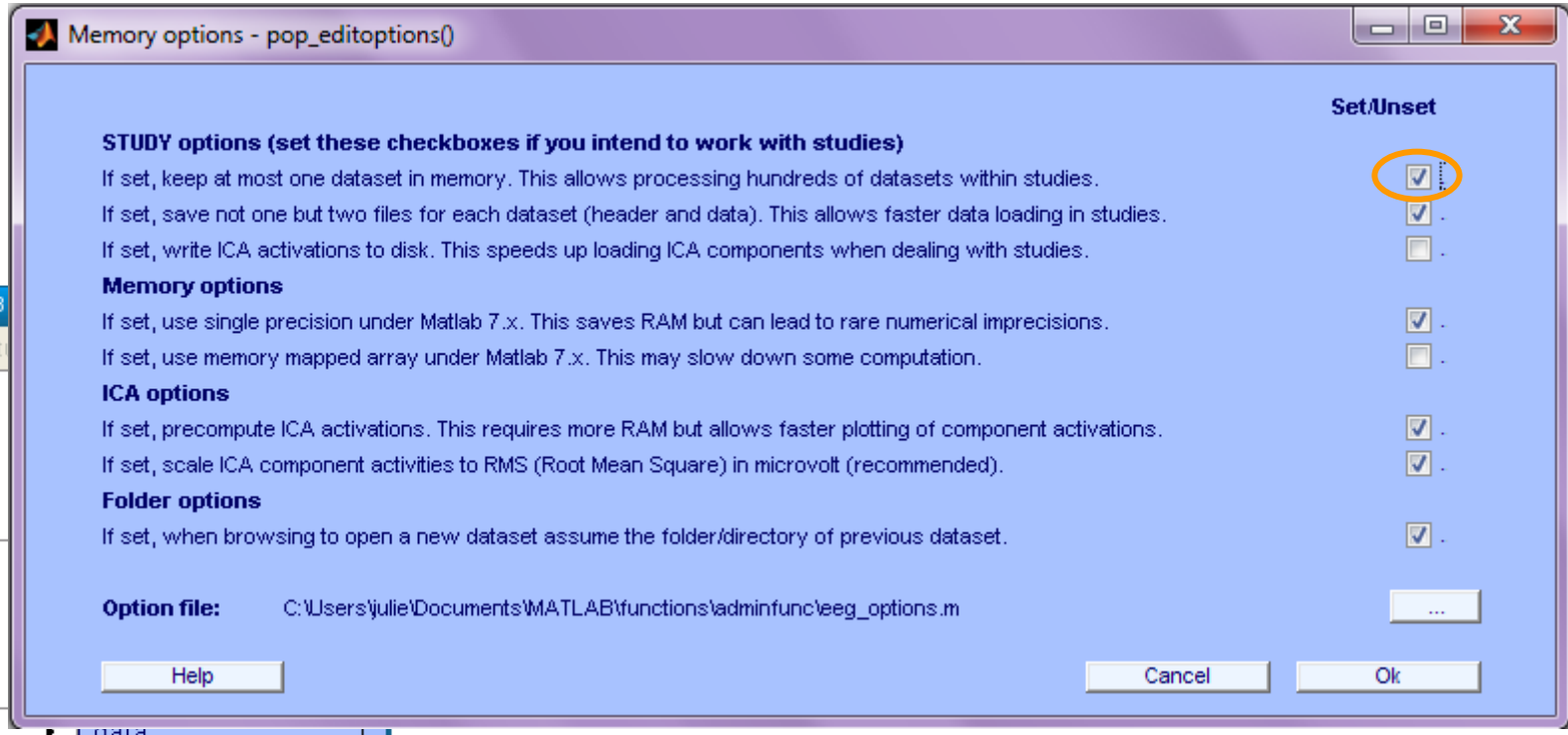
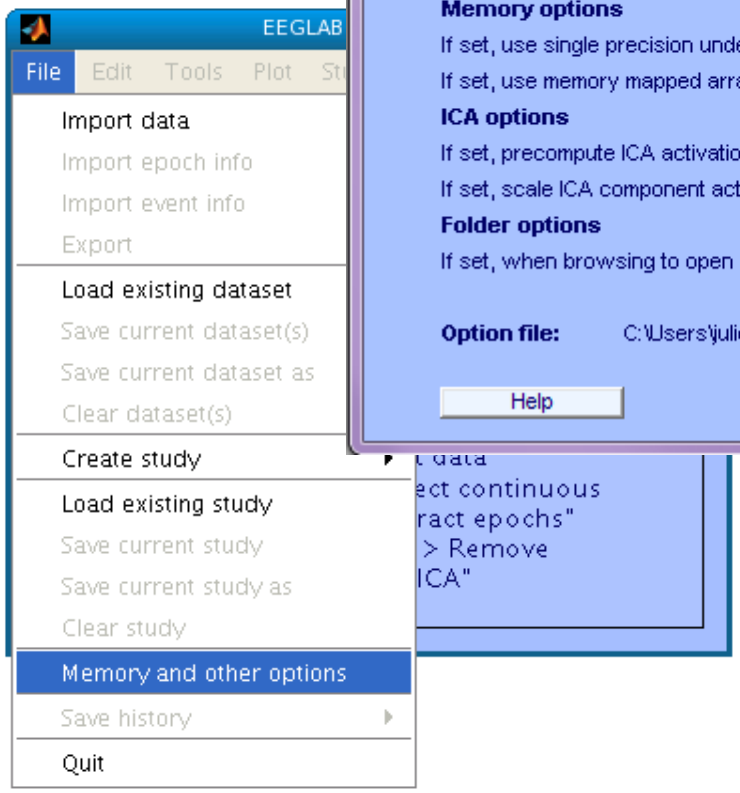
STUDY clustering overview

1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

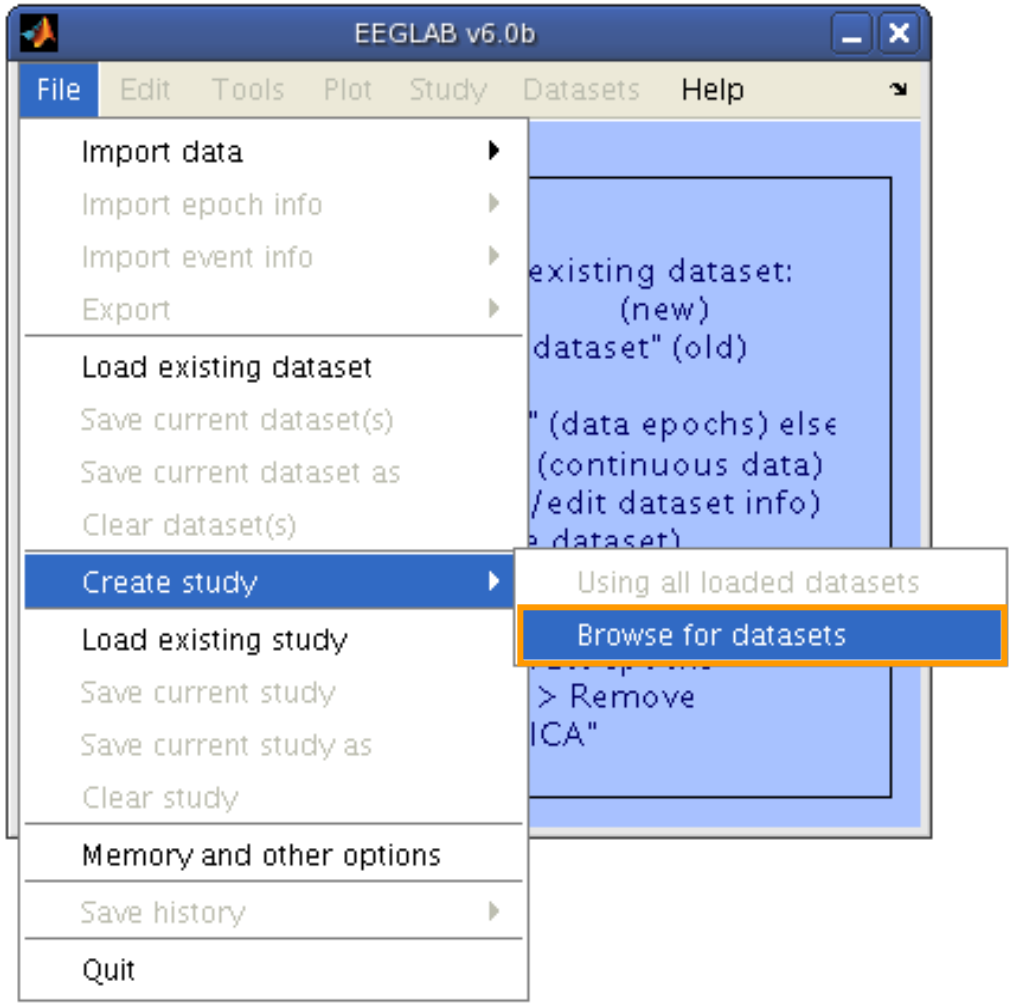
Memory options



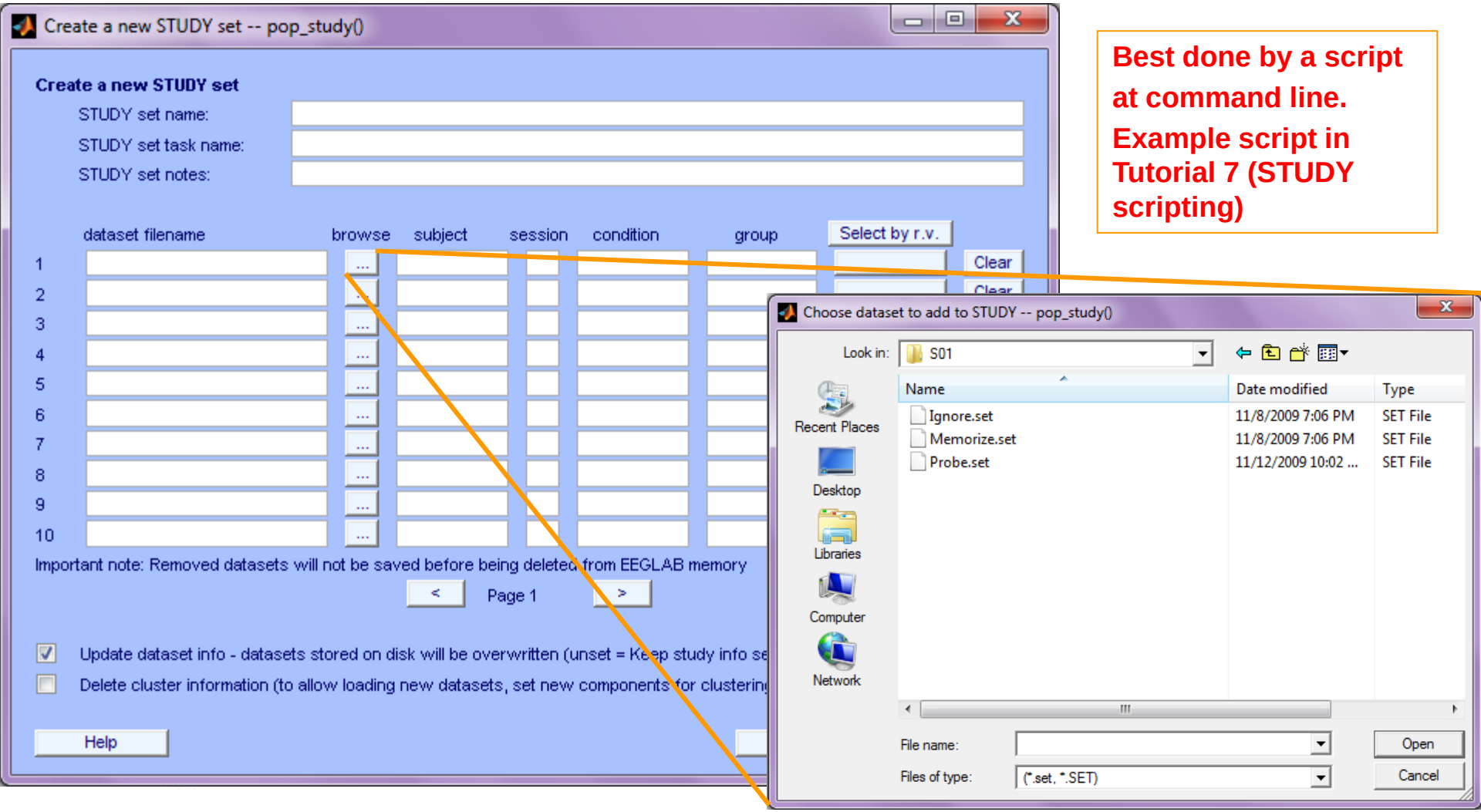


Memory options should change when using STUDY vs single dataset

Build a STUDY



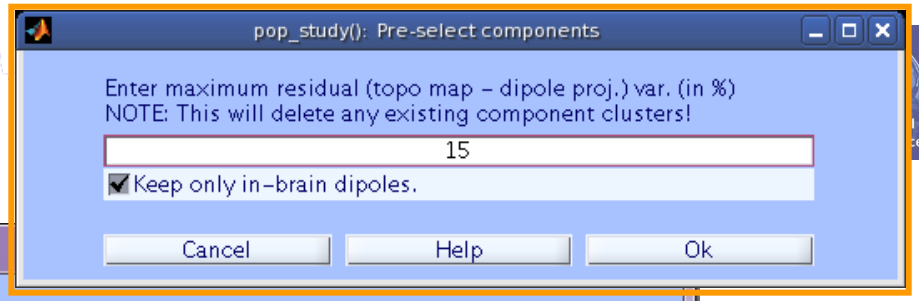
Build a STUDY, cont'd



Best done by a script at command line.

Example script in Tutorial 7 (STUDY scripting)

Edit dataset info



Create a new STUDY set -- pop_study()

Create a new STUDY set

STUDY set name: Sternberg

STUDY set task name: Sternberg

STUDY set notes:

	dataset filename	browse	subject	session	condition	group	Select by r.v.	
1	C:\Users\julie\Documents\Worl	...	S01	1	Memorize	1	All comp.	Clear
2	C:\Users\julie\Documents\Worl	...	S01	1	Ignore	1	All comp.	Clear
3	C:\Users\julie\Documents\Worl	...	S01	1	Probe	1	All comp.	Clear
4	C:\Users\julie\Documents\Worl	...	S02	1	Memorize	1	All comp.	Clear
5	C:\Users\julie\Documents\Worl	...	S02	1	Ignore	1	All comp.	Clear
6	C:\Users\julie\Documents\Worl	...	S02	1	Probe	1	All comp.	Clear
7	C:\Users\julie\Documents\Worl	...	S03	1	Memorize	1	All comp.	Clear
8	C:\Users\julie\Documents\Worl	...	S03	1	Ignore	1	All comp.	Clear
9	C:\Users\julie\Documents\Worl	...	S03	1	Probe	1	All comp.	Clear
10	C:\Users\julie\Documents\Worl	...	S04	1	Memorize	1	All comp.	Clear

Important note: Removed datasets will not be saved before being deleted from EEGLAB memory

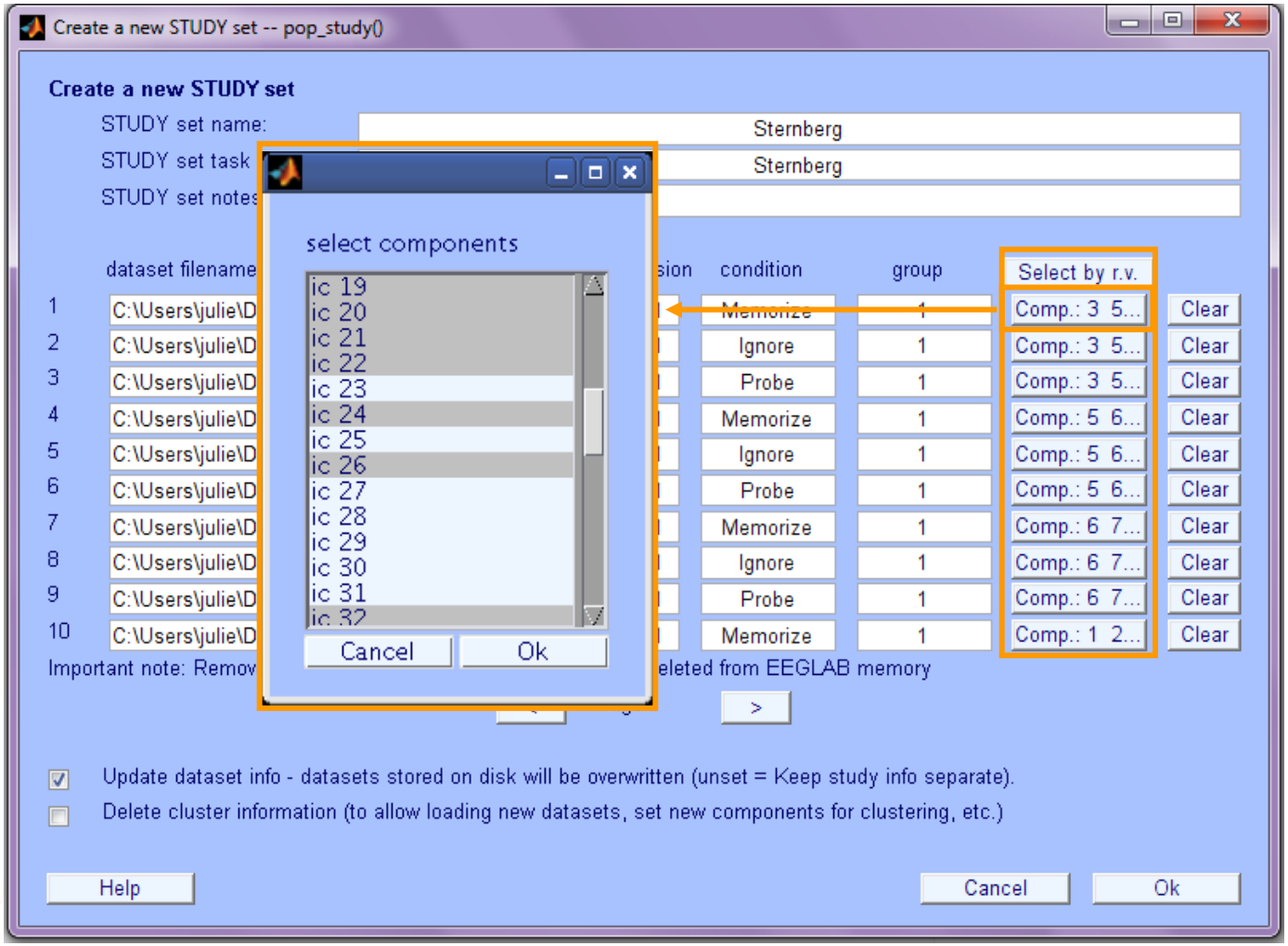
< Page 1 >

☑ Update dataset info - datasets stored on disk will be overwritten (unset = Keep study info separate).

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

Help Cancel Ok

ICs to cluster



STUDY structure

```
STUDY =
            name: 'Sternberg'
            task: 'Sternberg'
     datasetinfo: [1x39 struct]
           notes: ''
        filename: 'stern.study'
        filepath: 'C:\Users\julie\Documents\Workshops\Mallorca\STUDY'
         history: [1x7332 char]
         subject: {1x13 cell}
           group: {''}
         session: []
       condition: {'ignore'  'memorize'  'probe'}
          setind: [3x13 double]
             etc: [1x1 struct]
        preclust: [1x1 struct]
         cluster: [1x1 struct]
         changrp: [1x71 struct]
           saved: 'yes'
>>
```

