21st EEGLAB Workshop
Thursday 7th April 2016
Santa Margherita Ligure, Italy

Hierarchical Linear Modelling for EEG data

Cyril Pernet, PhD
Edinburgh Imaging
Centre for Clinical Brain Sciences

Context

- In all cases, 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) tree?

➢ 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)

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

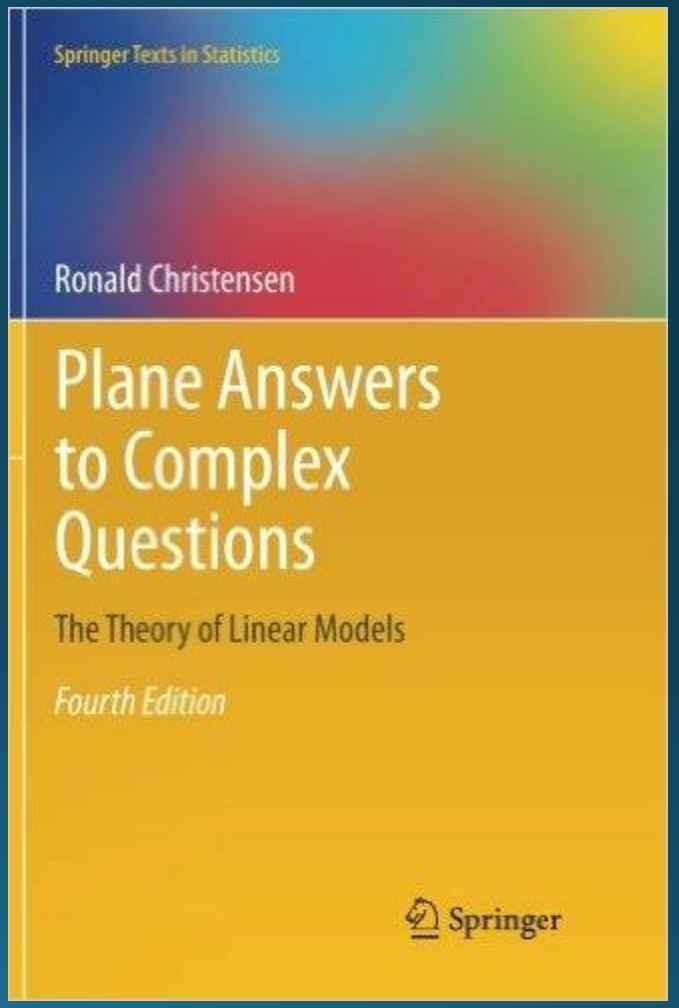
Overview

- → Setting up a study (again)
- Fixed, Random, Mixed and Hierarchical
- A extreme example
- GLM overview
- Weighted Least Squares for EEG

(not covered in the talk but you have it here anyway)

→ Review results / set 2nd level

- A word on designs



STUDY

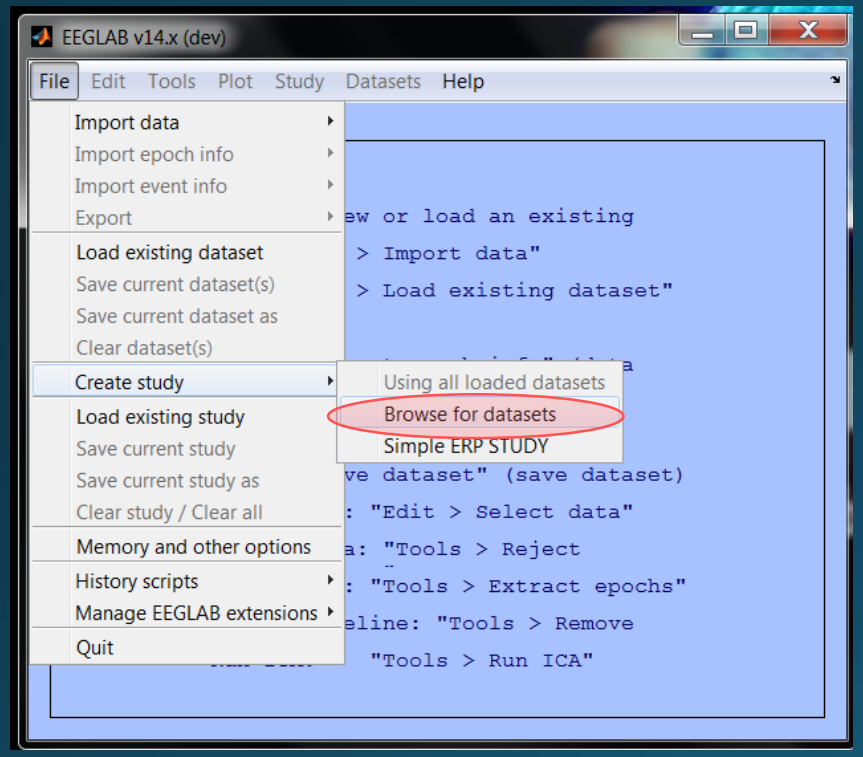
The Steinberg Experiment

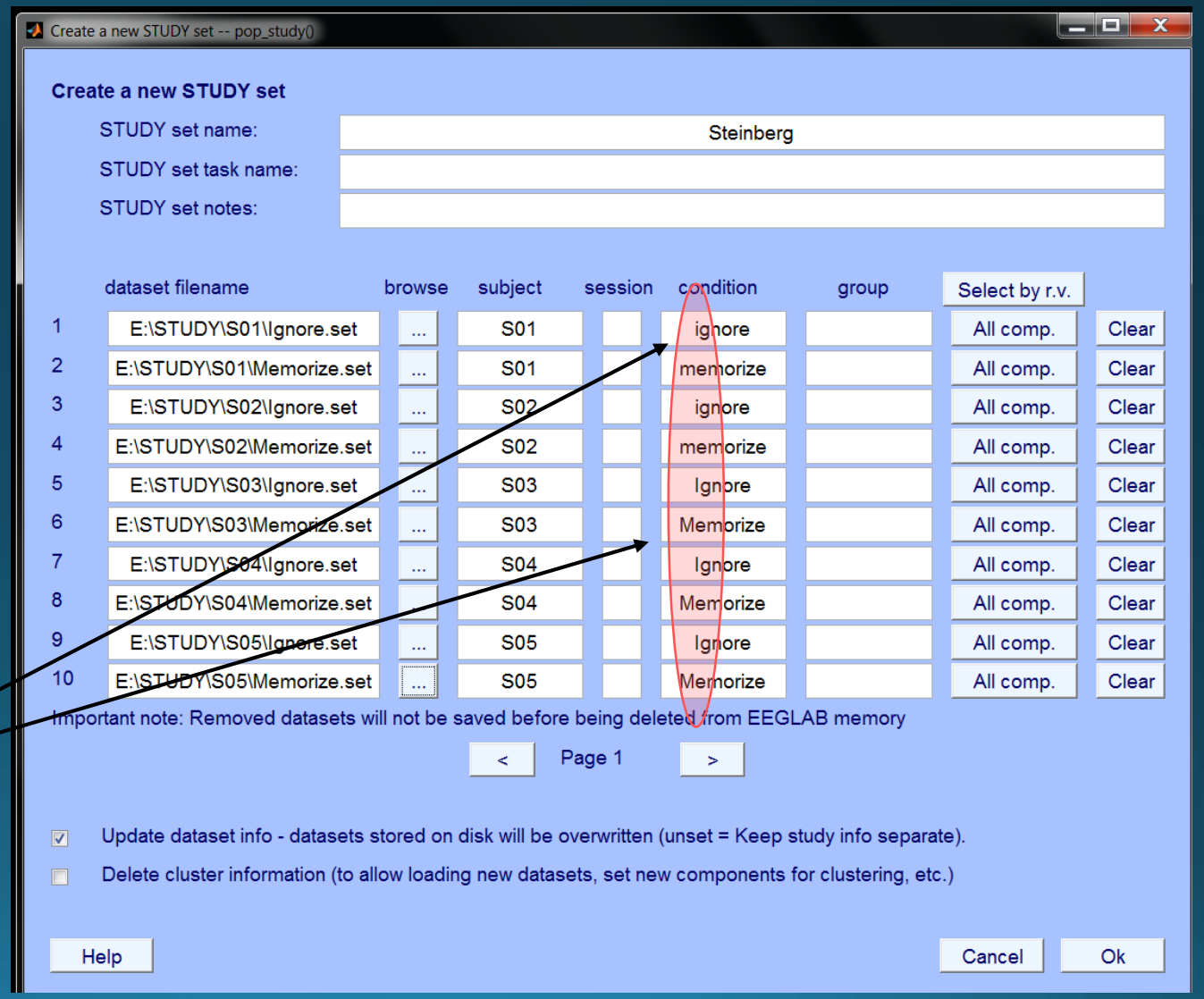
Quick overview

- Sternberg working memory task
- Ignore/Memorize → Maintain → Probe
- 8 items each time → variable load to memorize

- Load your study from yesterday
- Precompute single trials
- If ICA, do the clustering
- Create a design and run it through the LIMO EEG toolbox

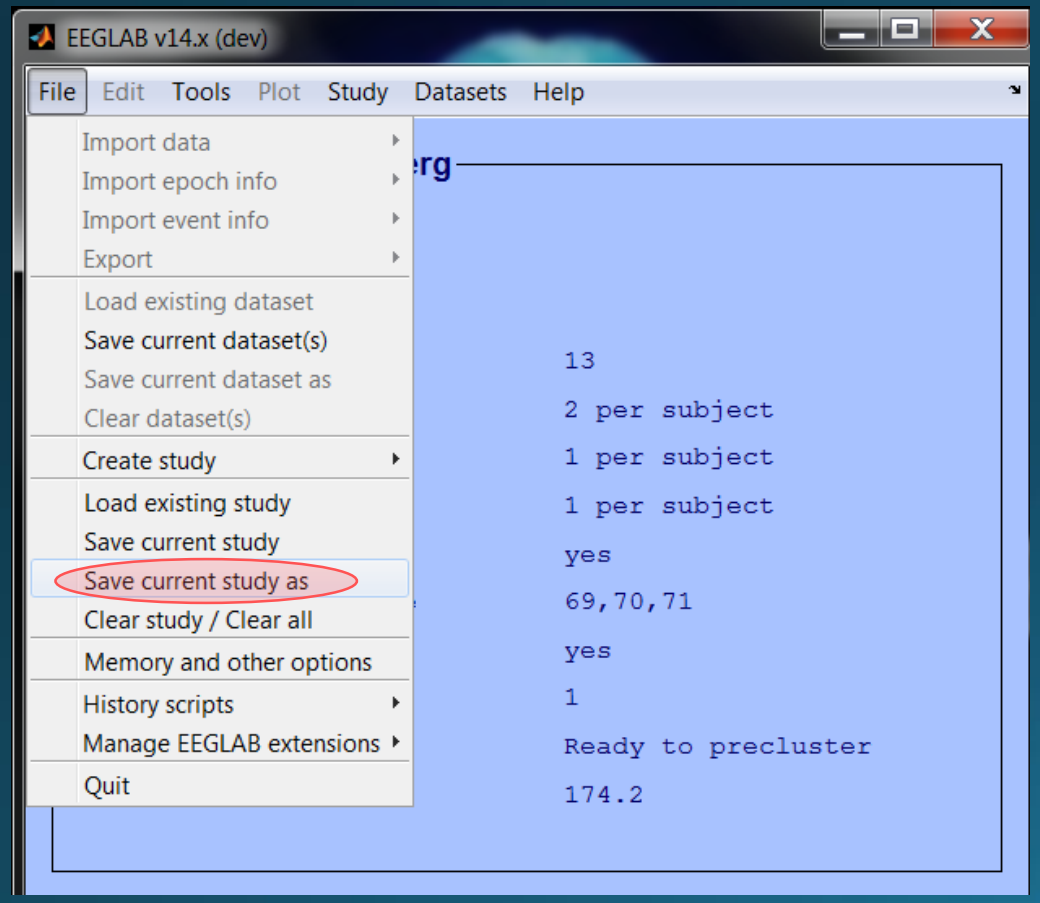
Creating a study and LIMO stats





ISSUE: Make sure the condition names are the same for all subjects (case sensitive)

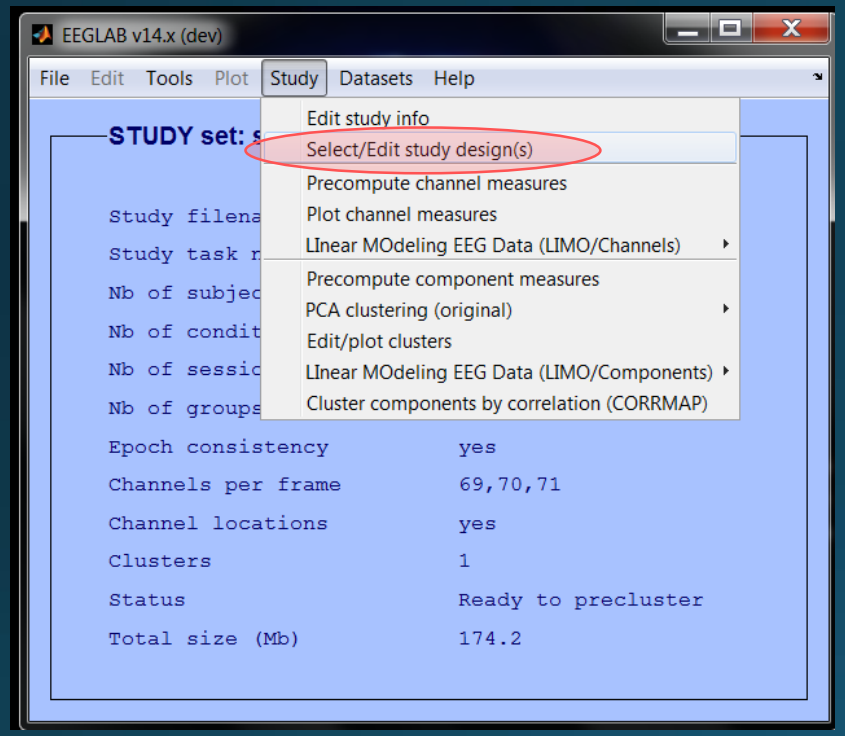
Creating a study and LIMO stats

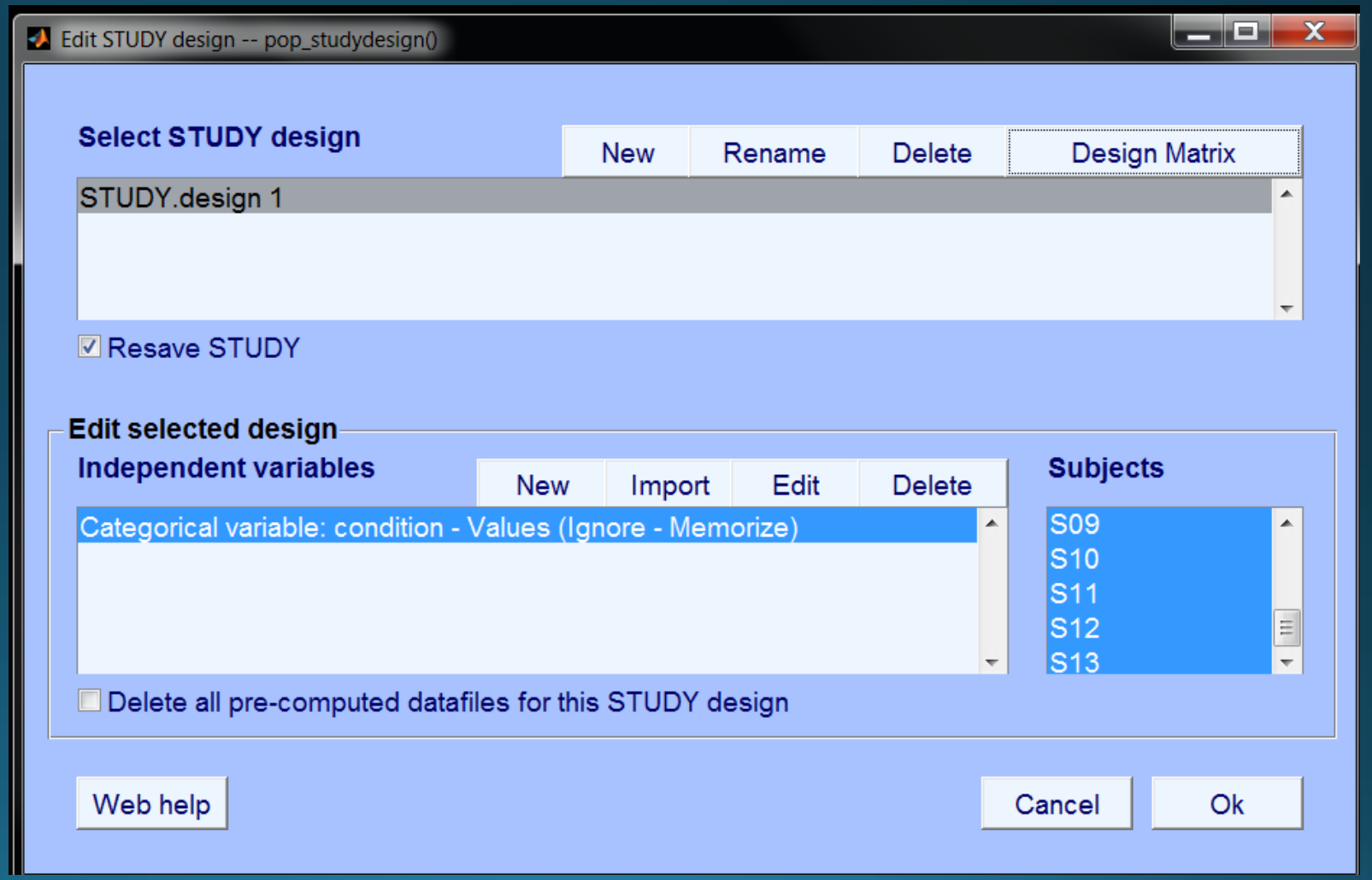


Save current study as

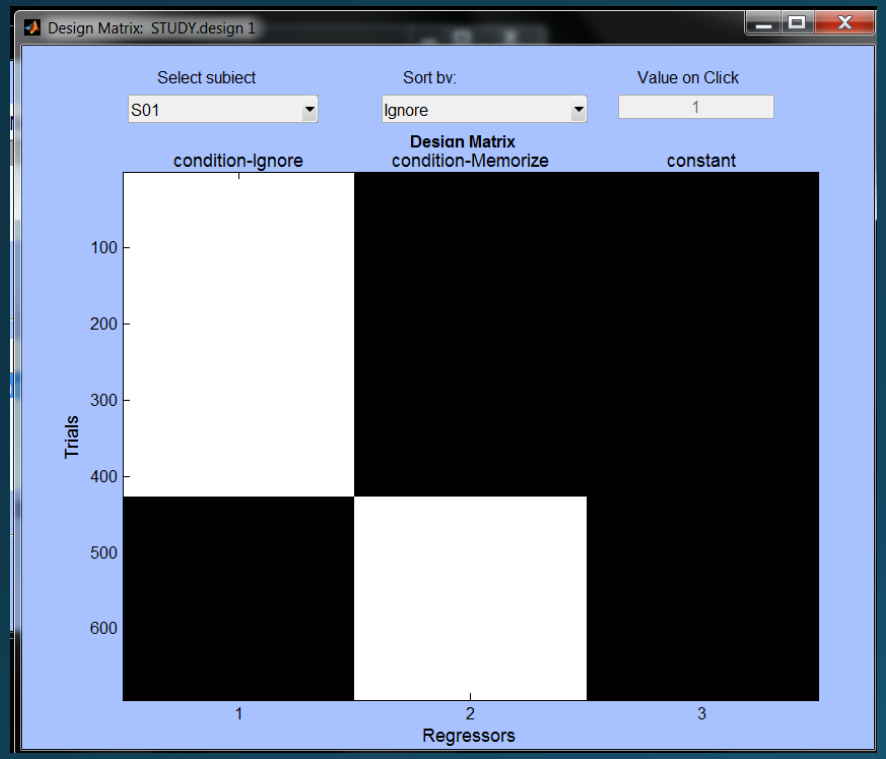
Currently a small bug, there is no name saved (ie 'save current study' doesn't work and nothing is saved by default)

