Edinburgh Imaging
www.ed.ac.uk/edinburgh-imaging

Statistical analysis of EEG data

Hierarchical modelling and multiple comparisons correction

10.6084/m9.figshare.4233977

Cyril Pernet, PhD
Centre for Clinical Brain Sciences
The university of Edinburgh, UK

22nd EEGLAB Workshop – San Diego, Nov. 2016

Context

- **Data collection** consists in recording electromagnetic events over the whole brain and for a relatively long period of time, with regards to neural spiking.
- In the majority of cases, **data analysis** consists in looking where we have signal and restrict our analysis to these channels and components.

➢ Are we missing the forest by choosing working on a single, or a few trees?

➢ By analysing where we see an effect, we increase the type 1 FWER because the effect is partly driven by random noise (solved if chosen based on prior results or split the data)

Rousselet & Pernet – It's time to up the Game Front. Psychol., 2011, 2, 107

Context

- Most often, we compute averages per condition and do statistics on peak latencies and amplitudes
- Several lines of evidence suggest that peaks mark the end of a process and therefore it is likely that most of the interesting effects lie in a component before a peak
- **Neurophysiology**: whether ERPs are due to additional signal or to phase resetting effects a peak will mark a transition such as neurons returning to baseline, a new population of neurons increasing their firing rate, a population of neurons getting on / off synchrony.
- **Neurocognition**: reverse correlation techniques showed that e.g. the N170 component reflects the integration of visual facial features relevant to a task at hand (Schyns and Smith) and that the peak marks the end of this process.

Rousselet & Pernet – It's time to up the Game Front. Psychol., 2011, 2, 107

Context

- Most often, we compute averages per condition and do statistics on peak latencies and amplitudes

➢Univariate methods extract information **among trials** in time and/or frequency across space

➢ Multivariate methods extract information across space, time, or both, **in individual trials**

➢Averages don't account for trial variability, fixed effect can be biased – these methods allow to get around these problems

Pernet, Sajda & Rousselet – Single trial analyses, why bother? Front. Psychol., 2011, 2, 322

Overview

- Fixed, Random, Mixed and Hierarchical
- Modelling subjects using a HLM
- Application to MEEG data
- Multiple Comparison correction for MEEG

Fixed, Random, Mixed and Hierarchical

Fixed effect: Something the experimenter directly manipulates

y=XB+e data = beta * effects + error
y=XB+u+e data = beta * effects + constant subject effect + error

Random effect: Source of random variation e.g., individuals drawn (at random) from a population. **Mixed effect**: Includes both, the fixed effect (estimating the population level coefficients) and random effects to account for individual differences in response to an effect

Y=XB+Zu+e data = beta * effects + zeta * subject variable effect + error

Hierarchical models are a mean to look at mixed effects.

Fixed vs Random

Fixed effects:

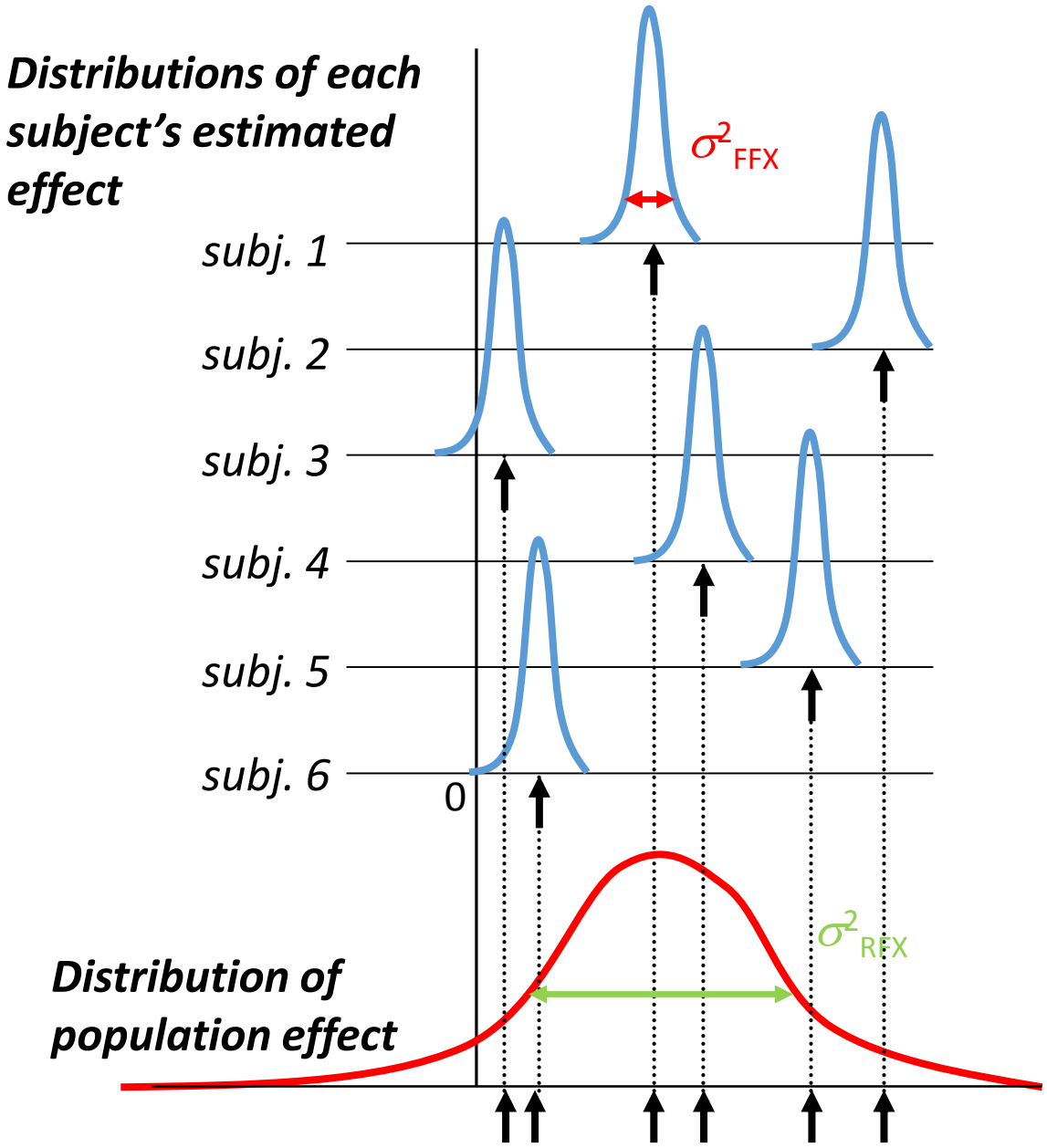
Intra-subjects variation

suggests all these subjects different from zero

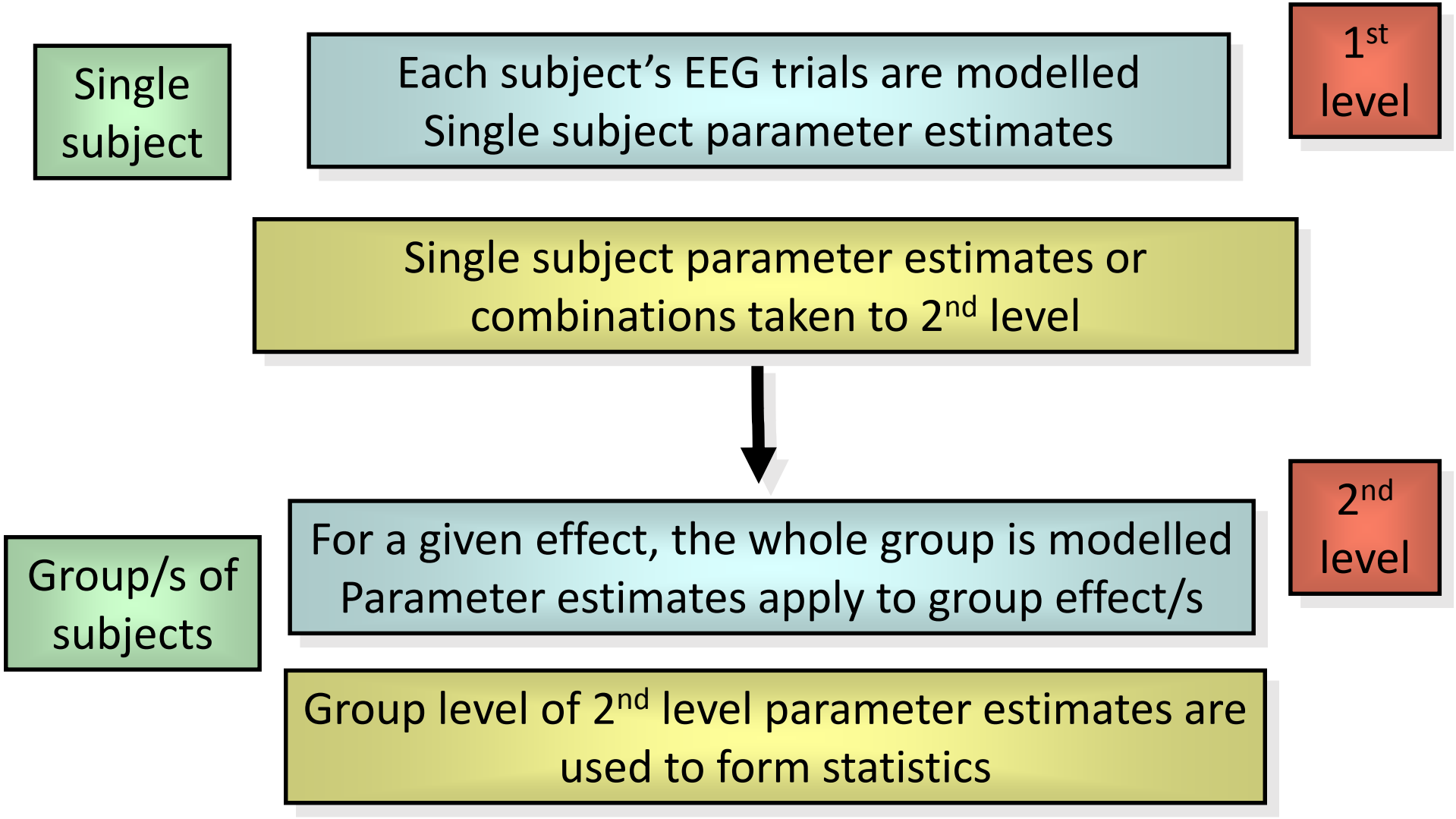
Random effects:

