LInear Modelling of EEG data

Theory and practice

Cyril R. Pernet, University of Edinburgh

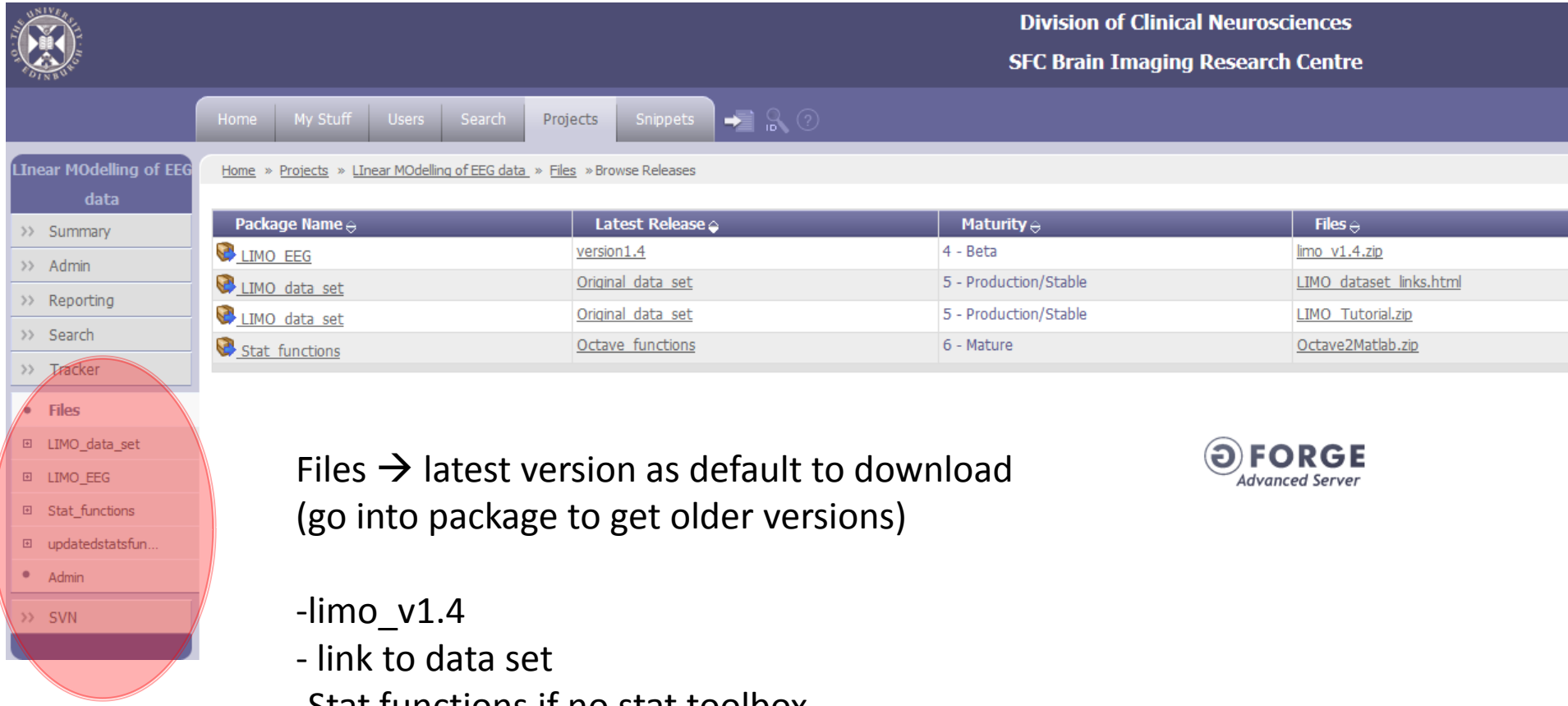
EEGLAB workshop, Aspet 2013

LIMO EEG:

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

LIMO EEG:

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



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/

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).

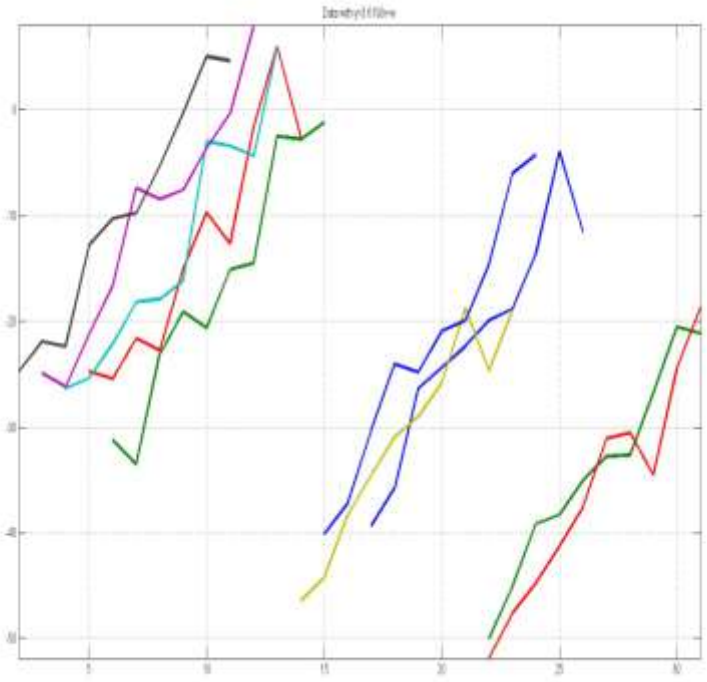
- Multiple comparisons correction: bootstrap used to evaluate the clustering of statistical tests.

Mixed effect Model

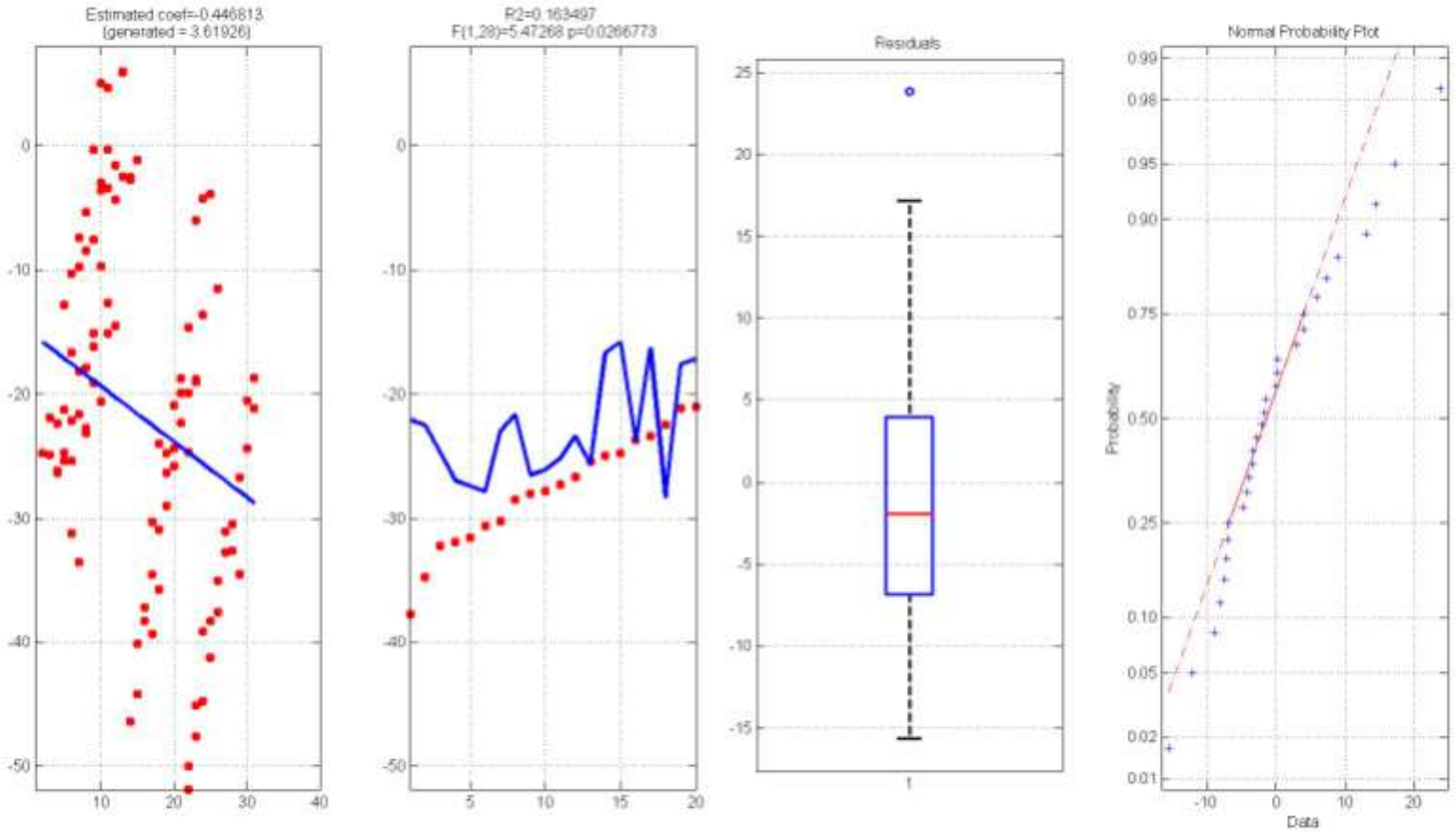
Model the data with fixed effects (the experimental conditions) and a random effect (subjects are allowed to have different overall values – considering subjects as a random variable)