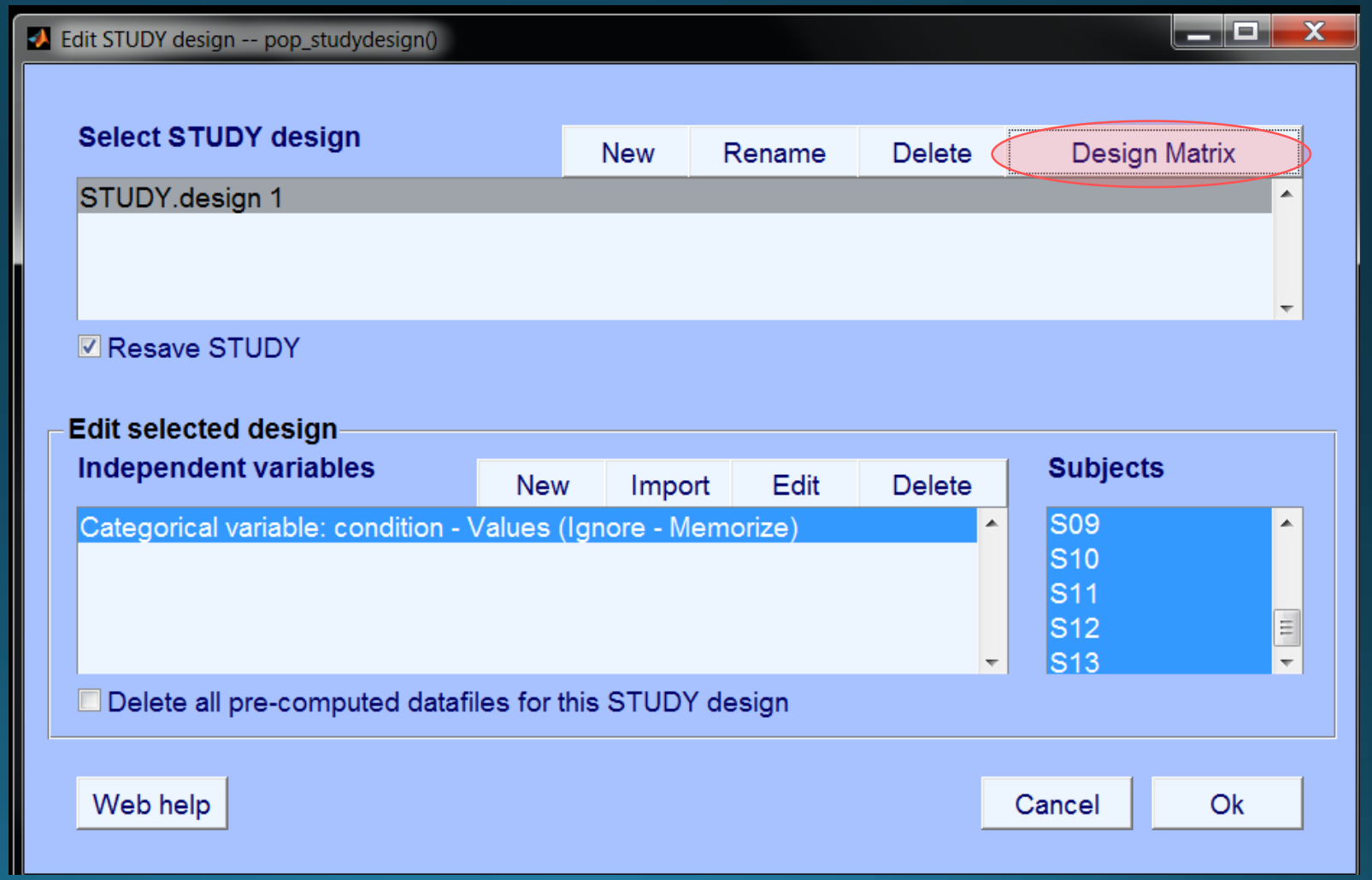
Creating a study and LIMO stats



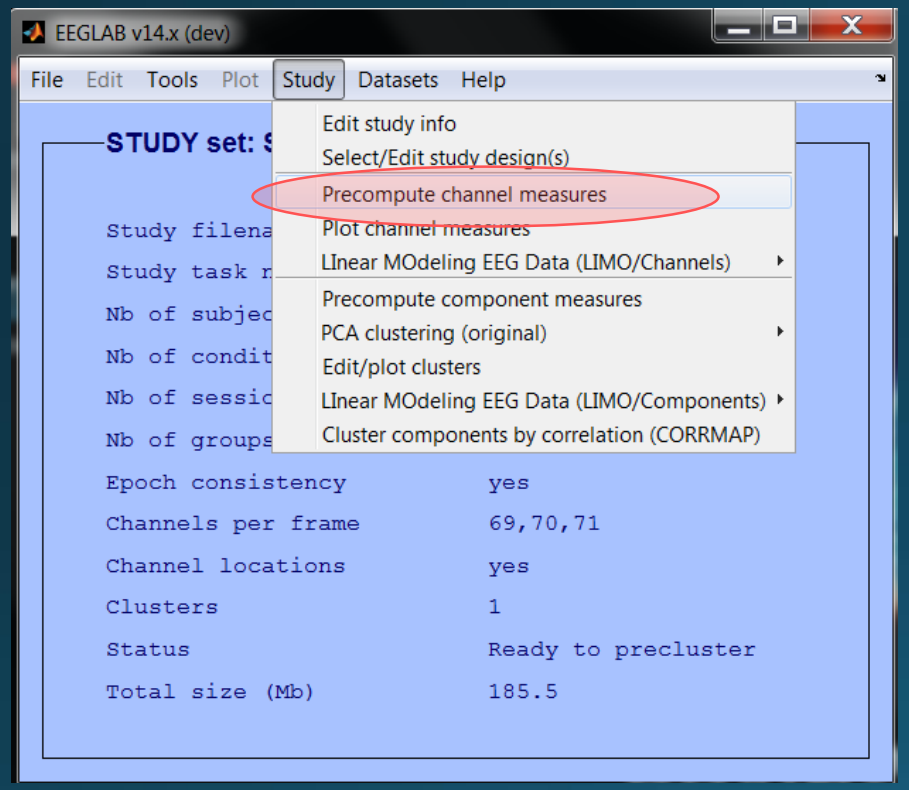


Creating a study and LIMO stats

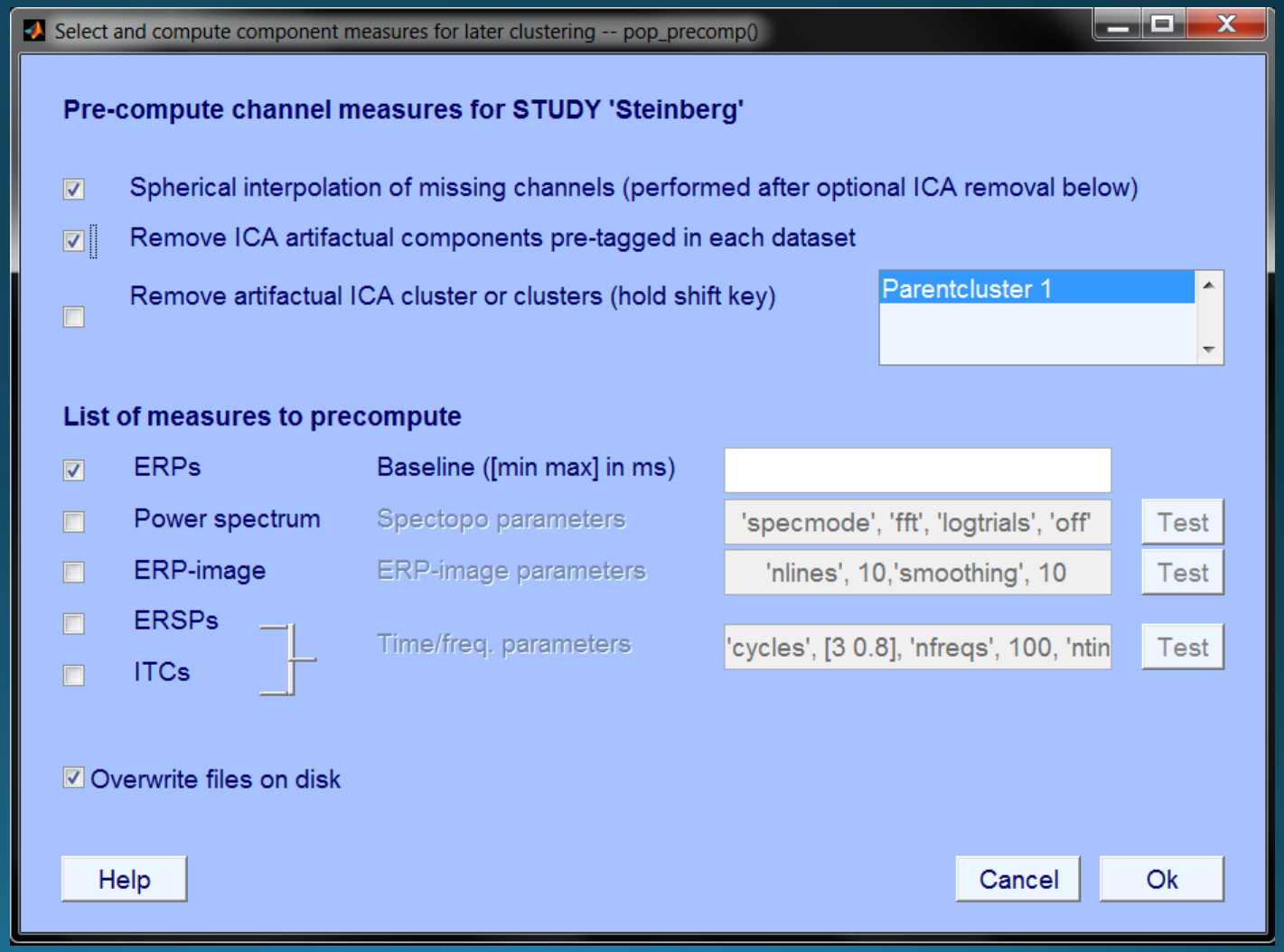




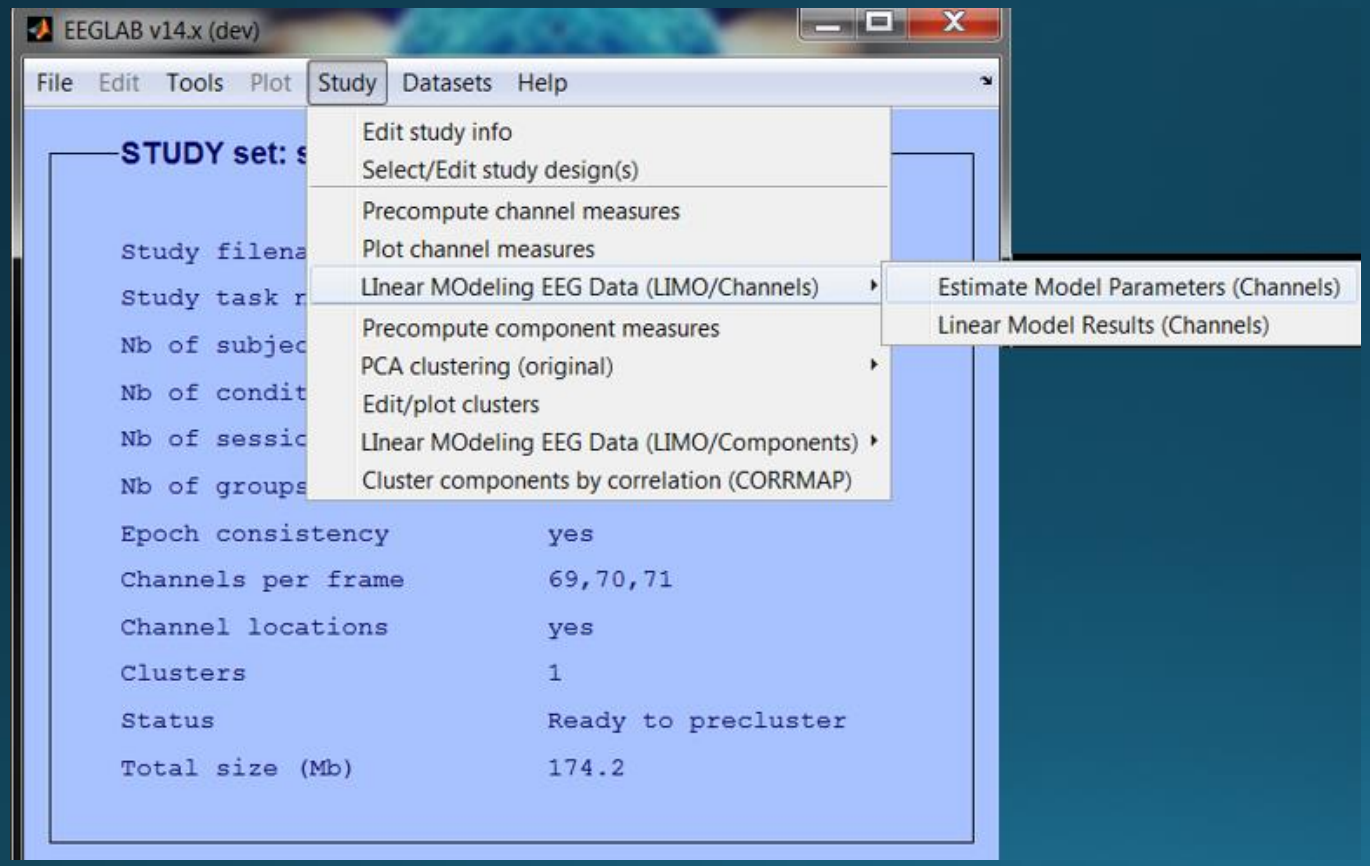
Creating a study and LIMO stats

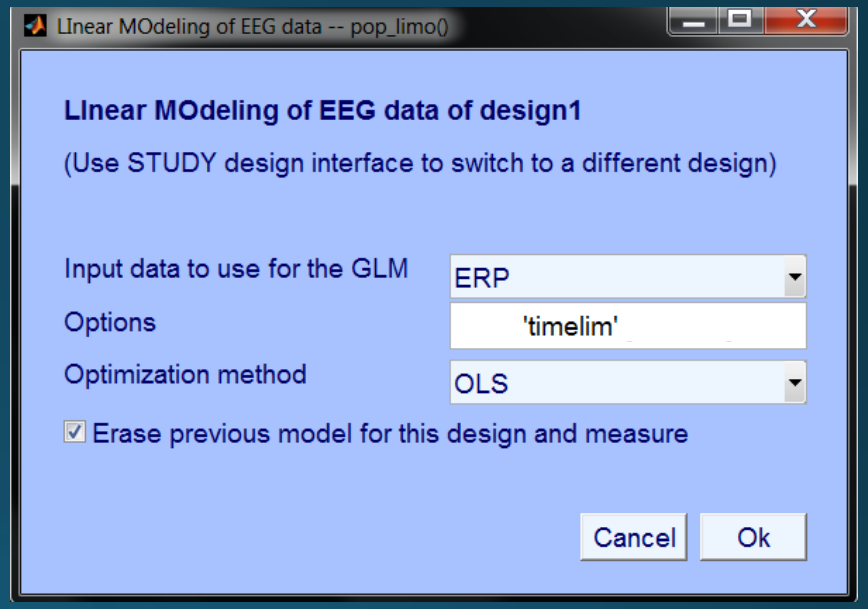


saves .daterp files = single trials



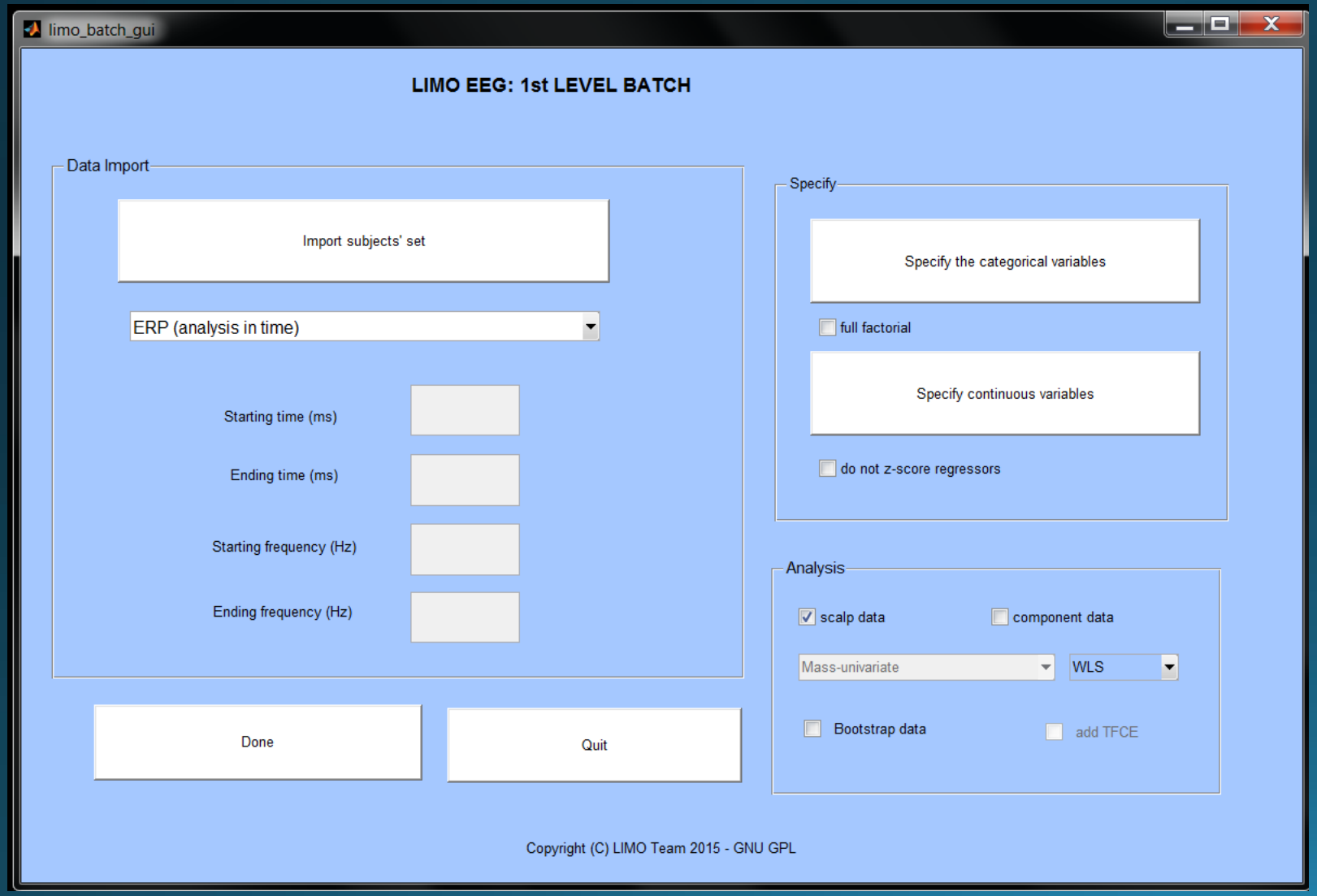
Creating a study and LIMO stats





Currently little bug –
'timelim' or 'freqlim' don't actually trim the data (will be fixed obviously)

