


Linear Modelling of EEG data

Theory and practice

Cyril R. Pernet, University of Edinburgh

LIMO EEG:

https://gforge.dcn.ed.ac.uk/gf/project/limo_eeg/



Division of Clinical Neurosciences
SFC Brain Imaging Research Centre

HomeMy StuffUsersSearchProjectsSnippets


LInear MOdelling of EEG data

- Summary
- Admin
- Reporting
- Search
- Tracker
- Files
- SVN

Home » Projects » LInear MOdelling of EEG data » Home

: LIMO EEG V1.4 now available for download

- 1st level analysis changed allowing factorial and full factorial and bootstrap
- LIMO Batch updated - process all subjects easily and automatically
- Threshold Free Cluster Enhancement now available along side Spatial-Temporal clustering
- 2nd level t-tests are now robust, i.e. rely on trimmed means and winsorized variances




Time	Activity Type	By
2013-Jun-14		
23:39:38	Tracker item "limo_ecluster_make" deleted	Cyril R. Pernet
2013-Jun-04		
10:11:41	Tracker item "limo_design_matrix" opened	Benedikt Ehinger
10:03:52	Tracker item "Bugs in limo_import_chanloc_import" opened	Benedikt Ehinger

Working on v1.5 already to get all 2nd level analyses robust

Preparing major upgrade for ERSF

General Sections

Tracker

 [Bug and Support](#) (2 Open / 2 Total)

Use this to report problems in LIMO EEG

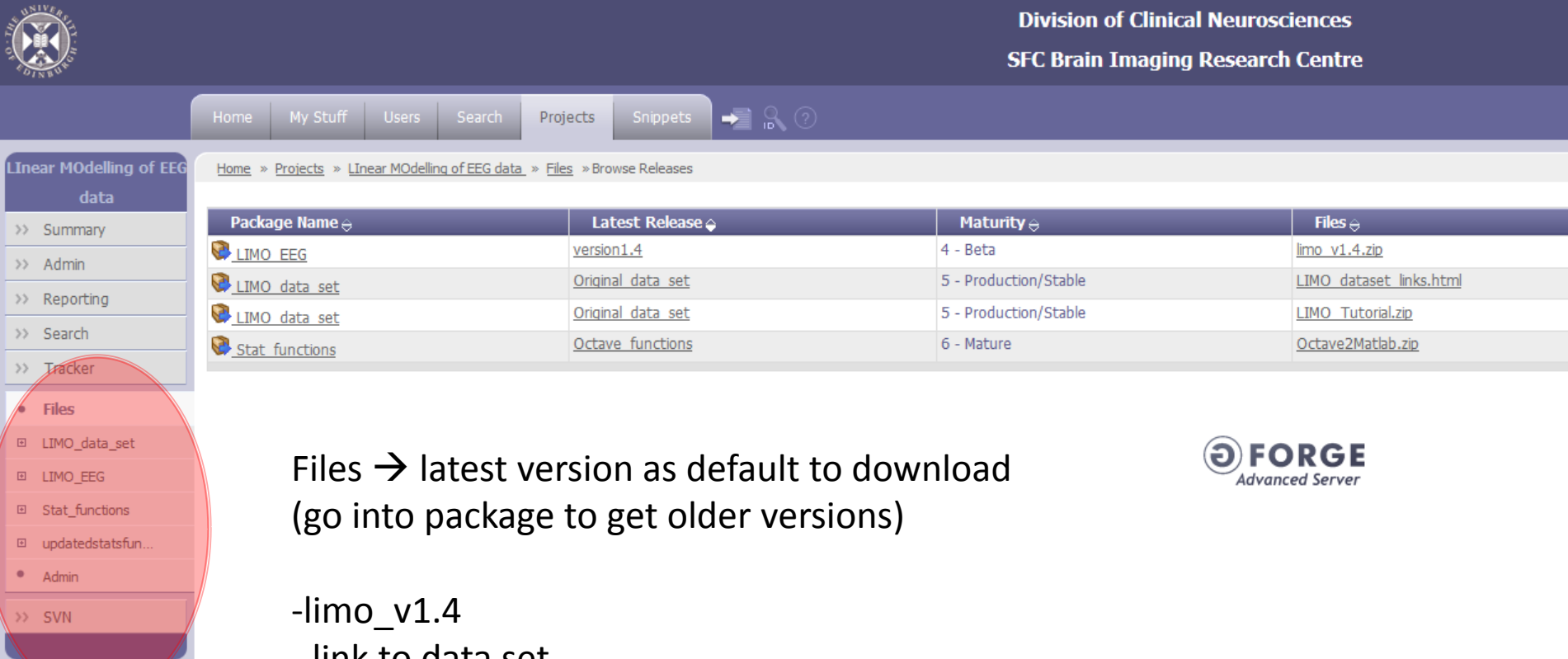
Please cite us when you use our tool. Ref: Pernet, C.R., Chauveau, N., Gaspar, C. & Rousselet, G.A. *LIMO EEG: a toolbox for hierarchical Linear Modeling of ElectroEncephaloGraphic data*. Computational Intelligence and Neuroscience, Volume 2011 (2011), Article ID 831409, 11 pages, doi:10.1155/2011/831409

You can also download our [OHBM 2011 poster](#) for a quick overview

Make sure you have the latest version of [EEGLAB](#) to benefit of the latest features compatible with LIMO EEG.

LIMO EEG:

https://gforge.dcn.ed.ac.uk/gf/project/limo_eeg/



The screenshot displays the GForge project page for LIMO EEG. The header includes the University of Edinburgh logo and the Division of Clinical Neurosciences SFC Brain Imaging Research Centre. The navigation bar shows 'Home', 'My Stuff', 'Users', 'Search', 'Projects', and 'Snippets'. The breadcrumb trail is 'Home » Projects » LLinear MOdelling of EEG data » Files » Browse Releases'. The main table lists the package name, latest release, maturity, and files.

Package Name	Latest Release	Maturity	Files
LIMO_EEG	version1.4	4 - Beta	limo_v1.4.zip
LIMO_data_set	Original data set	5 - Production/Stable	LIMO_dataset_links.html
LIMO_data_set	Original data set	5 - Production/Stable	LIMO_Tutorial.zip
Stat_functions	Octave functions	6 - Mature	Octave2Matlab.zip

The left sidebar shows the project structure with 'Files' highlighted. The files listed are LIMO_data_set, LIMO_EEG, Stat_functions, and updatedstatsfun... The 'Files' section is circled in red.


Files → latest version as default to download
(go into package to get older versions)

- limo_v1.4
- link to data set
- Stat functions if no stat toolbox



LIMO EEG:

https://gforge.dcn.ed.ac.uk/gf/project/limo_eeg/



Division of Clinical Neurosciences
SFC Brain Imaging Research Centre

HomeMy StuffUsersSearchProjectsSnippets


Linear Modelling of EEG data

- Summary
- >> Admin
- >> Reporting
- >> Search
- >> Tracker
- >> Files
- >> SVN

[Home](#) » [Projects](#) » [Linear Modelling of EEG data](#) » [Home](#)


: LIMO EEG V1.4 now available for download

- 1st level analysis changed allowing factorial and full factorial and bootstrap
- LIMO Batch updated - process all subjects easily and automatically
- Threshold Free Cluster Enhancement now available along side Spatial-Temporal clustering
- 2nd level t-tests are now robust, i.e. rely on trimmed means and winsorized variances



Time	Activity Type	By
2013-Jun-14		
23:39:38	Tracker item "limo_ecluster_make" deleted	Cyril R. Pernet
2013-Jun-04		
10:11:41	Tracker item "limo_design_matrix" opened	Benedikt Ehinger
10:03:52	Tracker item "Bugs in limo_import_chanloc_import" opened	Benedikt Ehinger

Working on v1.5 already to get all 2nd level analyses robust
Preparing major upgrade for ERSF

General Sections
Tracker
 [Bug and Support](#) (2 Open / 2 Total)
Use this to report problems in LIMO EEG

Please cite us when you use our tool. Ref: *Pernet, C.R., Chauveau, N., Gaspar, C. & Rousselet, G.A. LIMO EEG: a toolbox for hierarchical Linear Modeling of ElectroEncephaloGraphic data. Computational Intelligence and Neuroscience, Volume 2011 (2011), Article ID 831409, 11 pages, doi:10.1155/2011/831409*

You can also download our [OHBM 2011 poster](#) for a quick overview

Make sure you have the latest version of [EEGLAB](#) to benefit of the latest features compatible with LIMO EEG.