More STUDY structure details in Tutorial 7

STUDY clustering overview

1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

Experimental design

1x2 unpaired

Patients	Controls
Group A	Group B

2x2 unpaired

	Patients	Controls
Old	Group A	Group B
Young	Group C	Group D

1x2 paired

Stim A	Stim B

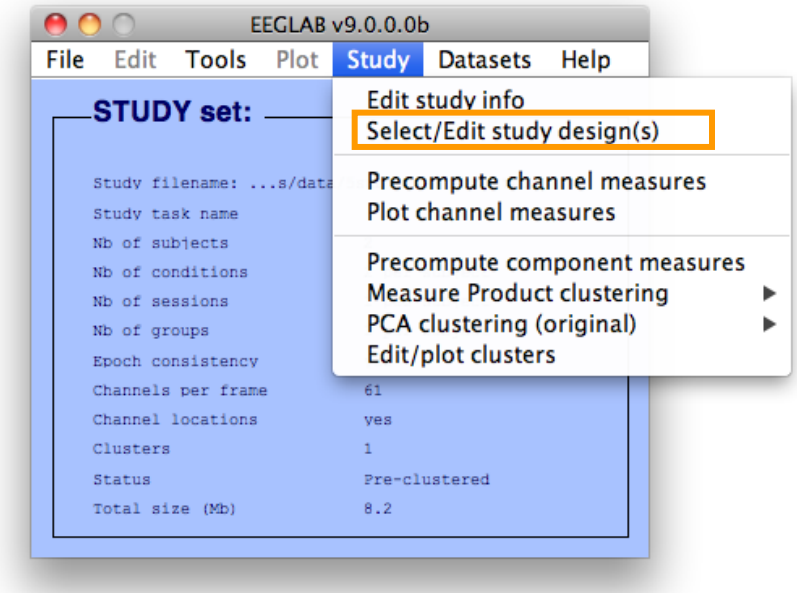
2x2 paired

	Stim A	Stim B
Drug A		
Drug B		

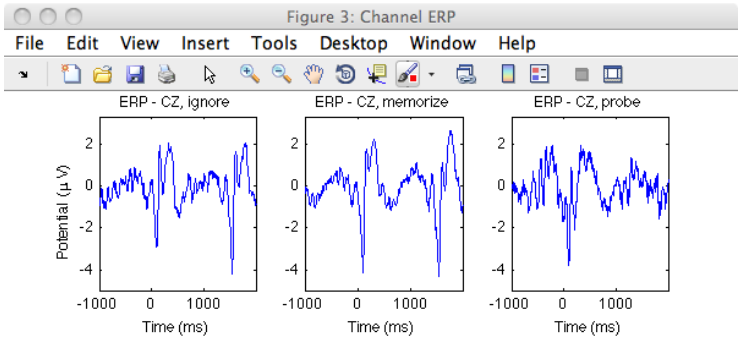
2x2 paired & unpaired

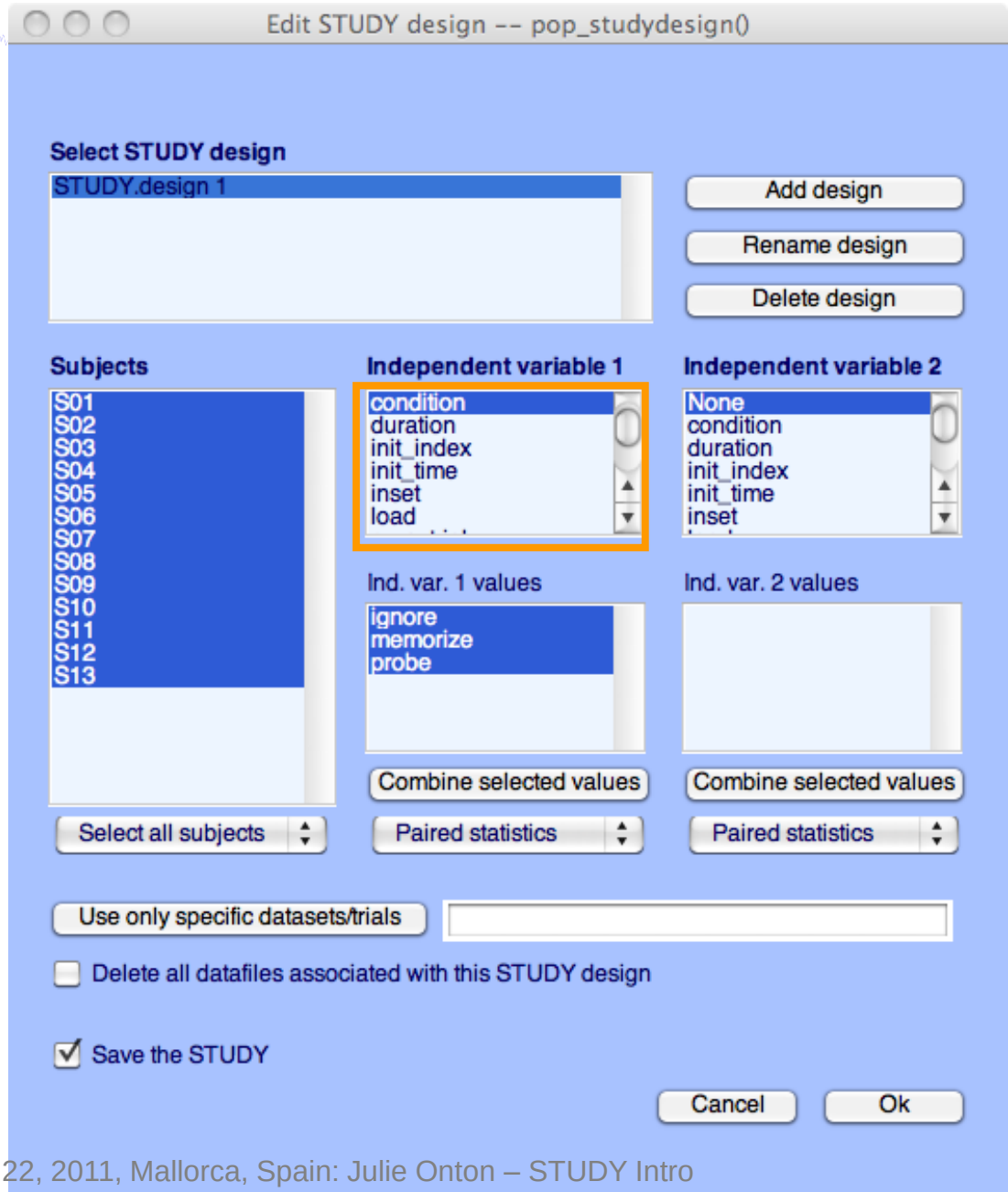
	Patients	Controls
Drug A		
Drug B		

Create design

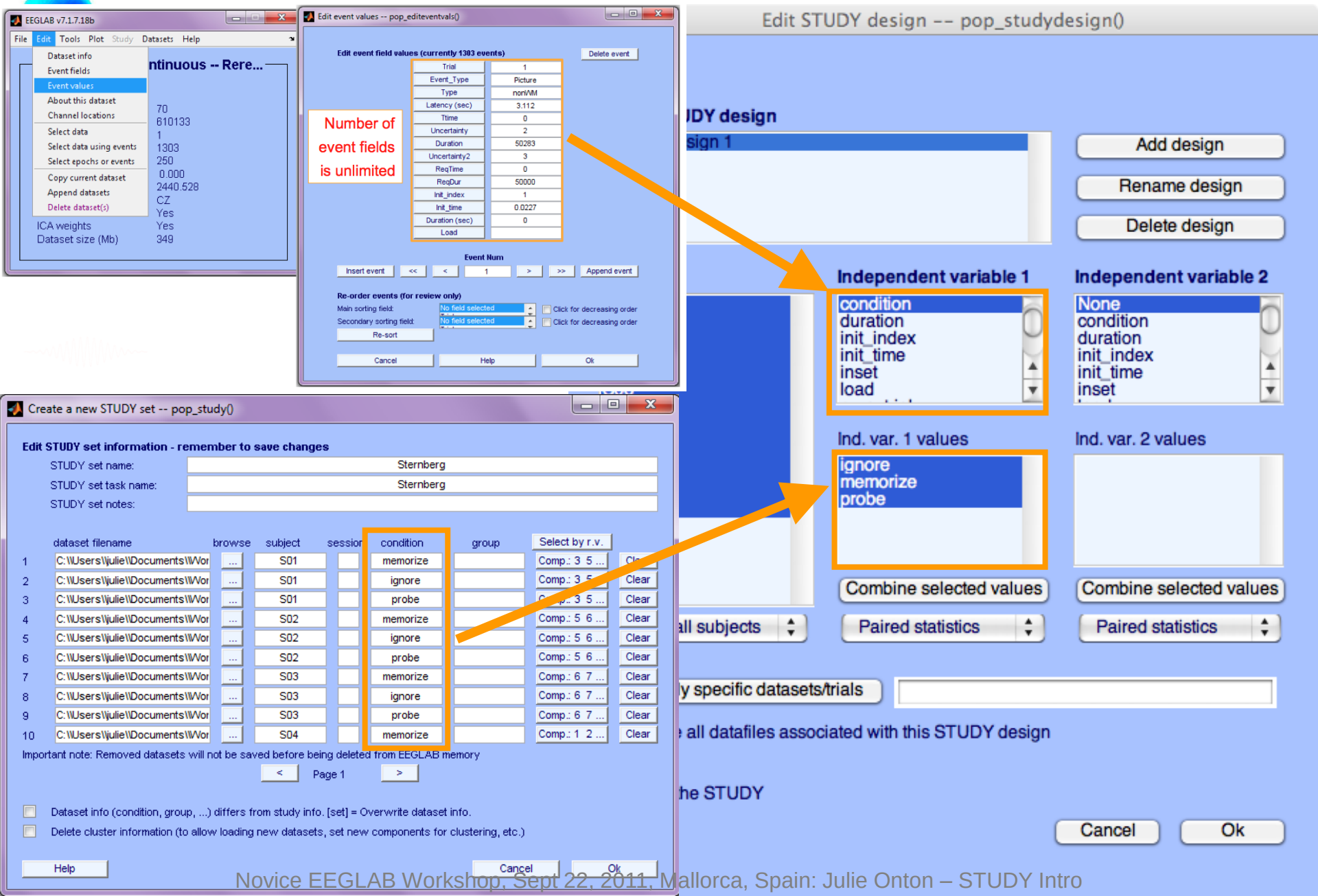


1x3 design

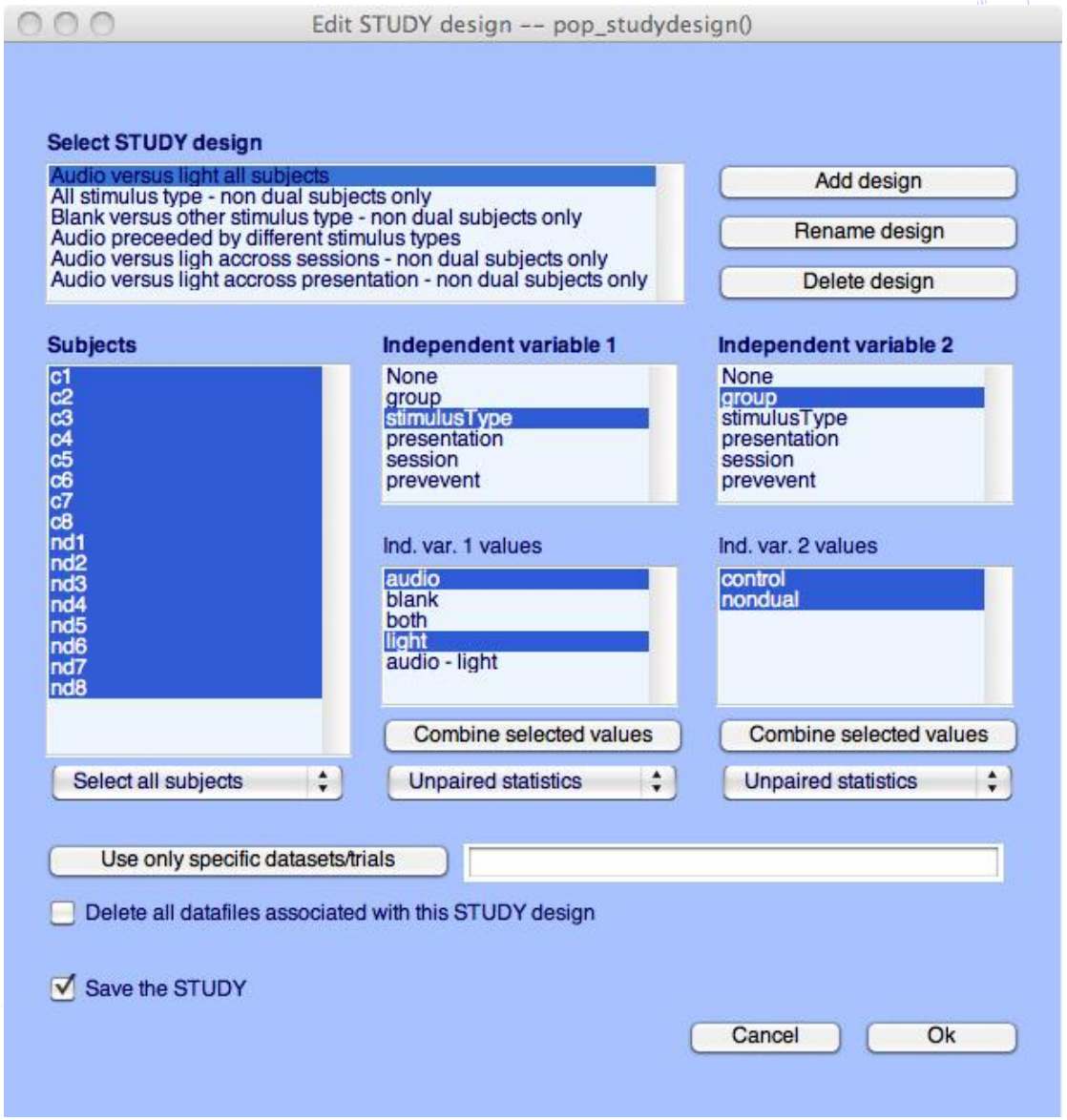




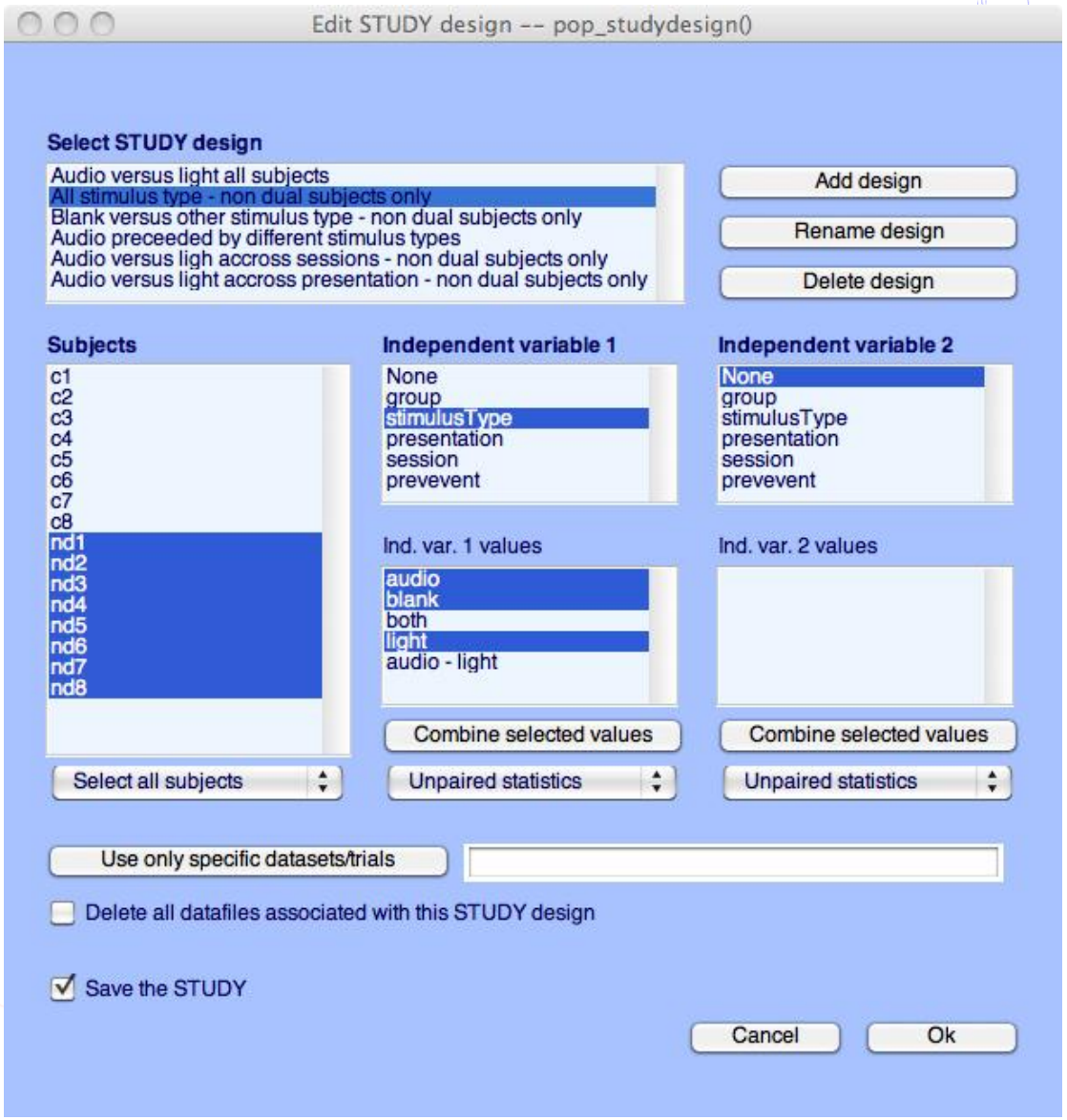
Create design



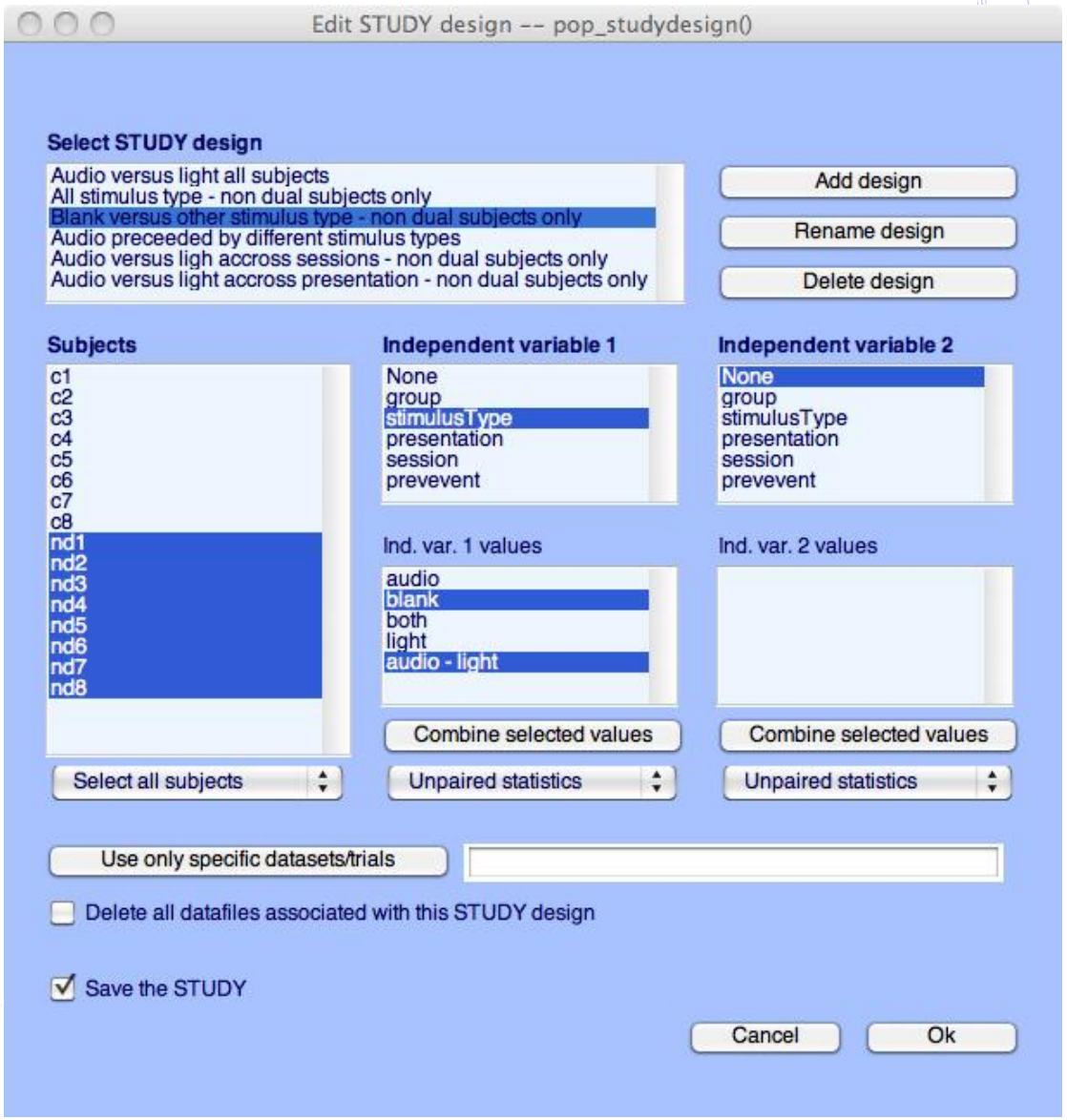
STUDY design examples



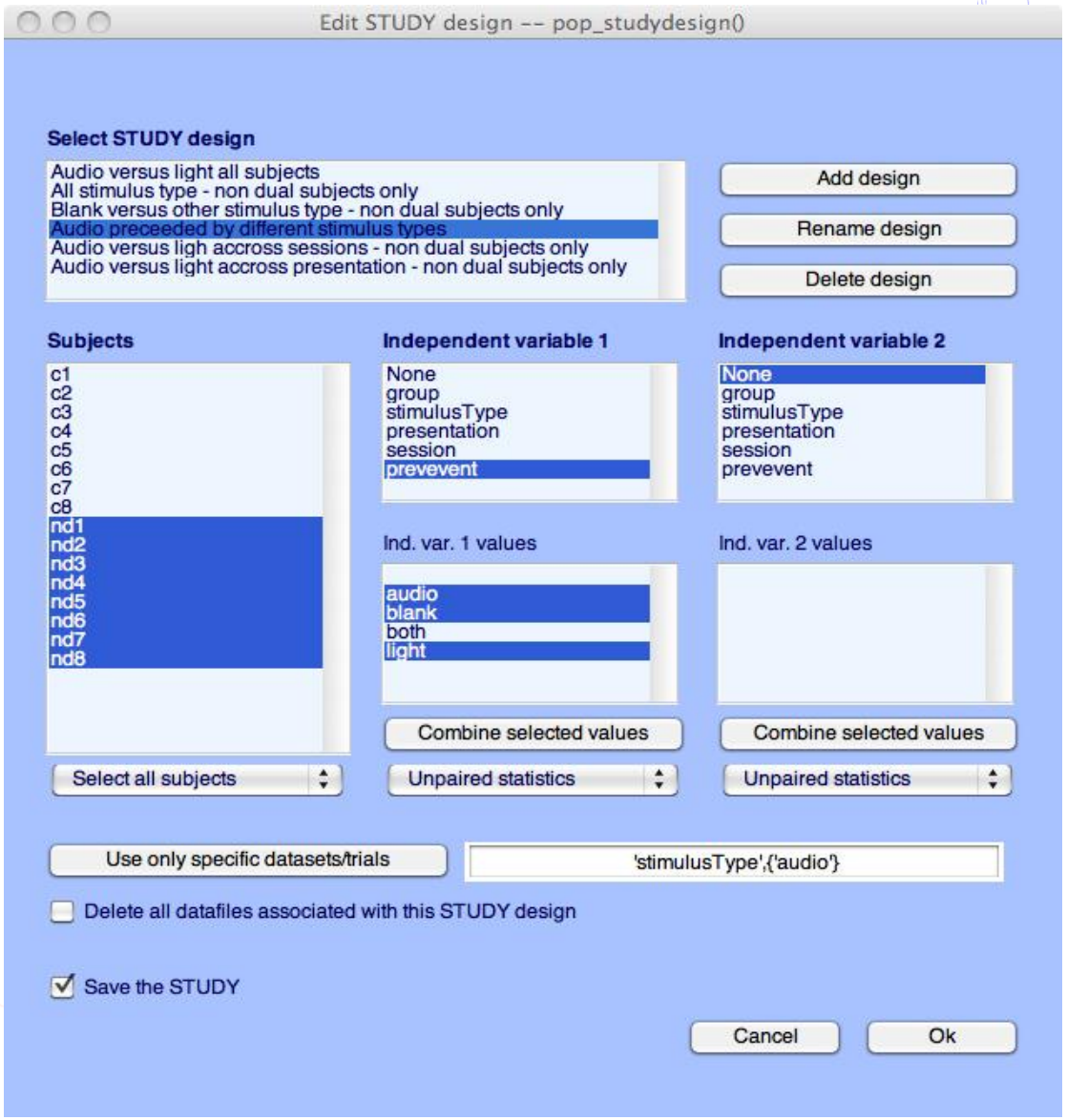
STUDY design examples



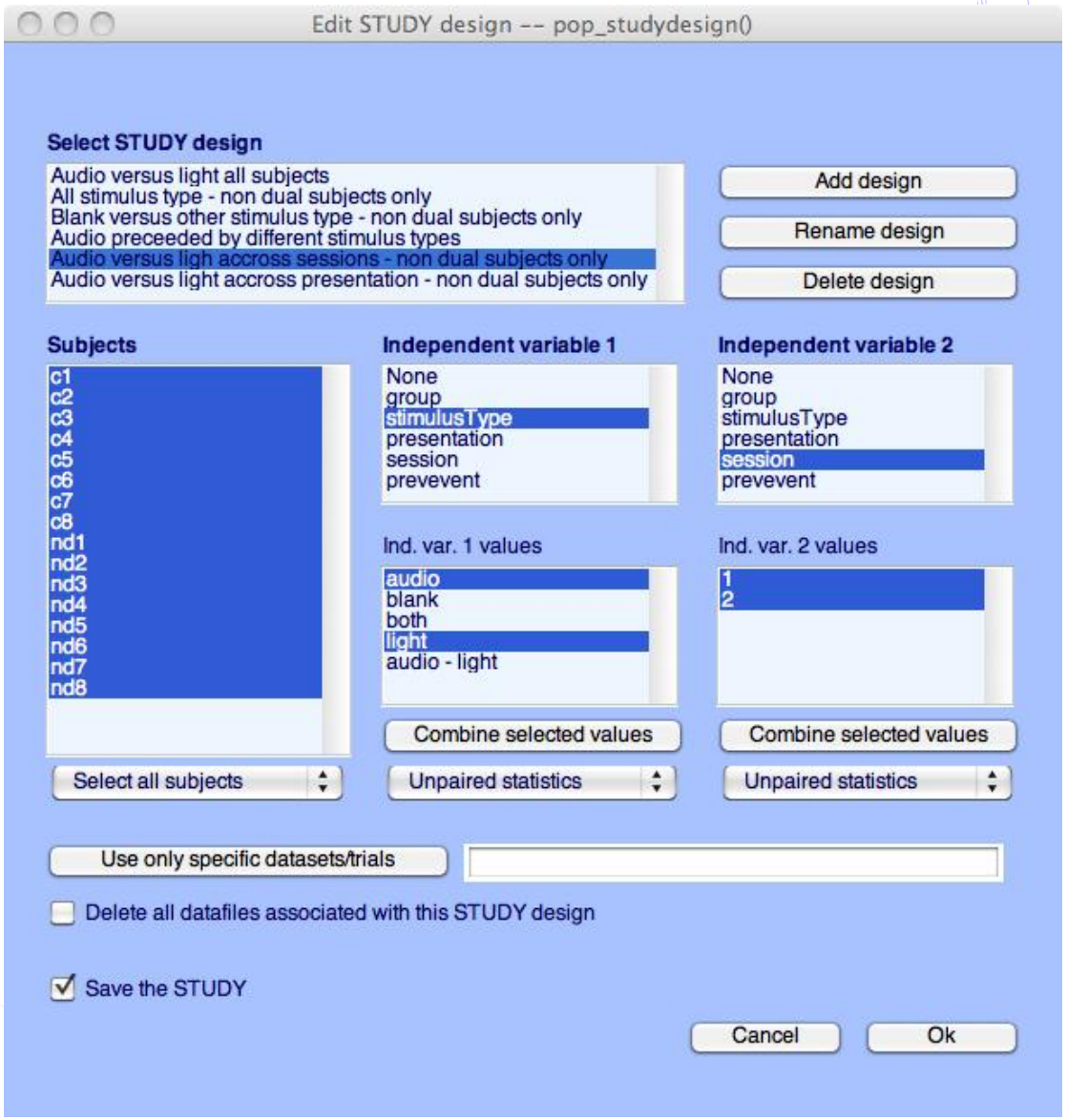
STUDY design examples



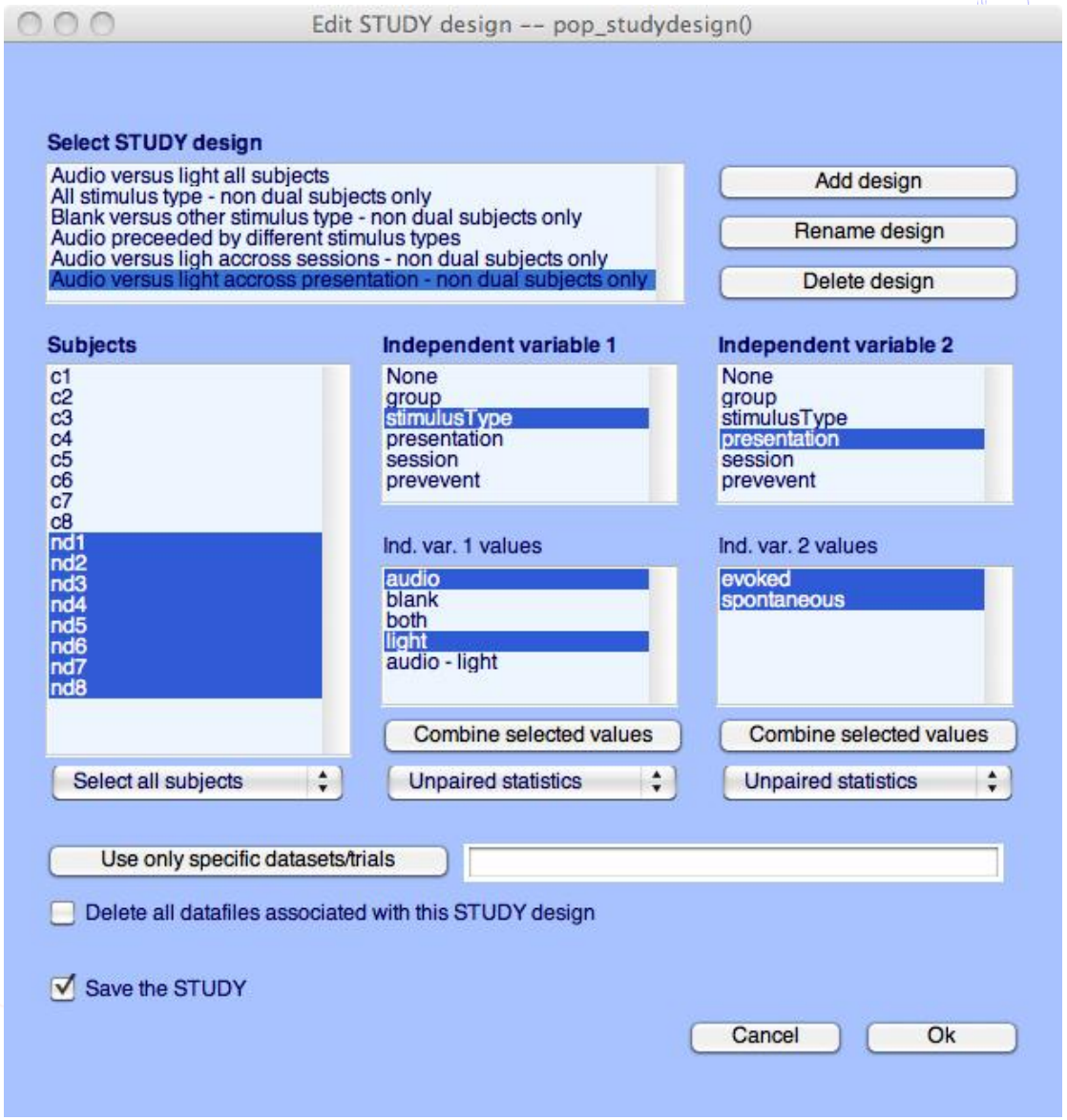
STUDY design examples



STUDY design examples



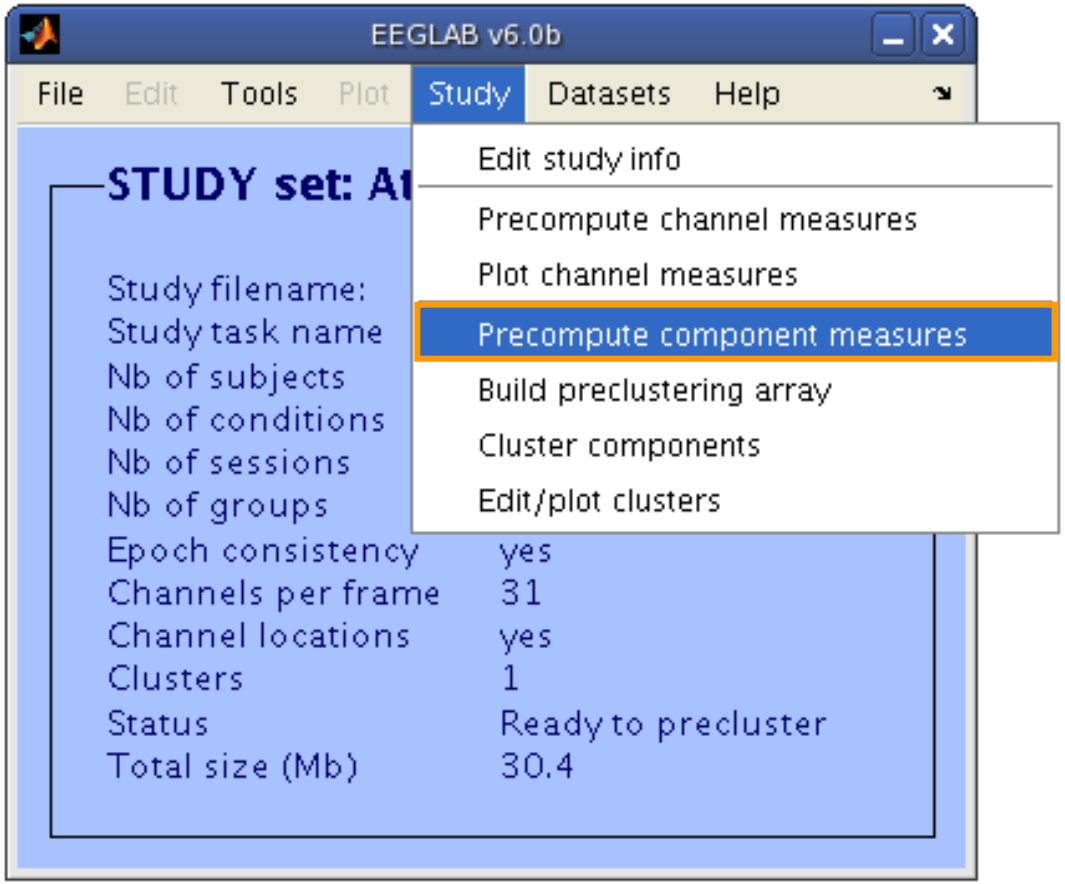
STUDY design examples



STUDY clustering overview

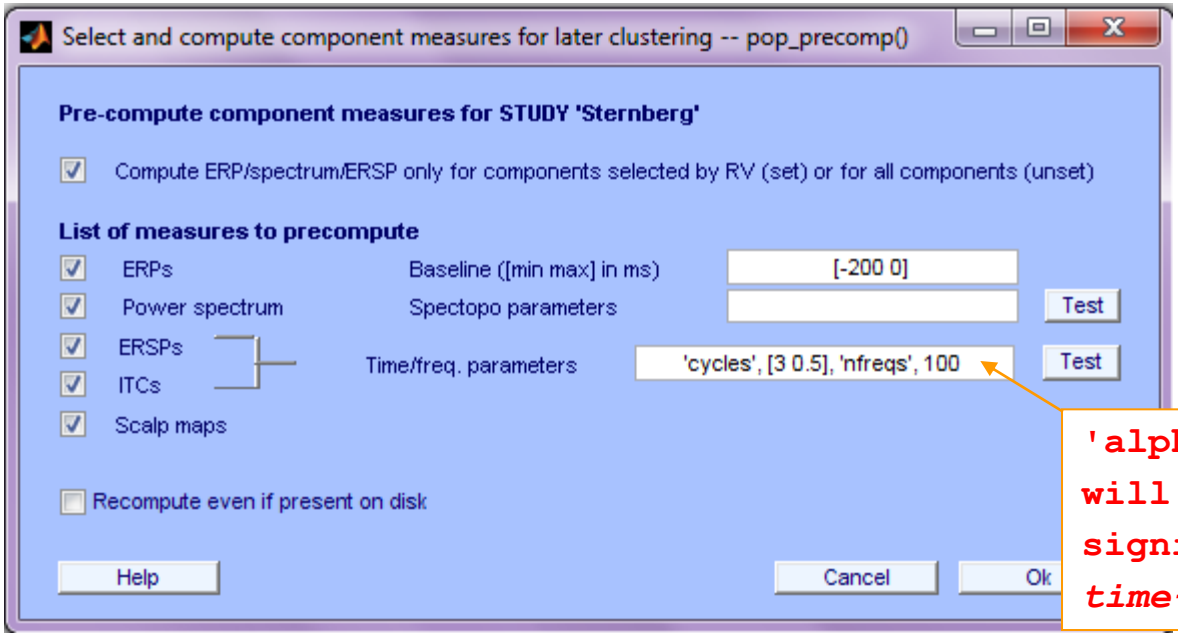
1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

Precompute data measures



Precompute data measures

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



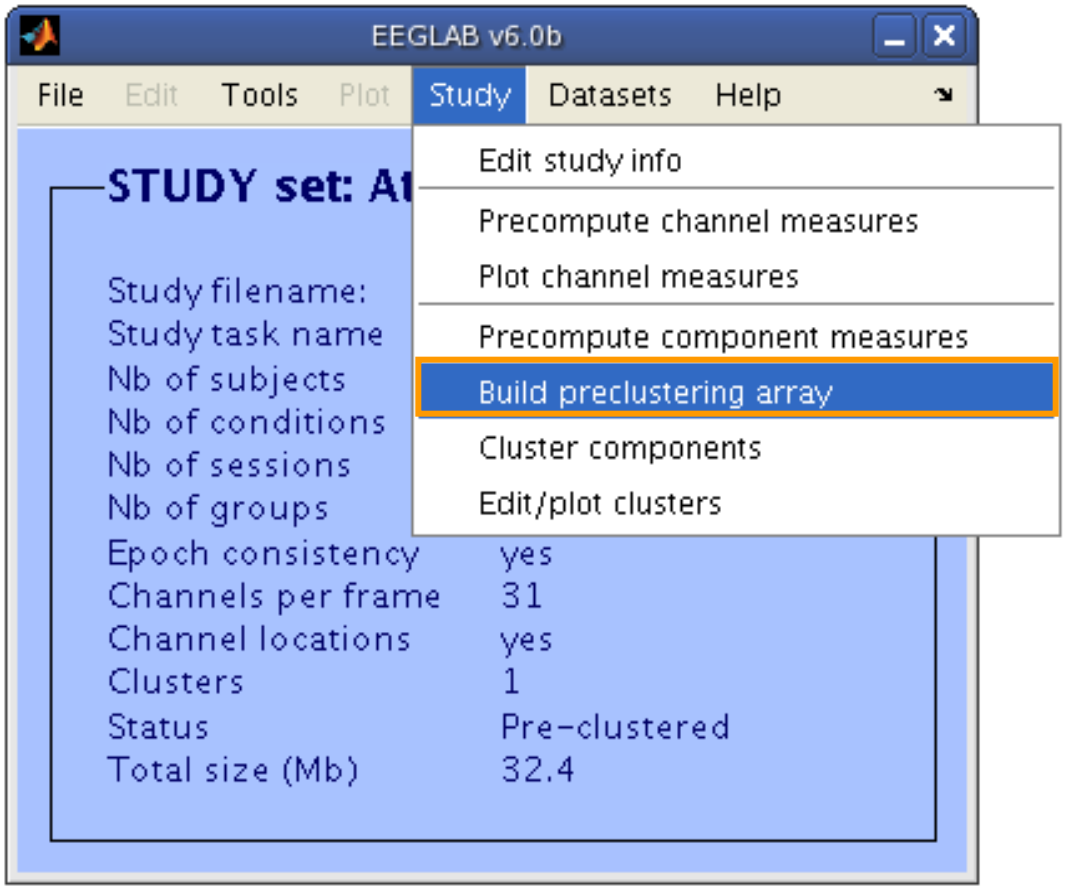
`'alpha',.01` will compute significance, but *time-consuming*

```