LIMO stats without study



Needs a single .set
Lists of text files
- Sets
- Categorical variables (conditions)
- Continuous varfiables

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:

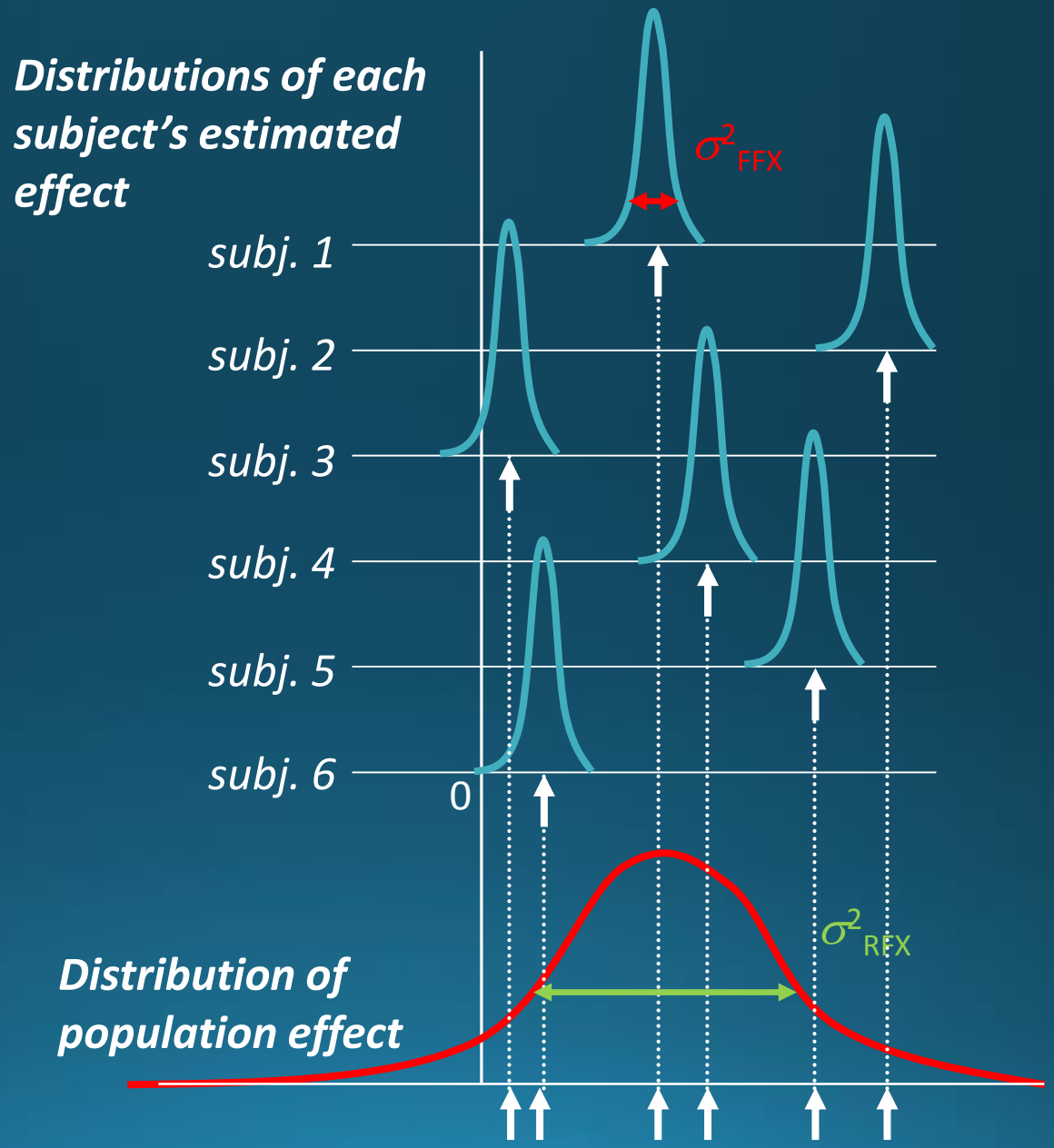
Intra-subjects variation

suggests all these subjects different from zero

Random effects:

Inter-subjects variation

suggests population not different from zero



Hierarchical model = 2-stage LM

Single subject

Each subject's EEG trials are modelled
Single subject parameter estimates

1st level

Single subject parameter estimates or combinations taken to 2nd level

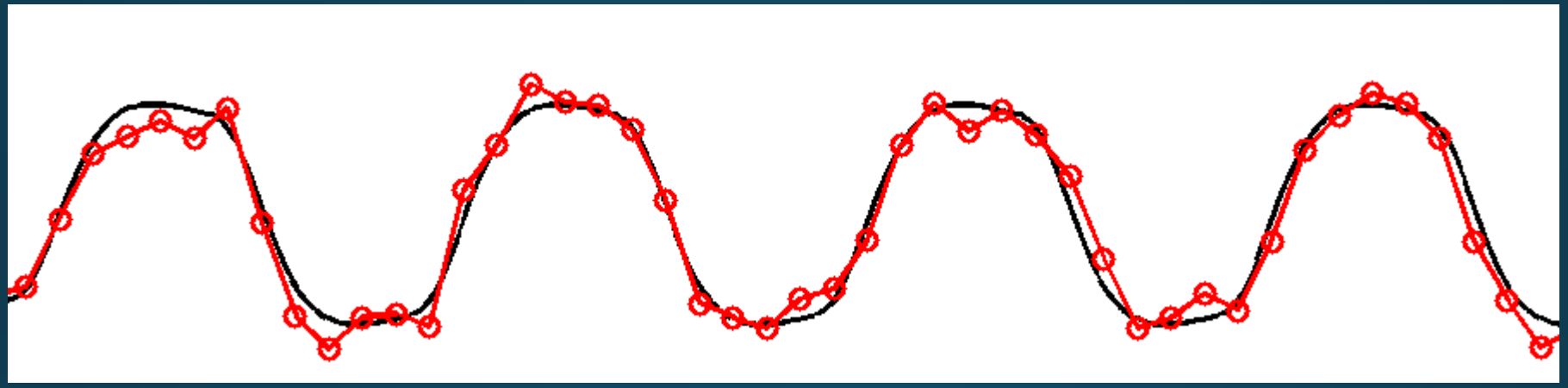
Group/s of subjects

For a given effect, the whole group is modelled
Parameter estimates apply to group effect/s

2nd level

Group level of 2nd level parameter estimates are used to form statistics

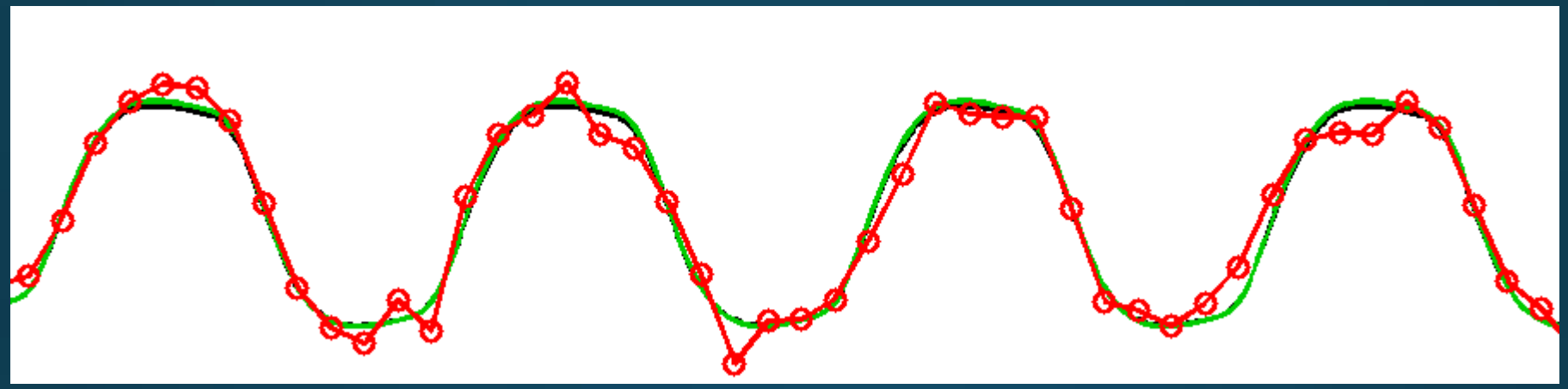
Fixed effects



Only source of variation (over trials) is measurement error

True response magnitude is *fixed*

Random effects



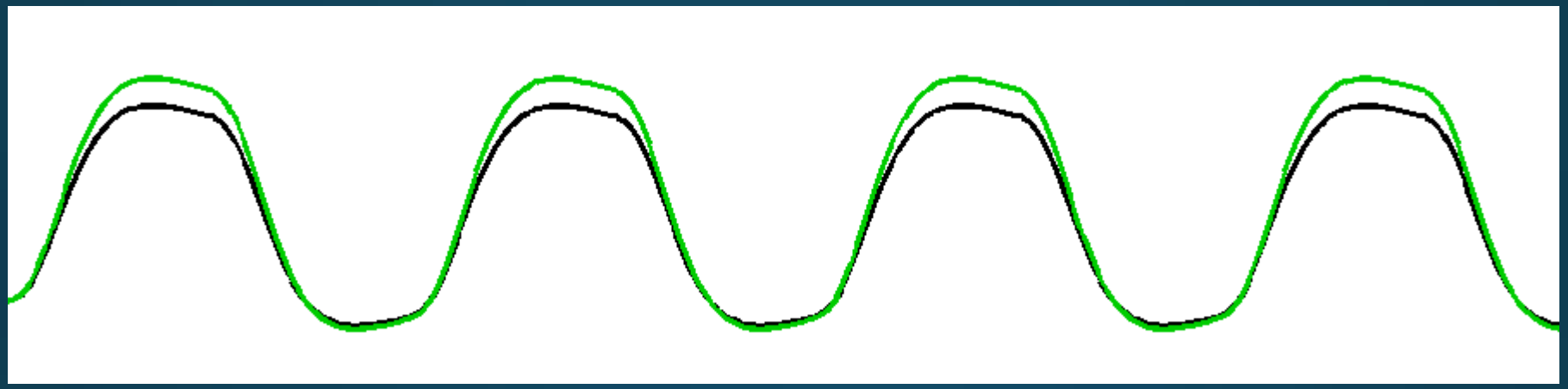
Two sources of variation

- measurement error
- response magnitude (over subjects)

Response magnitude is *random*

- each subject has random magnitude

Random effects



Two sources of variation

- measurement error
- response magnitude (over subjects)

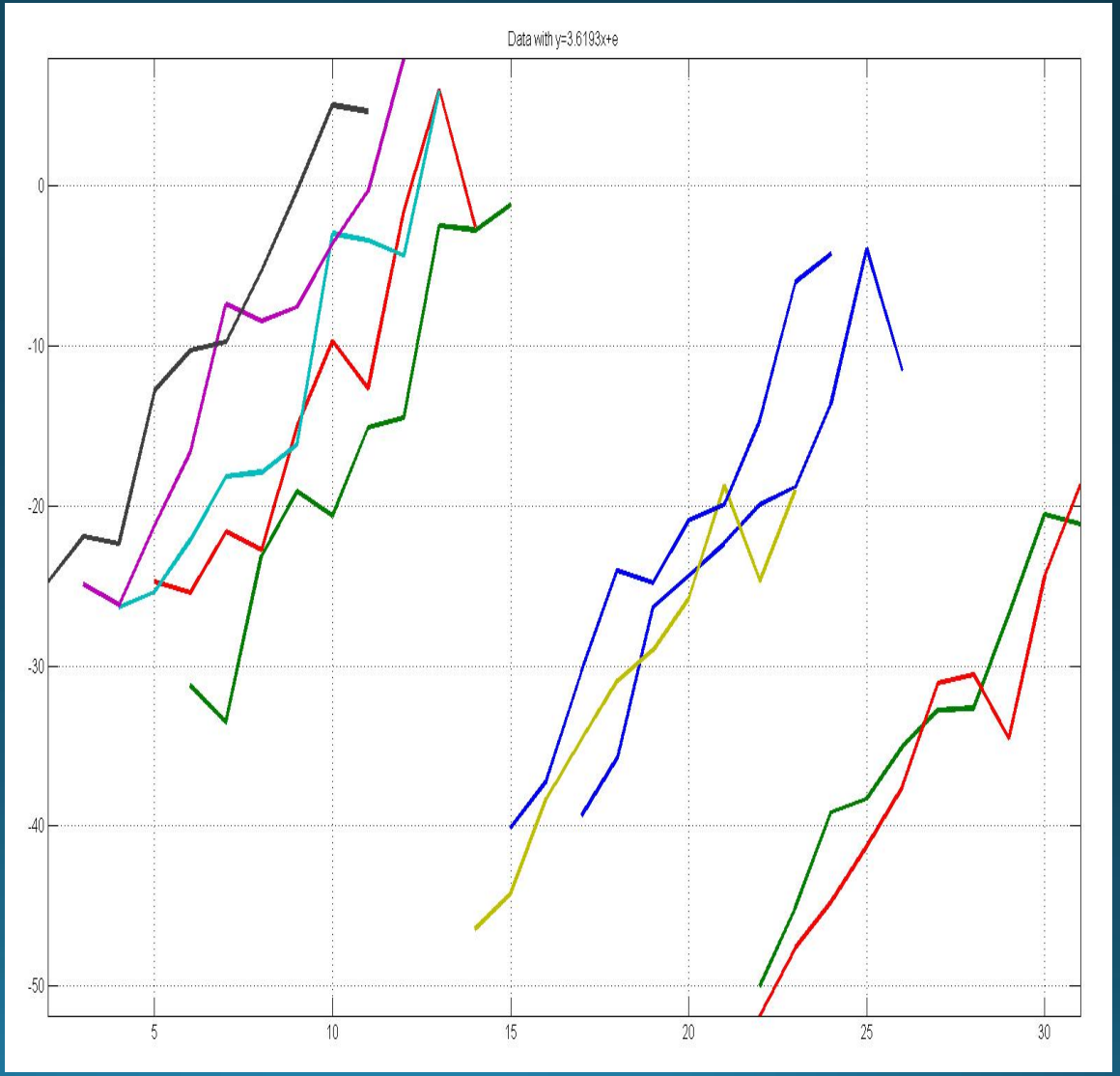
Response magnitude is *random*

- each subject has random magnitude
- but note, population mean magnitude is *fixed*

