Clustering of ICA components

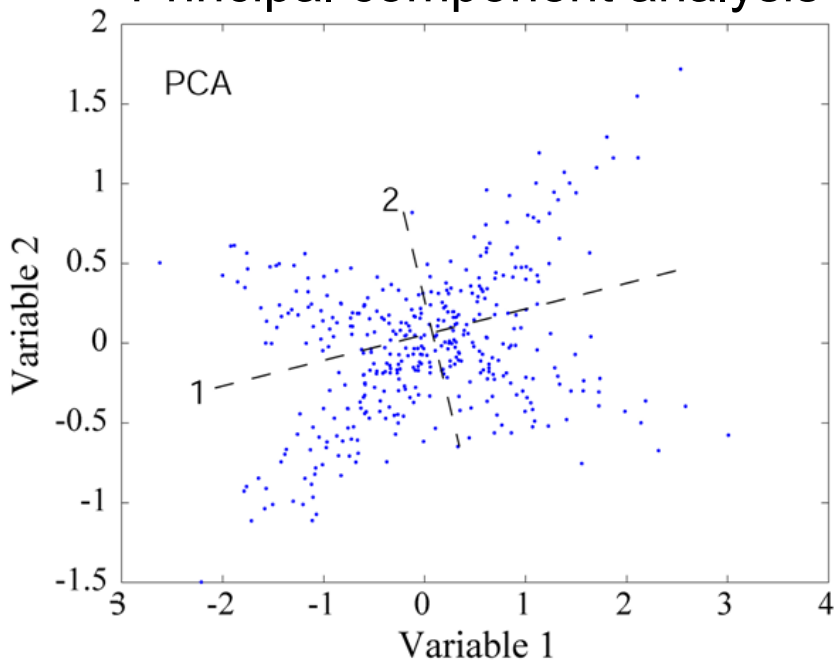
Arnaud Delorme

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

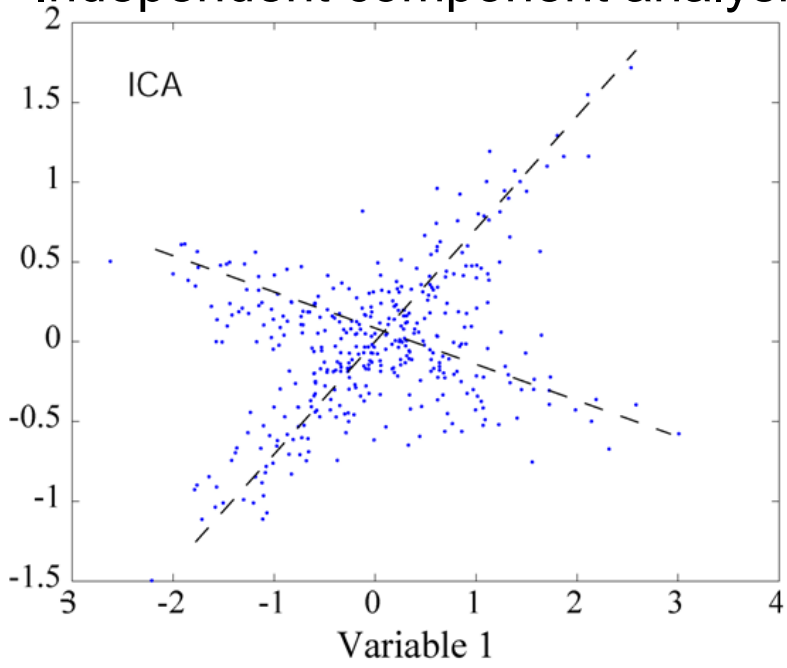
ICA and PCA

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

Principal component analysis

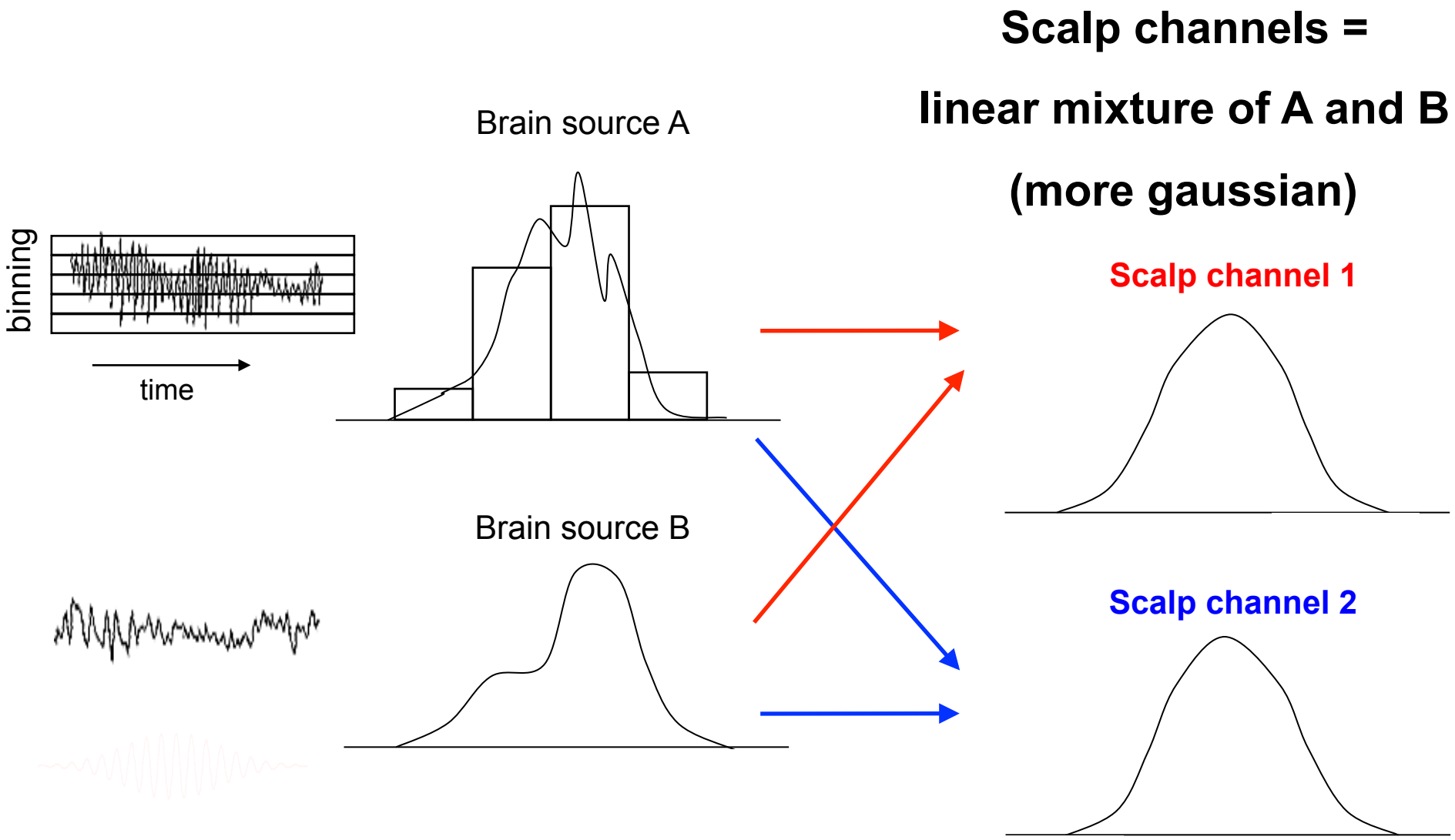


Independent component analysis

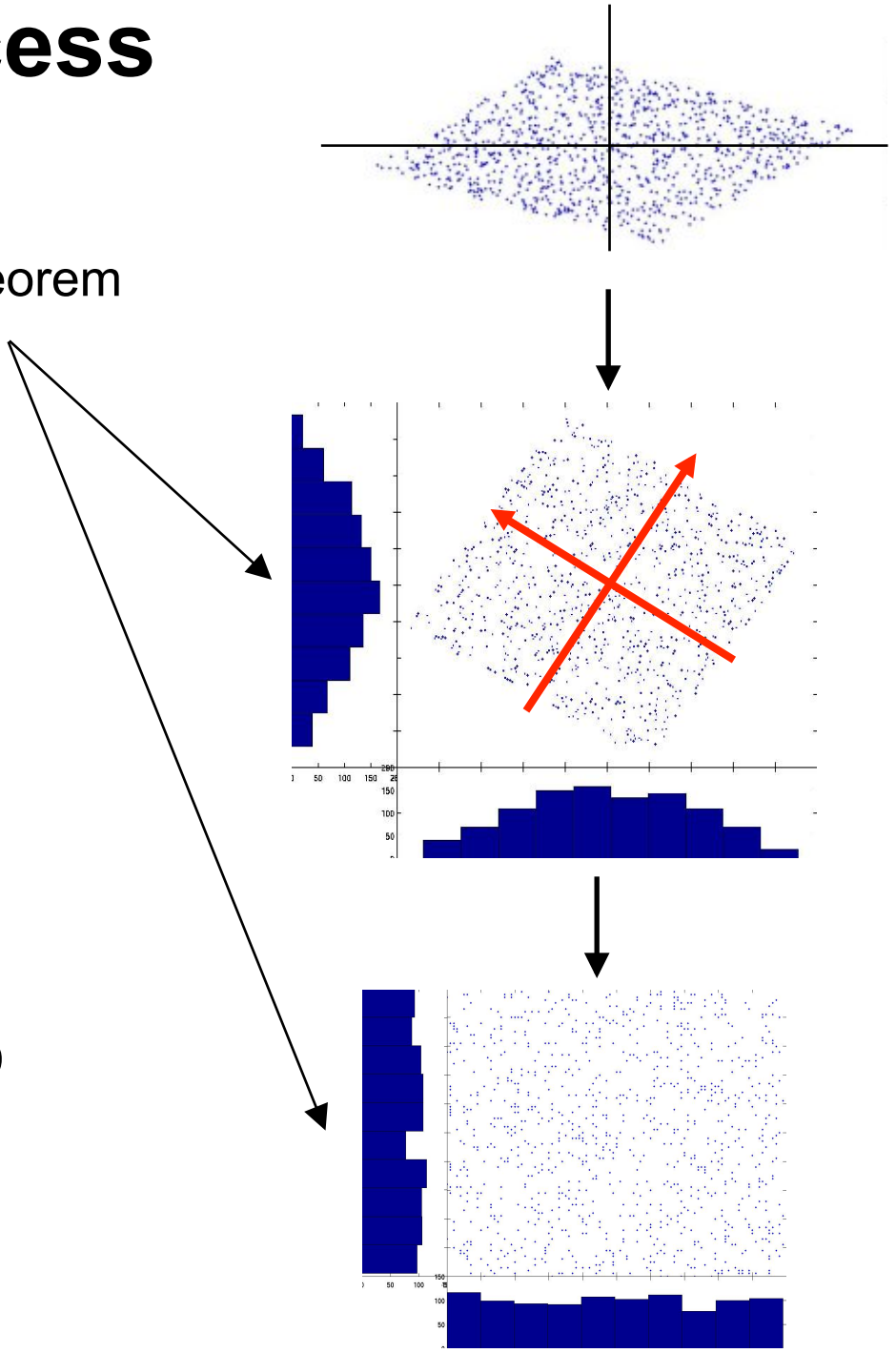


While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs statistically independent, while placing no constraints on the minxing matrix.

Central limit theorem



ICA Training Process



- Remove the mean
  $x = x - \langle x \rangle$
- 'Sphere' the data by diagonalizing its covariance matrix,

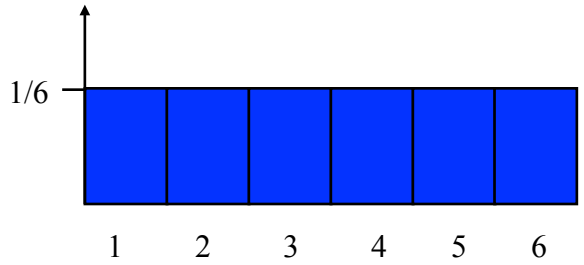
  $x = \langle xx^T \rangle^{-1/2}(x-\langle x \rangle)$.
- Update W according to

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

Entropy

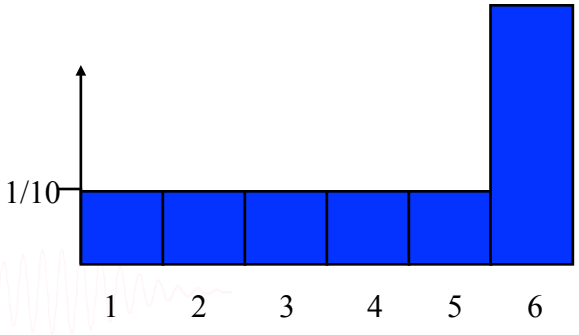
$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

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)



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

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 $H(y_1, y_2) \Longrightarrow$ **minimizing** $I(y_1, y_2)$.