Inter-subjects variation

suggests population not different from zero



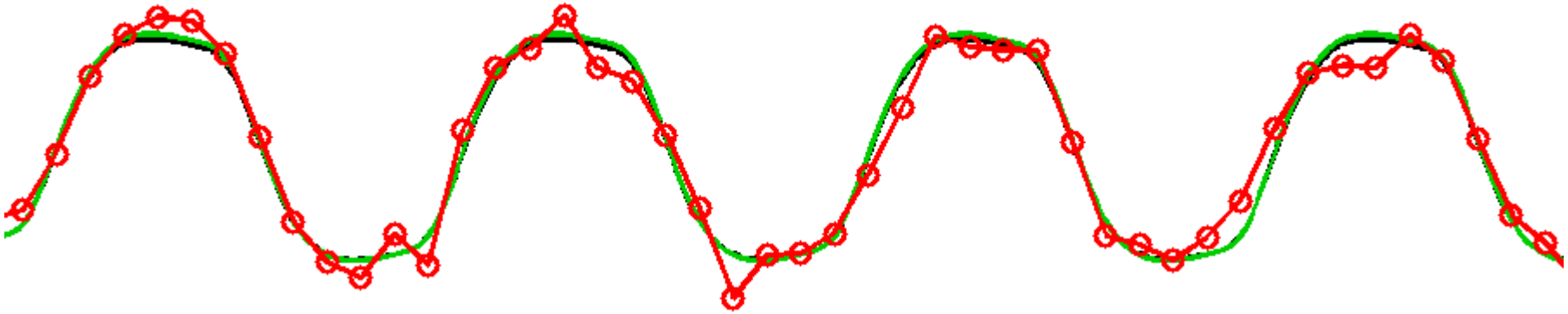
Hierarchical model = 2-stage LM



Fixed effects

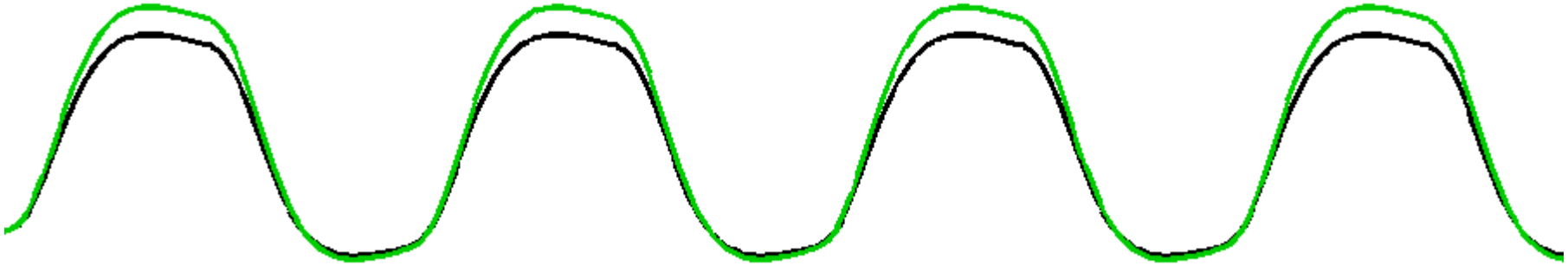
- Only source of variation (over trials) is measurement error
- True response magnitude is *fixed*

Random effects



- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject has random magnitude

Random effects

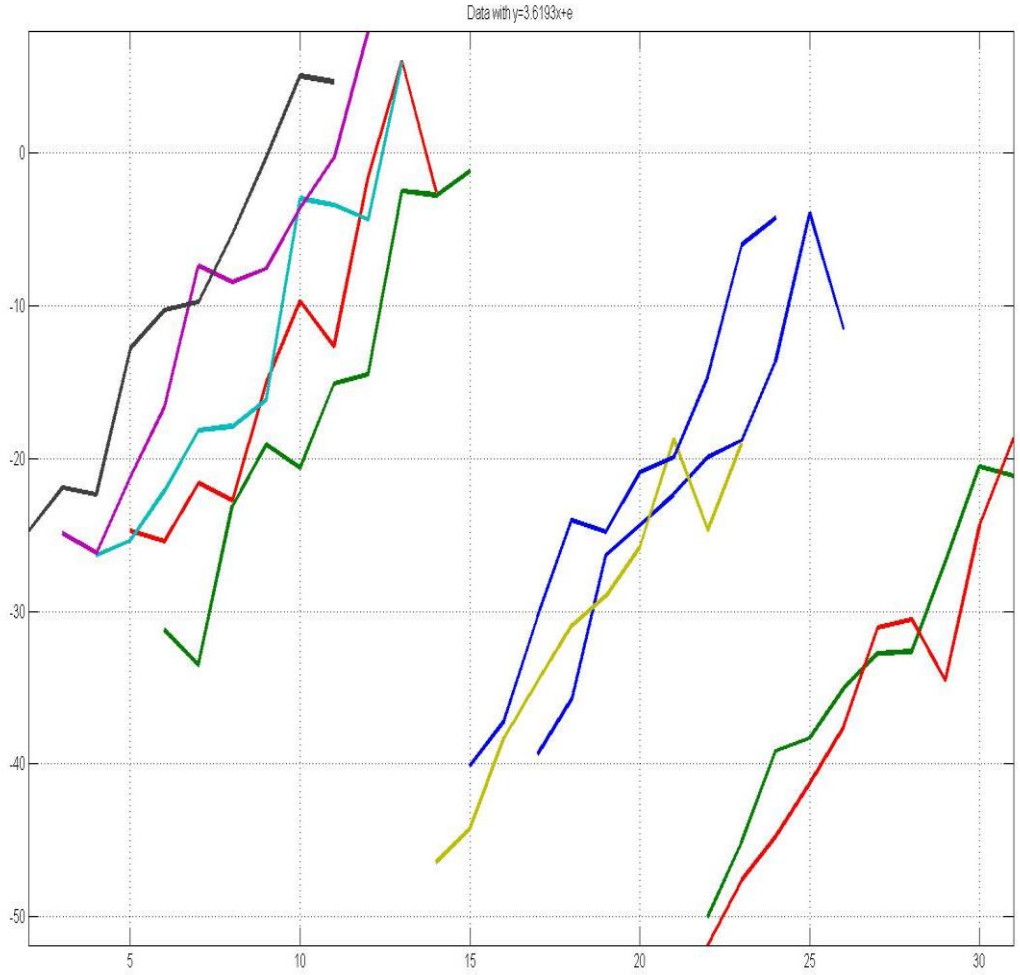


- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject has random magnitude
  - but note, population mean magnitude is *fixed*

