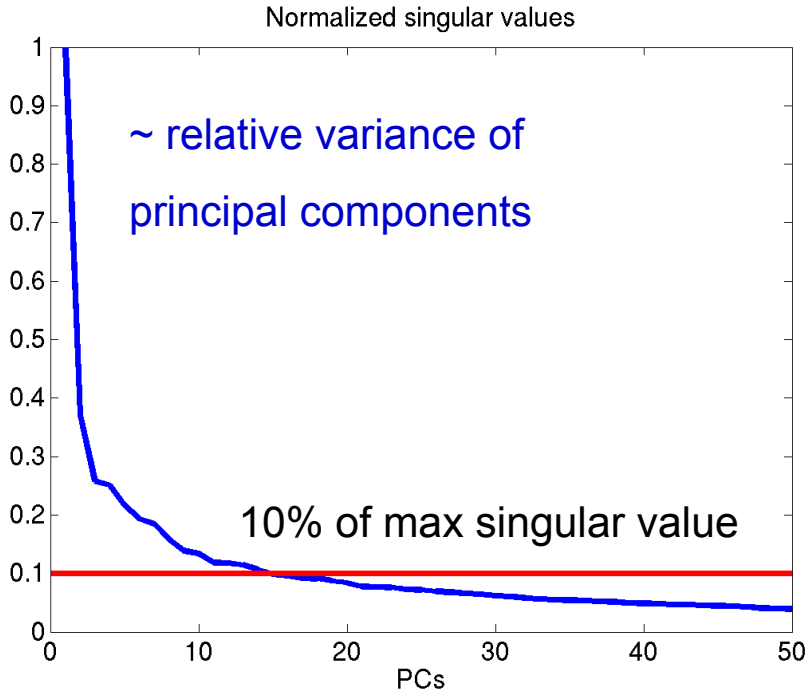


Precluster schematic



Precluster: Use singular values from PCA





Precluster schematic

ICs (all subj)

ERSP

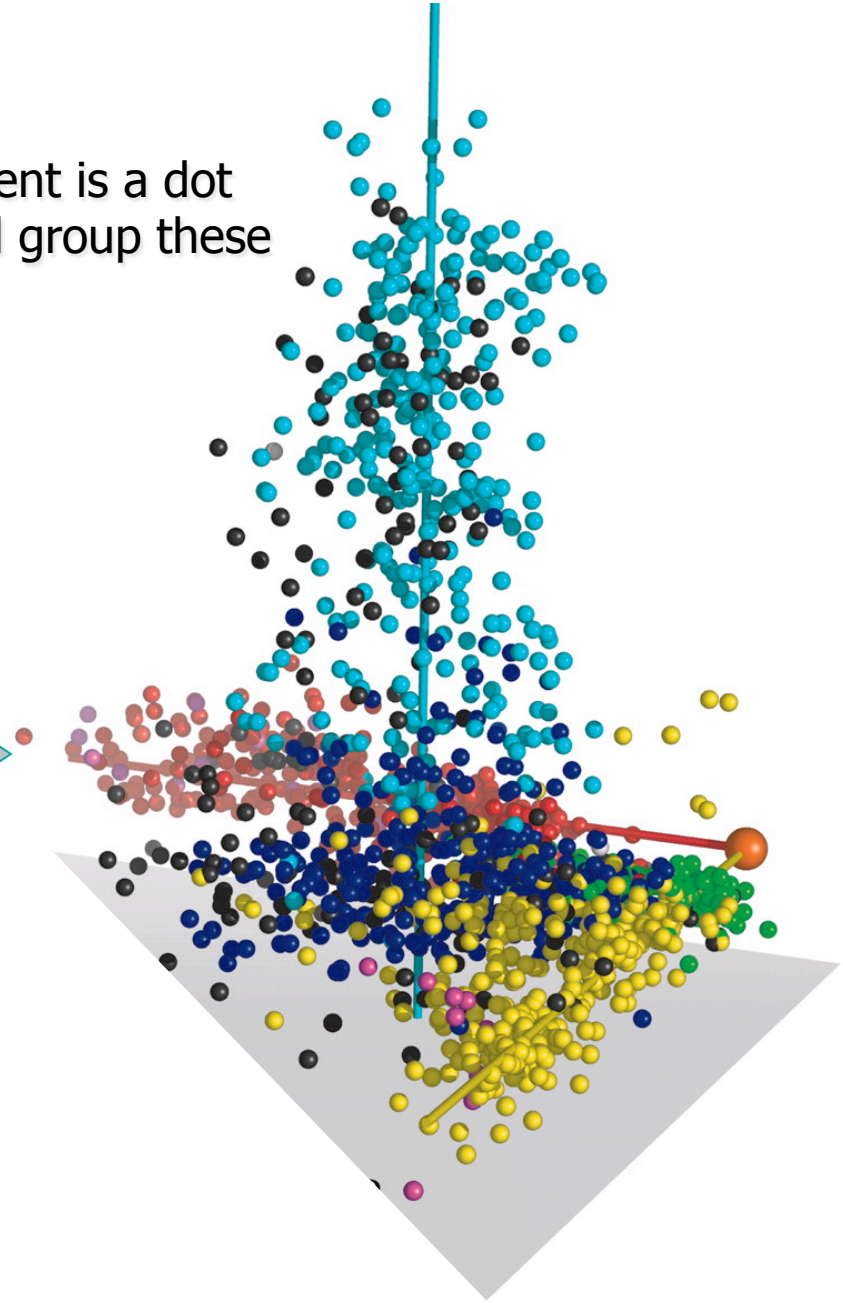
Spectrum

Dipoles

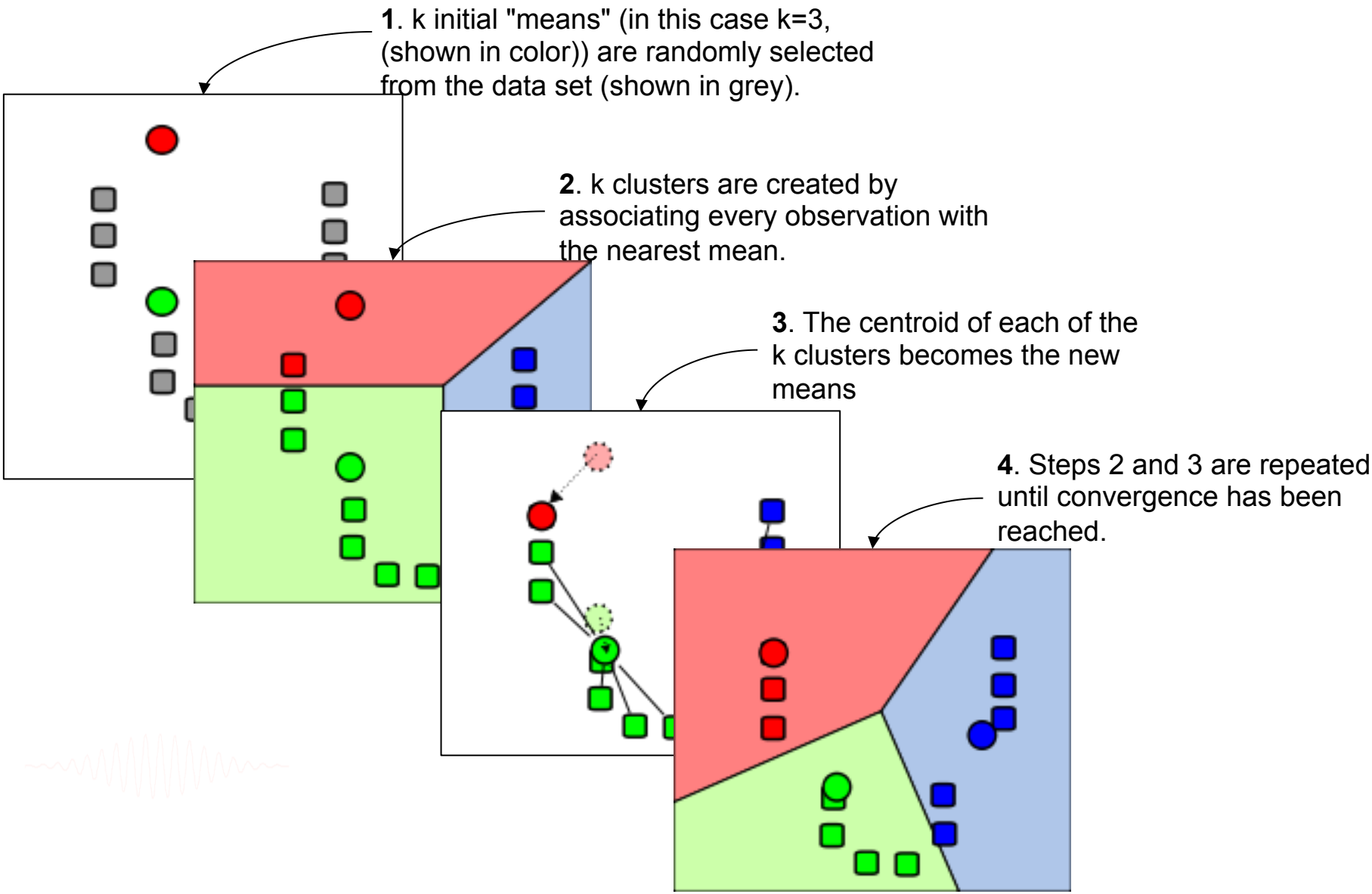
OR

ICs (all subj)

