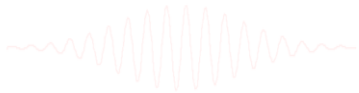
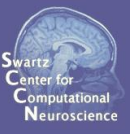
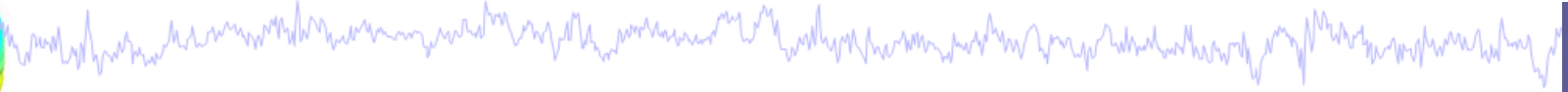


# Clustering of ICA components

Romain Grandchamp

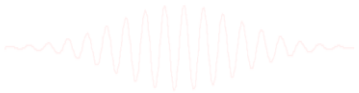
(with Arnaud Delorme, Julie Onton, Nima Bigdely Shamlo, Scott Makeig)

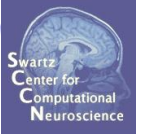




## Outline

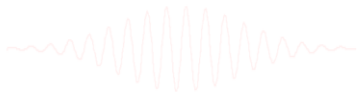
- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice

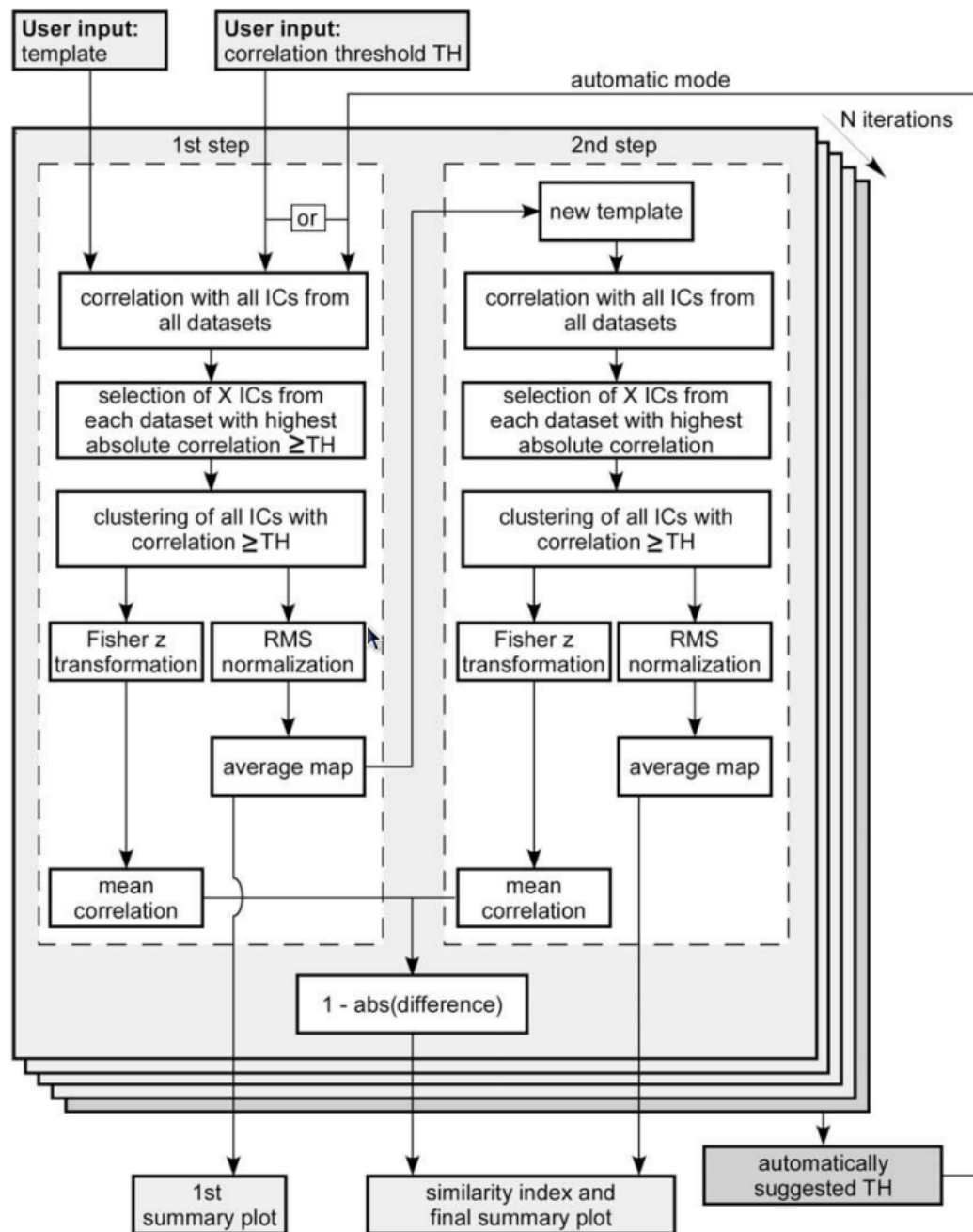




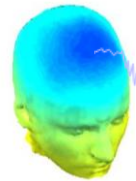
# ICA decomposition of multiple data sets from the same individuals

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





# 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)

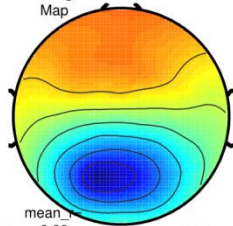
Max ICs from each dataset: 1

Cluster: 11 ICs from 11 sets

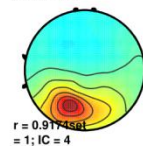
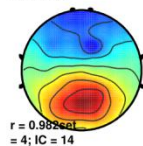
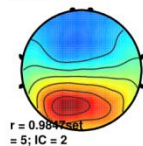
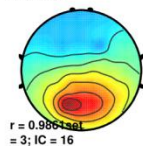
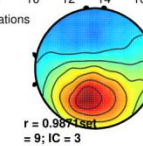
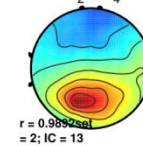
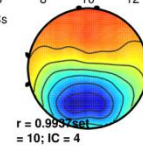
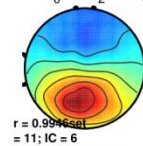
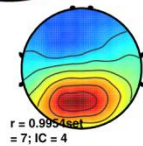
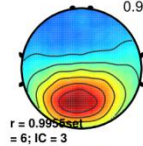
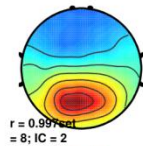
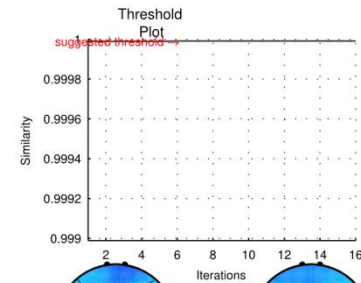
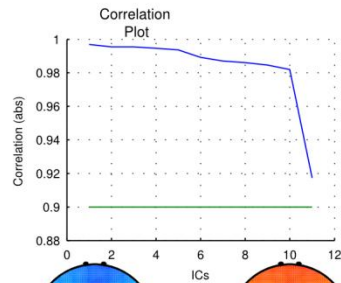
All datasets contribute.

Similarity = 1.0000

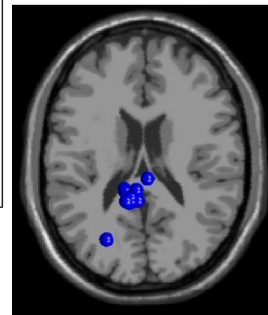
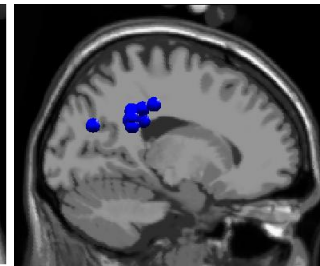
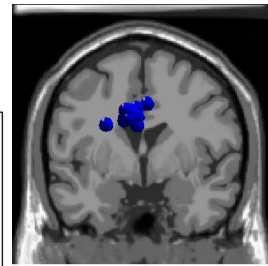
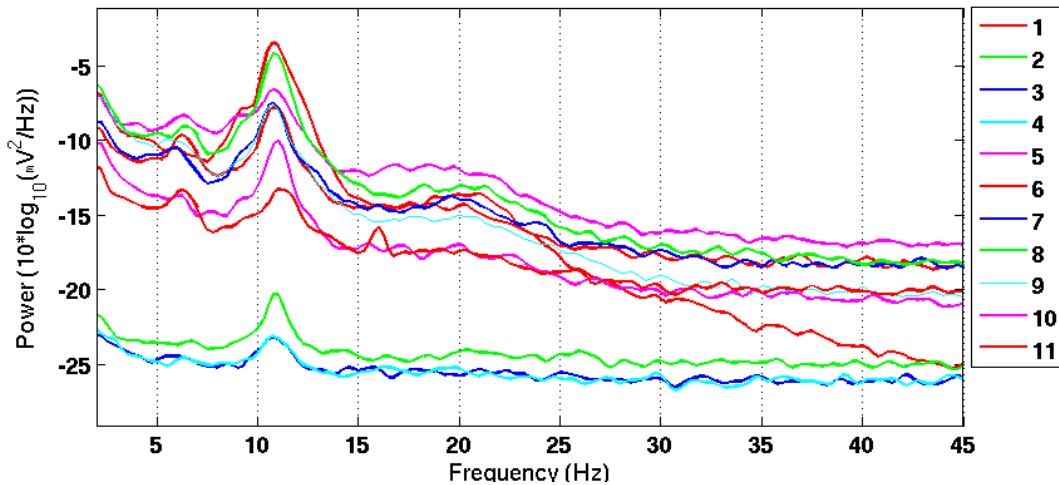
Average  
Map



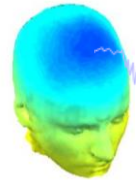
mean  
0.99



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)

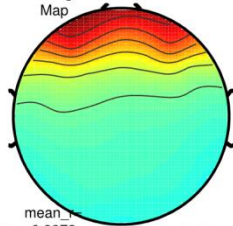
Max ICs from each dataset: 1

Cluster: 11 ICs from 11 sets

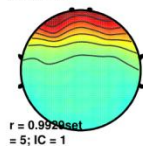
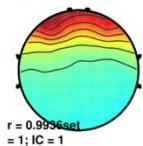
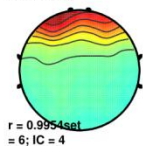
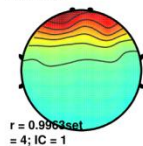
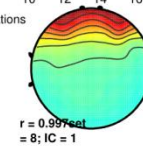
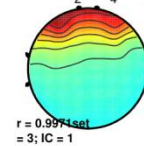
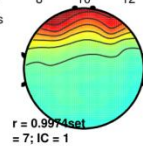
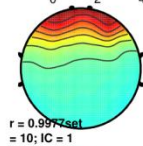
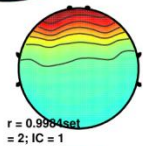
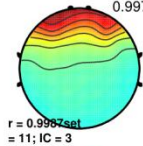
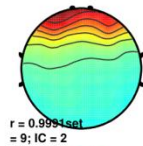
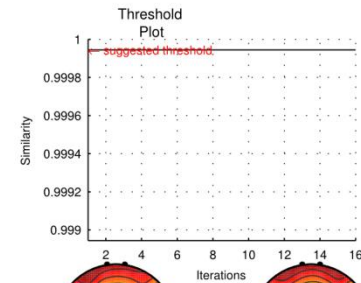
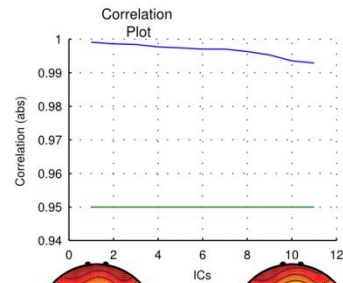
All datasets contribute.

Similarity = 0.9999

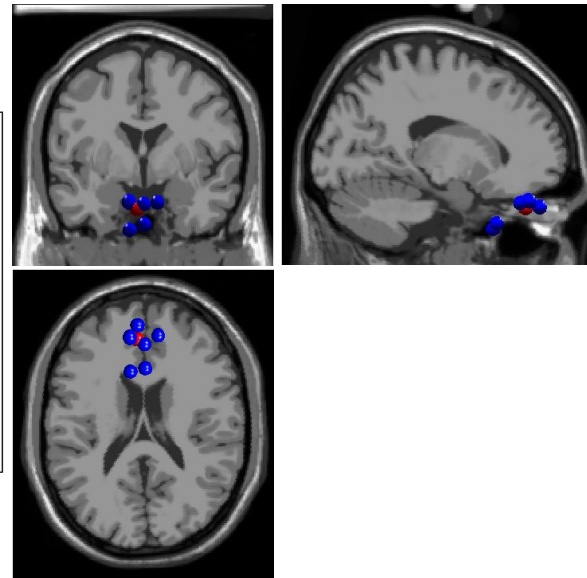
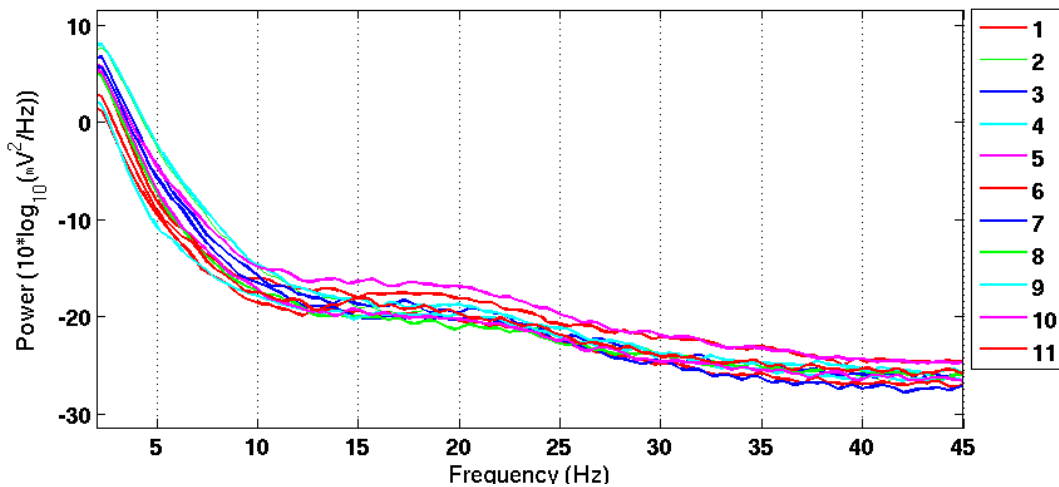
Average  
Map



mean =  
0.9972

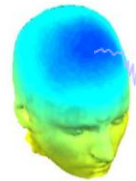


Cls 4 Spectrum





# 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)

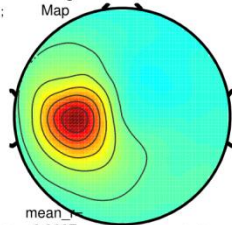
Max ICs from each dataset: 1

Cluster: 11 ICs from 11 sets

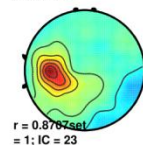
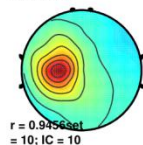
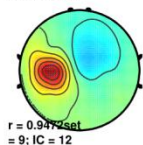
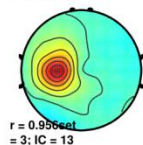
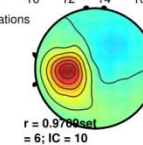
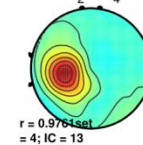
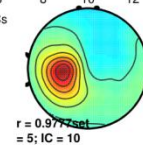
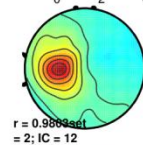
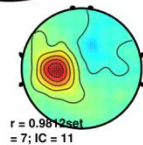
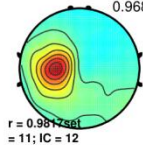
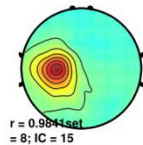
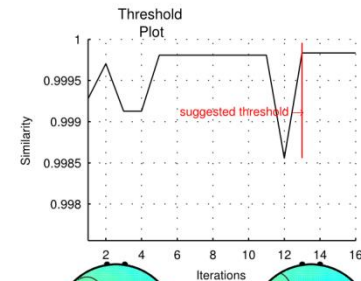
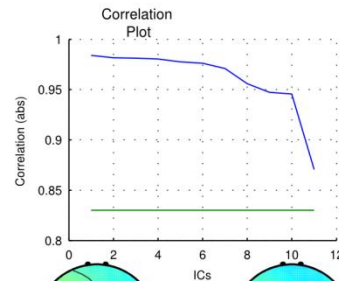
All datasets contribute.