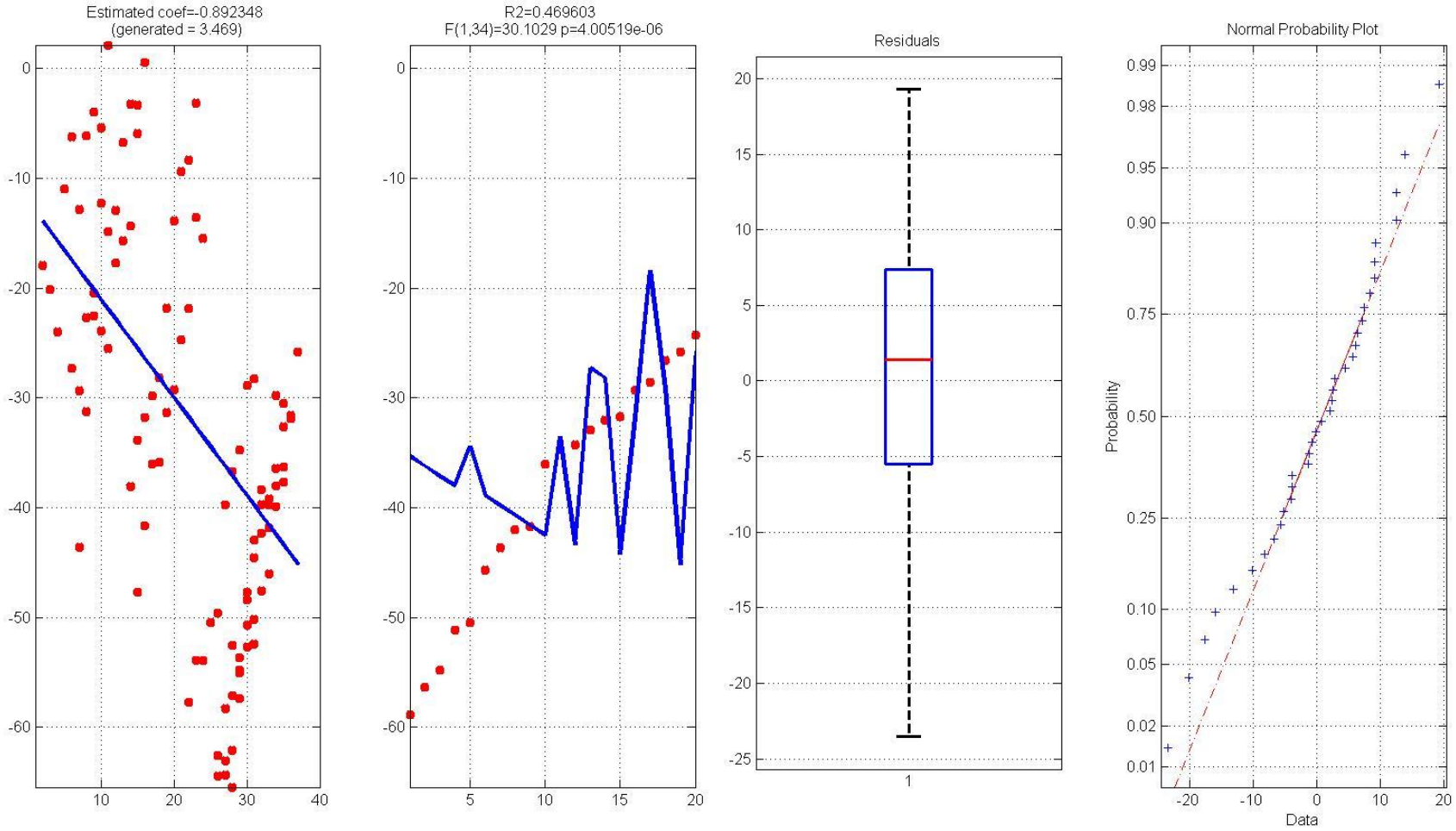
An example

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

→ There is a strong positive effect of intensity on responses

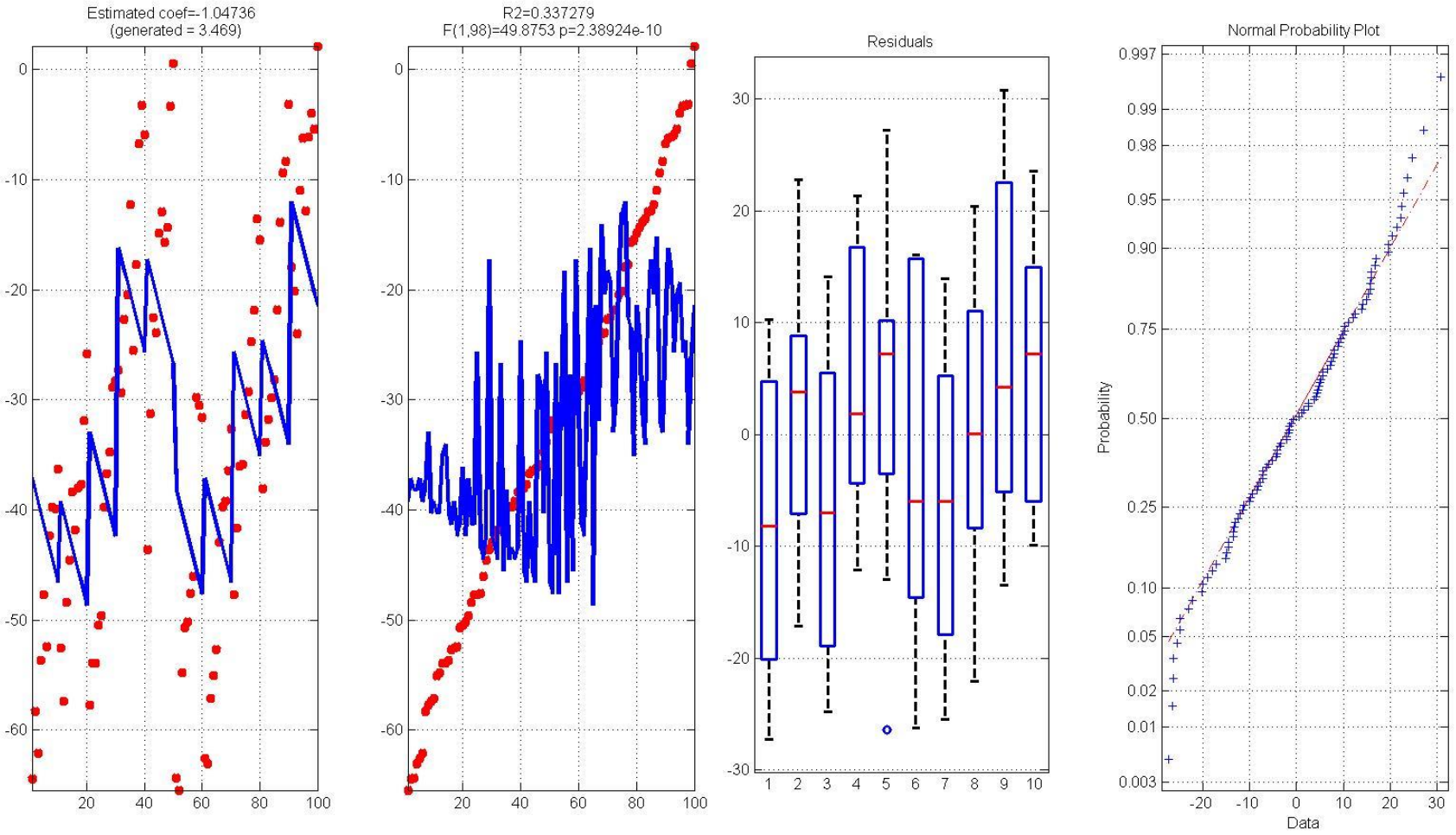


Fixed Effect Model 1: average subjects



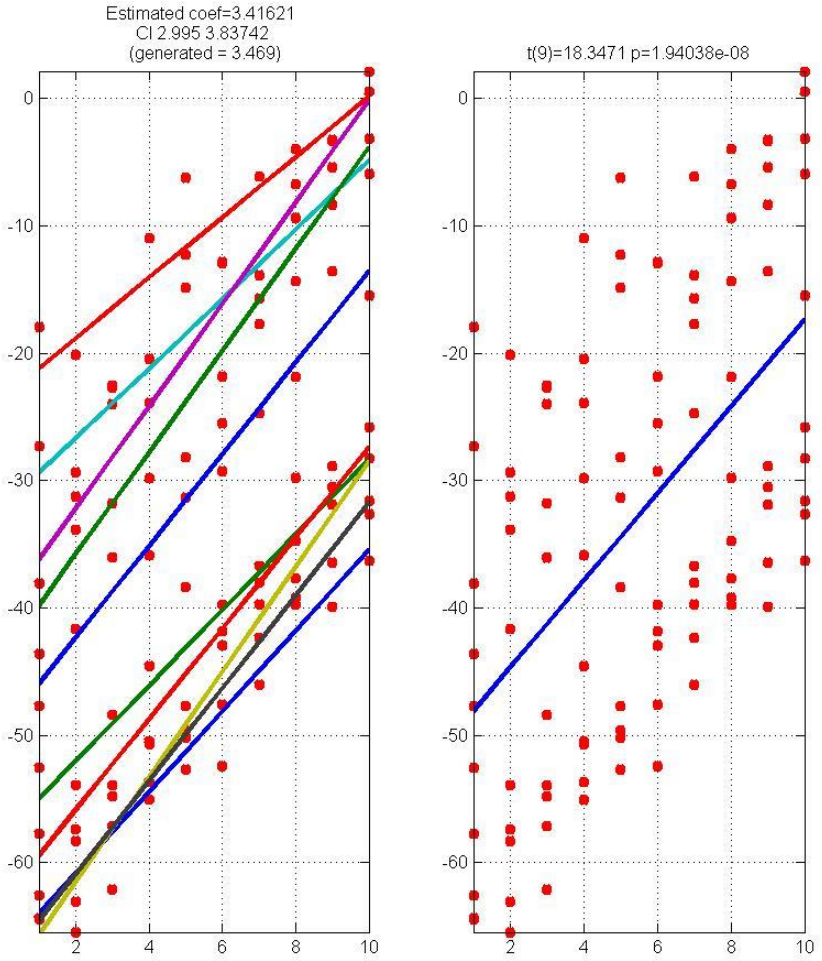
Fixed effect without subject effect → negative effect