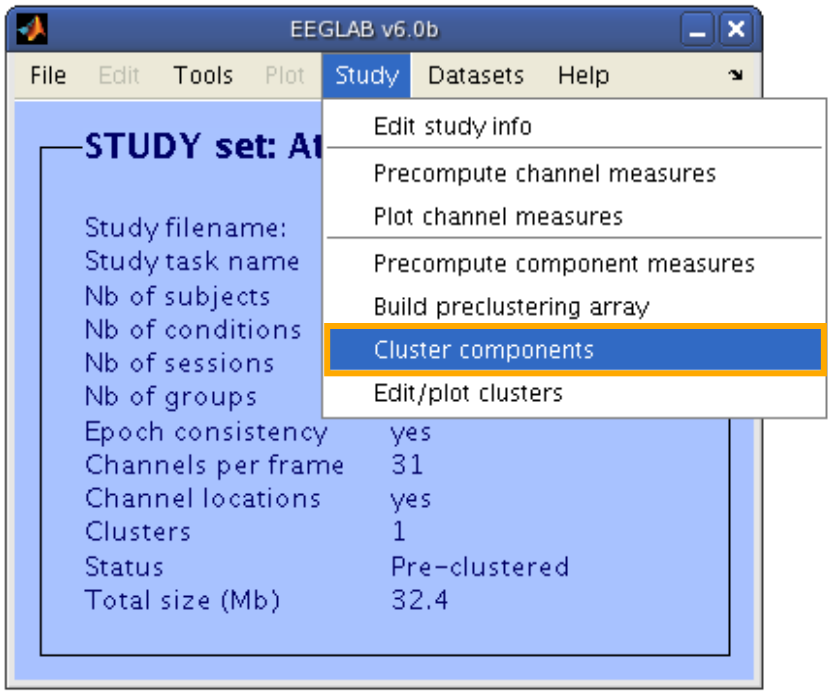
Each component is a dot
Clustering will group these dots

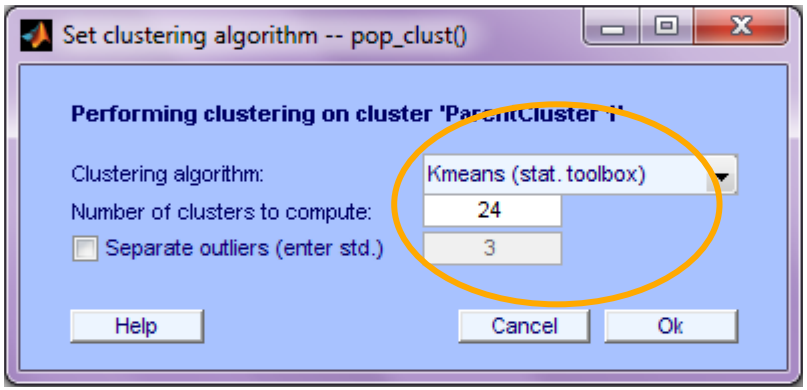


Classical KMean



Cluster components





Choosing data measures

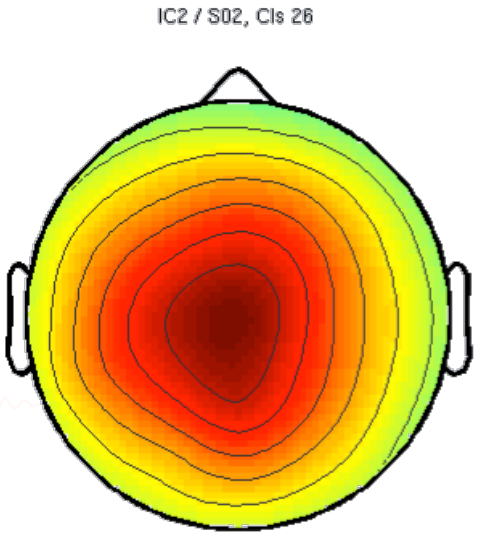
What measure(s) should you use?

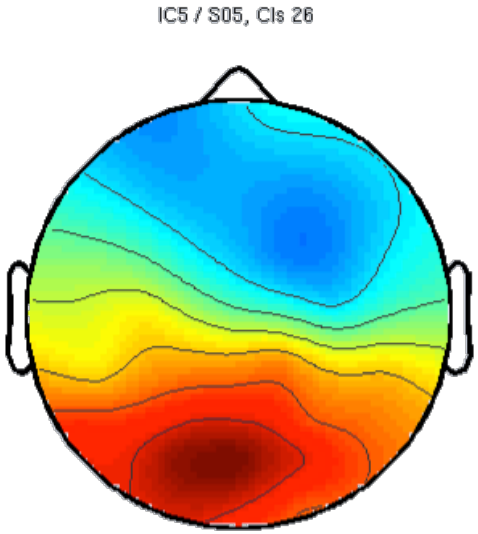
It depends on your final cluster criteria…

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

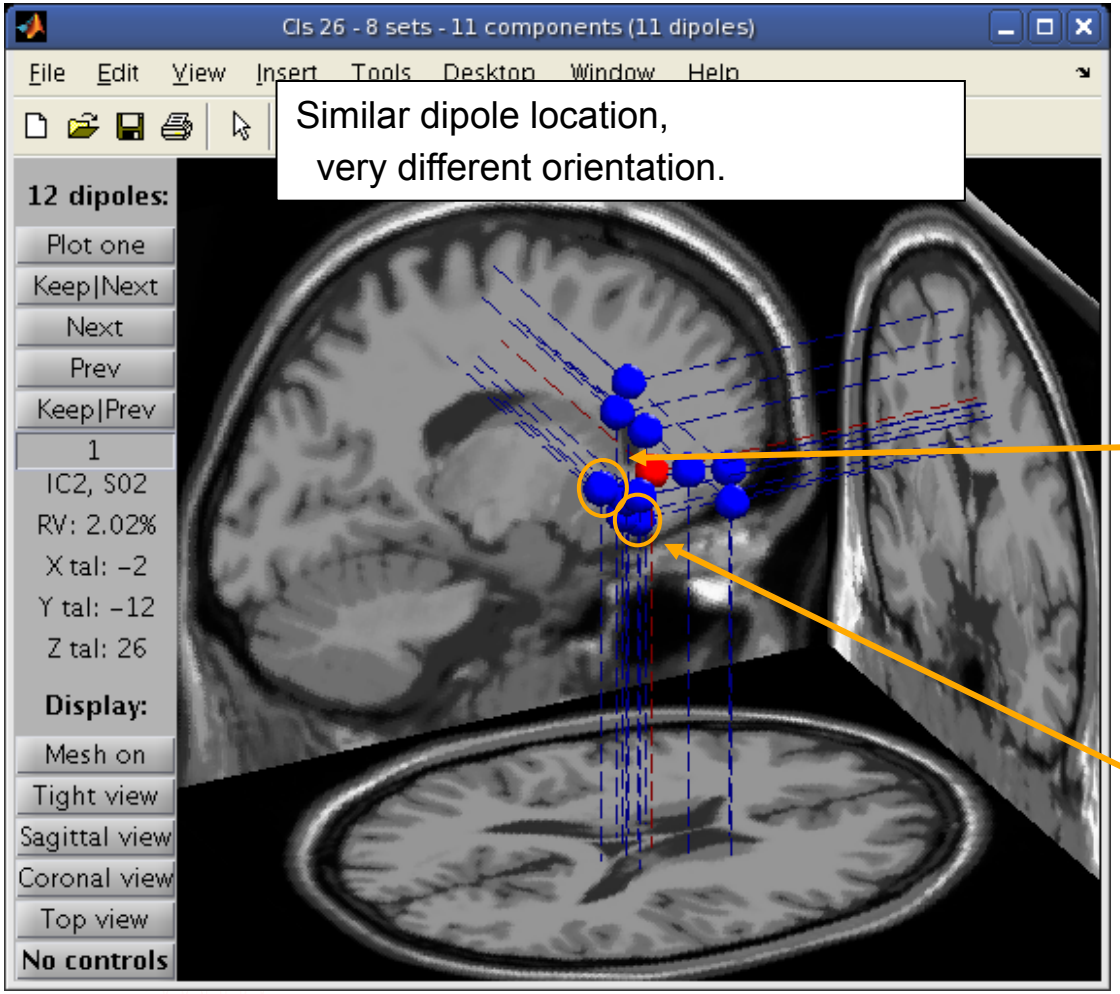
But consider:

- What is the difference between these two components?