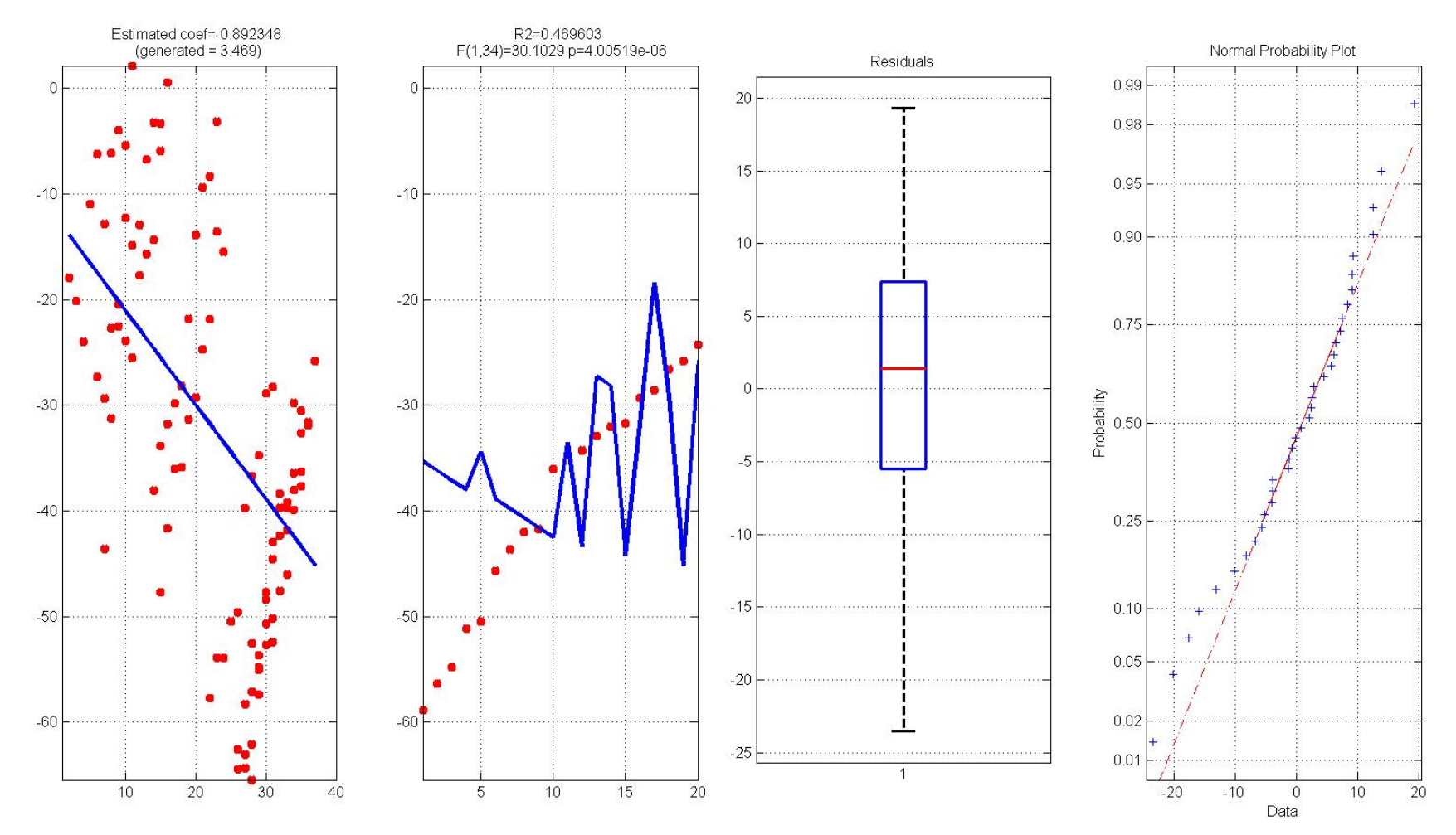
An extreme example

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

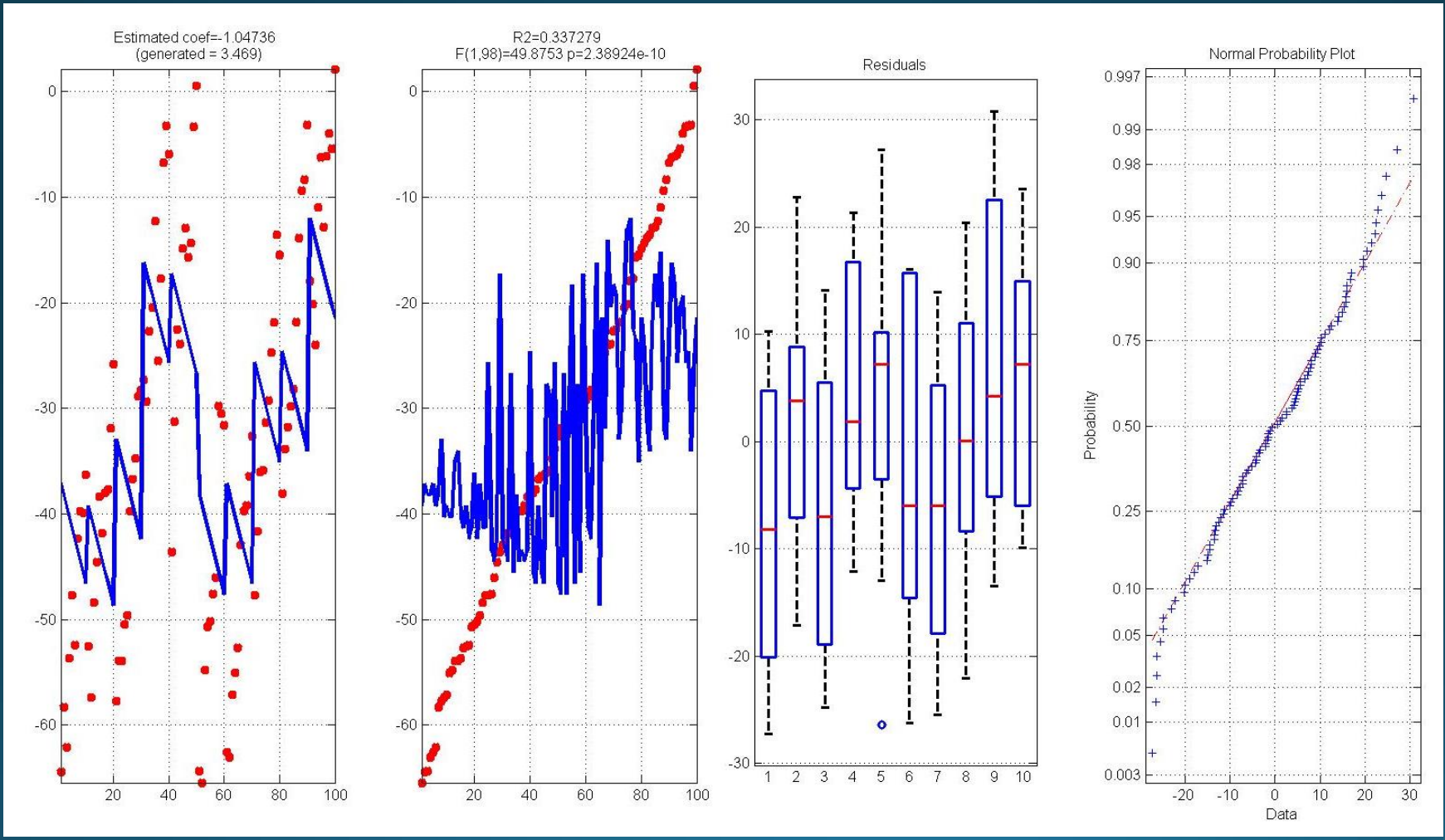


Fixed Effect Model 1: average subjects



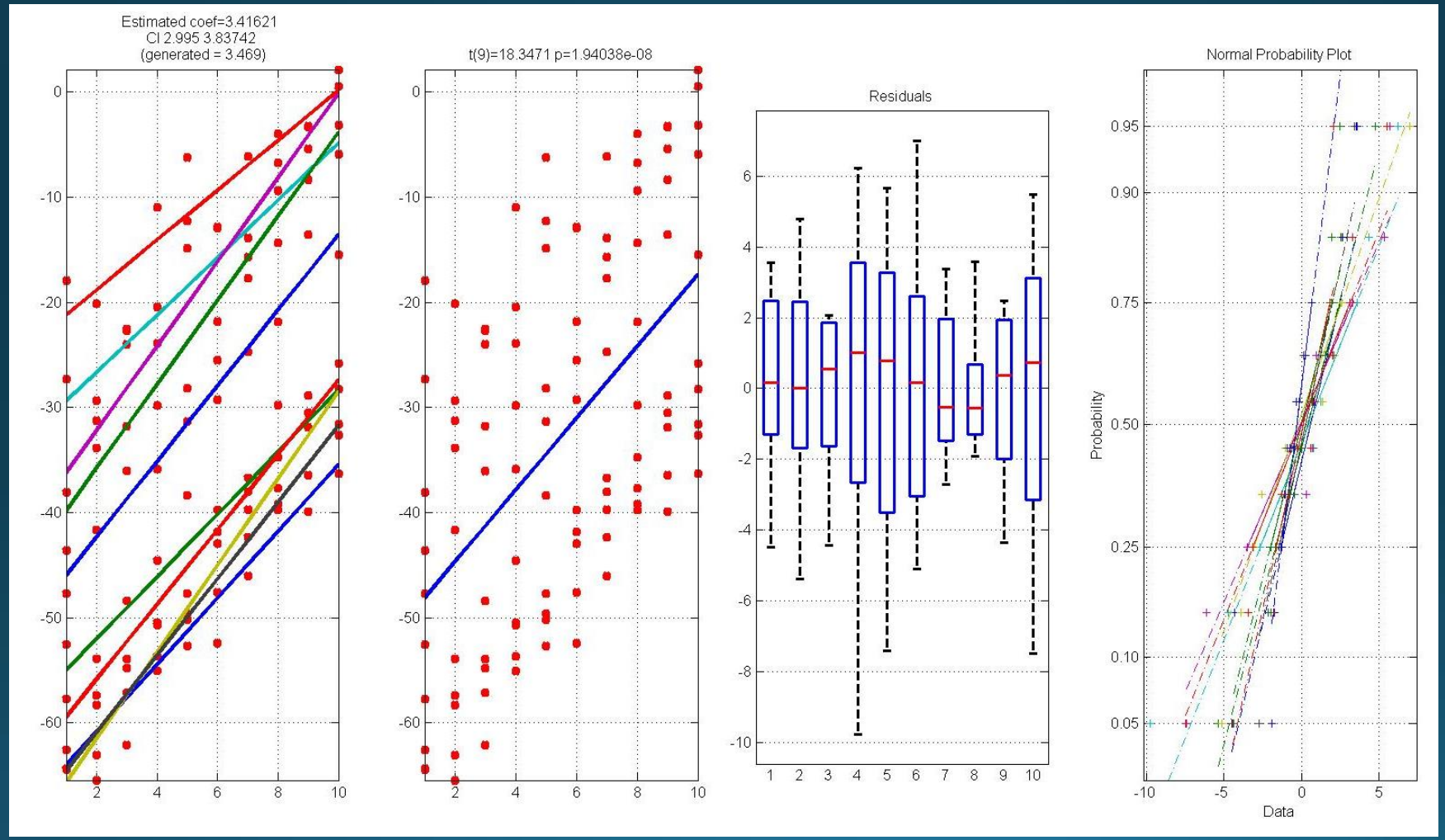
Fixed effect without subject effect → negative effect

Fixed Effect Model 2: constant over subjects



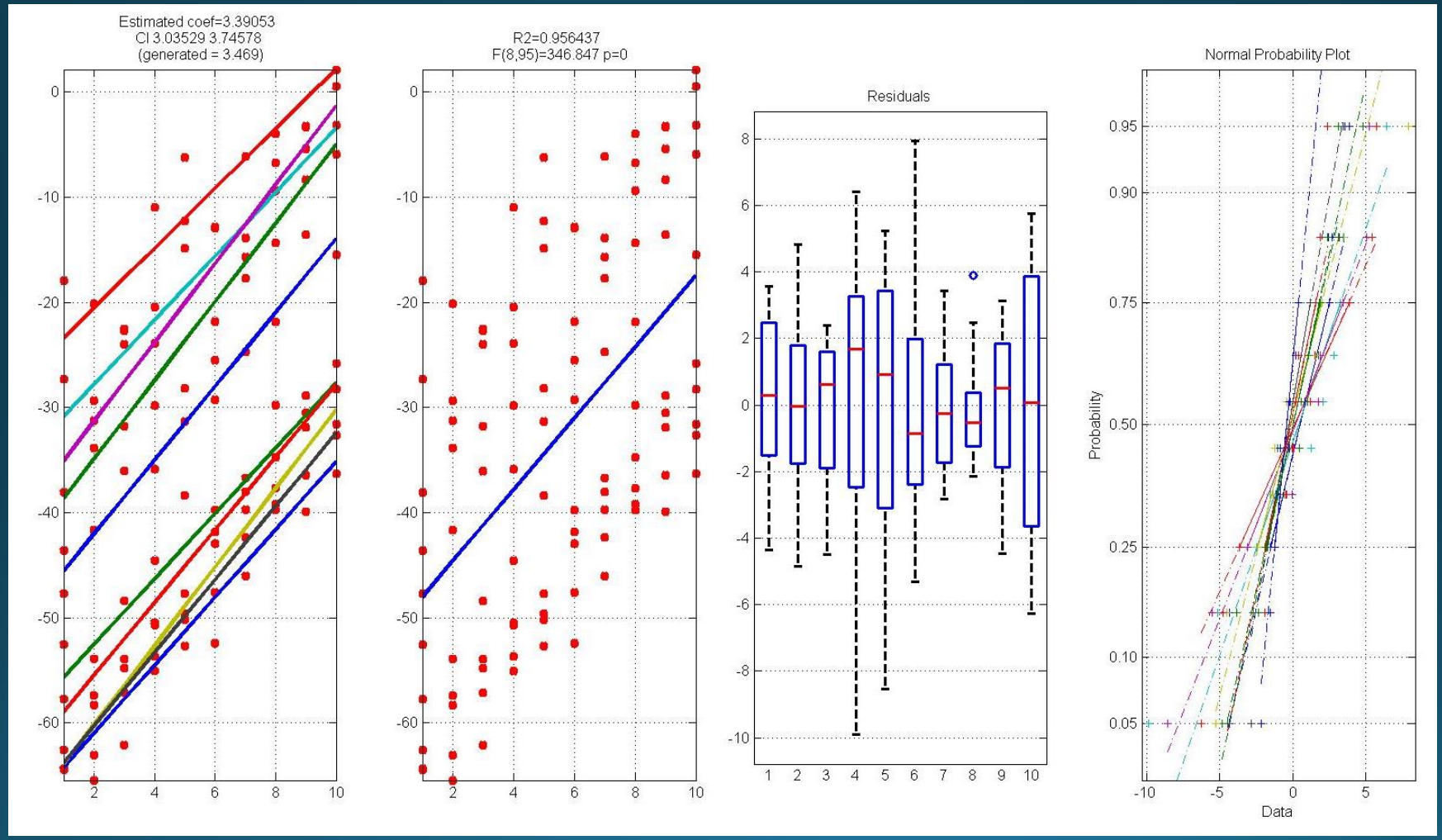
Fixed effect with a constant (fixed) subject effect → positive effect but biased result

HLM: random subject effect



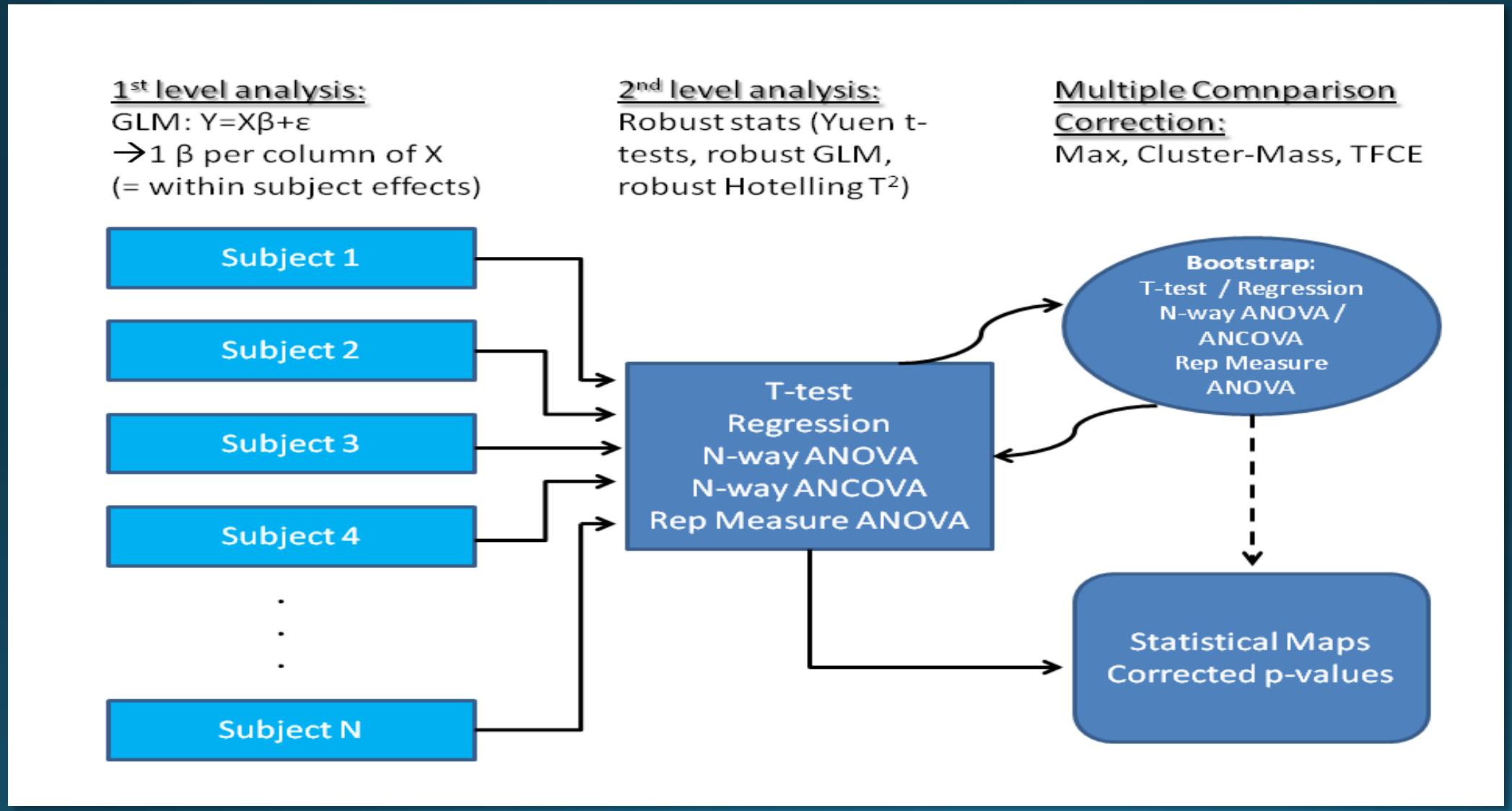
Mixed effect with a random subject effect → positive effect with good estimate of the truth

MLE: random subject effect



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

Hierarchical Linear Model for MEEG



The General Linear Model

What is a linear model?

- An equation or a set of equations that models data and which corresponds geometrically to straight lines, plans, hyperplans and satisfy the properties of additivity and scaling.
- Simple regression: $y = \beta 1x+\beta 2+\varepsilon$
- Multiple regression: $y = \beta 1x1+\beta 2x2+\beta 3+\varepsilon$
- One way ANOVA: $y = u+\alpha i+\varepsilon$
- Repeated measure ANOVA: $y=u+\alpha i+\varepsilon$
- …

What is the GLM?

- 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)

GLM examples

- EEG amplitude is modulated by the stimulus intensity:

$Y_1 = X_1 * B_1 + B_2 + e$

$Y_2 = X_2 * B_1 + B_2 + e$

$Y_3 = X_3 * B_1 + B_2 + e$

$Y_4 = X_4 * B_1 + B_2 + e$

.

.

.

$Y_n = X_n * B_1 + B_2 + e$

→

$$\begin{bmatrix} Y1 \\ Y2 \\ Y3 \\ Y4 \\ . \\ . \\ . \\ Yn \end{bmatrix} = \begin{bmatrix} X1 & 1 \\ X2 & 1 \\ X3 & 1 \\ X4 & 1 \\ . & 1 \\ . & 1 \\ . & 1 \\ Xn & 1 \end{bmatrix} * \begin{bmatrix} B1 \\ B2 \end{bmatrix} + \begin{bmatrix} e1 \\ e2 \\ e3 \\ e4 \\ . \\ . \\ . \\ en \end{bmatrix}$$

→ $Y = XB + e$

GLM examples

- EEG amplitude is modulated by the stimulus conditions (A vs B):

$Y1 = 1*B1+0*B2+B3+e$

$Y2 = 1*B1+0*B2+B3+e$

$Y3 = 0*B1+1*B2+B3+e$

$Y4 = 0*B1+1*B2+B3+e$

.

.

.