Similarity = 0.9998

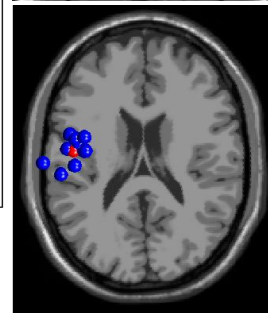
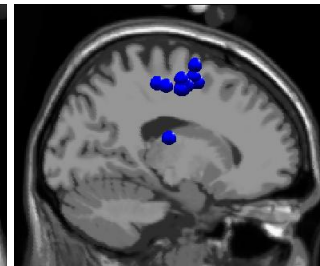
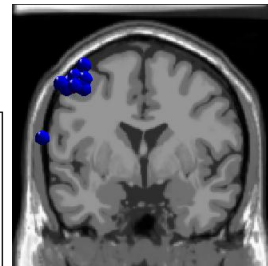
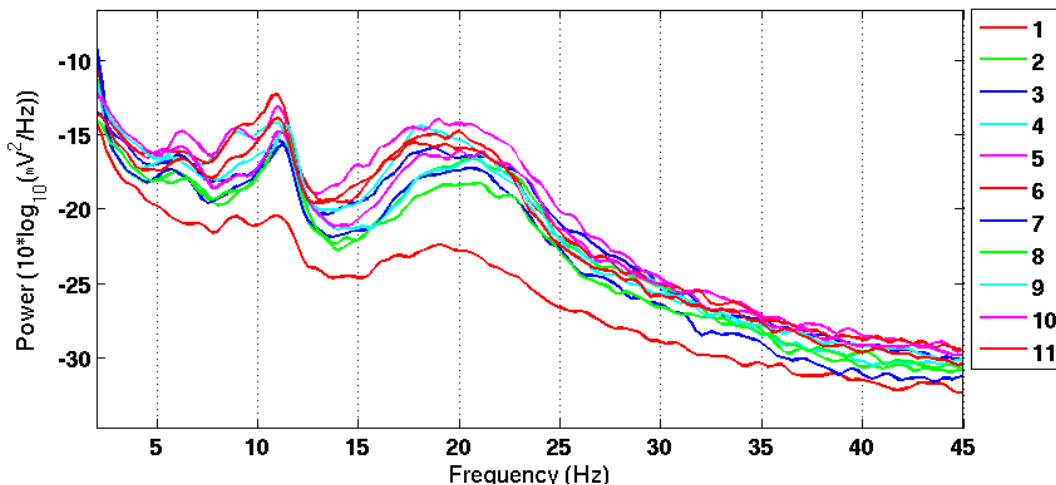
Average  
Map



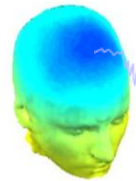
mean =  
0.9687



Cls 8 Spectrum



# Results (Cluster 13)



63.64% Sessions contribute

INFO:

Template: CB Session 2 PREPROC:STEP 2; Set 2; IC 24;

Number of datasets: 11

Correlation threshold: 0.8 (green line)

Max ICs from each dataset: 1

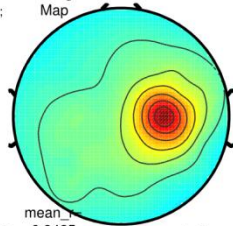
Cluster: 7 ICs from 7 sets

Sets not contributing:

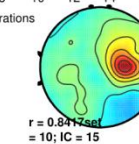
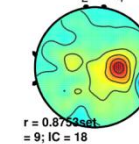
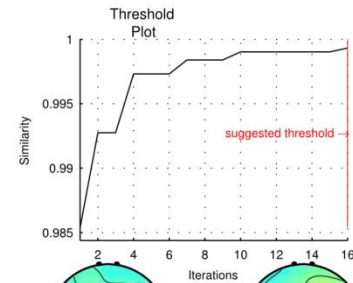
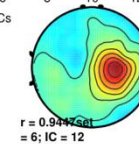
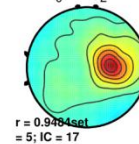
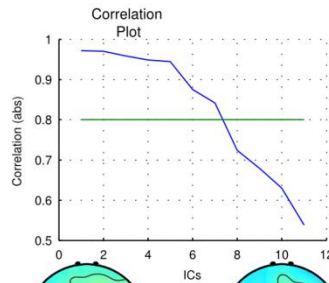
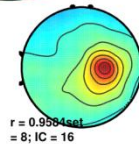
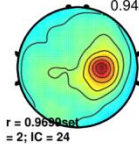
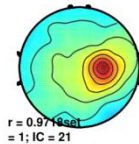
#3; #4; #7; #11;

Similarity = 0.9993

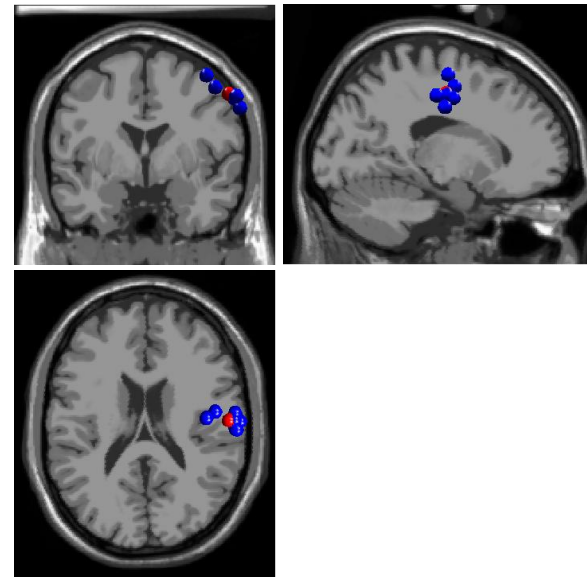
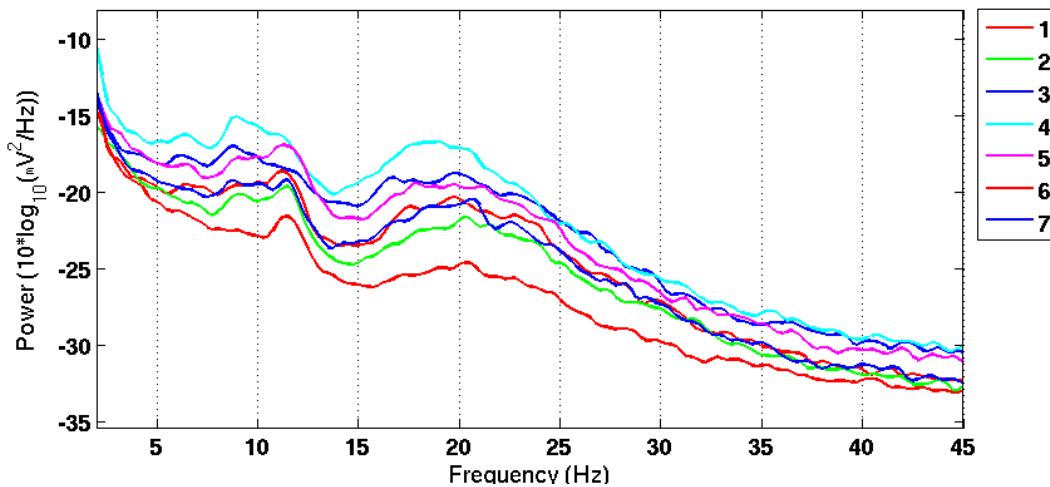
Average  
Map



mean  
0.9425

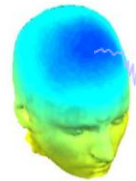


CIs 13 Spectrum





# Results (Cluster 14)



36.36% Sessions contribute

INFO:

Template: CB Session 1 PREPROC:STEP 2; Set 1; IC 5;

Number of datasets: 11

Correlation threshold: 0.93 (green line)

Max ICs from each dataset: 1

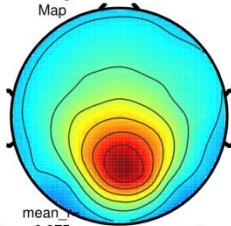
Cluster: 4 ICs from 4 sets

Sets not contributing:

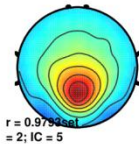
#5; #6; #7; #8; #9; #10#11;

Similarity = 0.9988

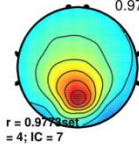
Average  
Map



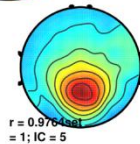
mean = 0.975



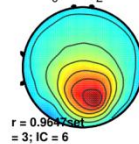
r = 0.9753  
Set = 2; IC = 5



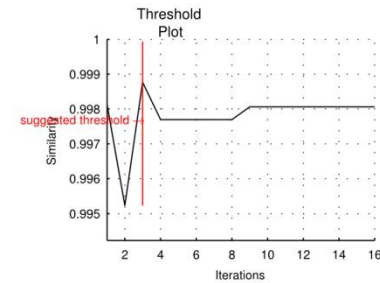
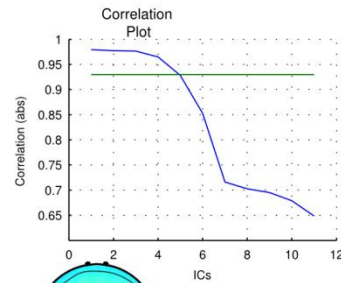
r = 0.9773  
Set = 4; IC = 7



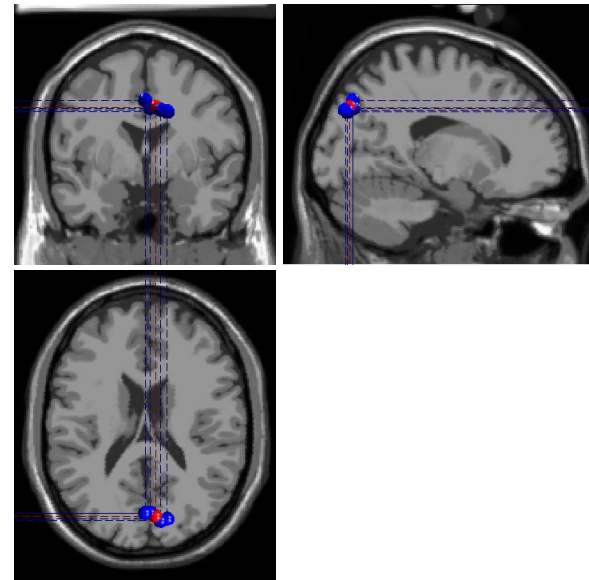
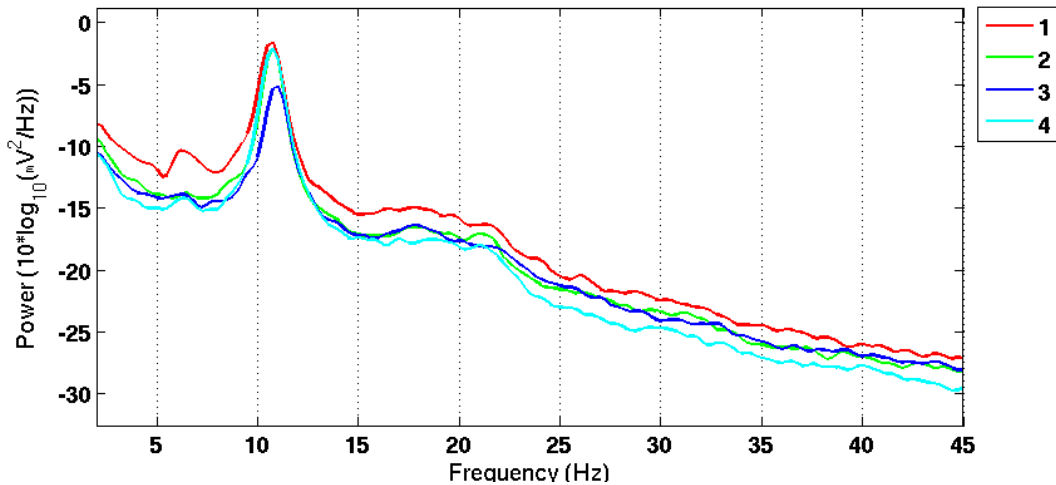
r = 0.9764  
Set = 1; IC = 5

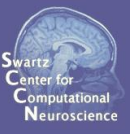
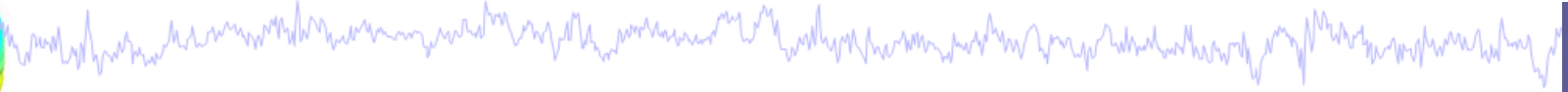


r = 0.9647  
Set = 3; IC = 6



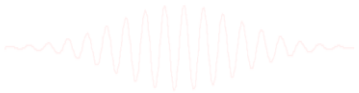
CIs 14 Spectrum



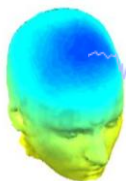


## Outline

- 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

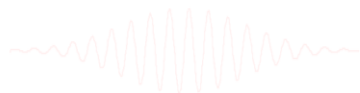
- 13 subjects performing a memory task
- 71 electrodes including EOGs
- more than 300,000 data points/subject

## Decomposition

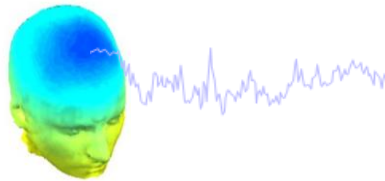
- 23 ICA algorithms plus PCA and Promax

## Analysis

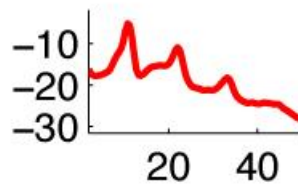
- Localization of all components with a single dipole (4-shell spherical model)



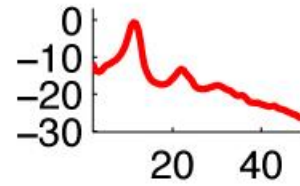
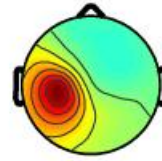
# Left $\mu$ cluster



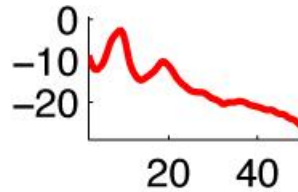
S2 IC47



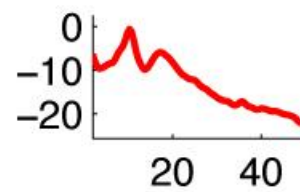
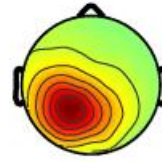
S3 IC47



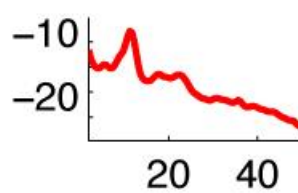
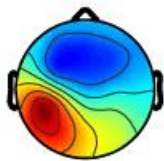
S4 IC37



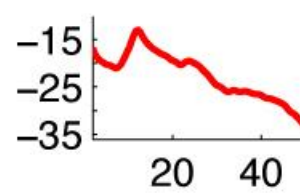
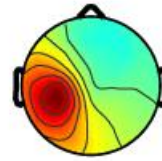
S5 IC48



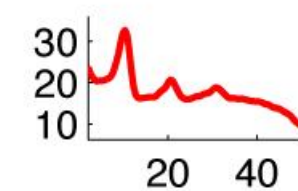
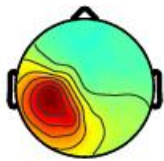
S6 IC46



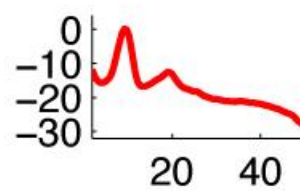
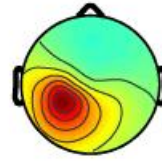
S7 IC35



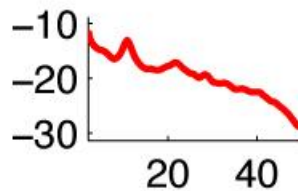
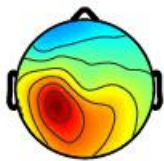
S9 IC7



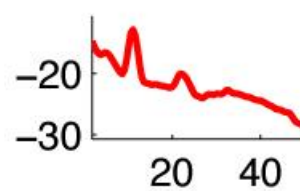
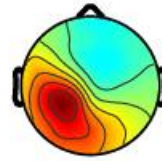
S11 IC45



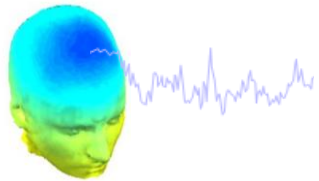
S12 IC45



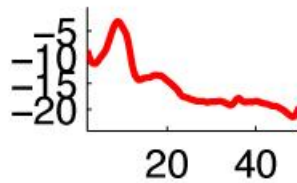
S14 IC45



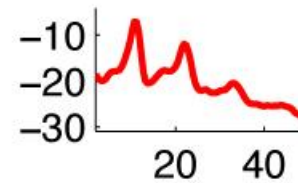
# Right $\mu$ cluster



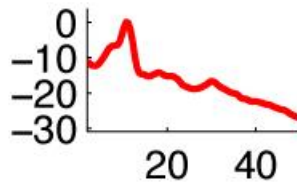
S1 IC51



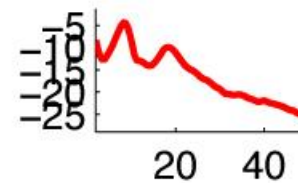
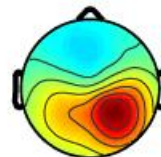
S2 IC41