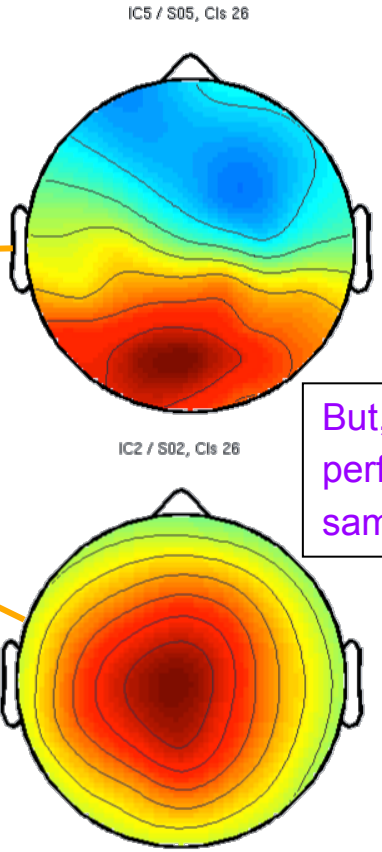


Choosing data measures



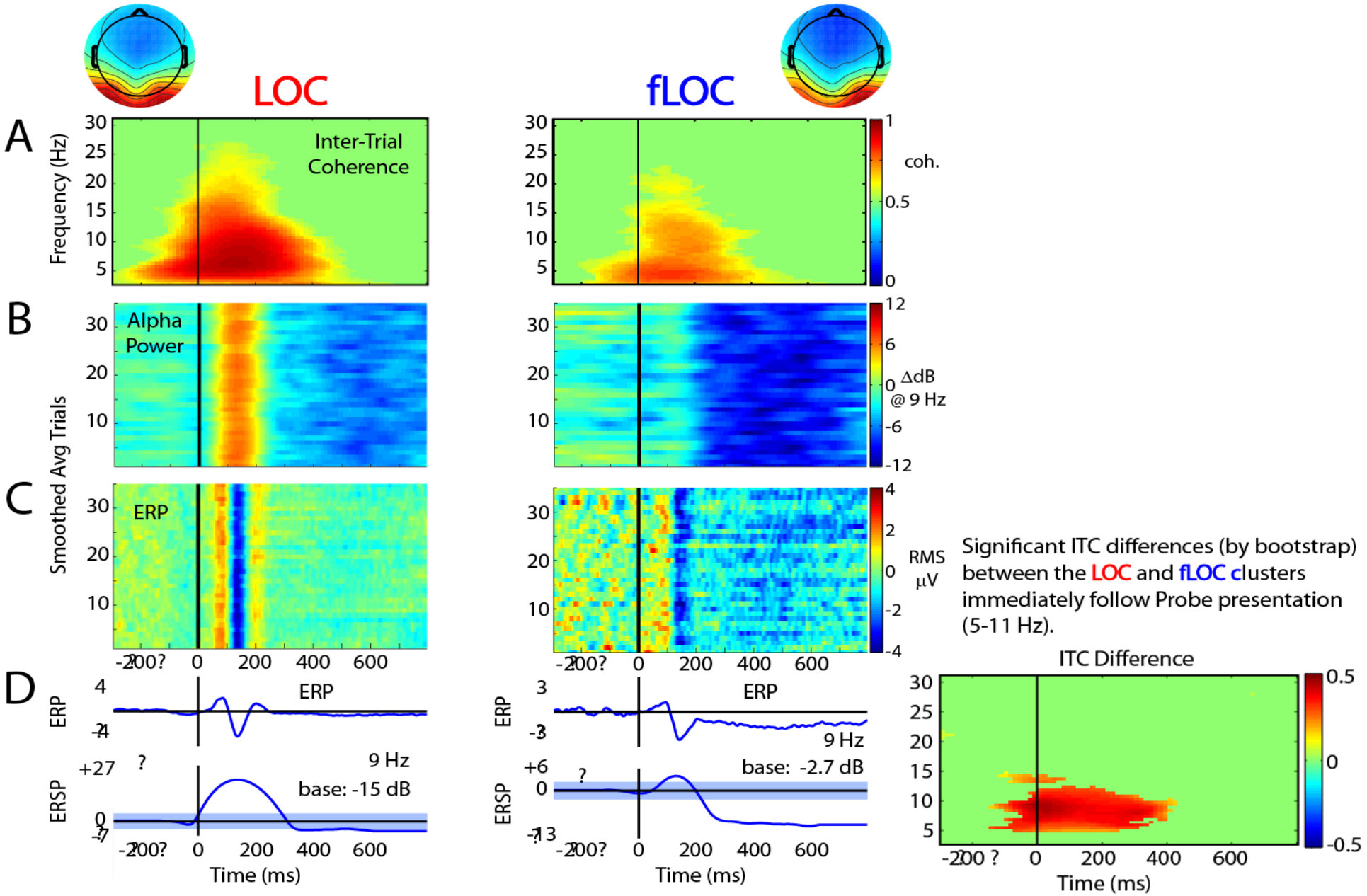
Similar dipole location,
very different orientation.

Obvious dramatic effect on scalp map topography:

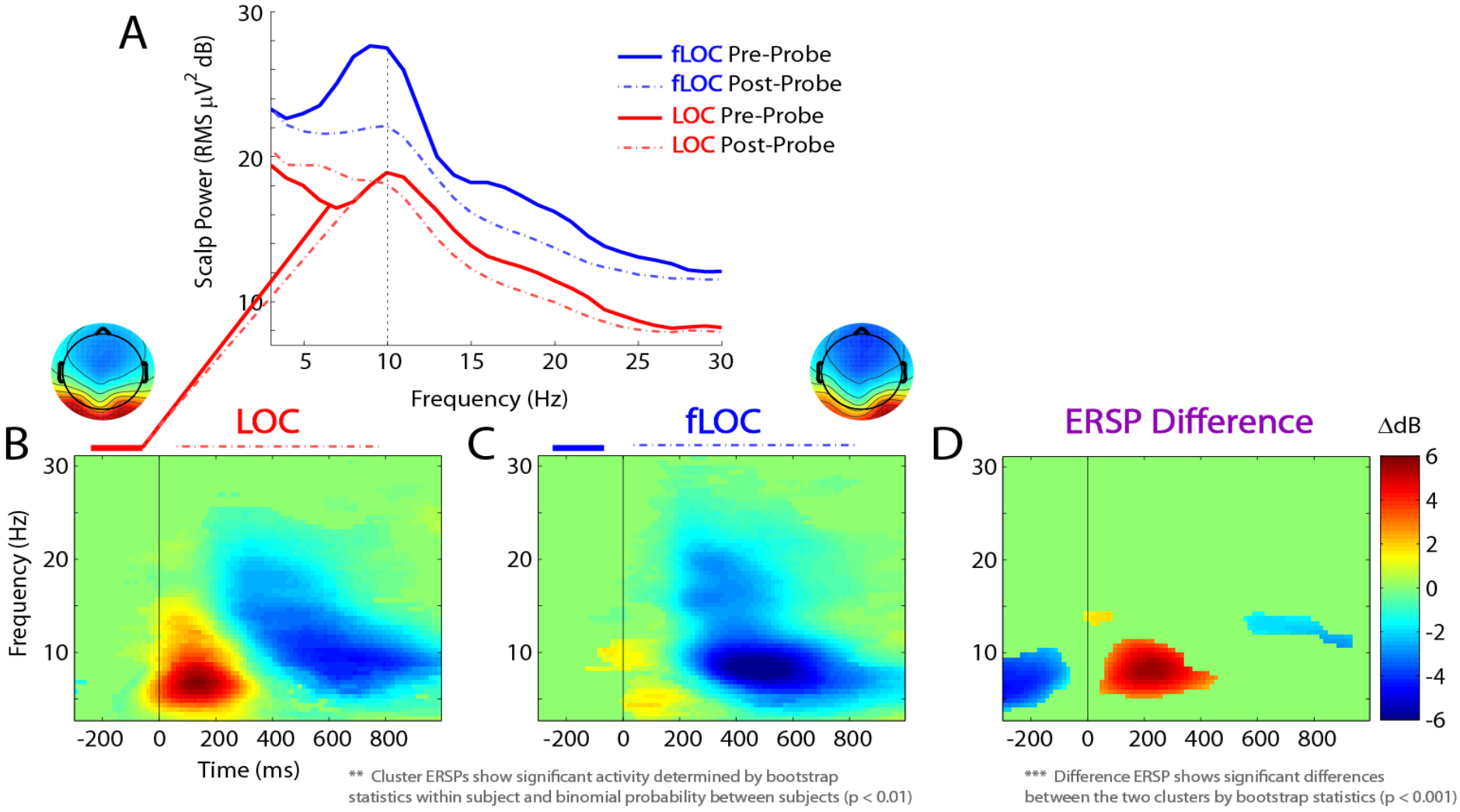


But, do they perform the same functions?

Subject differences?



Subject differences?

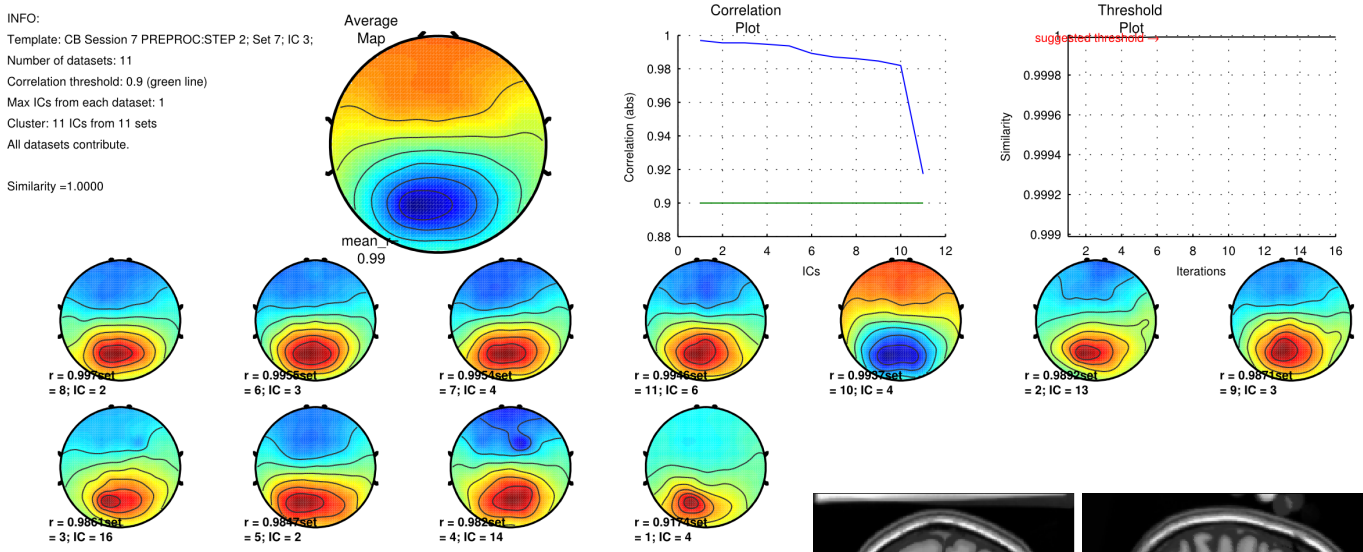


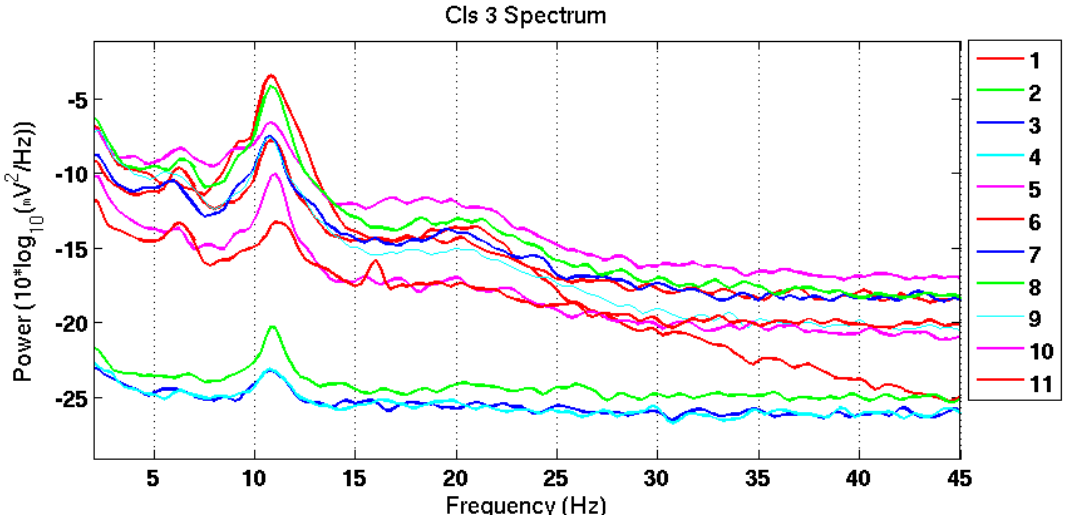
** Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects ($p < 0.01$)