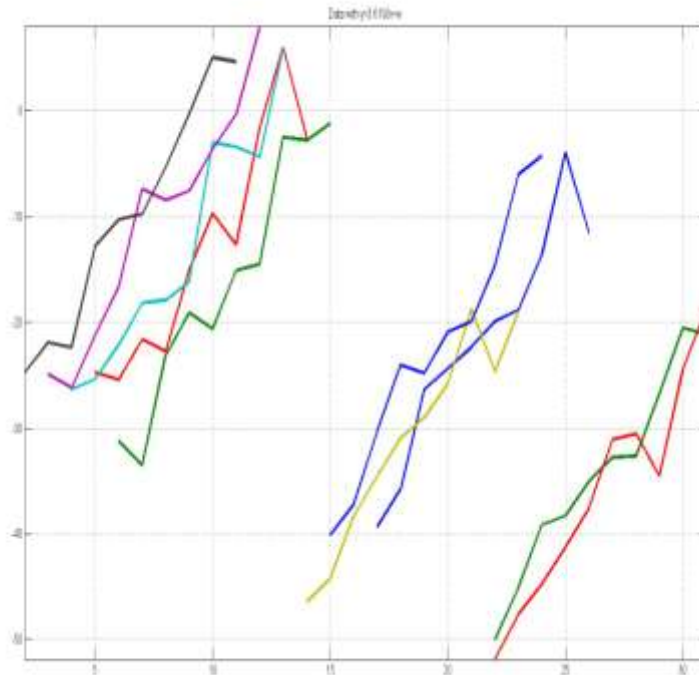
Linear Modelling

- **Framework:** Linear mixed effects through a hierarchical model (data within subjects, subjects within groups)
- **Statistics:** At the subject level, GLM (OLS) for each electrode and time frames separately (=inter trials variance). At the group level, bootstrapped robust tests (=inter subject variance, assumption free).
- **Multiple comparisons correction:** bootstrap used to evaluate the clustering of statistical tests.

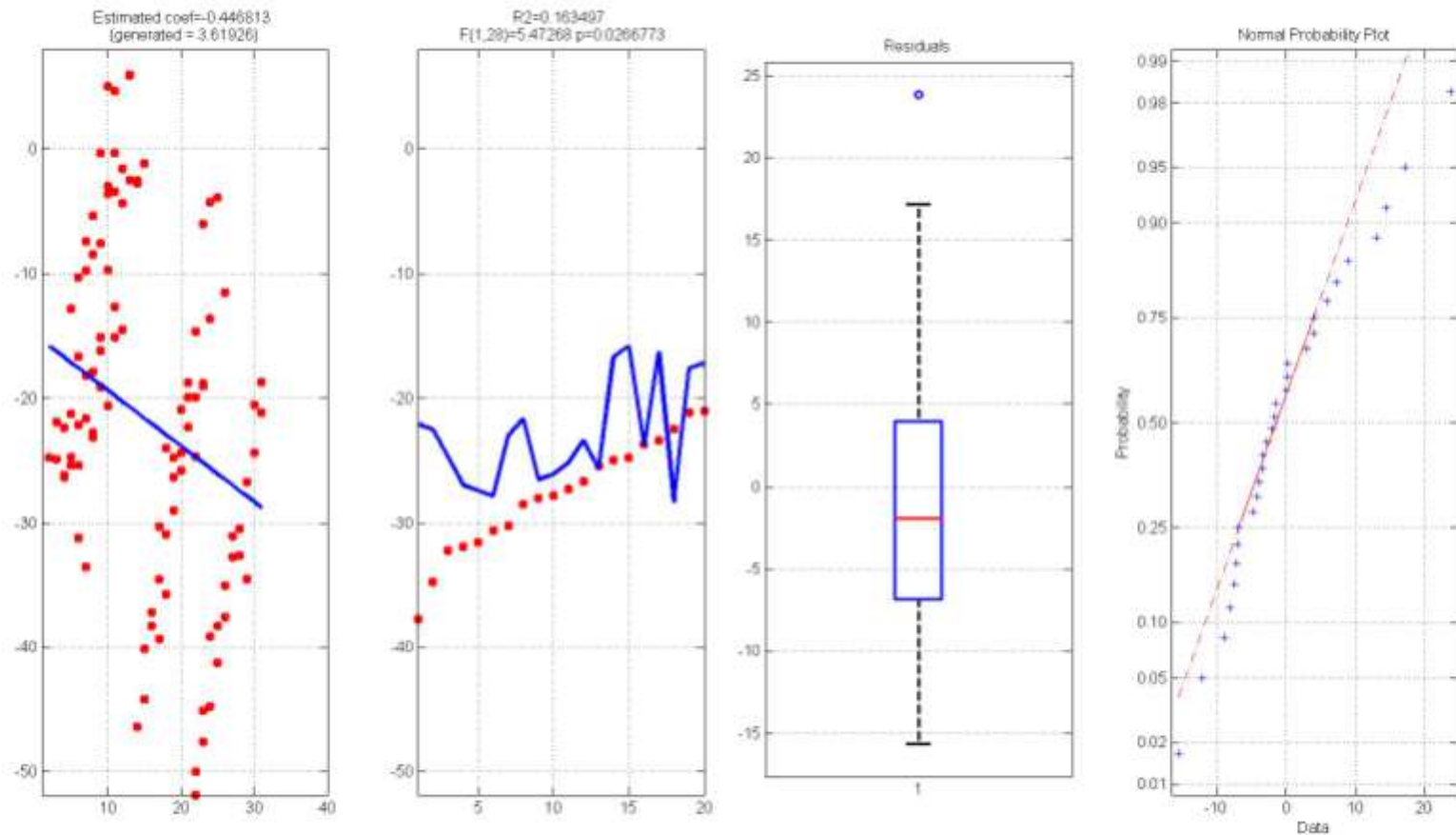
Mixed effect Model

Model the data with fixed effects (the experimental conditions) and a random effect (subjects are allowed to have different overall values – considering subjects as a random variable)

Example: present stimuli from intensity -5 units to +5 units around the subject perceptual threshold and measure RT
→ Plot the data per intensity