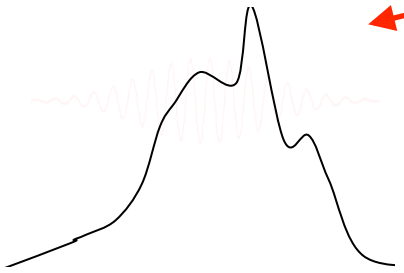
The learning rule:

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

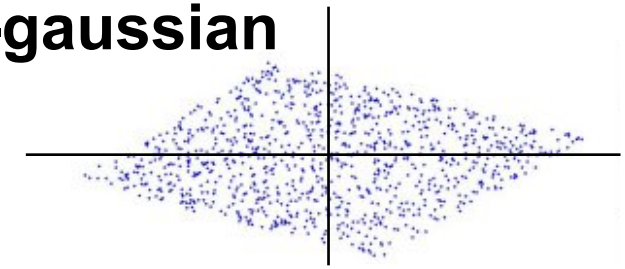
Natural gradient (Amari)

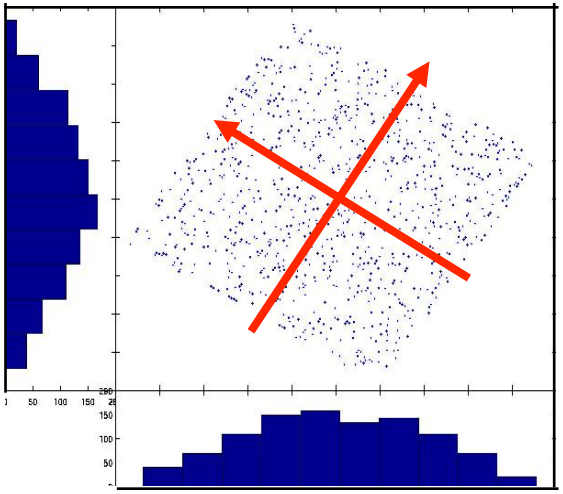
=0 if the two variables are independent

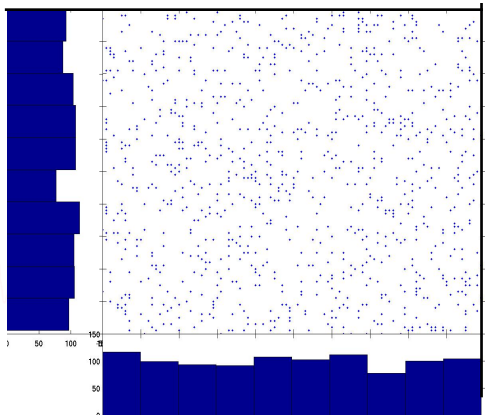
Entropy extremum



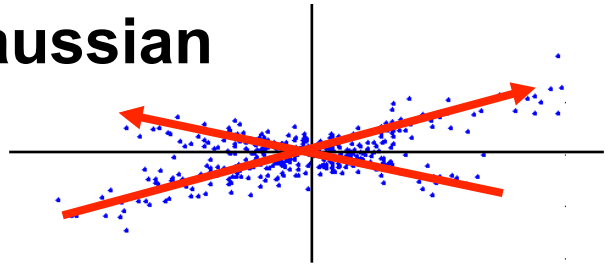
Sub-gaussian



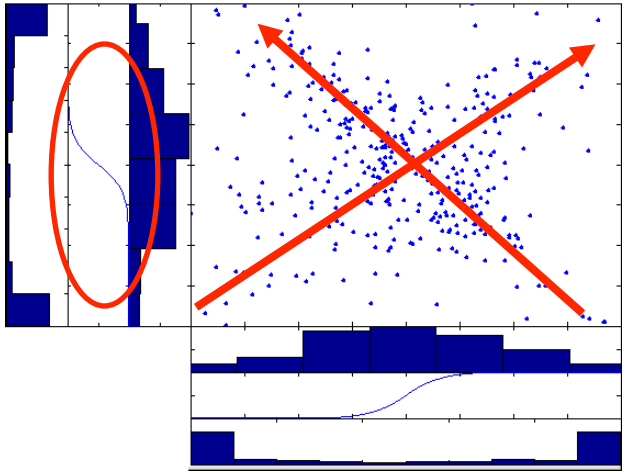




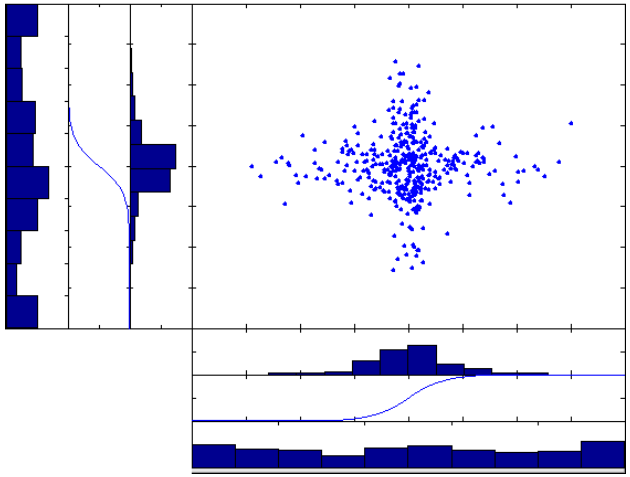
Super-gaussian