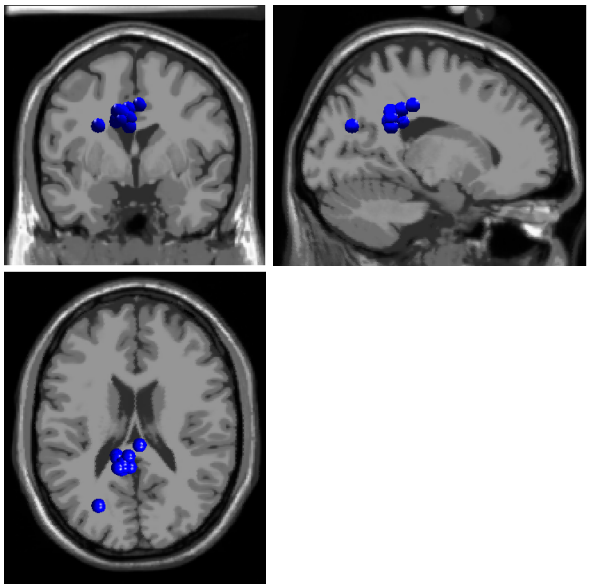
*** Difference ERSP shows significant differences between the two clusters by bootstrap statistics ($p < 0.001$)

Results (Cluster 1)

100 % Sessions contribute

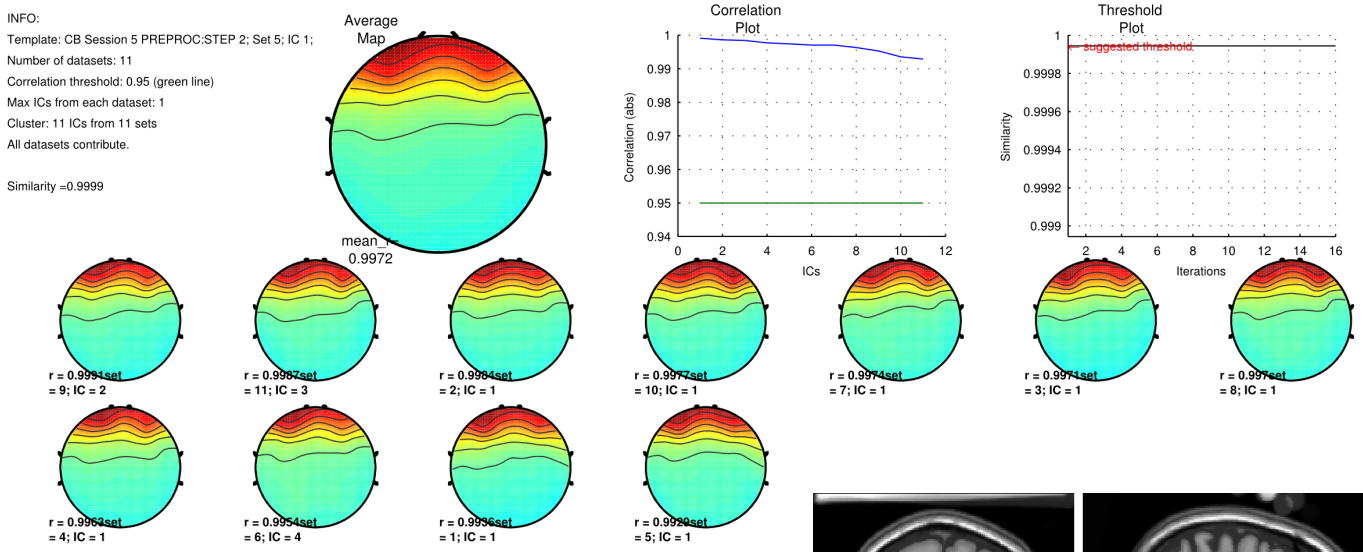


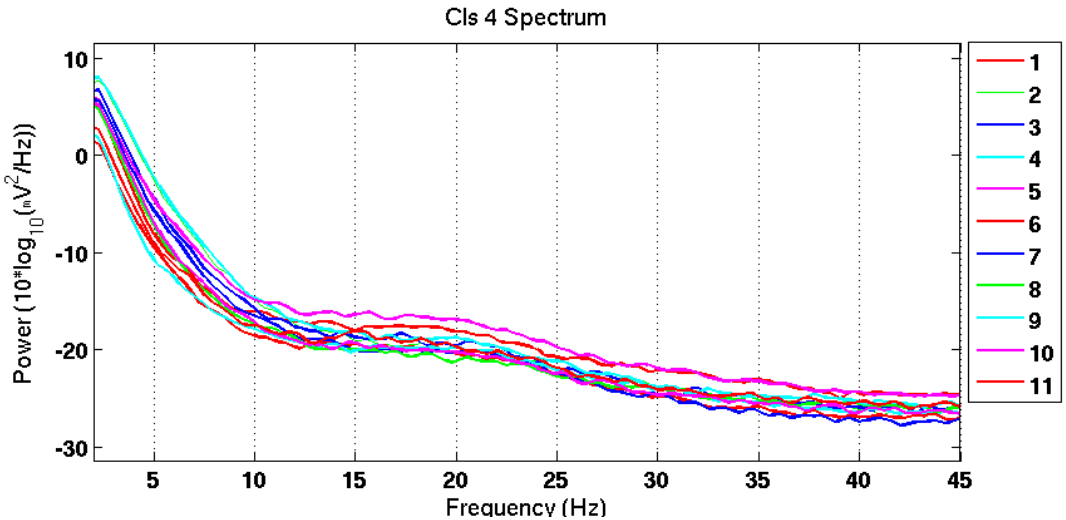


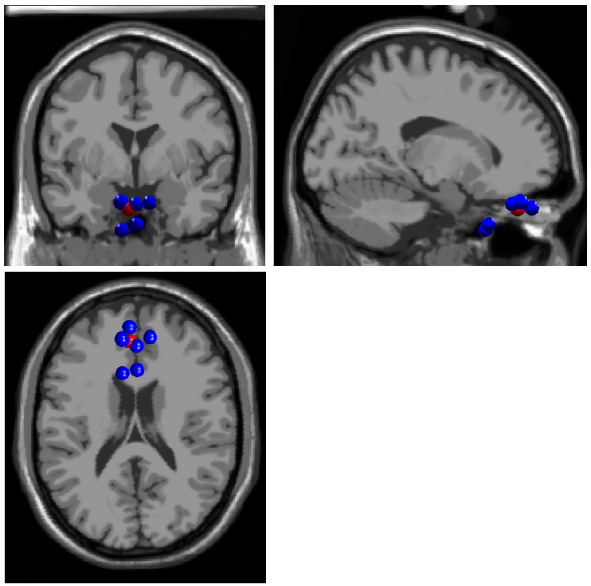


Results (Cluster 2)

