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 should change when using STUDY vs single dataset
Build a STUDY

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
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: 
STUDY set key name: 
STUDY set notes:

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

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

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.)

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
ICs to cluster

Create a new STUDY set -- pop_study()

Select components

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

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
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

1x2 paired

Stim A           Stim B

2x2 unpaired

Patients          Controls

Old

Group A          Group B

Young

Group C          Group D

2x2 paired

Stim A           Stim B

2x2 paired & unpaired

Patients          Controls

Drug A

Drug B

Drug A

Drug B
Create design

1x3 design

Select study design

Subjects

Independent variable 1

- Condition: S01, S02, S03, S04, S05, S06, S07, S08, S09, S10, S11, S12, S13
- Duration
- Init_index
- Init_time
- Inset
- Load...

Independent variable 2

- None
- Condition
- Duration
- Init_index
- Init_time
- Inset

Ind. var. 1 values
- Ignore
- Memorize
- Probe

Ind. var. 2 values
- Combine selected values

Use only specific datasets/trials
- Delete all datafiles associated with this STUDY design

Save the STUDY

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
Create design

Number of event fields is unlimited

Condition:
- ignore
- memorize
- probe

Independent variable 1:
- condition
- duration
- init_index
- init_time
- load

Independent variable 2:
- None
- condition
- duration
- init_index
- init_time
- load

Create a new STUDY set -- pop_study()

Edit STUDY set Information - remember to save changes

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

<table>
<thead>
<tr>
<th>condition</th>
<th>group</th>
<th>Select by r.v.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ignore</td>
<td>Comp. 3 S 5</td>
<td>Clear</td>
</tr>
<tr>
<td>memorize</td>
<td>Comp. 3 S 5</td>
<td>Clear</td>
</tr>
<tr>
<td>probe</td>
<td>Comp. 5 S 6</td>
<td>Clear</td>
</tr>
<tr>
<td>ignore</td>
<td>Comp. 6 7</td>
<td>Clear</td>
</tr>
<tr>
<td>memorize</td>
<td>Comp. 6 7</td>
<td>Clear</td>
</tr>
<tr>
<td>probe</td>
<td>Comp. 6 7</td>
<td>Clear</td>
</tr>
</tbody>
</table>

Important note: Removed datasets will not be saved before being deleted from EEGLAB memory.
STUDY design examples

Select STUDY design
- Audio versus light all subjects
- All stimulus type - non dual subjects only
- Blank versus other stimulus type - non dual subjects only
- Audio preceded by different stimulus types
- Audio versus light across sessions - non dual subjects only
- Audio versus light across presentation - non dual subjects only

Subjects
- c1, c2, c3, c4, c5, c6, c7, c8, nd1, nd2, nd3, nd4, nd5, nd6, nd7, nd8

Independent variable 1
- None
- group
- stimulusType
- presentation
- session
- preveent

Ind. var. 1 values
- audio
- blank
- both
- light
- audio - light

Independent variable 2
- None
- group
- stimulusType
- presentation
- session
- preveent

Ind. var. 2 values
- control
- nondual

Use only specific datasets/trials
- Delete all datafiles associated with this STUDY design
- Save the STUDY

Add design
Rename design
Delete design
Combine selected values
Unpaired statistics
Cancel
Ok
# STUDY design examples

**Select STUDY design**
- Audio versus light all subjects
- Blank versus other stimulus type - non dual subjects only
- Audio preceded by different stimulus types
- Audio versus light across sessions - non dual subjects only
- Audio versus light across presentation - non dual subjects only

**Subjects**
- c1
c2
c3
c4
c5
c6
c7
- nd1
- nd2
- nd3
- nd4
- nd5
- nd6
- nd7
- nd8

**Independent variable 1**
- None
- group
- stimulusType
- presentation
- session
- preevent

**Ind. var. 1 values**
- audio
- blank
- both
- light
- audio - light

**Independent variable 2**
- None
- group
- stimulusType
- presentation
- session
- preevent

**Ind. var. 2 values**

**Options**
- Select all subjects
- Unpaired statistics
- Use only specific datasets/trials
- Delete all datafiles associated with this STUDY design
- Save the STUDY
- Add design
- Rename design
- Delete design

---

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
STUDY design examples

Select STUDY design
- Audio versus light all subjects
- All stimulus type - non dual subjects only
- Audio preceded by different stimulus types
- Audio versus light across sessions - non dual subjects only
- Audio versus light across presentation - non dual subjects only

Subjects
- c1
- c2
- c3
- c4
- c5
- c6
- c7
- nd1
- nd2
- nd3
- nd4
- nd5
- nd6
- nd7
- nd8

Independent variable 1
- None
- group
- stimulusType
- presentation
- session
- prevevent