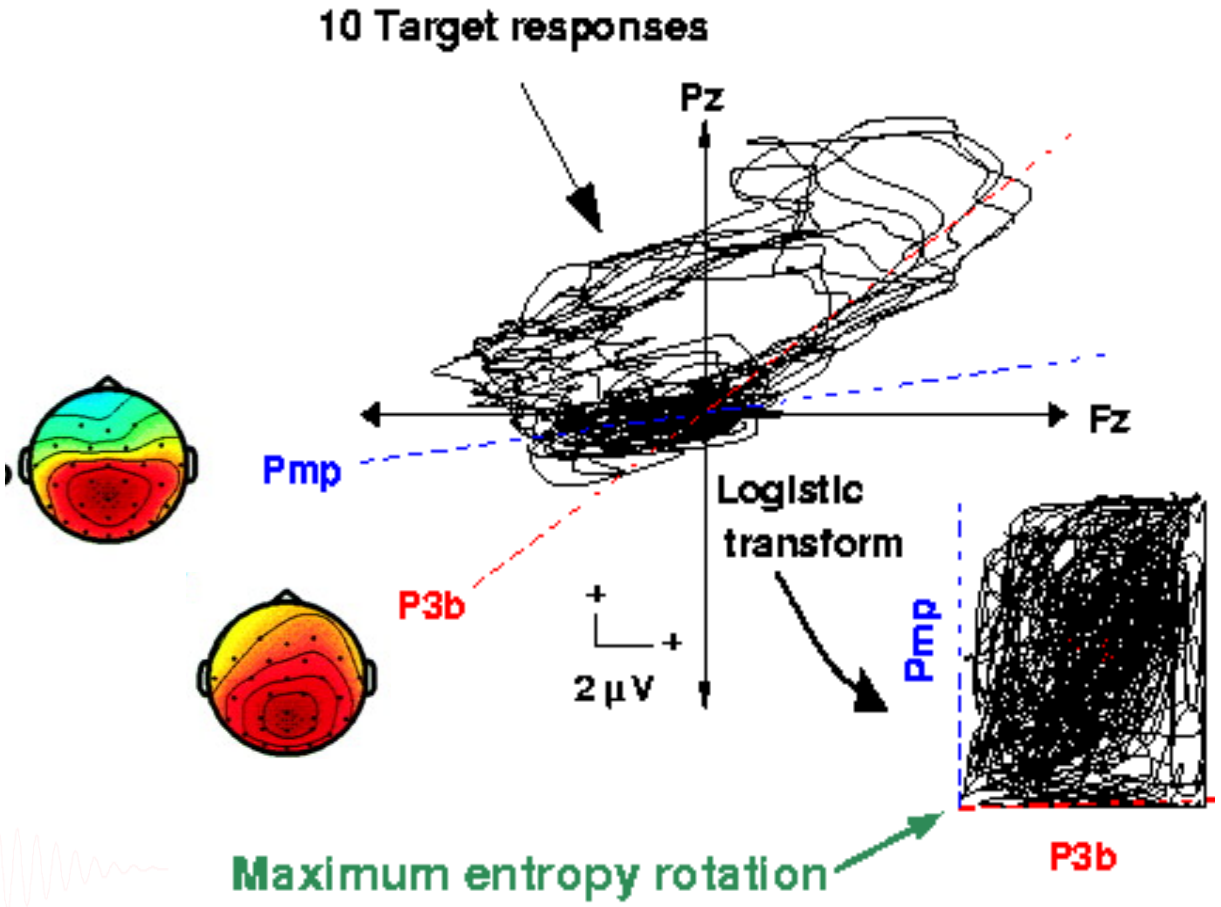
Sphering



ICA



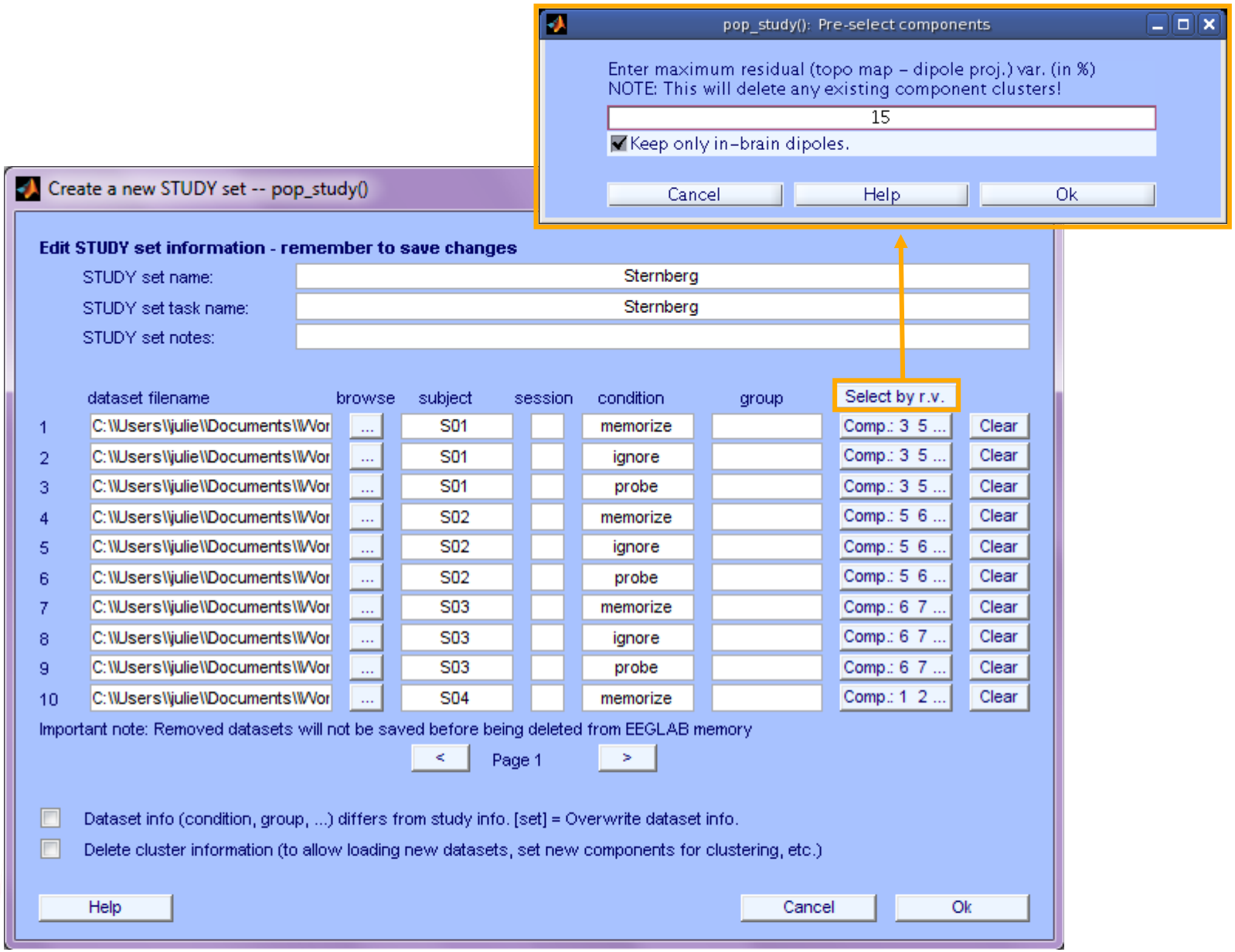
Independent components of EEG/ERP



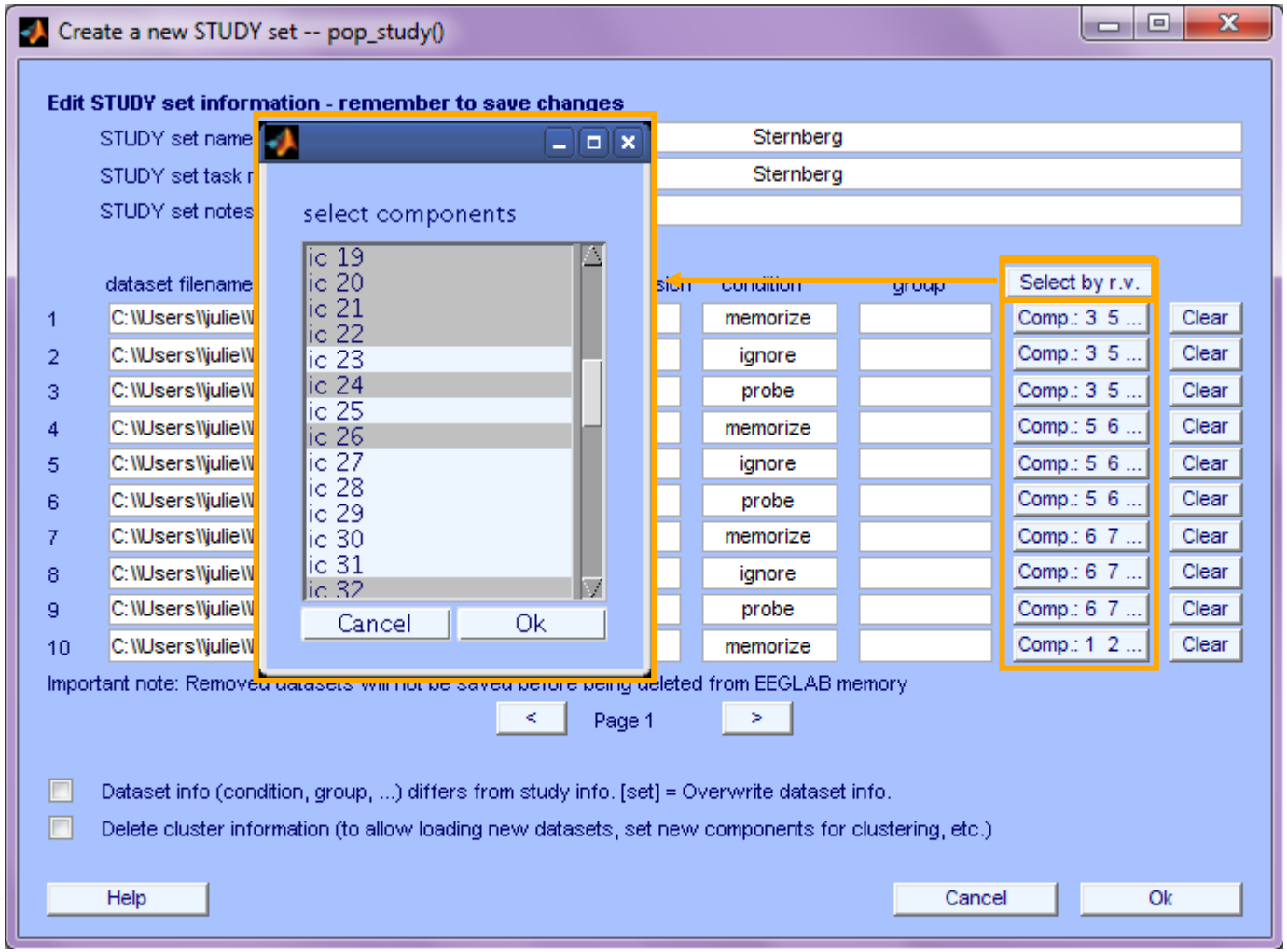
Steps of clustering

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

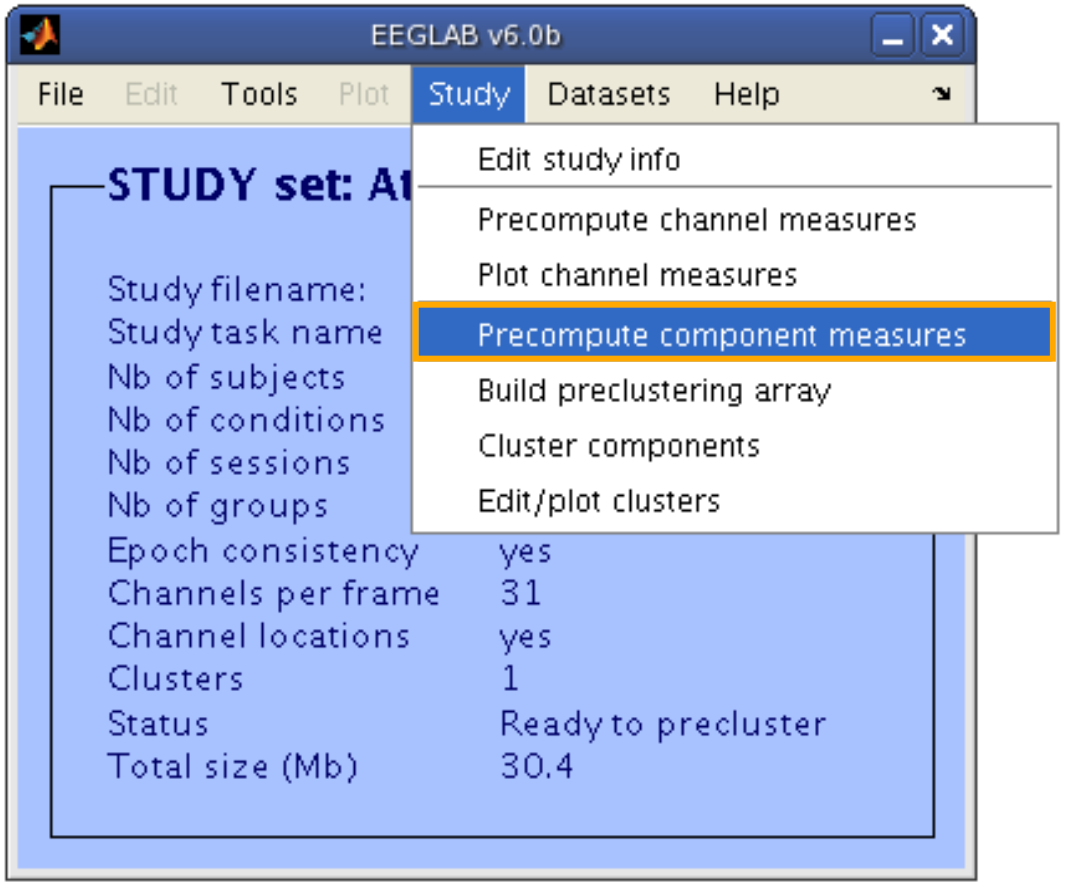
Edit dataset info



ICs to cluster

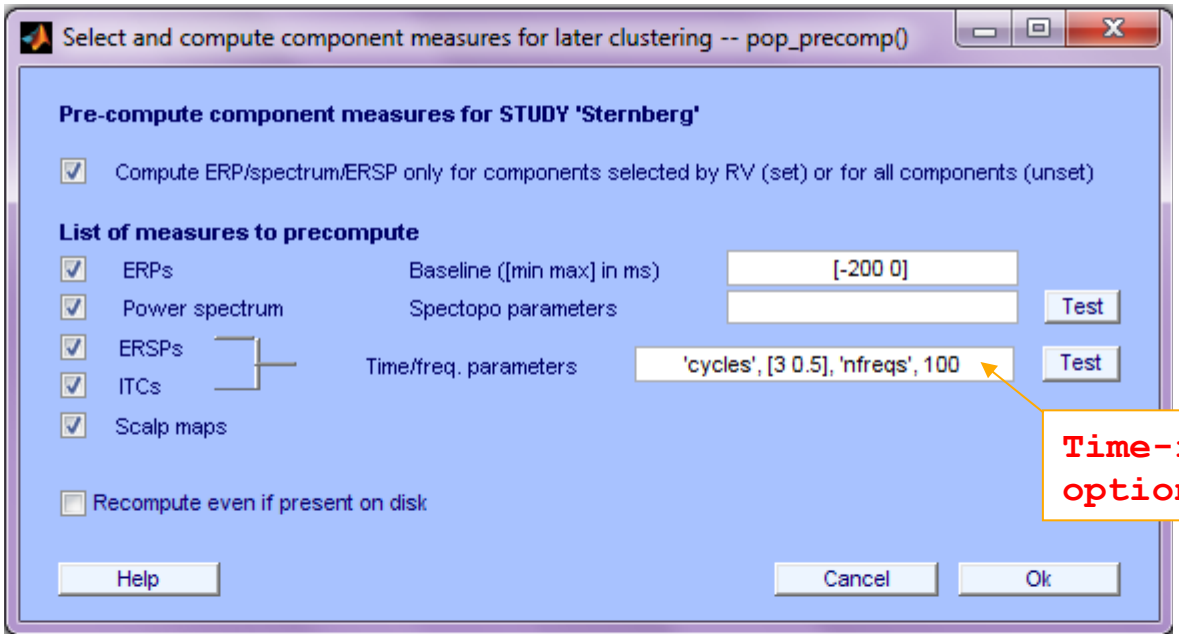


Precompute data measures



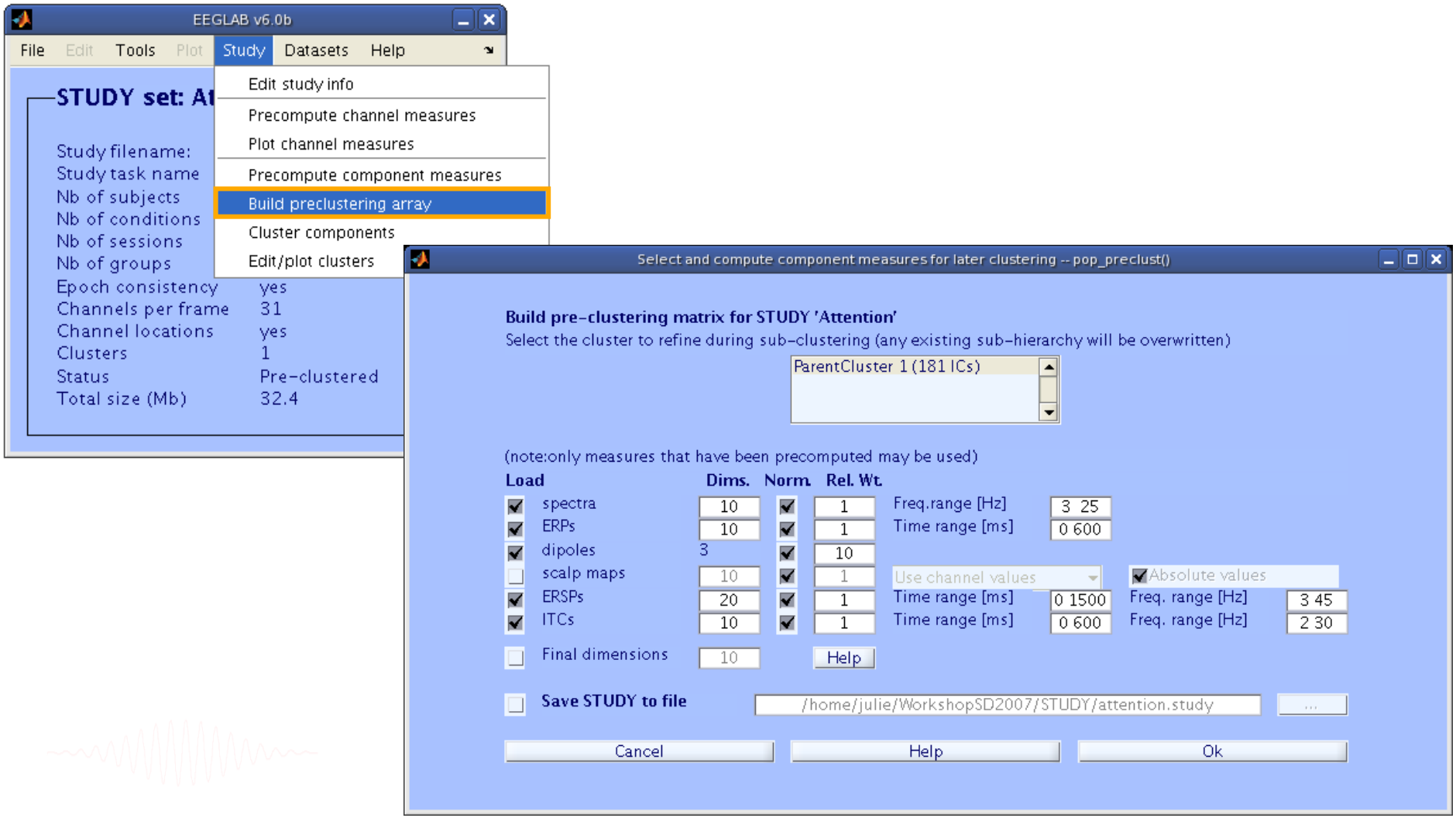
Precompute data measures

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