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

→ Plot the data per intensity

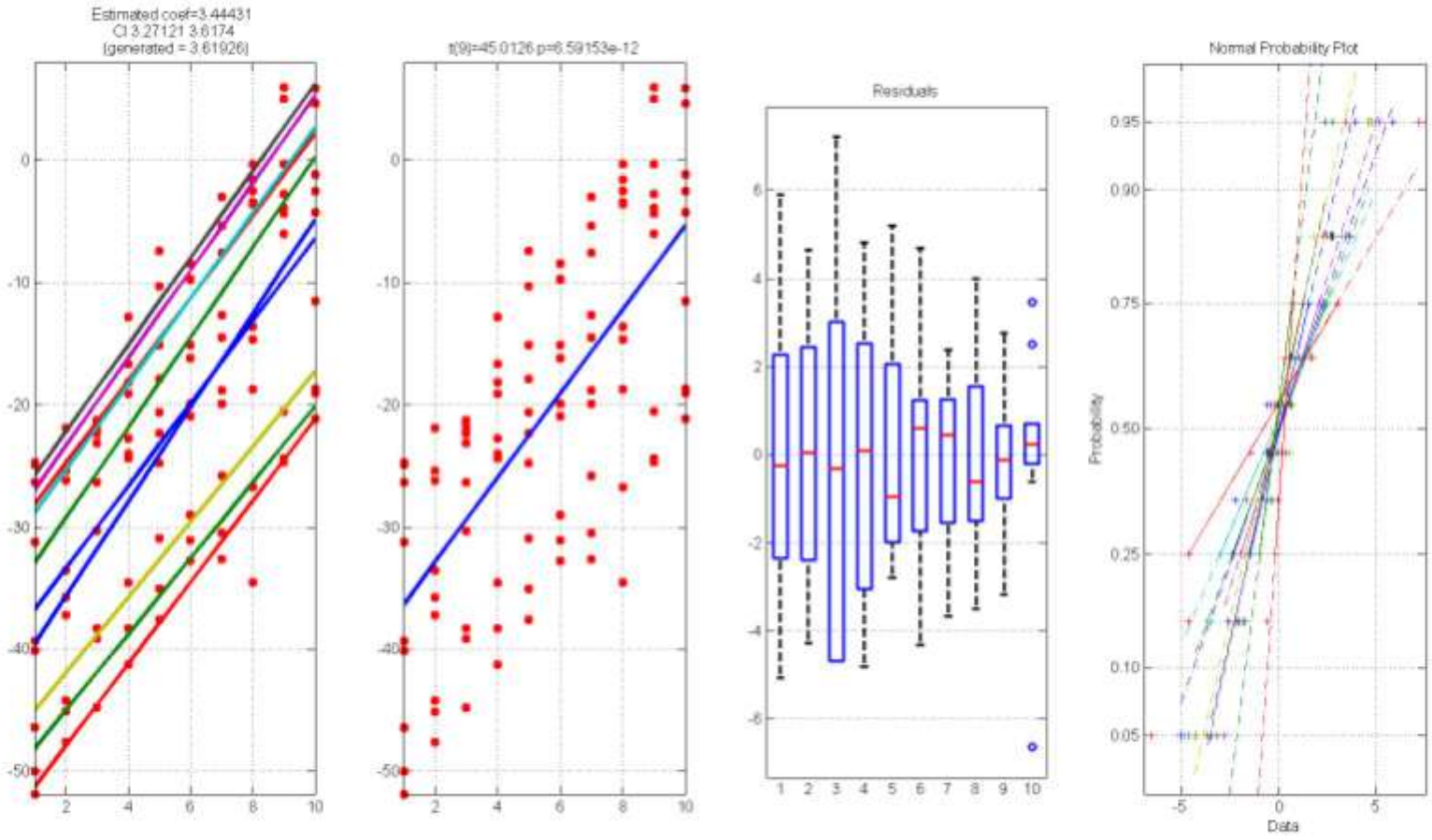


Mixed effect Model



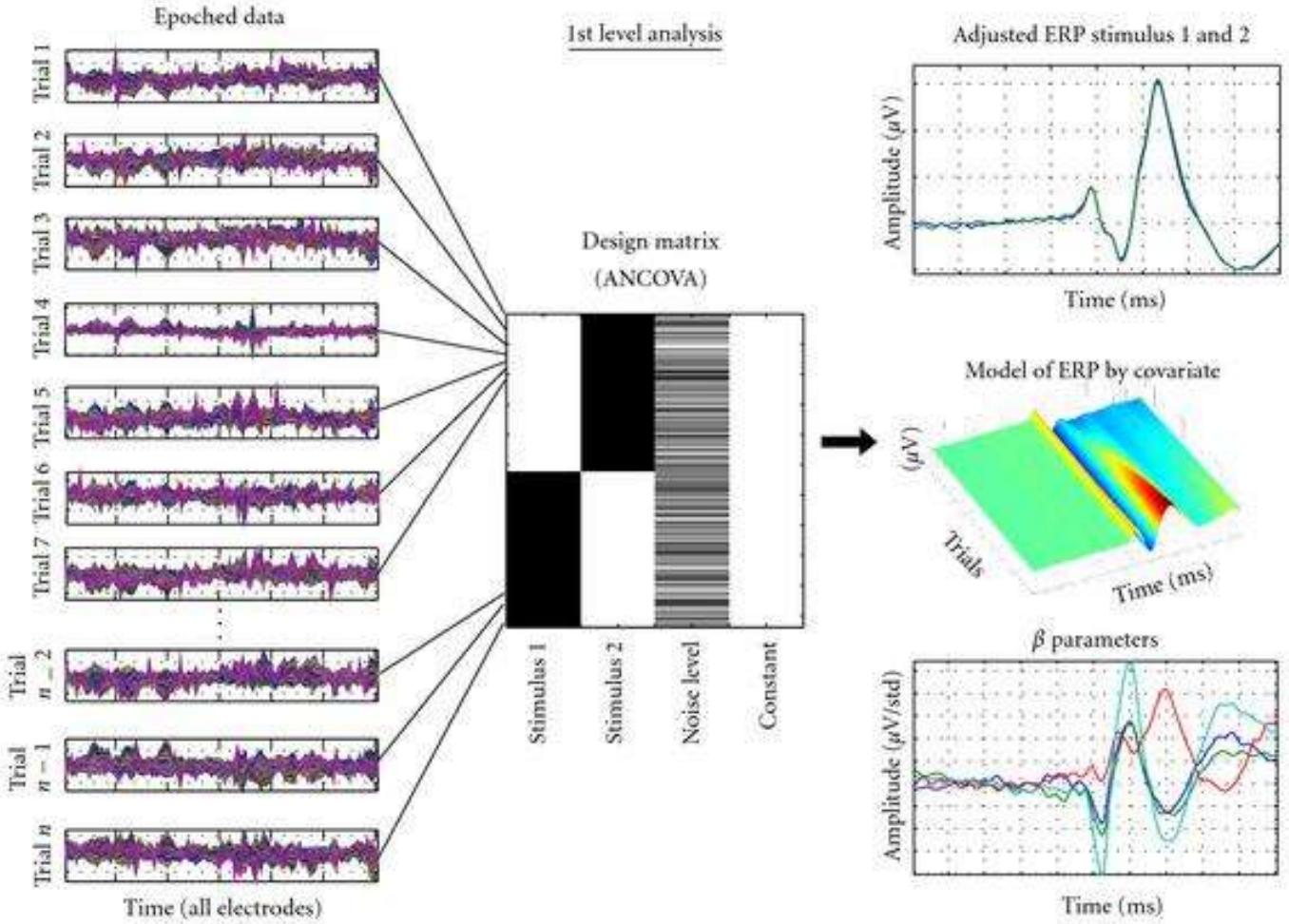
Fixed effect = average across subjects → negative correlation?