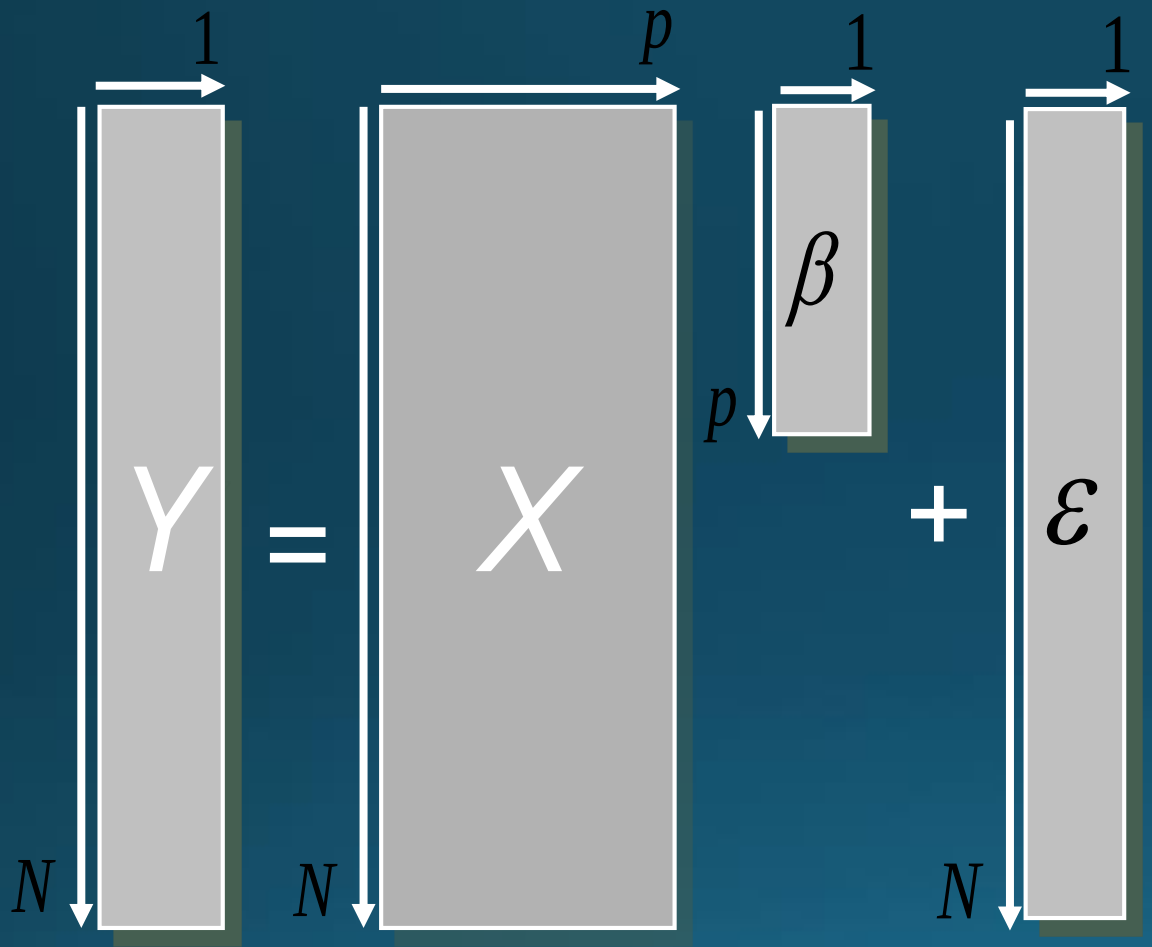
$Yn = X1n*B1+X2n*B2+B3+e$

→

$$\begin{bmatrix} Y1 \\ Y2 \\ Y3 \\ Y4 \\ . \\ . \\ . \\ Yn \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ . & . & 1 \\ . & . & 1 \\ . & . & 1 \\ X1n & X2n & 1 \end{bmatrix} * \begin{bmatrix} B1 \\ B2 \\ B3 \end{bmatrix} + \begin{bmatrix} e1 \\ e2 \\ e3 \\ e4 \\ . \\ . \\ . \\ en \end{bmatrix}$$

→ $Y = XB+e$

General Linear model



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

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

Estimate with Ordinary or Weighted Least Squares

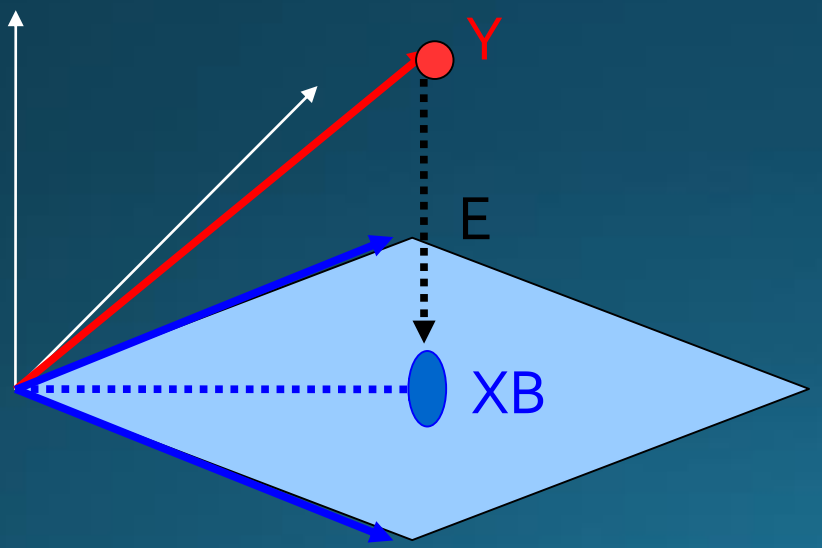
N: number of trials
p: number of regressors

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

Linear Algebra and Statistics

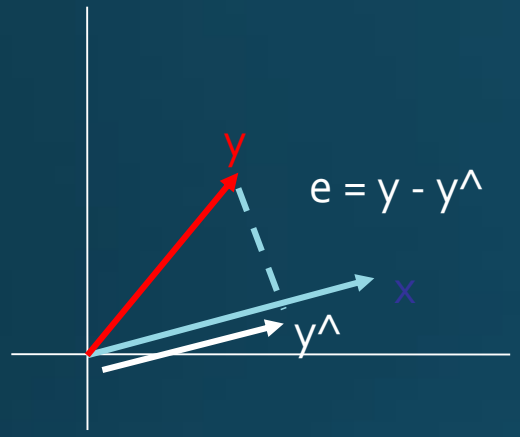
Y = 3 observations X = 2 regressors

Y = XB+E --> Y^=XB



SS total = variance in Y
SS effect = variance in XB
SS error = variance in E
R2 = SS effect / SS total
F = SS effect/df / SS error/dfe

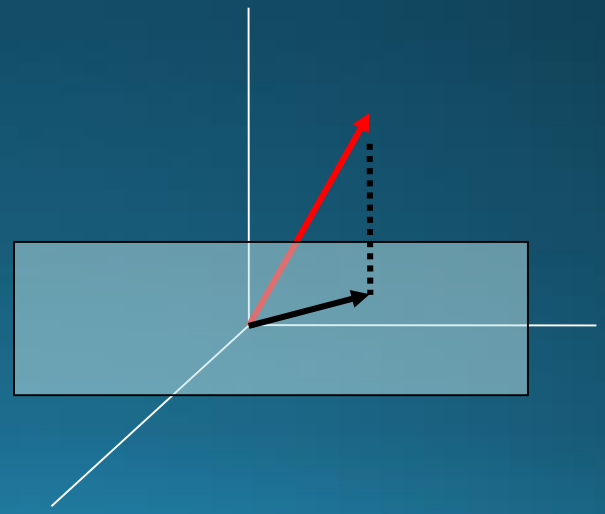
Linear Algebras: Projections



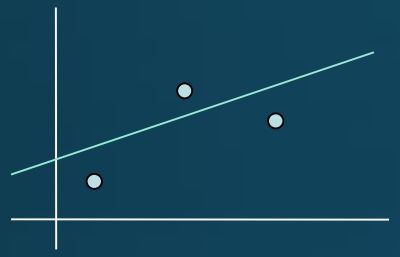
$\hat{y} = \beta x$
$x'(y - \beta x) = 0$
$\beta x'x = x'y$
$\beta = x'y / x'x$

$\hat{y} = (x'y / x'x)x$
$\hat{y} = Py \rightarrow P = xx' / x'x$

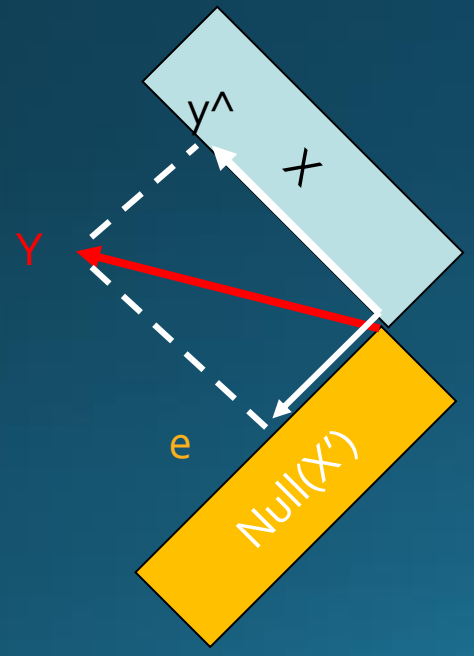
Why project? XB = Y may have no solution, the closest solution is a vector located in X space that is the closest to Y. In N dimensions:
P = X inv(X'X) X'
B = inv(X'X) X' Y



Projection and Least squares



$y = \beta x + c$
P projects the points on the line
Minimizing the distance (^2)
is projecting at perpendicular angles



Y = y^+e
y^ = PY
e = (I-P)Y

An 'effect' is defined by which part of X to test (i.e. project on a subspace)

```
Ro = I - (Xo*pinv(Xo));
P = Ro - R;
Effect  = (B'*X'*P*X*B);
```

Weighted Least Squares

The LIMO EEG approach: a single weight per trial

Mathematical issues

- Least Squares requires the error covariance to have o off diagonal ie Cov(e) = $\sigma^2 I$
- Deviations from that assumption can lead to substantial power reduction and increase in false positive rate
- Weighted Least Squares is the solution to these problems allowing Cov(e) = $\sigma^2 V$, with V a diagonal matrix

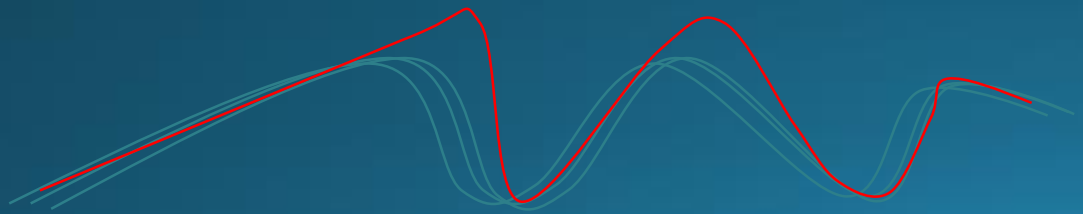
$$y = X\beta + e, \quad E(e) = 0, \quad Cov(e) = \sigma^2 V$$

$$Wy = WX\beta + We, \quad E(e) = 0, \quad Cov(e) = \sigma^2 I$$

$$\hat{\beta} = (X^T WX)^{-1} X^T Wy$$

How to apply weights?

- Weight reflect outlying data – does it make sense to be for some time frames and/or frequency bin?
- Noise + signal model → unlikely to have a background neural synch higher than signal
- Noise + signal model → unlikely to a single frame outlier, this an autoregressive process, many frames must be outlying
- Phase reset model → either a trial is out of phase or it as amplitude difference, in both cases many frames must be outlying



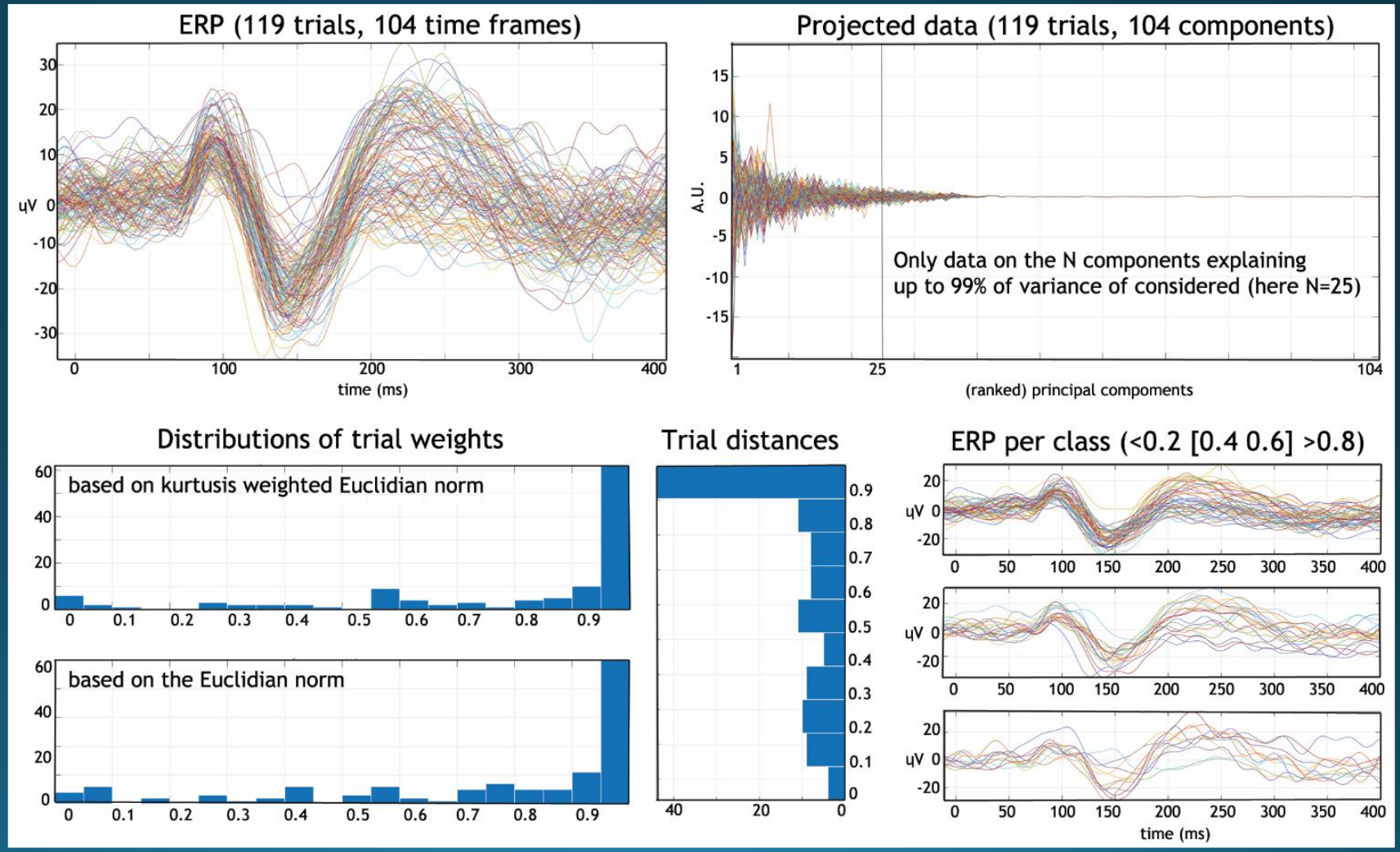
1111111111111.9.8.7.6..81111..8..7.8.9111111111

A trial can be both good and bad ?
What about information accumulation?

Weighted least squares in LIMO EEG

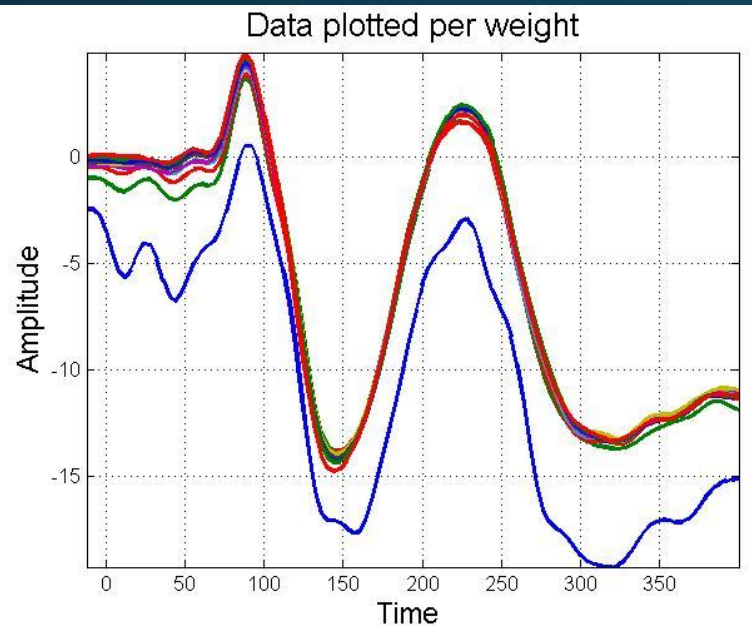
- Principal Component Projection method:

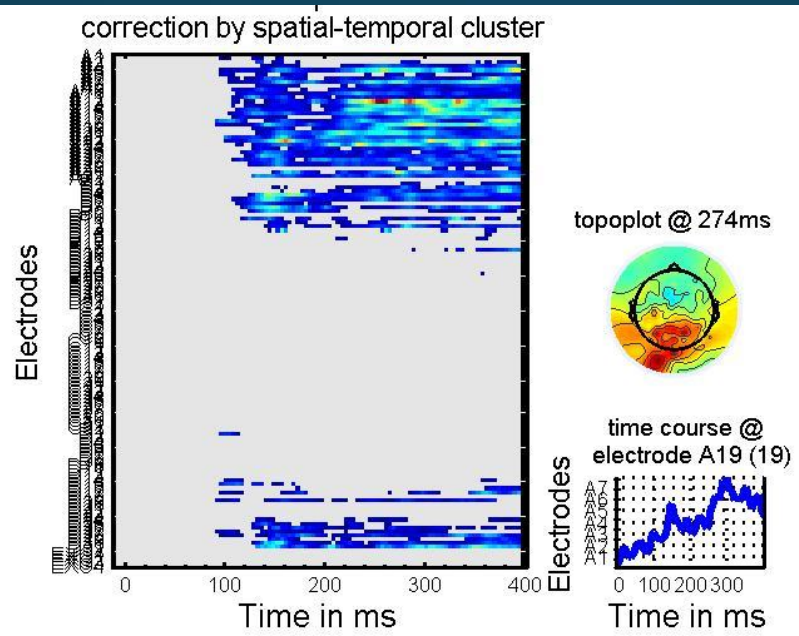
- PCA
- outlier detection on projected data points (Filzmoser et al., 2008)
- 1 weight per trial