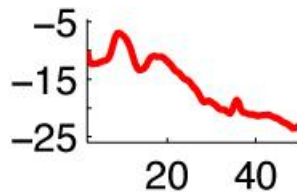
S3 IC41



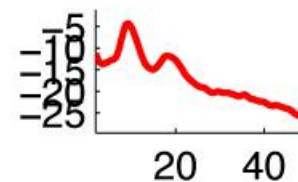
S4 IC50



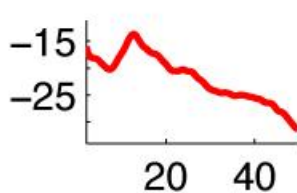
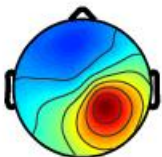
S5 IC51



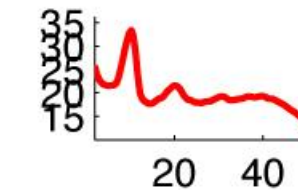
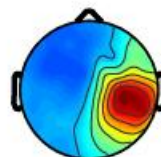
S6 IC6<sup>0</sup>



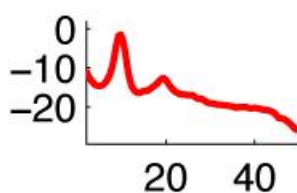
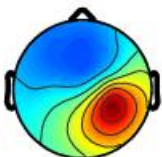
S7 IC48



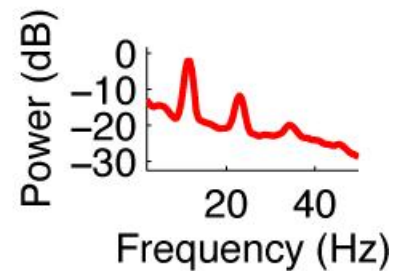
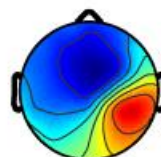
S9 IC39



S11 IC49



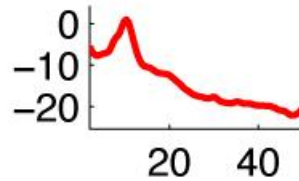
S14 IC49



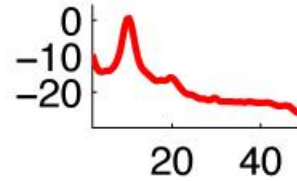
# Occipital $\alpha$ cluster



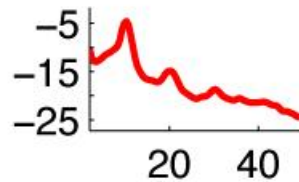
S1 IC67



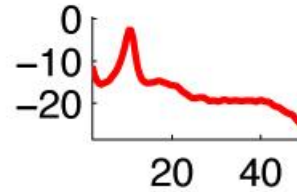
S2 IC67



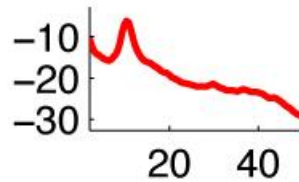
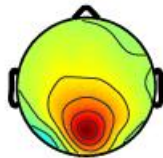
S3 IC51



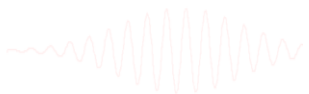
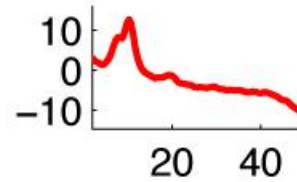
S11 IC65



S12 IC3<sup>8</sup>

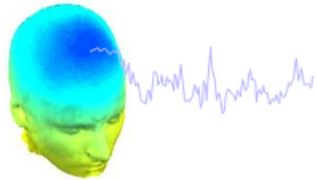


S13 IC65

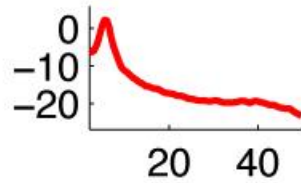




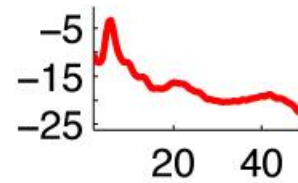
# Frontal Midline $\theta$ cluster



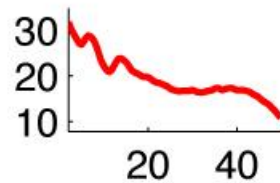
S1 IC63



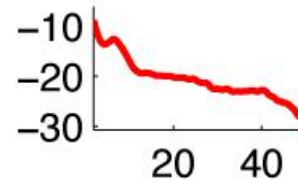
S2 IC18



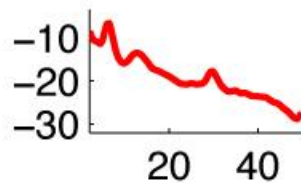
S9 IC16



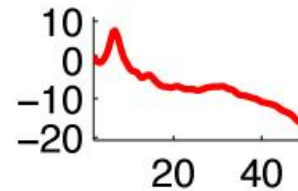
S11 IC16



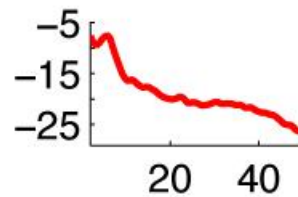
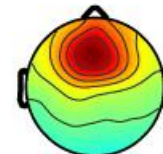
S12 IC15

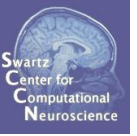
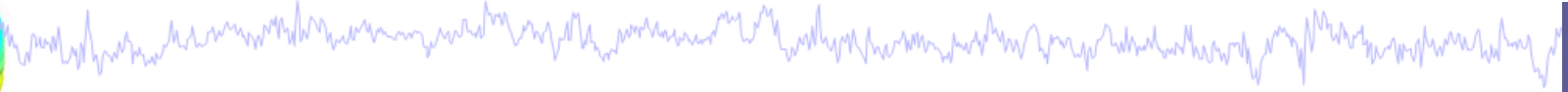


S13 IC15



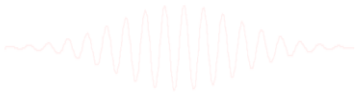
S14 IC16



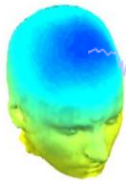


## Outline

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



# Edit dataset info



Create a new STUDY set -- pop\_study()

**Edit STUDY set information - remember to save changes**

STUDY set name:

STUDY set task name:

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

pop\_study(): Pre-select components

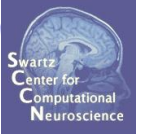
Enter maximum residual (topo map - dipole proj.) var. (in %)  
NOTE: This will delete any existing component clusters!

☒ Keep only in-brain dipoles.

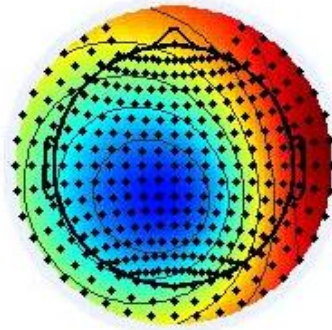
Cancel Help Ok



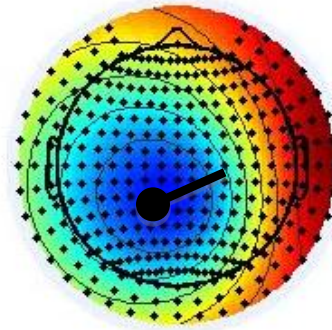
# Computing residual variance (%)



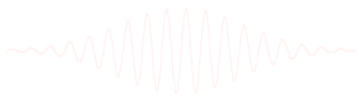
Actual



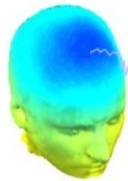
Dipole projection



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



# ICs to cluster



Create a new STUDY set -- pop\_study()

**Edit STUDY set information - remember to save changes**

STUDY set name: Sternberg  
STUDY set task: Sternberg  
STUDY set notes:

dataset filename:

	dataset filename	condition	group	Select by r.v.	
1	C:\Users\julieW\...	memorize		Comp.: 3 5 ...	Clear
2	C:\Users\julieW\...	ignore		Comp.: 3 5 ...	Clear
3	C:\Users\julieW\...	probe		Comp.: 3 5 ...	Clear
4	C:\Users\julieW\...	memorize		Comp.: 5 6 ...	Clear
5	C:\Users\julieW\...	ignore		Comp.: 5 6 ...	Clear
6	C:\Users\julieW\...	probe		Comp.: 5 6 ...	Clear
7	C:\Users\julieW\...	memorize		Comp.: 6 7 ...	Clear
8	C:\Users\julieW\...	ignore		Comp.: 6 7 ...	Clear
9	C:\Users\julieW\...	probe		Comp.: 6 7 ...	Clear
10	C:\Users\julieW\...	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

select components

- ic 19
- ic 20
- ic 21
- ic 22
- ic 23
- ic 24
- ic 25
- ic 26
- ic 27
- ic 28
- ic 29
- ic 30
- ic 31
- ic 32

Cancel Ok

Select by r.v.

Comp.: 3 5 ... Clear

Comp.: 3 5 ... Clear

Comp.: 3 5 ... Clear

Comp.: 5 6 ... Clear

Comp.: 5 6 ... Clear

Comp.: 5 6 ... Clear

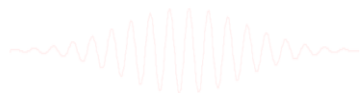
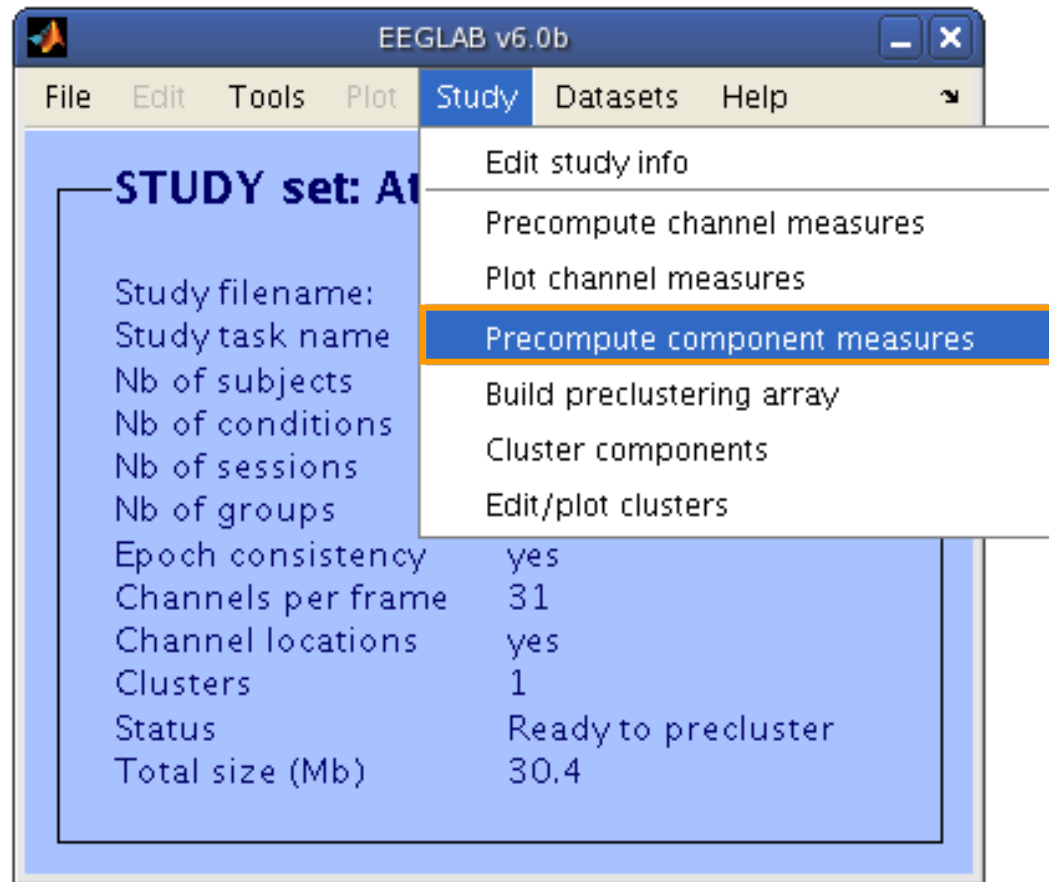
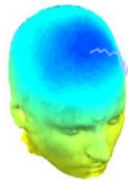
Comp.: 6 7 ... Clear

Comp.: 6 7 ... Clear

Comp.: 6 7 ... Clear

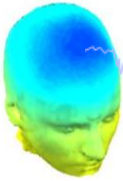
Comp.: 1 2 ... Clear

# Precompute data measures

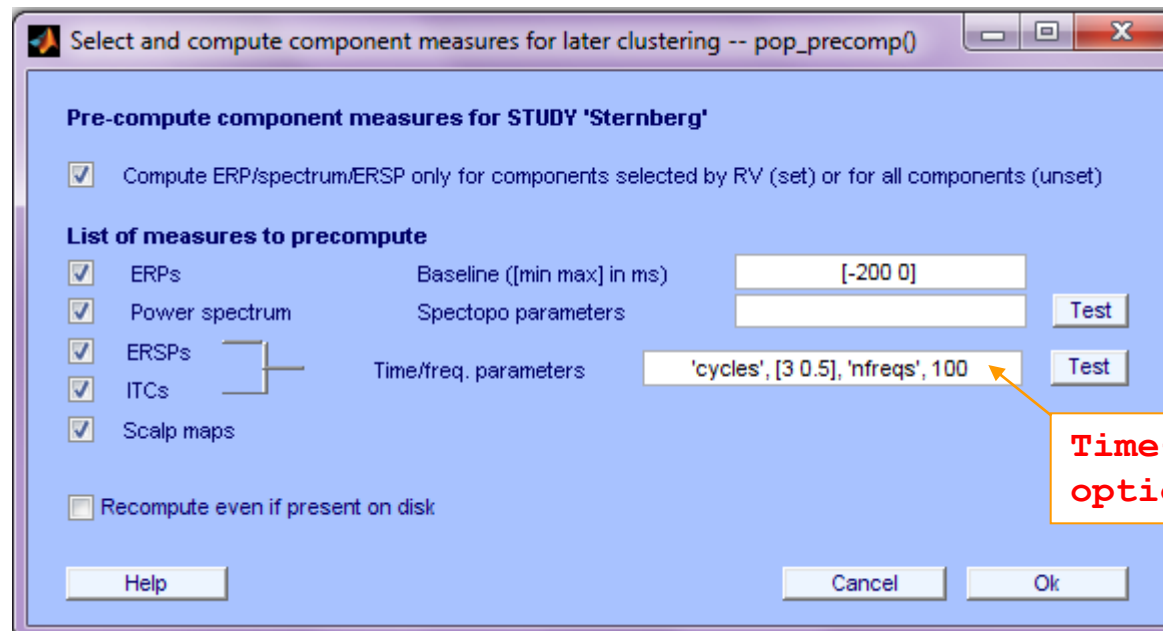




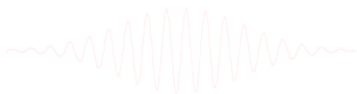
# Precompute data measures



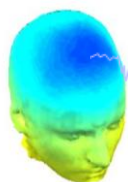
**TIP: Compute all measures so you can test different combinations for clustering**



**Time-frequency options**



# 3. Cluster components



EEGLAB v6.0b

File Edit Tools Plot Study Datasets Help

**STUDY set: Attention**

Study filename:  
Study task name  
Nb of subjects  
Nb of conditions  
Nb of sessions  
Nb of groups  
Epoch consistency yes  
Channels per frame 31  
Channel locations yes  
Clusters 1  
Status Pre-clustered  
Total size (Mb) 32.4

Edit study info  
Precompute channel measures  
Plot channel measures  
Precompute component measures  
**Build preclustering array**  
Cluster components  
Edit/plot clusters

Select and compute component measures for later clustering -- pop\_preclust()

**Build pre-clustering matrix for STUDY 'Attention'**  
Select the cluster to refine during sub-clustering (any existing sub-hierarchy will be overwritten)

ParentCluster 1 (181 ICs)