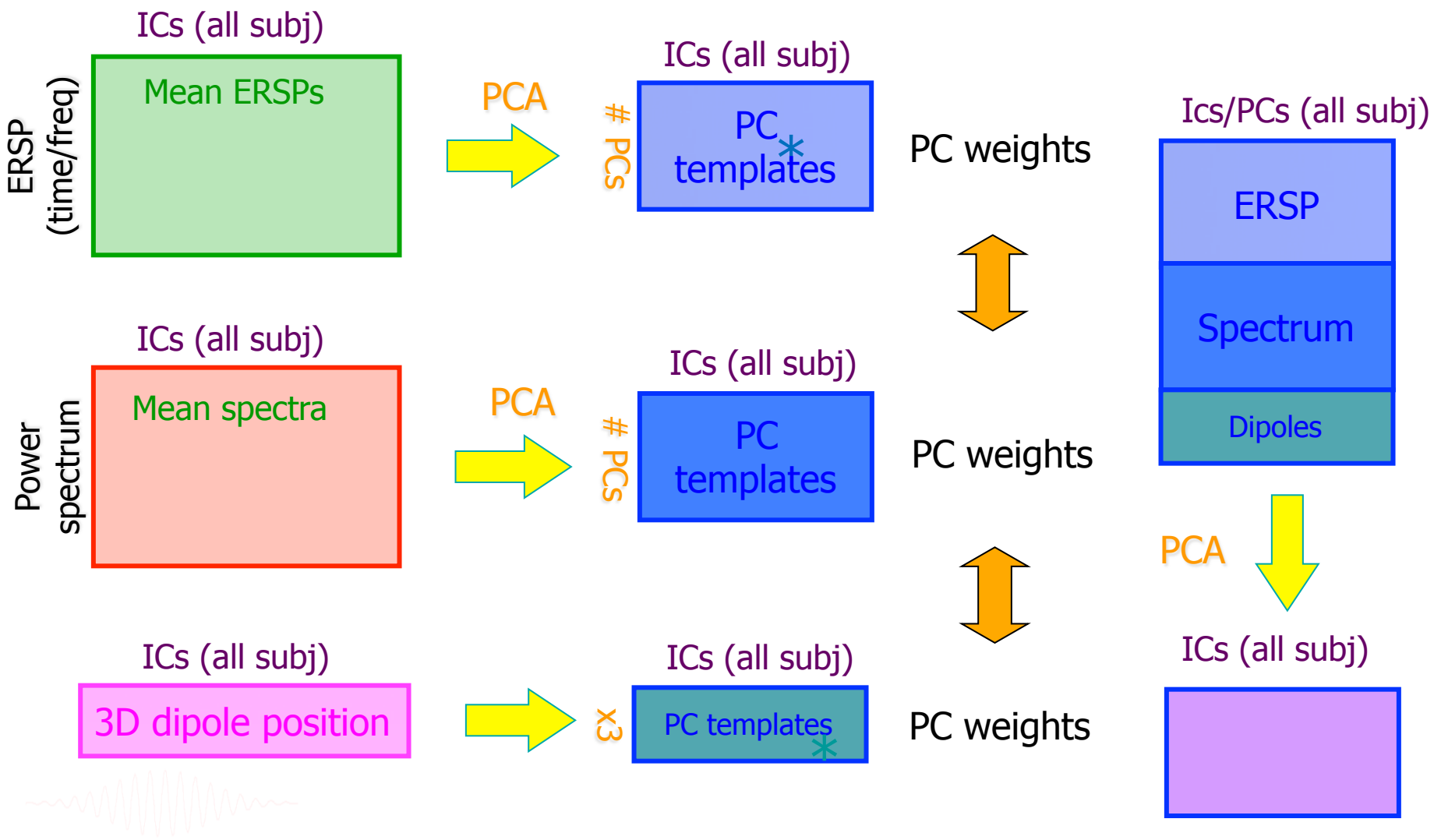


Time-frequency options

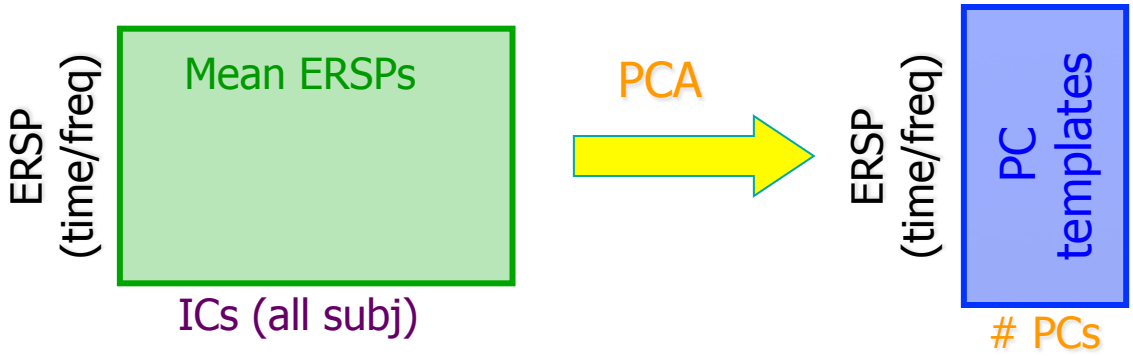
Cluster components

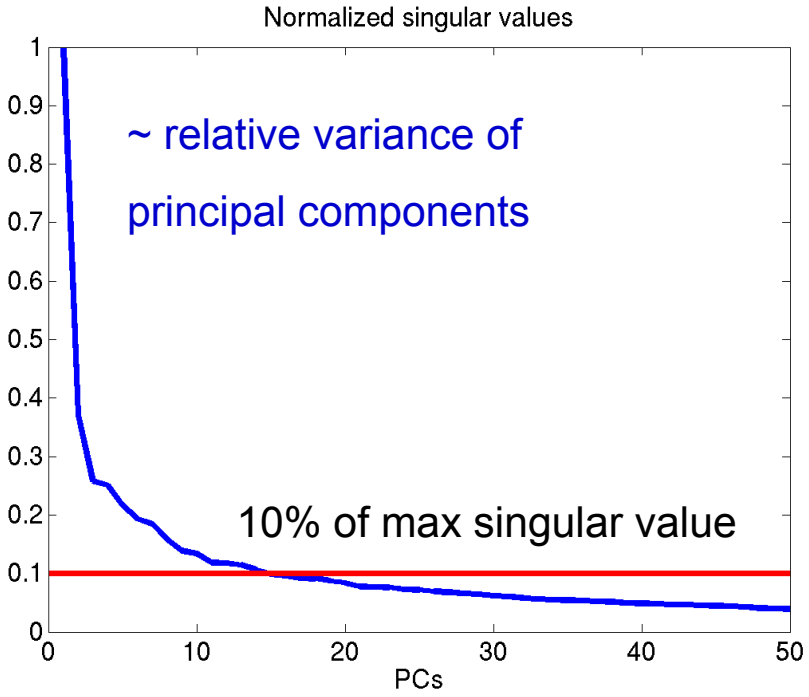


Precluster schematic



Precluster: Use singular values from PCA





Precluster schematic

ICs (all subj)

ERSP

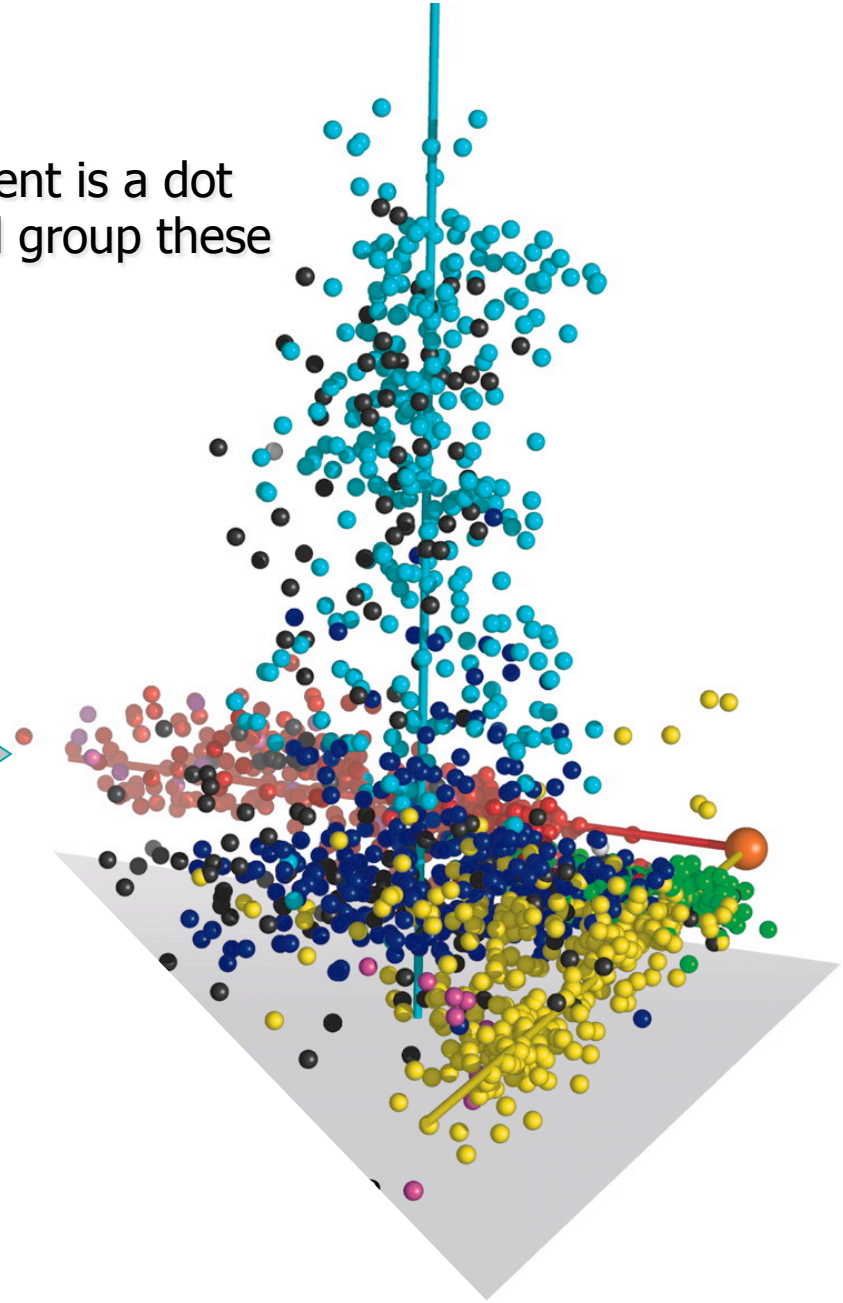
Spectrum

Dipoles

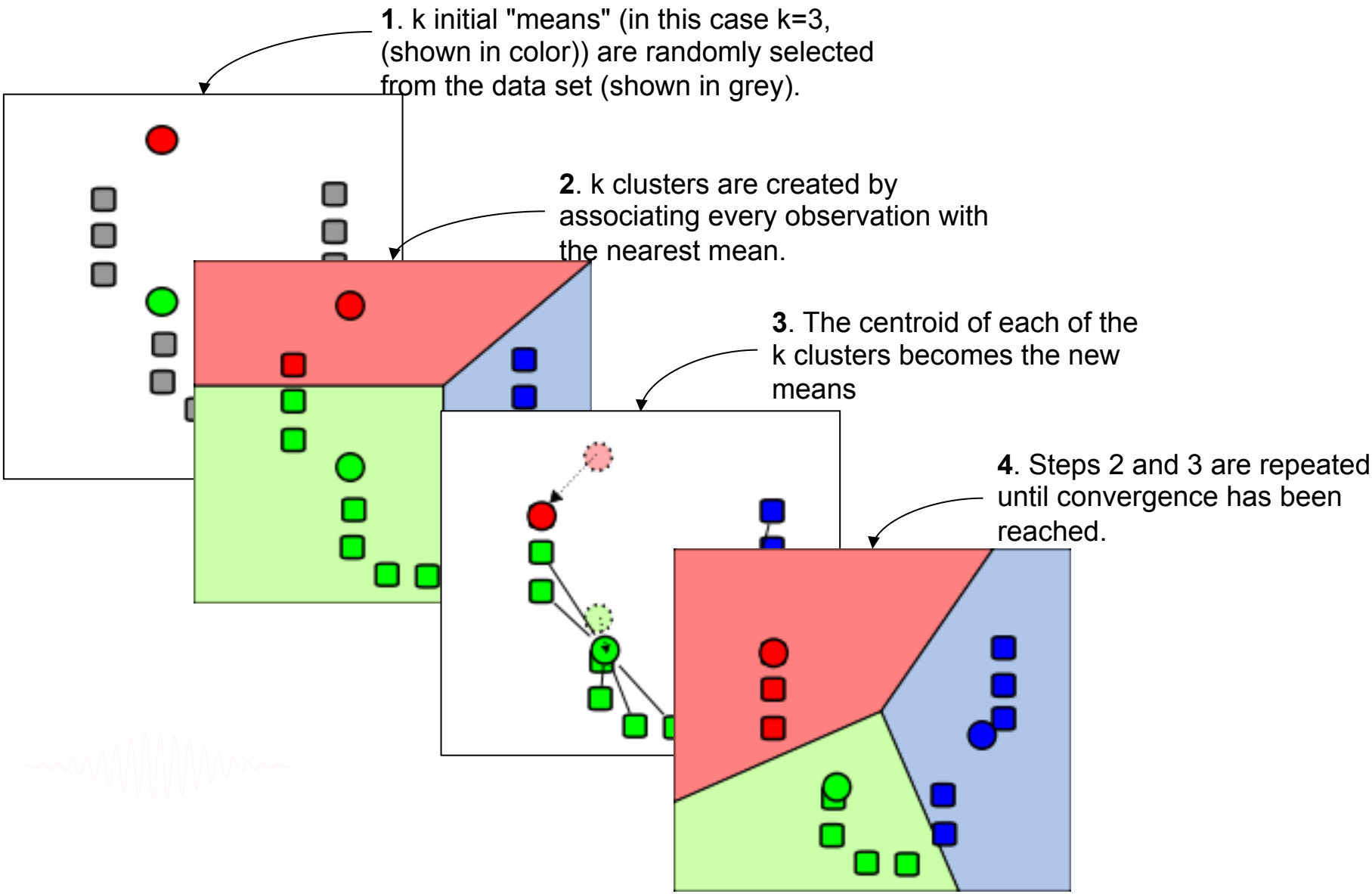
OR

ICs (all subj)

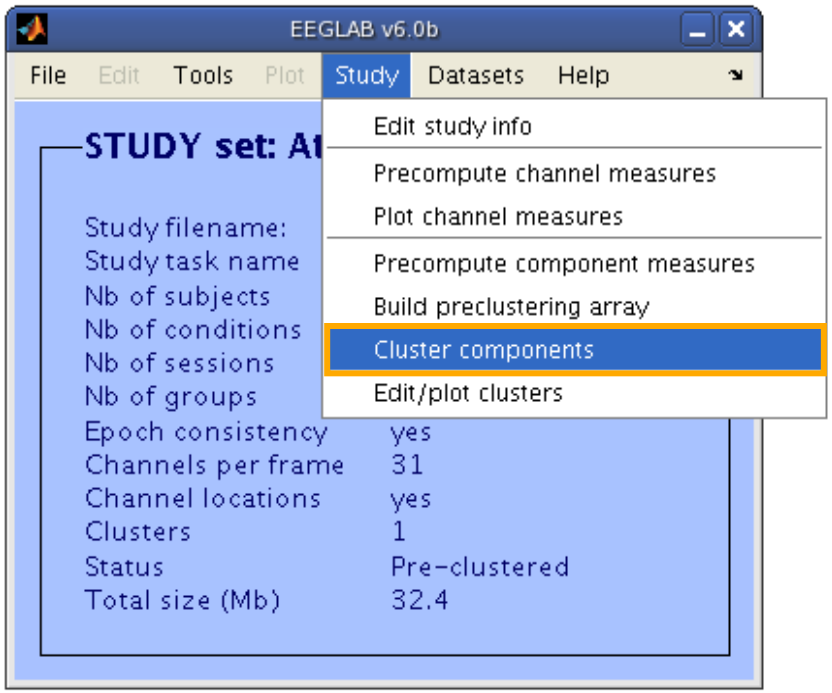
Each component is a dot
Clustering will group these dots