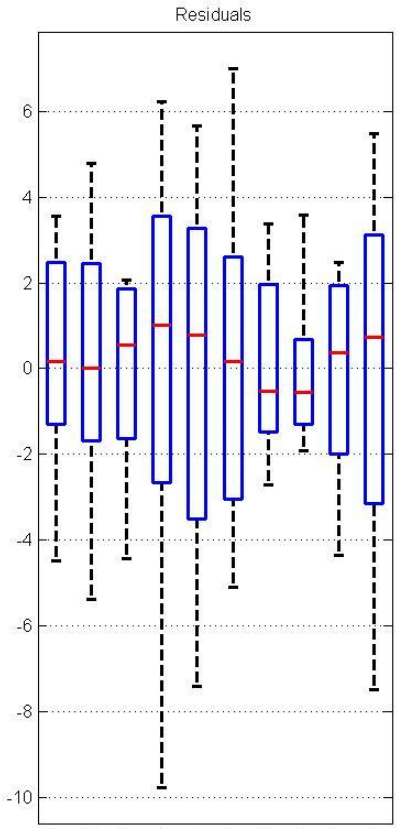
Fixed Effect Model 2: constant over subjects

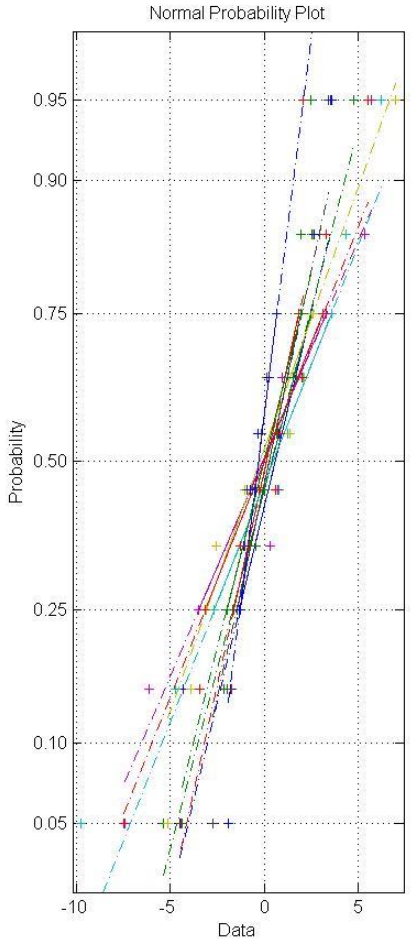


Fixed effect with a constant (fixed) subject effect → positive effect but biased result

HLM: random subject effect

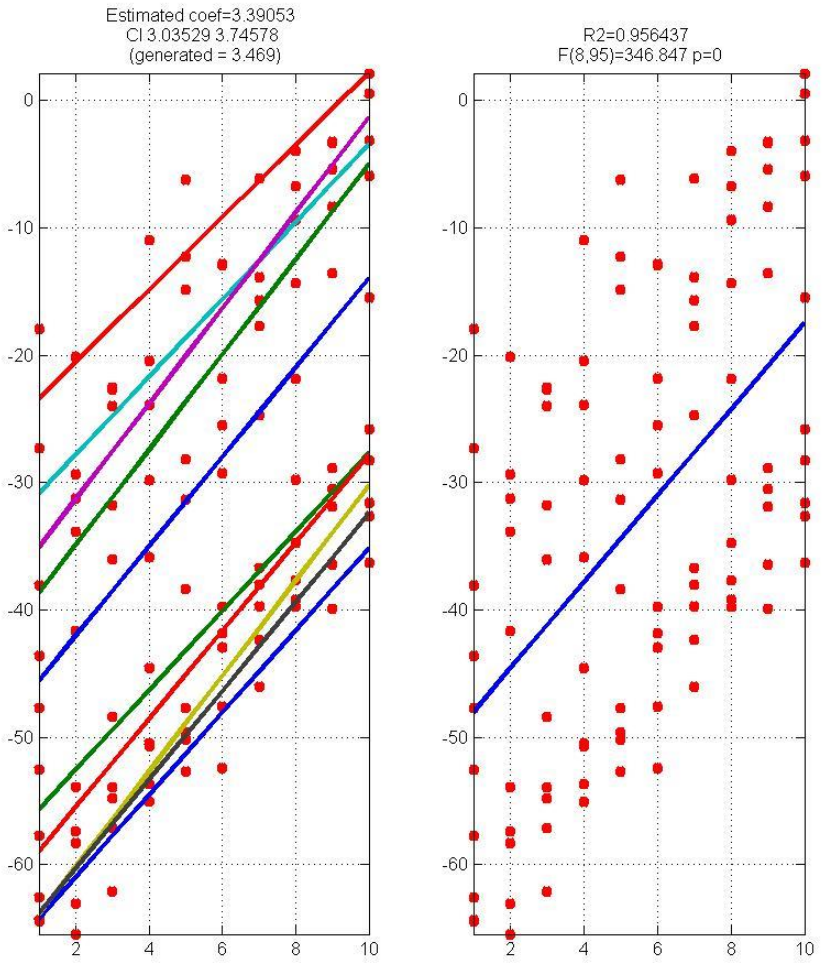


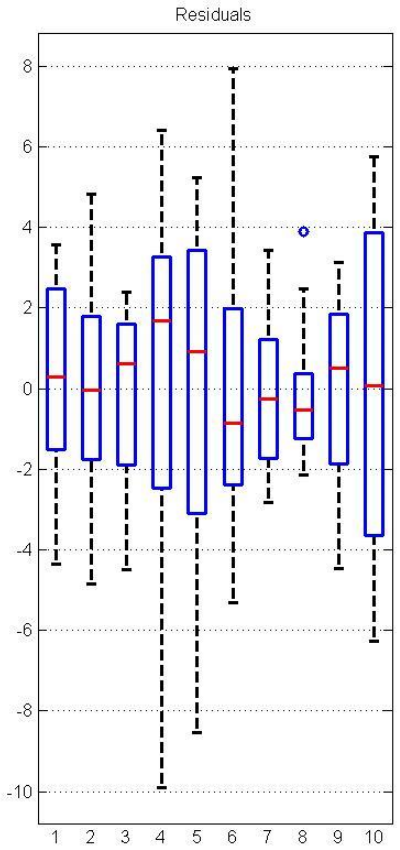


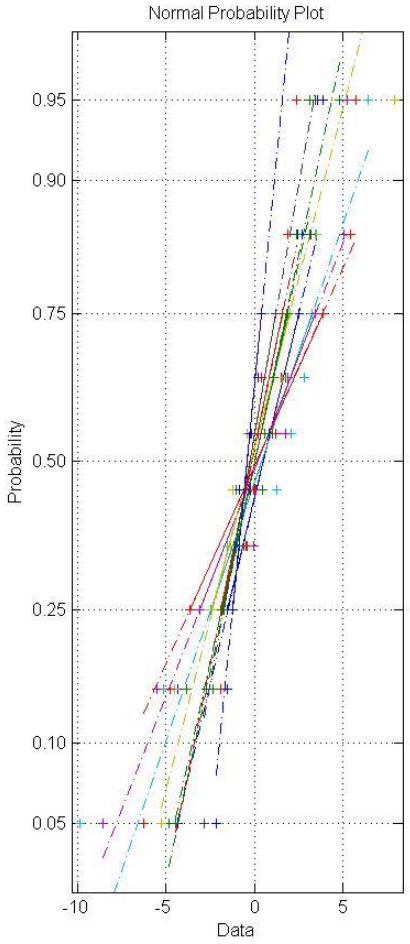


Mixed effect with a random subject effect → positive effect with good estimate of the truth