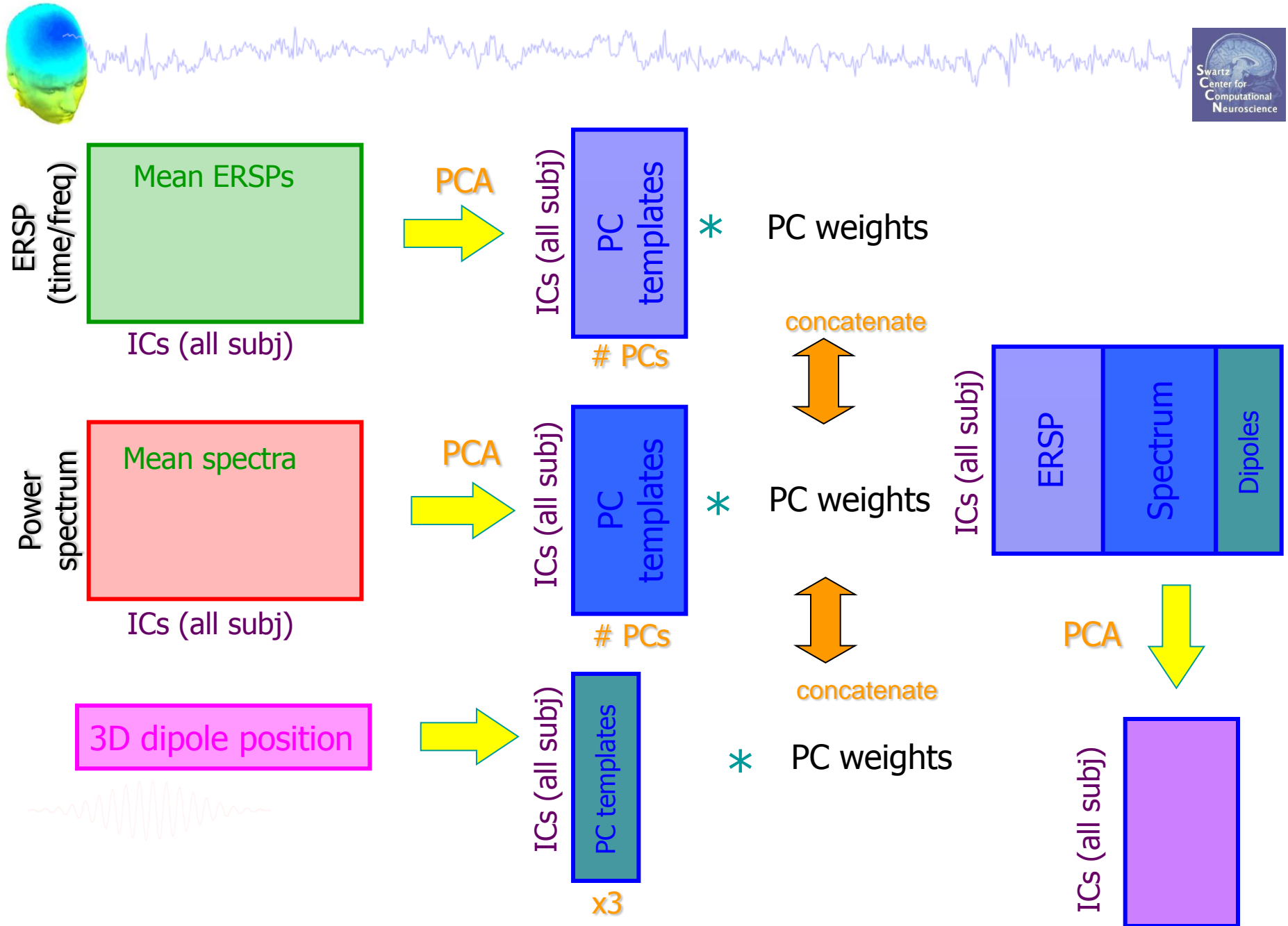
(note: only measures that have been precomputed may be used)

Load	Dims.	Norm.	Rel. Wt.	
<input checked="" type="checkbox"/> spectra	10	<input checked="" type="checkbox"/> 1		Freq. range [Hz] 3 25
<input checked="" type="checkbox"/> ERPs	10	<input checked="" type="checkbox"/> 1		Time range [ms] 0 600
<input checked="" type="checkbox"/> dipoles	3	<input checked="" type="checkbox"/> 10		
<input type="checkbox"/> scalp maps	10	<input checked="" type="checkbox"/> 1		Use channel values <input checked="" type="checkbox"/> Absolute values
<input checked="" type="checkbox"/> ERSPs	20	<input checked="" type="checkbox"/> 1		Time range [ms] 0 1500 Freq. range [Hz] 3 45
<input checked="" type="checkbox"/> ITCs	10	<input checked="" type="checkbox"/> 1		Time range [ms] 0 600 Freq. range [Hz] 2 30
<input type="checkbox"/> Final dimensions	10			

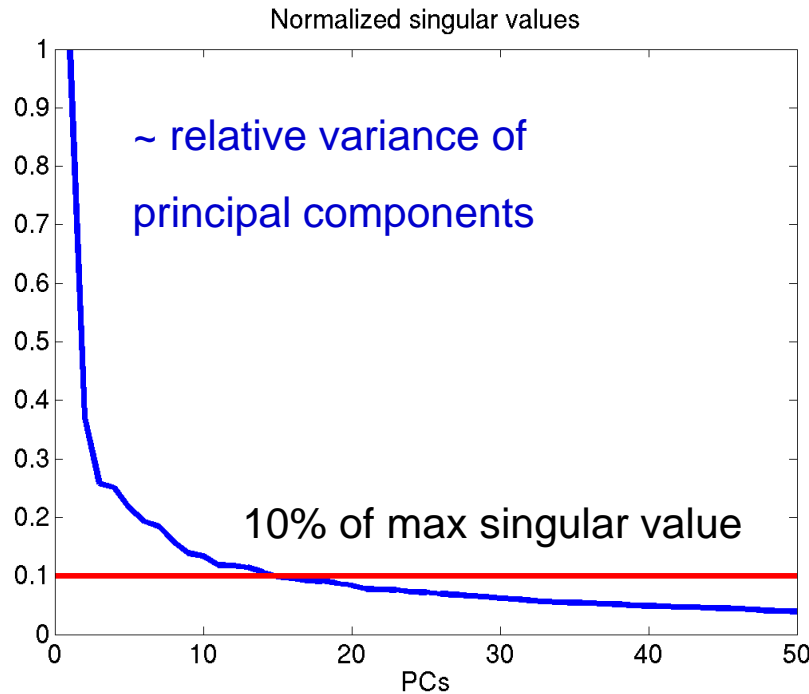
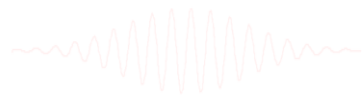
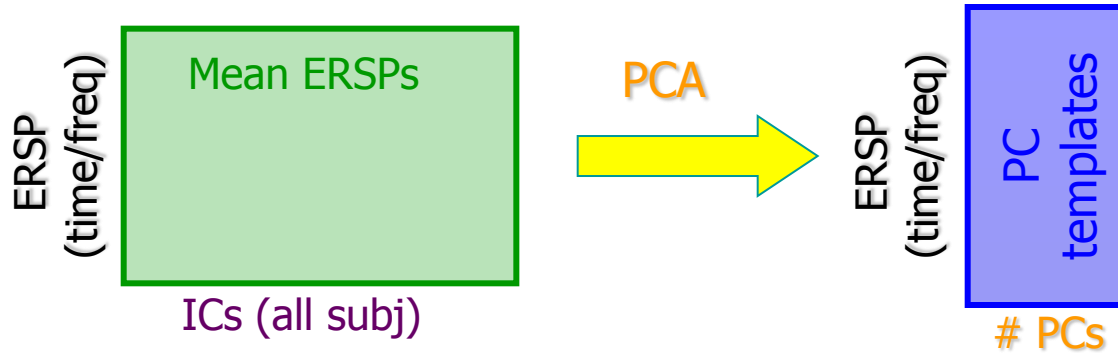
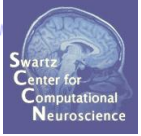
☐ Save STUDY to file /home/julie/WorkshopSD2007/STUDY/attention.study ...

Cancel Help Ok

# Precluster schematic

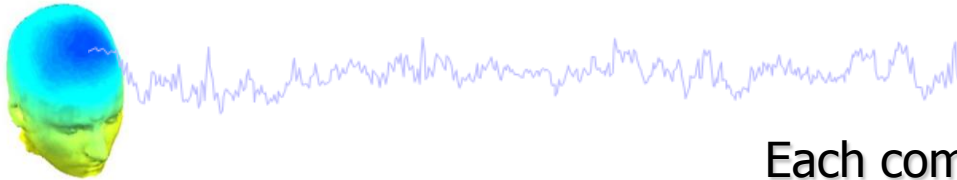


# Precluster: Use singular values from PCA

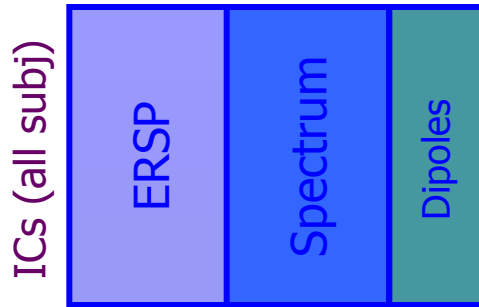


Credit: Julie Onton

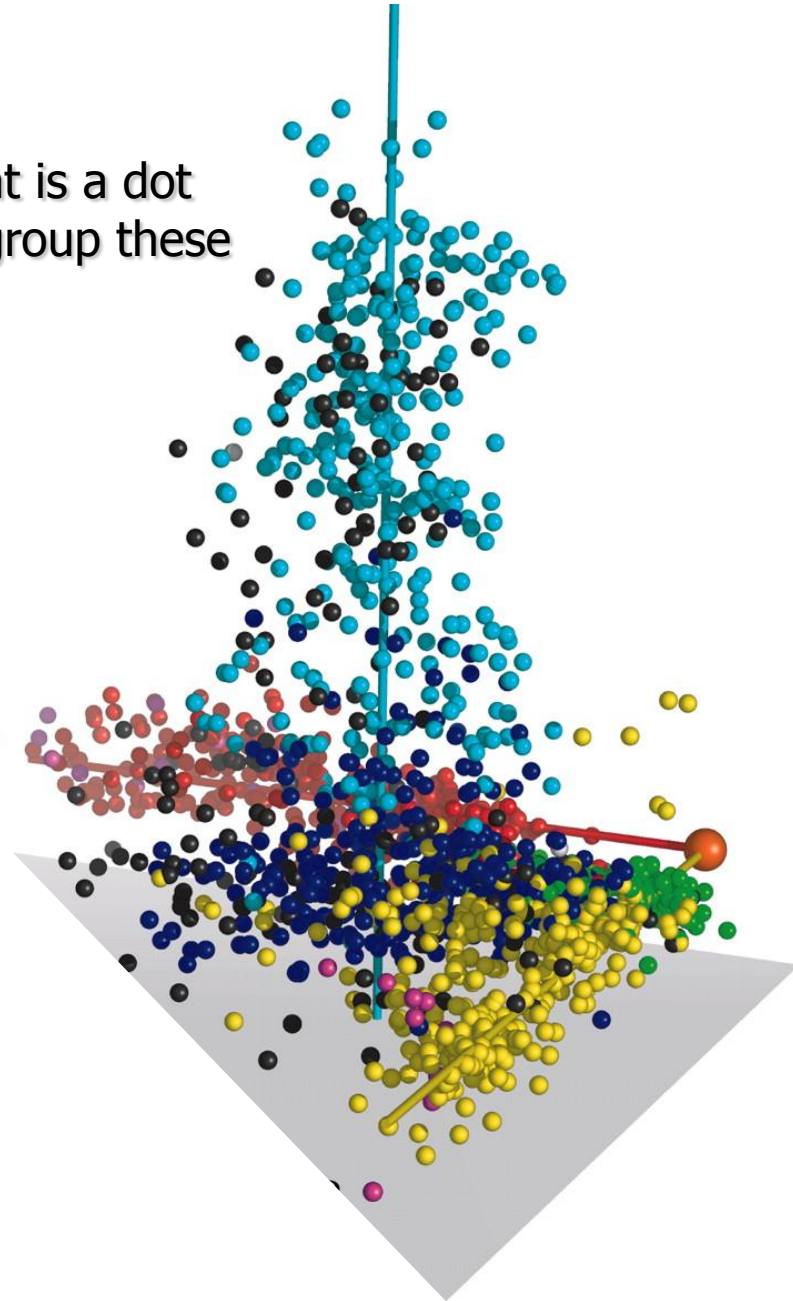
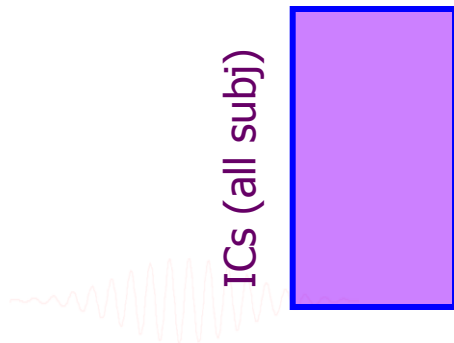
# Precluster schematic



Each component is a dot  
Clustering will group these  
dots



OR



# Classical KMean

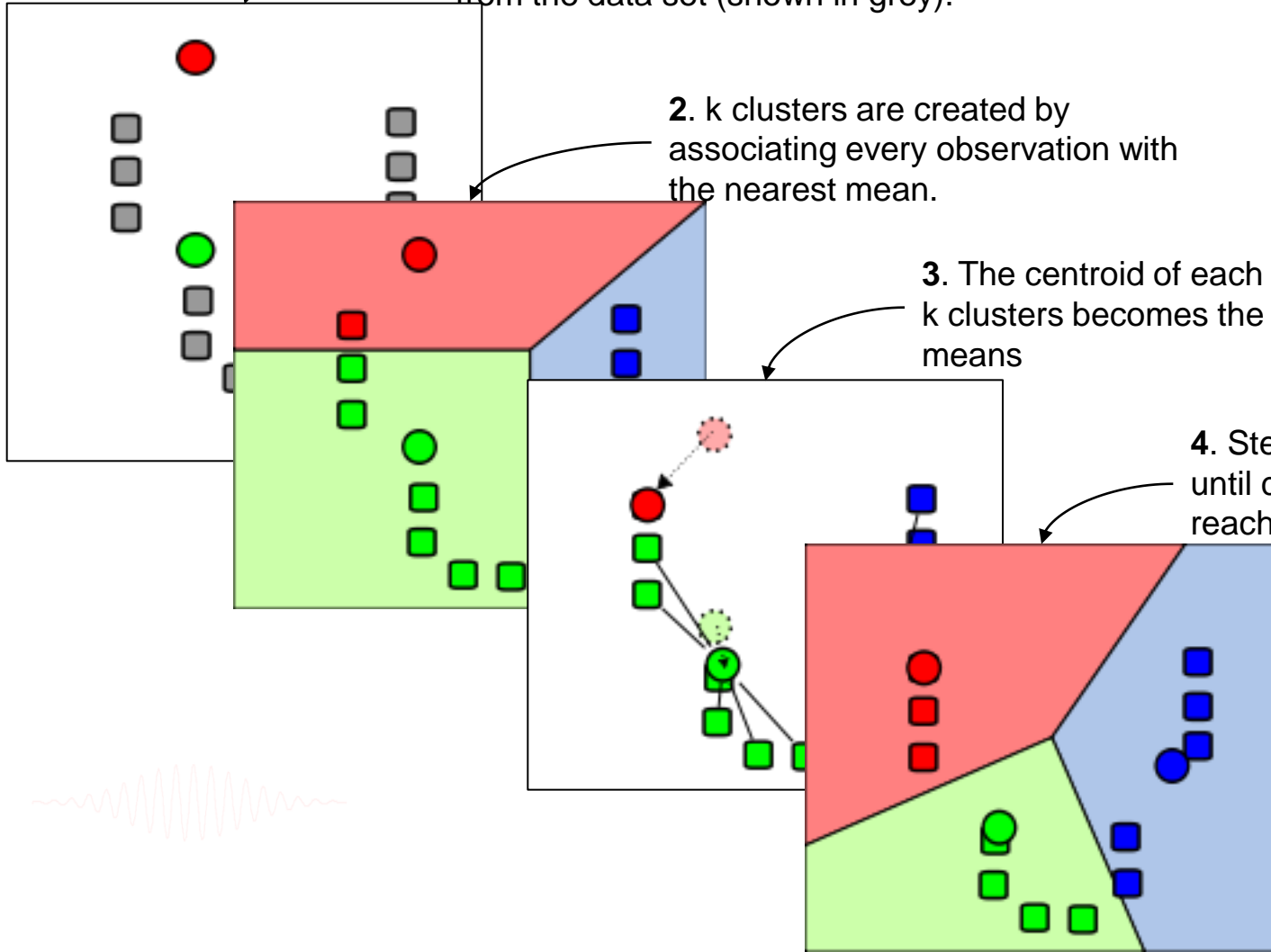


1.  $k$  initial "means" (in this case  $k=3$ , (shown in color)) are randomly selected from the data set (shown in grey).

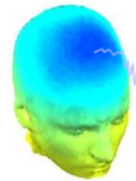
2.  $k$  clusters are created by associating every observation with the nearest mean.

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

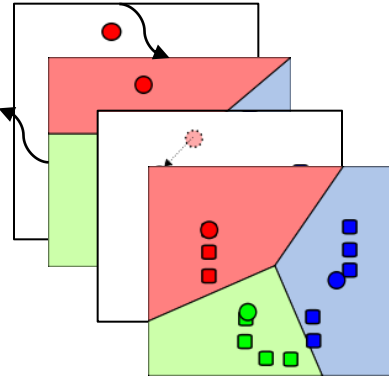
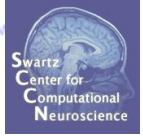
4. Steps 2 and 3 are repeated until convergence has been reached.