Mixed effect Model



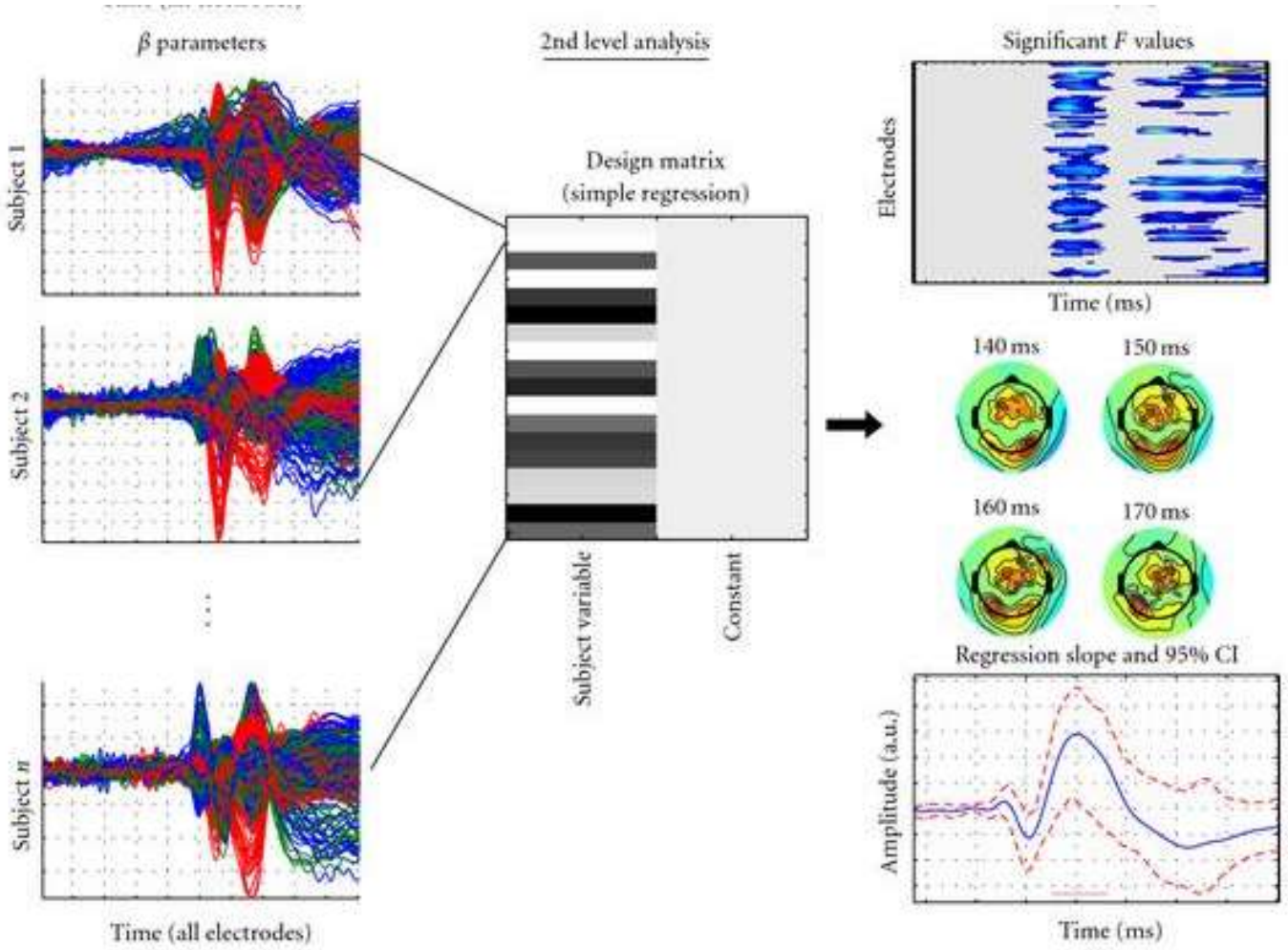
Mixed effect = effect per subject with variable (random) offsets

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*acuity + 2*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 squres), 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 $B = inv(X'X)X'Y \rightarrow pinv(X)Y$

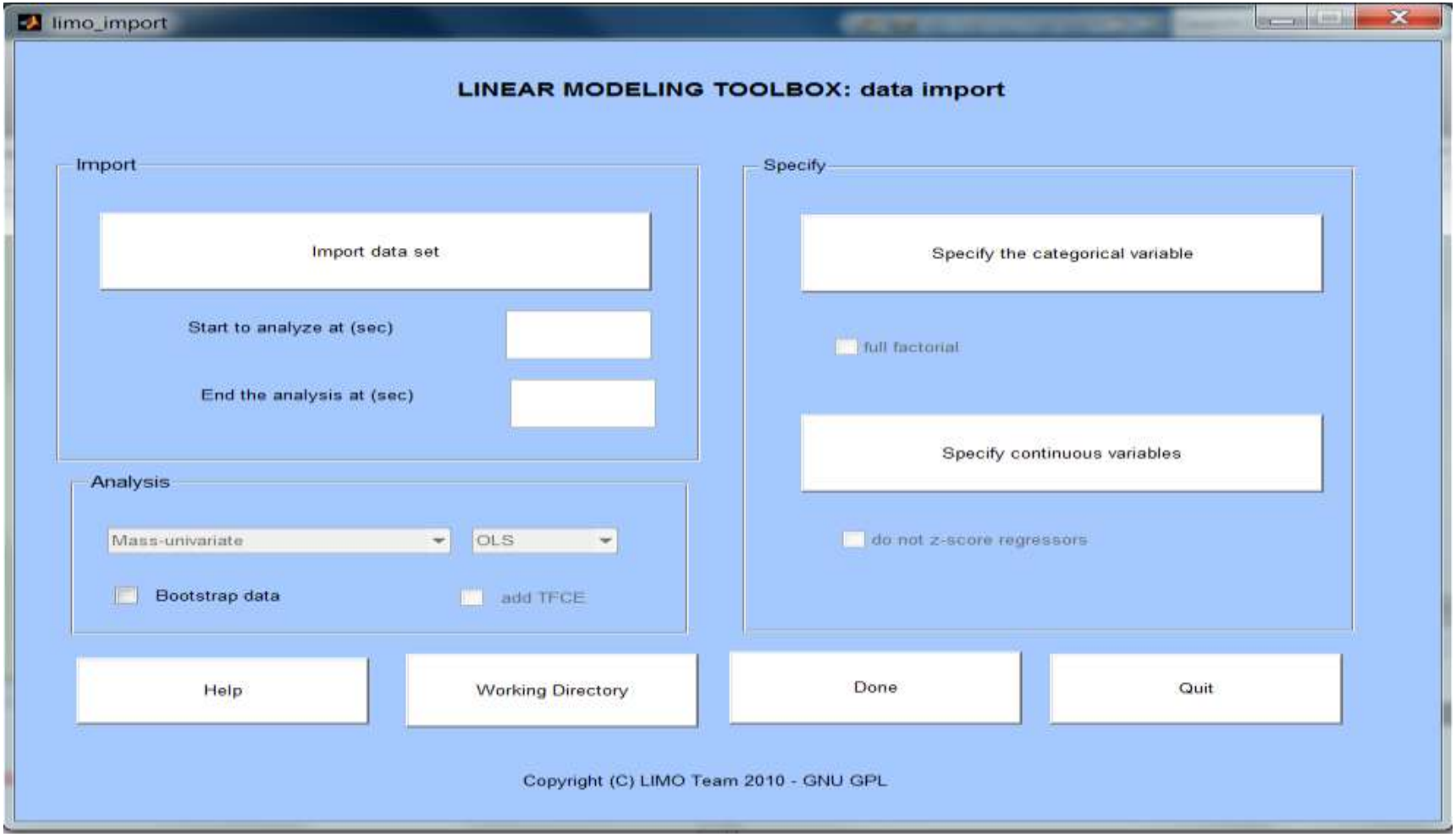
$R = I- (X*pinv(X))$ and thus $var = RY'*RY / dfe$