MLE: random subject effect

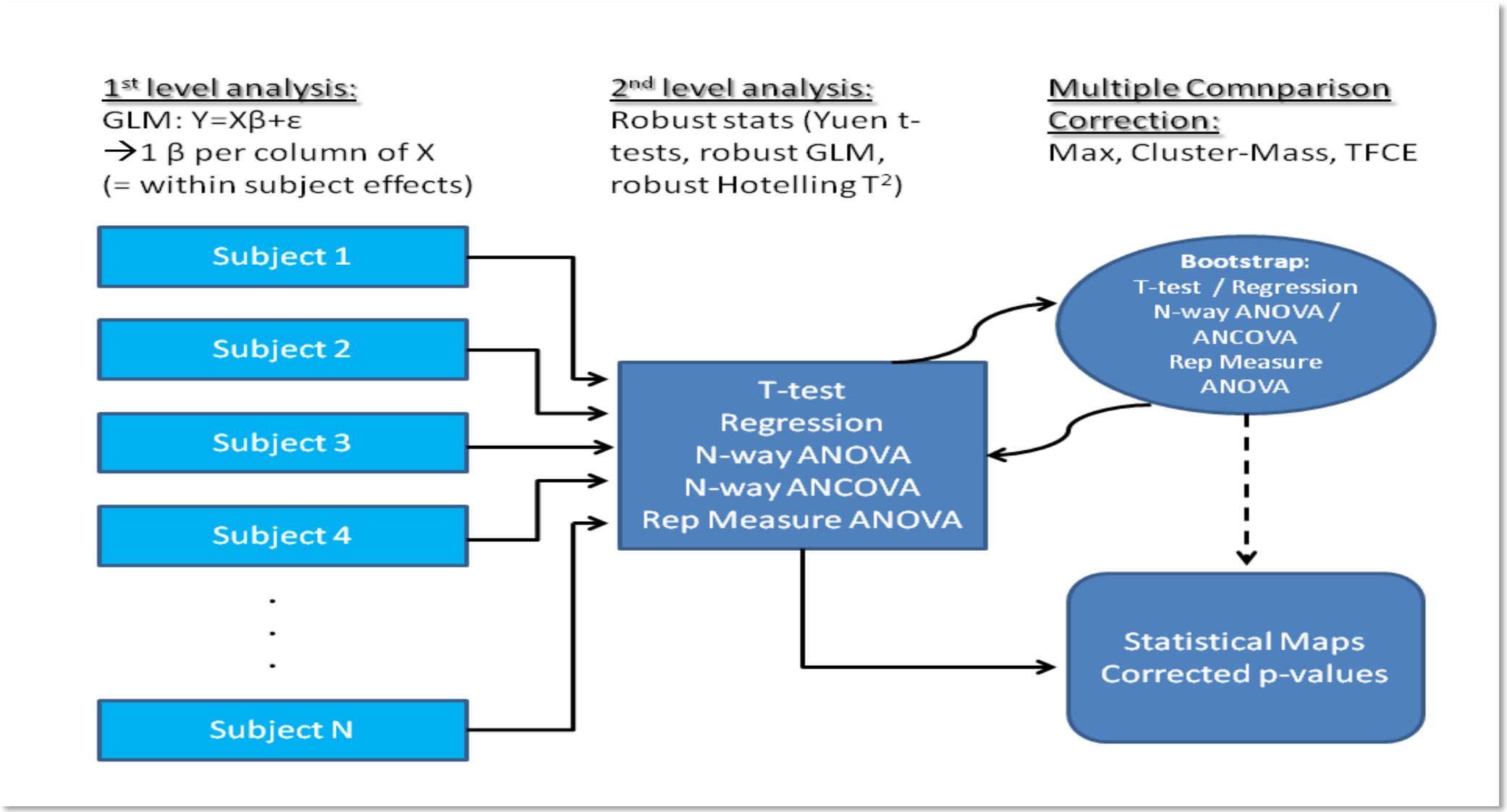






Mixed effect with a random subject effect → positive effect with good estimate of the truth

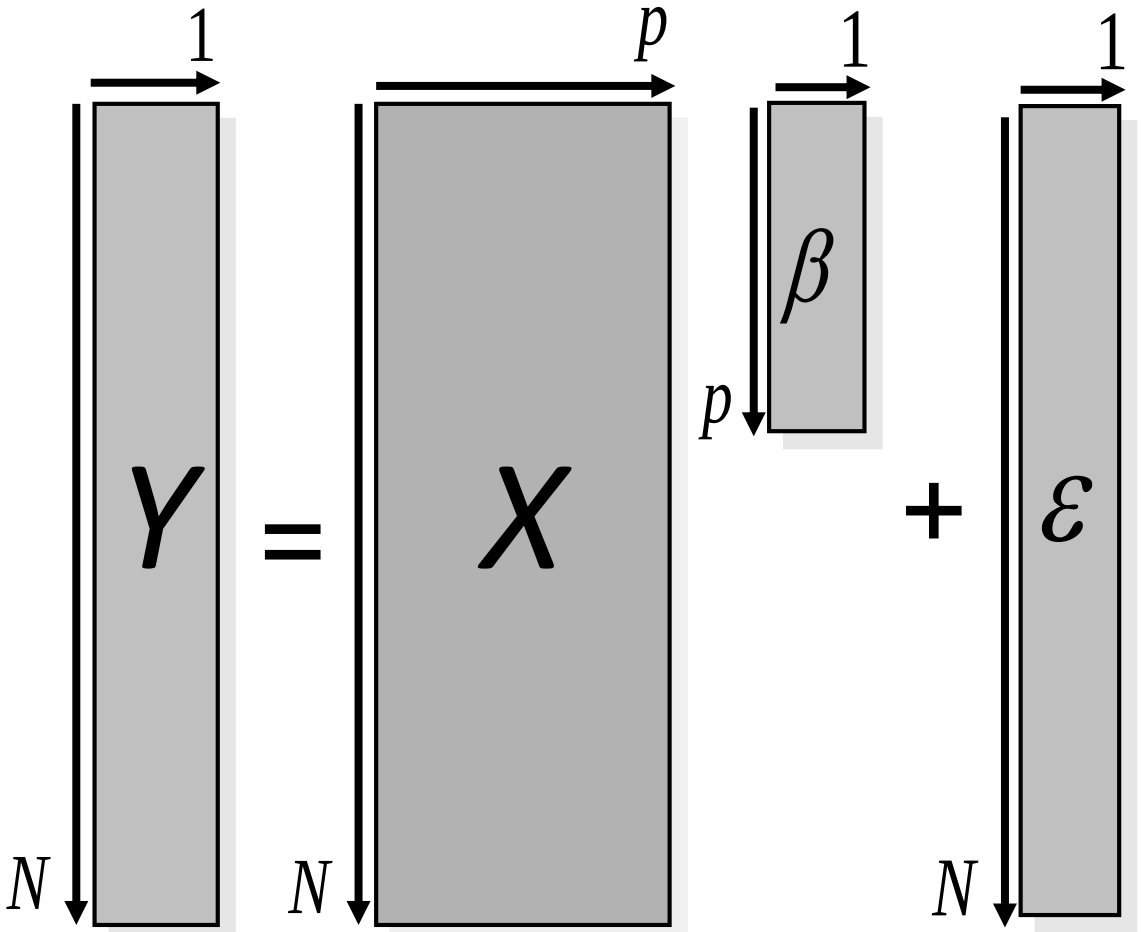
Hierarchical Linear Model for MEEG



General Linear model (reminder?)

- Model: assign to the data different effects / conditions ... All we have to do is find the parameters of this 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)
- 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)

General Linear model (reminder?)



$$y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

Estimate with Ordinary or Weighted Least Squares

N: number of trials
p: number of regressors

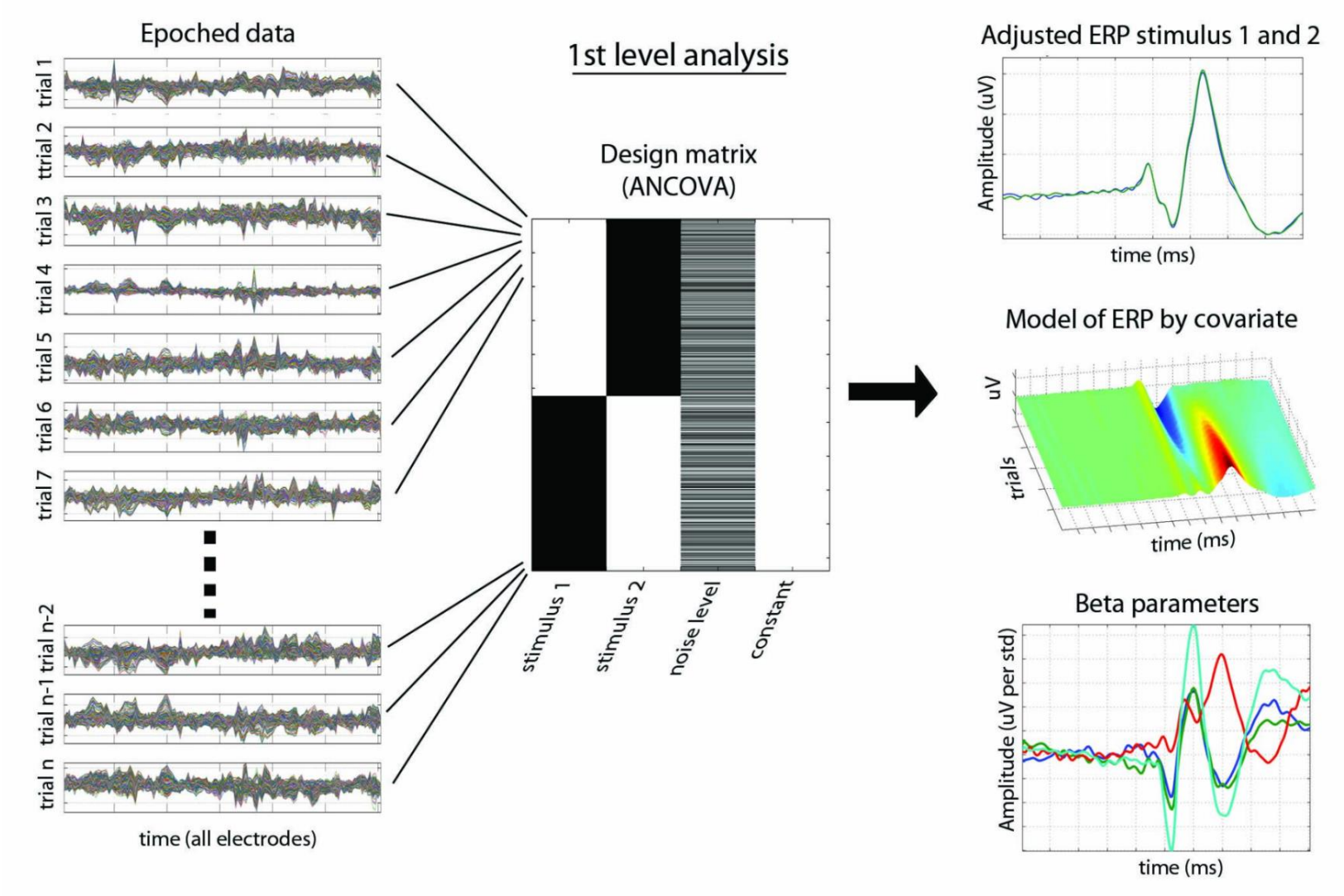
Model is specified by
1. Design matrix *X*
2. Assumptions about ε