# 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

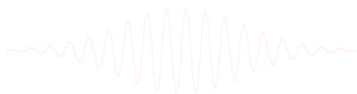
3. Keeps at most one component per session  
(min dist. to centroid)

4. Store the resulting cluster

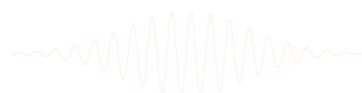
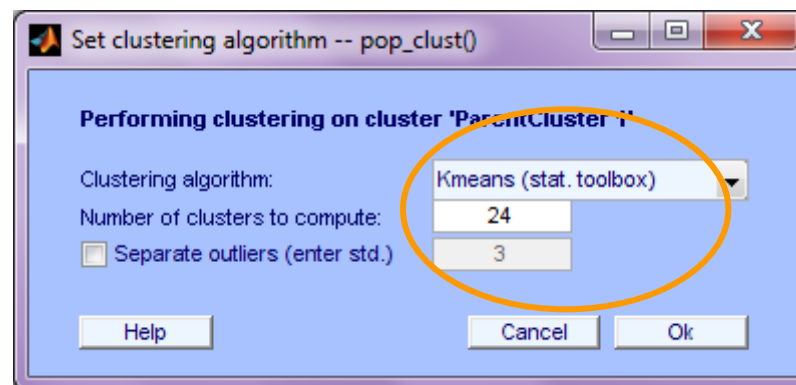
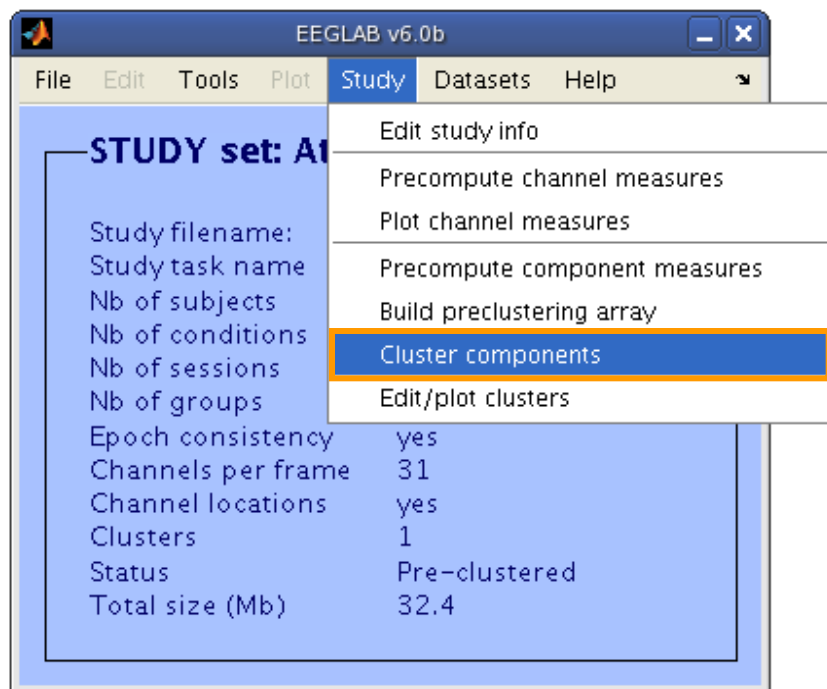
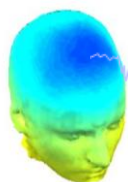
5. Remove the cluster's ICs from the pool of all ICs

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



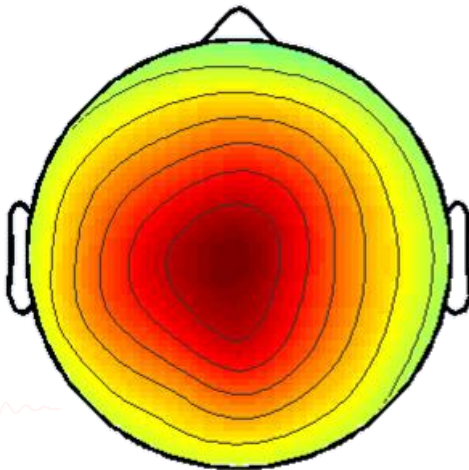
What measure(s) should you use?

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

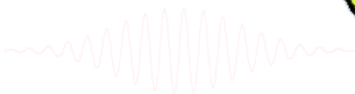
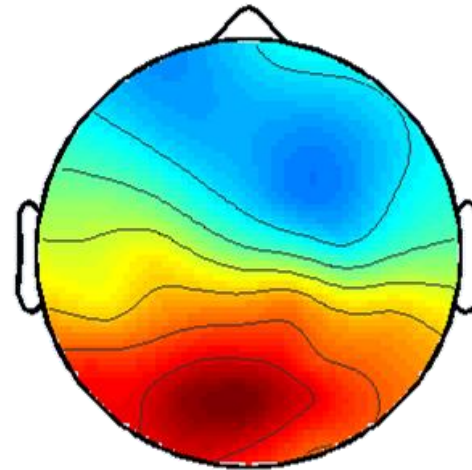
But consider:

- What is the difference between these two components?

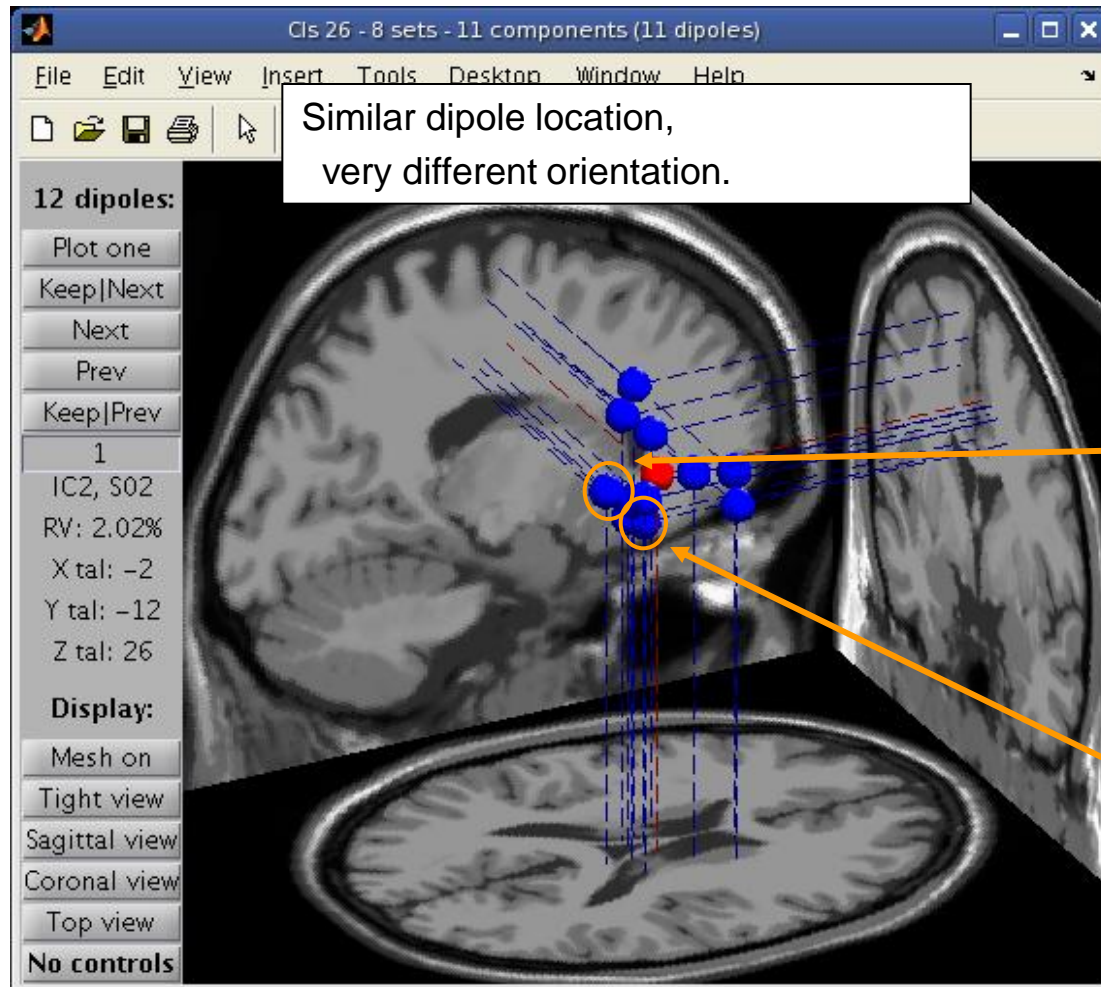
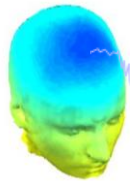
IC2 / S02, CIs 26



IC5 / S05, CIs 26

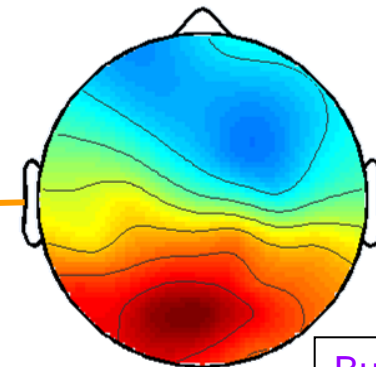


# Choosing data measures

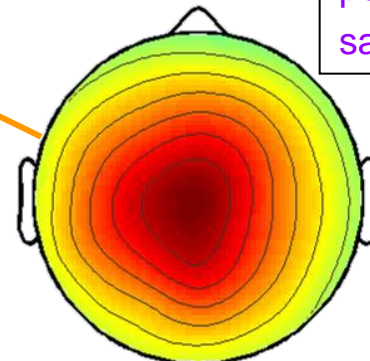


Obvious dramatic effect on  
scalp map topography:

IC5 / S05, Cls 26

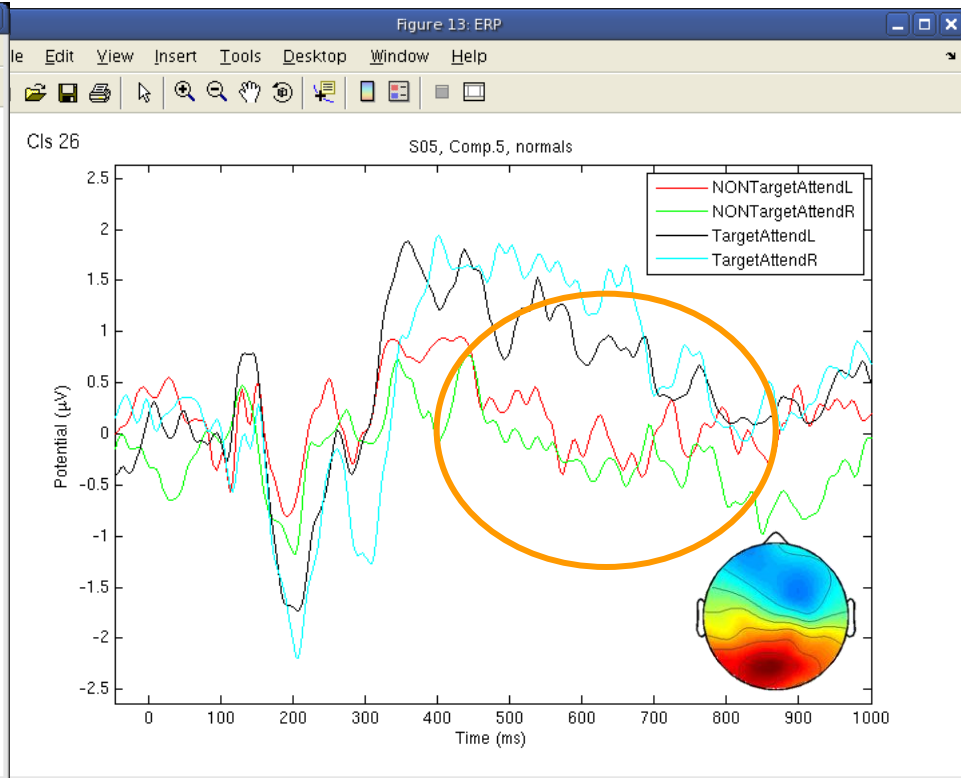
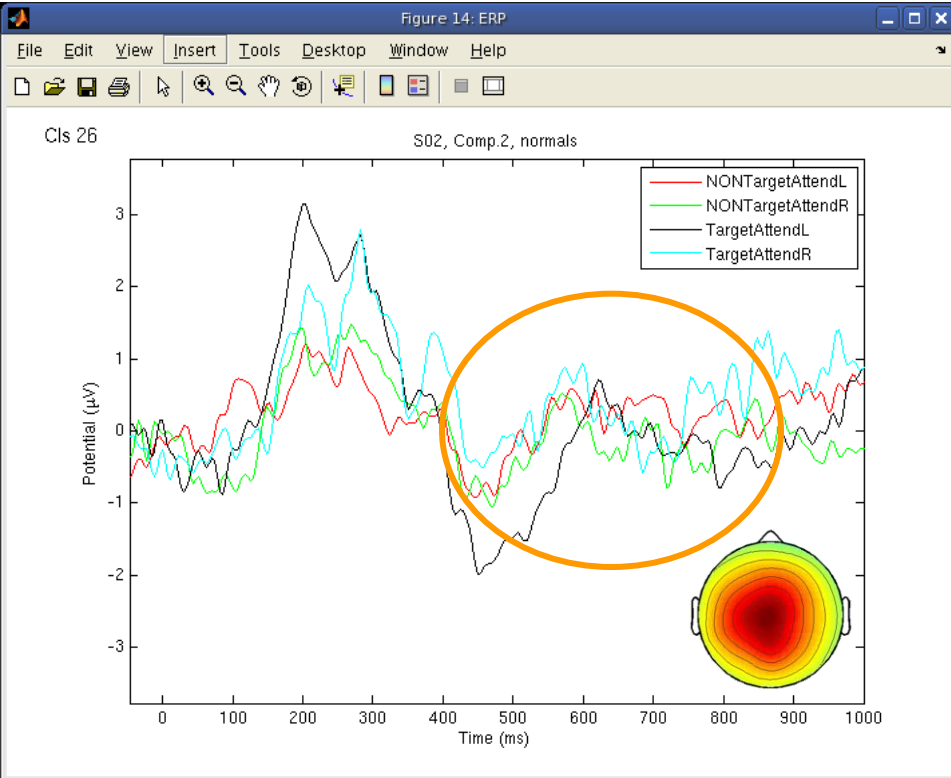


IC2 / S02, Cls 26

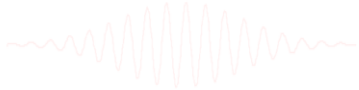


But, do they  
perform the  
same functions?

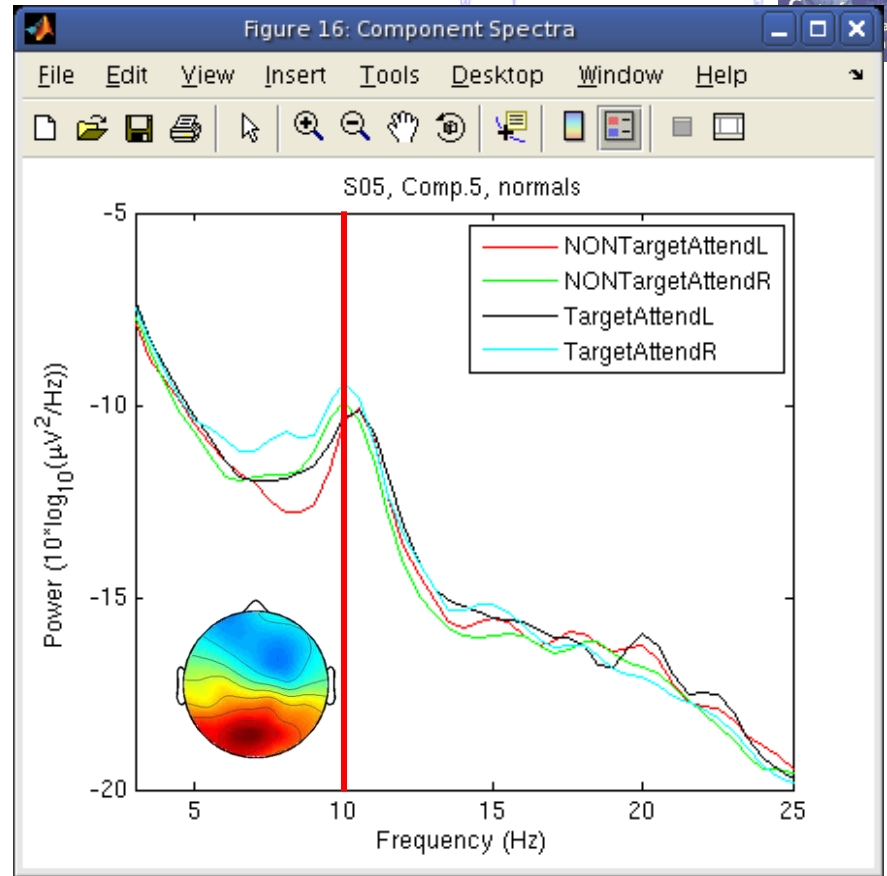
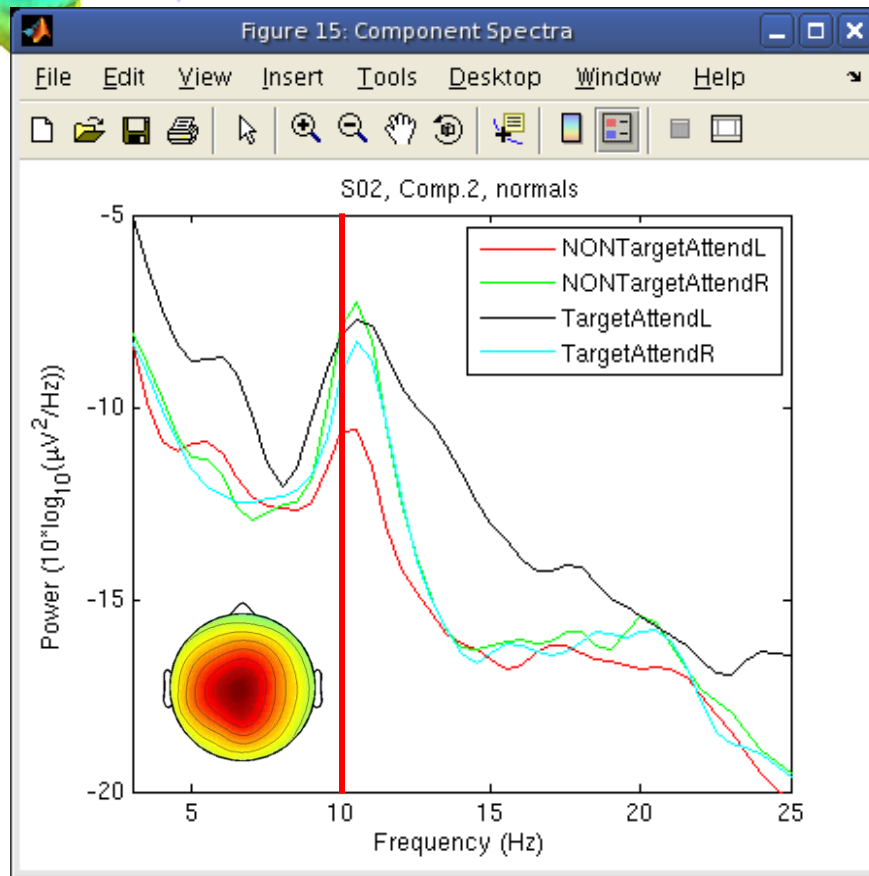
# Choosing data measures



ERPs seem different...