[STUDY ALLEEG] = std_precomp(STUDY, ALLEEG,  'components',…
'erp', 'on', 'rmbase',[-200 0] , 'scalp', 'on', 'spec',…
'on', 'specparams',{}, 'ersp', 'on', 'erspparams',…
{'cycles',[3 0.5] , 'nfreqs',100, 'freqs',[3 70] ,…
'alpha',0.01}, 'itc', 'on');
```

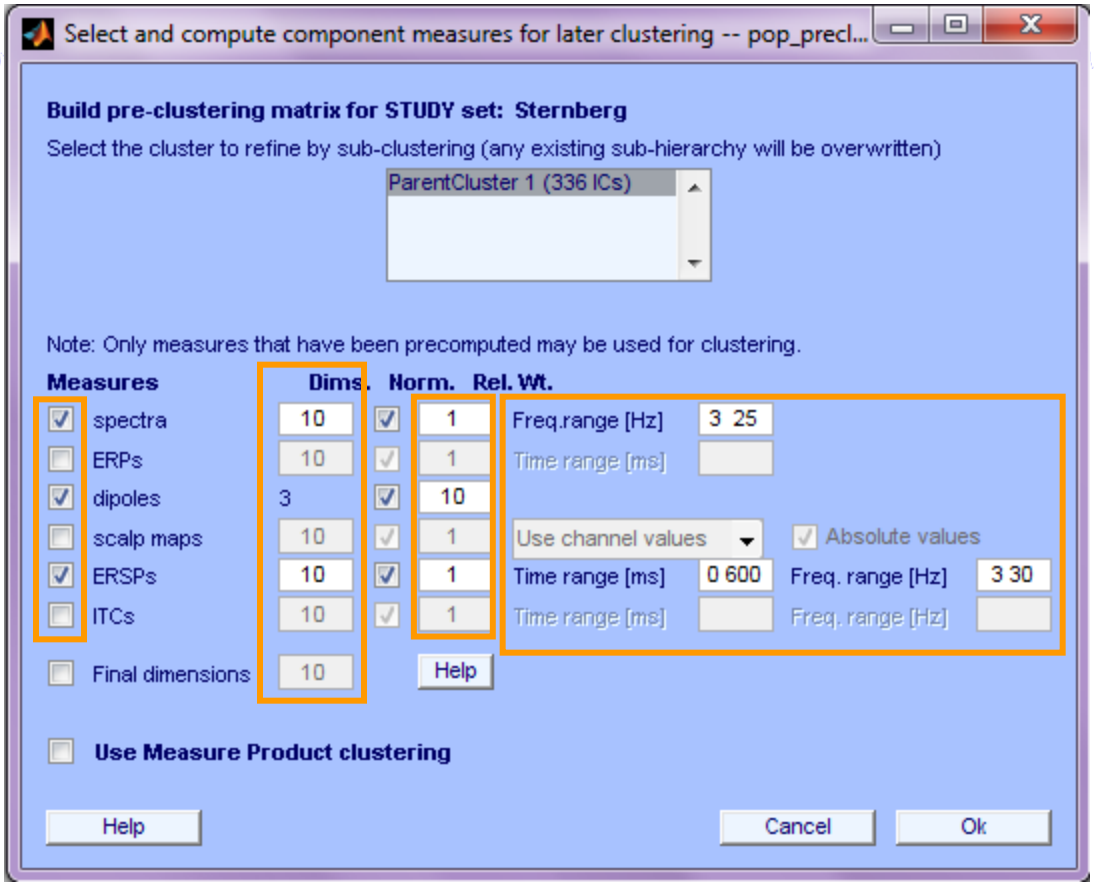
STUDY clustering overview

1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

Precluster the data



Precluster the data




```
parentclust = 1; % cluster 1 is always full parent cluster
[STUDY ALLEEG] = std_preclust(STUDY, ALLEEG, parentclust,{'spec','npca',5,…
   'norm',1,'weight',1,'freqrange',[3 25]},{'erp','npca',6,'norm',1,…
   'weight',1, 'timewindow',[0 400]},{'scalp','npca',10,'norm',1,'weight',1,…
   'abso',1},{'dipoles','norm',1,'weight',10},{'ersp','npca',20,…
   'freqrange',[3 30] ,'timewindow',[0 600],'norm',1,'weight',1},{'itc',…
   'npca',6,'freqrange',[3 30],'timewindow',[0 400] , 'norm',1, 'weight',1});
```

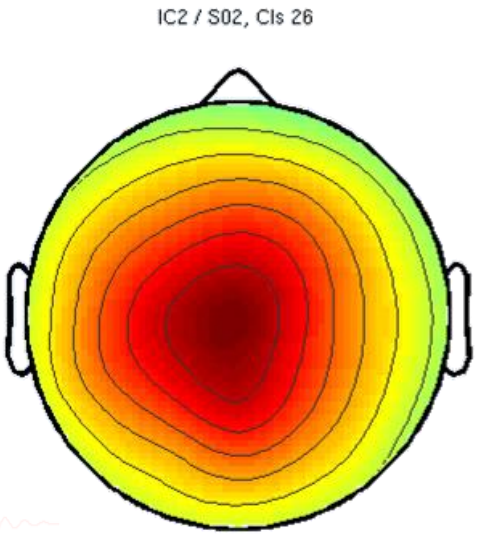
Choosing data measures

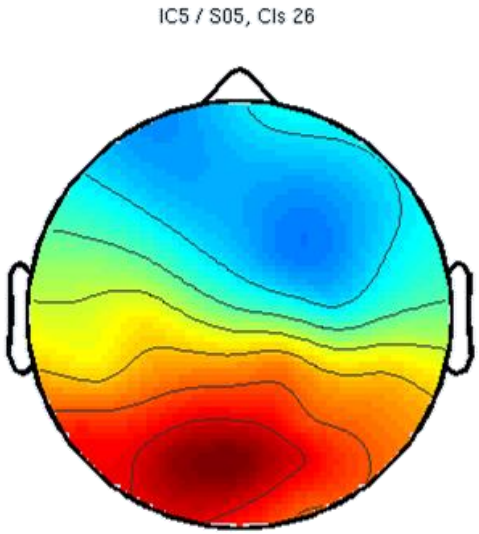
What measure(s) should you use?

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

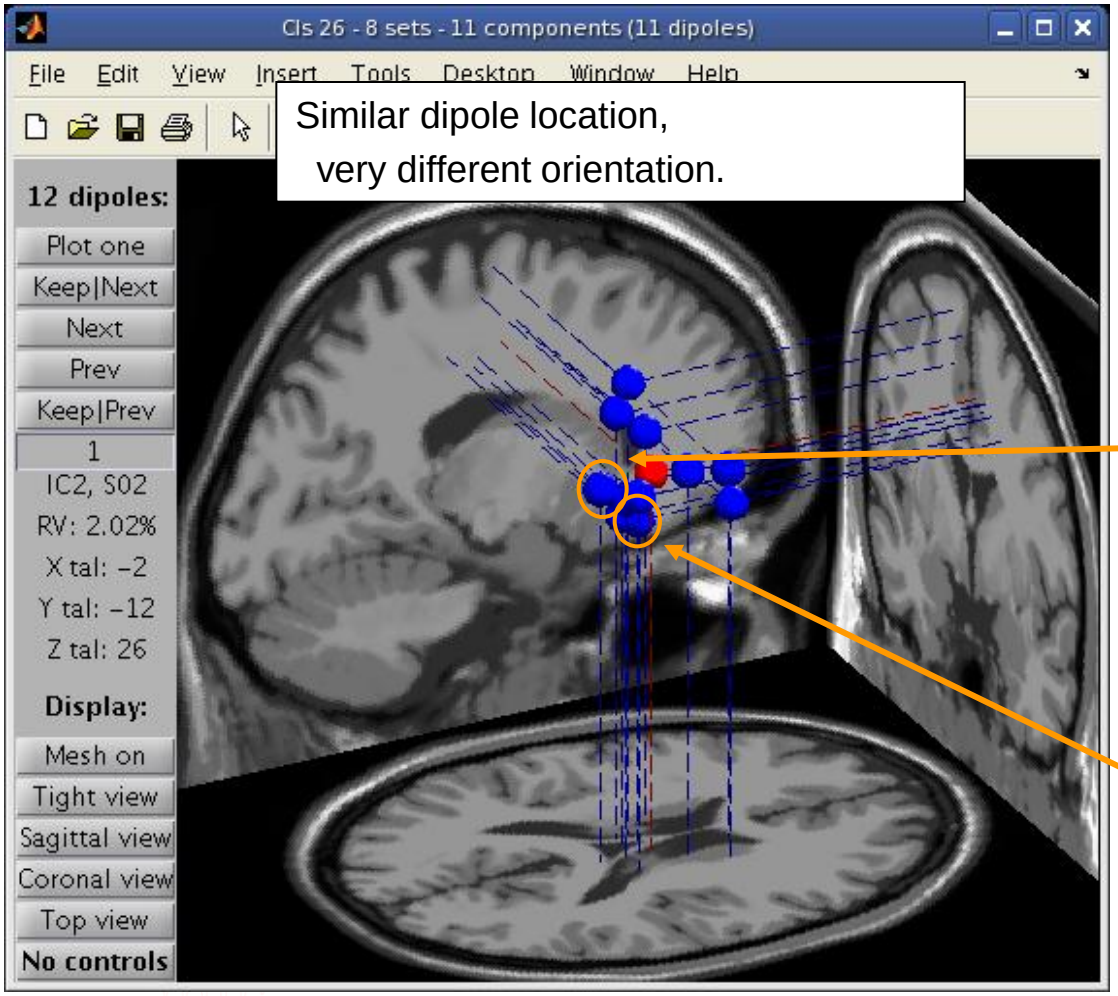
But consider:

- What is the difference between these two components?



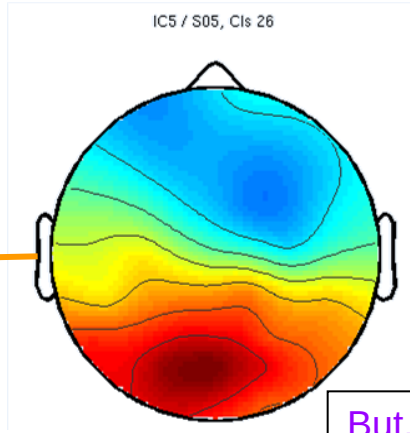


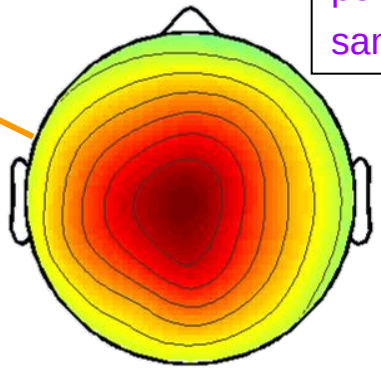
Choosing data measures



Similar dipole location,
very different orientation.