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

1st level model in practice

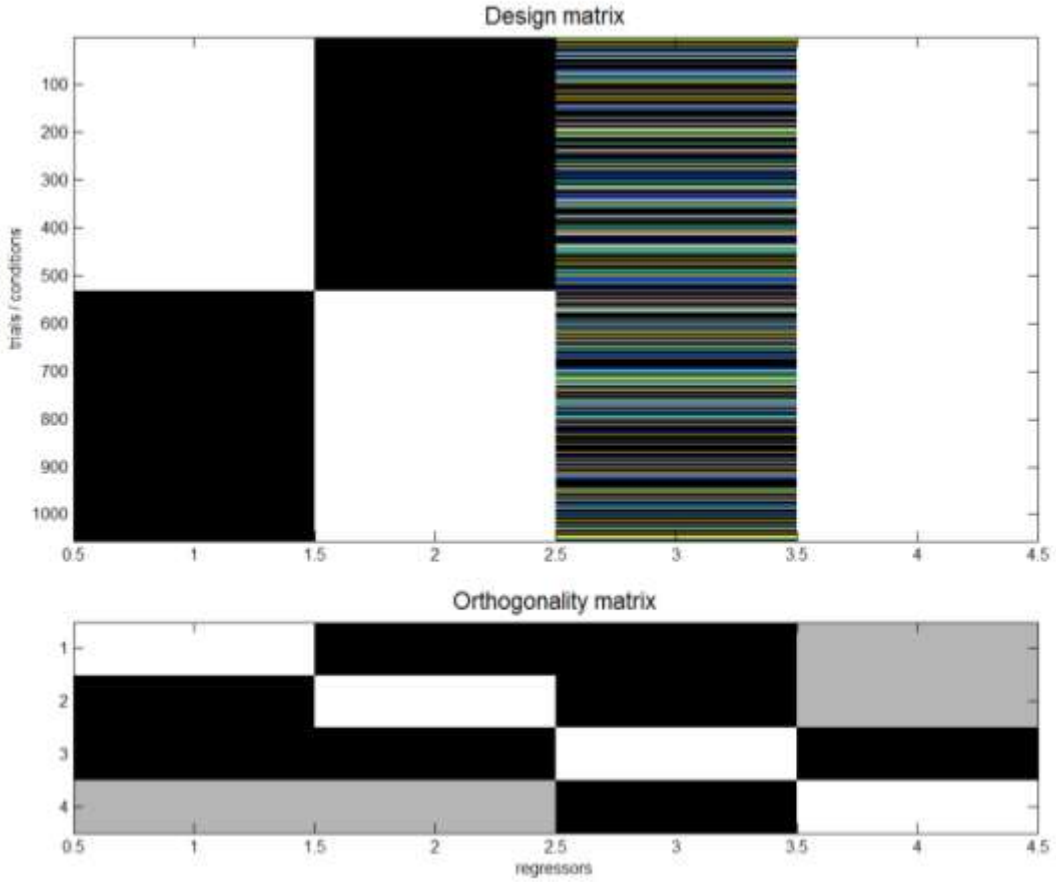
General Linear Model: Practice 1

- **Processing of subject 1.**



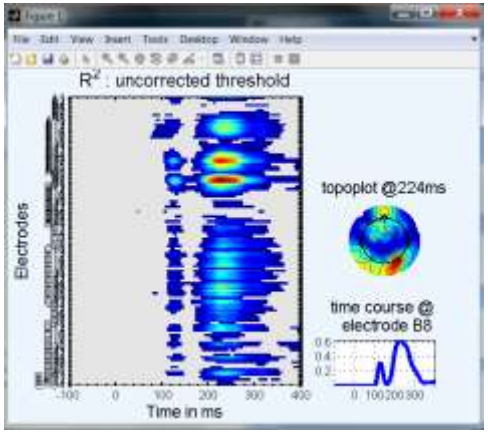
General Linear Model: Practice 1

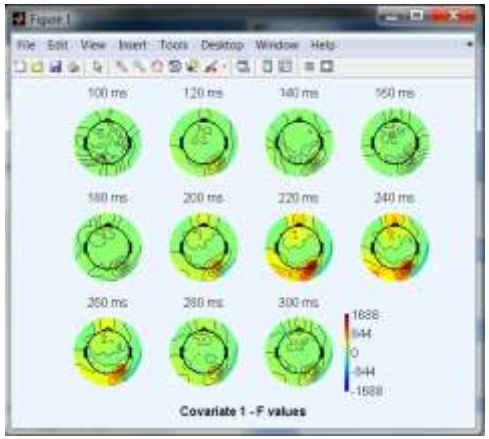
- Design matrix: accept if you think it looks like what the model should be

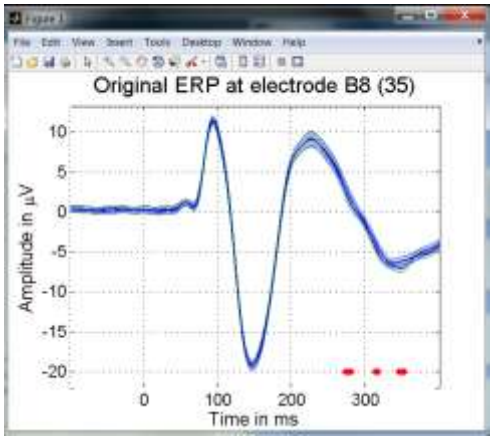


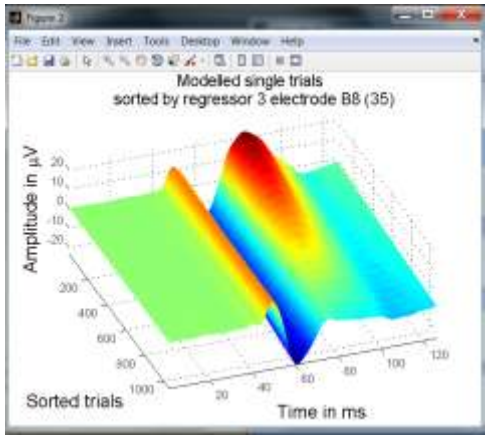
General Linear Model: Practice 1

- 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 $\sim N(0,s^2I)$

- 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 (Cov(ei,ej)=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:

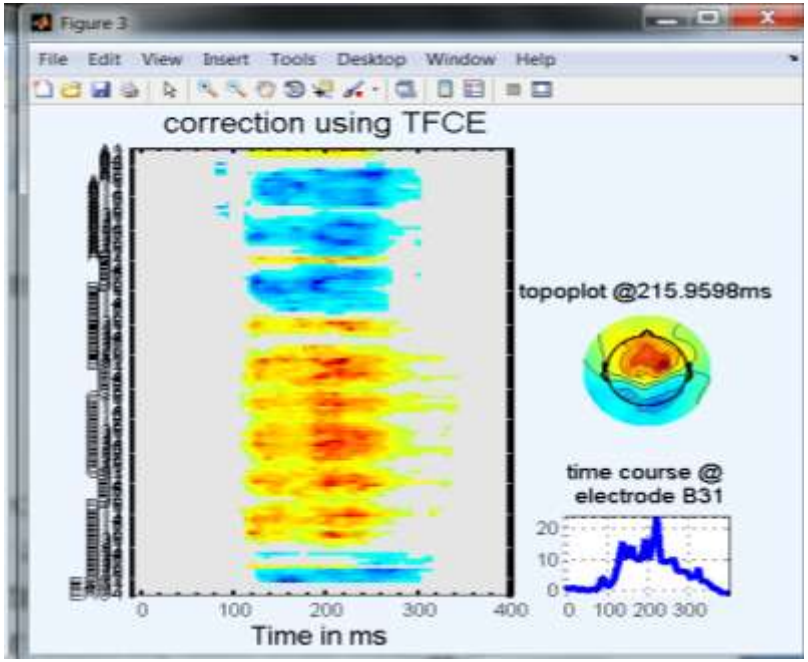
T-tests (robust)