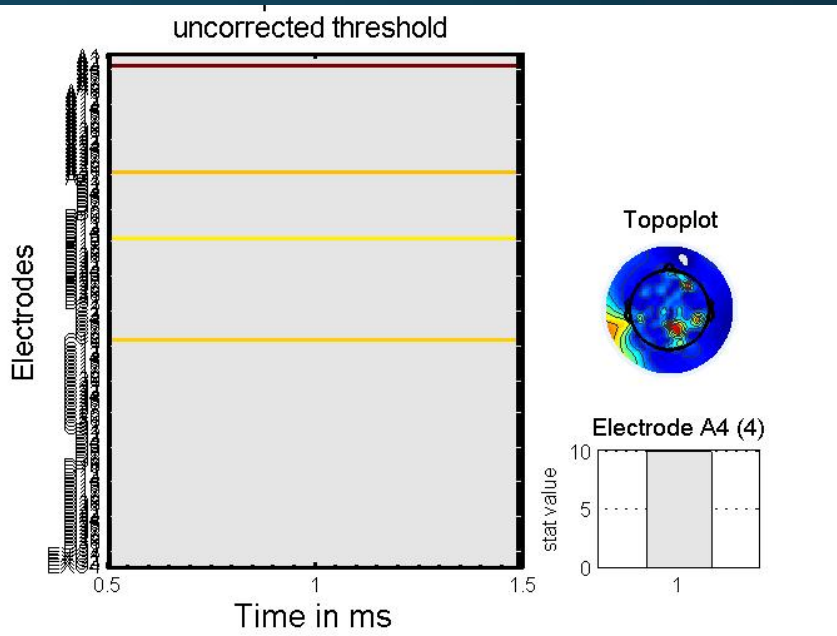


Weighted least squares in LIMO EEG

- LIMO EEG data set → limo_CheckWeight.m







Subject 2 electrode 41

good trials vs bad trials

Bias between conditions

Review results

What has been computed

- Input files: .set listing to single trial files generated using eeglab
- Yr the data for the design considered
- Yhat the modelled data
- Res thr residuals
- Condition and covariates effects
- Betas files
- Con files

→ only use betas and con for 2nd level

Now 2nd level

- Think of your betas are equivalent of your mean
- Cleaner because only show condition effect – accounting of the variance between trials
- Now we use eg paired t-test ignore vs memorize

→ Create of working directory paired_ttest

→Call limo_eeg

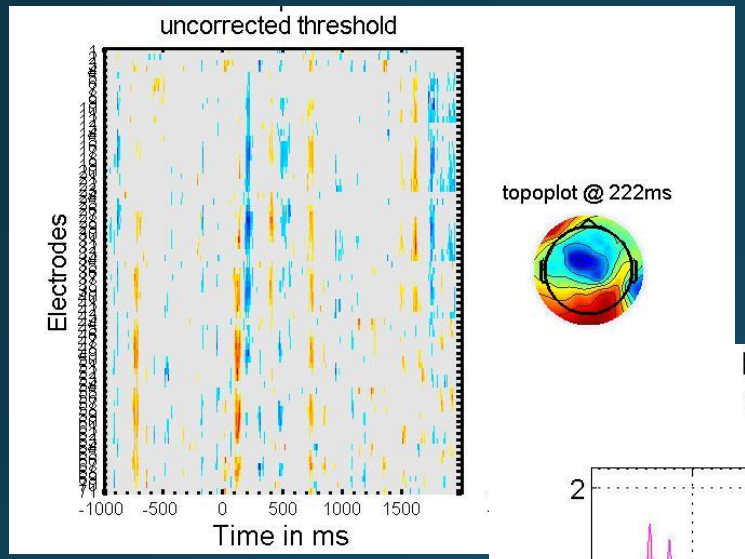
→Select random effect

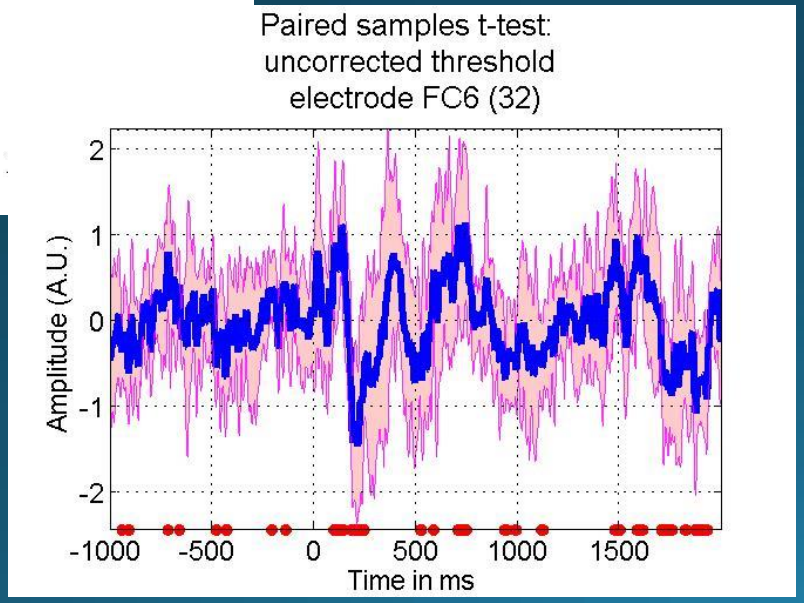
→Input neighbouring matrix

→do the paired t test (set bootstrap to 0): select beta list and select [1 2]

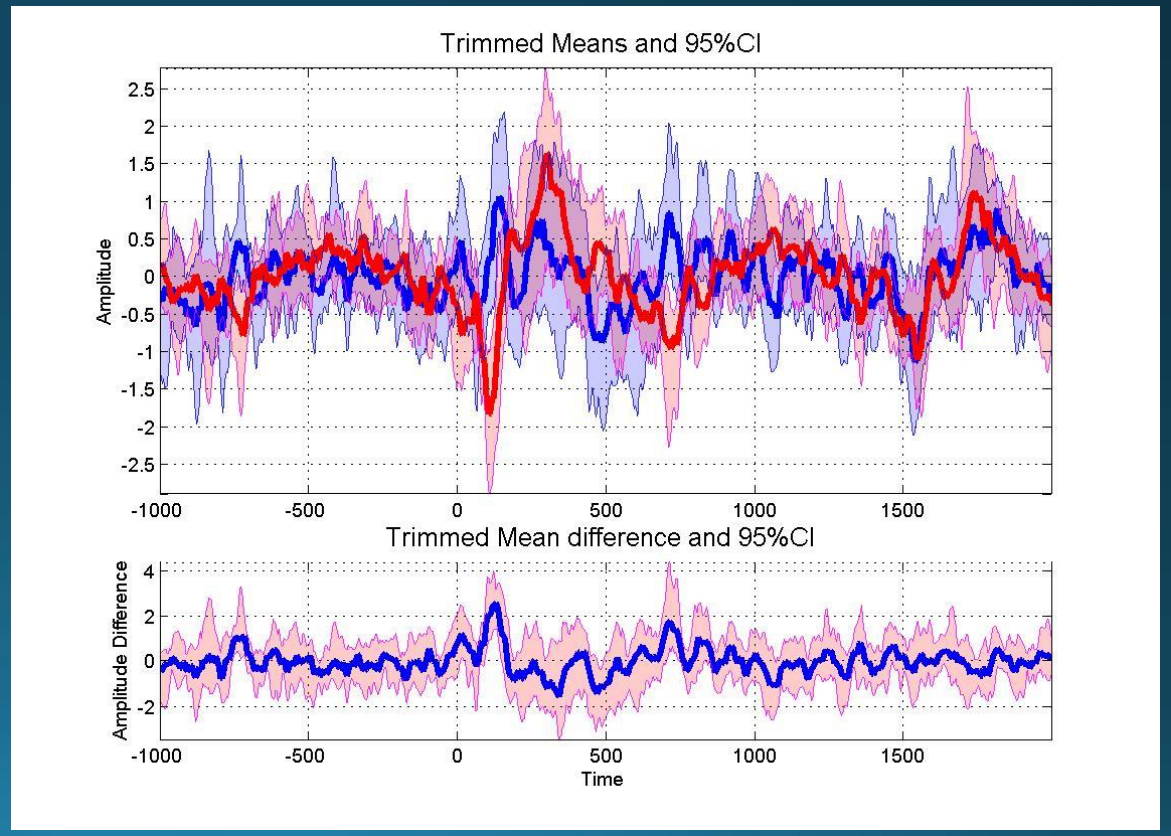
Now 2nd level

- You can also check what is going on raw data ! using summary stats





'Make and Plot a difference'



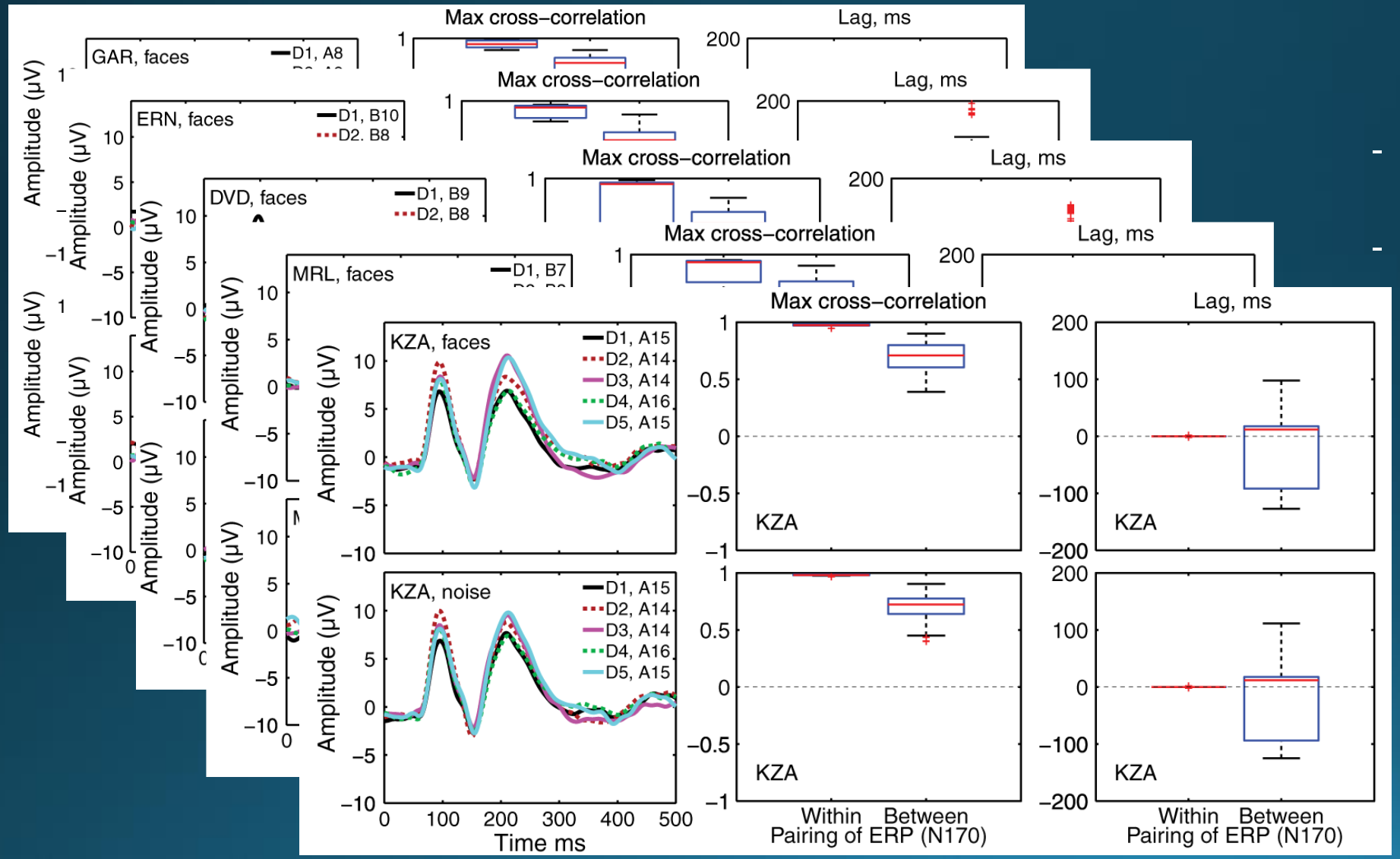
DESIGNS

So what now that we have a HLM?

EEG signals are idiosyncratic

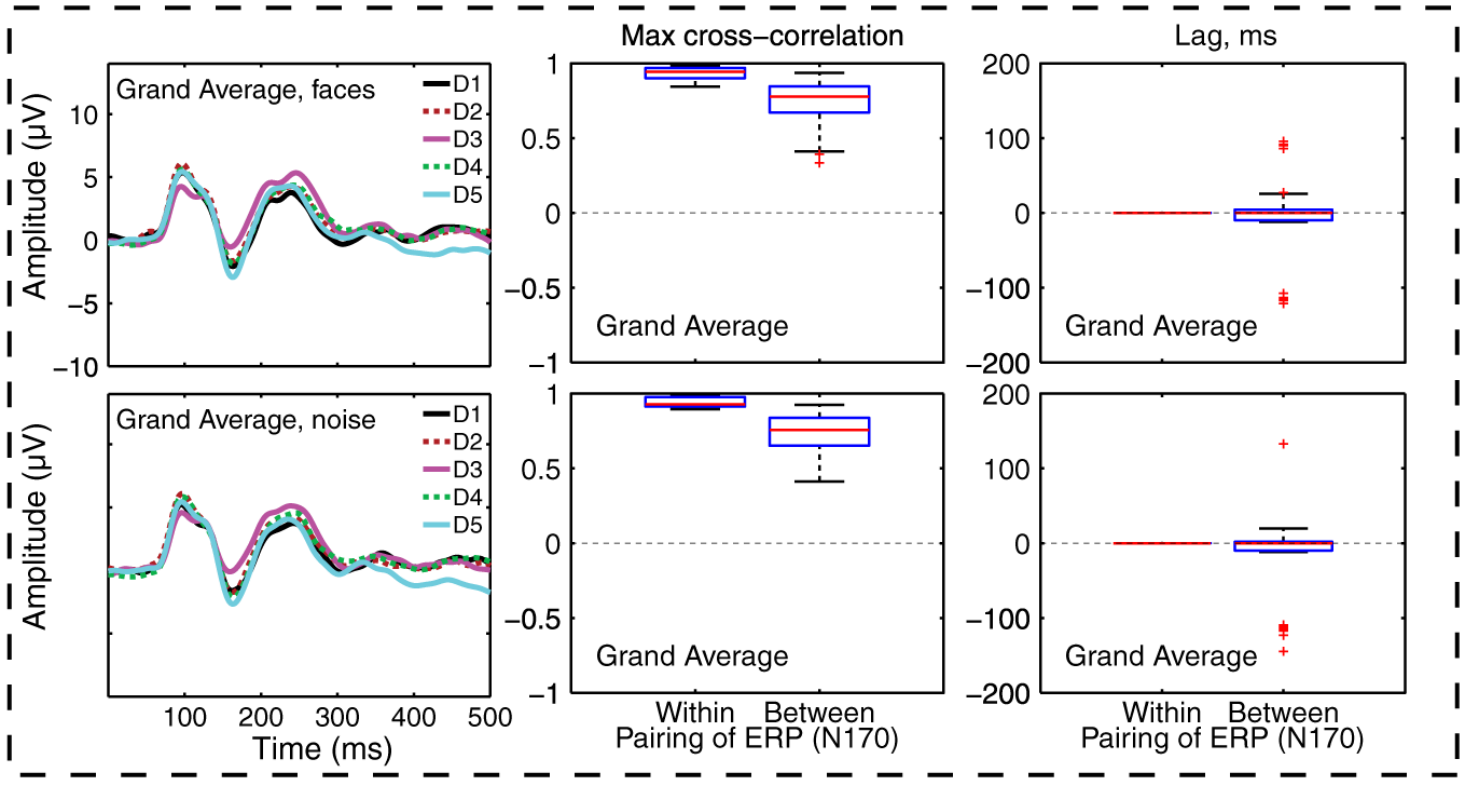
Gaspar et al. 2011 Reliability of ERP and single-trial analyses NeuroImage 58

Test-retest of ERPs



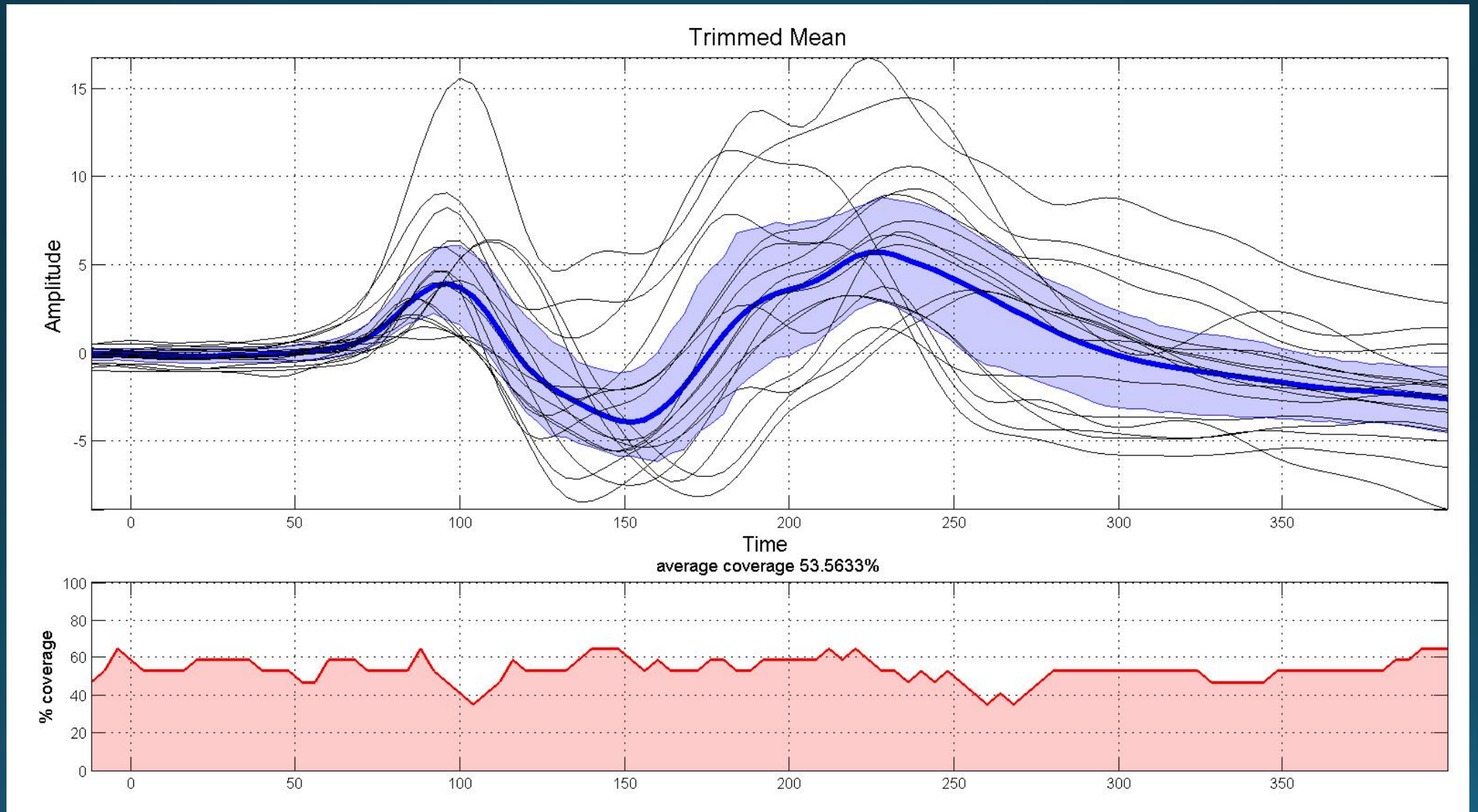
- ERPs are highly reliable within subjects
- xcorr >0.90 with ~4/6 ms lag