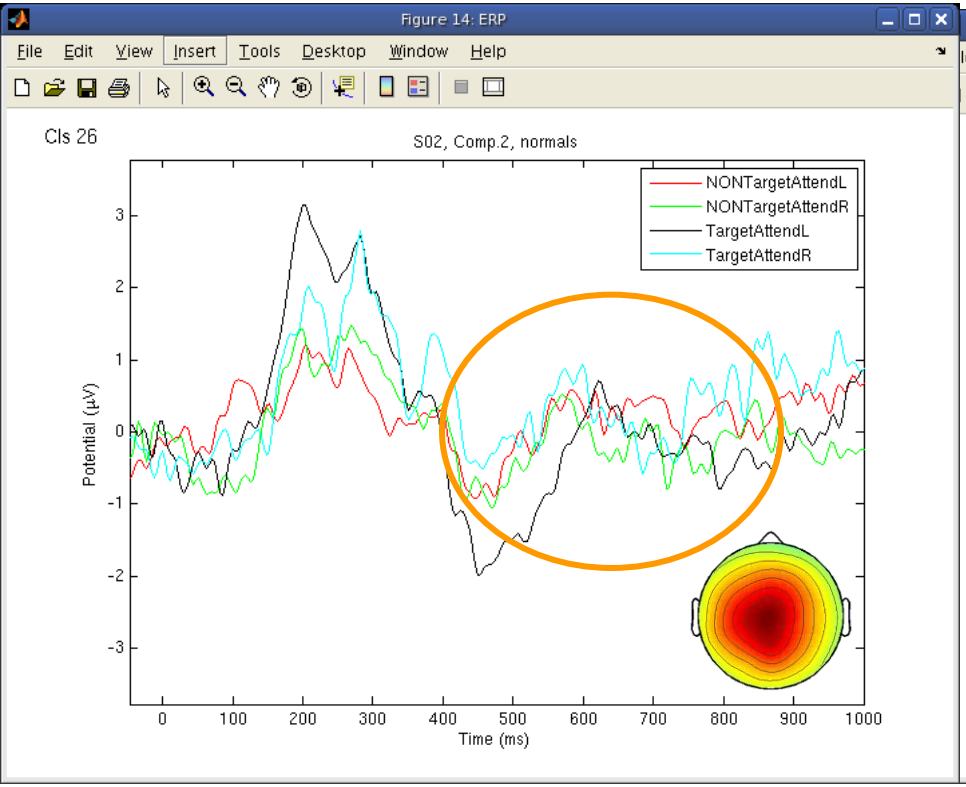
Obvious dramatic effect on scalp map topography:

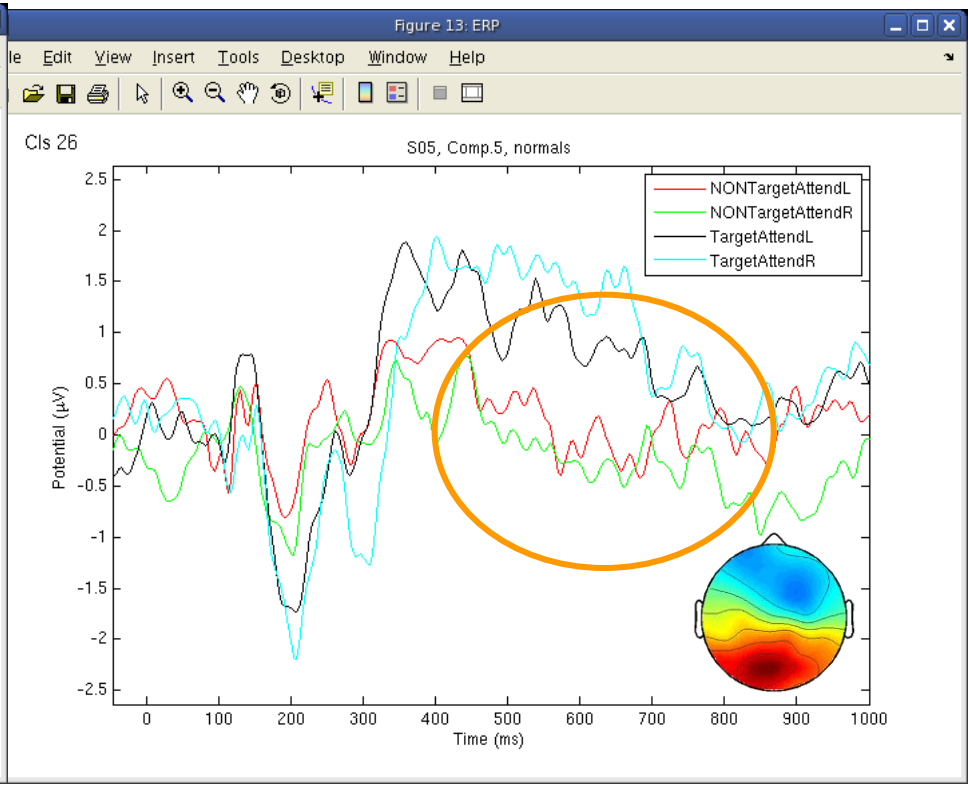




But, do they perform the same functions?

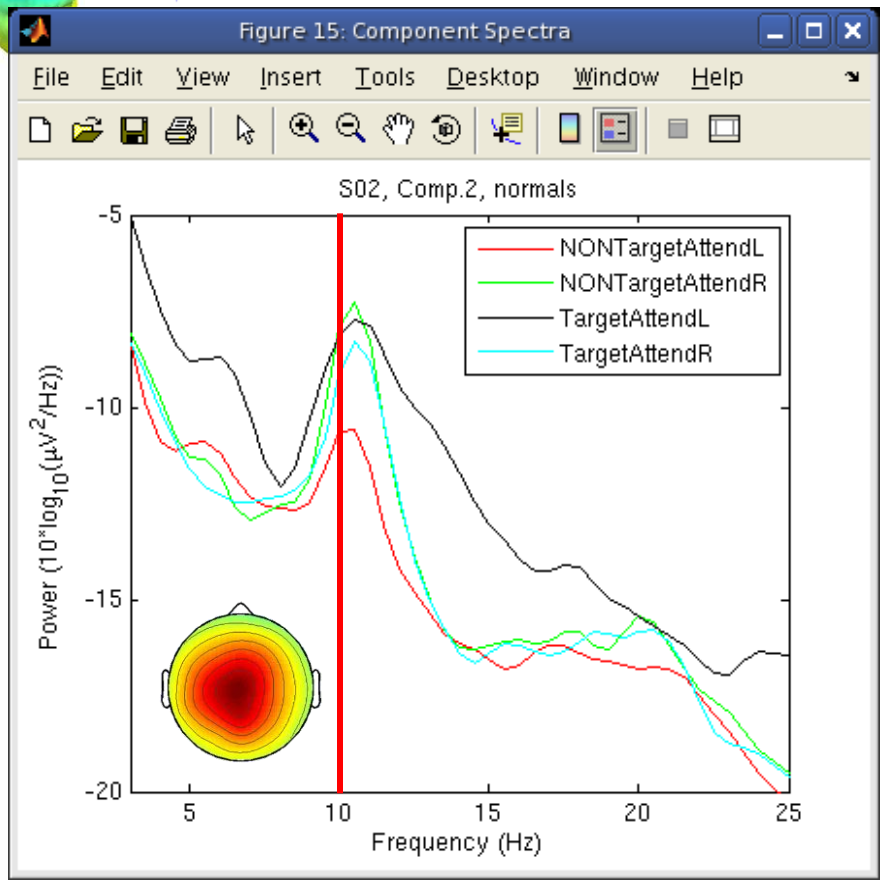
Choosing data measures

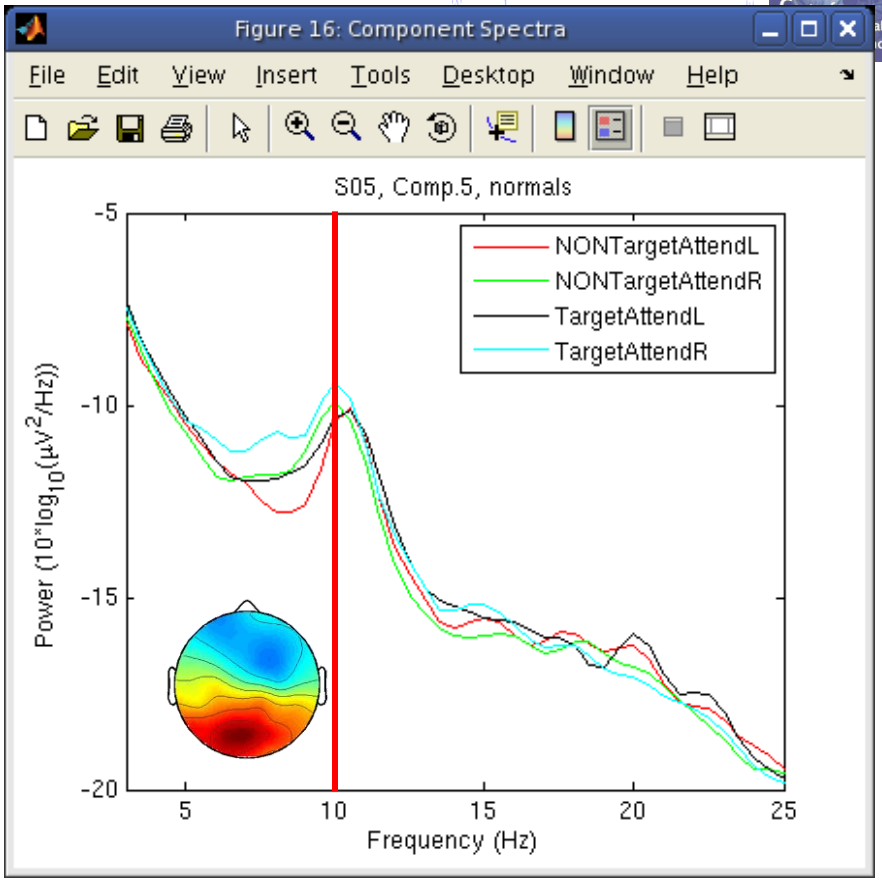




ERPs seem different…

Choosing data measures

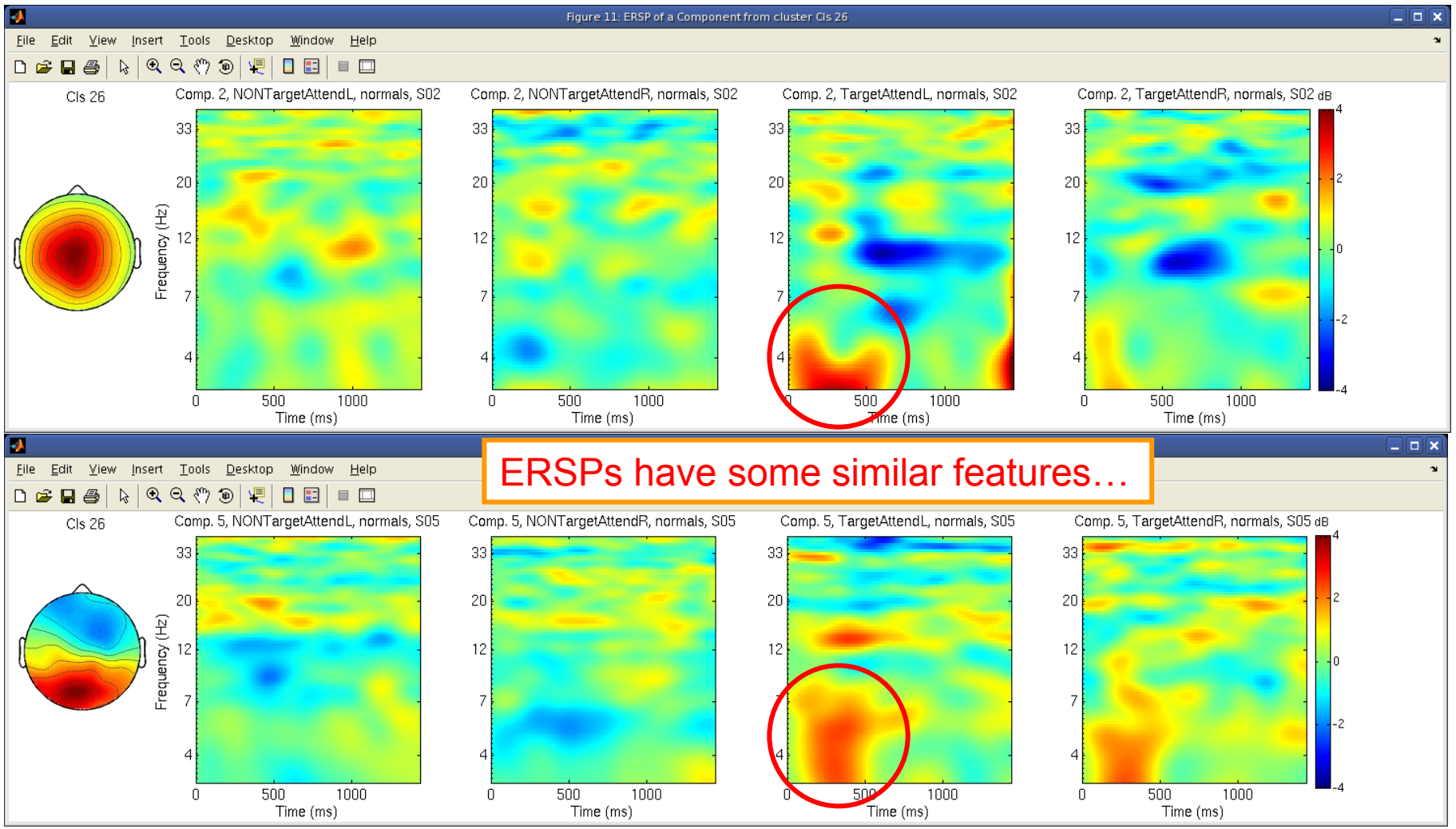




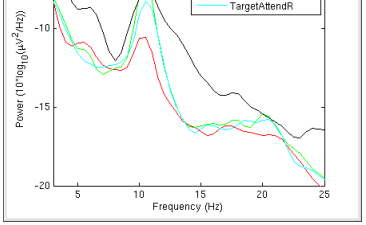
Spectra are similar, but they have

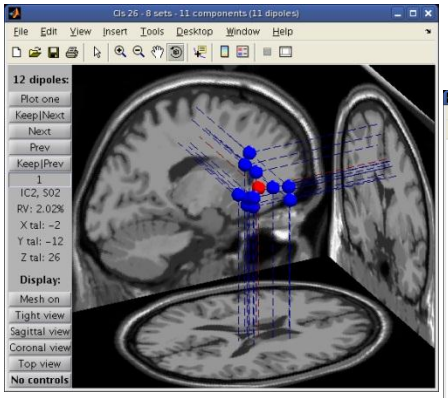
variable responses to different conditions…

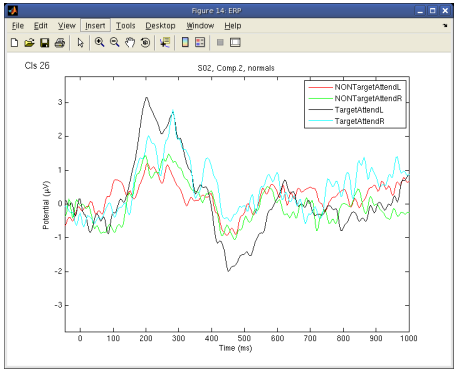
Choosing data measures

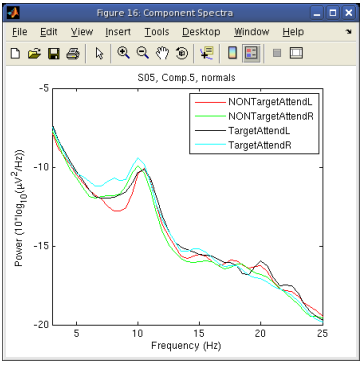


Choosing data measures





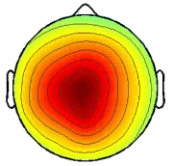


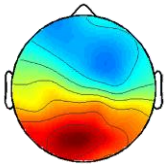


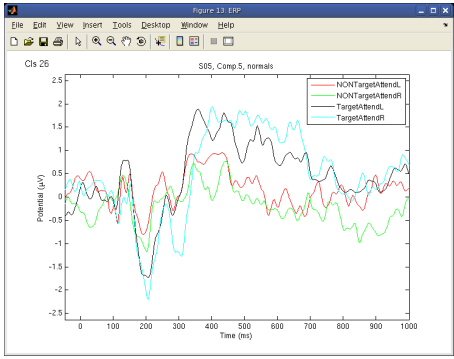
What data measures should you use?

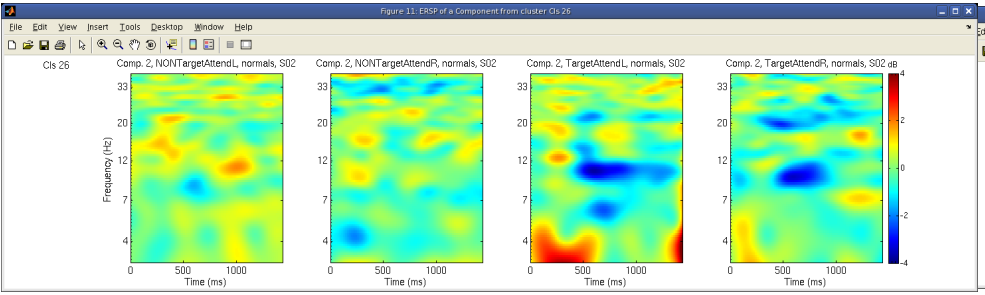
It depends…

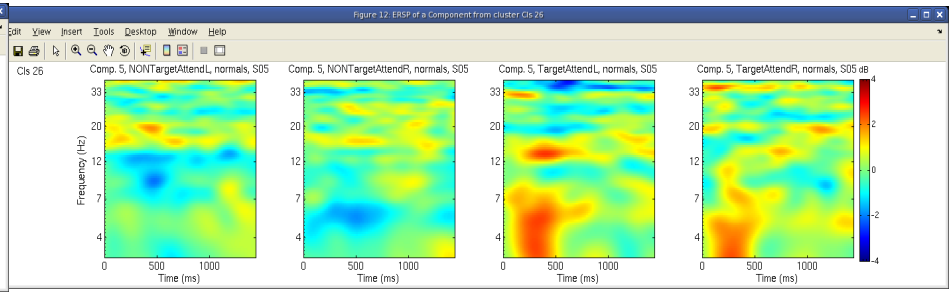
- broadly-matched ICs: use many/all of the measures.
- specifically-matched ICs: use one/few of the measures.



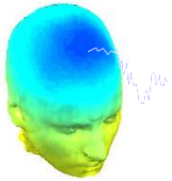




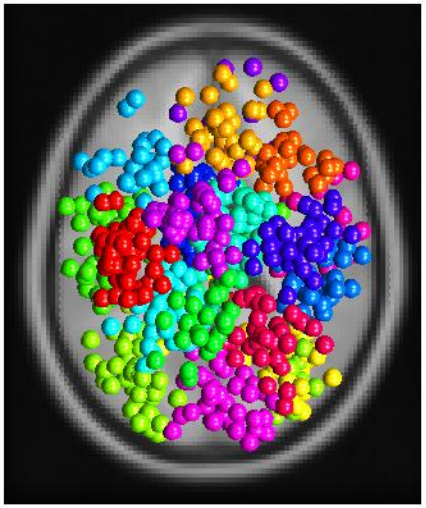


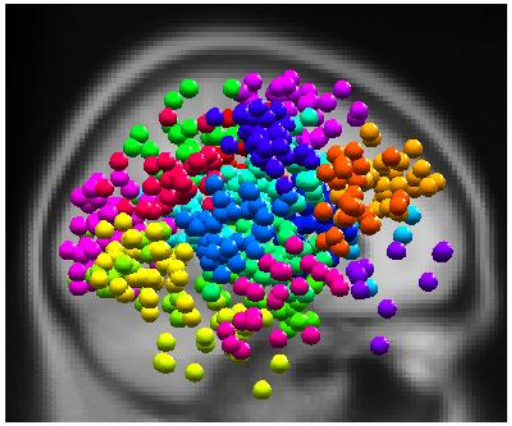


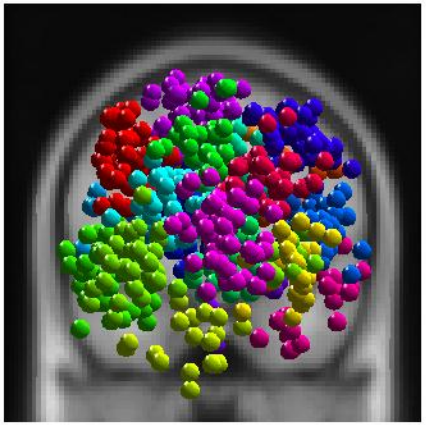
What should clusters look like?

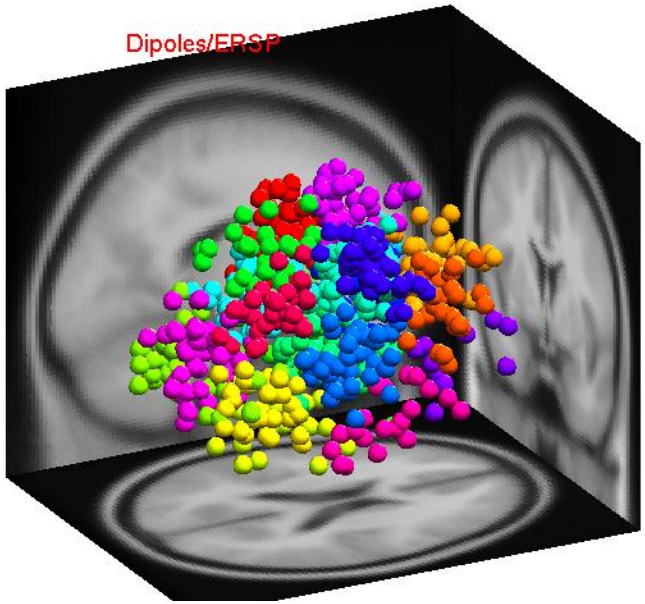


ICs clustered by dipole location and ERSP activity





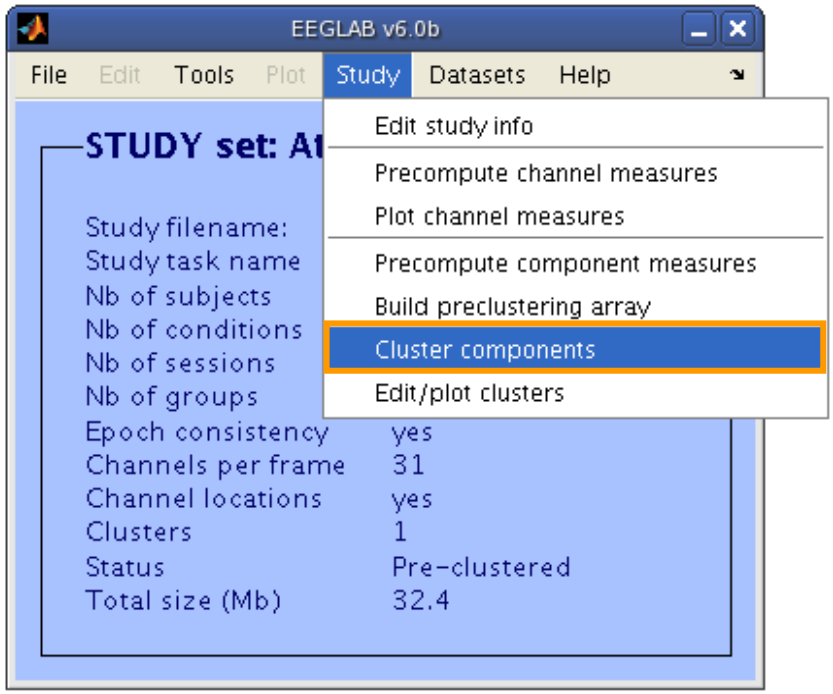




STUDY clustering overview

1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

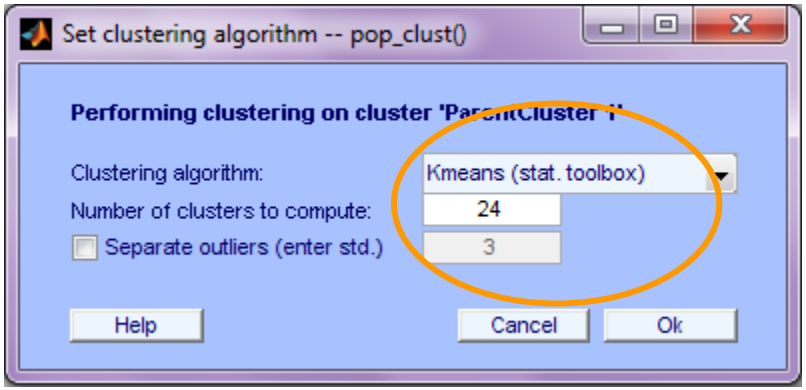
Cluster components



How to determine number of clusters?