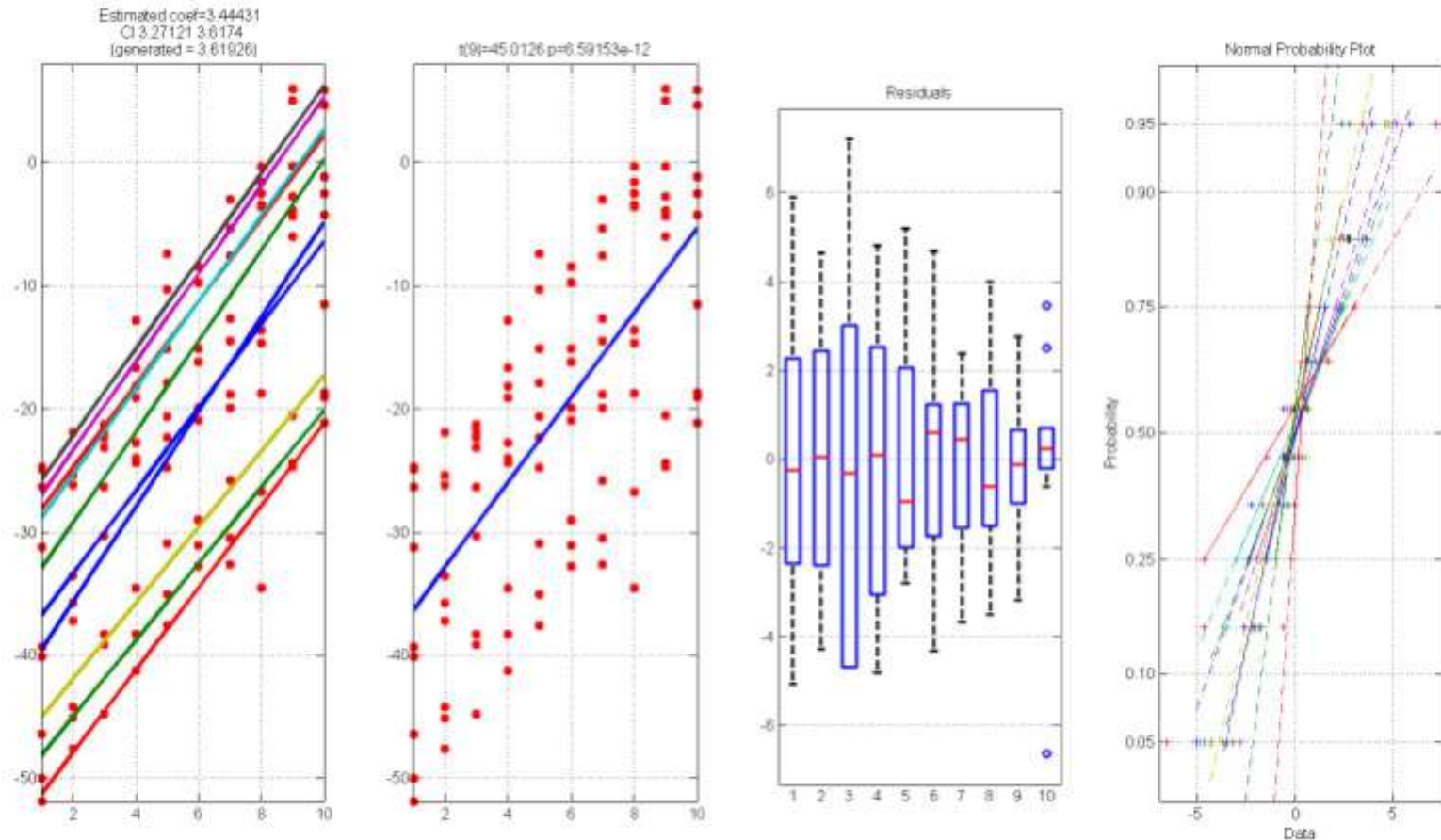


Mixed effect Model



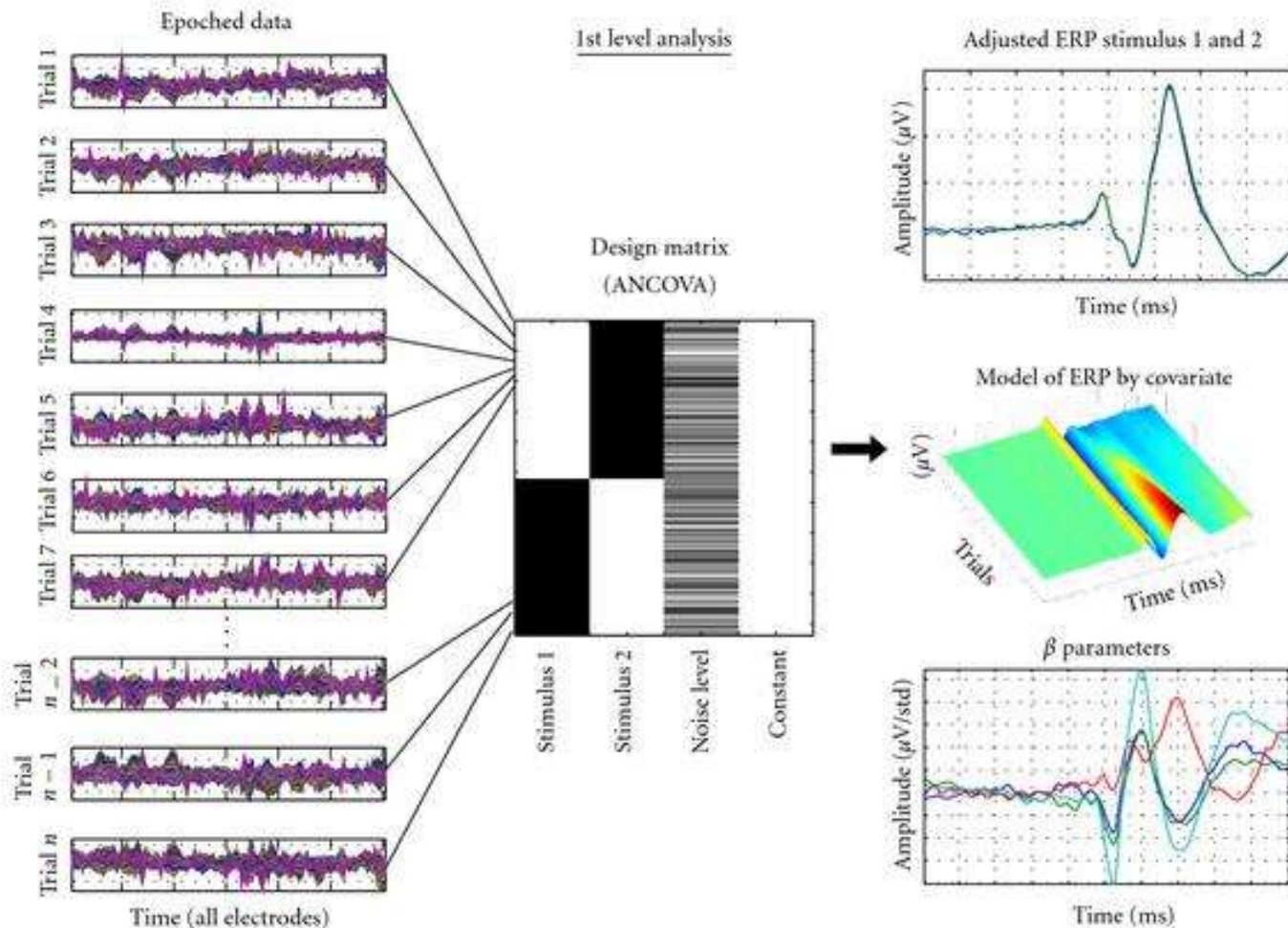
Fixed effect = average across subjects → negative correlation?

Mixed effect Model



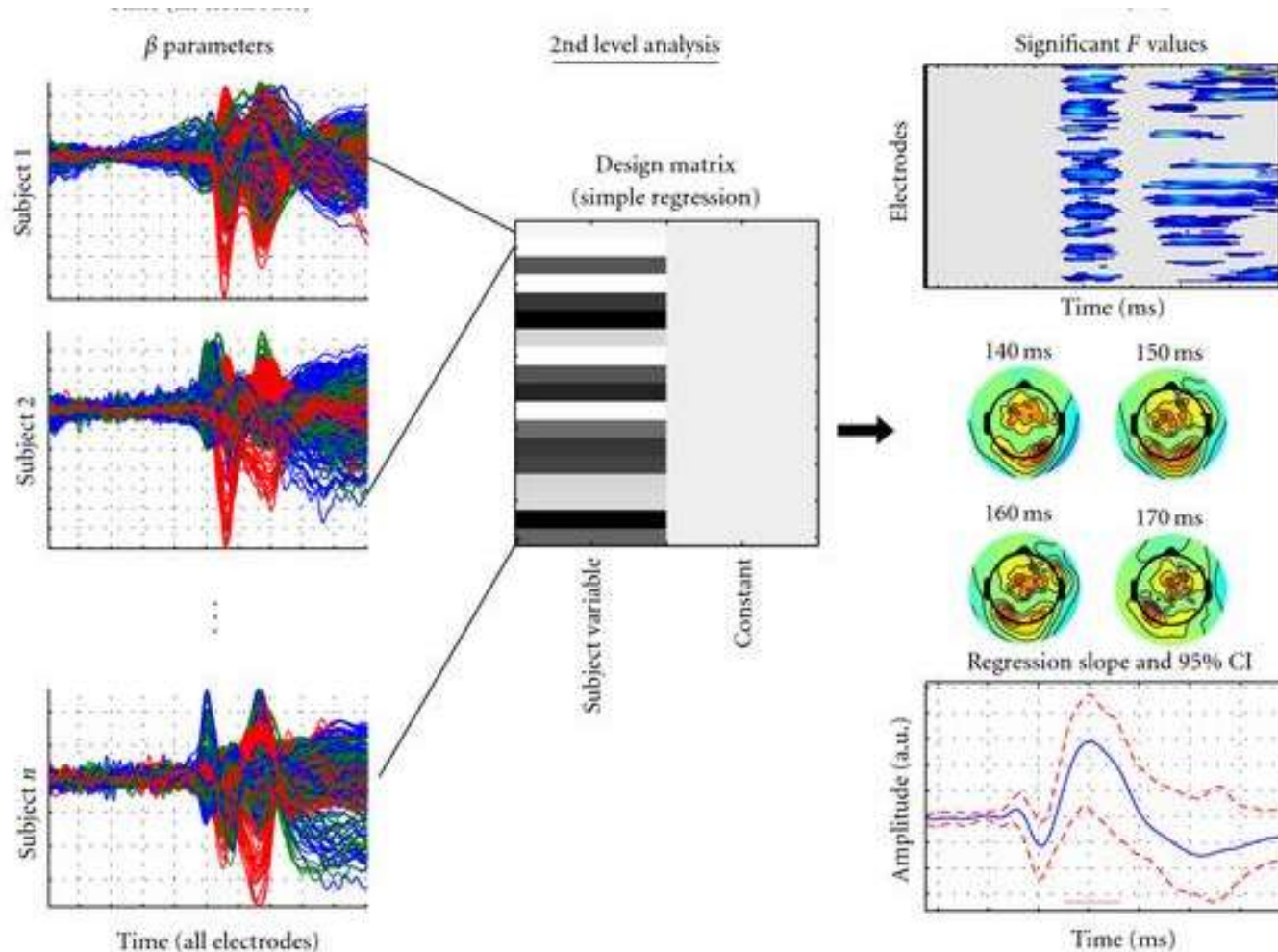
Mixed effect = effect per subject with variable (random) offsets