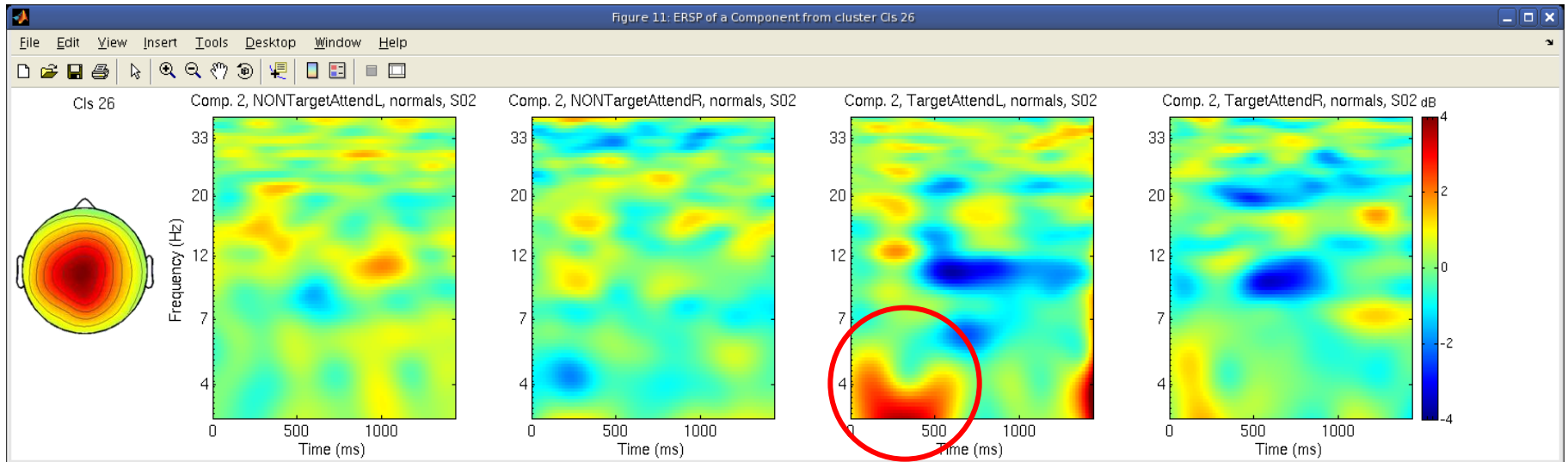
# Choosing data measures



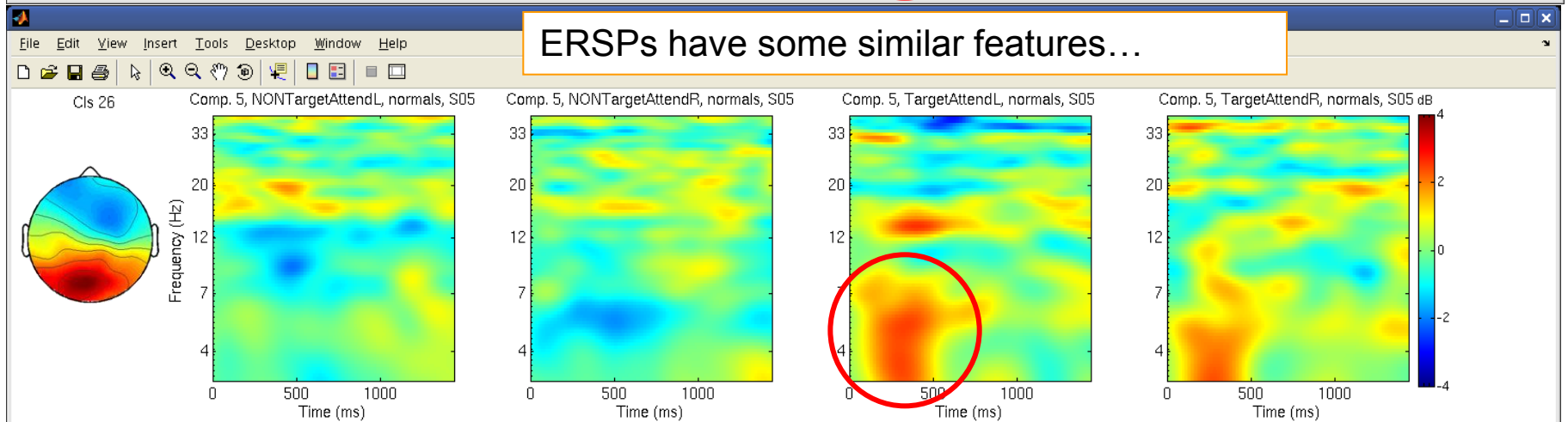
Spectra are similar, but they have  
variable responses to different conditions...



# Choosing data measures



ERSPs have some similar features...



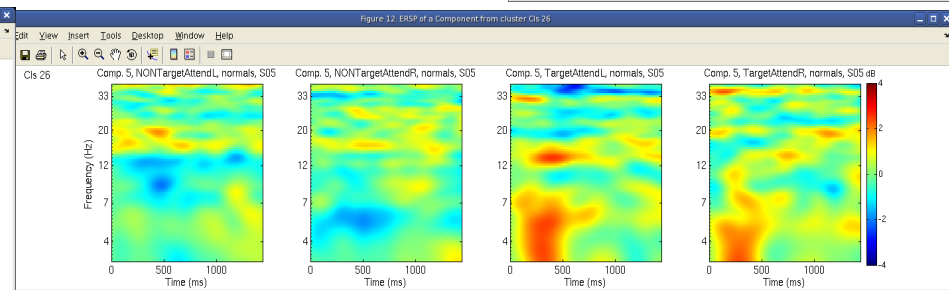
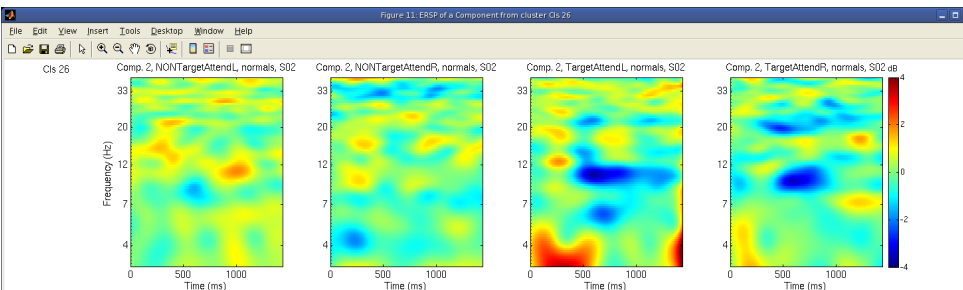
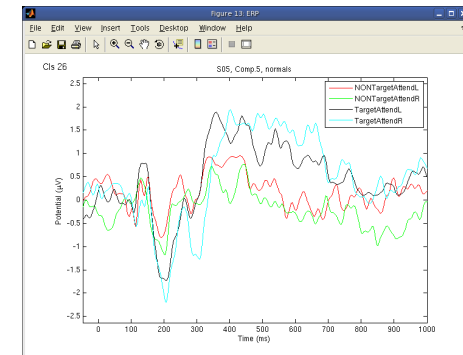
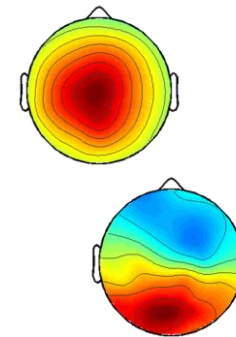
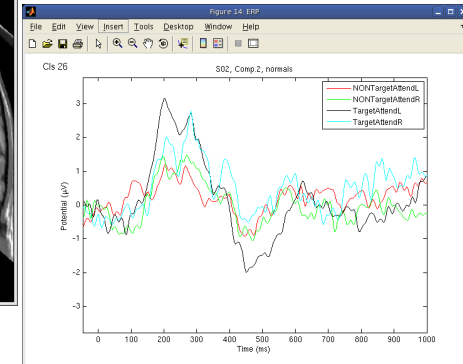
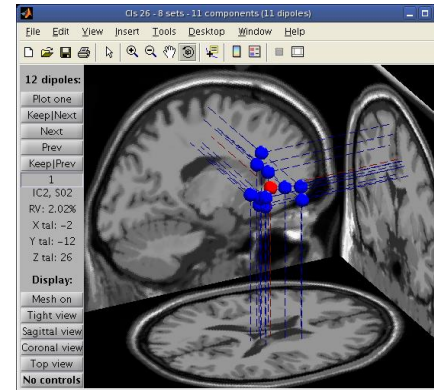
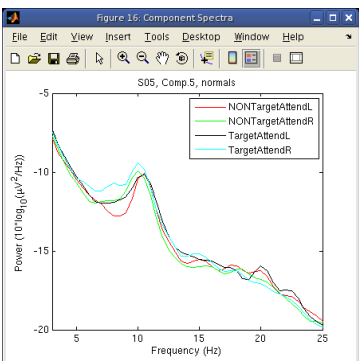
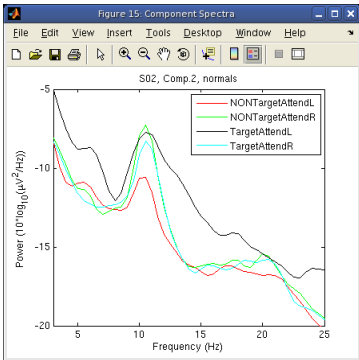
# Choosing data measures



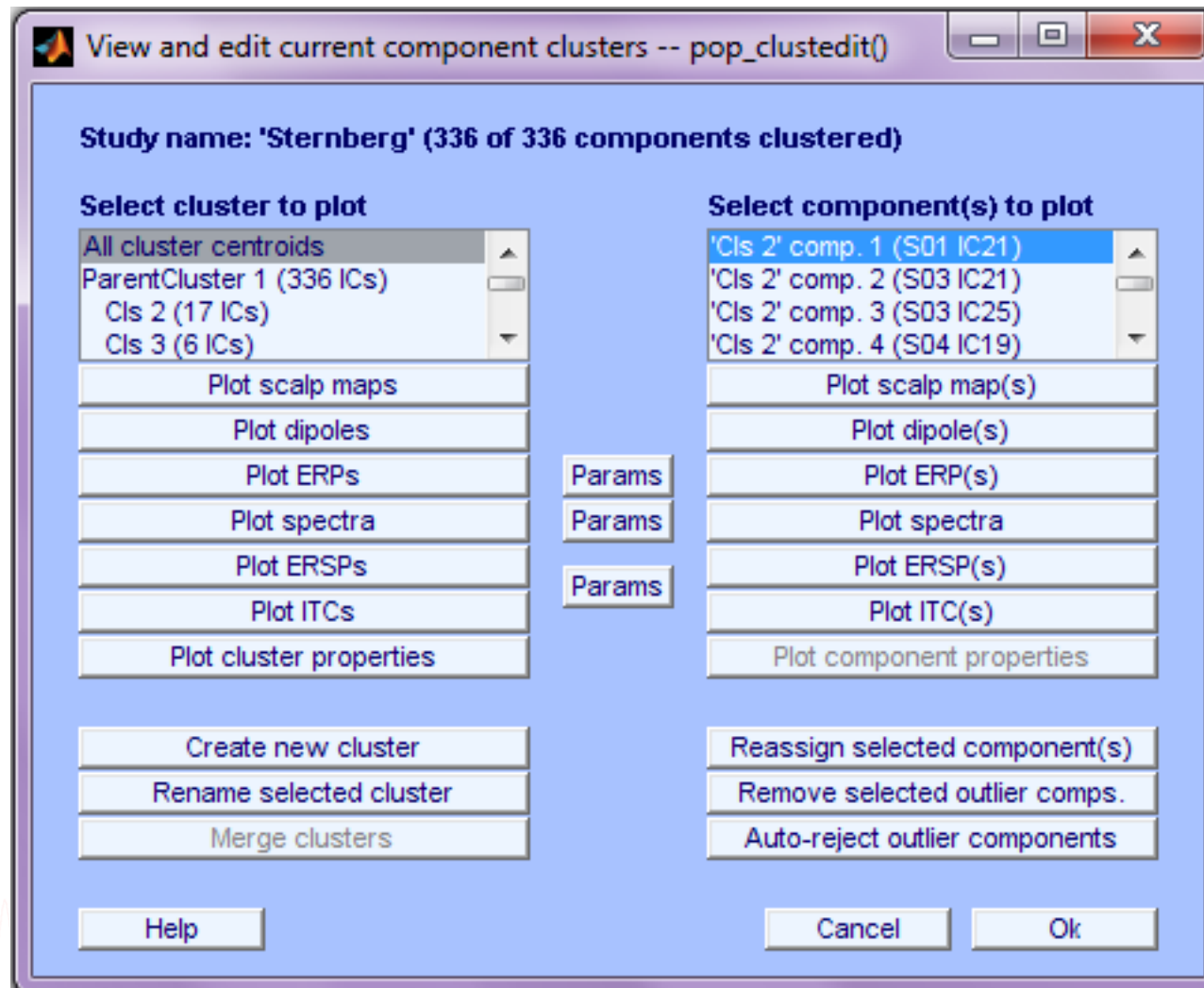
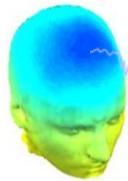
What data measures  
should you use?

It depends...