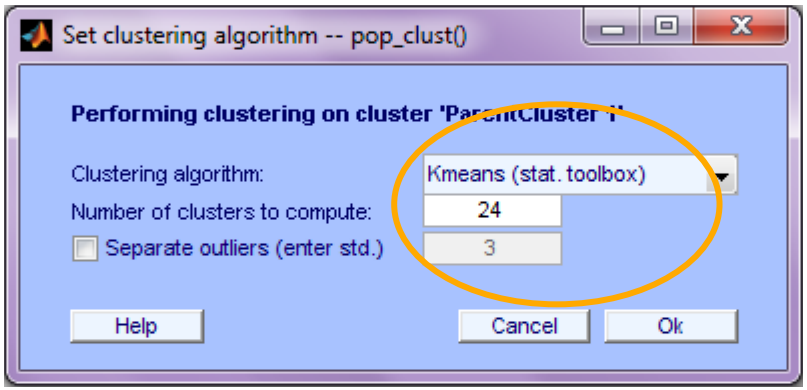


Classical KMean



Cluster components





Choosing data measures

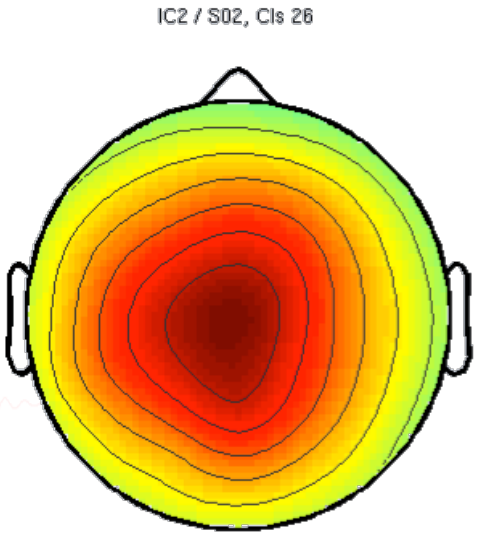
What measure(s) should you use?

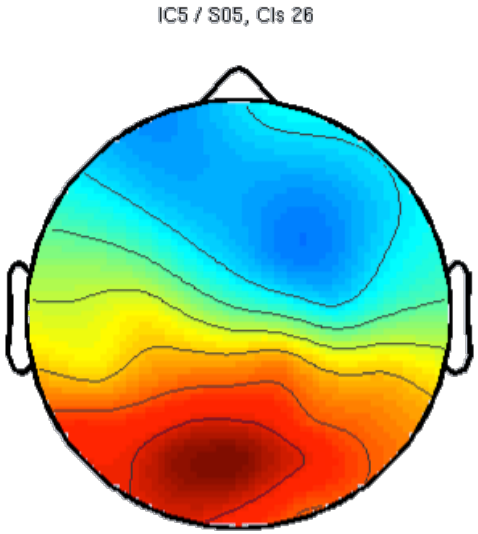
It depends on your final cluster criteria…

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

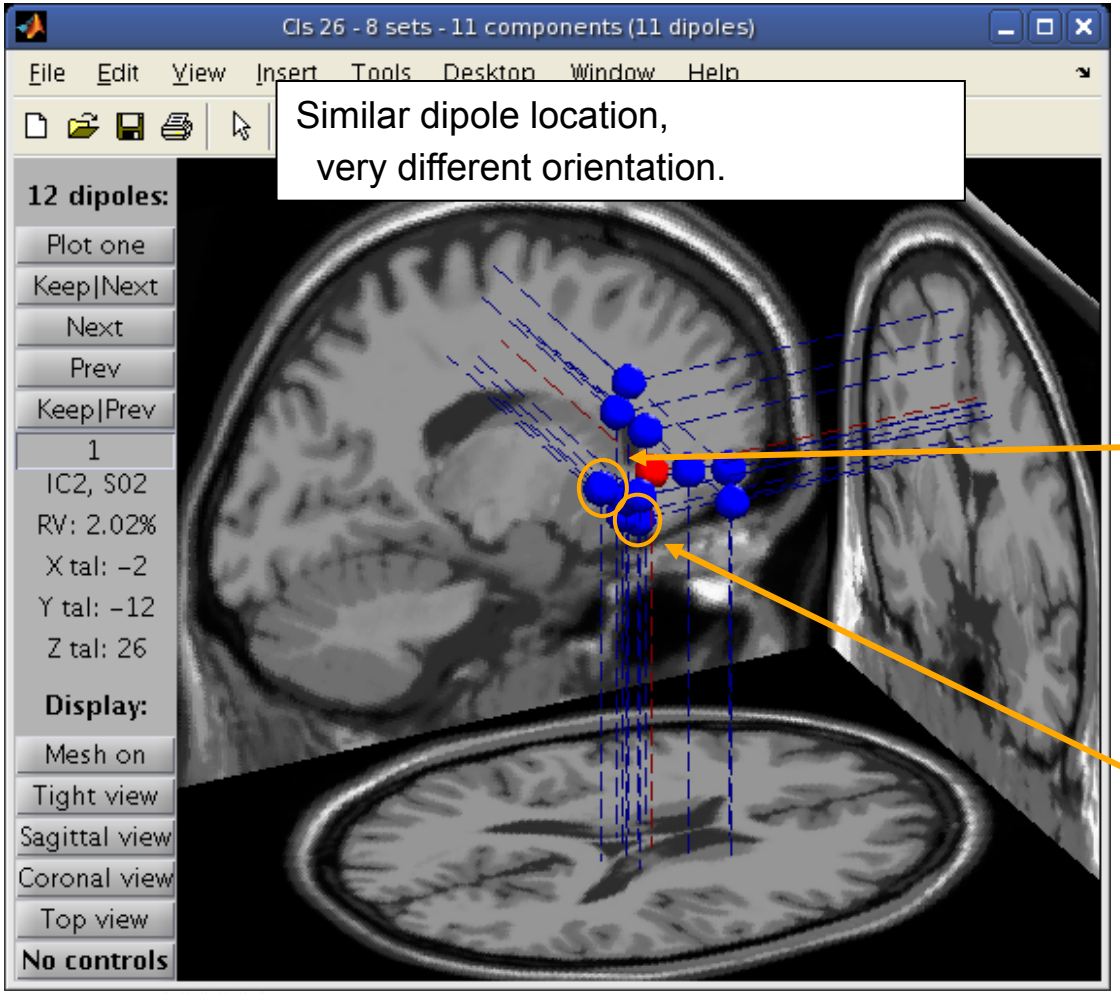
But consider:

- What is the difference between these two components?



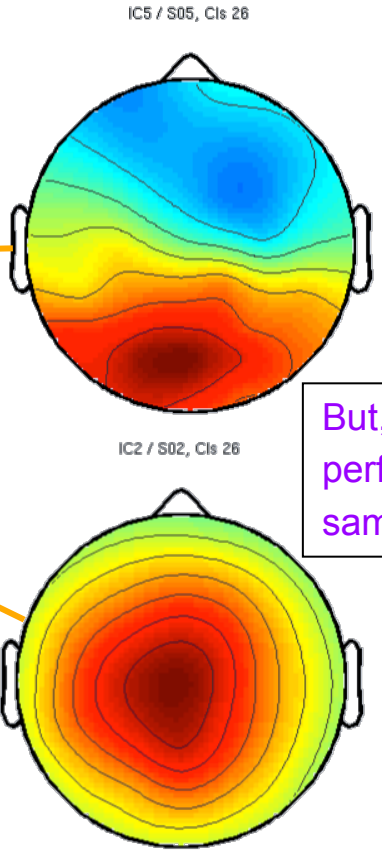


Choosing data measures



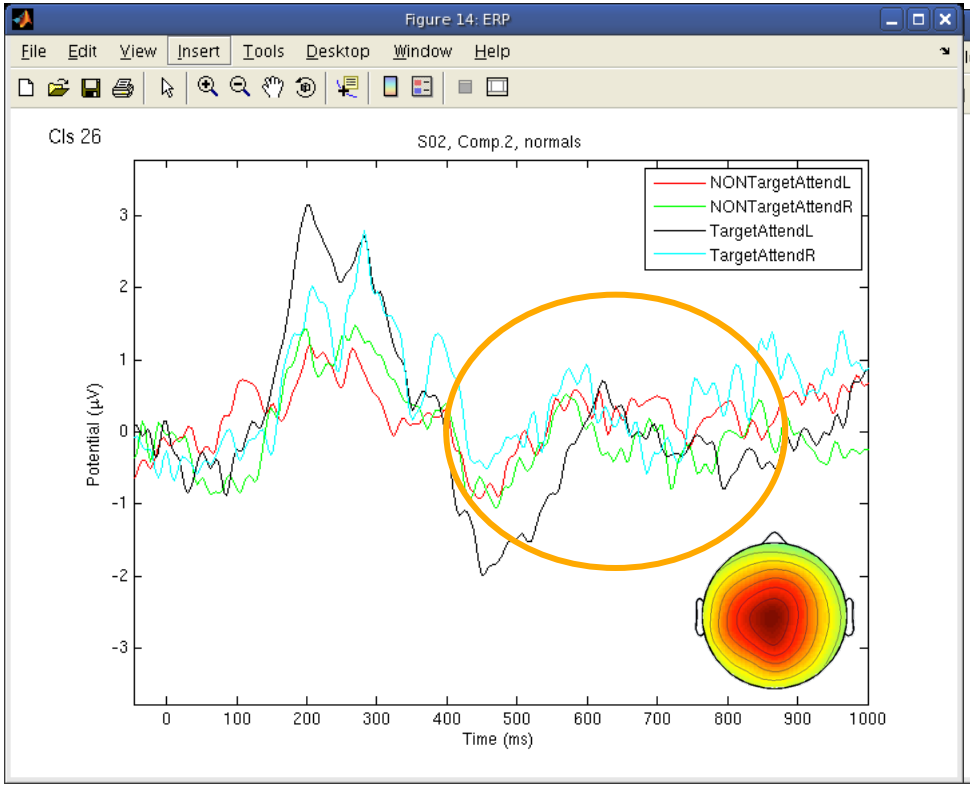
Similar dipole location,
very different orientation.

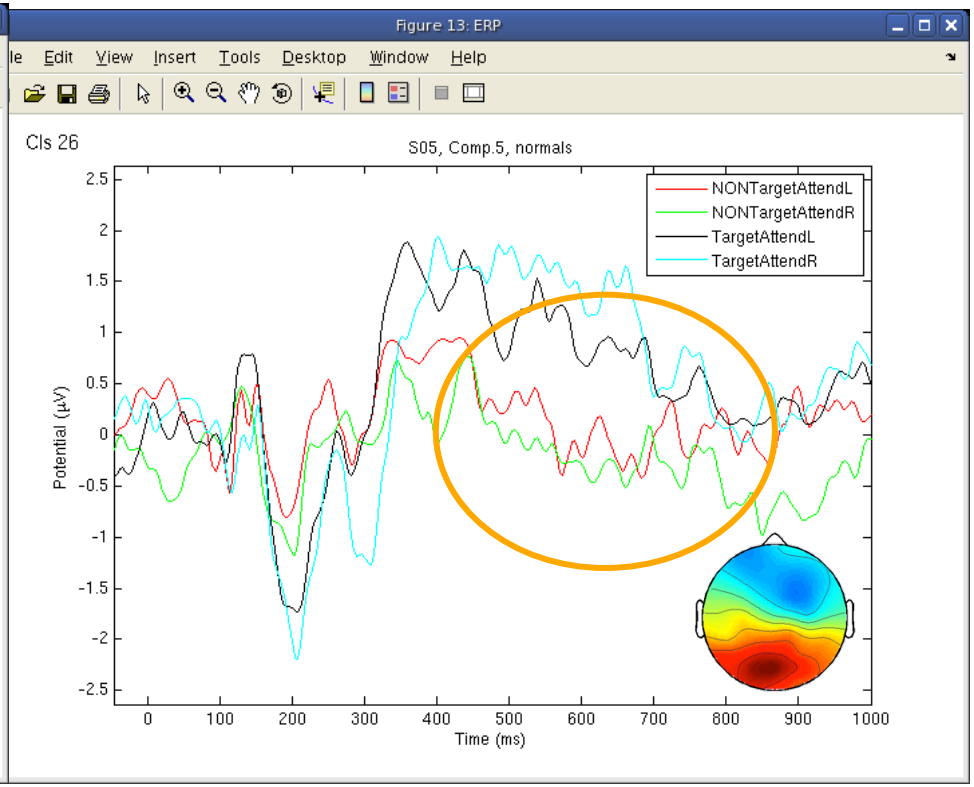
Obvious dramatic effect on scalp map topography:



But, do they perform the same functions?