Hierarchical Model



1st level = all trials – within subject variance

Hierarchical Model



2nd level = all subjects – between subjects variance

General Linear Model

- Linear: the output is a function of the input satisfying rules of scaling and additivity (e.g $RT = 3 \cdot \text{acuity} + 2 \cdot \text{vigilance} + 4 + e$)
- Model: assign to the data different effects / conditions ... All we have to do is find the parameters of this model
- General: applies to any known linear statistics (ttest, ANOVA, Regression, MANCOVA), can be adapted to be robust (ordinary least squares vs. weighted least squares), and can even be extended to non Gaussian data (Generalized Linear Model using link functions)

For those who like maths: $Y = BX + e$ and thus $B = \text{inv}(X'X)X'Y \rightarrow \text{pinv}(X)Y$
 $R = I - (X \cdot \text{pinv}(X))$ and thus $\text{var} = RY' \cdot RY / \text{dfe}$

The weighted least square proceeds with $B = \text{inv}(X'WX)X'WY$

1st level model in practice

General Linear Model: Practice 1

- Processing of subject 1.

The screenshot shows the 'limo_import' window titled 'LINEAR MODELING TOOLBOX: data import'. The interface is divided into three main sections: 'Import', 'Specify', and 'Analysis'.

Import Section:

- A text box labeled 'Import data set'.
- Two input fields for 'Start to analyze at (sec)' and 'End the analysis at (sec)'.

Specify Section:

- A text box labeled 'Specify the categorical variable'.
- A checkbox labeled 'full factorial'.
- A text box labeled 'Specify continuous variables'.
- A checkbox labeled 'do not z-score regressors'.

Analysis Section:

- Two dropdown menus: the first is set to 'Mass-univariate' and the second to 'OLS'.
- Two checkboxes: 'Bootstrap data' and 'add TFCE'.

Buttons:

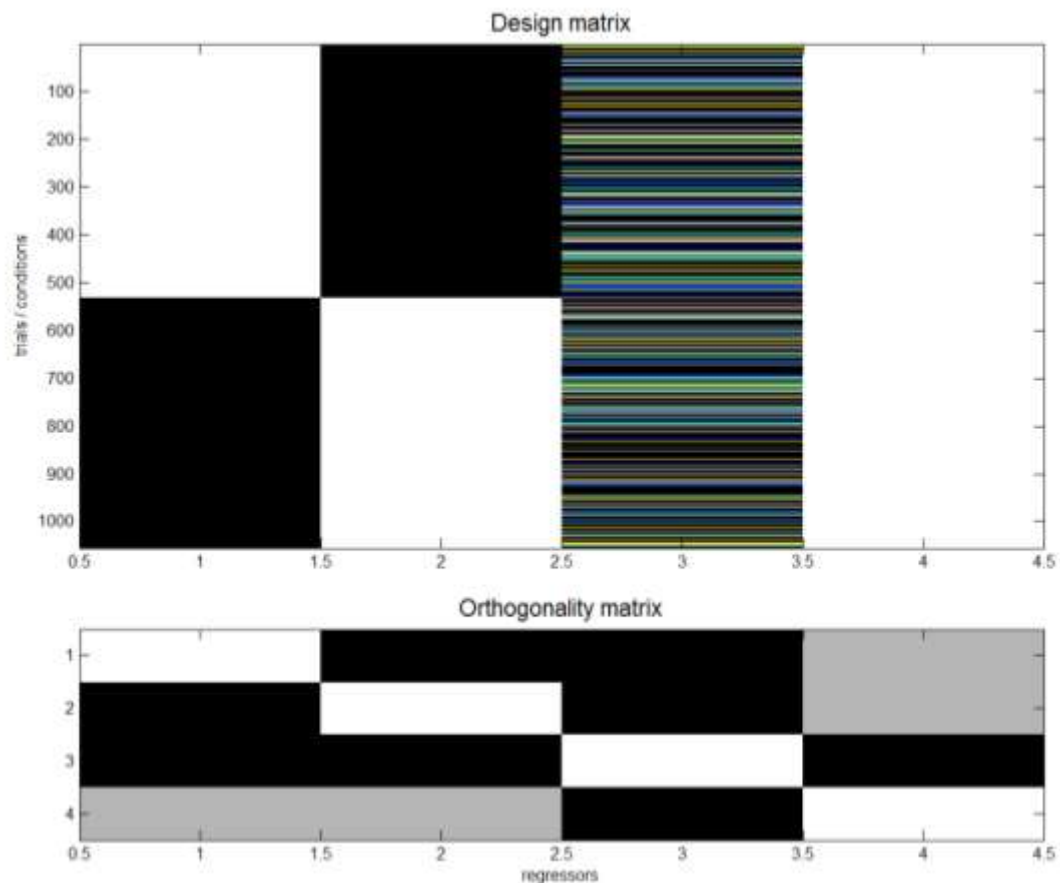
- 'Help', 'Working Directory', 'Done', and 'Quit' buttons are located at the bottom.

Footer:

Copyright (C) LIMO Team 2010 - GNU GPL

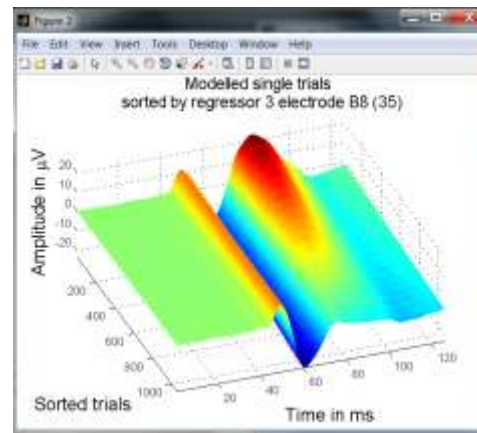
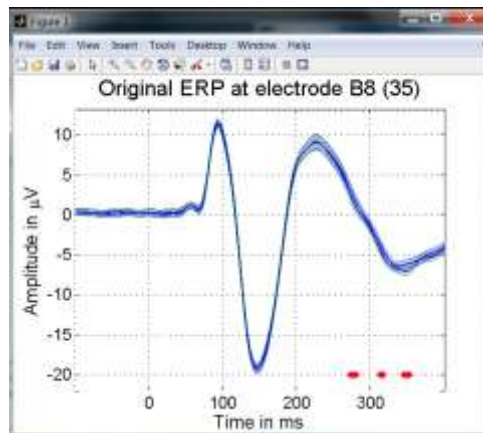
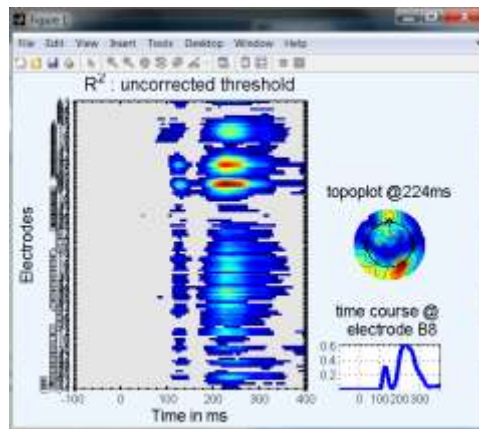
General Linear Model: Practice 1

- Design matrix: accept if you think it looks like what the model should be



General Linear Model: Practice 1

- Results (non corrected for multiple comparisons)



General Linear Model

- Gauss-Markov theorem: OLS estimates are the best linear unbiased estimator (BLUE) of the population parameters, providing
- A. regressors in the design matrix X are independent from the error, non-stochastic and known
- B. regressors are independent from each other.
- C. errors are independently and identically distributed $\sim N(0, s^2 I)$
- Your job as users to create experiments satisfying conditions A and B, our job as programmers to 'ensure' C

General Linear Model

- Errors should be i.i.d.
 - C1. Errors from different observations are not correlated ($\text{Cov}(e_i, e_j) = 0$)
 - C2. The expected value of the error term is zero ($E(e) = 0$)
 - = if C1 and C2 not satisfied, beta estimates are not BLUE
- C3. The variance of the error is s^2 at all observations. That is the off-diag element of the covariance matrix are 0 i.e. errors are homoscedastic.
- = if C3 not satisfied, t/F values do not follow known distributions

2nd level model

2nd level analysis

- Usually assumes 1st level variances are homoscedastic (impact statistics), but taken care of by the bootstrap.