100 % Sessions contribute

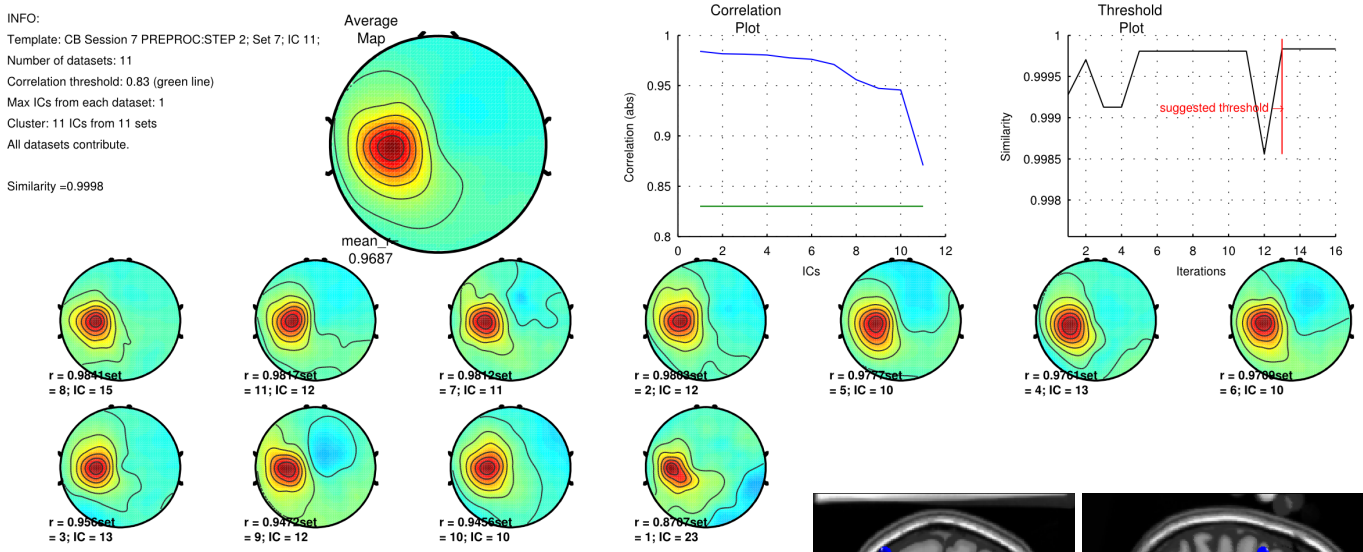


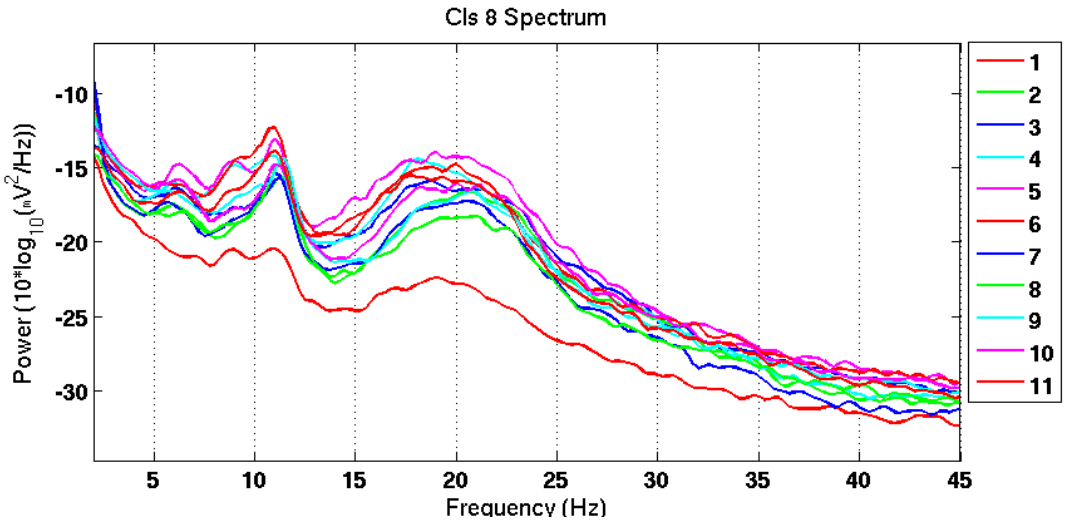


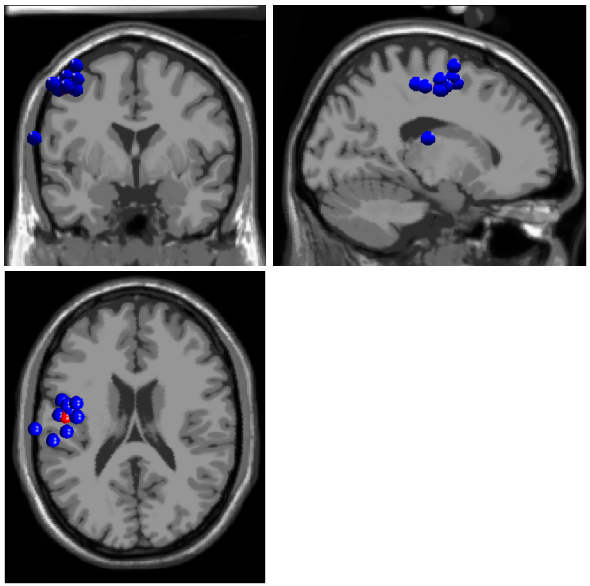


Results (Cluster 8)

100 % Sessions contribute

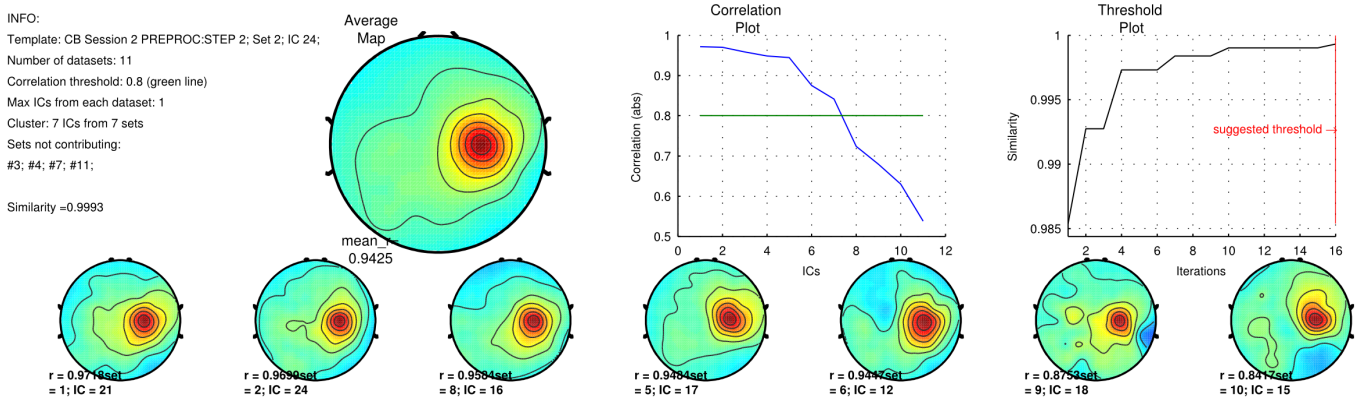


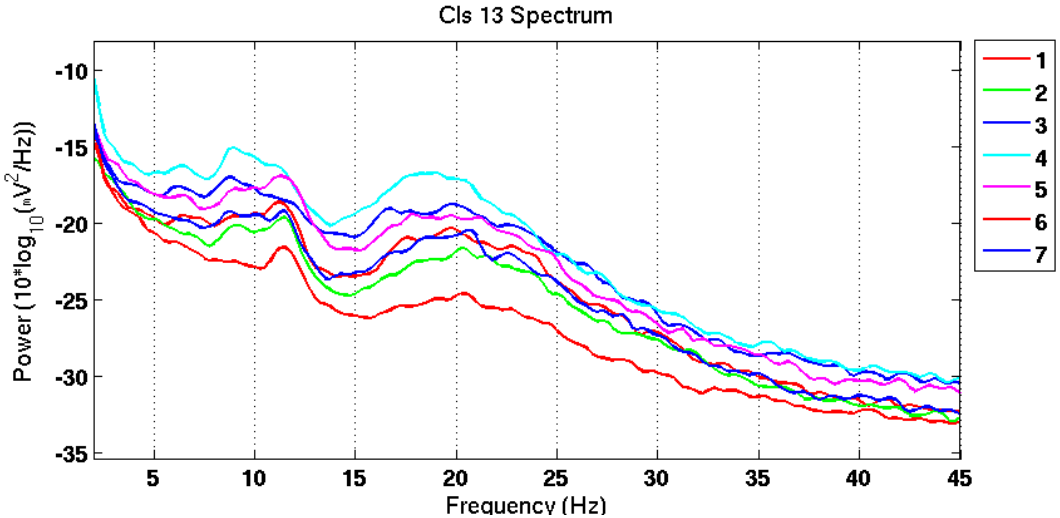


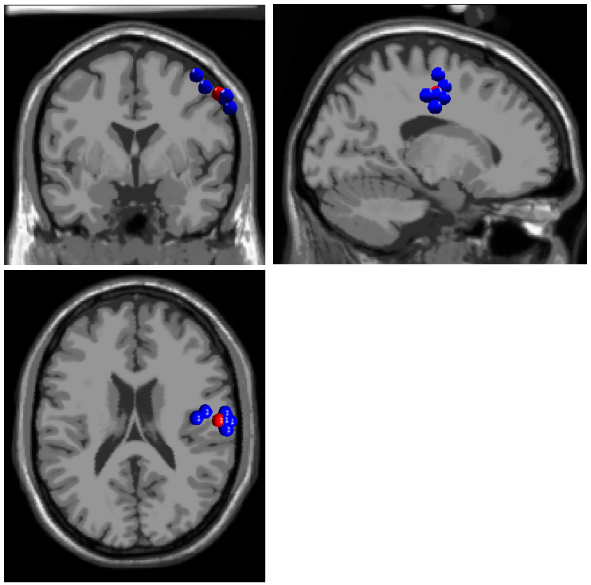


Results (Cluster 13)