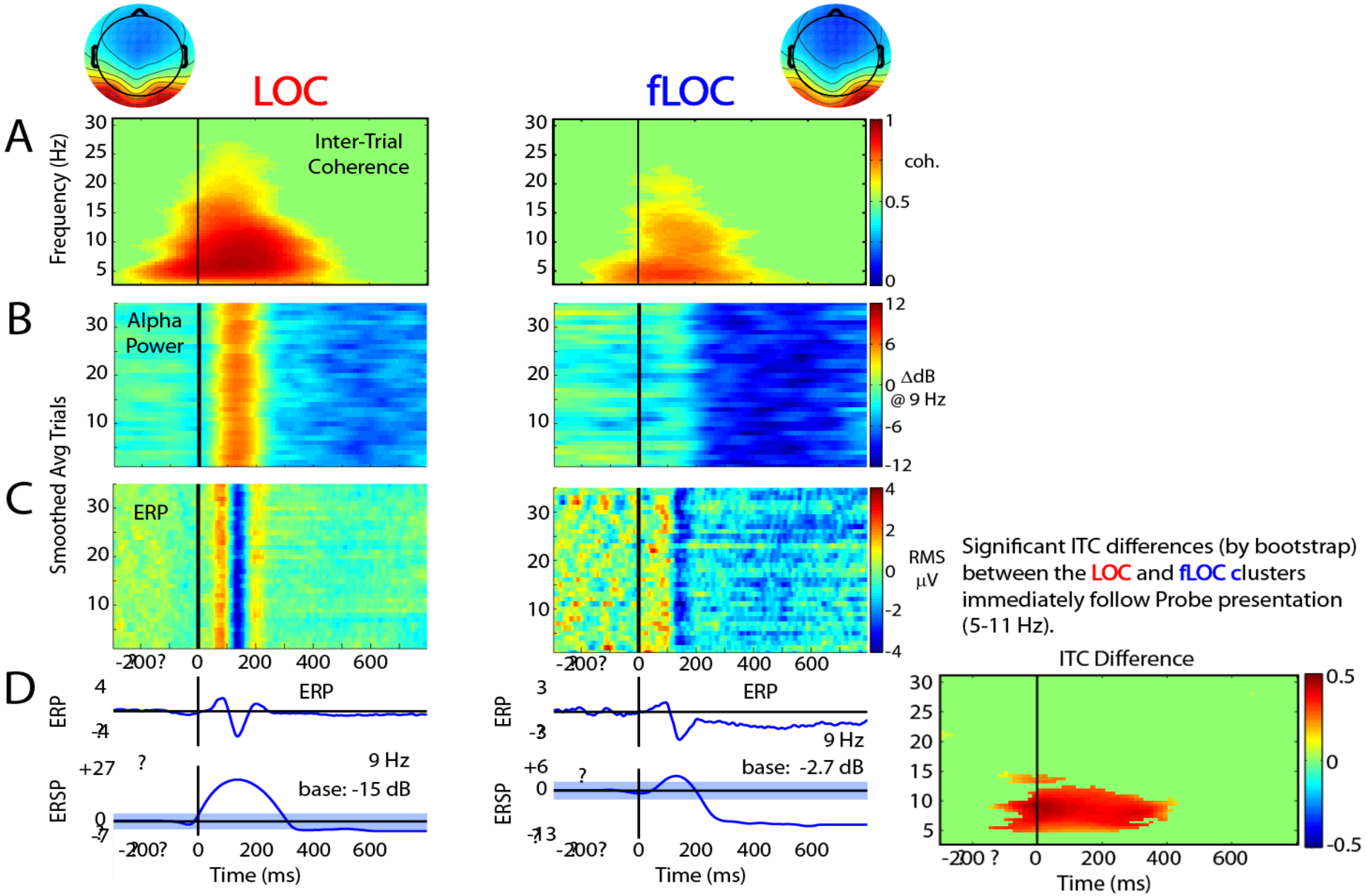
Choosing data measures



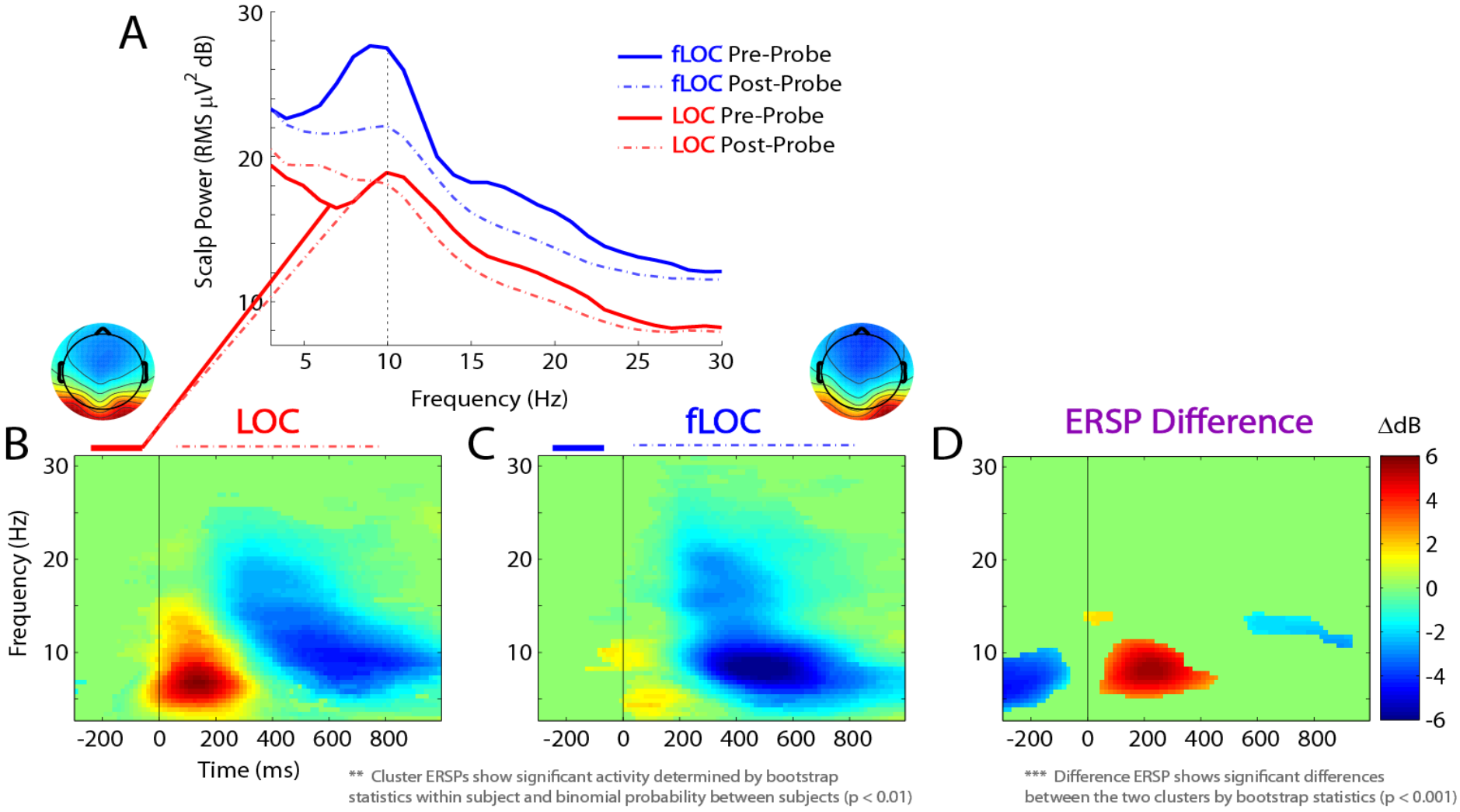


ERPs seem different…

Subject differences?



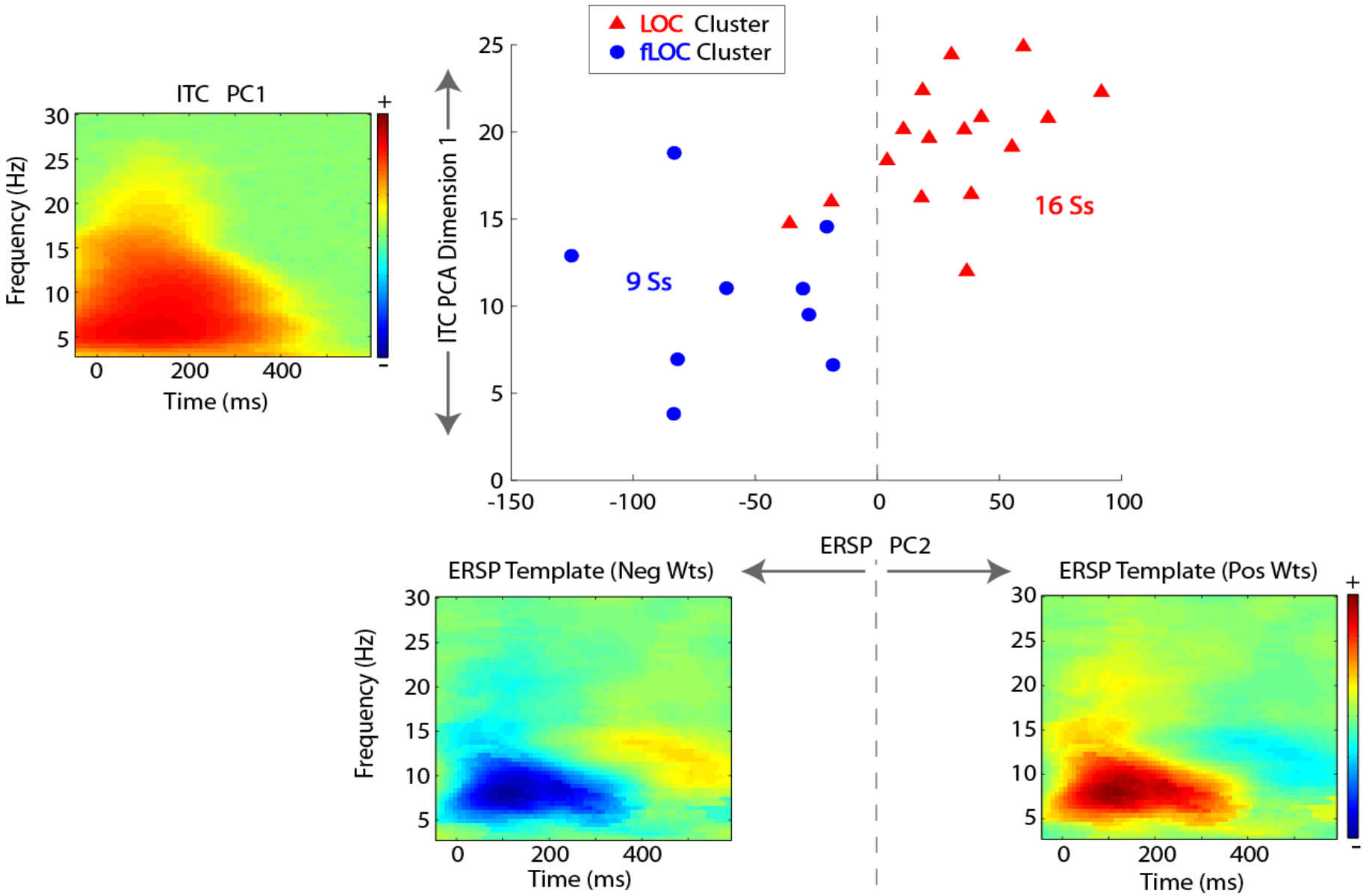
Subject differences?



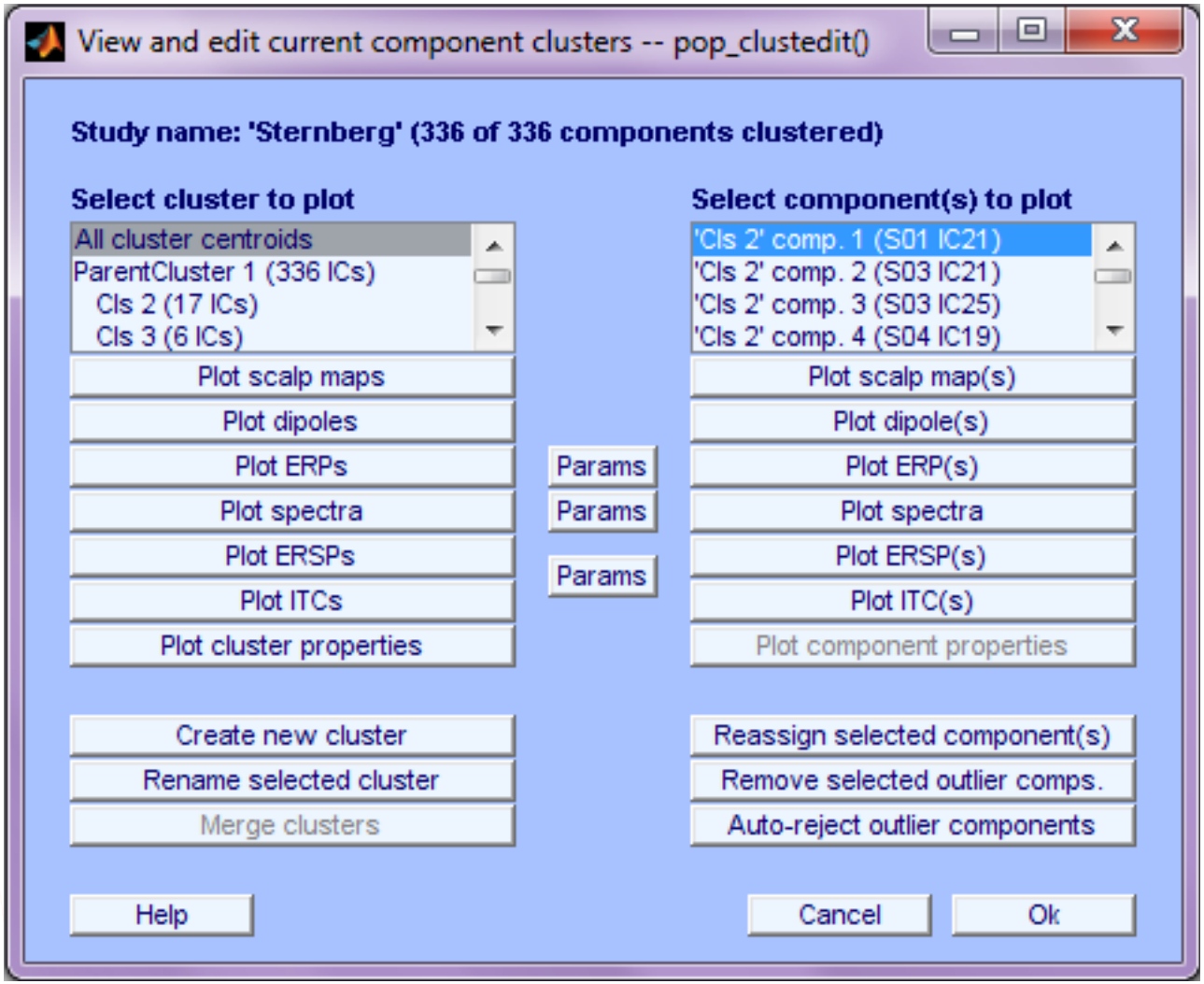
** Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects ($p < 0.01$)