Grand averages do not reflect ERP dynamics



- Because ERPs are highly reliable within subjects, grand averages are also highly reliable.
- However, this 'within-subject' reliability also means that grand averages ERPs are significantly different from individual subjects' ERPs.
- Plots of grand average can be misleading

Grand averages do not reflect ERP dynamics

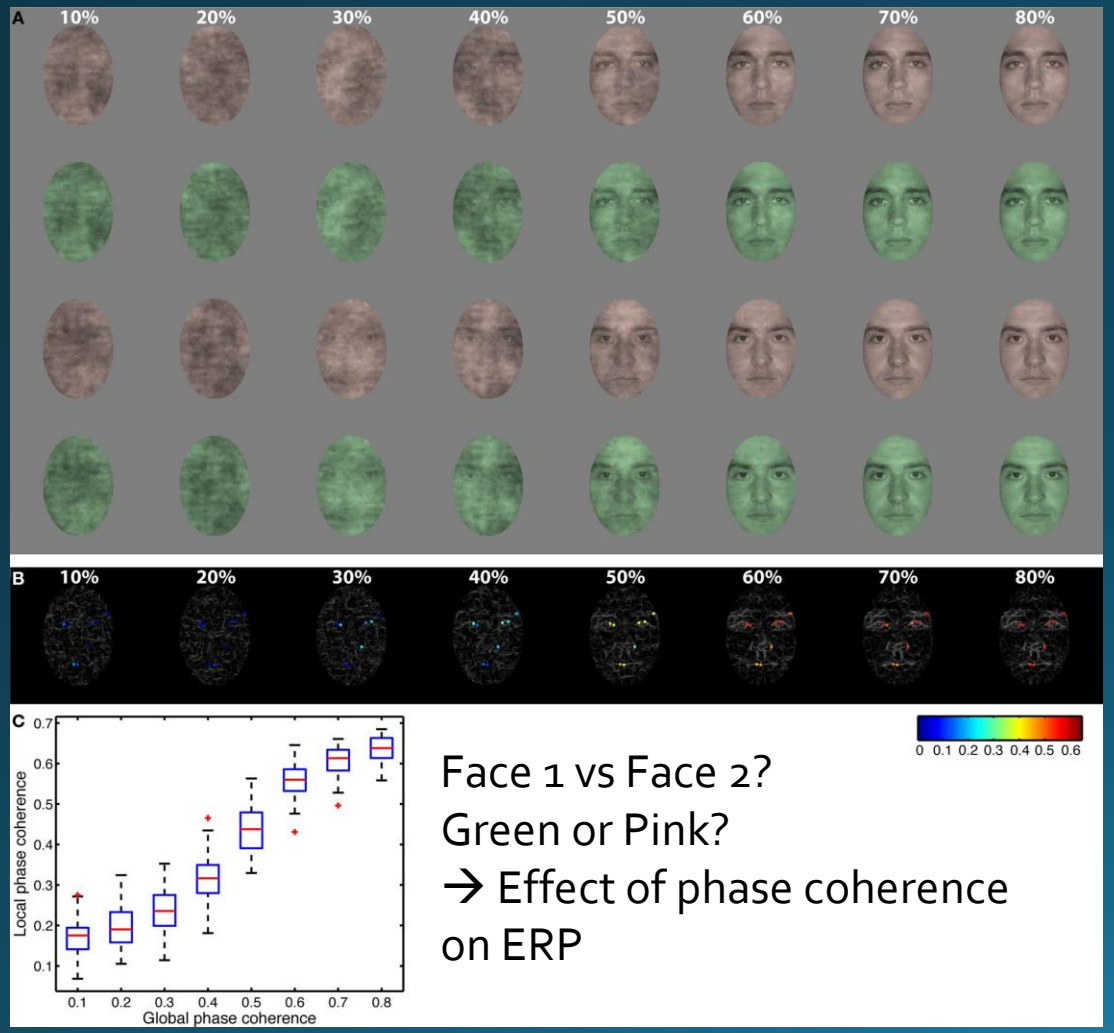


Single subjects or group analysis

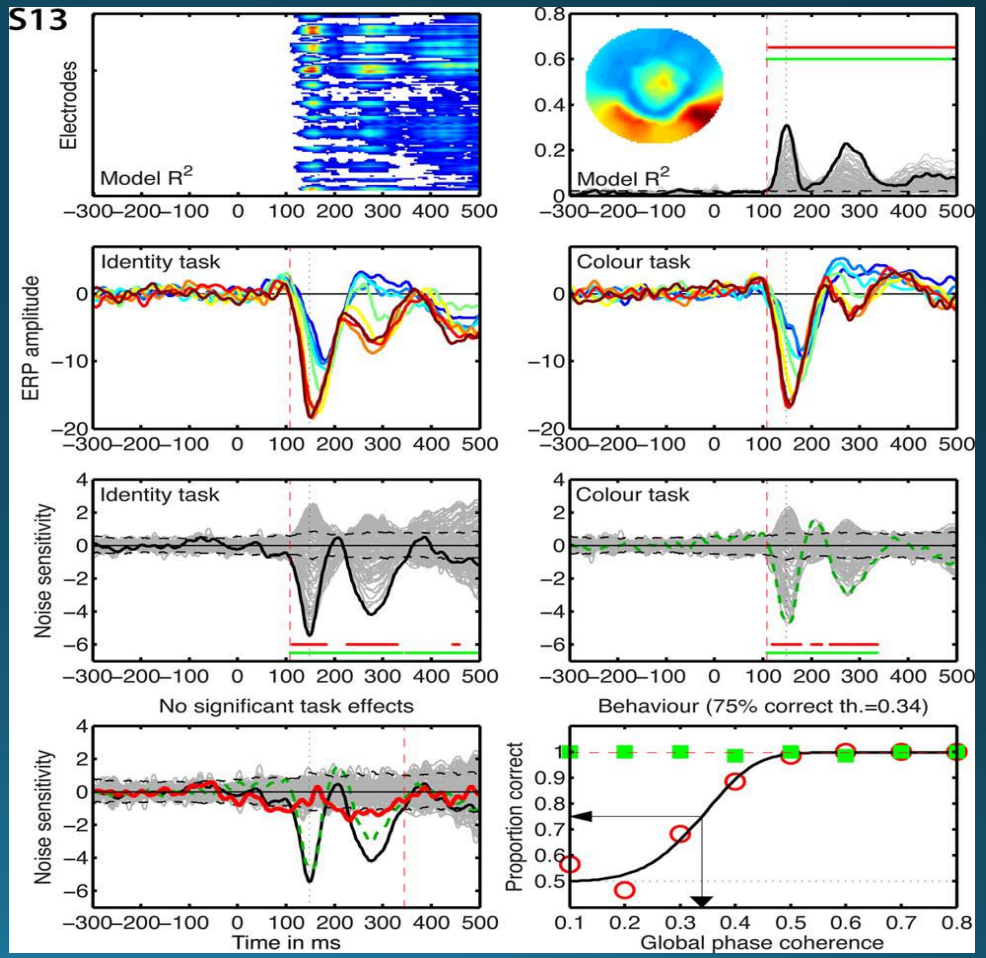
What is the question?

- If the question pertains to dynamic analyses (when things happen) and/or quantitative aspects (how much this variable explains of the data), then single subjects analyses make more sense given the idiosyncratic nature of EEG.
- Yet some group stats are needed for inference – e.g. average cluster onset, average number of subject showing an effect, etc .. + derive group level effect sizes
- If the question is general in nature (is there a measurable difference between these conditions) or pertains to group differences and/or attributes, then group analyses makes sense.

How task constraints modulate the ERP response?

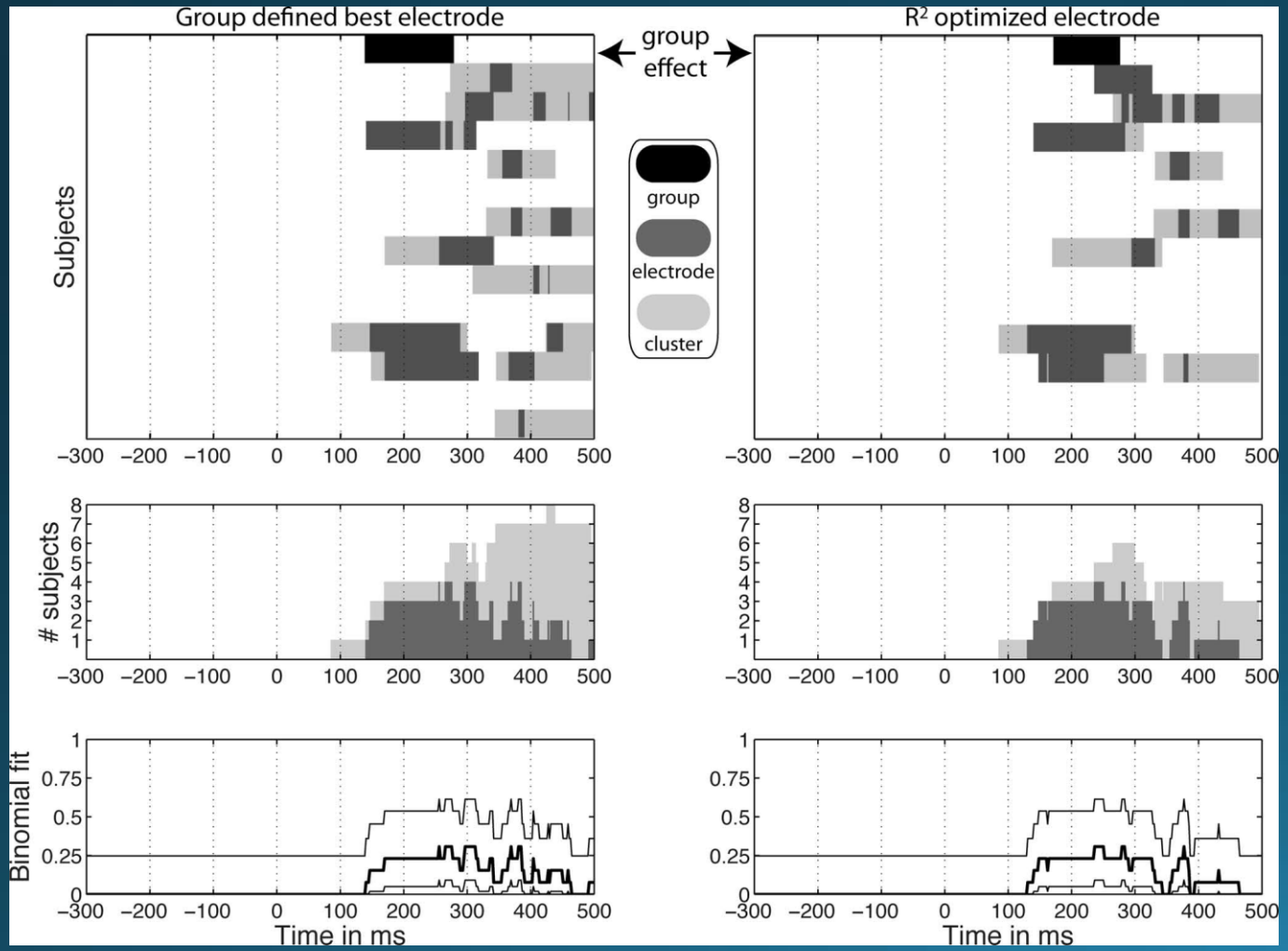


Face 1 vs Face 2?
Green or Pink?
→ Effect of phase coherence on ERP



Rousselet et al. Front Psy 2011

How task constraints modulate the ERP response?

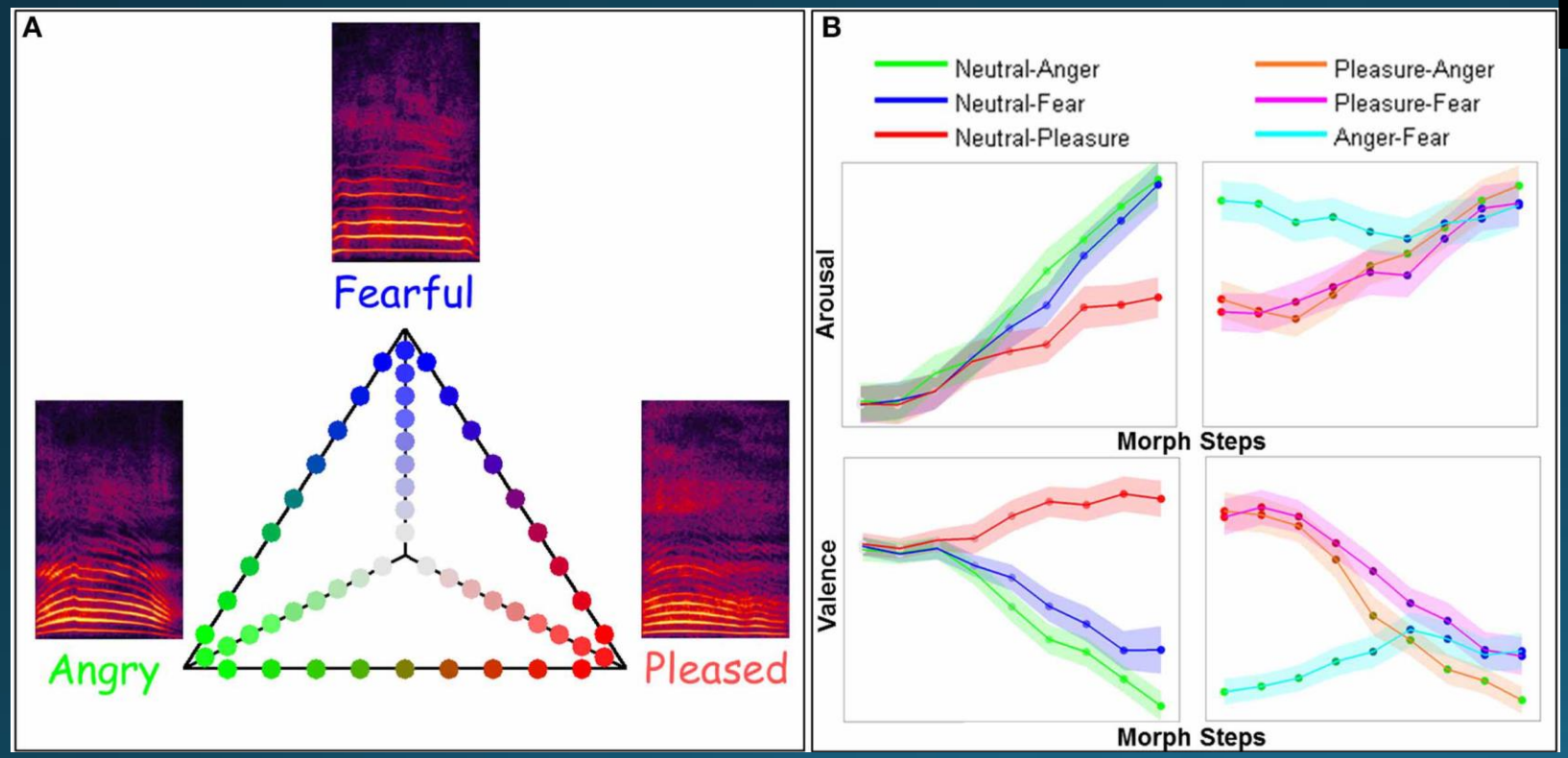


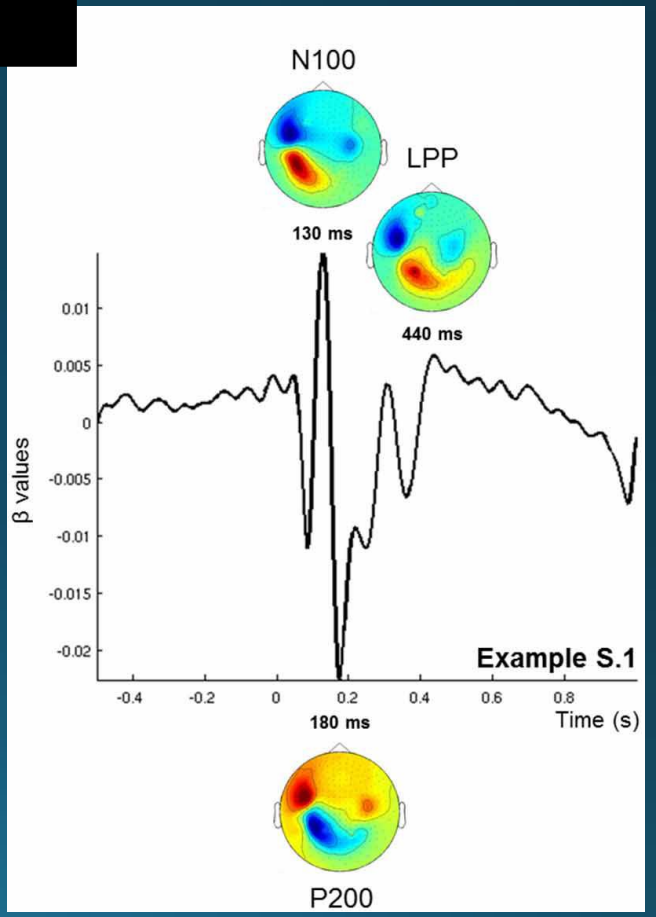
At the group level, ERP sensitivity to phase noise was reduced between about 140 and 300 ms when stimulus phase information was task irrelevant.

we observed a significant task effect in only 60% of subjects, and at any time point only 31% of subjects showed results consistent with group analyses