Regression, ANOVA, ANCOVA

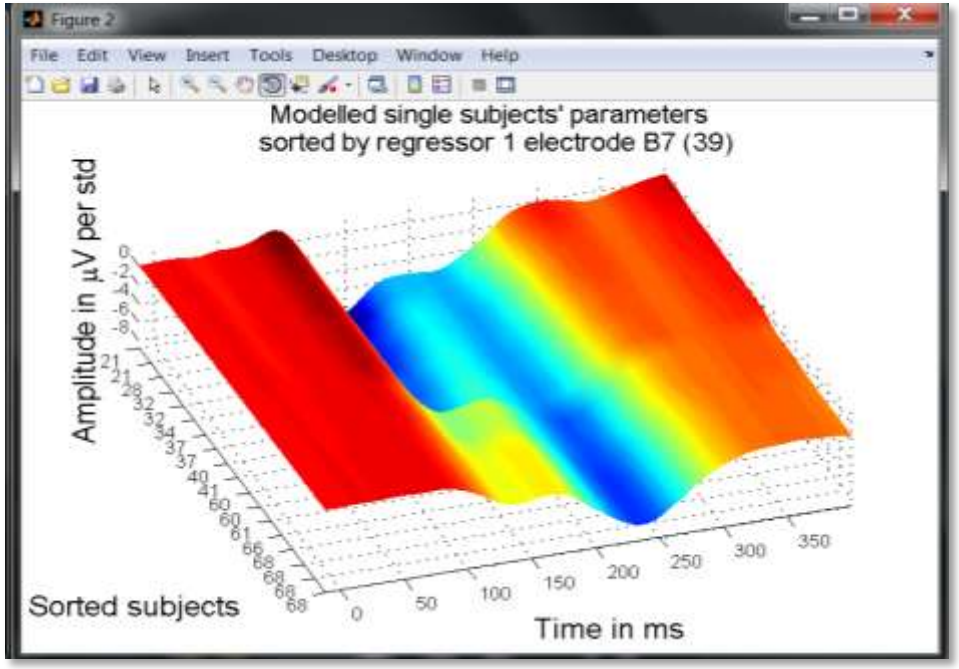
Repeated Measures ANOVA (Hotelling T2)

General Linear Model: Practice 2

One-sample t-test



Regression



Multiple Comparisons Correction

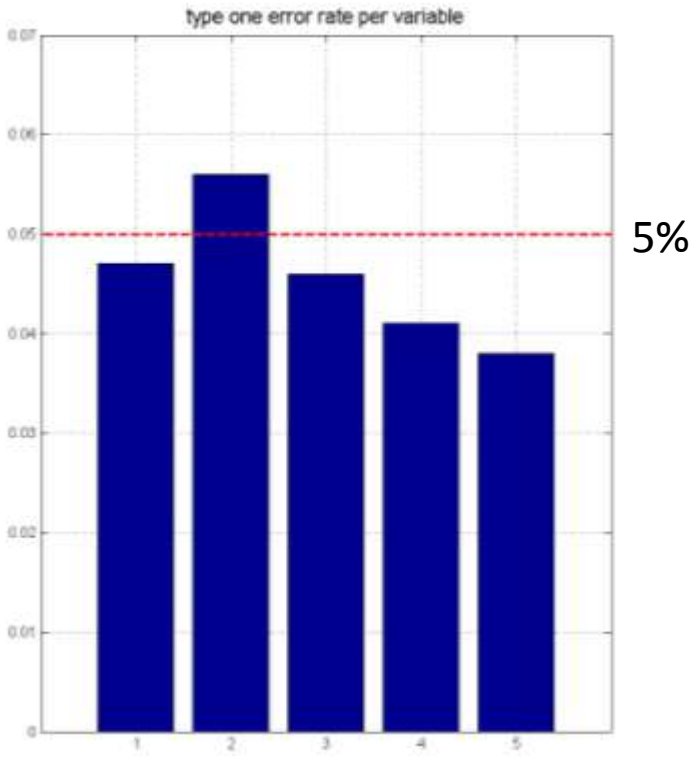
Multiple comparisons correction

- When we do a statistical test, we set alpha, the probability to reject H0 (under H0) – this is also known as type I error rate

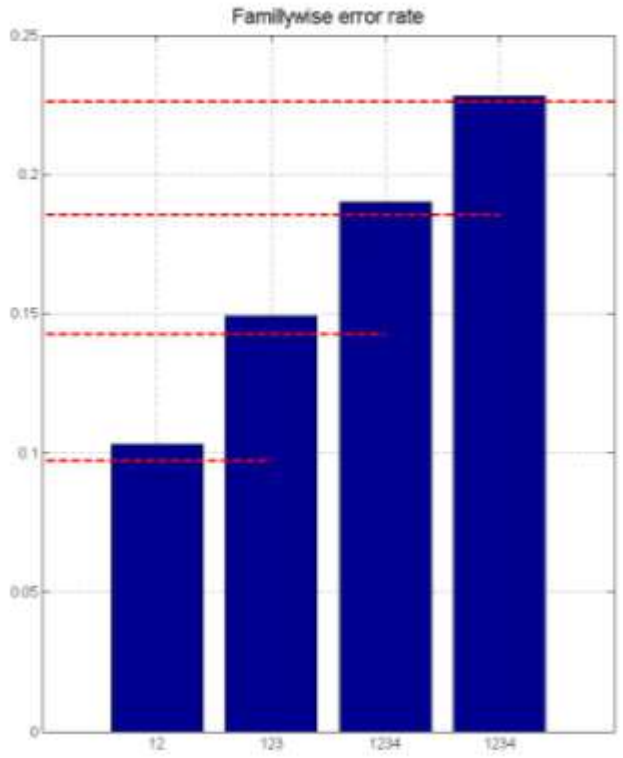
- The famillywise error rate has to do with the number of tests performed, and assuming tests are independents from each other, the FWER = 1 - (1 - alpha)^n
- eg. so for alpha =5/100, if we do 2 tests we should get about 1-(1-5/100)^2 ~ 9% false positives, if we do 126 electrodes * 150 time frames tests, we should get about 1-(1-5/100)^18900 ~ 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