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

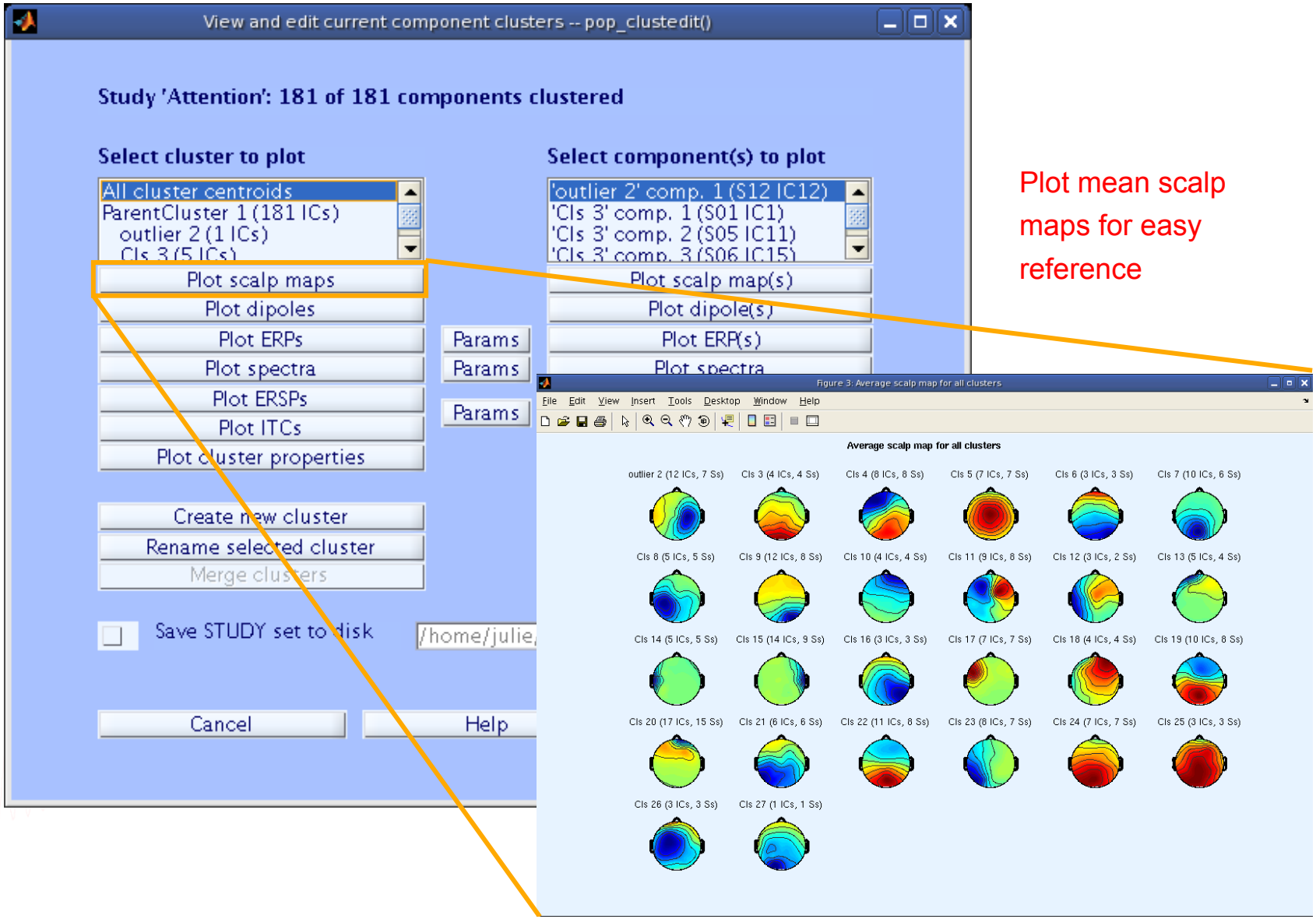
Subject differences?