Rousselet et al. Front Psy 2011

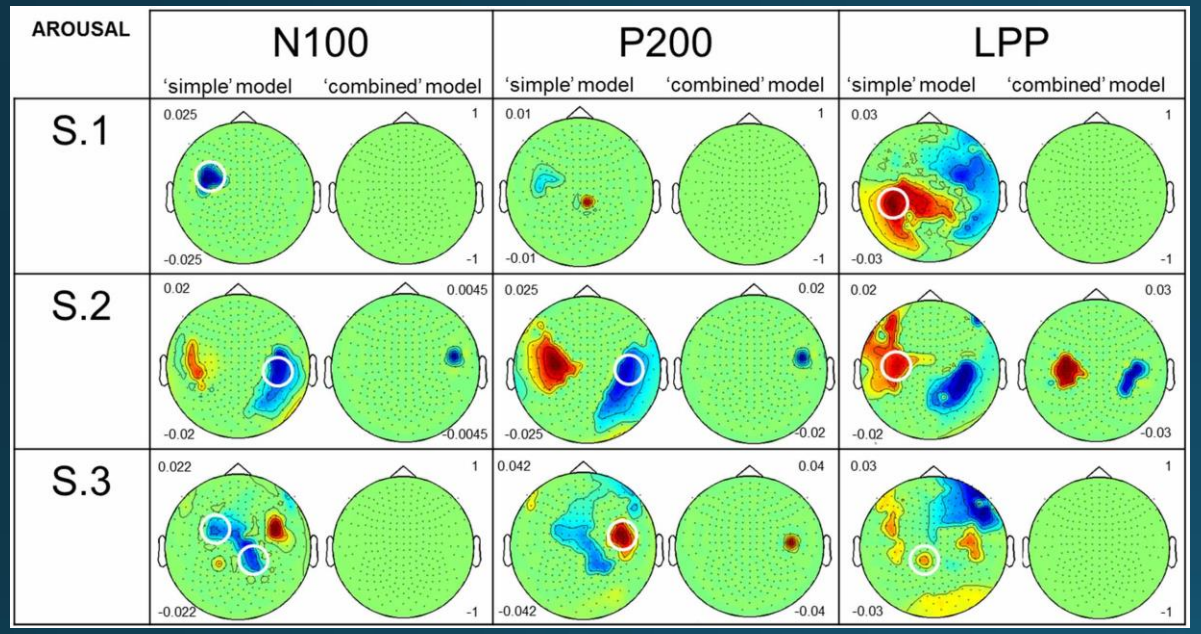
MEG of acoustic properties in affective vocalizations





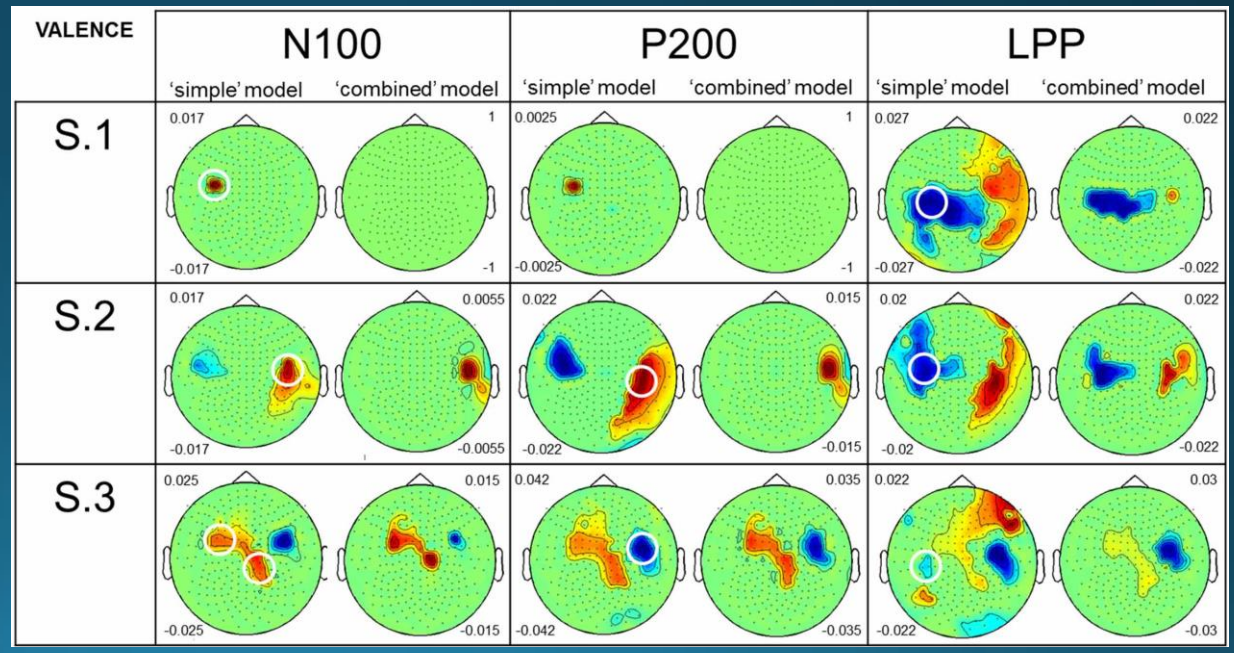
Salvia et al. Front Neurosc 2014

MEG of acoustic properties in affective vocalizations



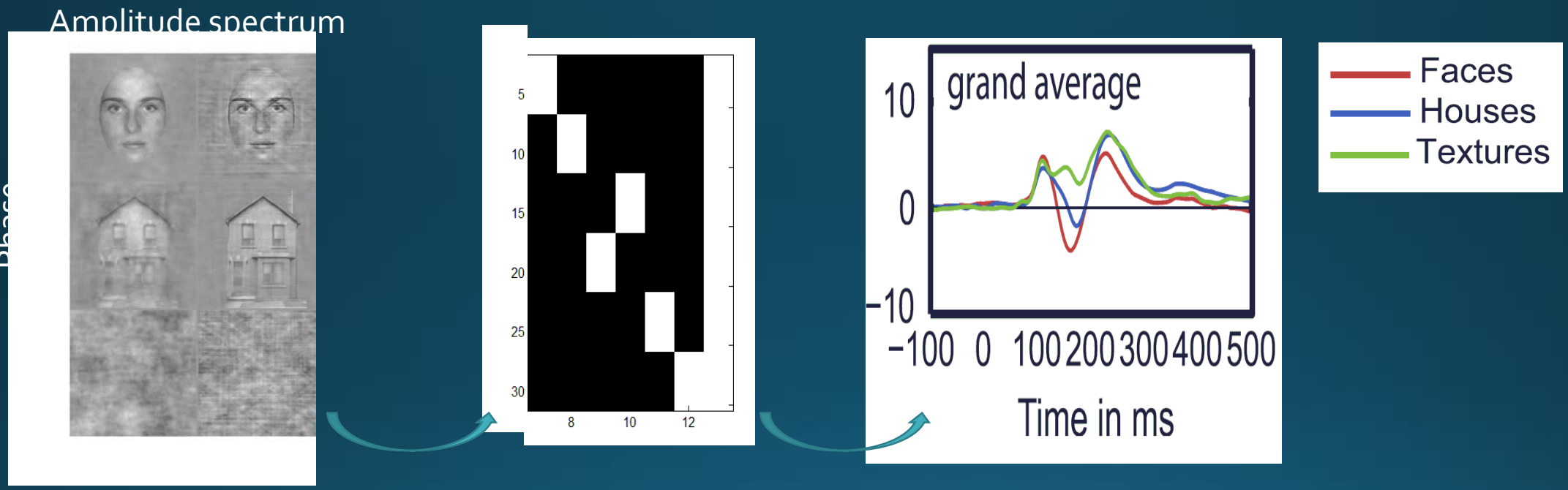
Simple model: for each sound, input the arousal and valence value - *Combined model*: valence, arousal, and 2 components of a PCA (72% var) from six acoustic parameters: mean/ SD of fo, HNR and percentages of unvoiced frame, jitter and shimmer.

Early effects are largely driven by acoustical variations
Once the variance explained by acoustic properties is accounted for, the remaining effects of emotionalv variables (especially valence) are mostly observed at late stages (~400–600 ms).



Categorical designs

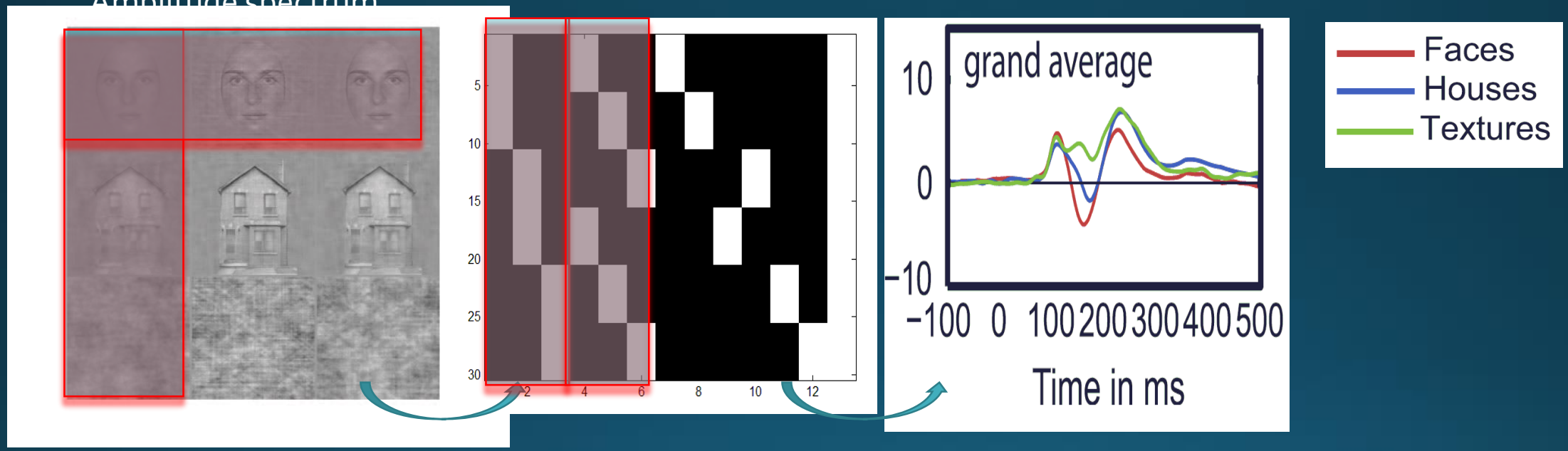
Factorial Designs: 3*3



For group analyses, all you need is an estimate for each condition per subject
Level 1: Y = XB1→6, each beta is a mixture of the factors at that stage, but estimate the condition
Level 2: Y = XB1→12, the beta of the 1st level are now split into factors (3*3) and interaction (6)

Bienek, et al (2012). Phase vs Amplitude Spectrum. Journal of Vision 12

Factorial Designs: N*N*N*…

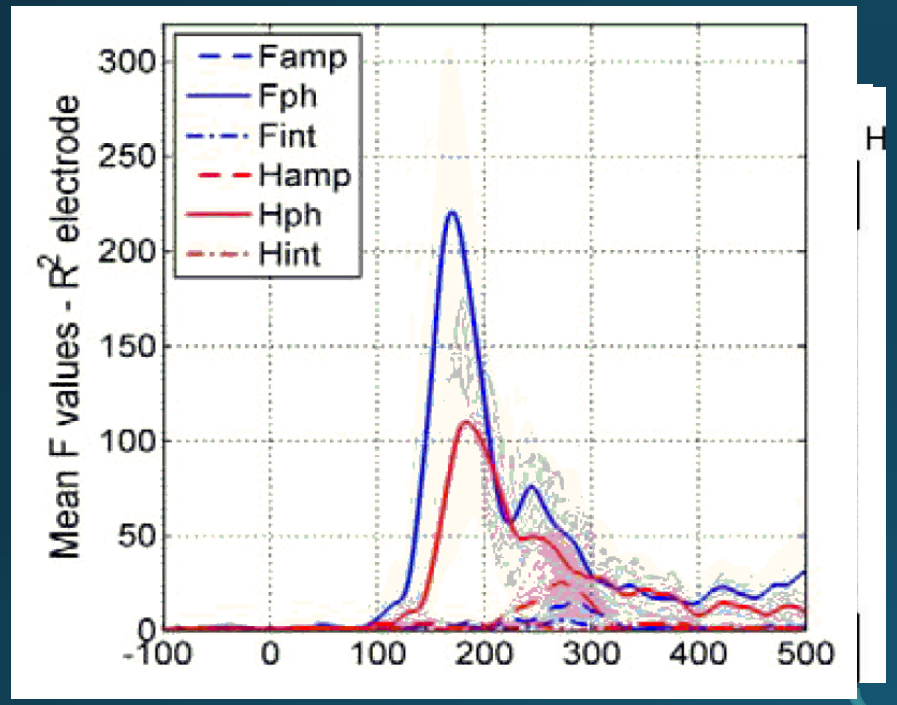


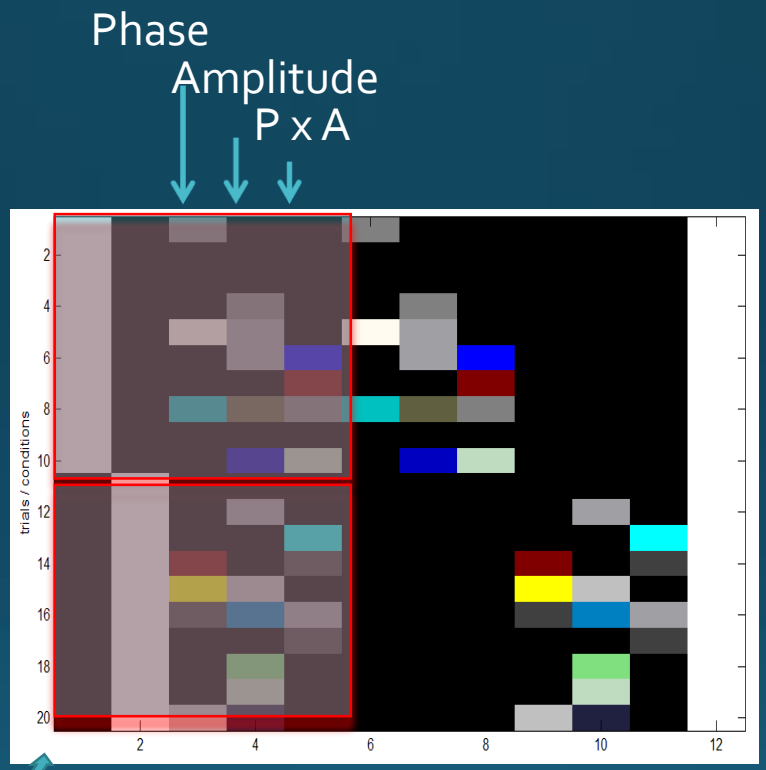
For single subject analyses, you need all effects
Level 1: Y = XB1→12, the data of each subject are split into factors (3*3) and interaction (6)
Level 2: nothing left to explain (stats on attributes)

Bienek, et al (2012). Phase vs Amplitude Spectrum. Journal of Vision 12

Continuous designs

Regression based designs – 1st level

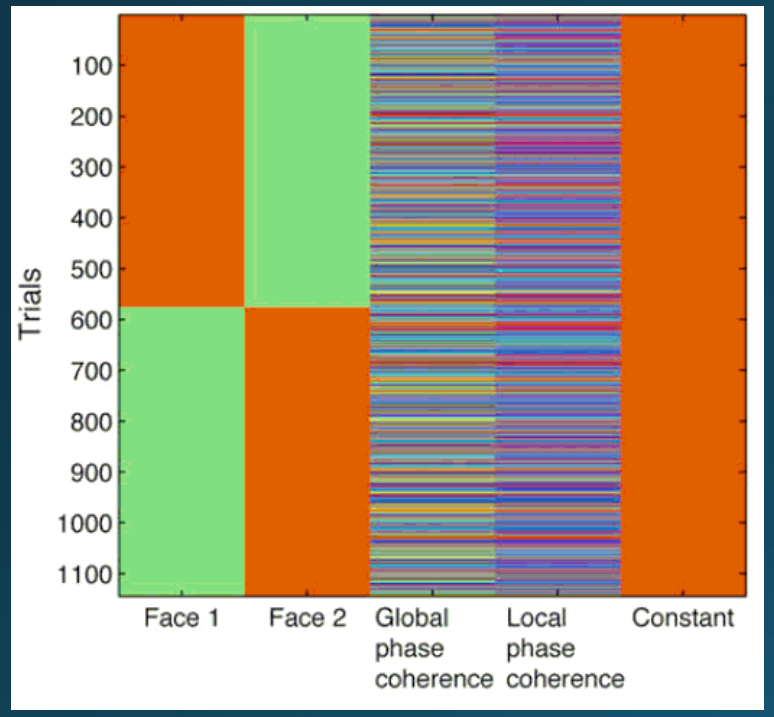


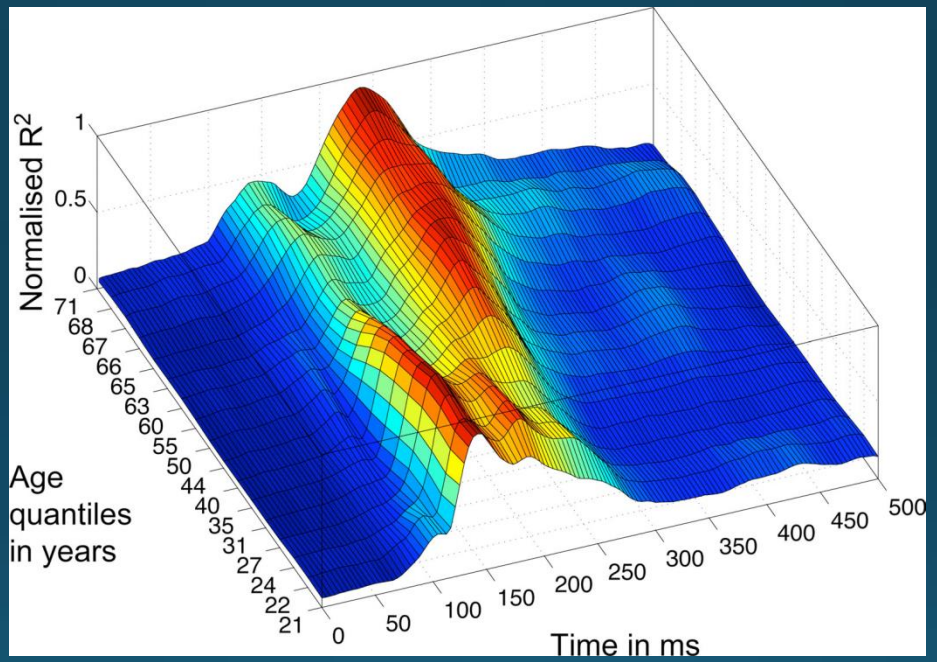


Split continuous variables like factors to control low level physical properties

Bienek, et al (2012). Phase vs Amplitude Spectrum. Journal of Vision 12

Regression based designs – 2nd level





study the effect of stimulus properties within subjects
effect of aging between subjects

Rousselet, et al. (2010). Aging and face perception. Front Psy