- Almost all designs available:

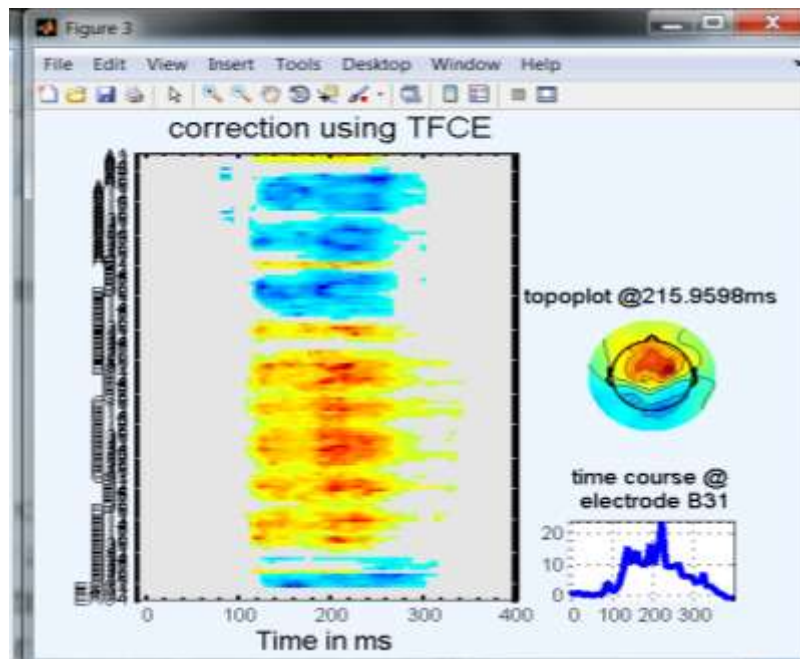
T-tests (robust)

Regression, ANOVA, ANCOVA

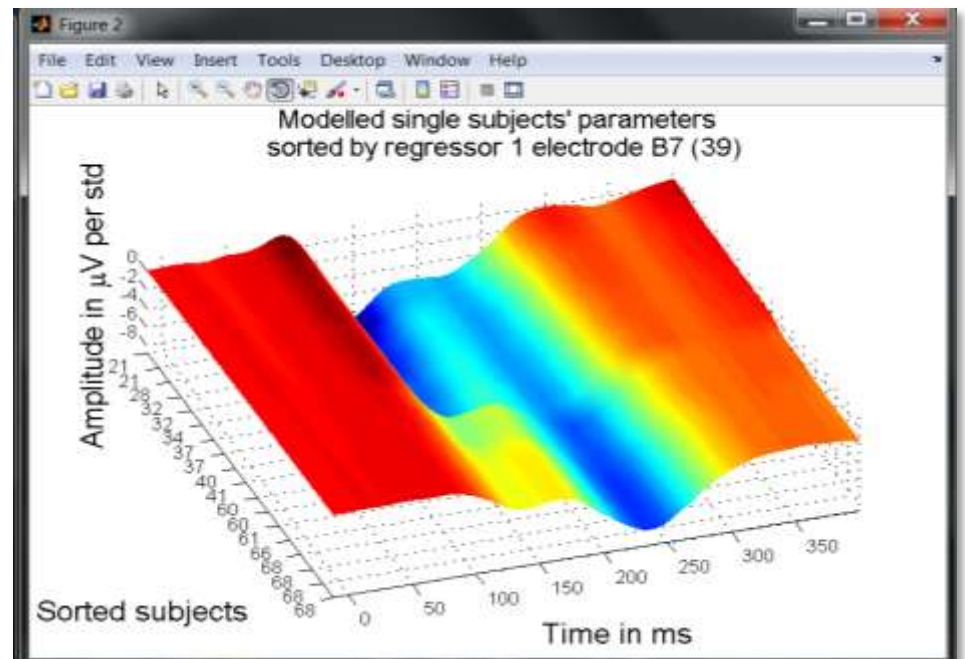
Repeated Measures ANOVA (Hotelling T²)

General Linear Model: Practice 2

One-sample t-test



Regression



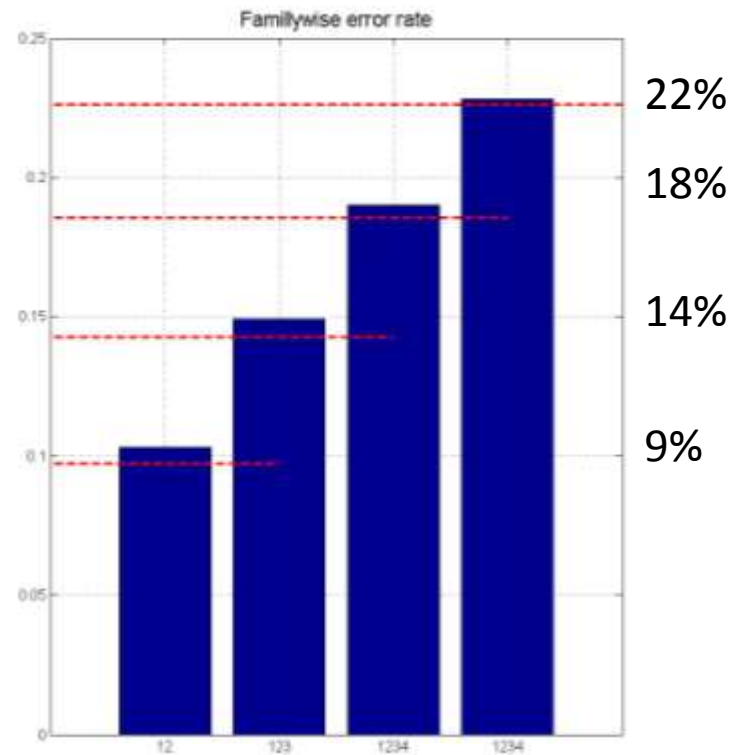
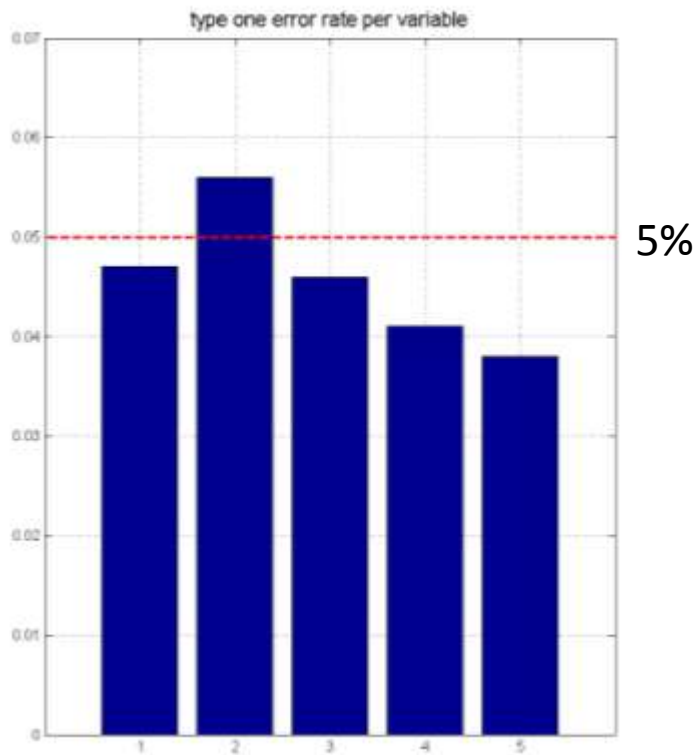
Multiple Comparisons Correction

Multiple comparisons correction

- When we do a statistical test, we set alpha, the probability to reject H_0 (under H_0) – this is also known as type I error rate
- The familywise error rate has to do with the number of tests performed, and assuming tests are independent from each other, the $\text{FWER} = 1 - (1 - \alpha)^n$
- eg. so for $\alpha = 5/100$, if we do 2 tests we should get about $1 - (1 - 5/100)^2 \sim 9\%$ false positives, if we do 126 electrodes * 150 time frames tests, we should get about $1 - (1 - 5/100)^{18900} \sim 100\%$ false positives! i.e. **you can't be certain of any of the statistical results you observe**

Multiple comparisons correction

- Illustration with 5 independent variables from $N(0,1)$
- Repeat 1000 times and measures type 1 error rate