Ind. var. 1 values
- audio
- blank
- both
- light
- audio - light

Independent variable 2
- None
- group
- stimulusType
- presentation
- session
- prevevent

Ind. var. 2 values

Use only specific datasets/trials

Delete all datafiles associated with this STUDY design

Save the STUDY

Add design
Rename design
Delete design
Combine selected values
Unpaired statistics
Cancel
Ok
## STUDY design examples

### Select STUDY design
- Audio versus light all subjects
- All stimulus type - non dual subjects only
- Blank versus other stimulus type - non dual subjects only
- Audio preceded by different stimulus types
- Audio versus light across sessions - non dual subjects only
- Audio versus light across presentation - non dual subjects only

### Subjects
- c1
- c2
- c3
- c4
- c5
- c6
- c7
- c8
- nd1
- nd2
- nd3
- nd4
- nd5
- nd6
- nd7
- nd8

### Independent variable 1
- None
- group
- stimulusType
- presentation
- session
- presevent

### Ind. var. 1 values
- audio
- blank
- both
- light

### Independent variable 2
- None
- group
- stimulusType
- presentation
- session
- presevent

### Ind. var. 2 values
STUDY design examples
STUDY design examples

Select STUDY design
Audio versus light all subjects
All stimulus type - non dual subjects only
Blank versus other stimulus type - non dual subjects only
Audio preceded by different stimulus types
Audio versus light across sessions - non dual subjects only
Audio versus light across presentation - non dual subjects only

Subjects
- c1
- c2
- c3
- c4
- c5
- c6
- c7
- nd1
- nd2
- nd3
- nd4
- nd5
- nd6
- nd7
- nd8

Independent variable 1
- None
group
stimulusType
presentation
session
prevevent

Ind. var. 1 values
- audio
- blank
- both
- light
- audio - light

Combine selected values
Unpaired statistics

Independent variable 2
- None
group
stimulusType
presentation
session
prevevent

Ind. var. 2 values
- evoked
- spontaneous

Combine selected values
Unpaired statistics

Use only specific datasets/trials
Delete all datfiles associated with this STUDY design
Save the STUDY

Add design
Rename design
Delete design

Cancel Ok
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 component measures