The LIMO EEG data set

- 18 subjects
- Simple discrimination task face 1 vs face 2
- Variable level of noise for each stimulus – noise here is in fact a given amount of phase coherence in the stimulus

Rousselet, Pernet, Bennet, Sekuler (2008). Face phase processing. *BMC Neuroscience 9:98*

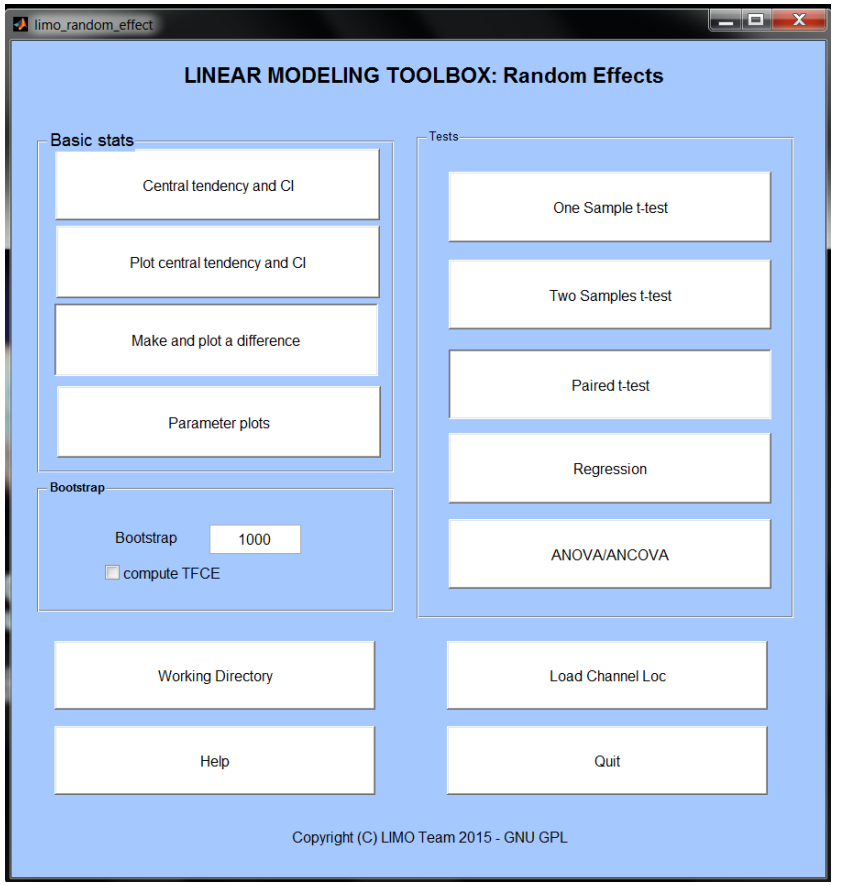
EEG 1st level = GLM (any designs !)



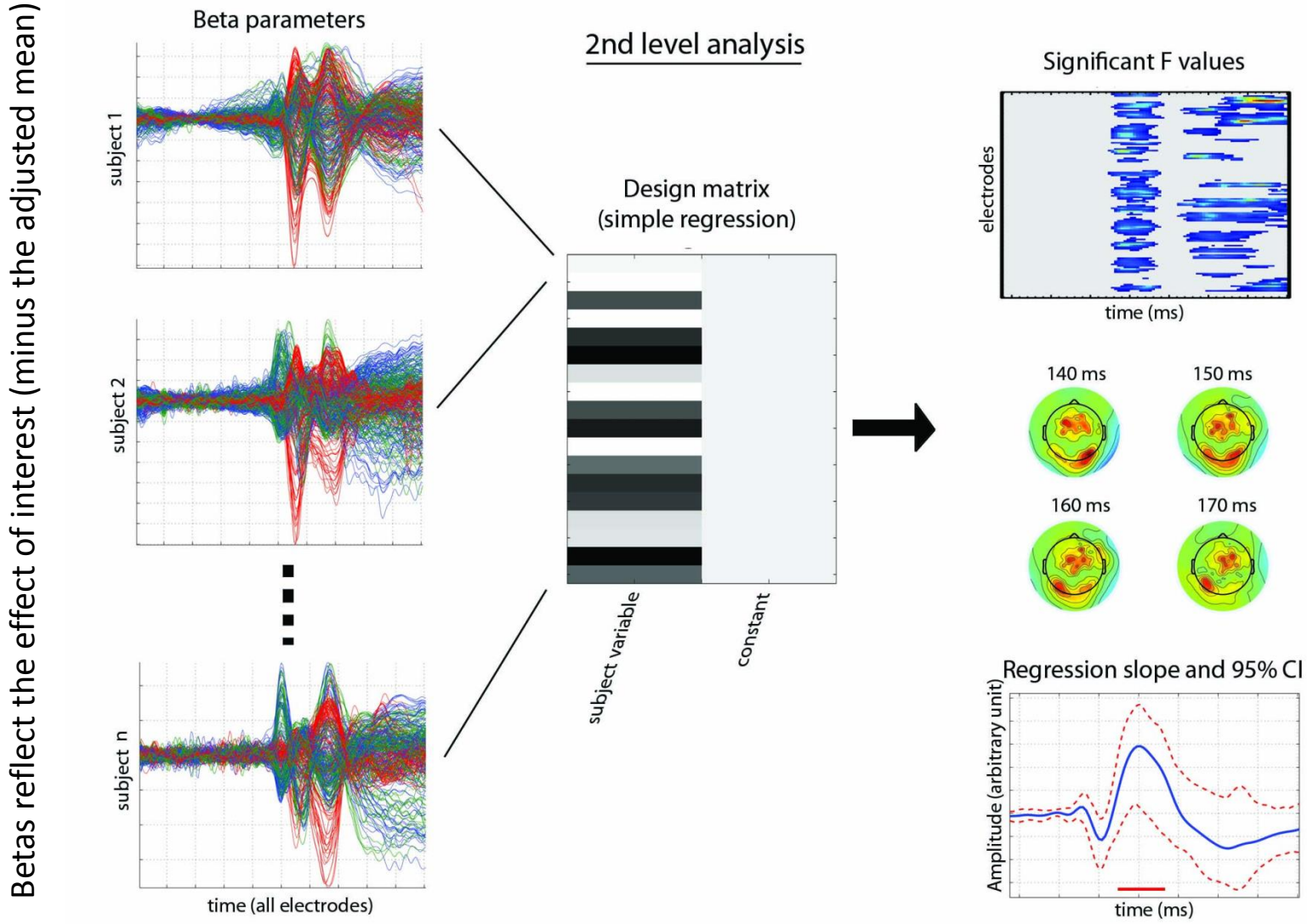
EEG 2nd level (usual tests but robust)

- We have 18 subjects of various ages -> how is the processing of phase information (beta 3) influenced by age.

- 2nd level analysis GUI
- Use the same channel location file across subjects (no channel interpolation)
- Regress the effect of age (2nd level variable) on the effect of phase on the EEG (1st level variable)
- Use multiple comparison correction using bootstrap



EEG 2nd level



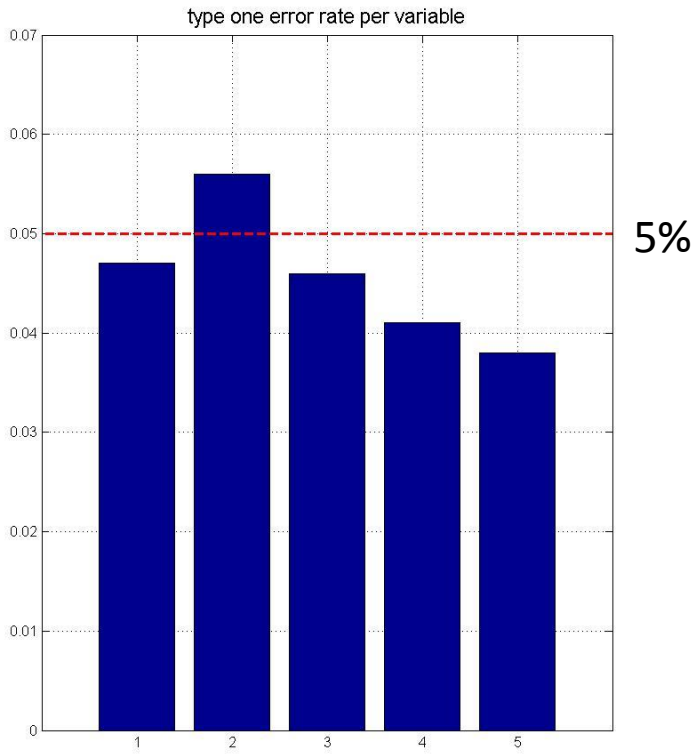
Multiple Comparison Correction for MEEG

- Assuming tests are independents from each other, the family-wise error rate FWER = 1 - (1 - alpha)^n