63.64% Sessions contribute

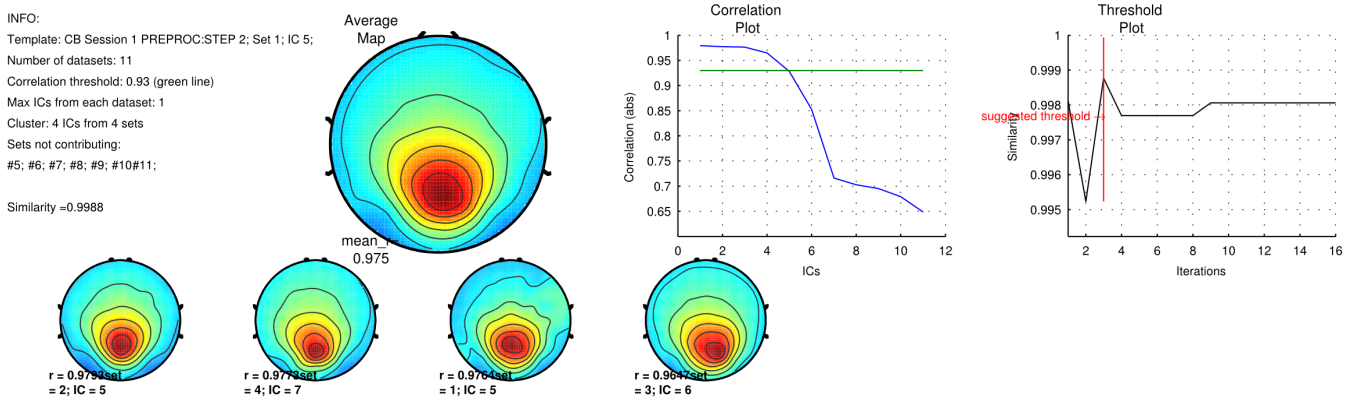


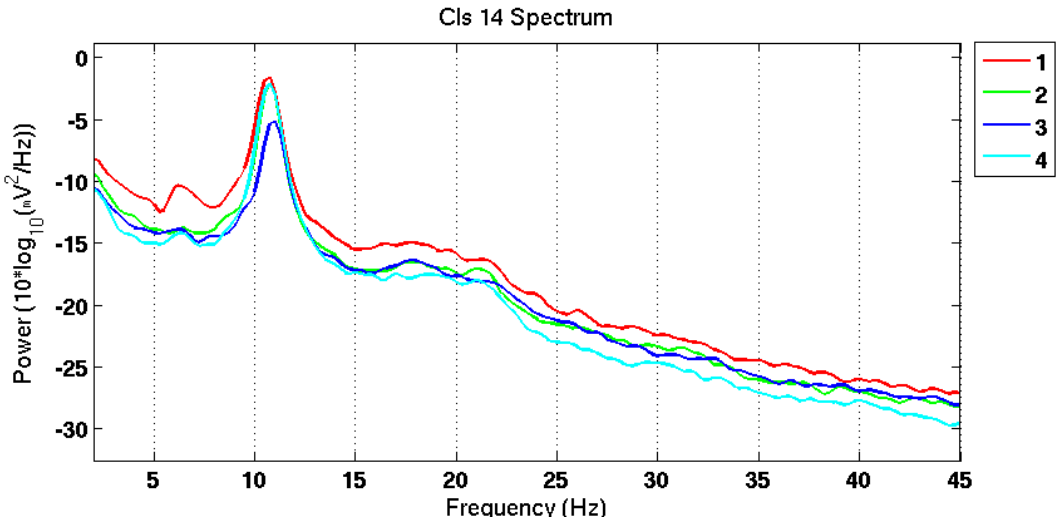


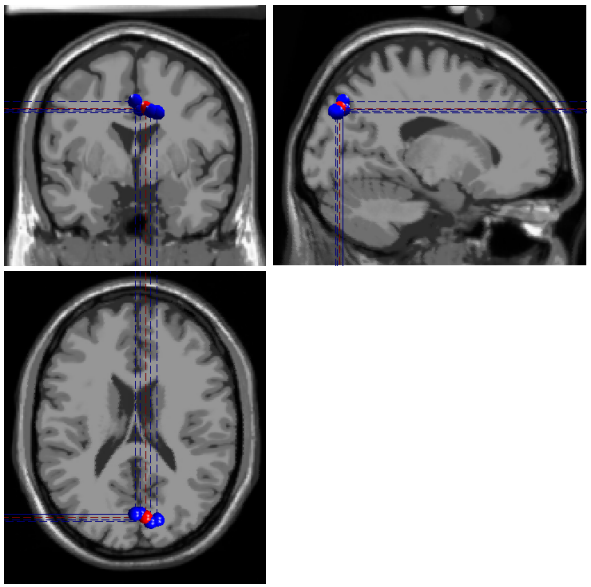


Results (Cluster 14)