Try the average number of ICs per subject.

Then adjust up or down from there.

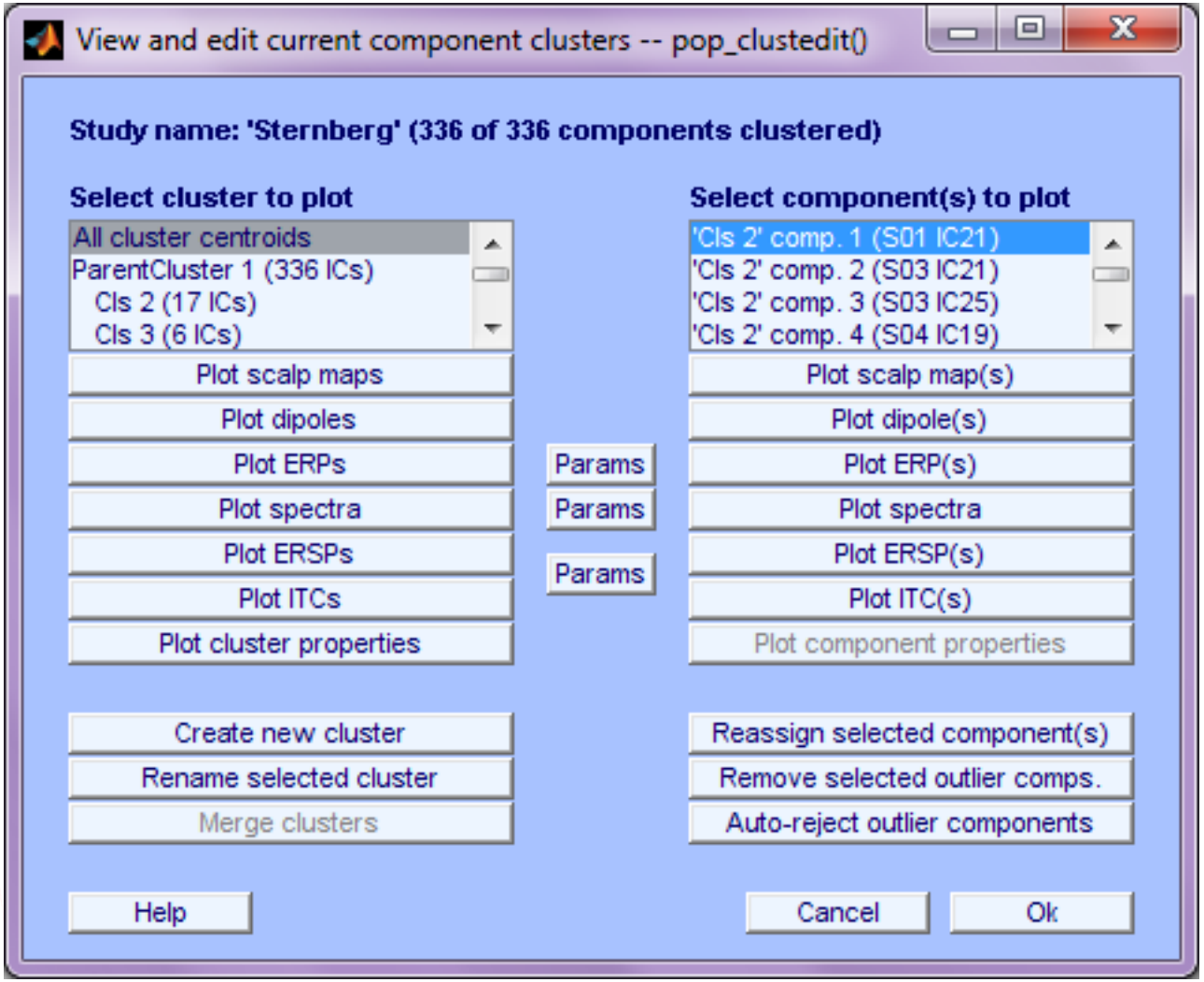



```
nclusts = 24; % choose # of clusters to create (no 'right' answer)
[STUDY] = pop_clust(STUDY, ALLEEG,'algorithm','kmeans','clus_num',nclusts);
```

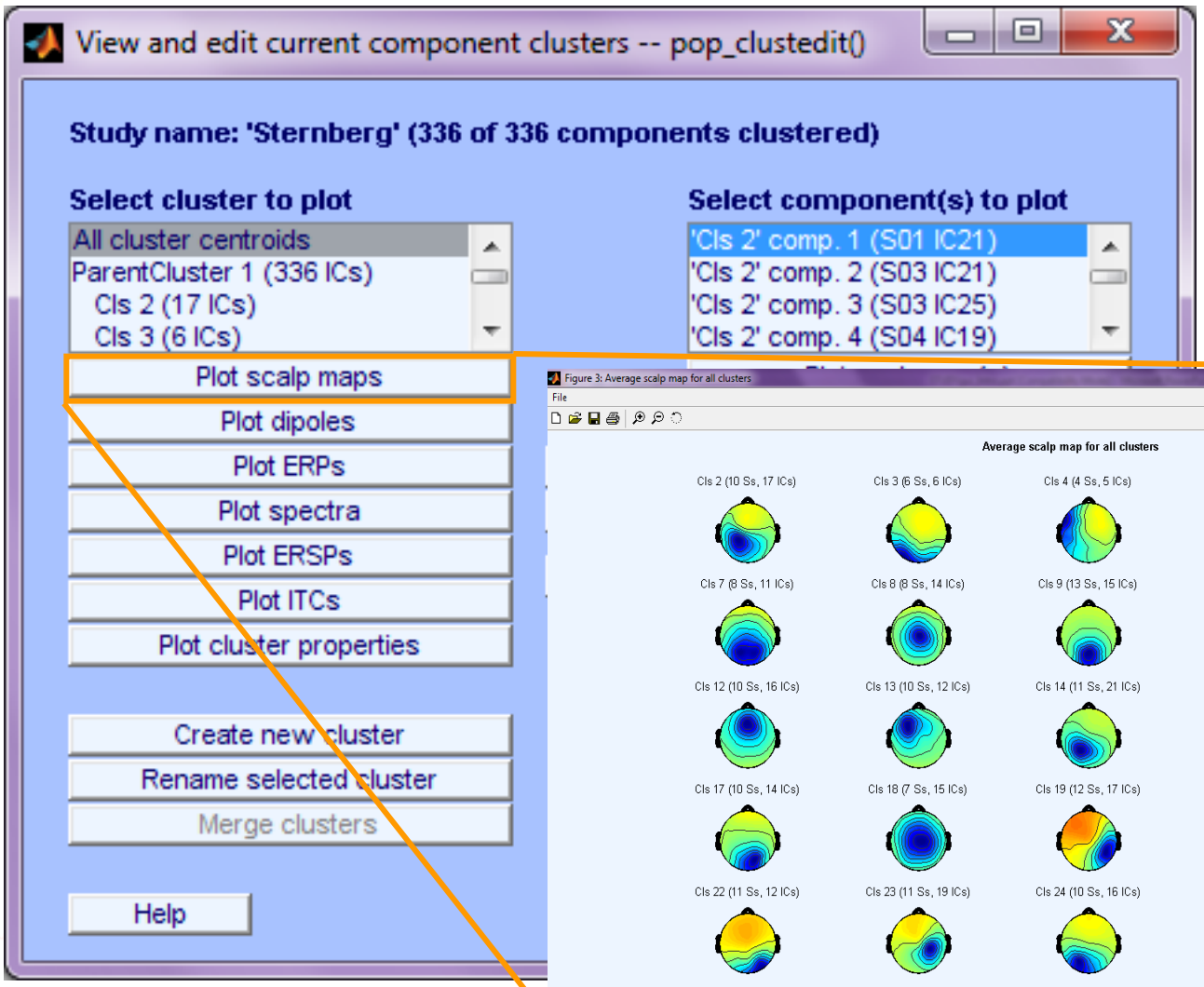
STUDY clustering overview

1. Build a STUDY
2. Create STUDY design
3. Precompute data measures
4. Precluster data measures
5. Cluster data measures
6. Plot/edit clusters