5%



22%

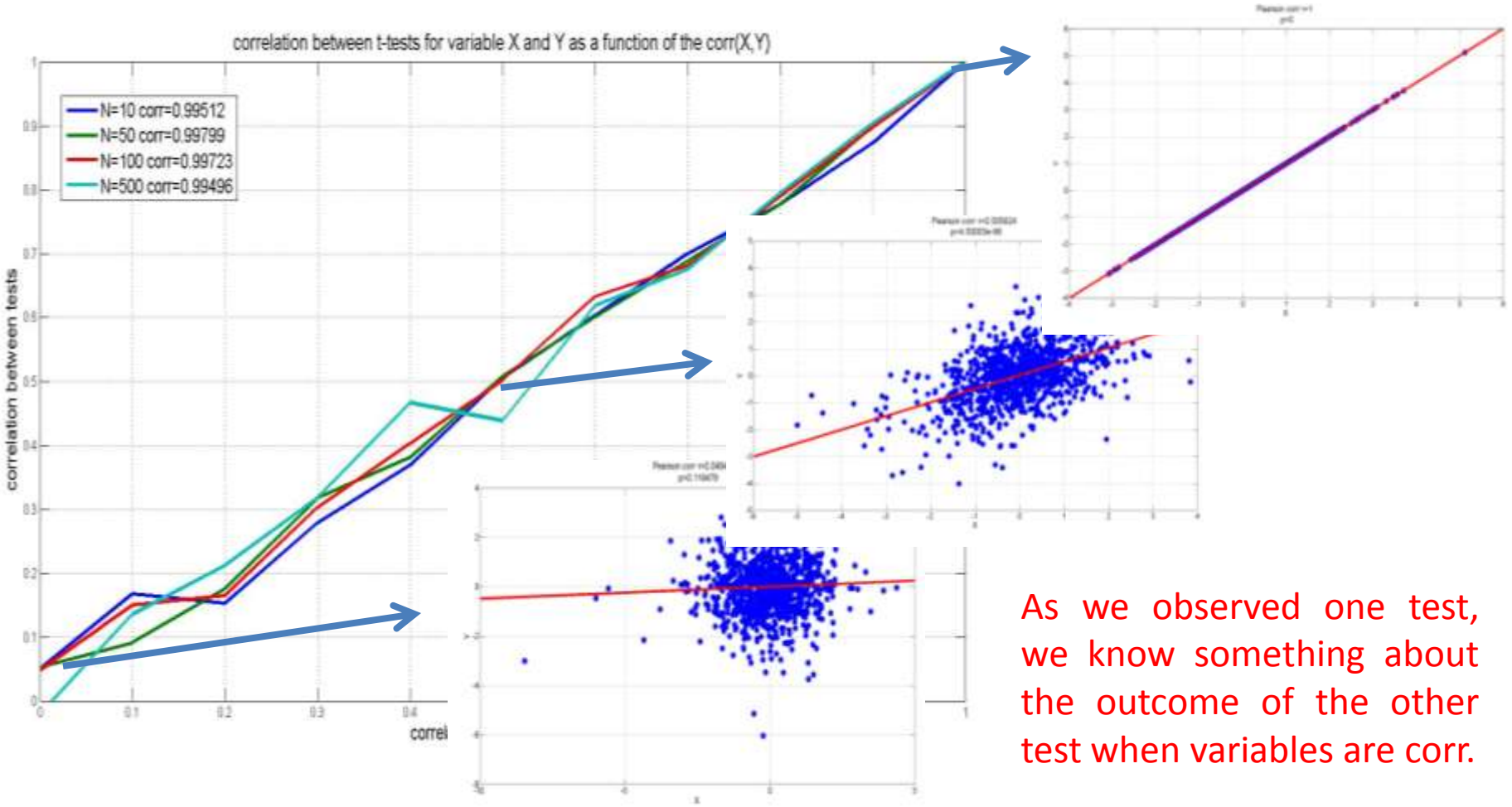
18%

14%

9%

Multiple comparisons correction

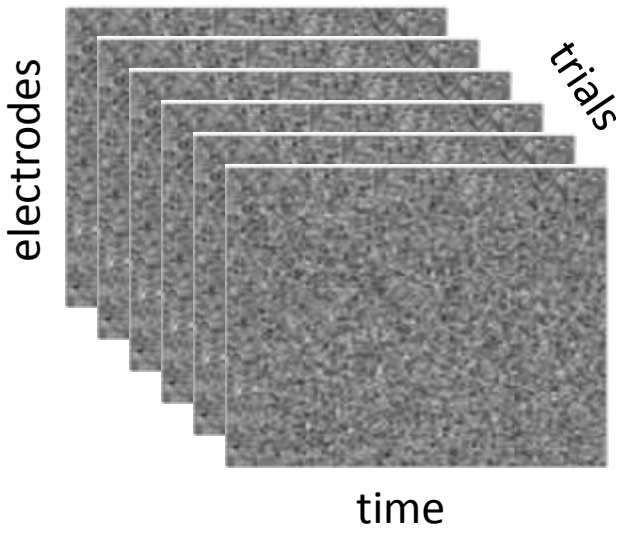
- Illustration with 2 variables with Pearson‘s r=0:1



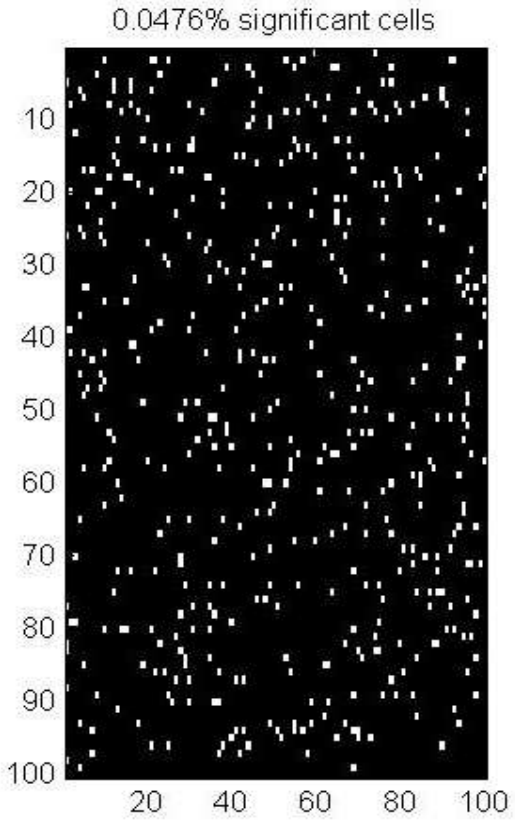
As we observed one test, we know something about the outcome of the other test when variables are corr.

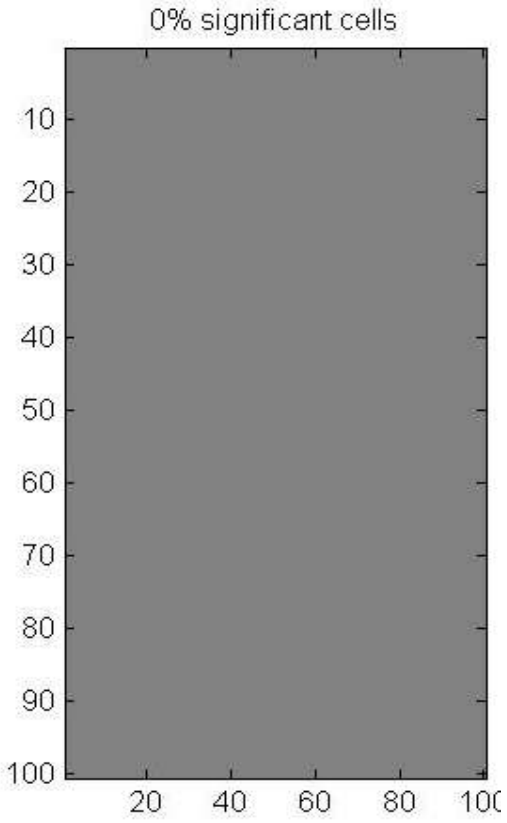
Multiple comparisons correction

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



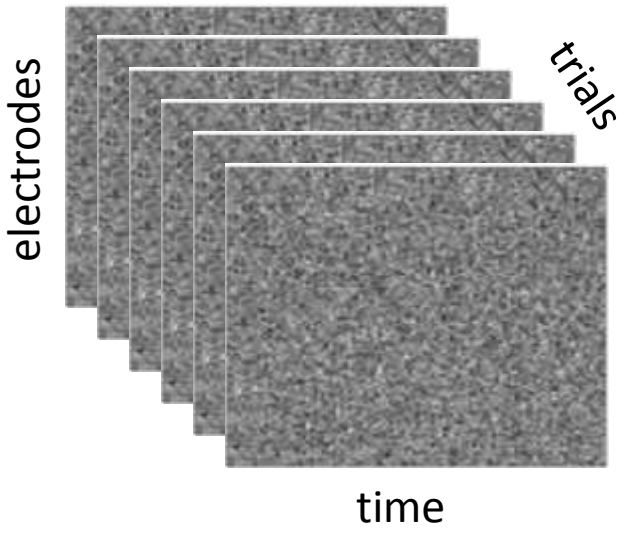
One sample t test > 0 ?



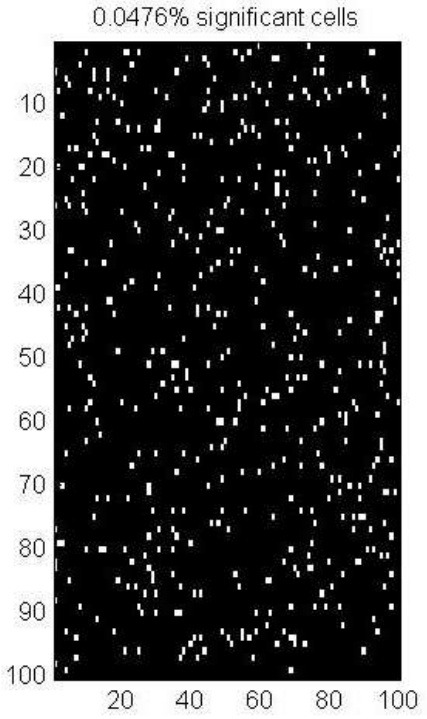


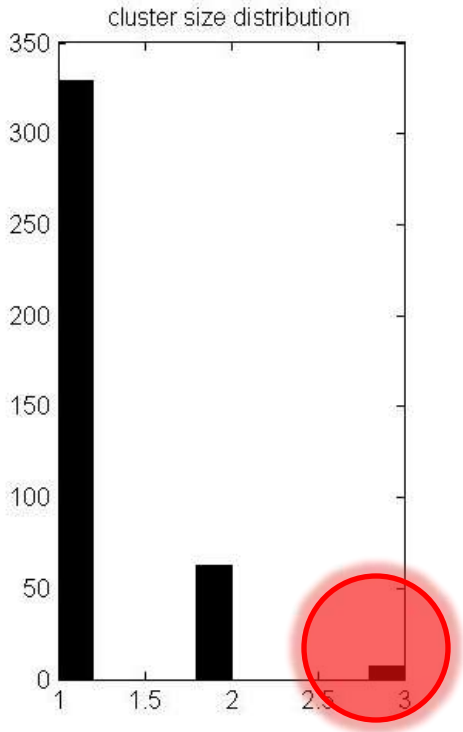
Multiple comparisons correction

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



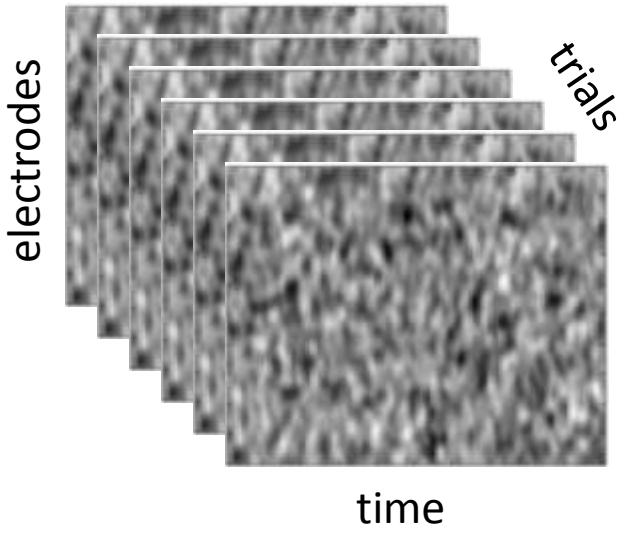
One sample t test > 0 ?