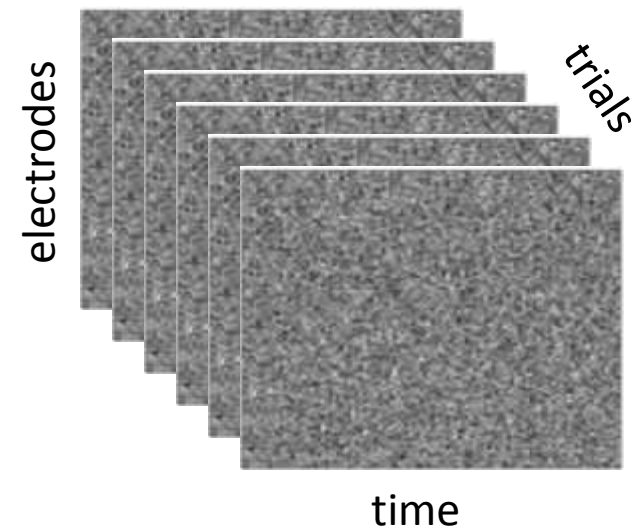
Multiple comparisons correction

- Illustration with 2 variables with Pearson's $r=0:1$

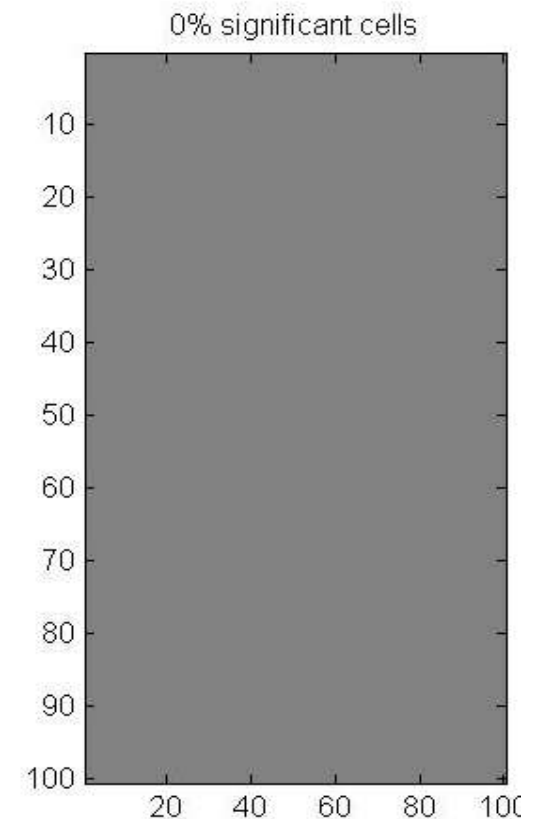
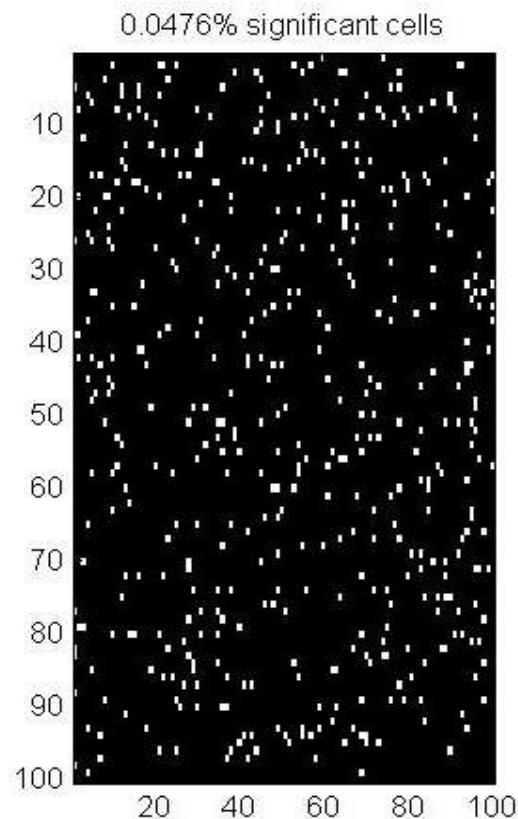


Multiple comparisons correction

- Bonferroni correction allows to keep the FWER at 5% by simply dividing alpha by the number of tests

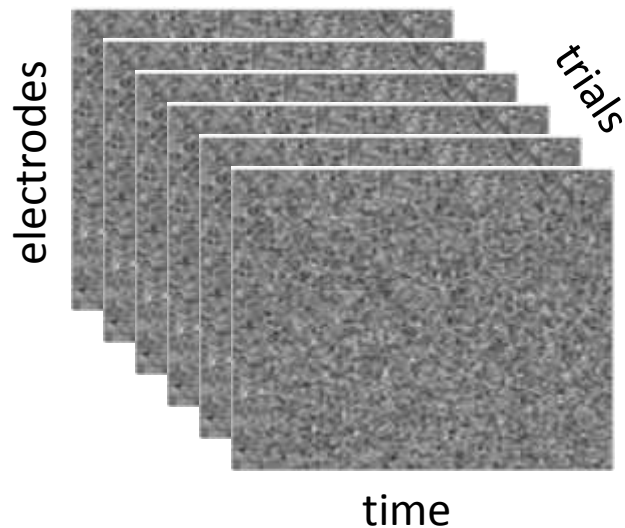


One sample t test > 0 ?

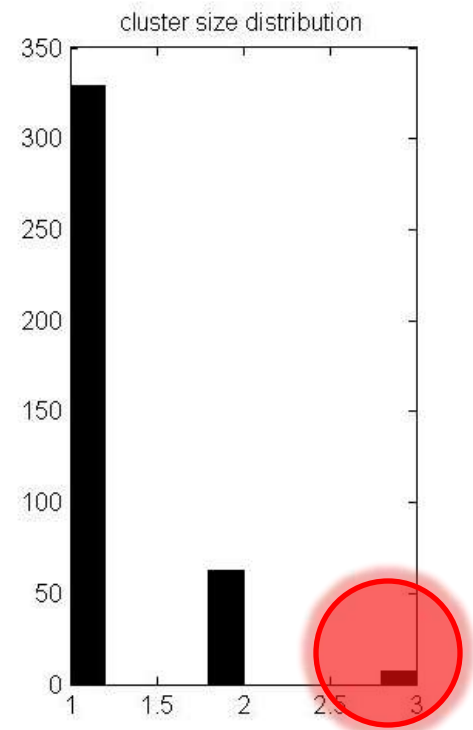
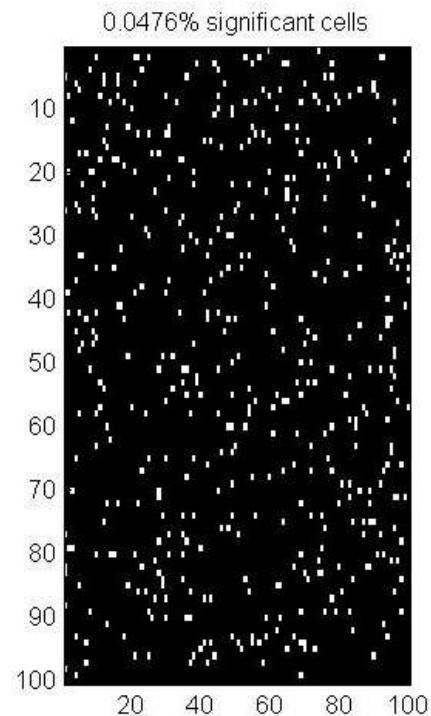


Multiple comparisons correction

- Wait, it is well known that **Bonferroni is too conservative**, i.e. the $\text{FWER} < \alpha$. In EEG we instead **consider cluster** because is much less likely that statistics are significant in groups

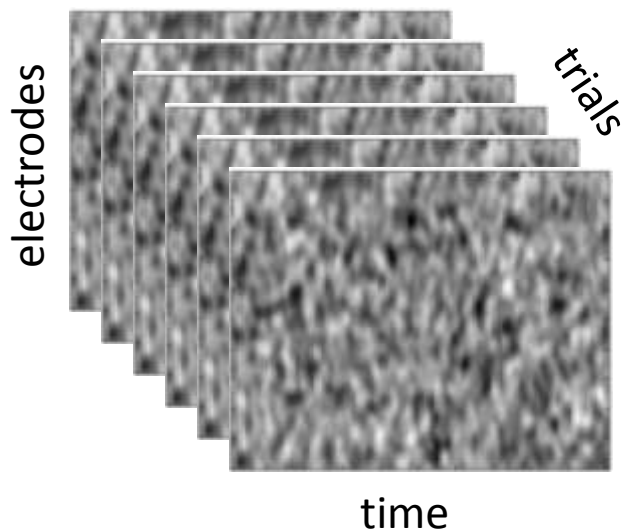


One sample t test > 0 ?