Plot/edit clusters

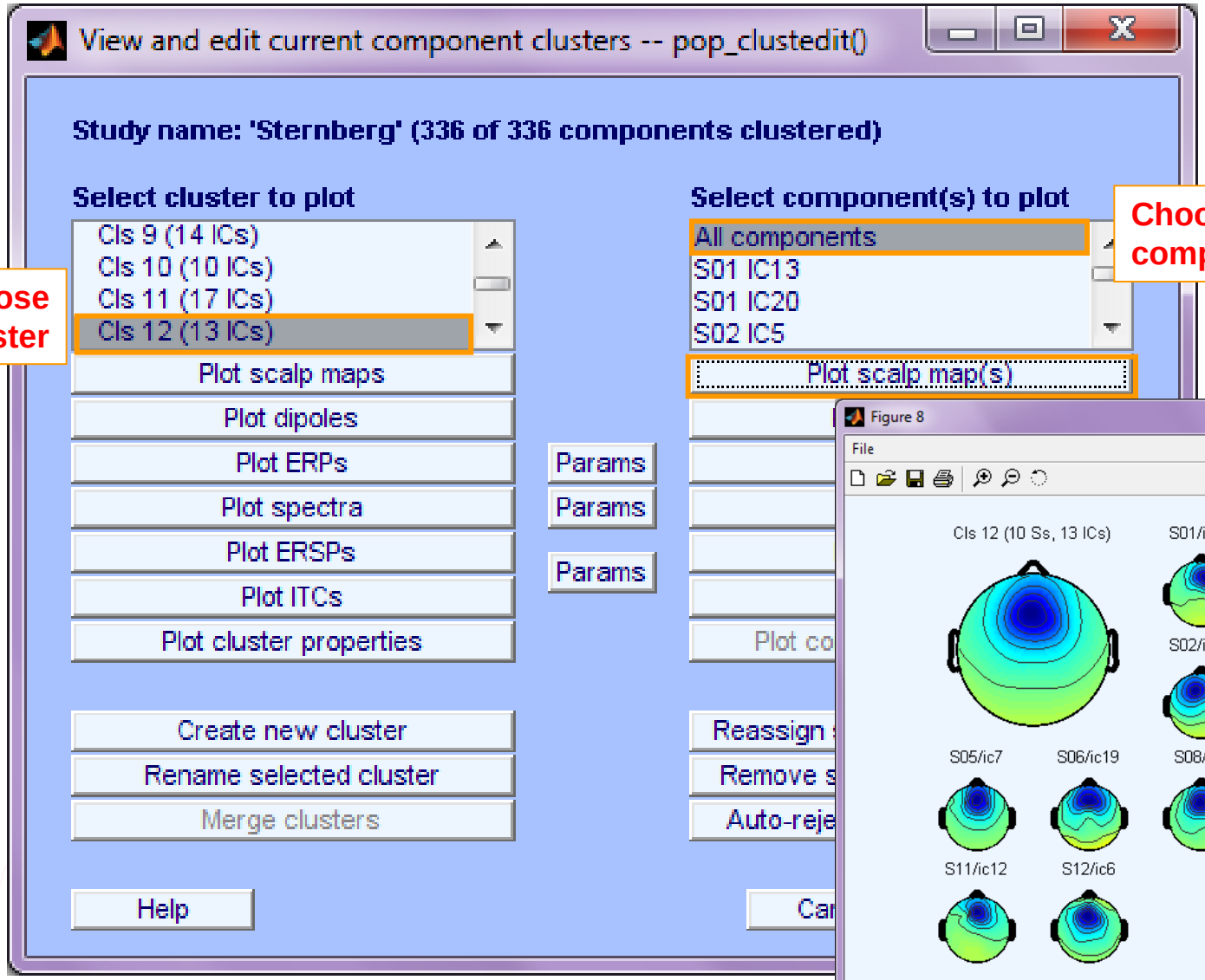


Plot cluster data



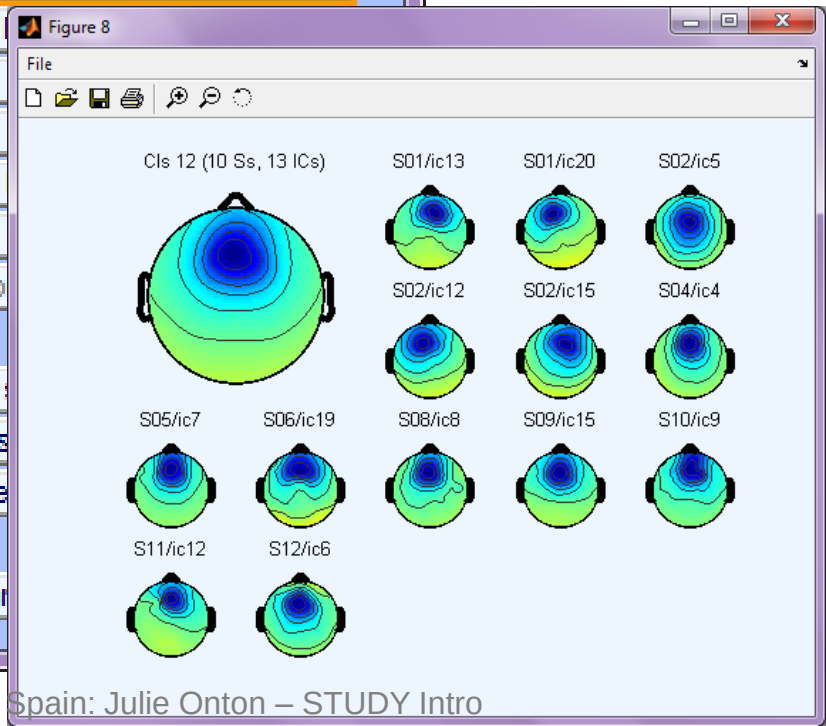
Plot mean scalp maps for easy reference

Plot cluster data

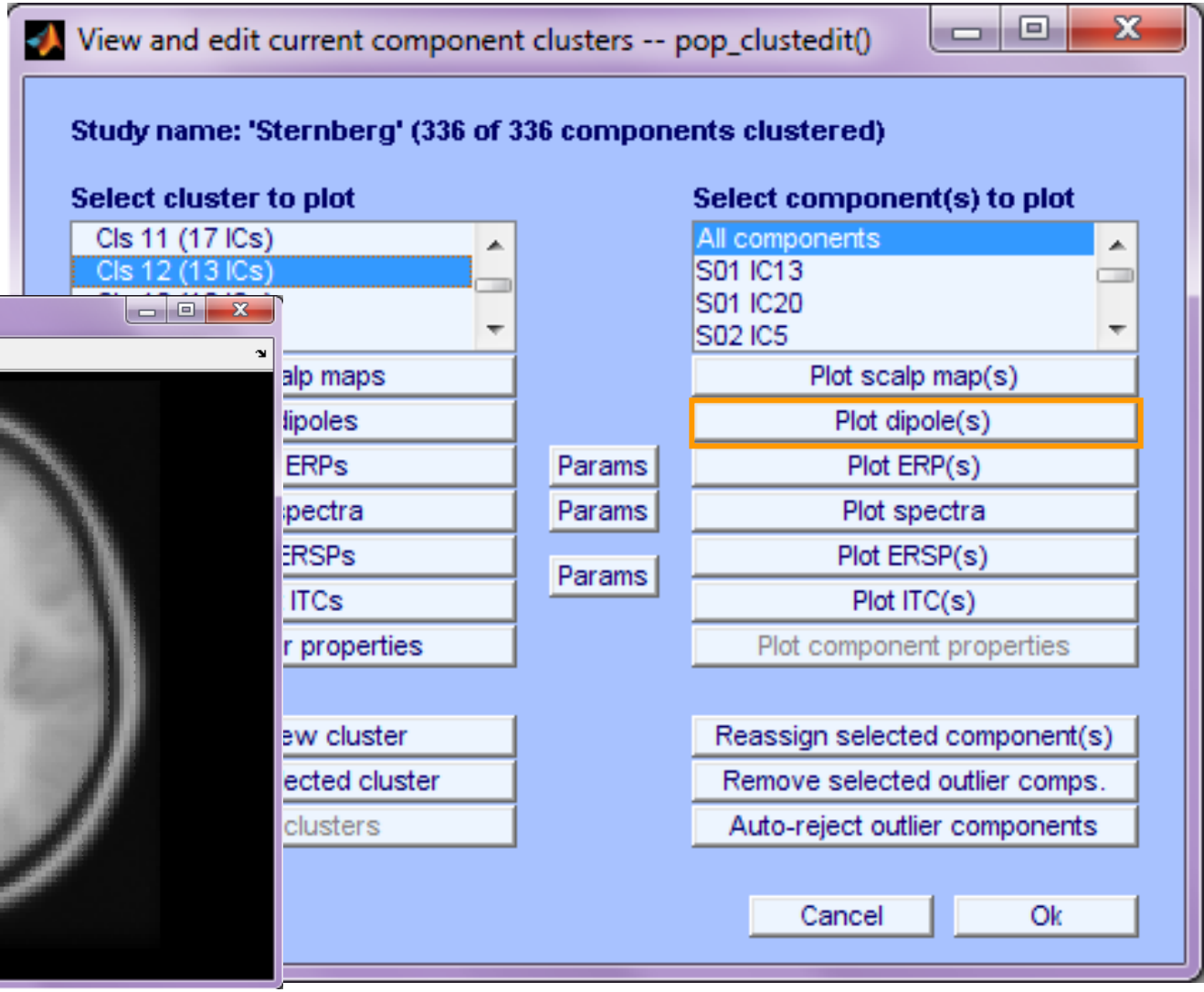


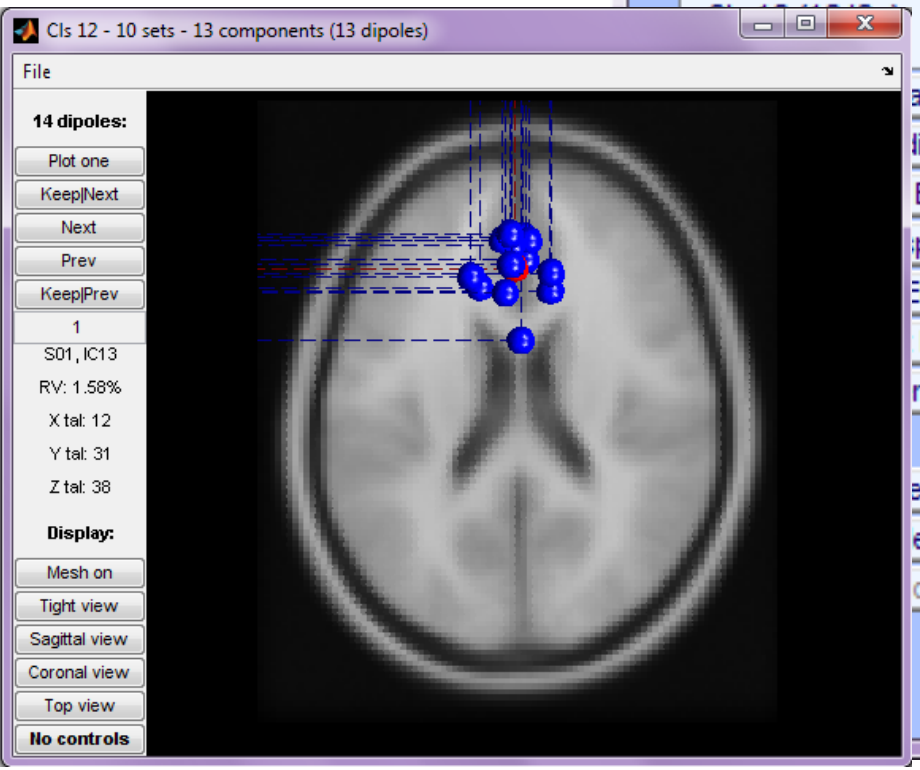
Choose a cluster

Choose which components

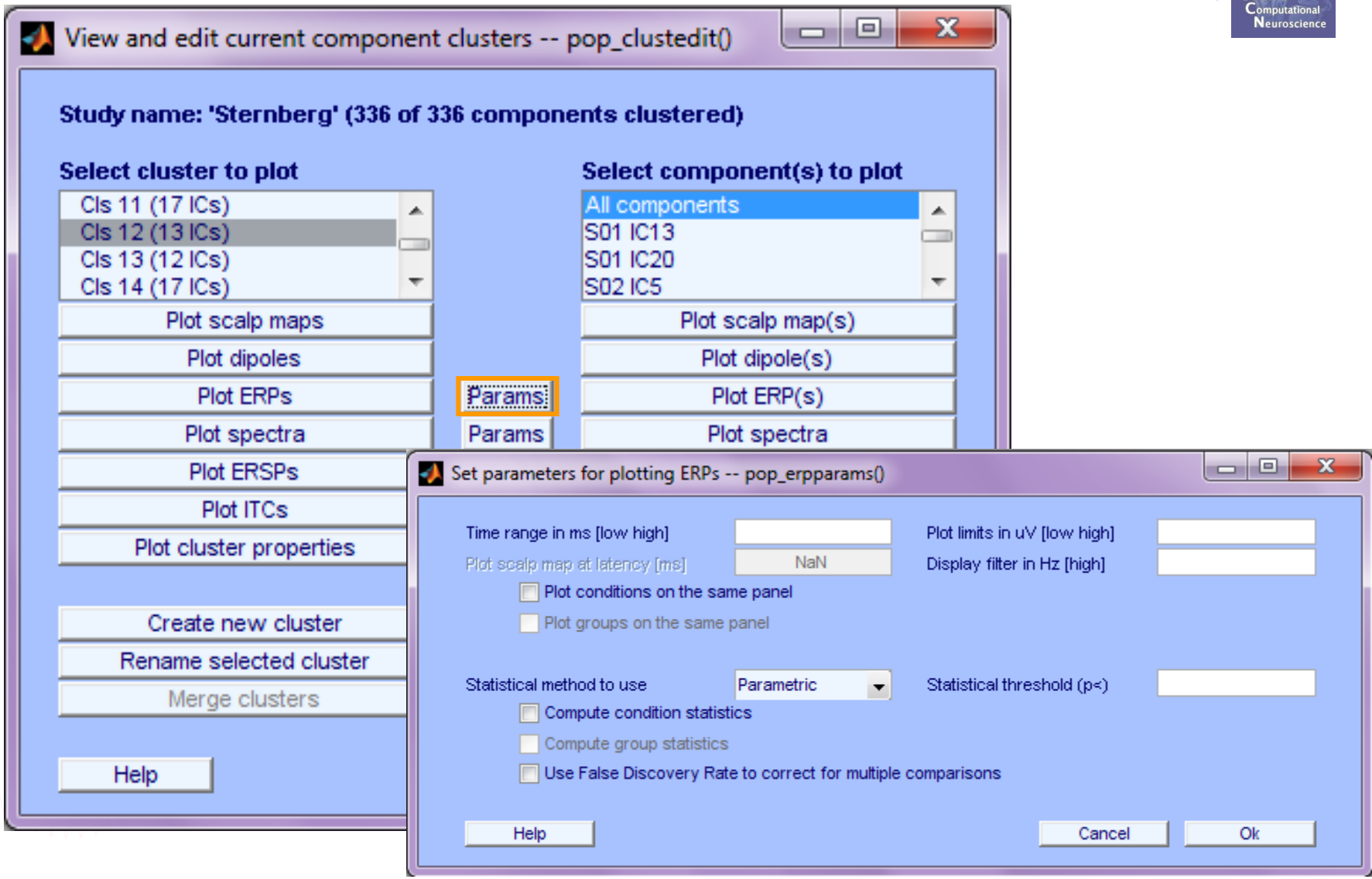


Plot clusters

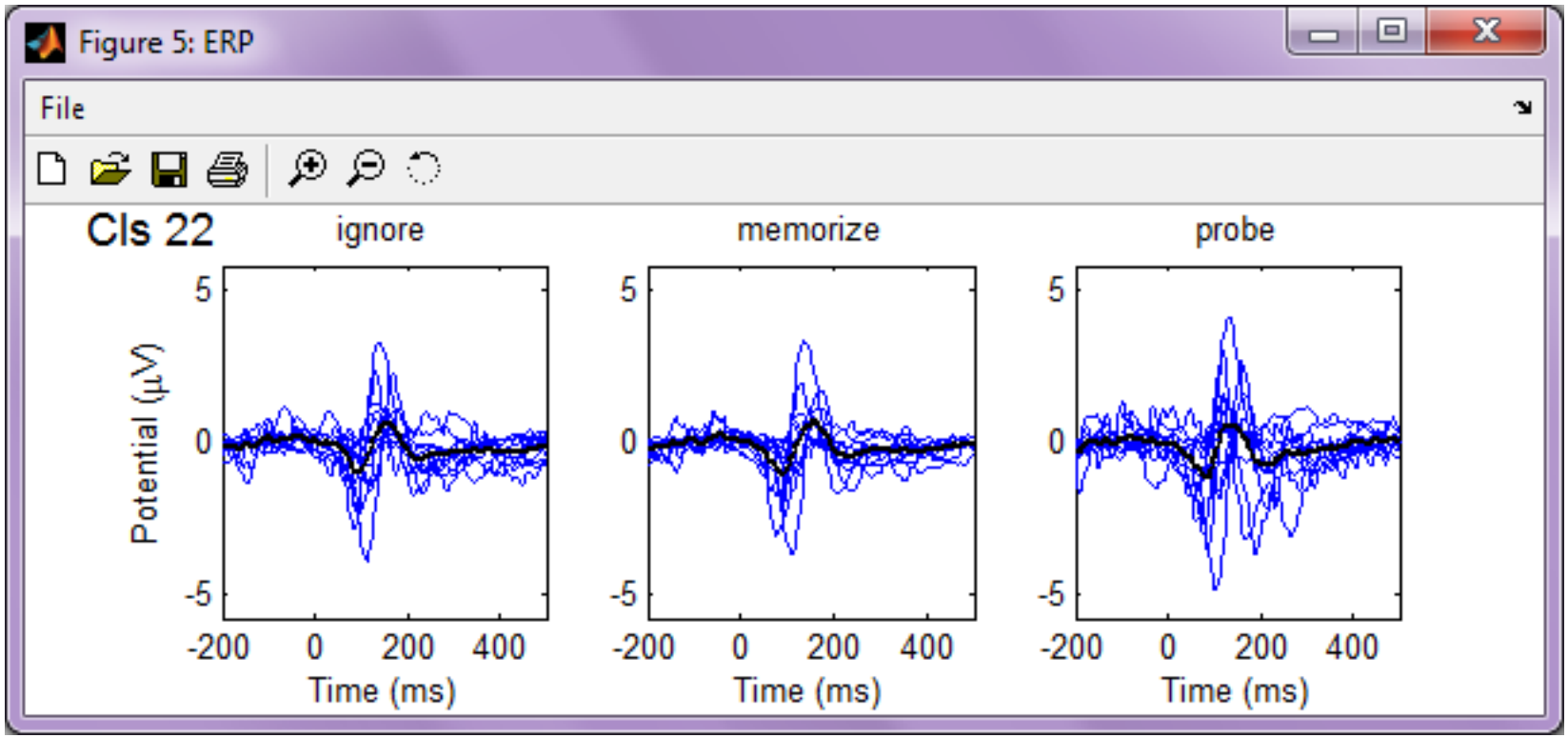




Plot ERPs

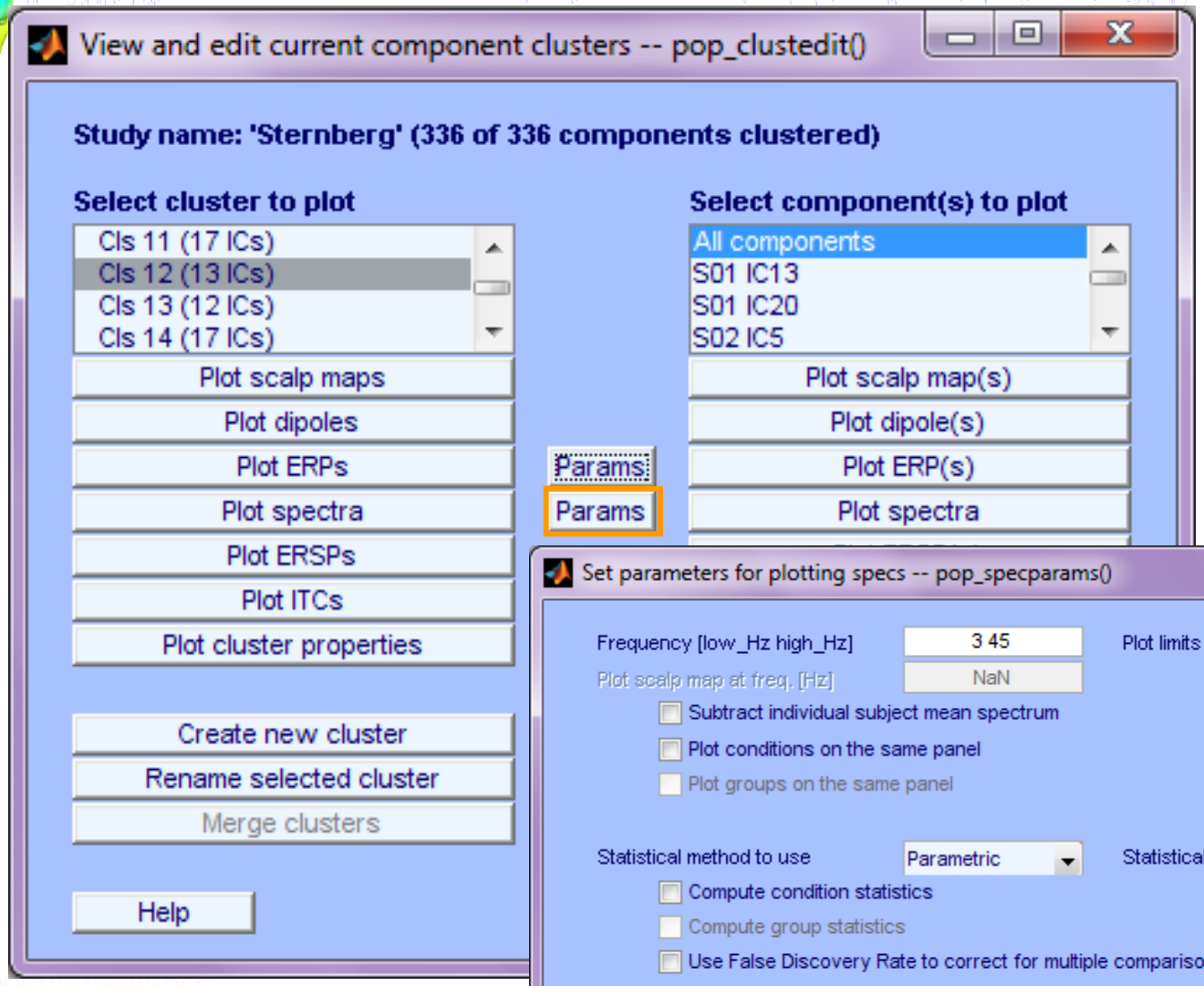

View and edit current component clusters -- pop_clustedit()
Study name: 'Sternberg' (336 of 336 components clustered)
Select cluster to plot
Cls 11 (17 ICs)
Cls 12 (13 ICs)
Cls 13 (12 ICs)
Cls 14 (17 ICs)
Select component(s) to plot
All components
S01 IC13
S01 IC20
S02 IC5
Plot scalp maps
Plot dipoles
Plot ERPs
Plot spectra
Plot ERSPs
Plot ITCs
Plot cluster properties
Plot scalp map(s)
Plot dipole(s)
Plot ERP(s)
Plot spectra
Params
Params
Create new cluster
Rename selected cluster
Merge clusters
Help
Set parameters for plotting ERPs -- pop_erpparams()
Time range in ms [low high]
Plot limits in uV [low high]
Plot scalp 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 False Discovery Rate to correct for multiple comparisons
Help
Cancel
Ok

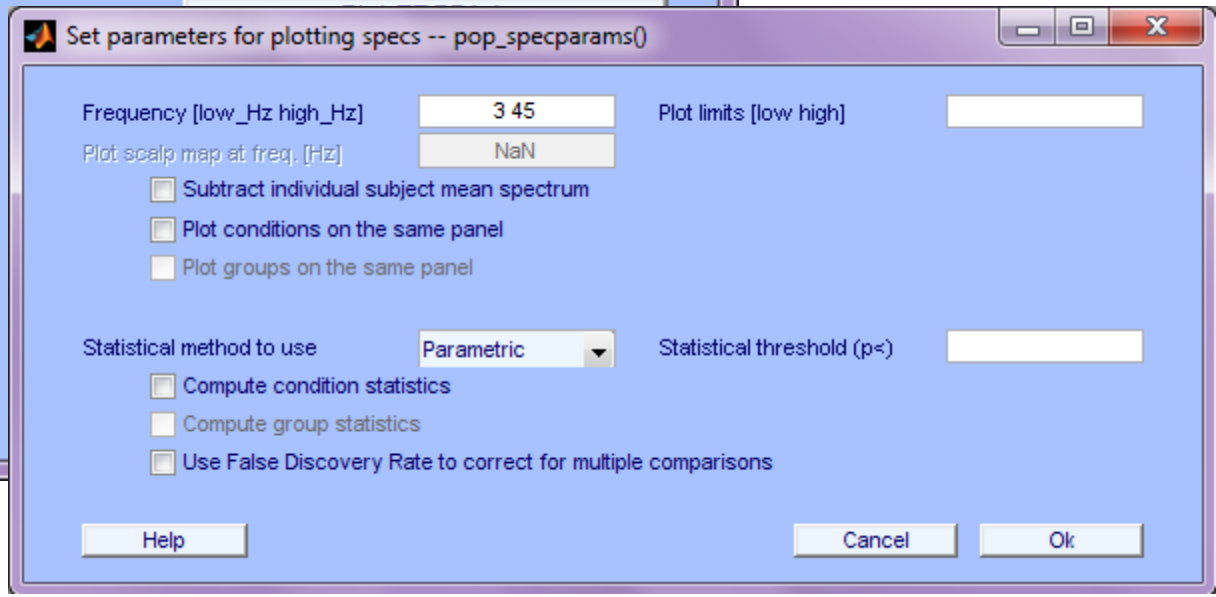
Plot cluster ERP



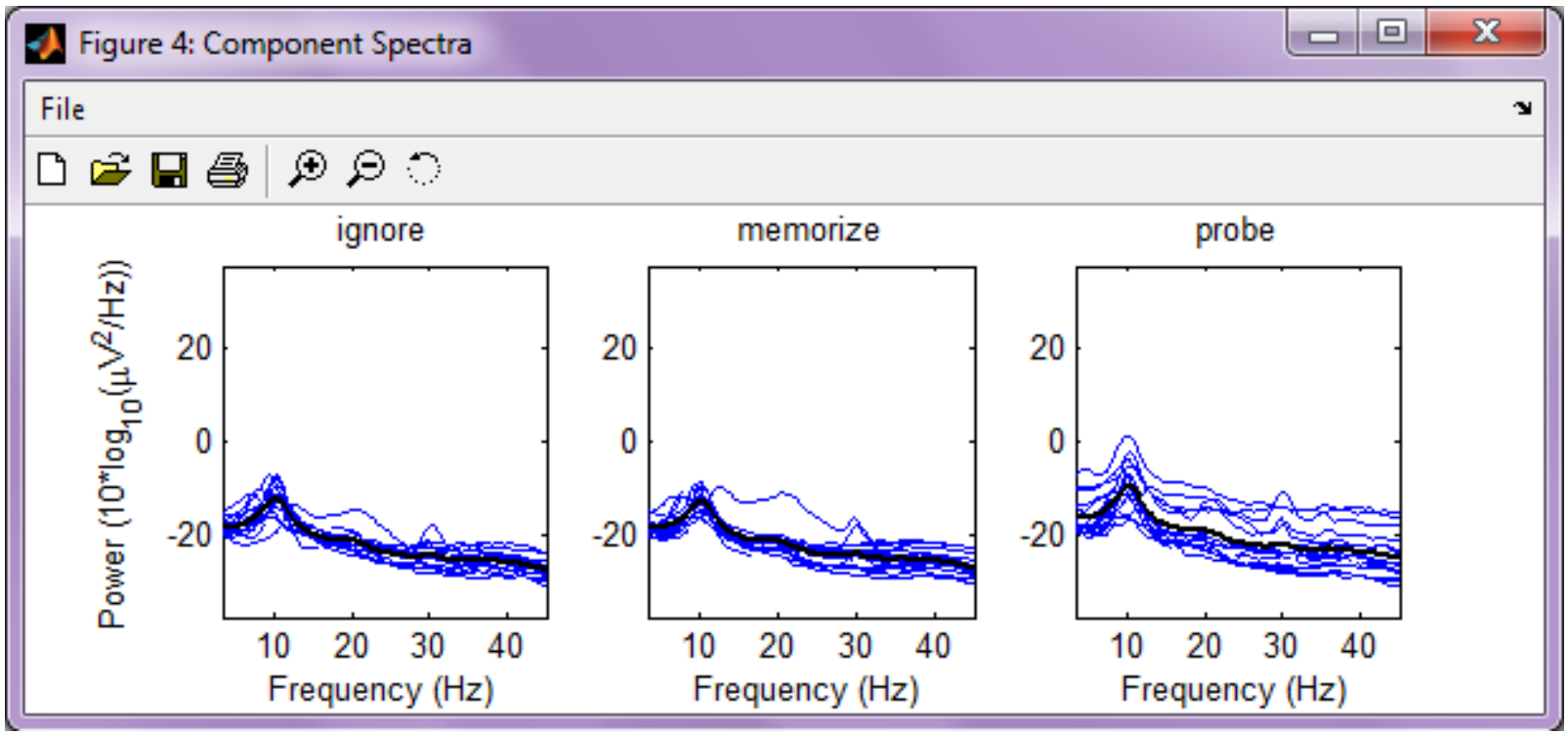
Each blue trace is the ERP of a different component