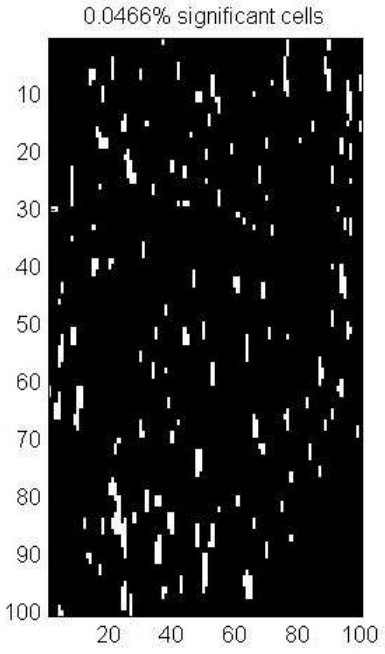


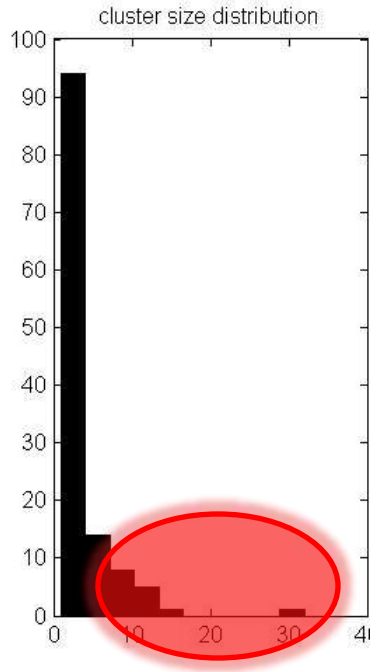
Multiple comparisons correction

- 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 ?





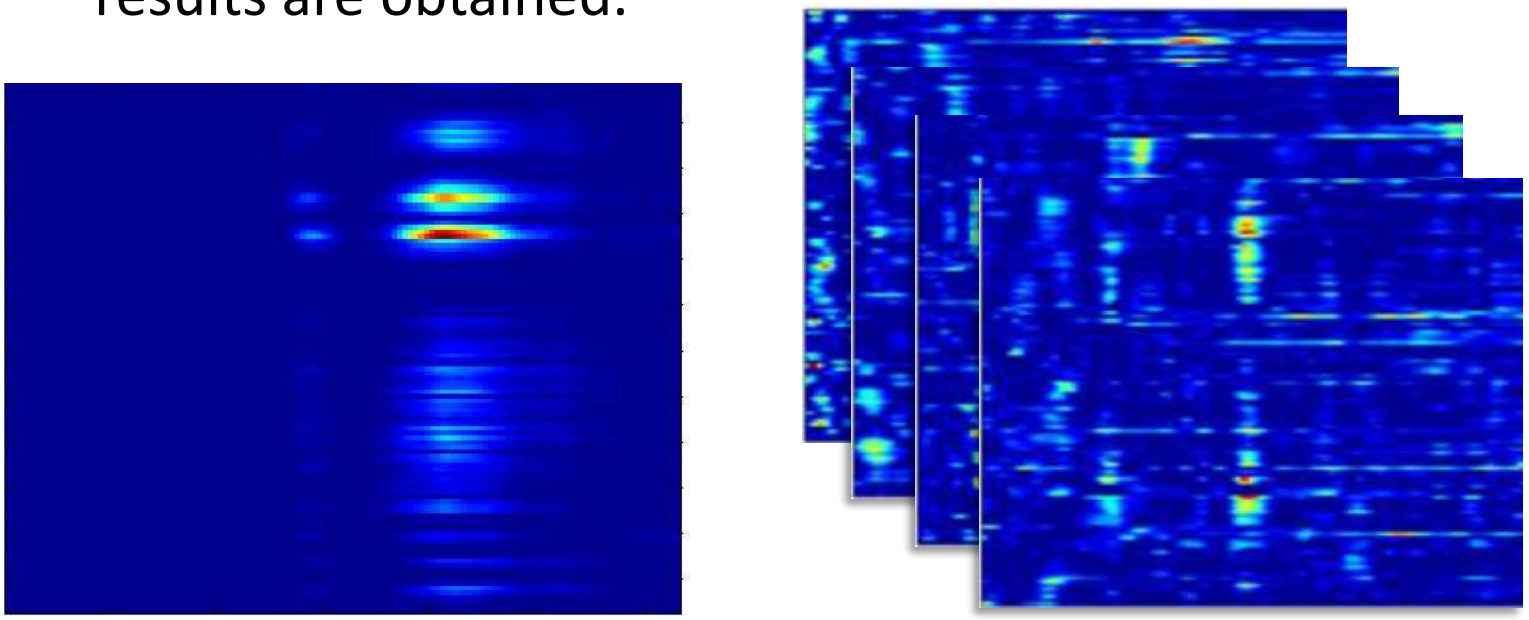
Multiple comparisons correction

- 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

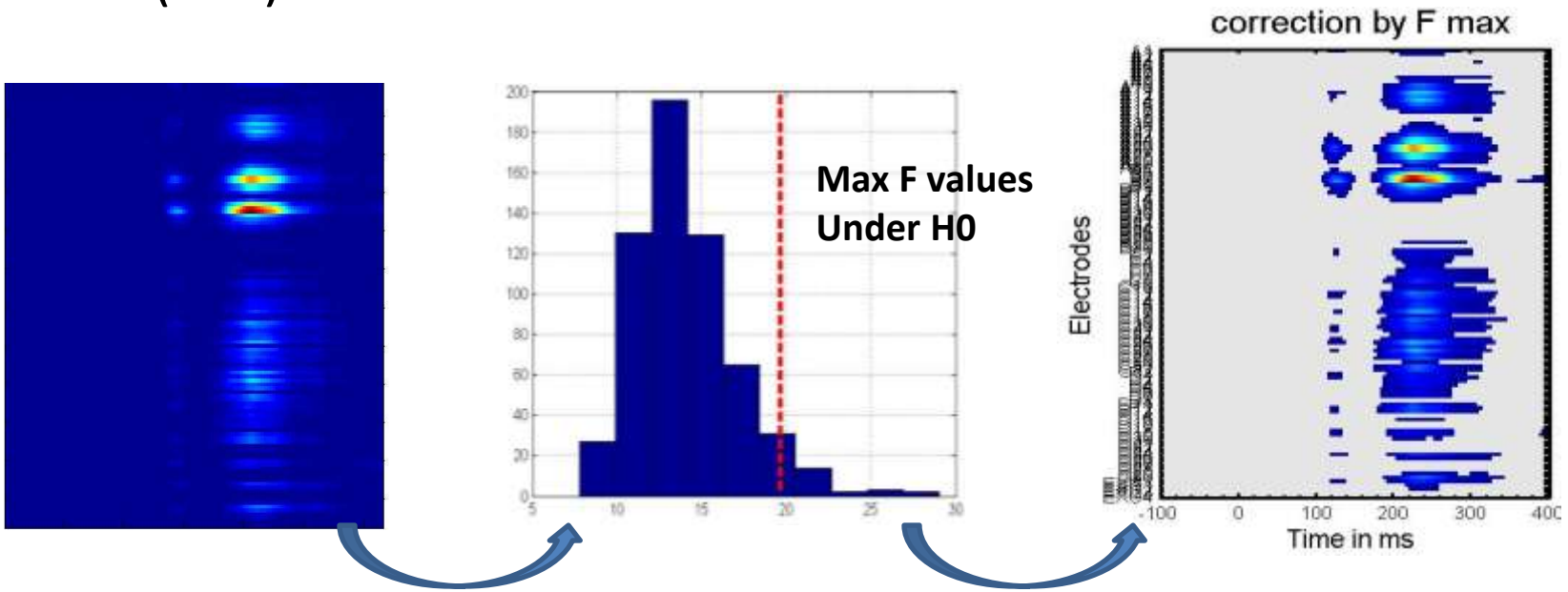
- In LIMO EEG, we bootstrap the data under H0: e.g. if you have 2 conditions, center the data for each condition (H0 is true) and then resample and test – by chance some significant results are obtained.