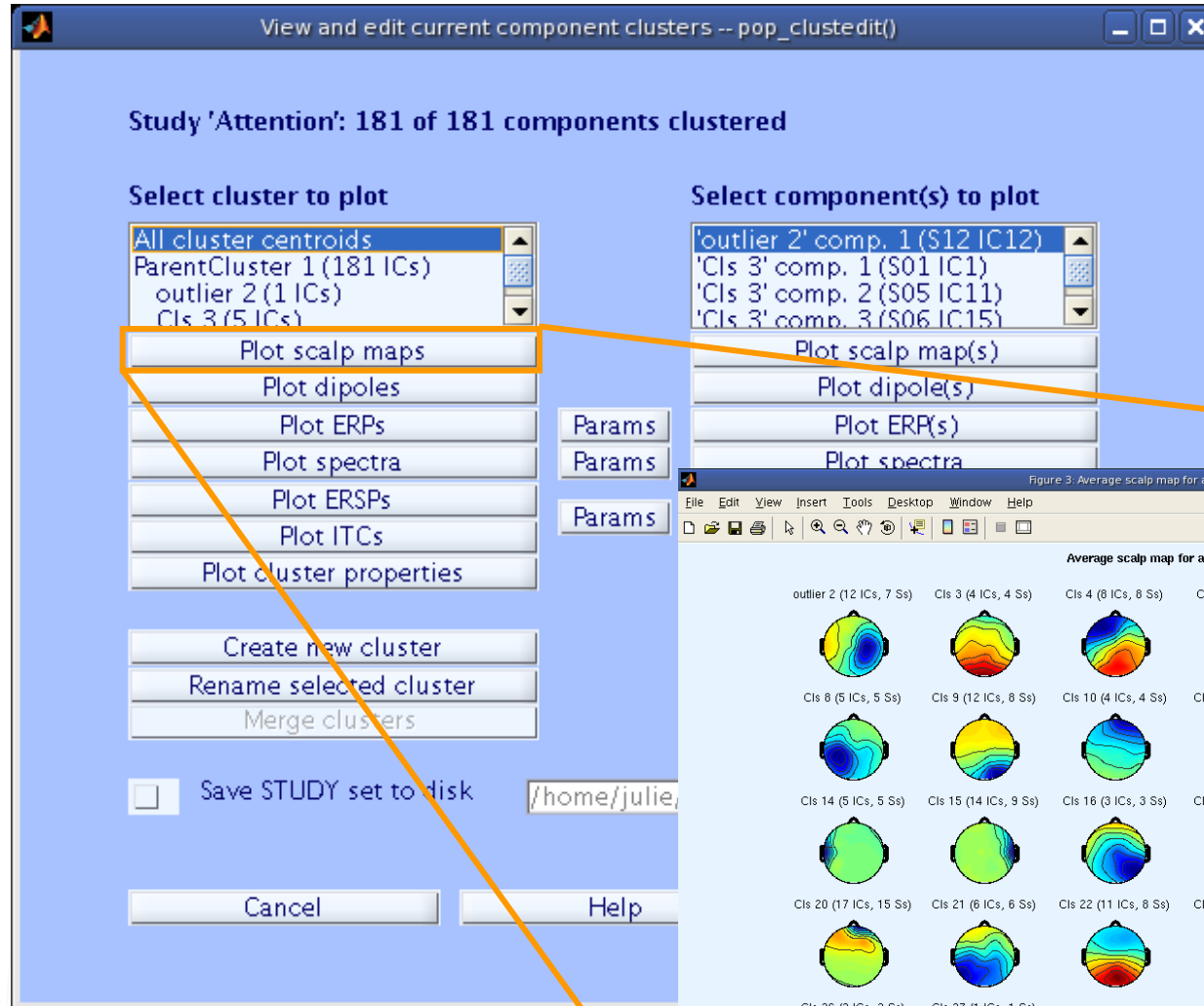
- broadly-matched ICs: use many/all of the measures.
- specifically-matched ICs: use one/few of the measures.



# Plot/edit clusters

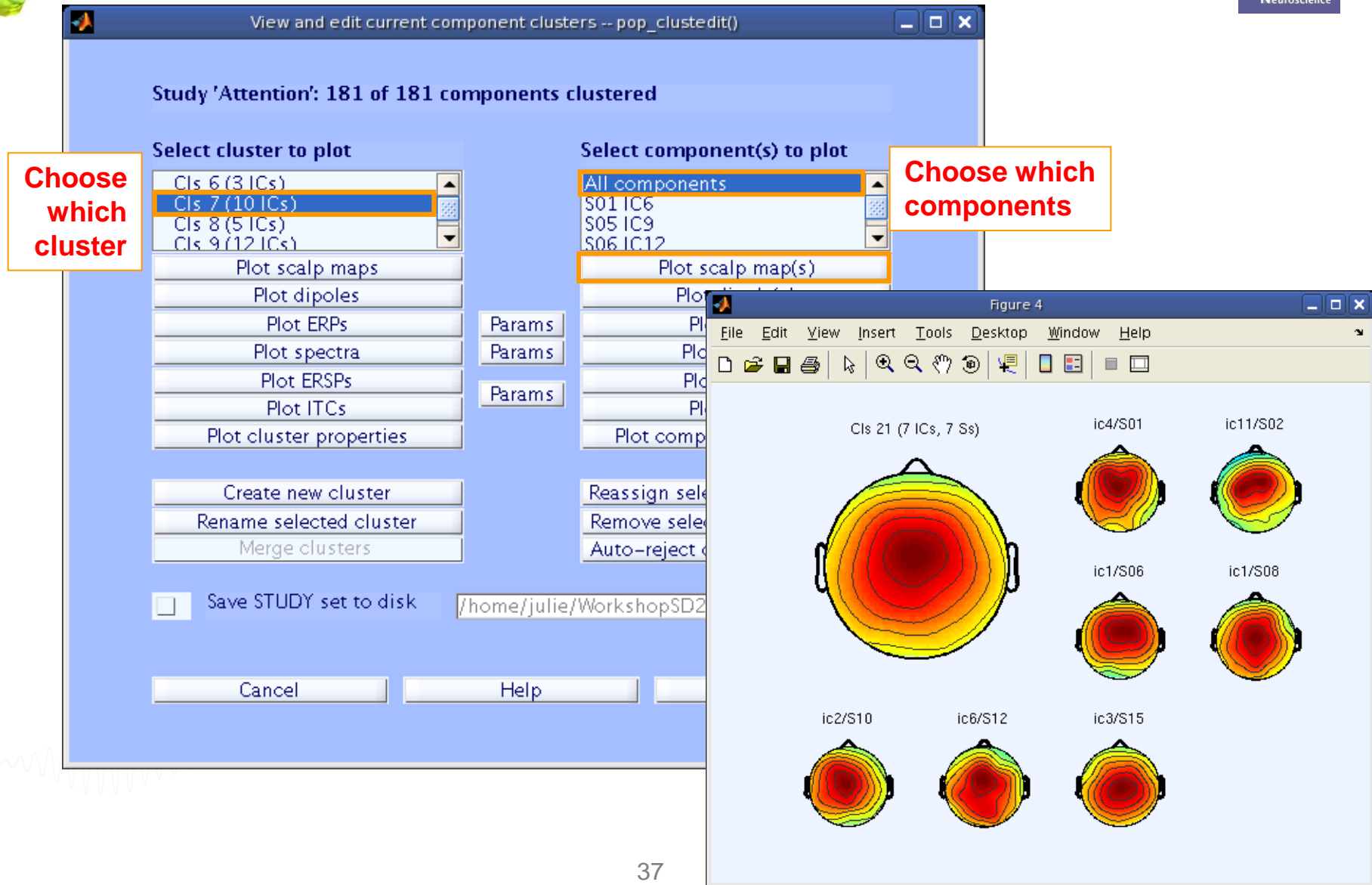
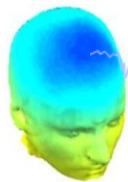


# Plot cluster data

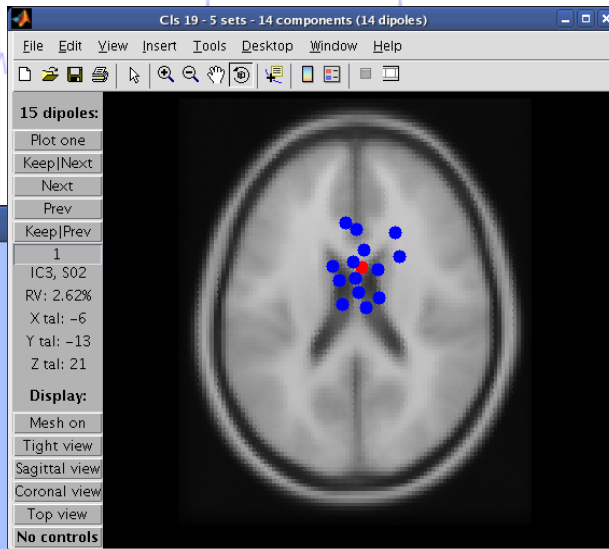


Plot mean scalp maps for easy reference

# Plot cluster data



# Plot cluster data



component clusters -- pop\_clustedit()

Select component(s) to plot

All components  
S02 IC3  
S02 IC11  
S02 IC12  
S02 IC17

Plot scalp map(s)  
**Plot dipole(s)**  
Plot ERP(s)  
Plot spectra  
Plot ERSP(s)  
Plot ITC(s)  
Plot component properties

Params  
Params  
Params

Create new cluster  
Rename selected cluster  
Merge clusters

Reassign selected component(s)  
Remove selected outlier comps.  
Auto-reject outlier components

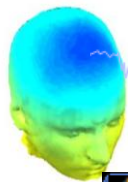
☐ Save STUDY set to disk

/home/julie/workshop06/5subjects/WSstudy.study

Cancel Help Ok



# Plot cluster data



View and edit current component clusters -- pop\_clustedit()

Study #: 151 of 151 components clustered

Select cluster to plot

- Cls 15 (8 ICs)
- Cls 16 (6 ICs)
- Cls 17 (4 ICs)
- Cls 18 (14 ICs)
- Cls 19 (14 ICs)

Plot scalp maps

Plot dipoles

Plot ERPs

Plot spectra

Params

Params

Select component(s) to plot

- All components
- S02 IC3
- S02 IC11
- S02 IC12
- S02 IC17

Plot scalp map(s)

Plot dipole(s)

Plot ERP(s)

Plot spectra

Set ERP plotting parameters -- pop\_erpparams()

Time range in ms [low high]

Plot scalp map at latency [ms]

☐ Plot conditions on the same panel

☐ Plot groups on the same panel

Statistical method to use

☐ Compute condition statistics

☐ Compute group statistics

☐ Use single trials (when available)

☐ Use False Discovery Rate to correct for multiple comparisons

Plot limits in uV [low high]

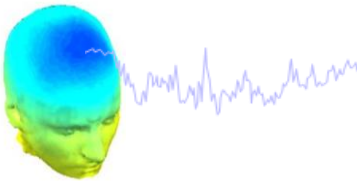
Display filter in Hz [high]

Statistical threshold (p<)

Help Cancel Ok



# Other plotting options...



Set ERP plotting parameters -- pop\_erpparams()

Time range in ms [low high]

Plot limits in uV [low high]

Plot scale map at latency [ms]  NaN

Display filter in Hz [high]

☐ Plot conditions on the same panel

☐ Plot groups on the same panel

Statistical method to use  Parametric

Statistical threshold (p<)

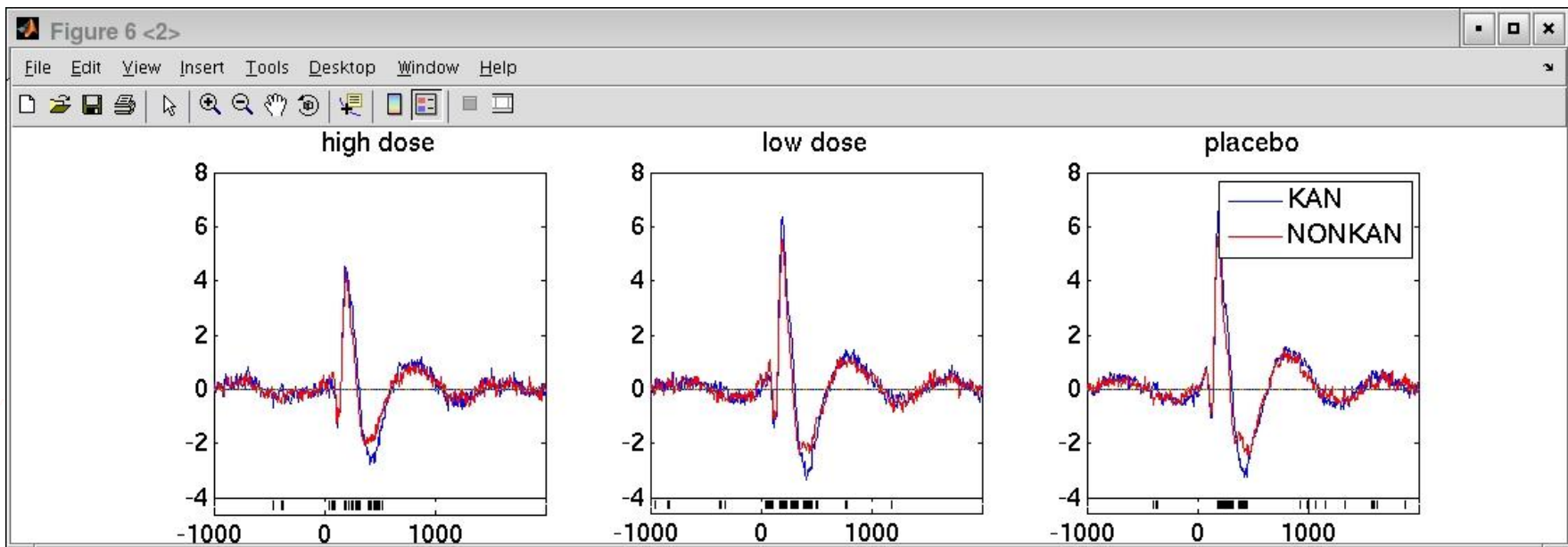
☐ Compute condition statistics

☐ Compute group statistics

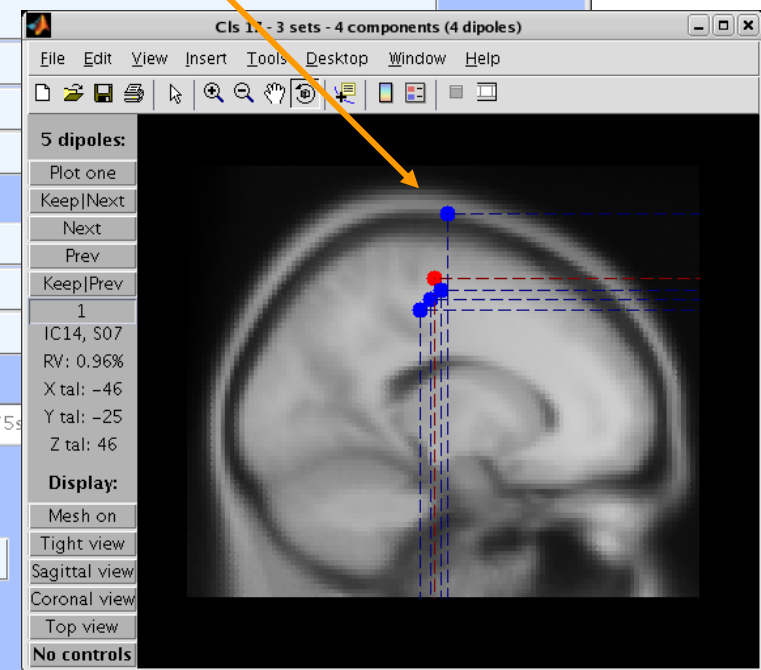
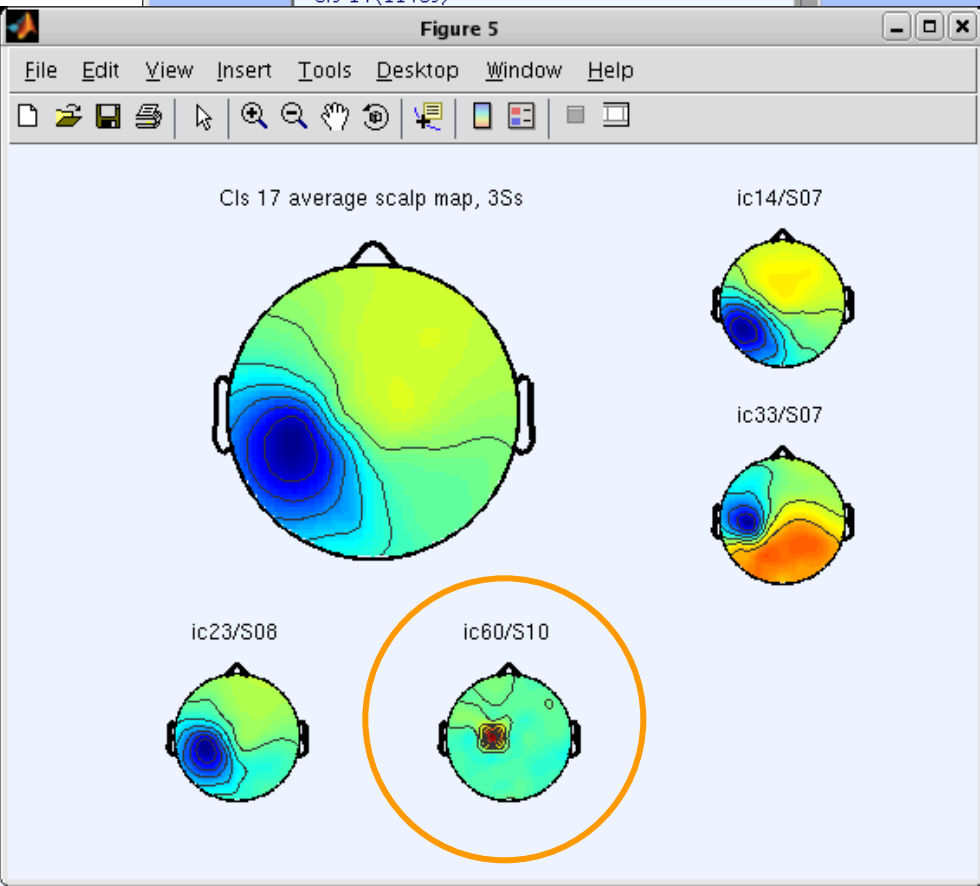
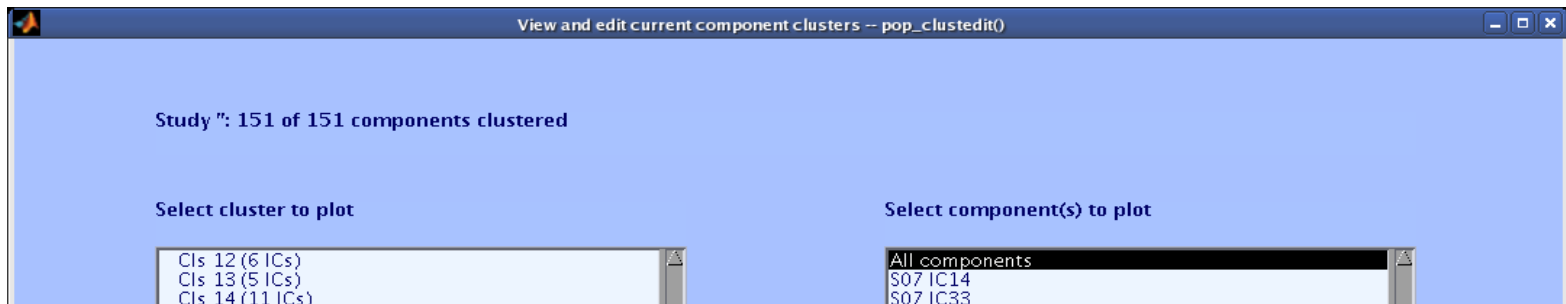
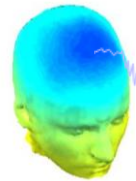
☐ Use single trials (when available)