Plot cluster spectra



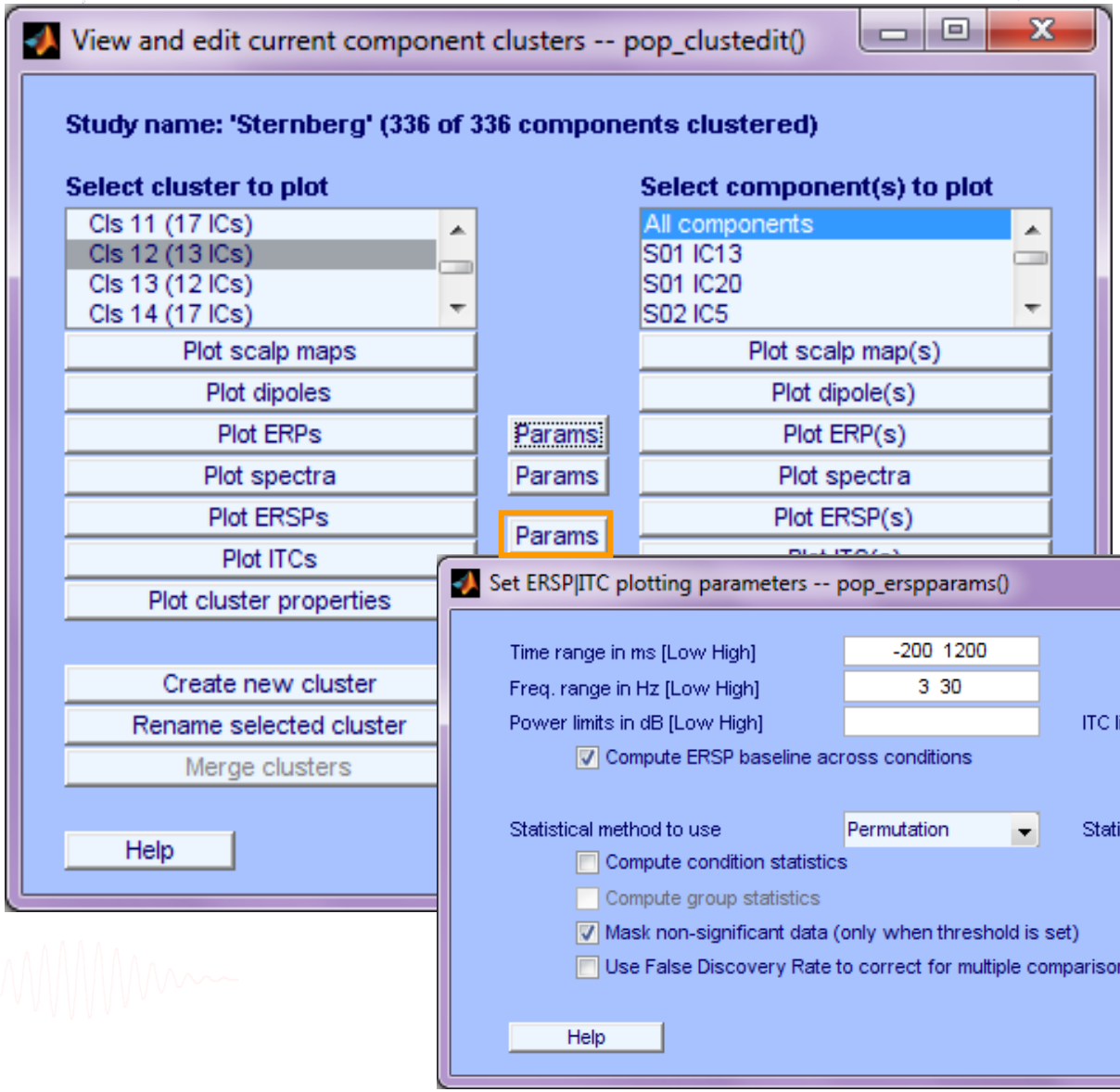


Plot cluster spectra



Each blue trace is the power spectrum of a different component

Plot cluster ERSPs

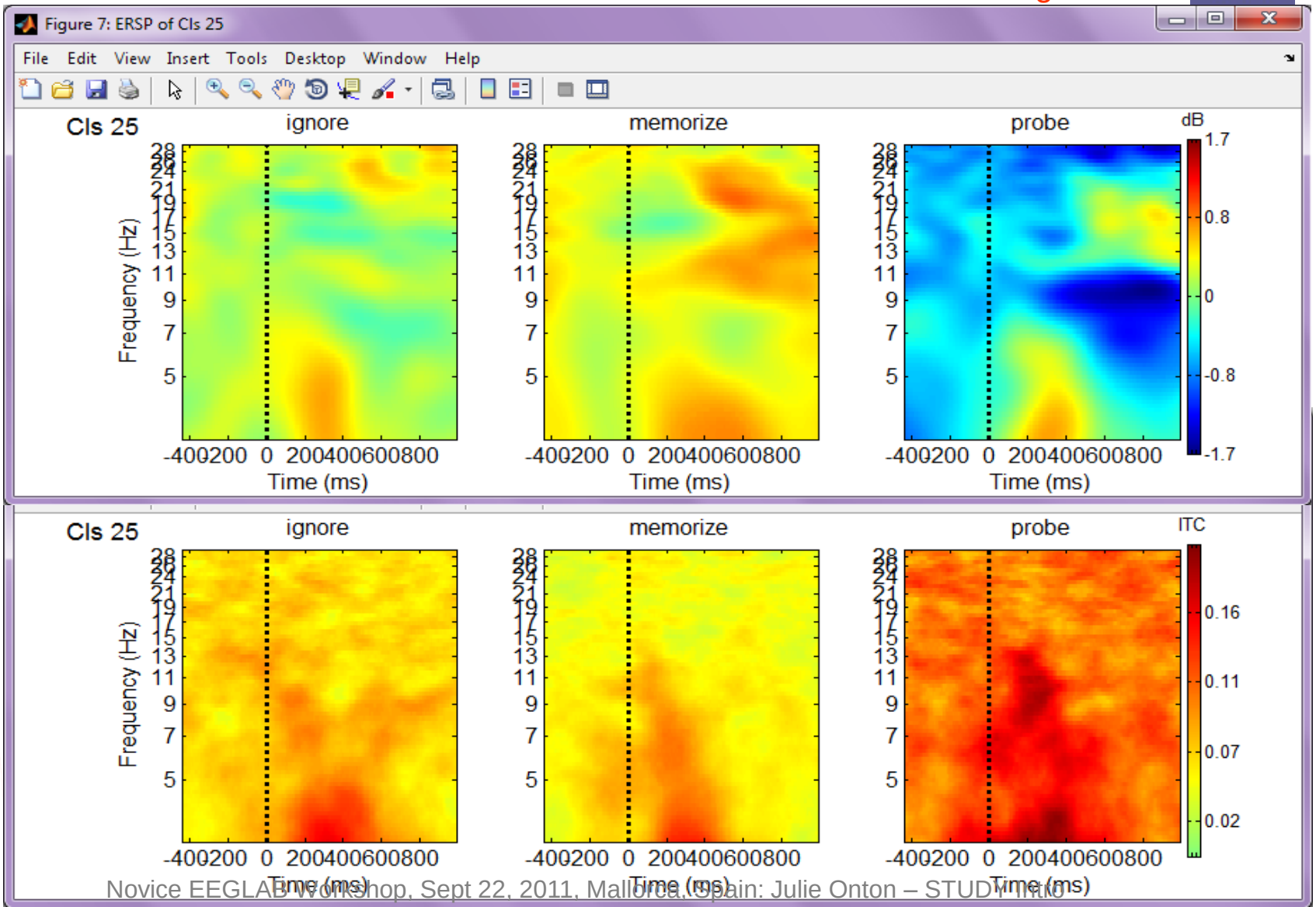


Plot cluster ERSPs and ITC

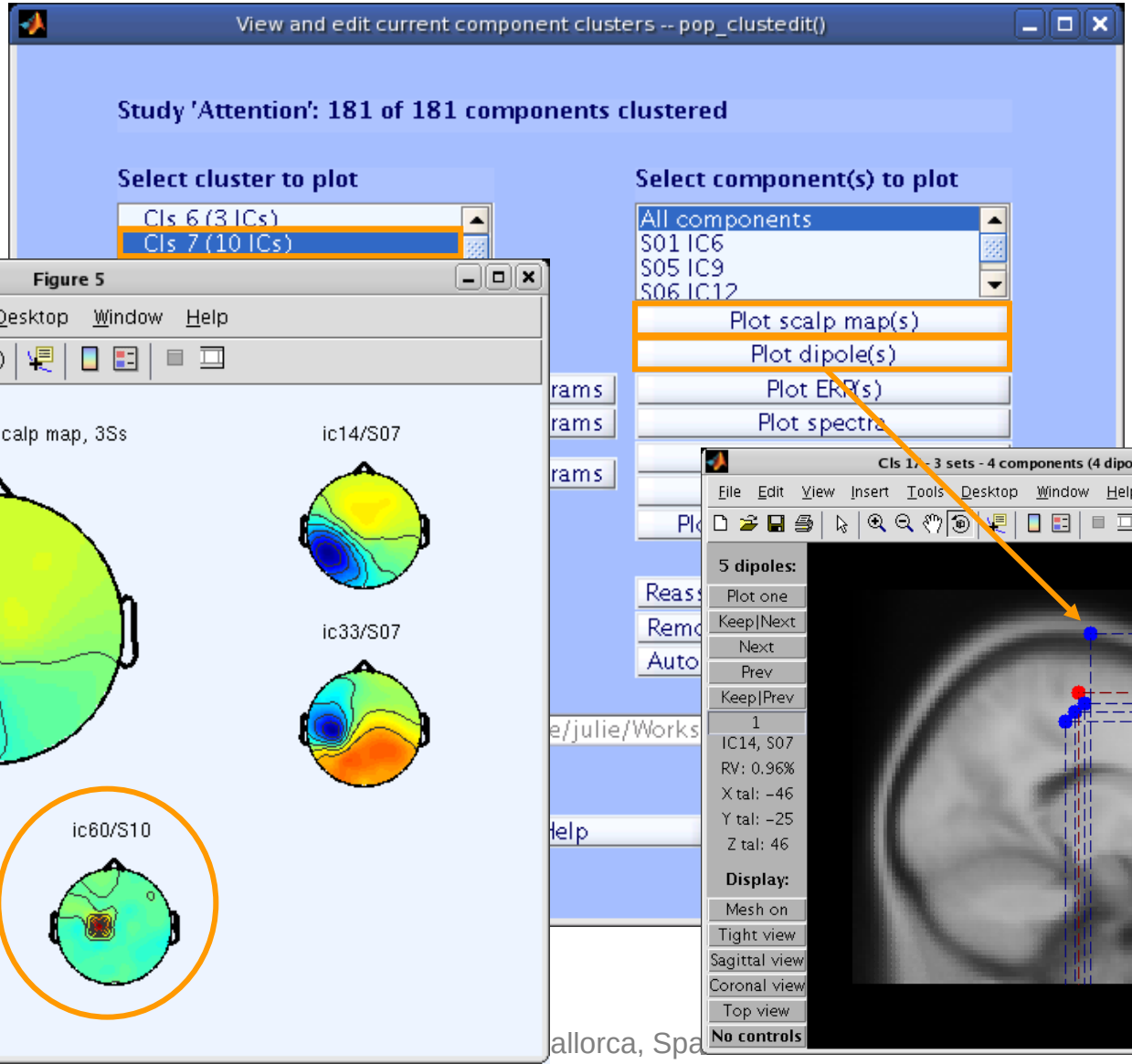
Not masked for significance

ERSP

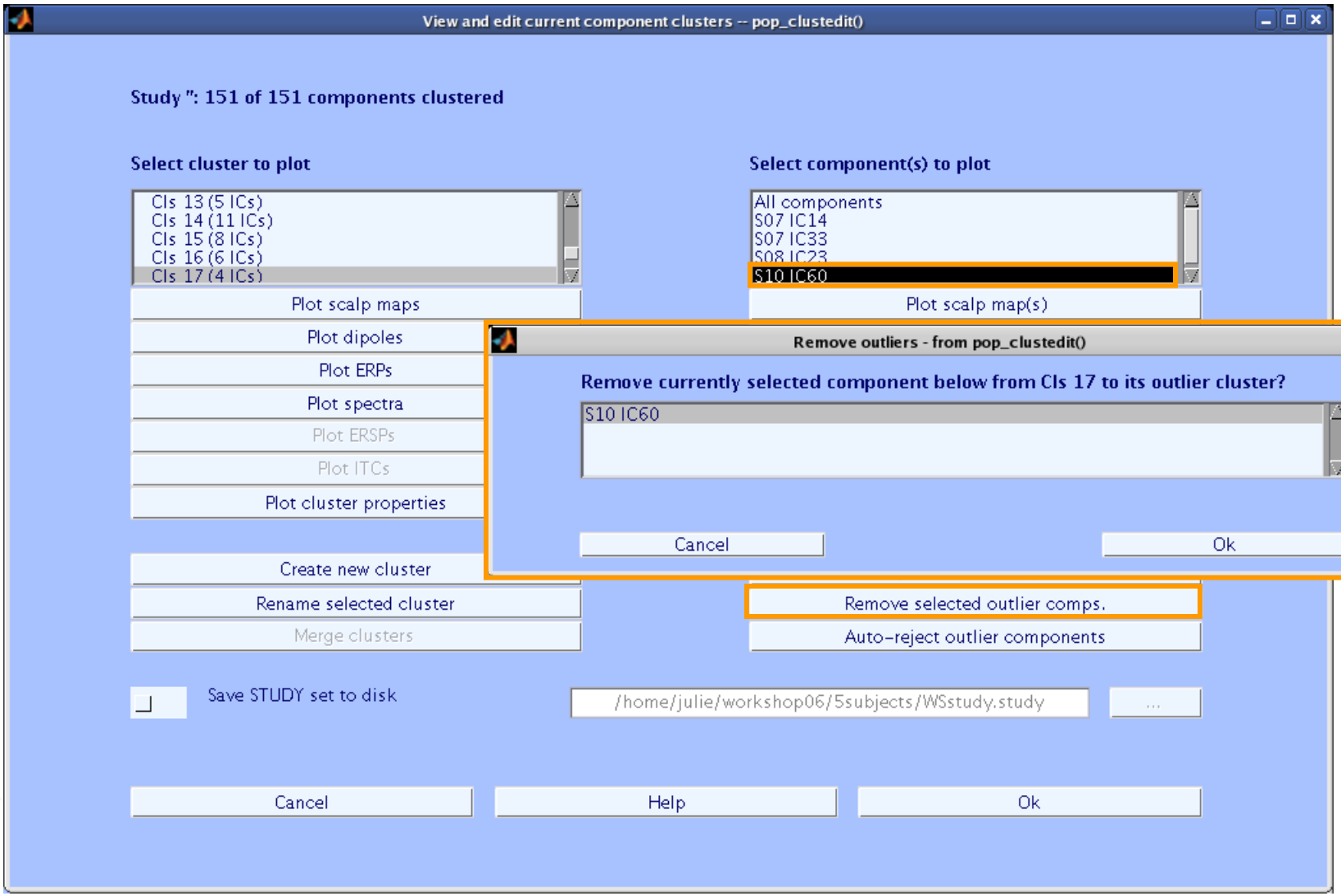
ITC



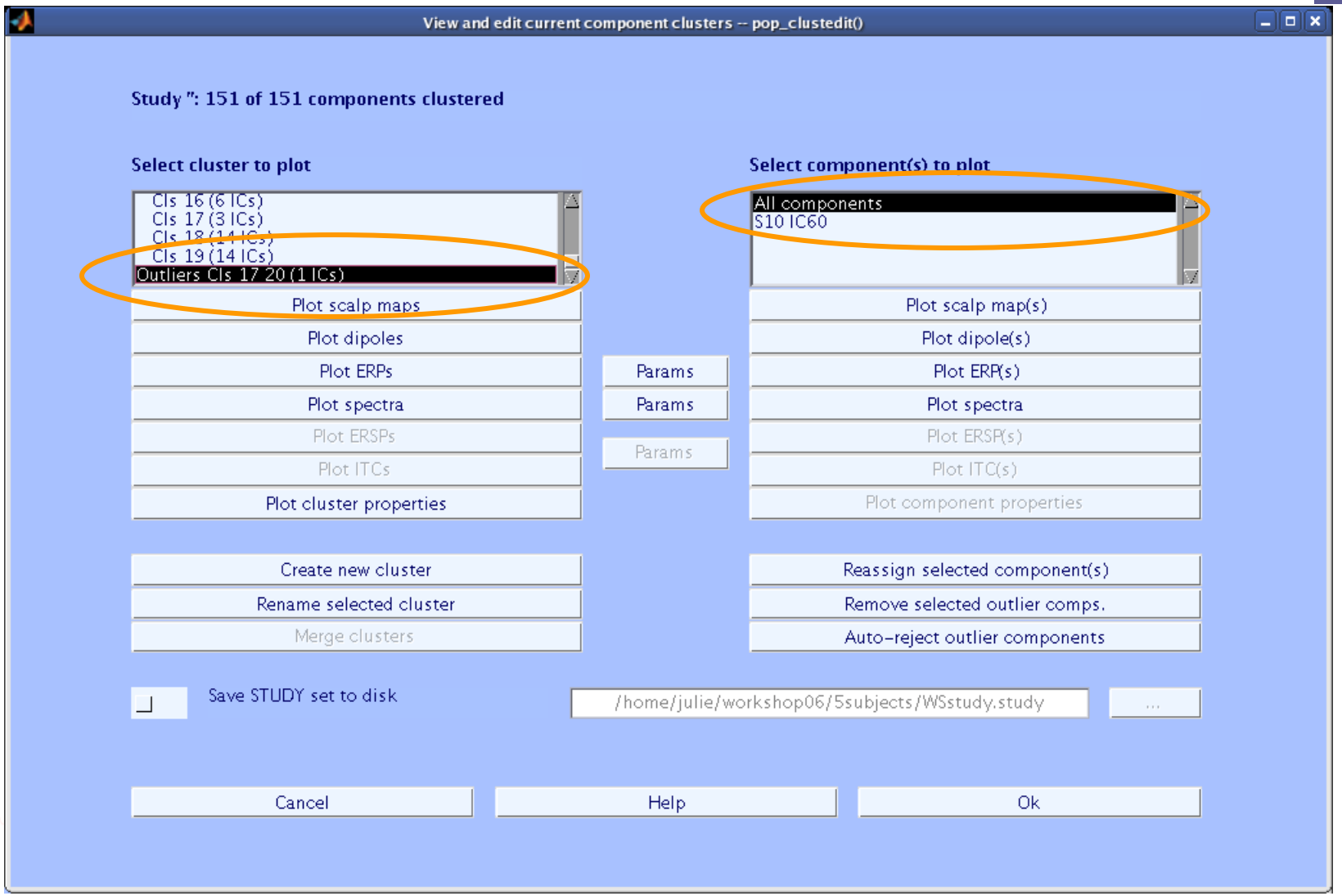
Remove outlier components



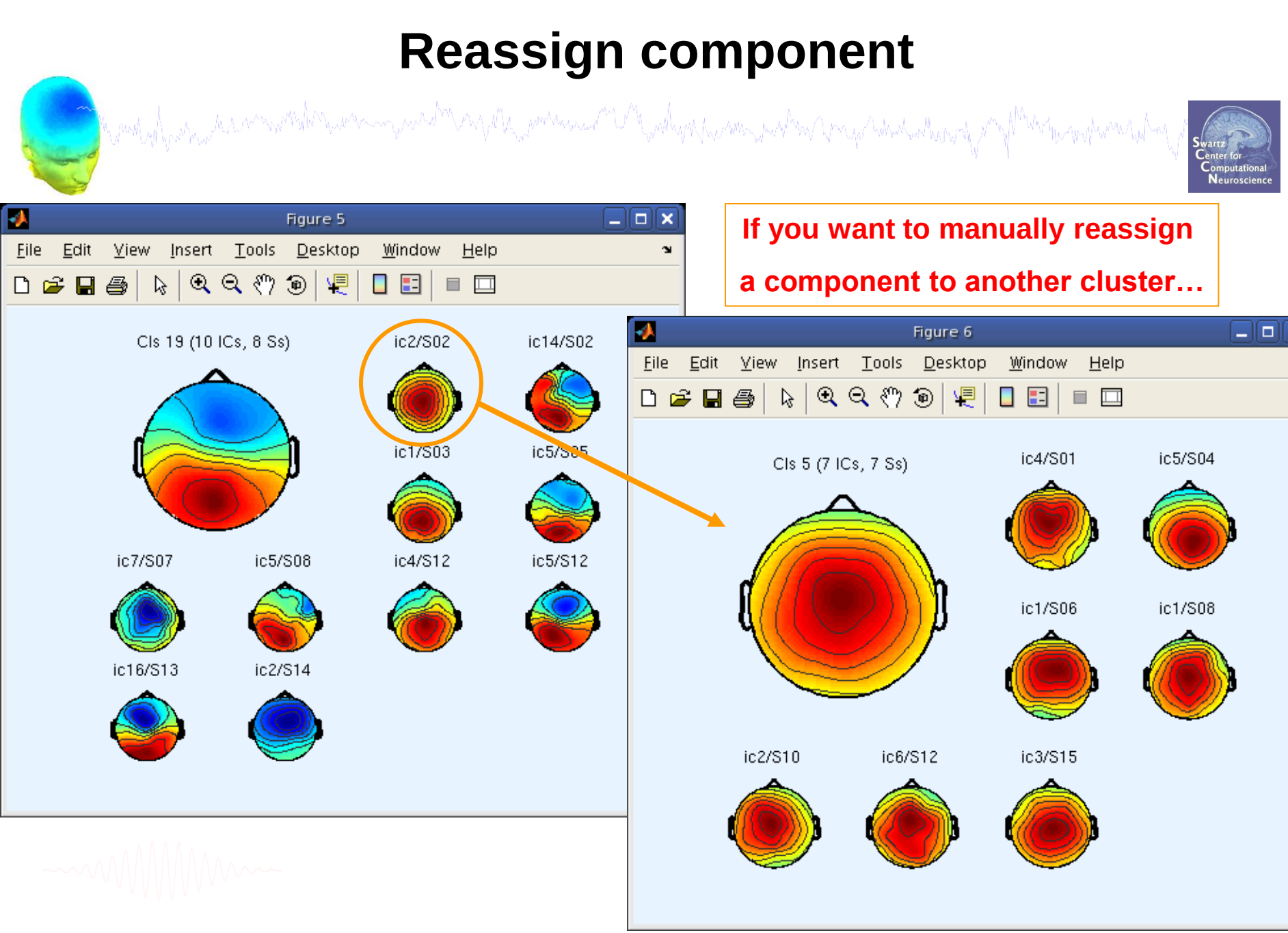
Remove outlier components



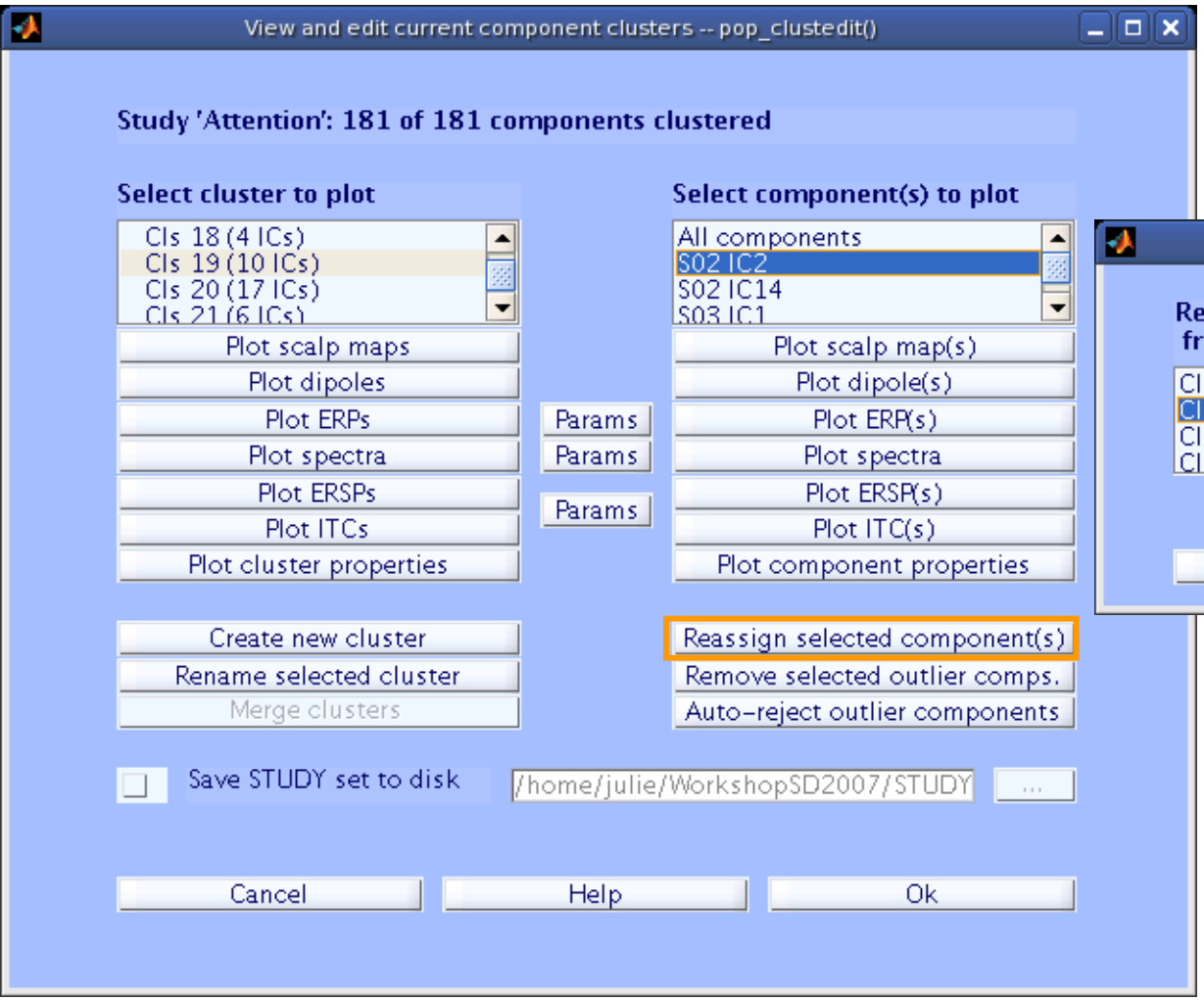
Remove outlier components

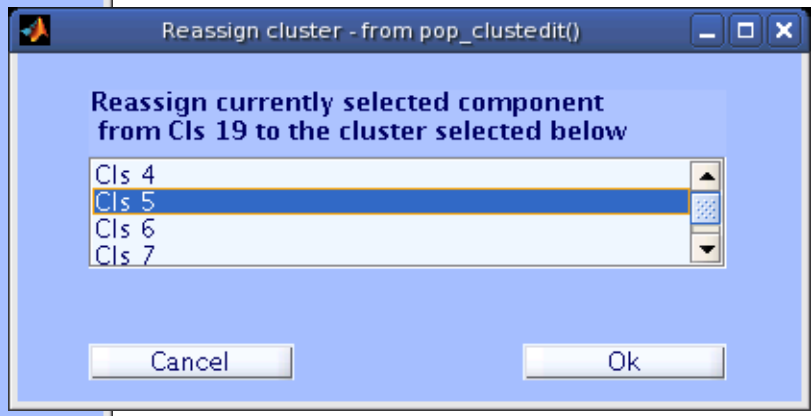


Reassign component

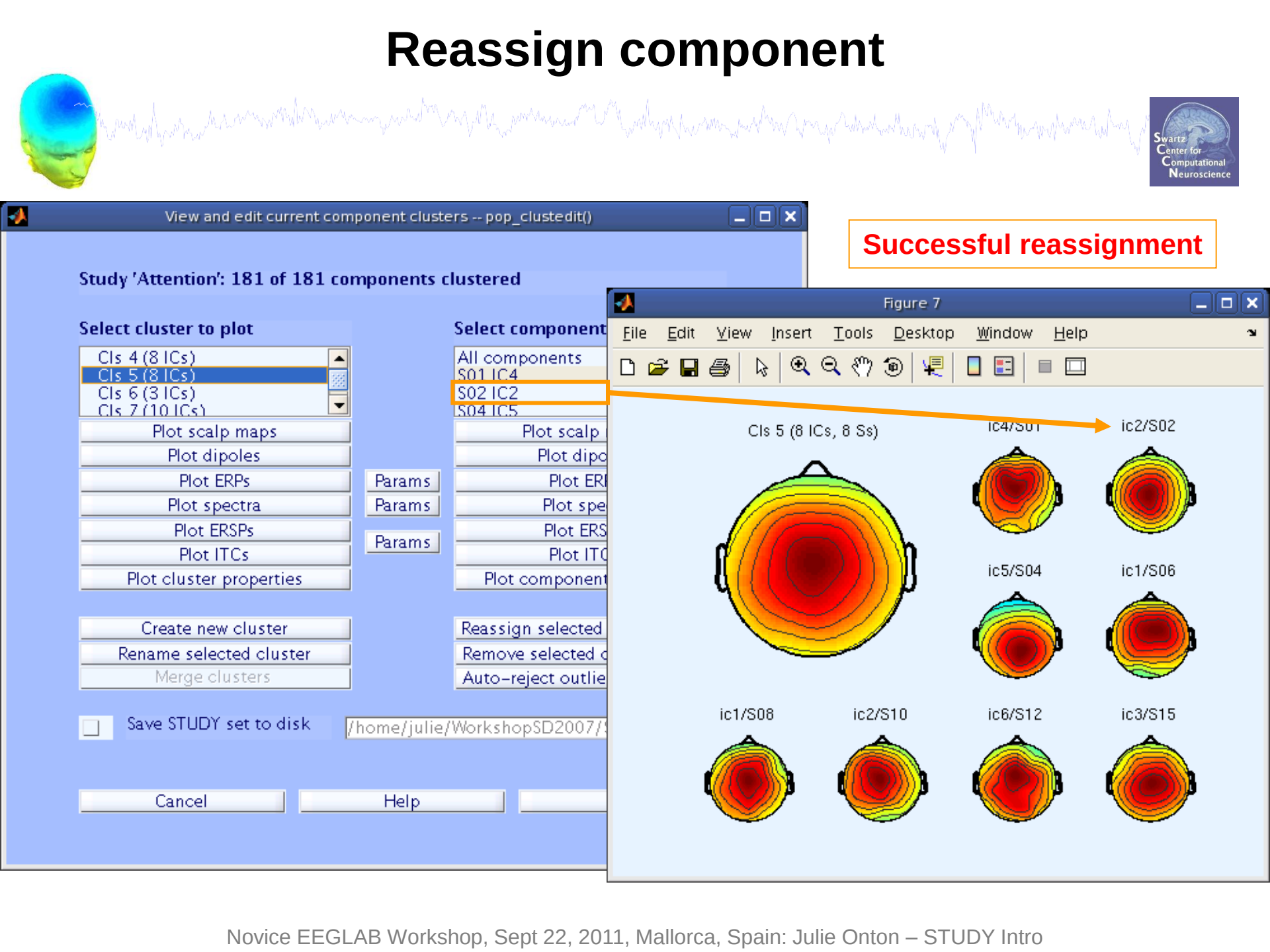


Reassign component

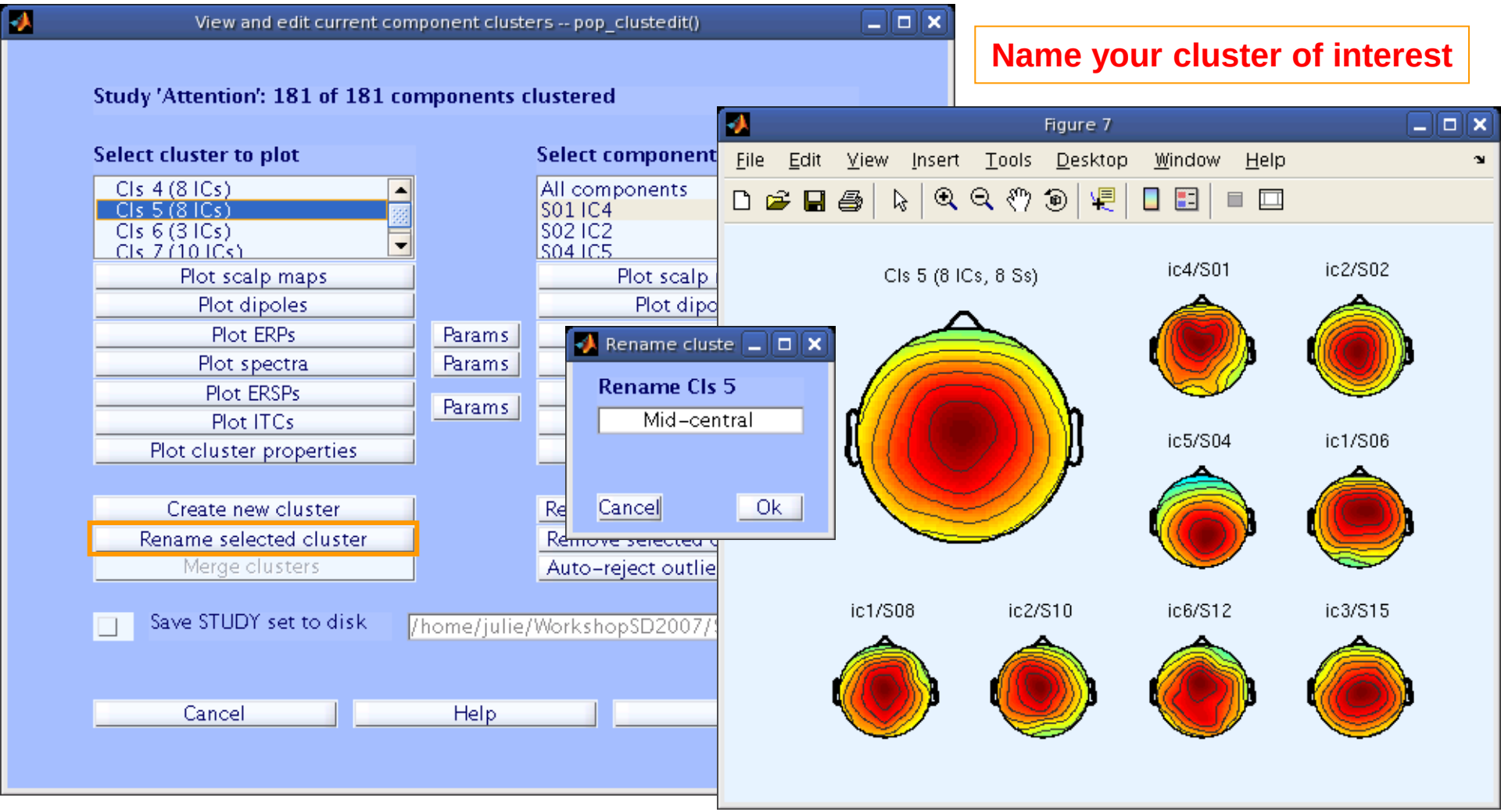




Reassign component



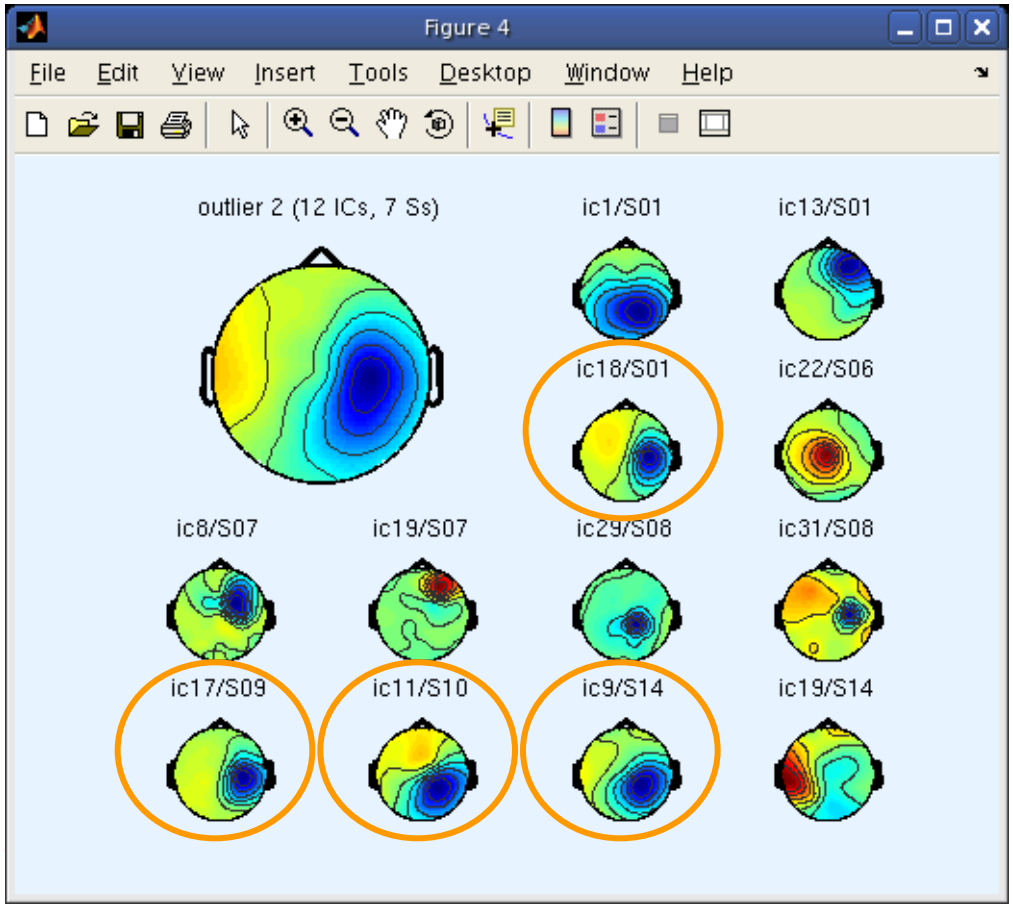
Rename a cluster

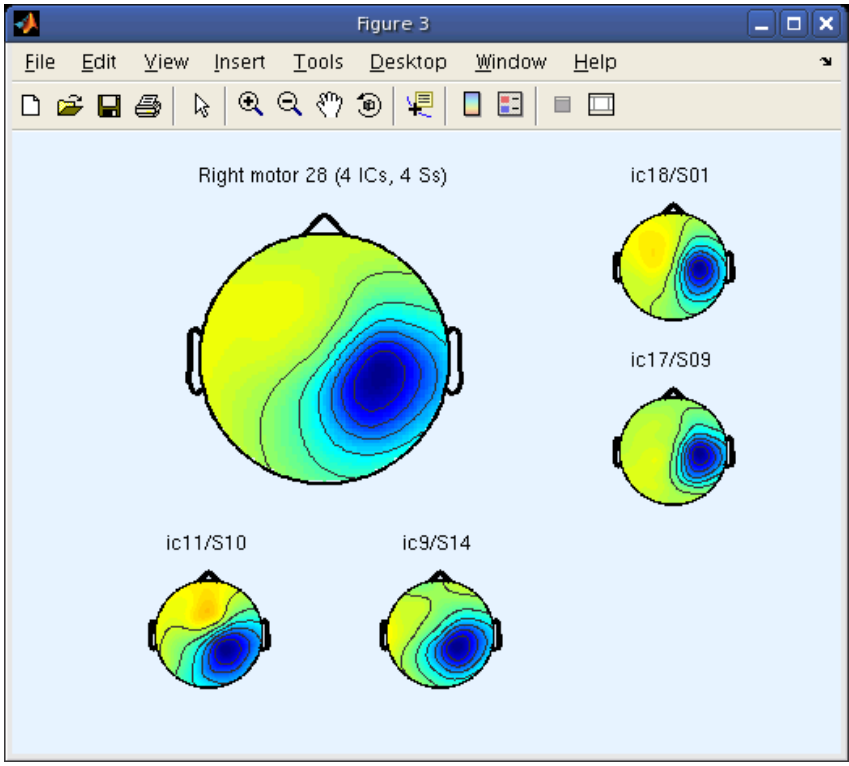


Name your cluster of interest

Create a new cluster

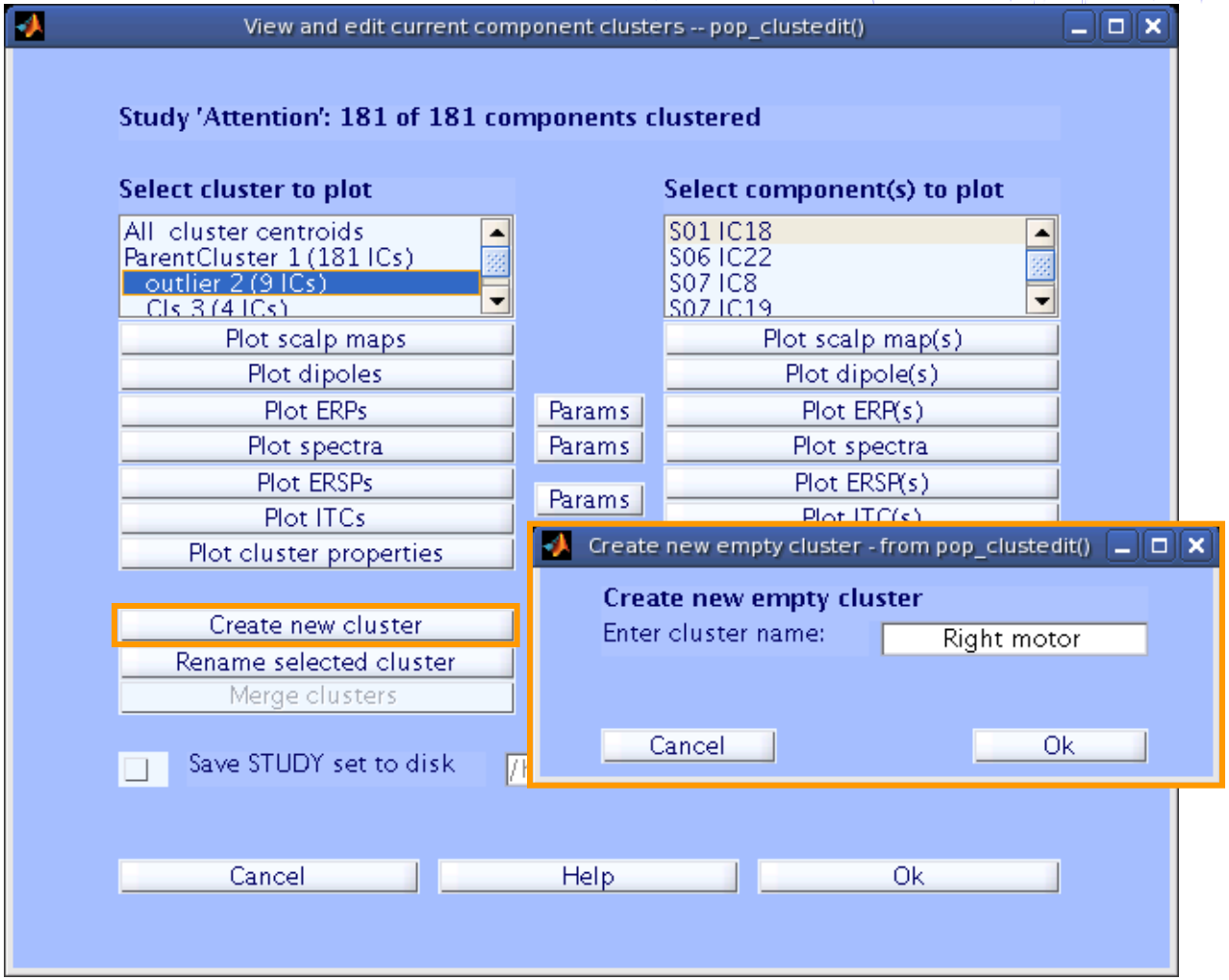
You found a bunch of 'outliers' that seem well-matched





Create a new cluster…

Create a new cluster



Create a new cluster

View and edit current component clusters -- pop_clustedit()

Study 'Attention': 181 of 181 components clustered

Select cluster to plot

All cluster centroids
ParentCluster 1 (181 ICs)
outlier 2 (9 ICs)
Cls 3 (4 ICs)

Plot scalp maps
Plot dipoles

Select component(s) to plot

S01 IC18
S06 IC22
S07 IC8
S07 IC19

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

Params
Params
Params

Reassign selected component(s)
Remove selected outlier comps.

/home/julie/Work

Help

Figure 4

File Edit View Insert Tools Desktop Window Help

outlier 2 (12 ICs, 7 Ss)
ic1/S01
ic13/S01
ic18/S01
ic22/S06
ic8/S07
ic19/S07
ic29/S08
ic31/S08
ic17/S09
ic11/S10
ic9/S14
ic19/S14

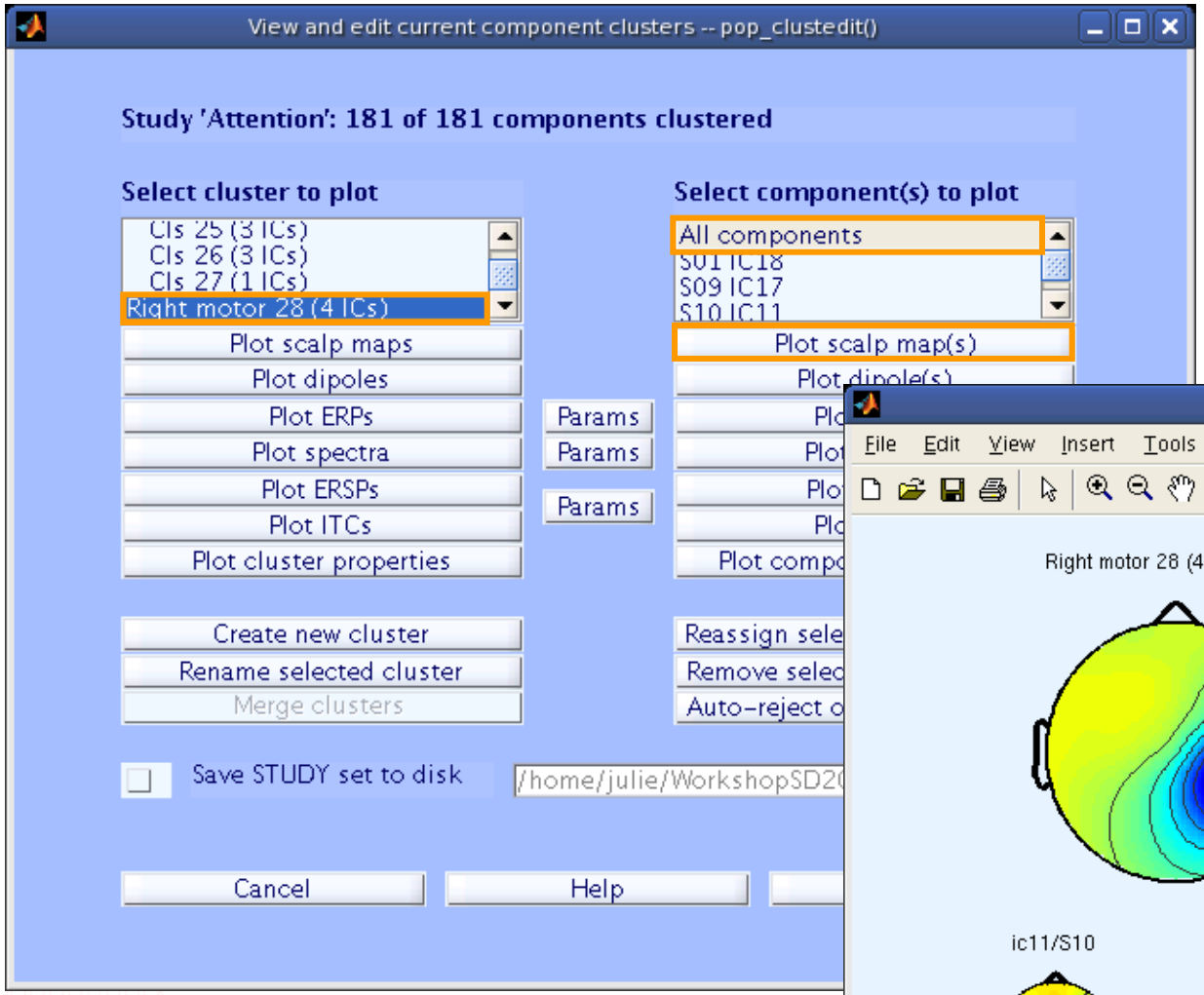
Reassign cluster - from pop_clustedit()

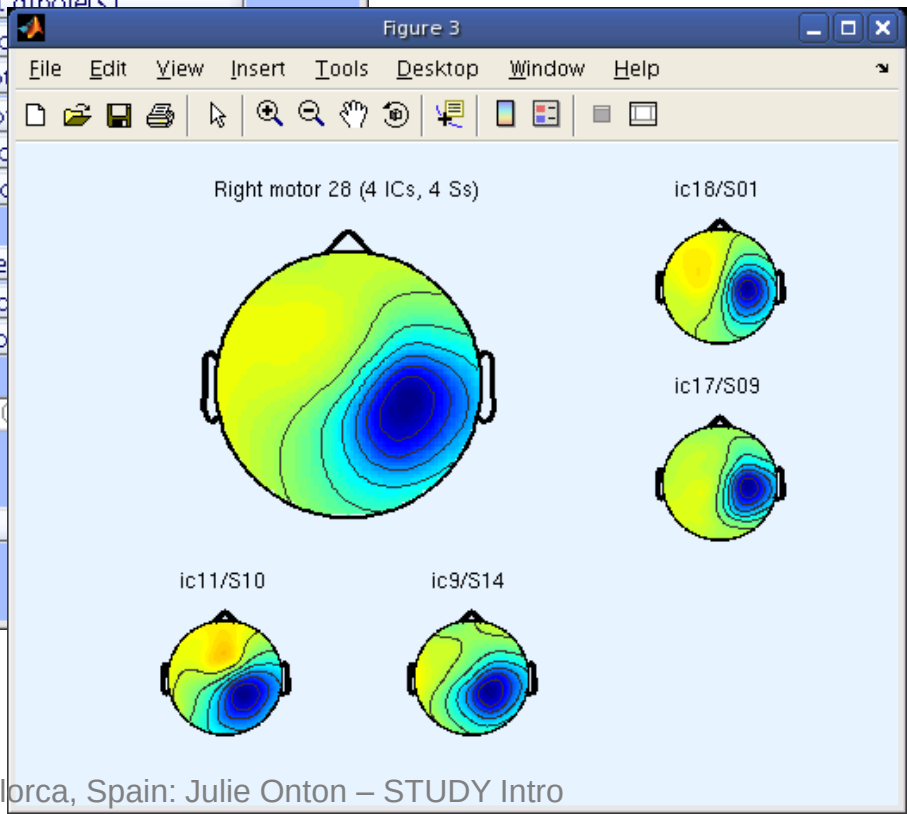
Reassign currently selected component
from outlier 2 to the cluster selected below

Cls 25
Cls 26
Cls 27
Right motor 28

Cancel
Ok

New cluster created





Exercise

- **Novice**

  -Open stern.study and practice plotting the existing clusters

  -Try removing outliers ICs or moving to another cluster

  -Rename clusters; change plotting parameters… etc

- **Intermediate / Advanced**

  - Precluster (using existing design) and cluster components using measures of your choice.

  > How does the number of dimensions and the weighting for each measure affect the results?

  - Choose a STUDY design of your choice and recompute measures (dipoles, scalp maps, ERPs are fastest; ERSPs will take several days)