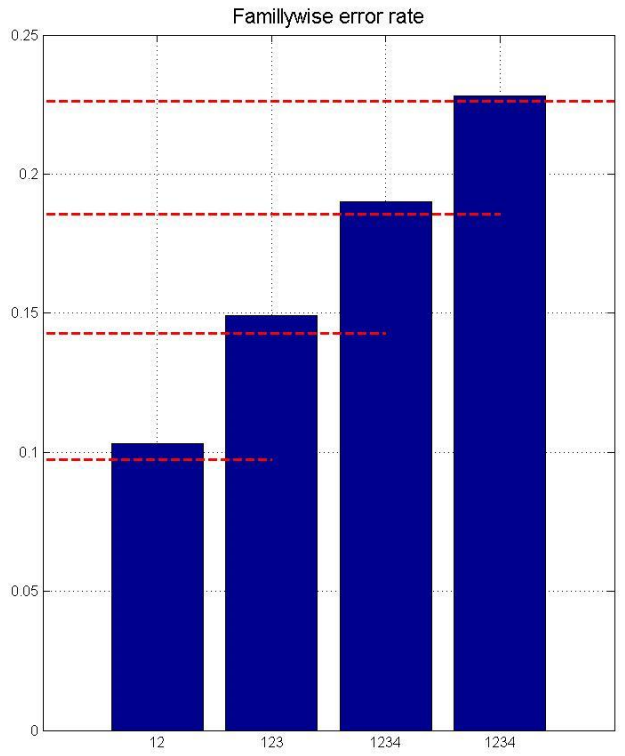
- for alpha =5/100, if we do 2 tests we should get about 1-(1-5/100)^2 ~ 9% false positives, if we do 126 electrodes * 150 time frames tests, we should get about 1-(1-5/100)^18900 ~ 100% false positives! i.e. you can’t be certain of any of the statistical results you observe

What is the problem?

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



5%



22%

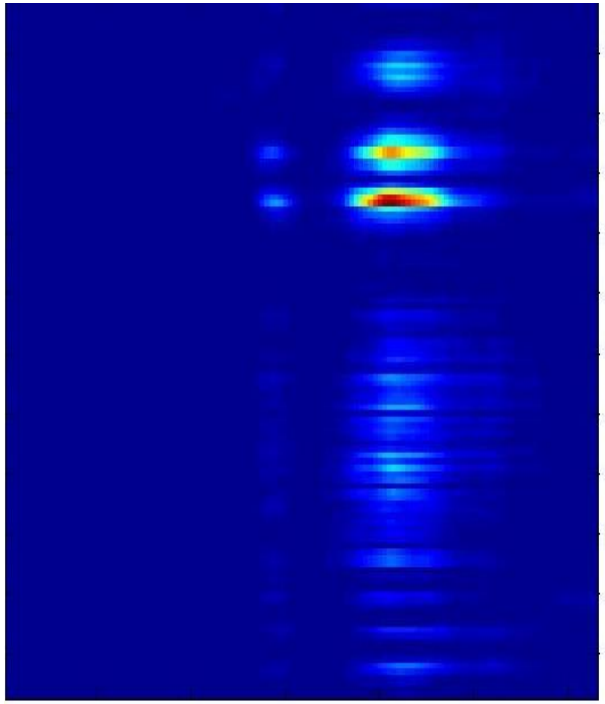
18%

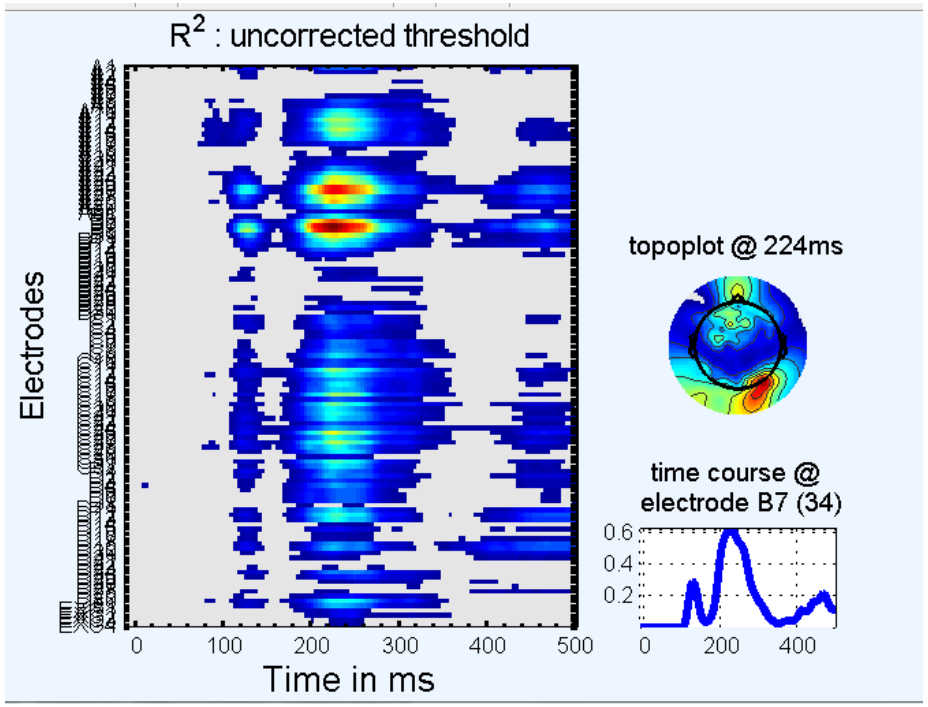
14%

9%

What is the problem?

- Illustration with 18900 independent variables (126 electrodes and 150 time frames)





we know there are false positives – which ones is it?

Family Wise Error rate

- FWER is the probability of making one or more Type I errors in a family of tests, under H0
- H0 = no effect in any channel/time and/or frequency bins → implies that rejecting a single bin null hyp. is equal to rejecting H0

$$P(\mathrm{U}_{i \in V}\{T_i \geq u\}|H_0) \leq \propto$$

We want to find the threshold u such the prob of any false positives under H0 is controlled at value alpha

Bonferroni Correction

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

$$P(T_i \geq u|H0) \leq \frac{\propto}{m}$$ *Find u to keep the FWER < α/m*

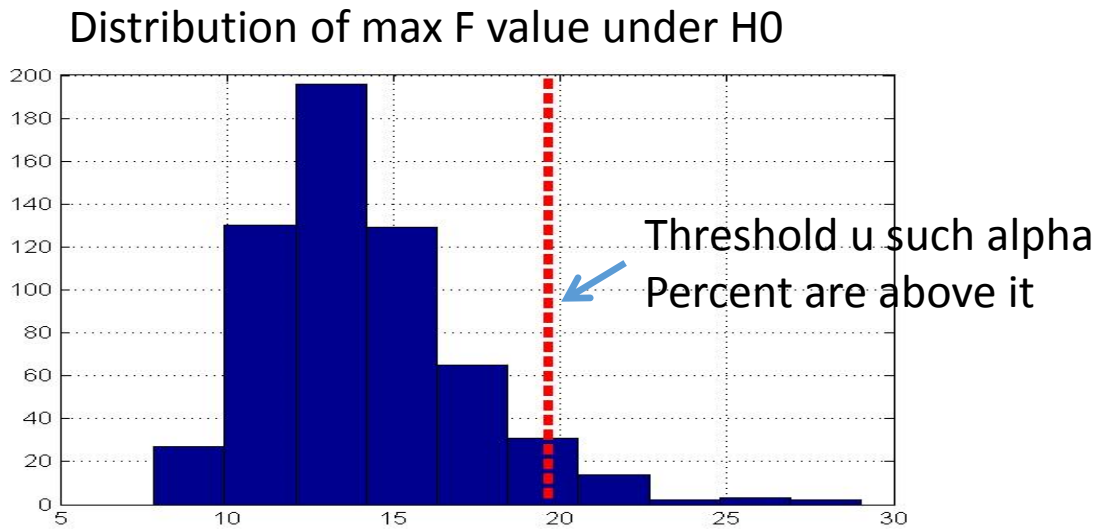
$$\text{FWER} = P(\text{U}_{i \in V}\{T_i \geq u\}|H_0) \leq \propto$$