Observed F values → F values under H0

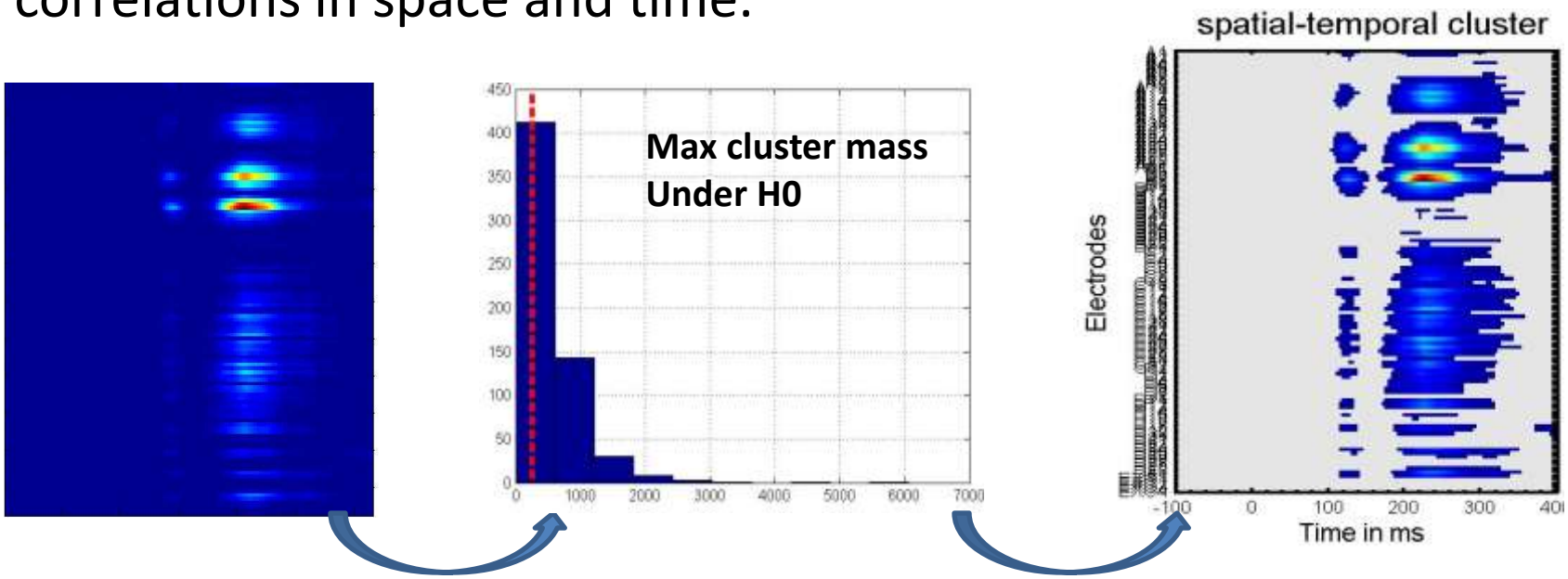
Multiple comparisons correction

- Max(stat): for each bootstrap record the max(t) or max(F) to build the distribution of max under H0. 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 trough 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.

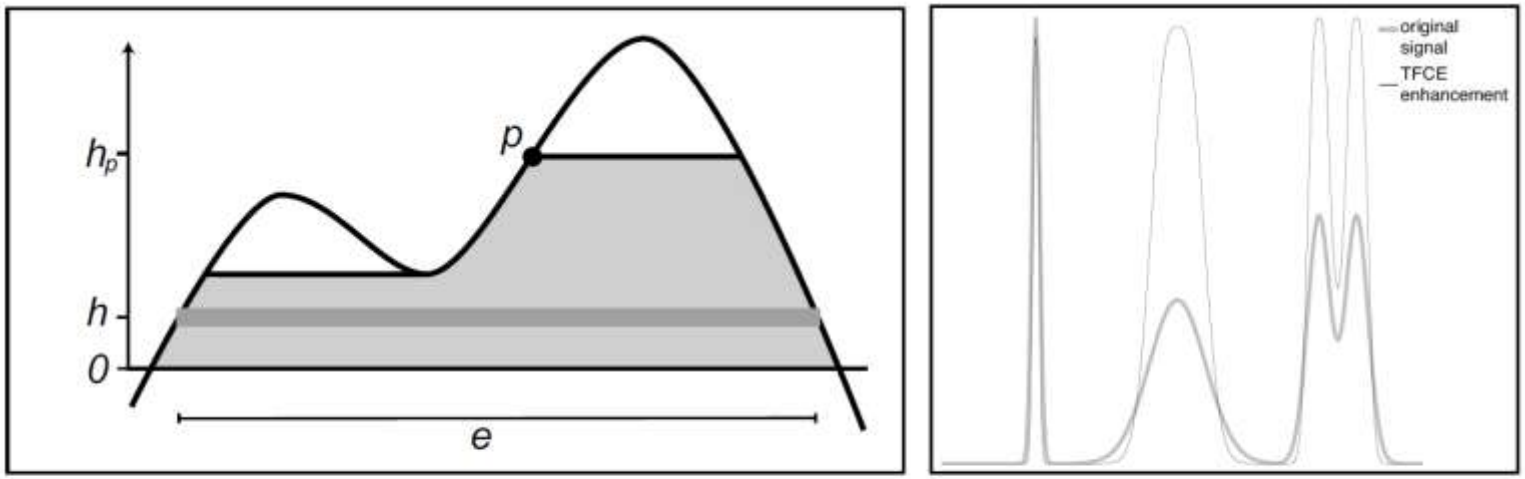
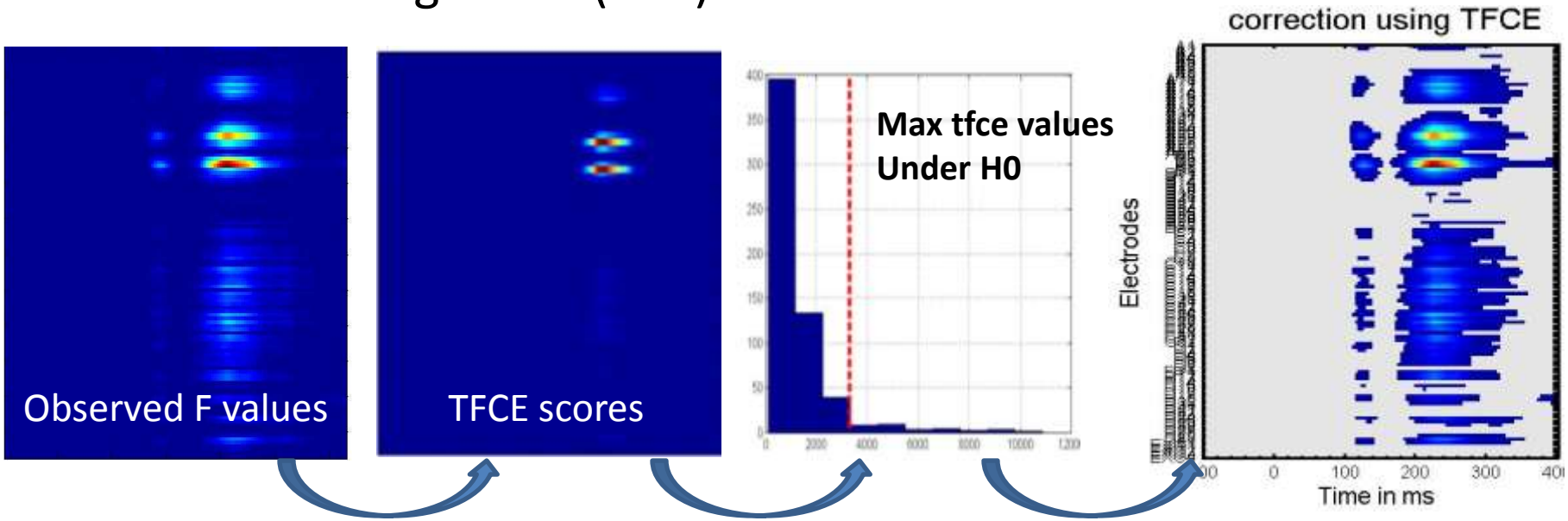


Figure 1: Illustration of the TFCE approach. Left: The TFCE score at voxel p is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark grey band) within the area of "support" of p (light grey). The score for each section is a simple function of its height 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.

Smith & Nichols 2009 NeuroImage 44

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 H0 and get max(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→Y) and discriminant analyses (Y→X) both linear qnd quadratic

- 2nd level robust for all tests

- *Andrew* → All Components *time frames

→ 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