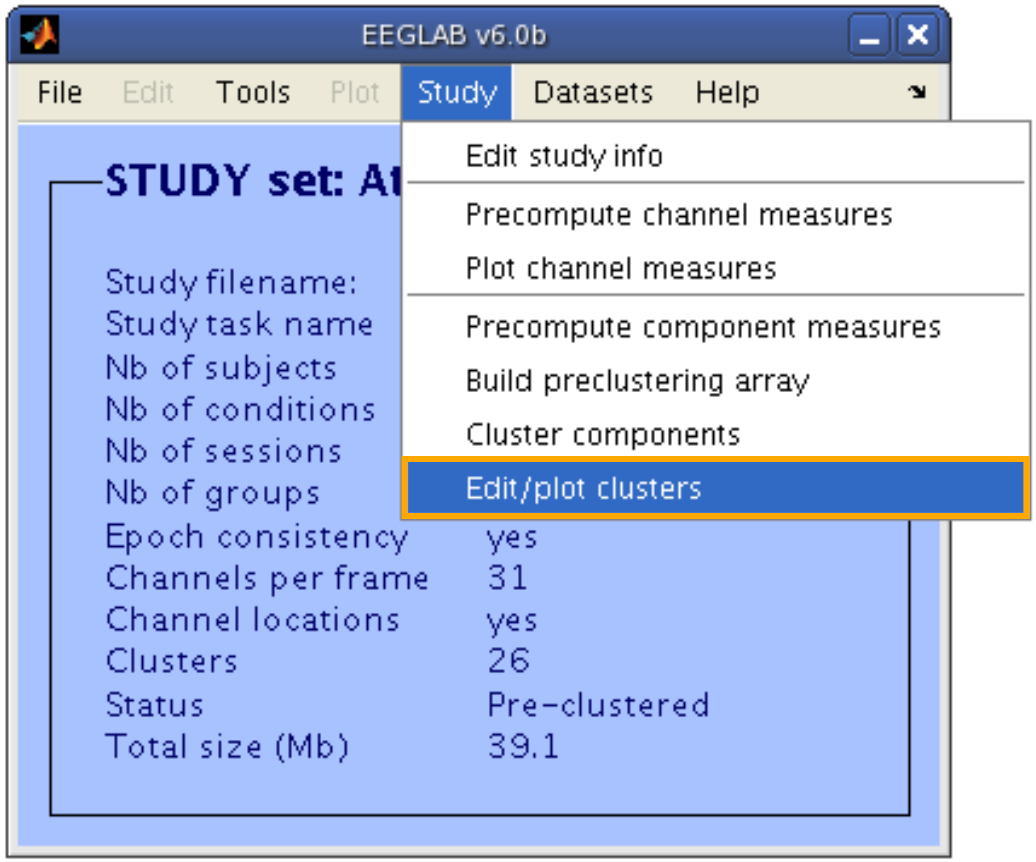
36.36% Sessions contribute



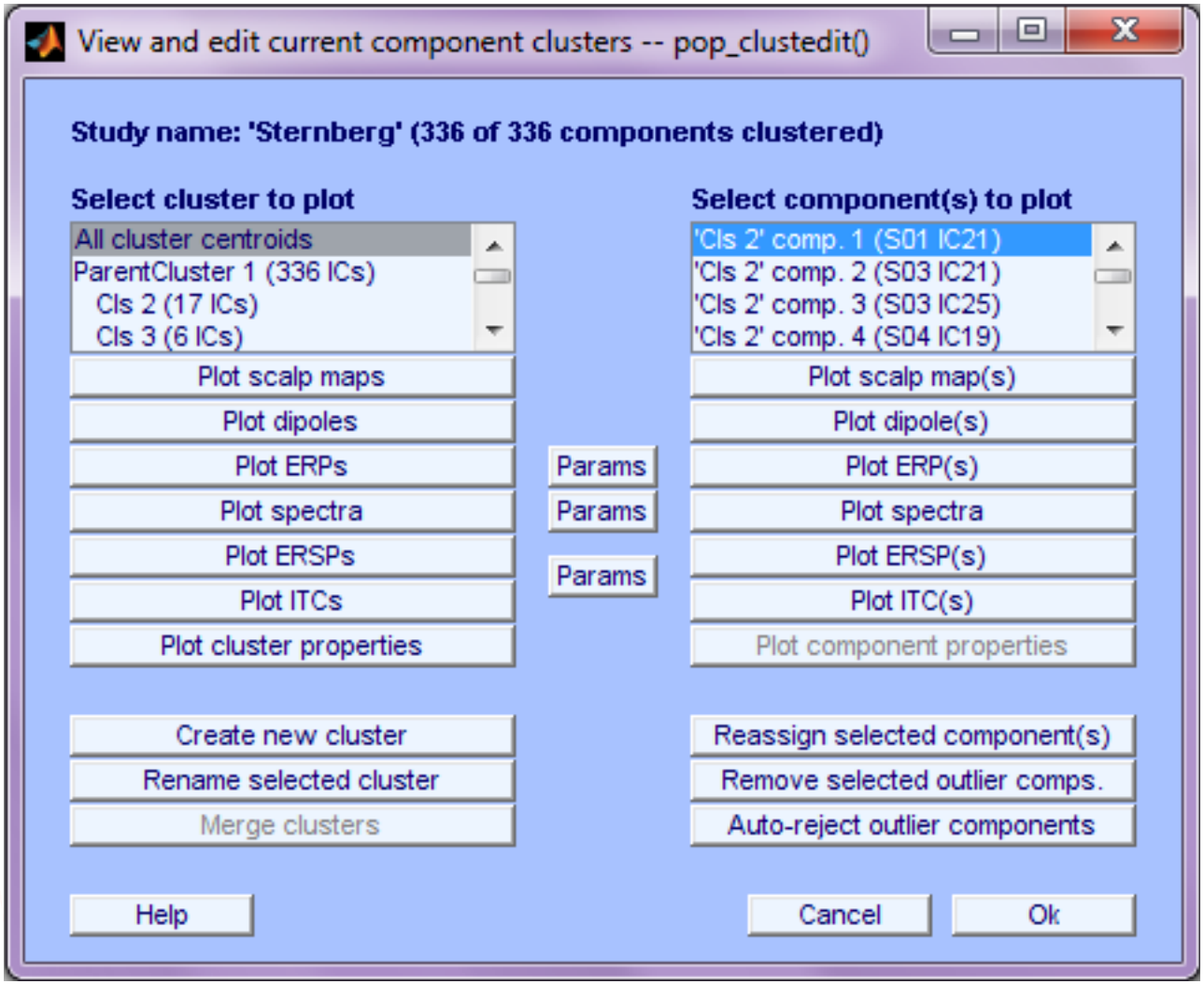




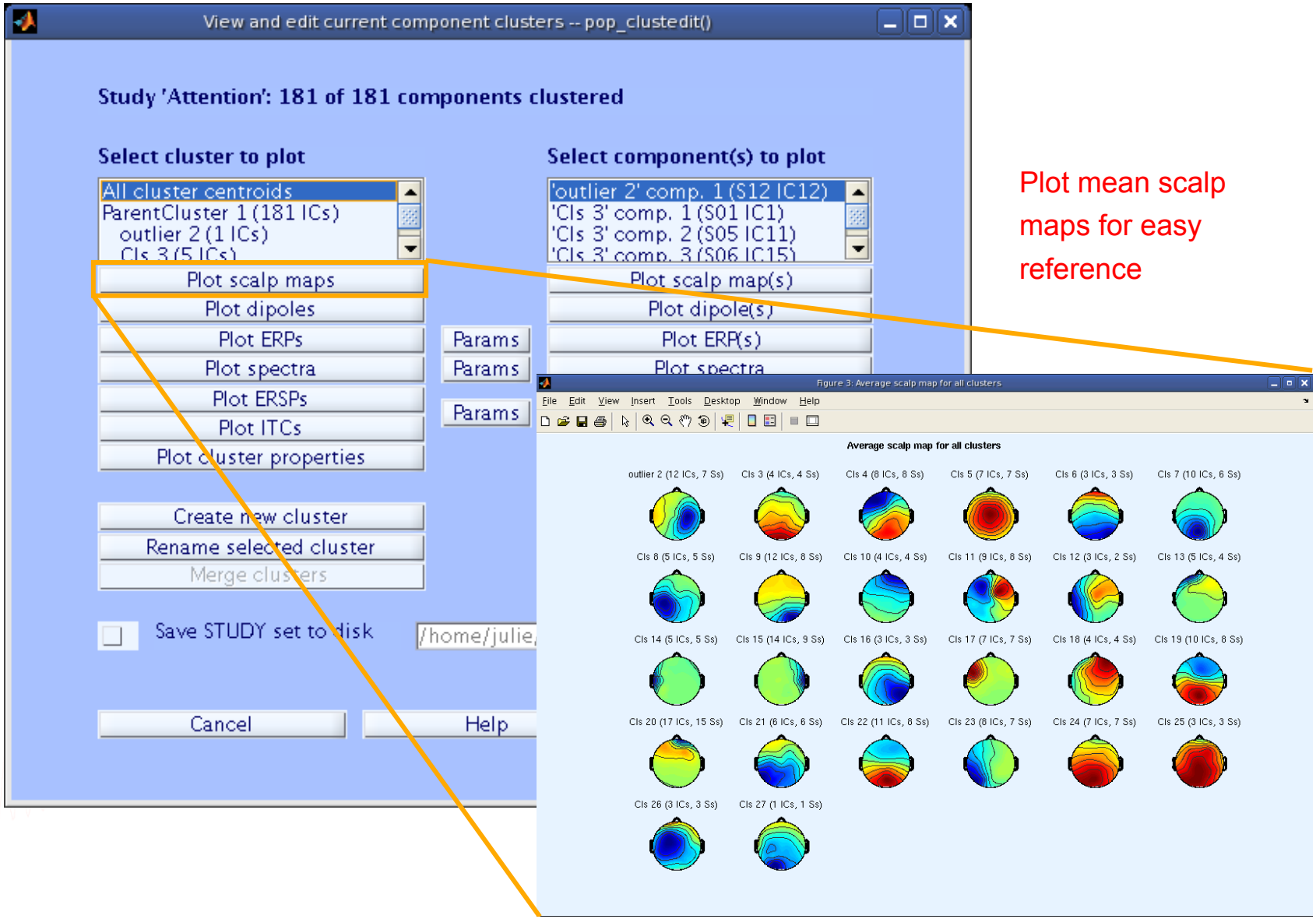
View and edit clusters



Plot/edit clusters

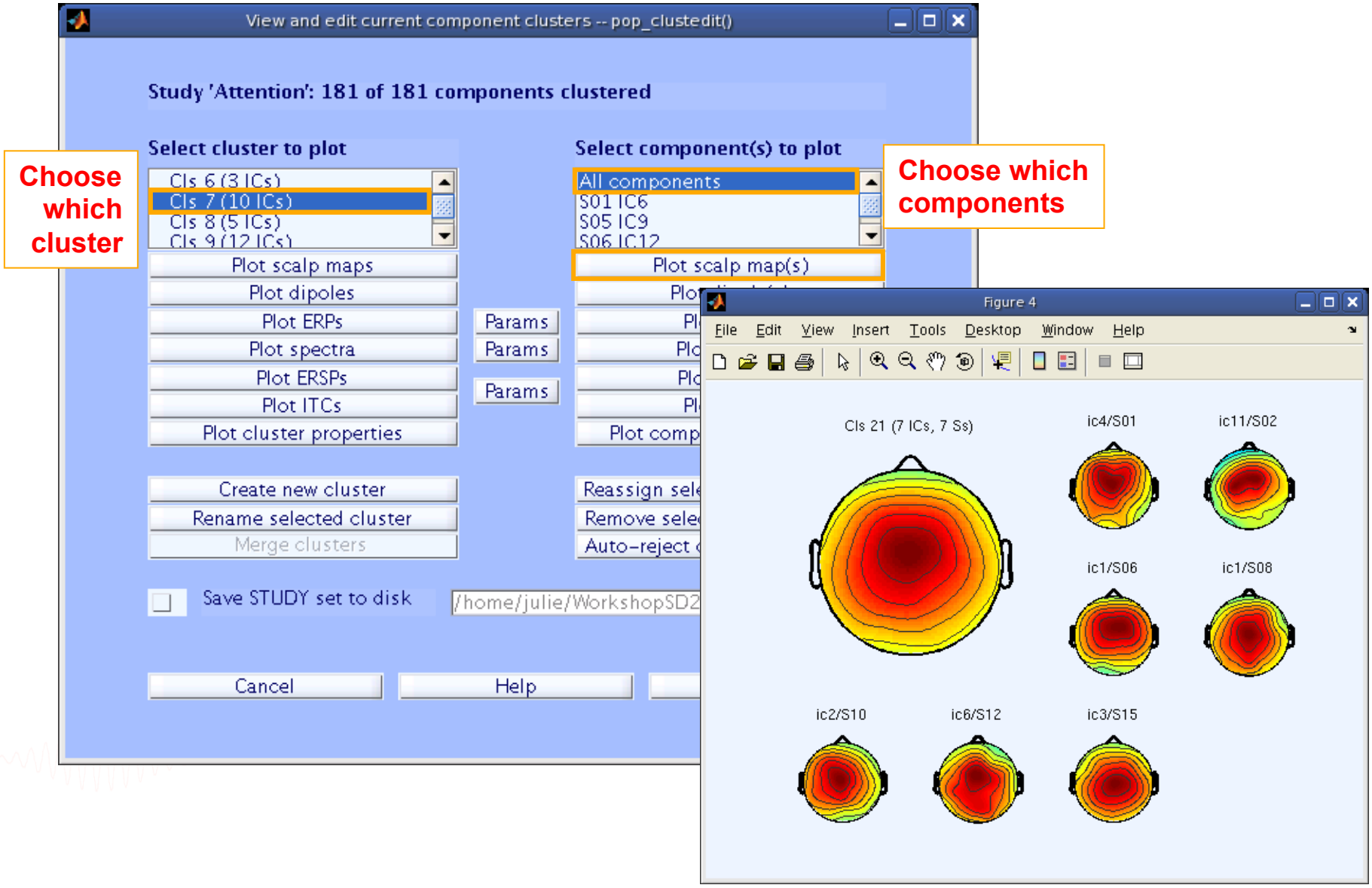


Plot cluster data

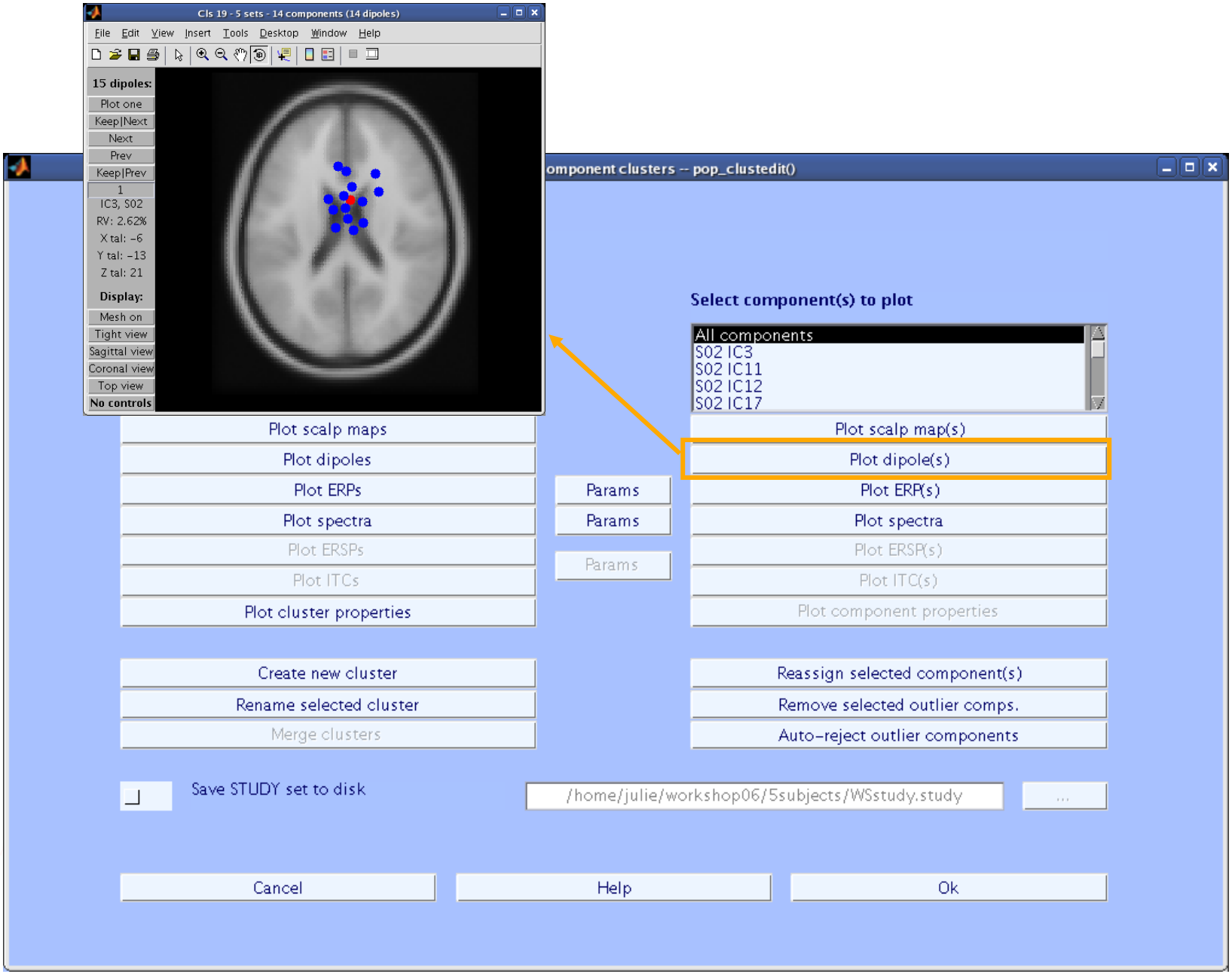


Plot mean scalp maps for easy reference

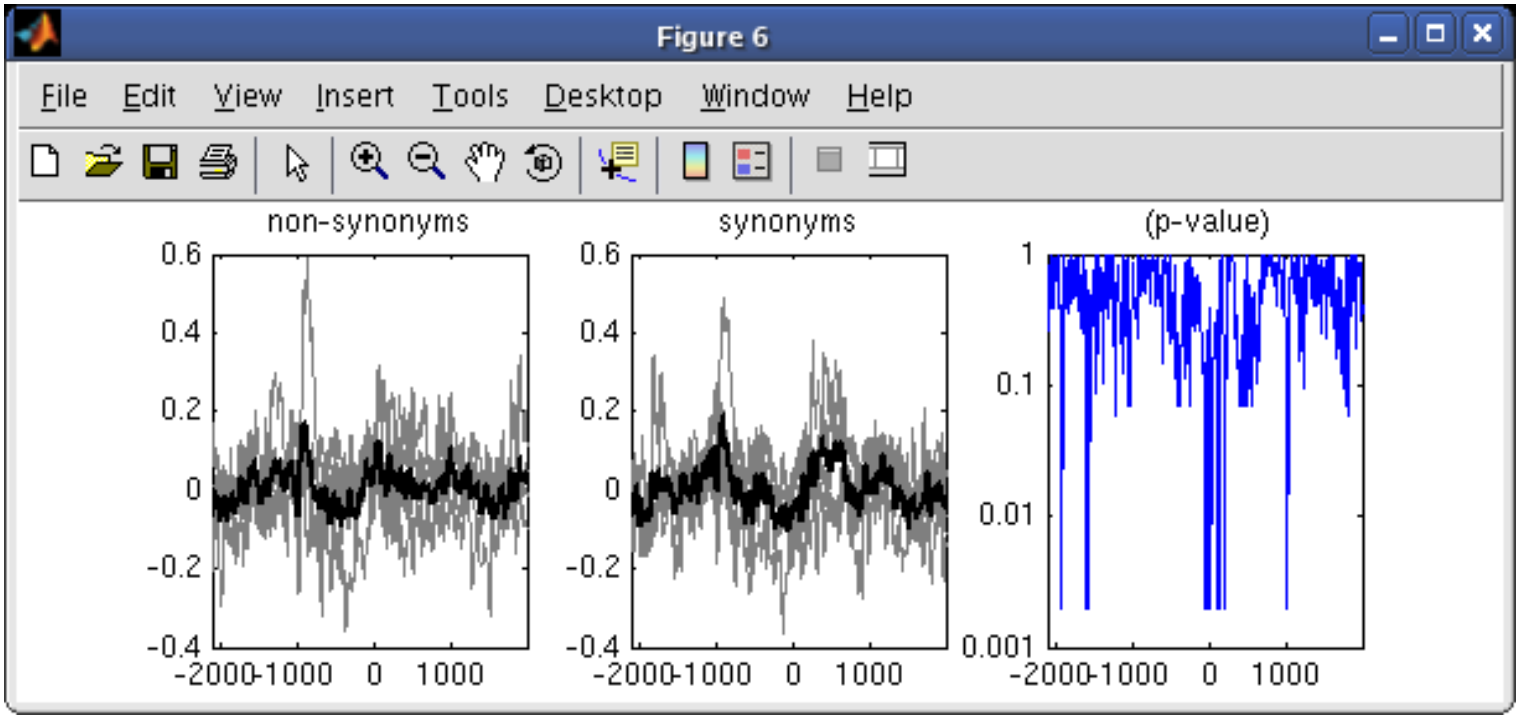
Plot cluster data