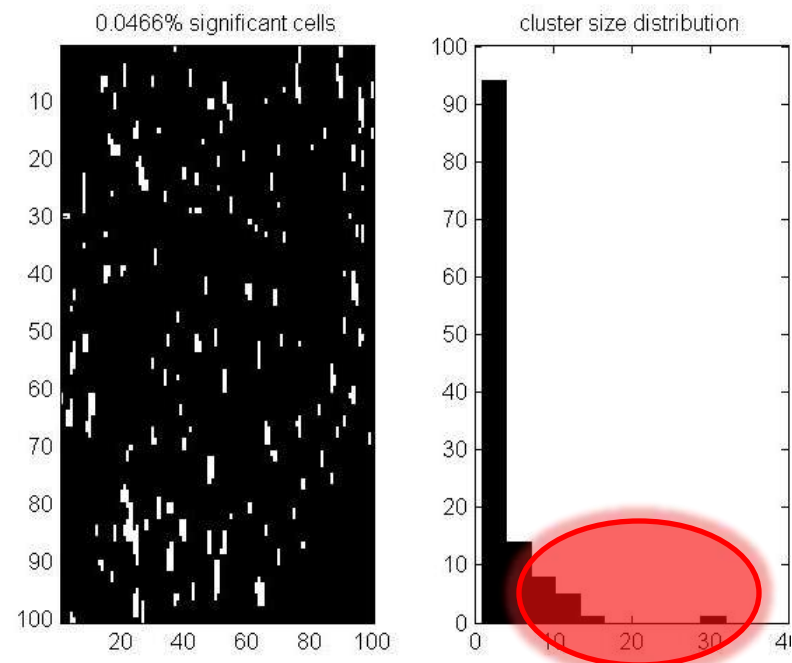


Multiple comparisons correction

- Wait, it is well known that Bonferroni is too conservative, i.e. the FWER $< \alpha$. In EEG we instead consider cluster because is much less likely that statistics are significant in groups – **but data are smooth in space and time!**



One sample t test > 0 ?

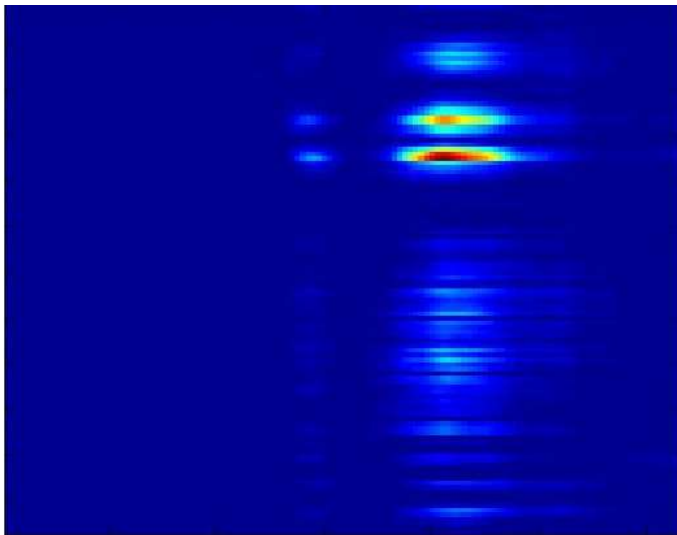


Multiple comparisons correction

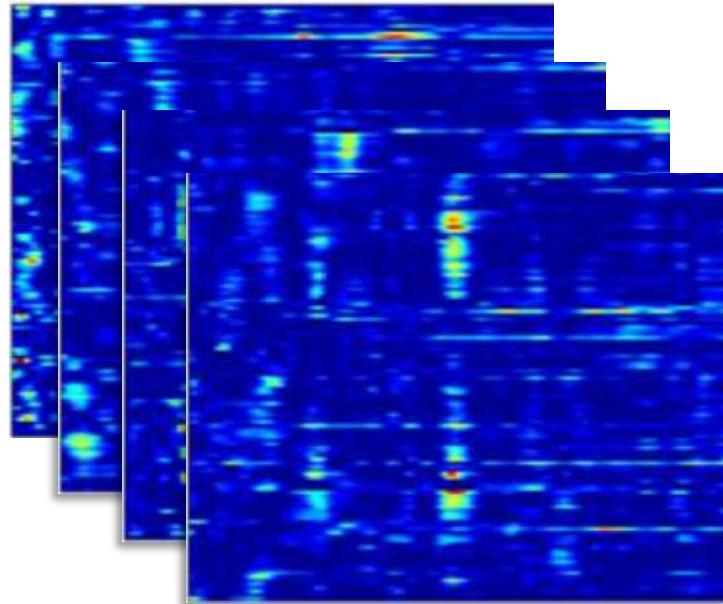
- Clustering is a good option because it accounts for topological features in the data. Techniques like Bonferroni, FDR, max(stats) control the FWER but independently of the correlation between tests.
- To use clustering we need to consider cluster statistics rather than individual statistics
- Cluster statistics depend on (i) the cluster size, which depends on the data at hand (how correlated data are in space in time), and (ii) the strength of the signal (how strong are the t, F values in a cluster) or (iii) a combination of both.

Multiple comparisons correction

- In LIMBO EEG, we bootstrap the data under H_0 : e.g. if you have 2 conditions, center the data for each condition (H_0 is true) and then resample and test – by chance some significant results are obtained.