☐ Use False Discovery Rate to correct for multiple comparisons

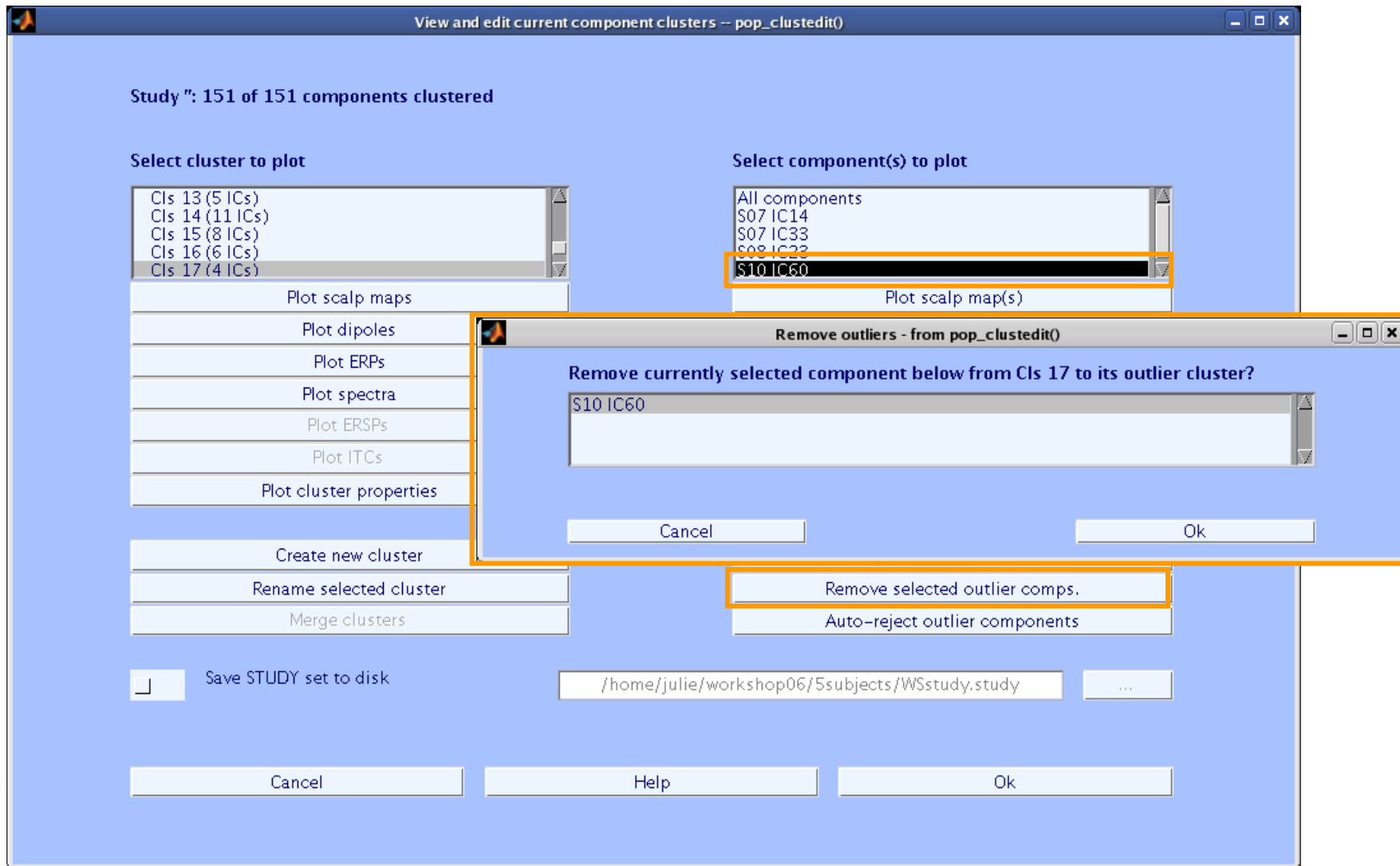
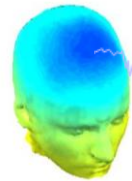
Help Cancel Ok



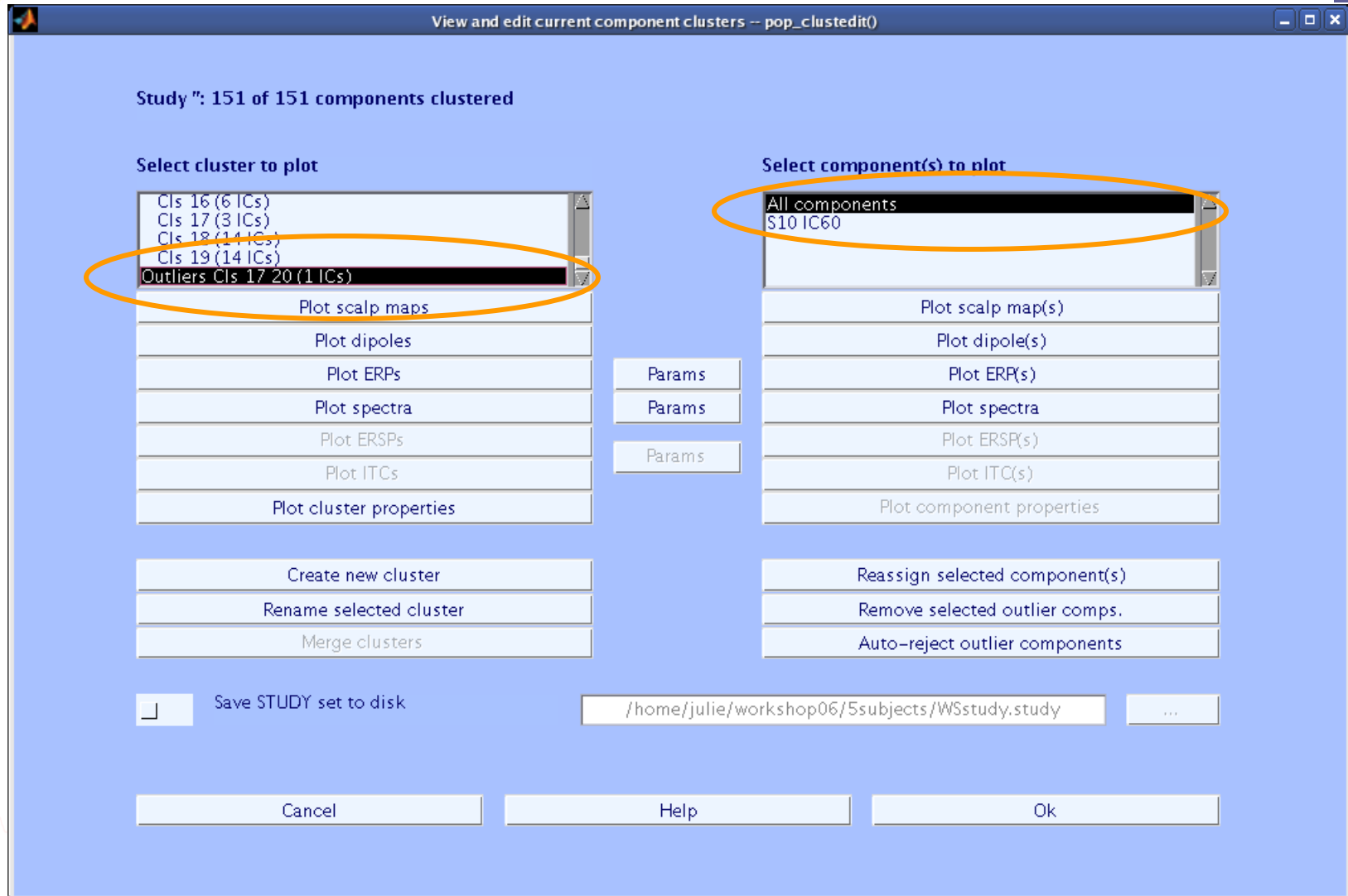
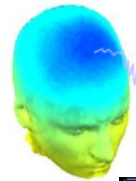
# Reassigning components



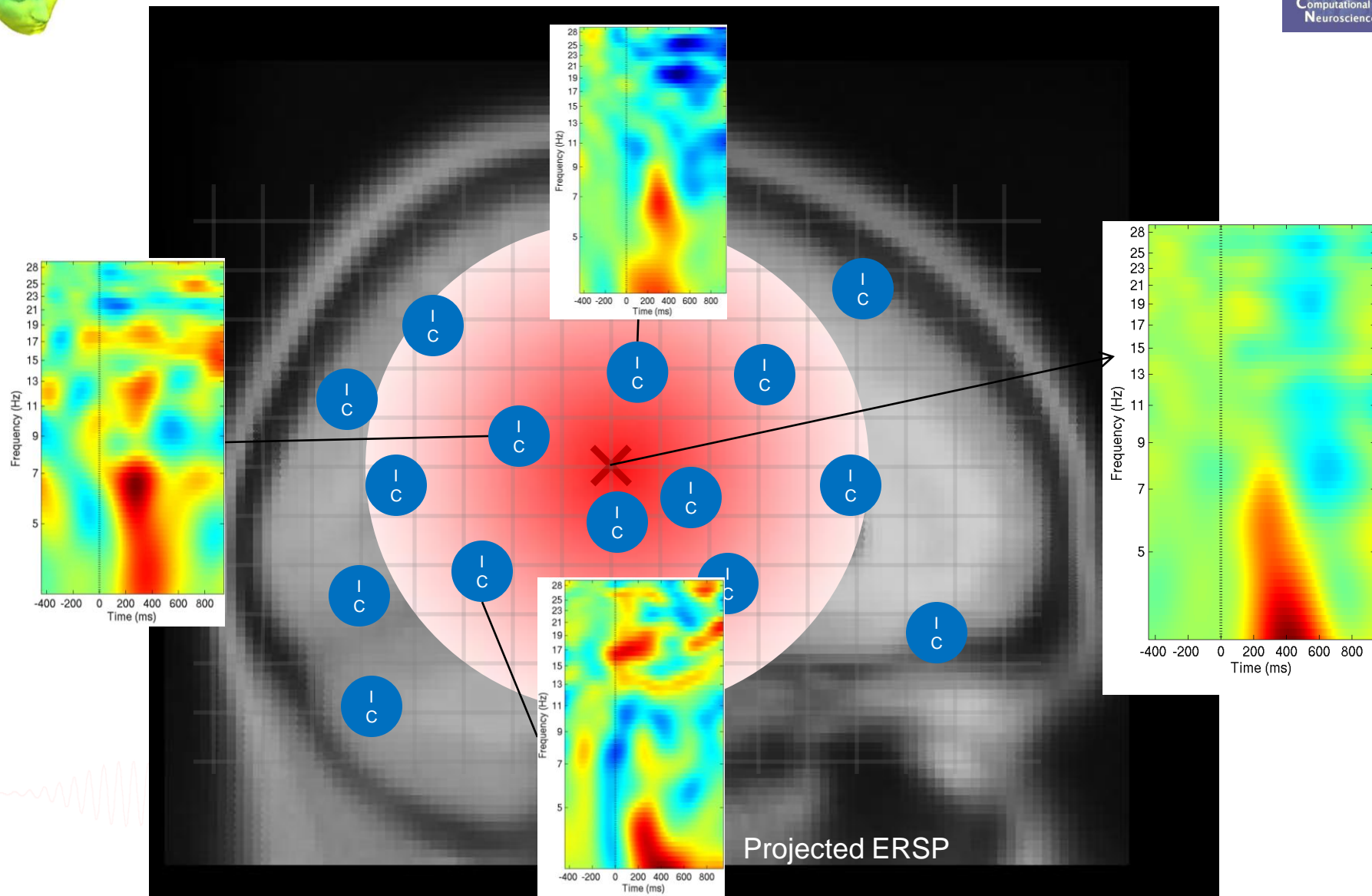
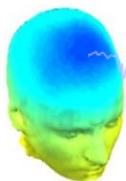
# Reassigning components



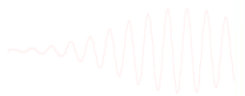
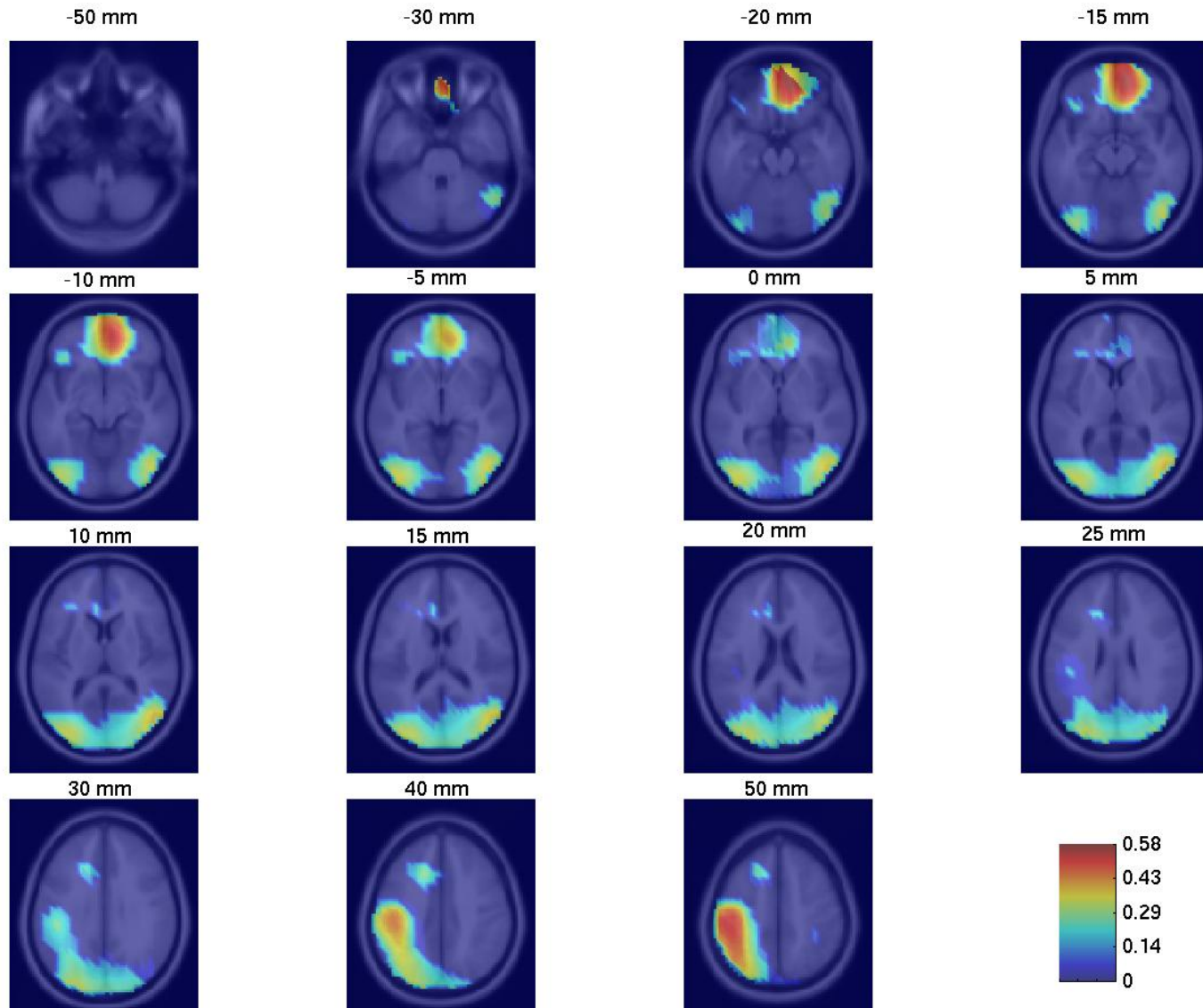
# Outlier cluster reassignment

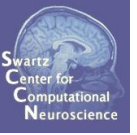


# Measure Projection Toolbox



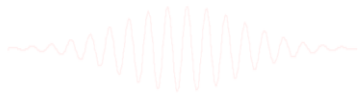
# Measure Projection Toolbox





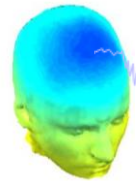
## Exercise

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





# Inter iteration Cluster Consistency



Iterations

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

Clusters