Plot cluster data



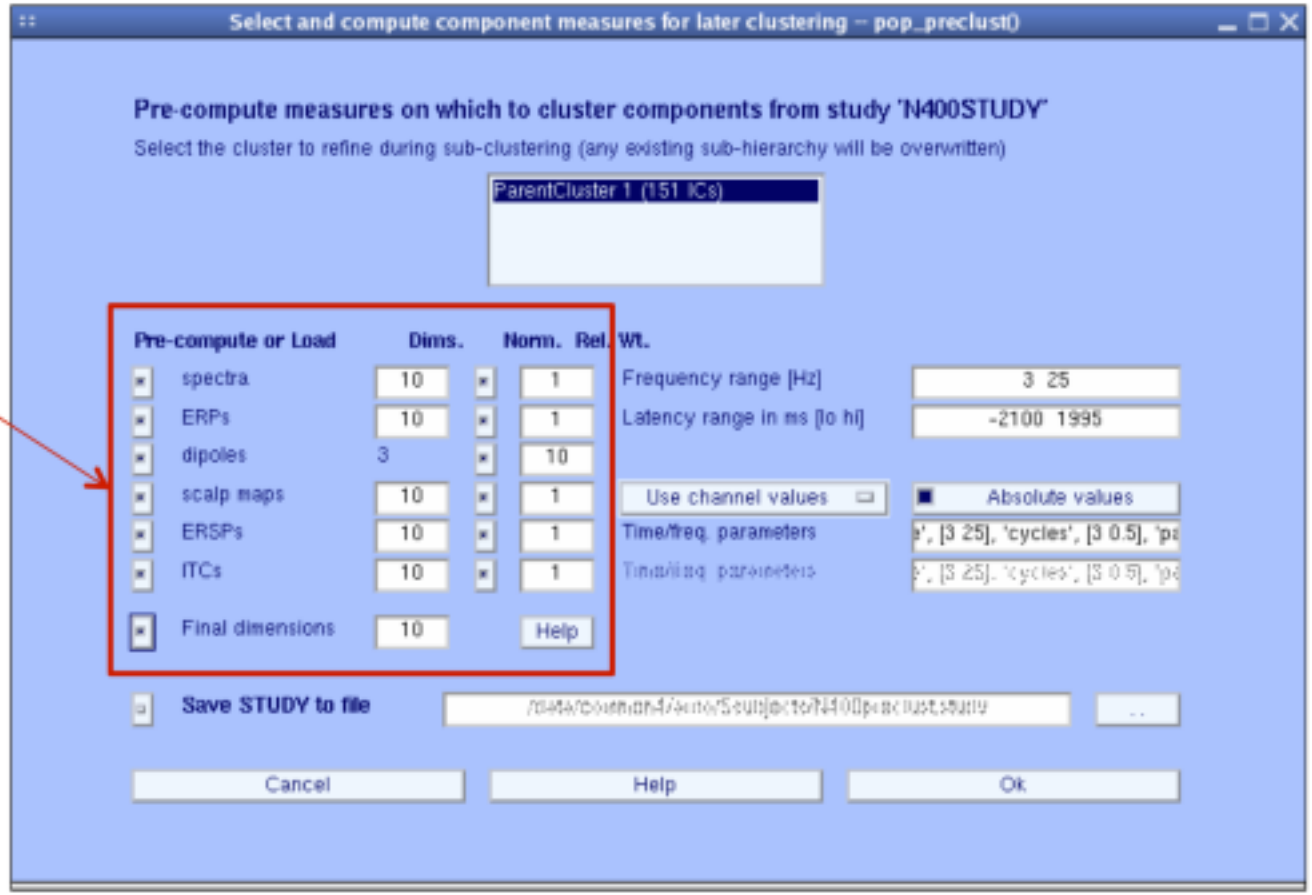
Plot cluster ERP



Issue with standard clustering

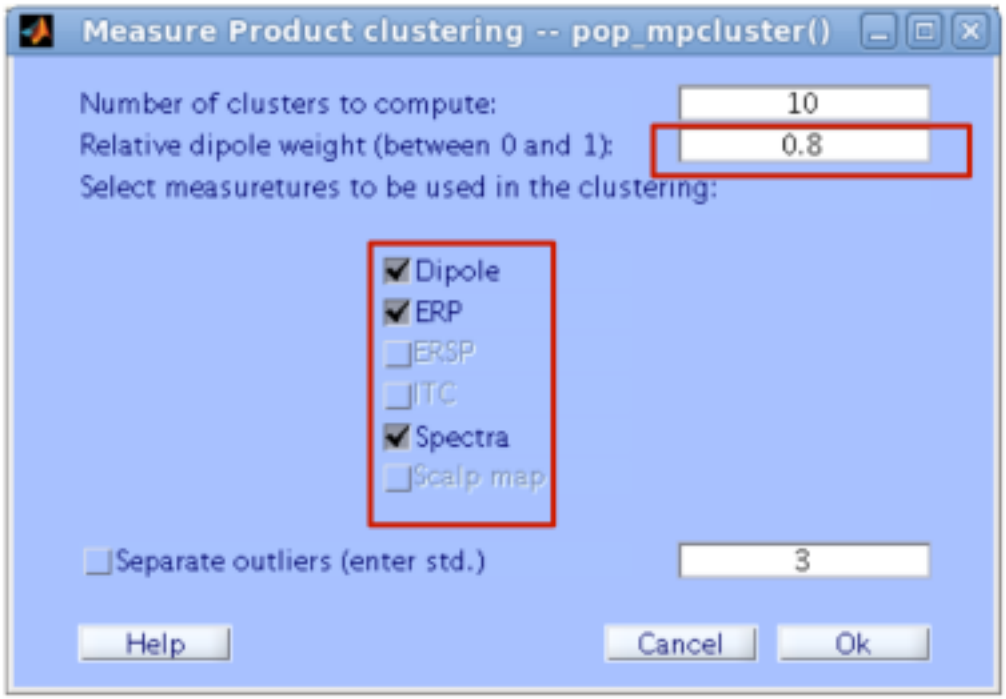
Large parameter space problem: many different clustering solutions can be produced by changing parameters and measure subsets. Which one should we choose?

EEGLAB clustering has ~12 parameters



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