  -What are your criteria for a 'good' cluster? Can you imagine a study design that would change your criteria?

  -Are all subjects in all clusters? Is this crucial?

Supplementary lessons

Plot STUDY dipoles

```
% std_diopleclusters() variables:

clusters = [3:length(STUDY.cluster)]; % clusters to plot
title = 'Cluster Dipoles'; % figure title
plot_params = [2,2,1]; % [nrows,ncols,subplot]
views = [1,2,3,4]; % 1=top,2=side,3=rear,4=oblique
cols = hsv(length(clusters));

% std_dipoleclusters function call:

std_dipoleclusters(STUDY,ALLEEG,'clusters',clusters,…
'title',title,'viewnum',views,'rowcolplace',plot_param,…
'centroid','off','colors',cols);
```

Precluster schematic

ERSP (time/freq)

Mean ERSPs

ICs (all subj)

PCA

ERSP (time/freq)

PC templates

PCs

*

PCs

PC weights

ICs (all subj)

concatenate

Power spectrum

Mean spectra

ICs (all subj)

PCA

Power spectrum

PC templates

PCs

*

PCs

PC weights

ICs (all subj)

Rel. weight

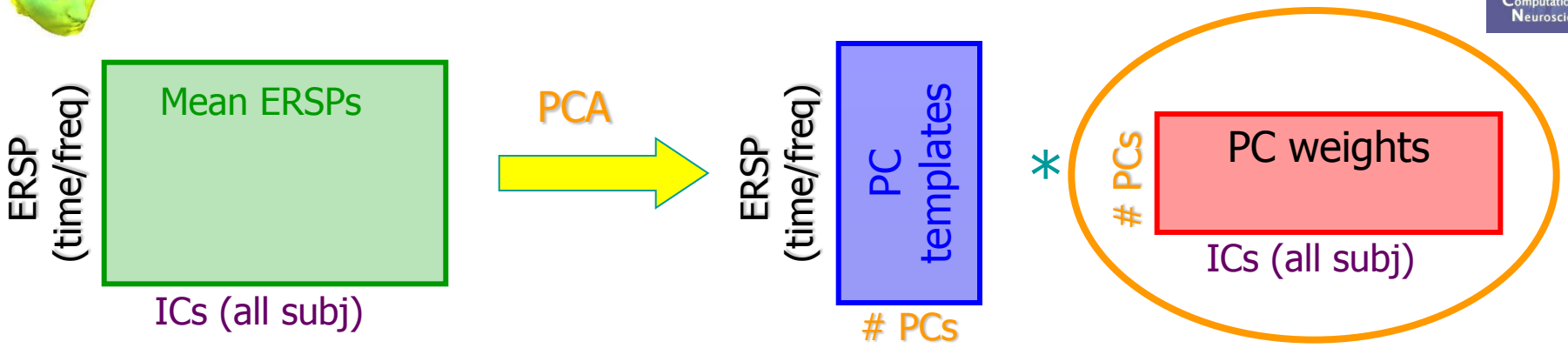
*

Dipole location

3D dipole position

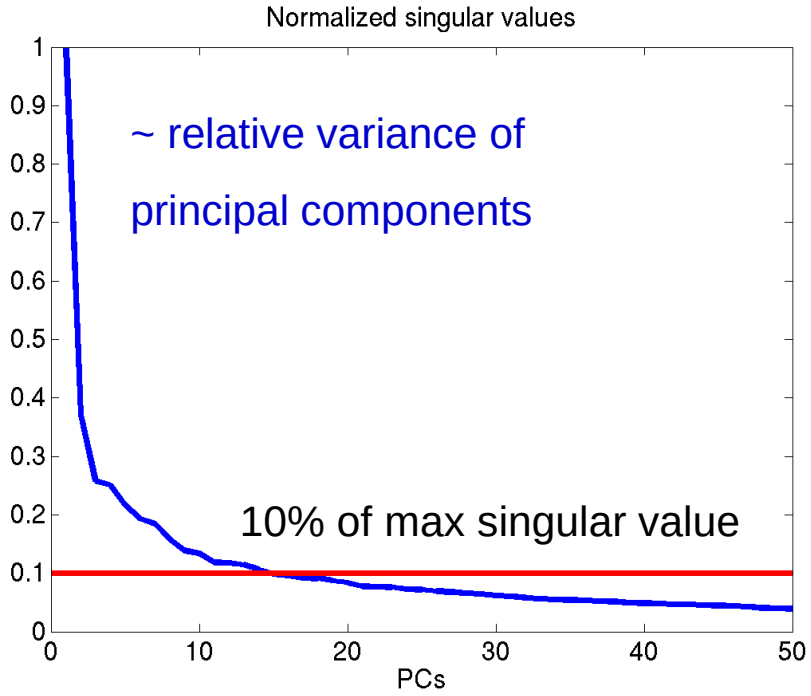
ICs (all subj)

Precluster: Use singular values from PCA

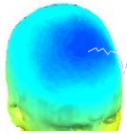


Do it yourself:

- Load all ERSP data
- Decompose with PCA
- Plot singular values



Find variance of ERSP PCA dimensions




```
% Call in raw data and run PCA, then plot singular values:
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
cond = 1; % experimental condition (corresponds to STUDY.condition)

clear logersp
for ic = 1:size(STUDY.cluster(1).sets,2)
    [logersp(:,:,ic), logfreqs, timevals, params, baseersp] = …
    std_readersp(ALLEEG, STUDY.cluster(1).sets(cond,ic), STUDY.cluster(1).comps(ic), [0 1000],[0 40] );
end;
ersp2d = reshape(logersp,size(logersp,1)*size(logersp,2),size(logersp,3));

npcs = 50; % limit the number and speed up PCA
[pc,eigvec,sv] = runpca(double(ersp2d),npcs); % decompose ERSP x ICs

% PLOT the singluar values:
figure; sbplot(2,2,1); plot(max(sv));set(gca,'xlim',[1 size(sv,1)]);
title('Raw singular values'); xlabel('PCs');

sbplot(2,2,2);plot(max(sv)/max(sv(:)));hold on;
set(gca,'xlim',[1 size(sv,1)]);
title('Normalized singular values'); xlabel('PCs');
plot([get(gca,'xlim')],[.1 .1],'r-');% line showing 10% of max

sbplot(2,2,3); imagesc(eigvec); xlabel('PCs');  title('ERSP templates');
ylabel('ERSP time/freq points');cbar;

sbplot(2,2,4); imagesc(pc); title('Component weightings');
ylabel('PCs'); xlabel('Trials'); cbar;
textsc('ERSP PCA dimensions','title');
```