Plot/edit clusters

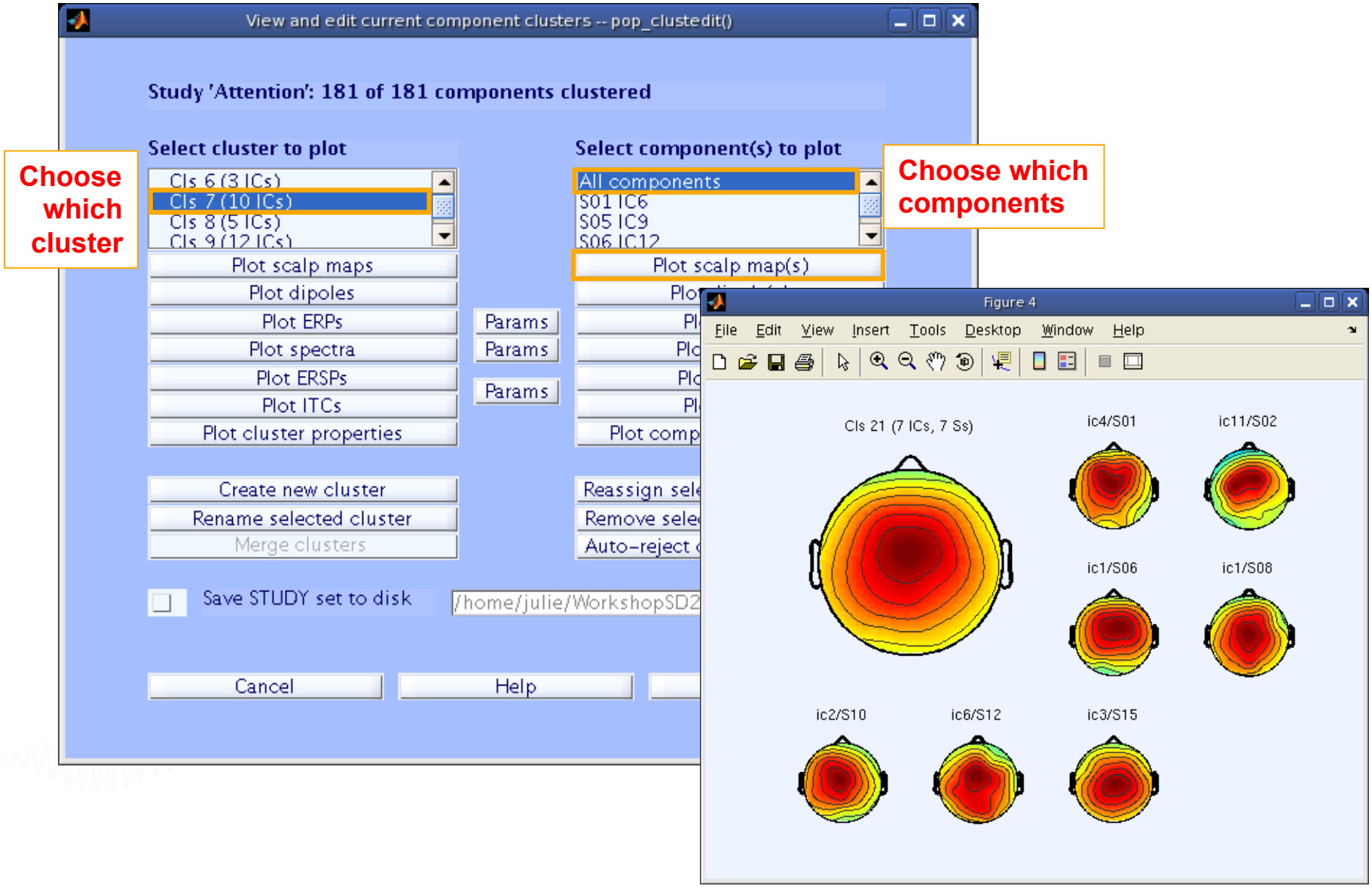


Plot cluster data

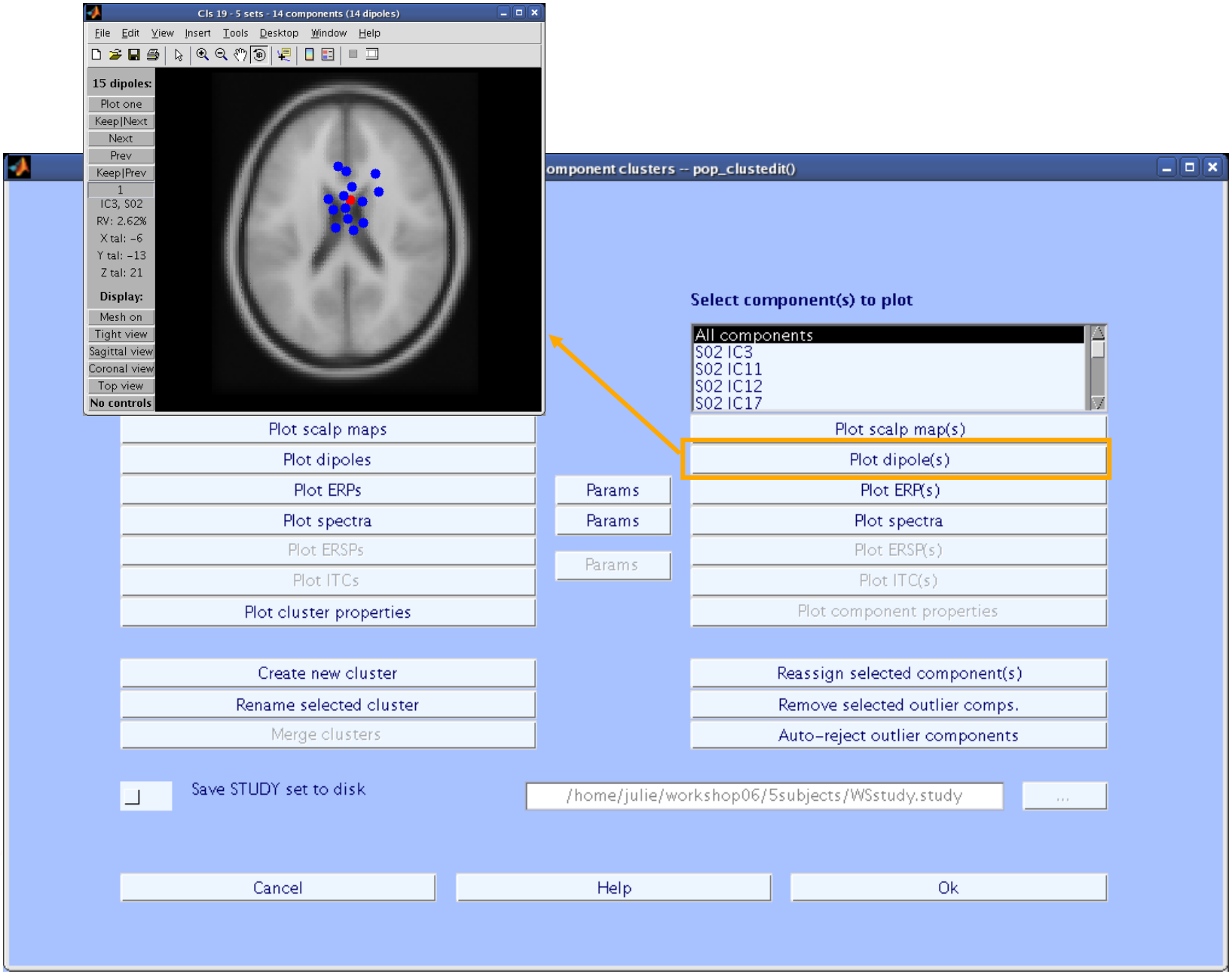


Plot mean scalp maps for easy reference

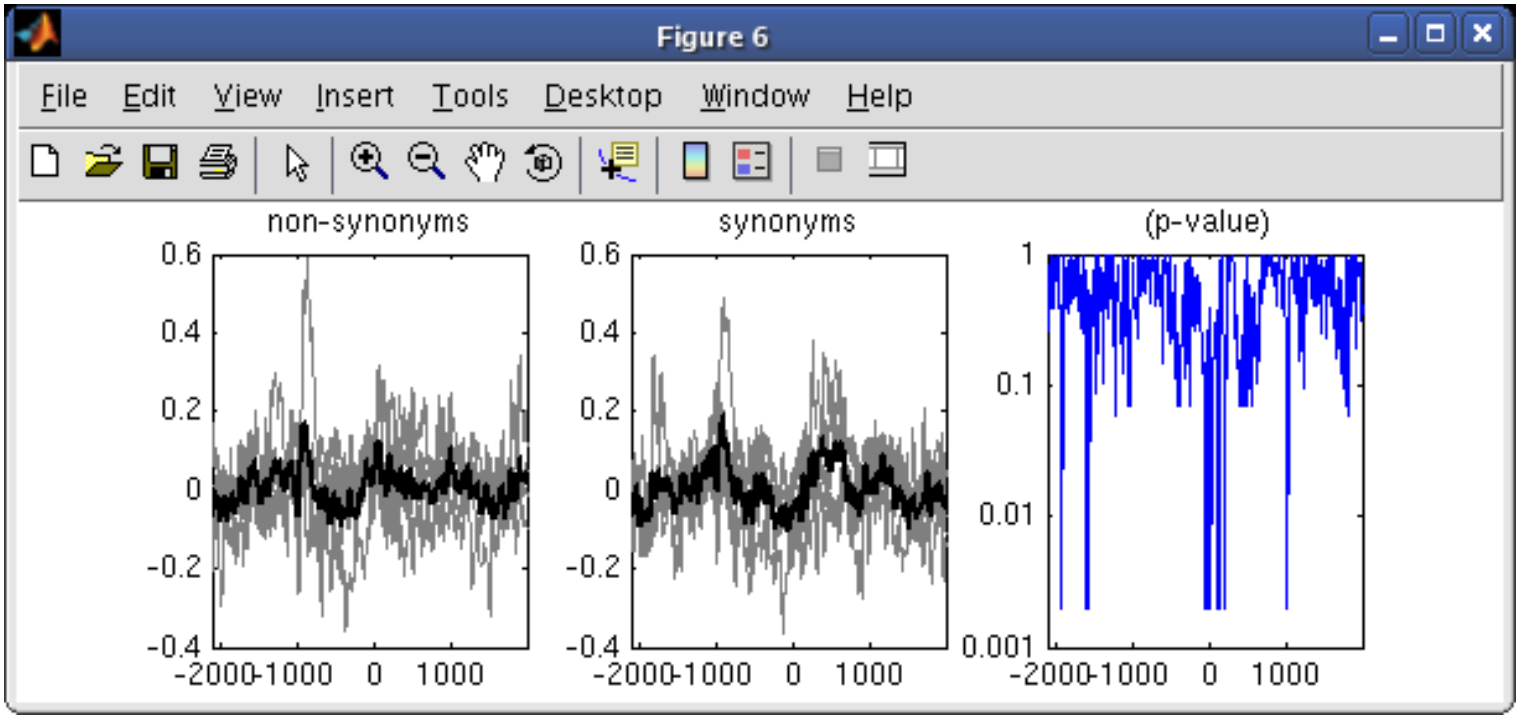
Plot cluster data



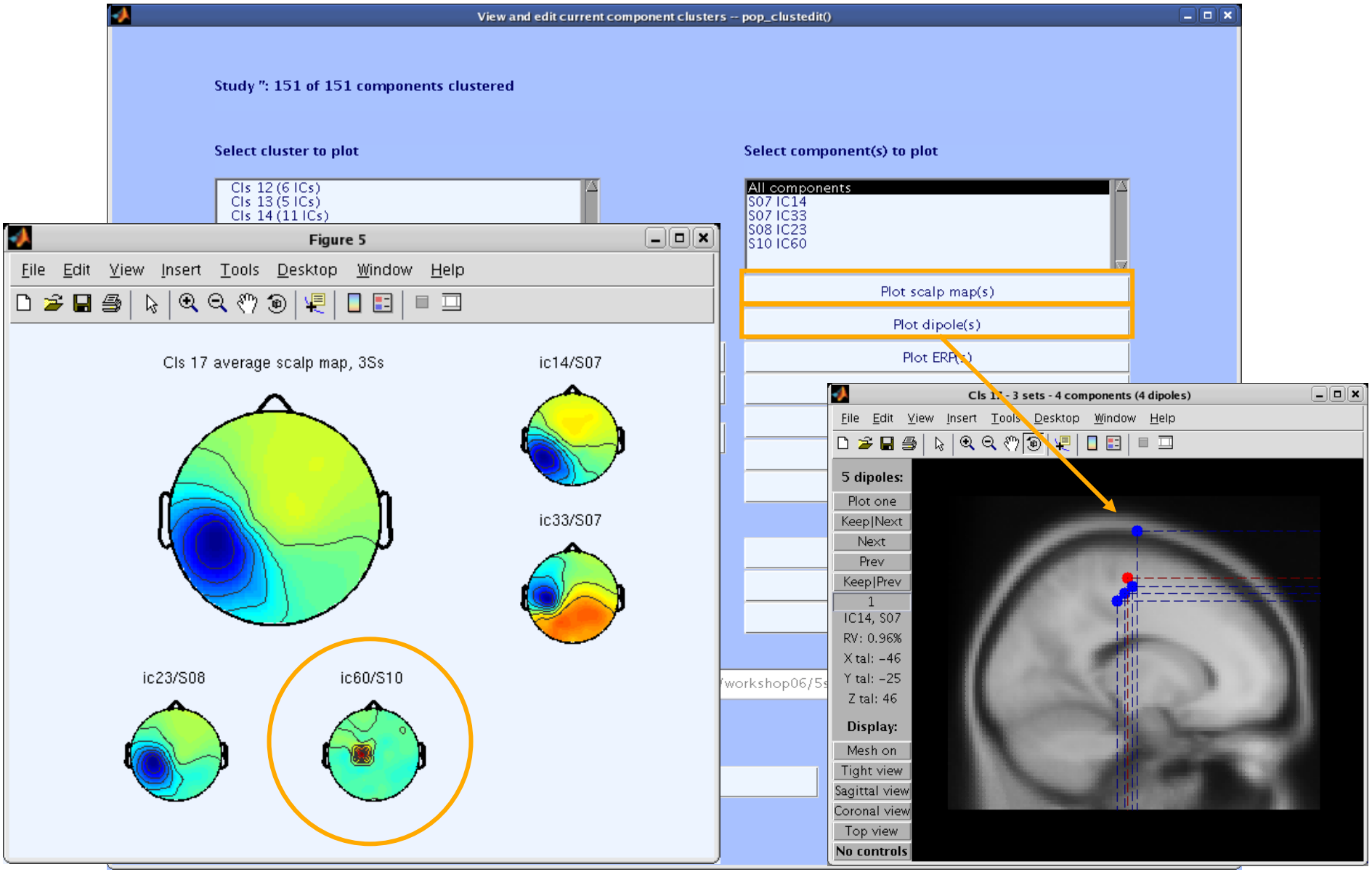
Plot cluster data



Plot cluster ERP



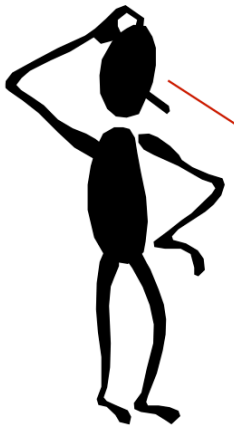
Reassigning components

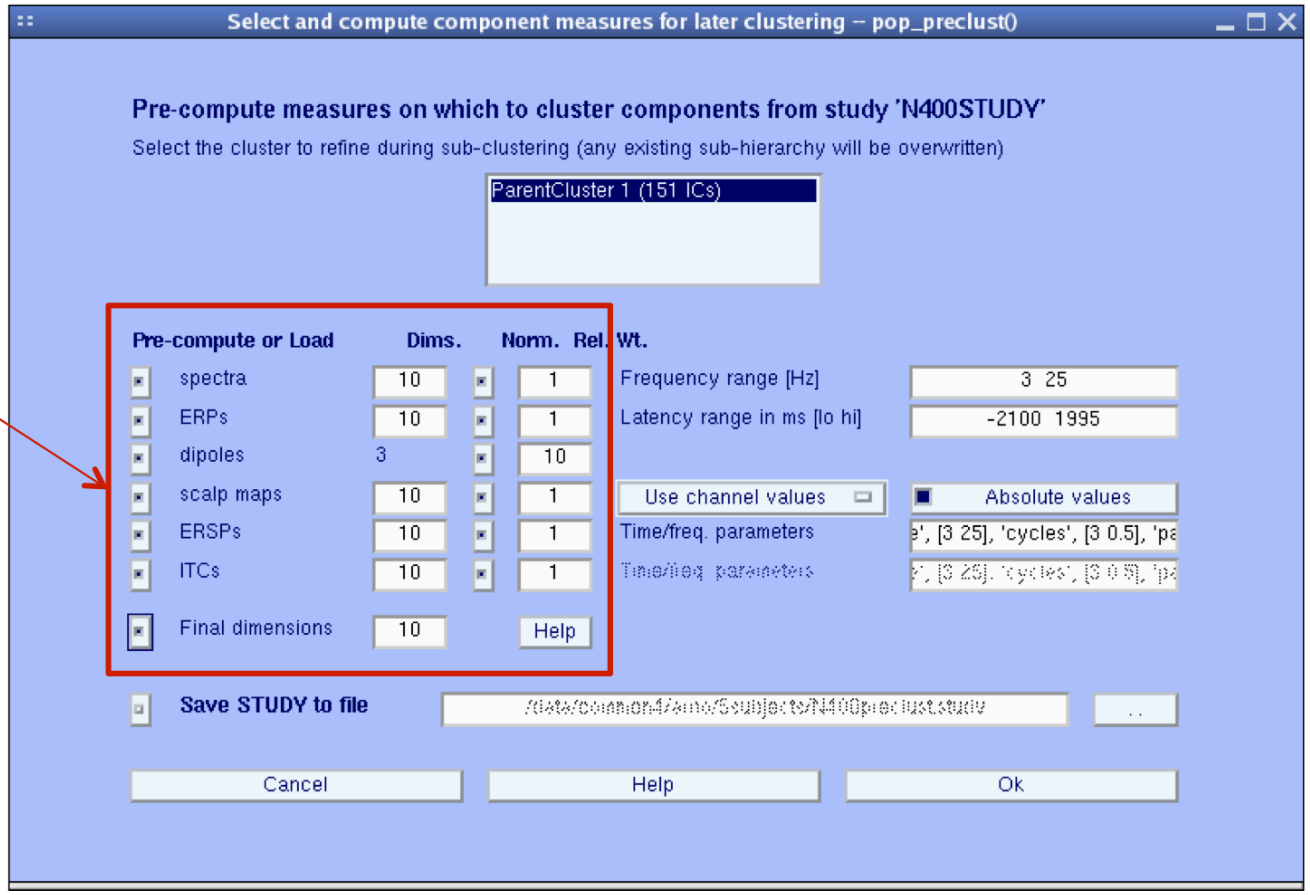


Issue with standard clustering

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

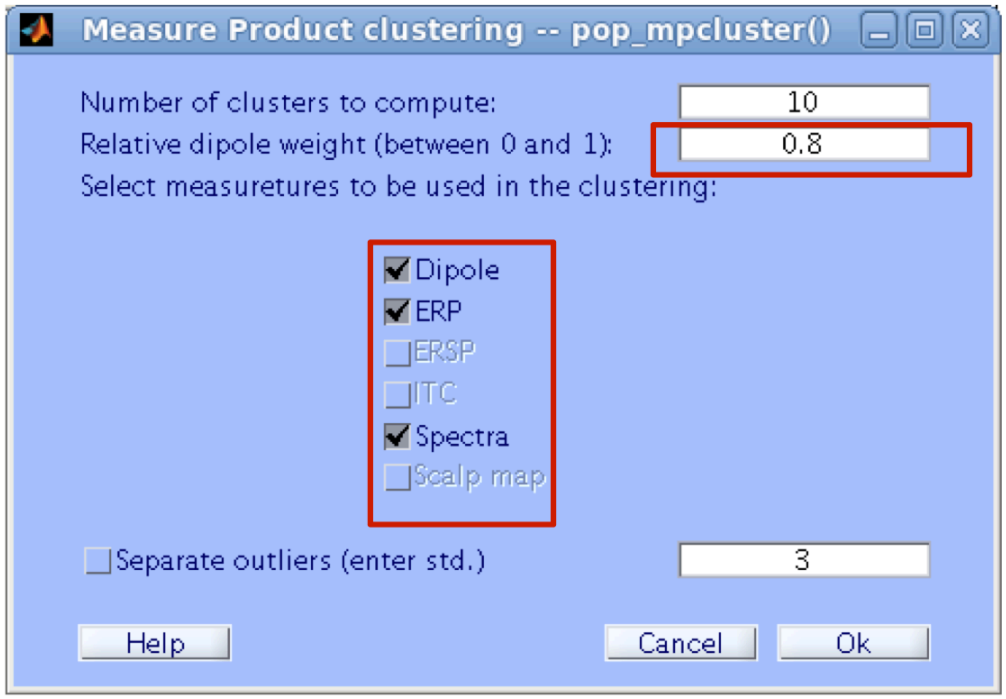
EEGLAB clustering has ~12 parameters





Measure projection

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



(Affinity clustering by Pernet, Martinez, Delorme)

Exercise

- Load the STUDY
- 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