$$\leq \sum P(T_i \geq u|H0)$$ *Boole's inequality*

$$\leq \sum_i \frac{\propto}{m} = \propto$$

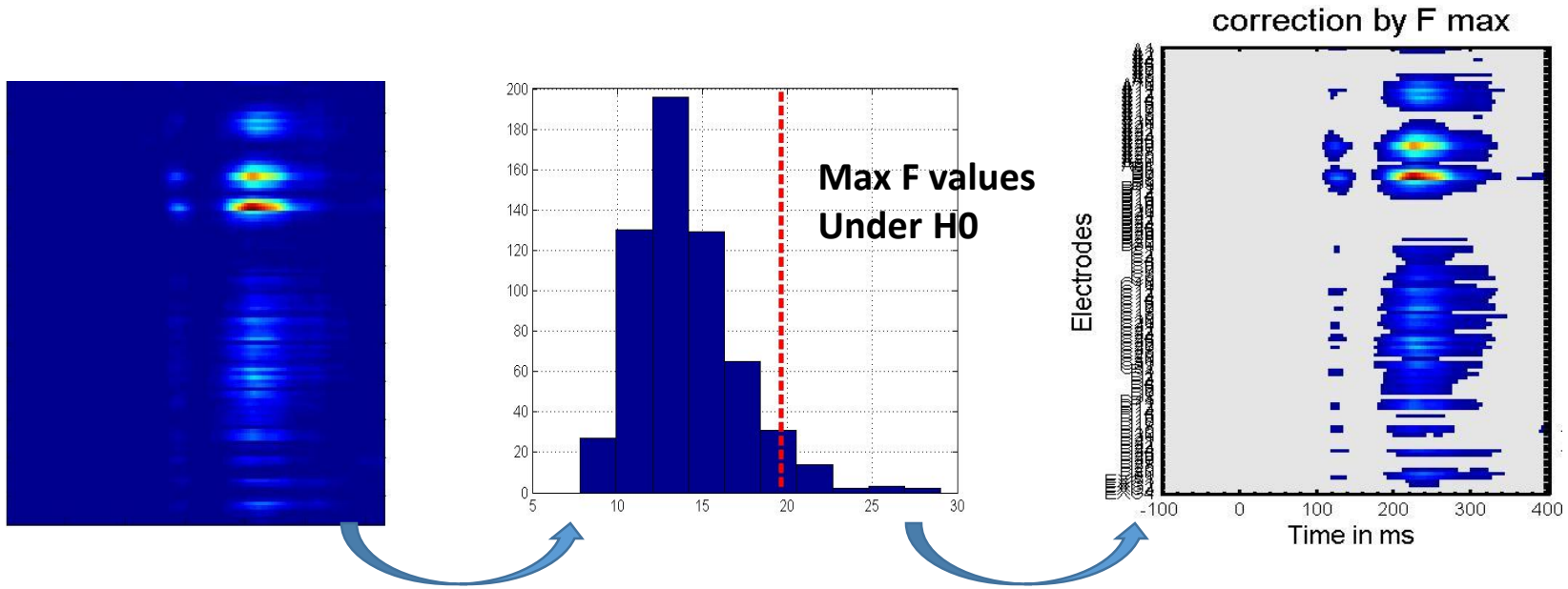
Maximum Statistics

- Since the FWER is the prob that any stats > u, then the FWER is also the prob. that the max stats > u
- All we have to do, is thus to find a threshold u such that the max only exceed u alpha percent of the time.



Maximum Statistics

- Estimate the distribution of max under H0 (bootstrap) and simply threshold the observed results a threshold u
- Still assumes all tests are independent



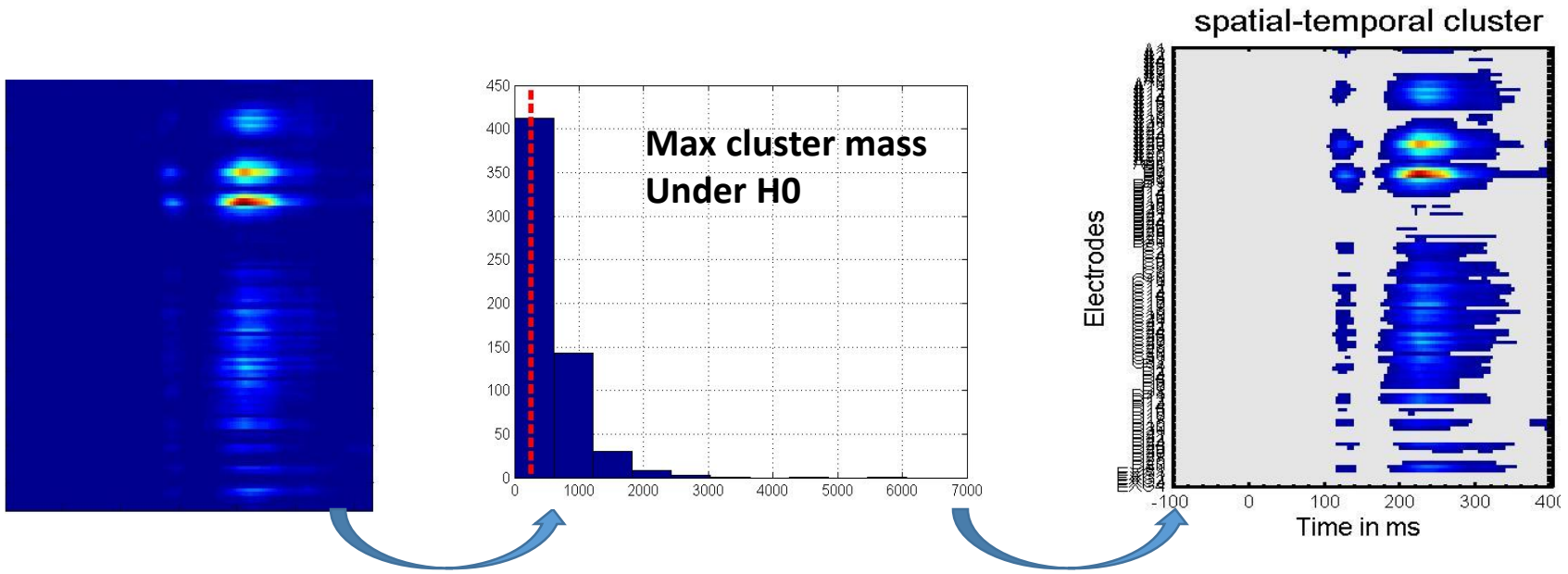
The clustering solution

- Clustering is an alternative, more powerful option that accounts for topological features in the data. Techniques like Bonferroni, FDR, max(stats) control the FWER but independently of the correlations (in time / frequency / space) 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 and in time/frequency), and (ii) the strength of the signal (how strong are the t, F values in a cluster) or (iii) a combination of both.

The clustering solution

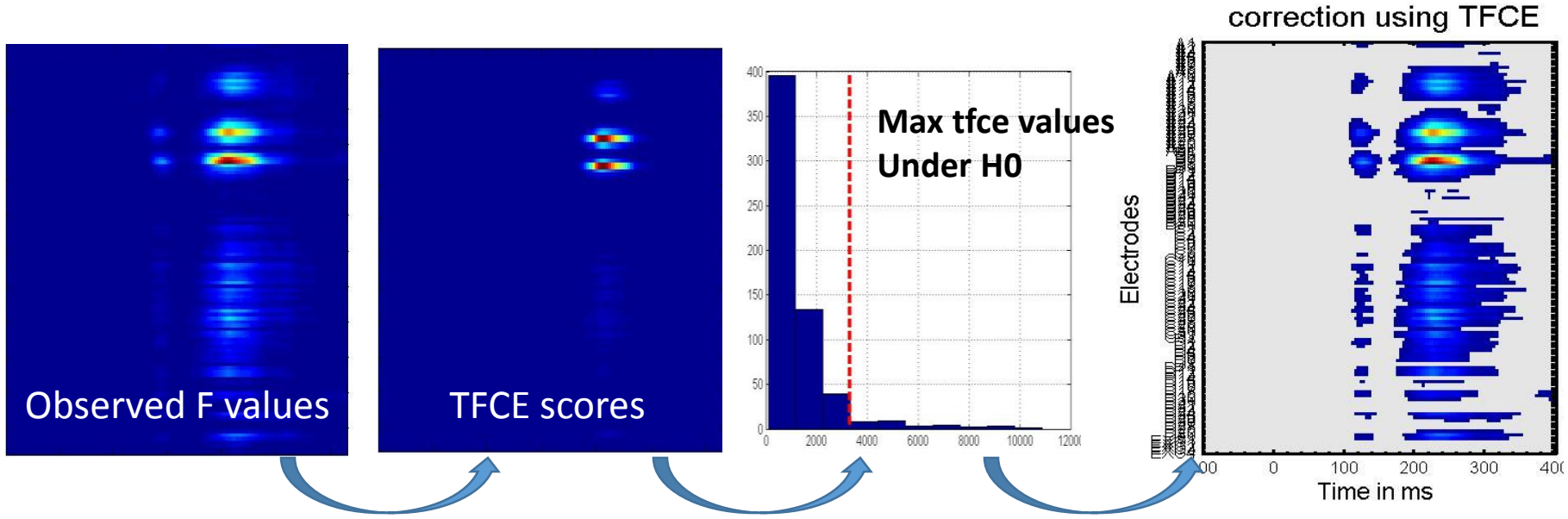
- Spatial-Temporal clustering: for each bootstrap, threshold at alpha and record the max(cluster mass), i.e. sum of F values within a cluster. Then threshold the observed clusters based on there mass using this distribution → accounts for correlations in space and time.



Loss of resolution: inference is about the cluster, not max in time or a 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. As before, bootstrap under H0 and get max(tfce).



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

Modern Analysis of EEG data

- Selection of channels and frequency bins must be independent – without good priors, we can analyse the whole space
- Amplitude and Peaks are related, simply analyse the whole space continuously
- Use HLM to account for variance across trials and model the random subject effect
- Use a (robust) GLM at 1st level to model data – any designs and covariates can be accounted for.
- Use (robust) group level statistics to infer effects in space / time / frequency while controlling the type 1 FWER.

References

- **Maris, E. & Oostenveld, R. (2007).** Nonparametric statistical testing of EEG- and MEG-data. Journal of *Neuroscience Methods, 164*, 177-190
- **Pernet, C., Chauveau, N., Gaspar, C. & Rousselet, G (2011).** Linear Modelling of MEEG. Comp. Intel. Neurosc. Article ID 831409
- **Pernet, C., Latinus, M., Nichols, T. & Rousselet, G.A. (2015).** Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. Journal of *Neuroscience Methods, 250*, 85-93