- Build preclustering array
- Cluster components
- Edit/plot clusters

Study filename: [Field]
Study task name: [Field]
Nb of subjects: [Field]
Nb of conditions: [Field]
Nb of sessions: [Field]
Nb of groups: [Field]
Epoch consistency: yes
Channels per frame: 31
Channel locations: yes
Clusters: 1
Status: Ready to precluster
Total size (Mb): 30.4
TIP: Compute all measures so you can test different combinations for preclustering later.

```
[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

The image shows a screenshot of the EEGLAB v6.0b software. In the screenshot, the window is open for the 'STUDY set: At...' study. The study information includes:

- **Study filename:** [Study info is not visible in the screenshot]
- **Study task name:** [Task name is not visible in the screenshot]
- **Nb of subjects:** [Information is not visible in the screenshot]
- **Nb of conditions:** [Information is not visible in the screenshot]
- **Nb of sessions:** [Information is not visible in the screenshot]
- **Nb of groups:** [Information is not visible in the screenshot]
- **Epoch consistency:** Yes
- **Channels per frame:** 31
- **Channel locations:** Yes
- **Clusters:** 1
- **Status:** Pre-clustered
- **Total size (Mb):** 32.4

Under the 'Study' tab, there are options to:

- Edit study info
- Precompute channel measures
- Plot channel measures
- Precompute component measures
- Build preclustering array
- Cluster components
- Edit/plot clusters
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,'frequrange',[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,...
    'frequrange',[3 30], 'timewindow',[0 600], 'norm',1,'weight',1},{'itc',...
    'npca',6,'frequrange',[3 30], 'timewindow',[0 400], 'norm',1, 'weight',1});
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…

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
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.
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);
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

Study name: 'Sternberg' (336 of 336 components clustered)

Select cluster to plot

- All cluster centroids
- ParentCluster 1 (336 ICs)
  - Cls 2 (17 ICs)
  - Cls 3 (6 ICs)
  - Plot scalp maps
  - Plot dipoles
  - Plot ERP(s)
  - Plot spectra
  - Plot ERSPs
  - Plot ITCs
  - Plot cluster properties

Select component(s) to plot

- 'Cls 2' comp. 1 (S01 IC21)
  - 'Cls 2' comp. 2 (S03 IC21)
  - 'Cls 2' comp. 3 (S03 IC25)
  - 'Cls 2' comp. 4 (S04 IC19)
  - Plot scalp map(s)
  - Plot dipole(s)
  - Plot ERP(s)
  - Plot spectra
  - Plot ERSP(s)
  - Plot ITC(s)
  - Plot component properties

- Params

Create new cluster
- Rename selected cluster
- Merge clusters

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

Help

Cancel

Ok
Plot cluster data

Plot mean scalp maps for easy reference
Plot cluster data

Choose which components

Choose a cluster
Plot clusters

Study name: 'Sternberg' (336 of 336 components clustered)

Select cluster to plot

Cls 11 (17 ICs)
Cls 12 (13 ICs)

Select component(s) to plot

All components
S01 IC13
S01 IC20
S02 IC5

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

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

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
Plot ERPs

[Image of EEGlab software interface showing options for selecting clusters and components to plot, along with parameters for setting plot characteristics such as time range, plot map latency, statistical methods, and threshold values.]
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

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
Create new cluster
Rename selected cluster
Merge clusters

Set ERSP|ITC plotting parameters -- pop_erspparams()

- Time range in ms [Low High]: -200 1200
- Freq. range in Hz [Low High]: 3 30
- Power limits in dB [Low High]:
- ITC limit (0-1) [High]:
- Compute ERSP baseline across conditions
- Statistical method to use: Permutation
- Statistical threshold (p<)

Options:
- Compute condition statistics
- Compute group statistics
- Mask non-significant data (only when threshold is set)
- Use False Discovery Rate to correct for multiple comparisons

Help Cancel Ok
Plot cluster ERSPs and ITC

Not masked for significance

ERSP

ITC

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
Remove outlier components
Remove outlier components

Study " : 151 of 151 components clustered

Select cluster to plot

C13 (5 ICs)
C14 (11 ICs)
C15 (8 ICs)
C16 (6 ICs)
C17 (4 ICs)

Plot scalp maps
Plot dipoles
Plot ERPs
Plot spectra
Plot ERSPs
Plot ITCs
Plot cluster properties

Select component(s) to plot

All components
S07 IC14
S07 IC33
S08 IC23

$10$ IC60

Remove outliers - from pop_clustedit()

Remove currently selected component below from C1s 17 to its outlier cluster?

$10$ IC60

Cancel
Ok

Remove selected outlier comps.
Auto-reject outlier components

Save STUDY set to disk

Canceled

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

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
Remove outlier components

Study: 151 of 151 components clustered

Select cluster to plot:
- Cls 16 (6 ICs)
- Cls 17 (3 ICs)
- Cls 18 (1 ICs)
- Cls 19 (14 ICs)
- Outliers Cls 17 20 (1 ICs)

Select component(s) to plot:
- All components
- S10 IC60

Options:
- Create new cluster
- Rename selected cluster
- Merge clusters

Save STUDY set to disk: /home/julie/workshop06/5subjects/WStudy/study

Options:
- Cancel
- Help
- Ok
Reassign component

If you want to manually reassign a component to another cluster...
Reassign component

Study 'Attention': 181 of 181 components clustered

Select cluster to plot

<table>
<thead>
<tr>
<th>Cls 18 (4 ICs)</th>
<th>Cls 19 (10 ICs)</th>
<th>Cls 20 (17 ICs)</th>
<th>Cls 21 (6 ICs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot scalp maps</td>
<td>Plot dipoles</td>
<td>Plot ERPs</td>
<td>Plot ITCs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Select component(s) to plot

<table>
<thead>
<tr>
<th>All components</th>
<th>S02 IC2</th>
<th>S02 IC4</th>
<th>S03 IC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot scalp map(s)</td>
<td>Plot dipole(s)</td>
<td>Plot ERP(s)</td>
<td>Plot ITCs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reassign currently selected component from Cls 19 to the cluster selected below

- Cls 4
- Cls 5
- Cls 6
- Cls 7

Reassign selected component(s)

- Remove selected outlier comps.
- Auto-reject outlier components

Save STUDY set to disk: /home/julie/WorkshopSD2007/STUDY

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
Reassign component

Successful reassignment
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 new empty cluster - from pop_clustedit()

Enter cluster name: Right motor
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?
% 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_diopleclusters function call:

std_diopleclusters(STUDY,ALLEEG,'clusters',clusters,...
        'title',title,'viewnum',views,'rowcolplace',plot_param,...
        'centroid','off','colors',cols);
Precluster: Use singular values from PCA

Mean ERSPs

ICs (all subj)

**PCA**

ERSP (time/freq)

**PC templates**

# PCs

ICs (all subj)

PC weights

Do it yourself:
- Load all ERSP data
- Decompose with PCA
- Plot singular values

Normalized singular values

~ relative variance of principal components

10% of max singular value

Novice EEGLAB Workshop, Sept 22, 2011, Mallorca, Spain: Julie Onton – STUDY Intro
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 singular 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');