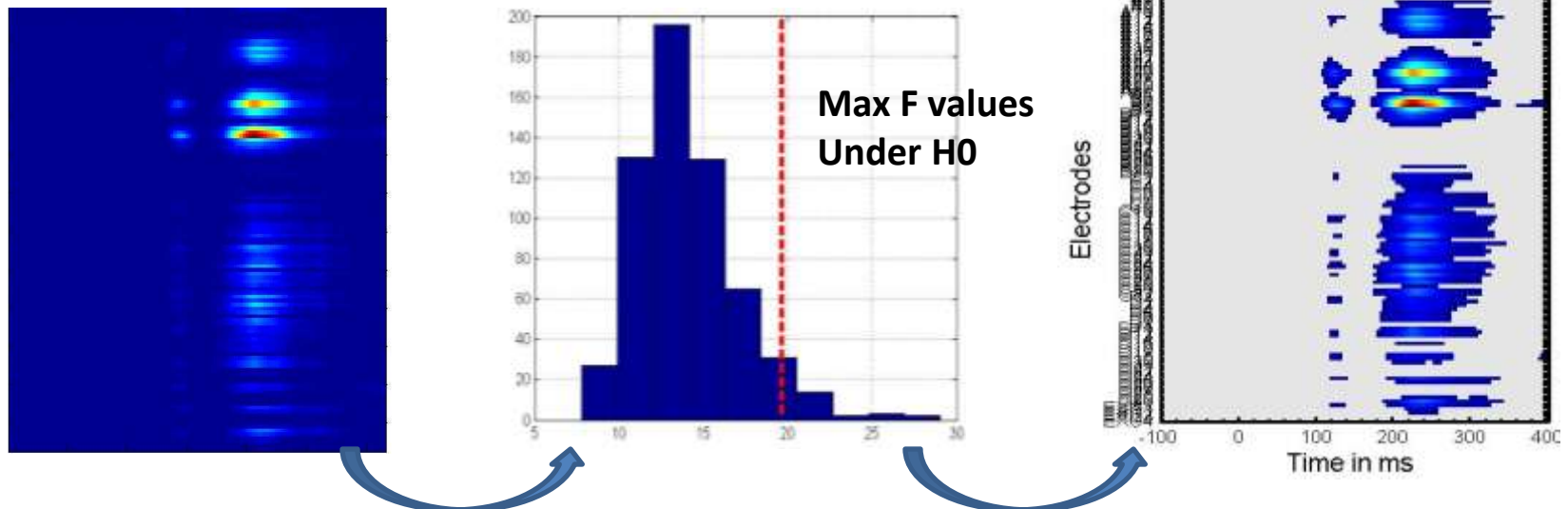
Observed F values



F values under H_0

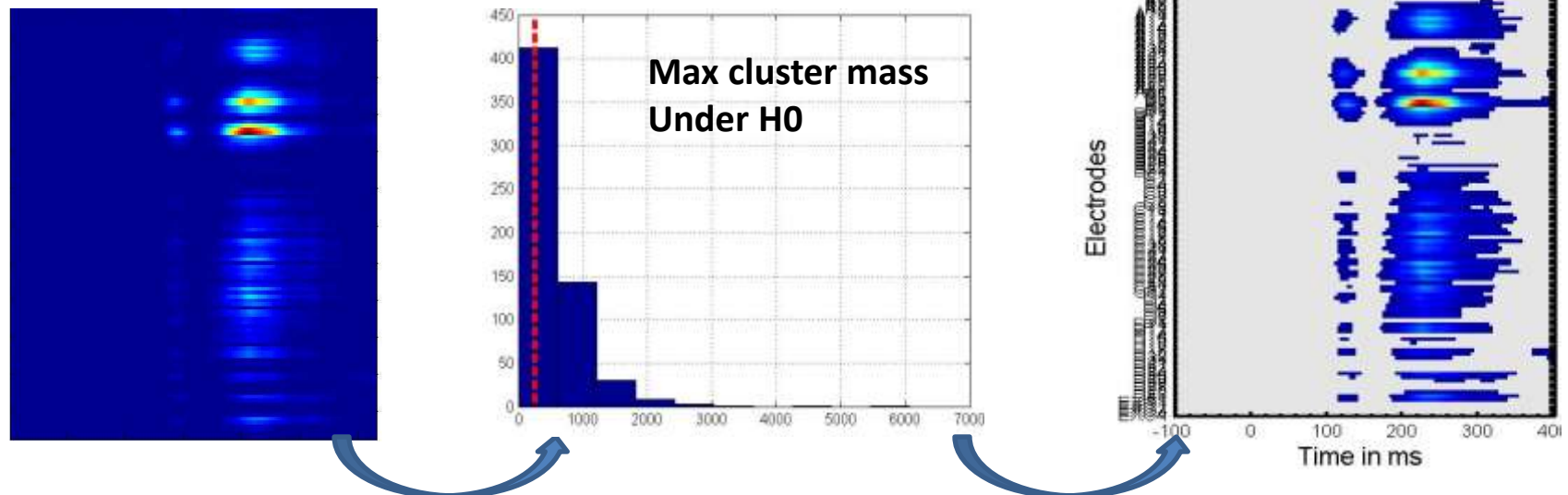
Multiple comparisons correction

- **Max(stat)**: for each bootstrap record the max(t) or max(F) to build the distribution of max under H_0 . Then threshold the observed results using this distribution. Because the max value is obtained across all electrodes and time frames, it corrects to thresholding data through this whole space. Max(stat) doesn't account for clusters.



Multiple comparisons correction

- **Spatial-Temporal clustering**: for each bootstrap, threshold at alpha and record the max(cluster mass), i.e. sum of F values within a cluster. Then threshold the observed clusters based on there mass using this distribution → accounts for correlations in space and time.



Loss of resolution: inference is about the cluster, not max in time or specific electrode !

Threshold Free Cluster Enhancement

- **Threshold Free Cluster Enhancement (TFCE)**: Integrate the cluster mass at multiple thresholds. A TFCE score is thus obtain per cell but the value is a weighted function of the statistics by it's belonging to a cluster.

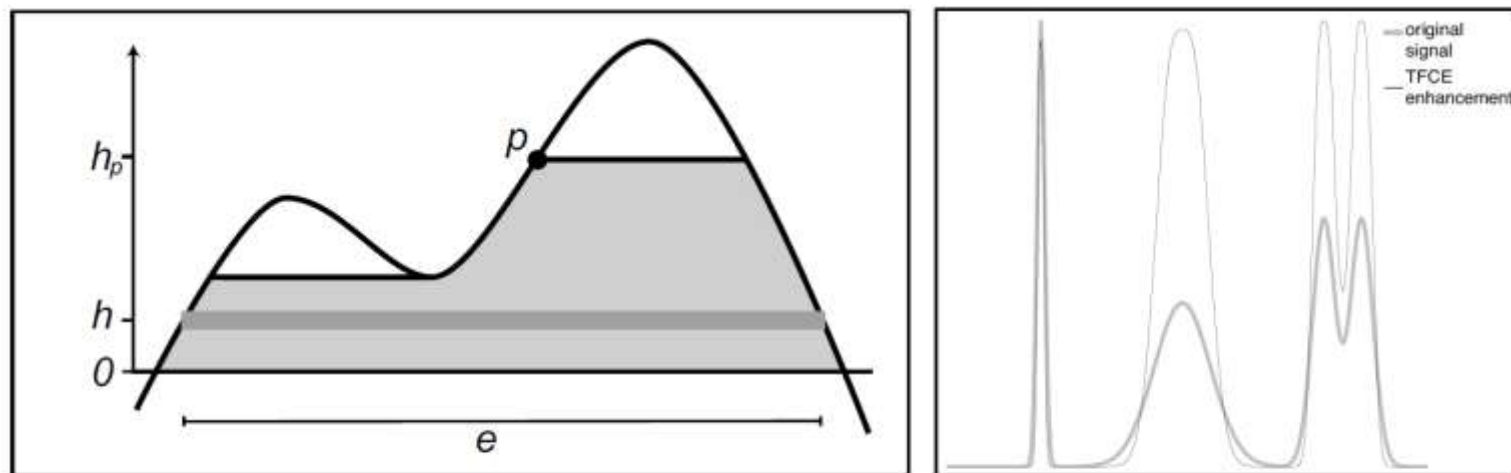
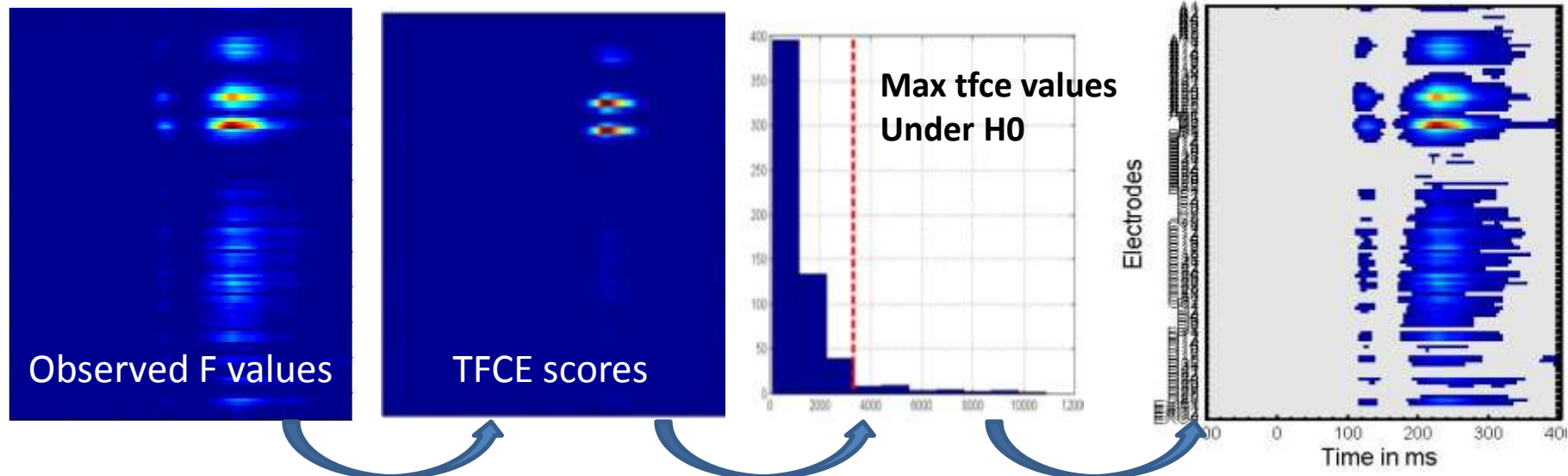


Figure 1: Illustration of the TFCE approach. Left: The TFCE score at voxel p is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark grey band) within the area of “support” of p (light grey). The score for each section is a simple function of its height h and extent e . Right: Example input image and TFCE-enhanced output. The input contains a focal, high signal, a much more spatially extended, lower, signal and a pair of overlapping signals of intermediate extent and height. The TFCE output has the same maximal values for all three cases, and preserves the distinct local maxima in the third case.

Multiple comparisons correction

- **Threshold Free Cluster Enhancement (TFCE)**: Integrate the cluster mass at multiple thresholds. A TFCE score is thus obtain per cell but the value is a weighted function of the statistics by it's belonging to a cluster. As before, bootstrap under H_0 and get $\max(\text{tfce})$.



Excellent resolution: inference is about cells, but we accounted for space/time dependence

The future of your analyses

What's next

- 1st level analysis revised with WLS
- Multivariate analyses: MANCOVA ($X \rightarrow Y$) and discriminant analyses ($Y \rightarrow X$) both linear and quadratic
- 2nd level robust for all tests
- *Andrew* \rightarrow All Components *time frames
 \rightarrow All Components/Channels *time frames * freq

LIMO Support

- Software use → EEGLAB mailing list (Arno has lots of time to manage it)
- Software / data issues → email me (cyril.pernet@ed.ac.uk GLM / stats related questions) or Guillaume (guillaume.rousselet@gla.ac.uk MCC and data handling issues)
- Software bugs / changes → use tracker on the website, remember to assign the bug to me or Guillaume