University of Glasgow | Institute of Neuroscience & Psychology

CENTRE FOR COGNITIVE NEUROIMAGING

Robust statistics in EEGLAB

Guillaume A. Rousselet

with extra slides from:
Arnaud Delorme, Robert Oostenveld & Cyril Pernet

Robust statistics

Classic statistics make unrealistic assumptions about the distribution of data samples that lead to biased estimates of location and scale, which in turn, lead to low statistical power.

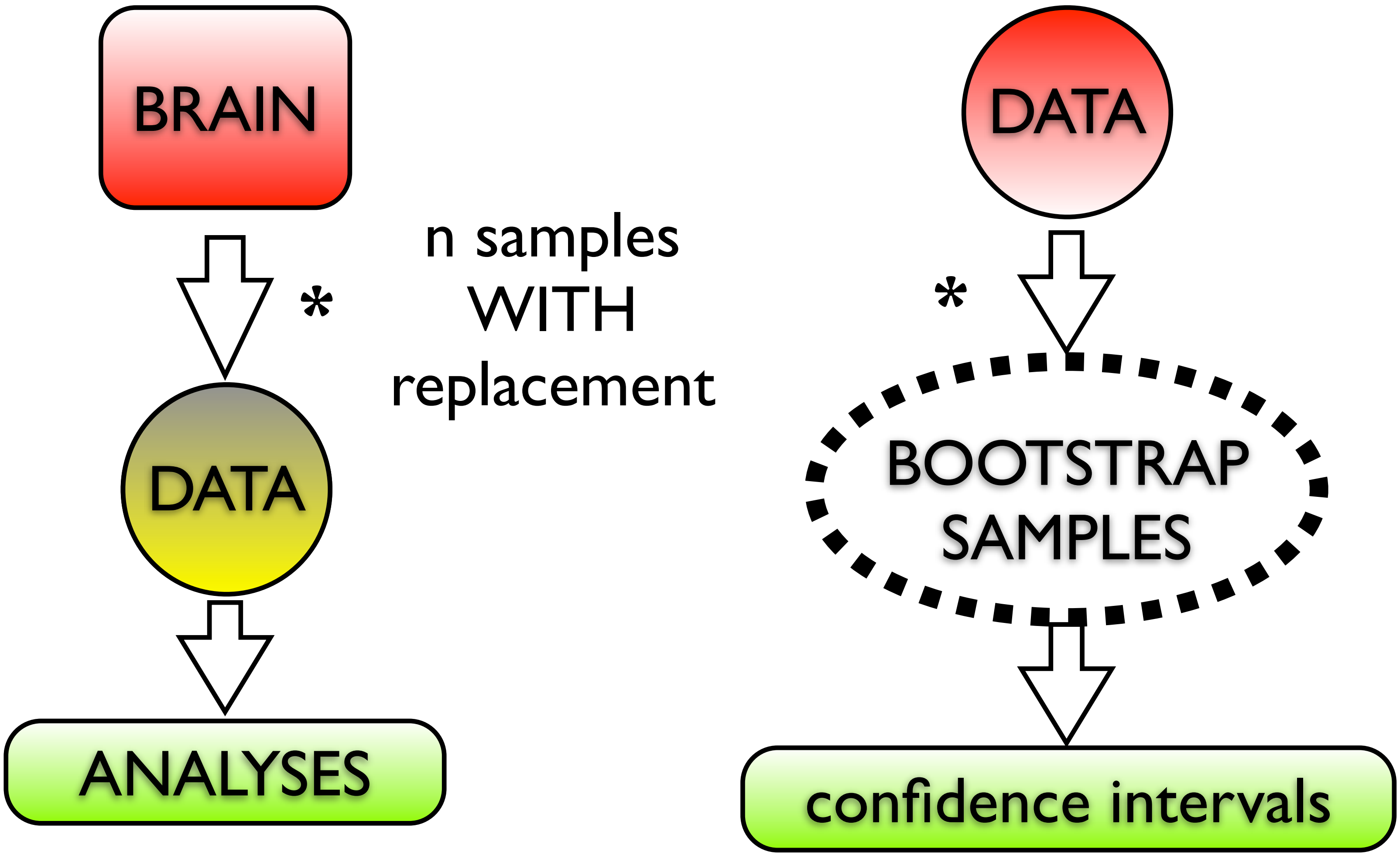
Robust statistics provide a range of tools that perform well under normality and continue to perform relatively well under non normality: control type II error rate.

Robust estimators: central tendency, entire distributions, dispersion;

Bootstrap and permutation methods: shuffle/bootstrap data and recompute estimators;

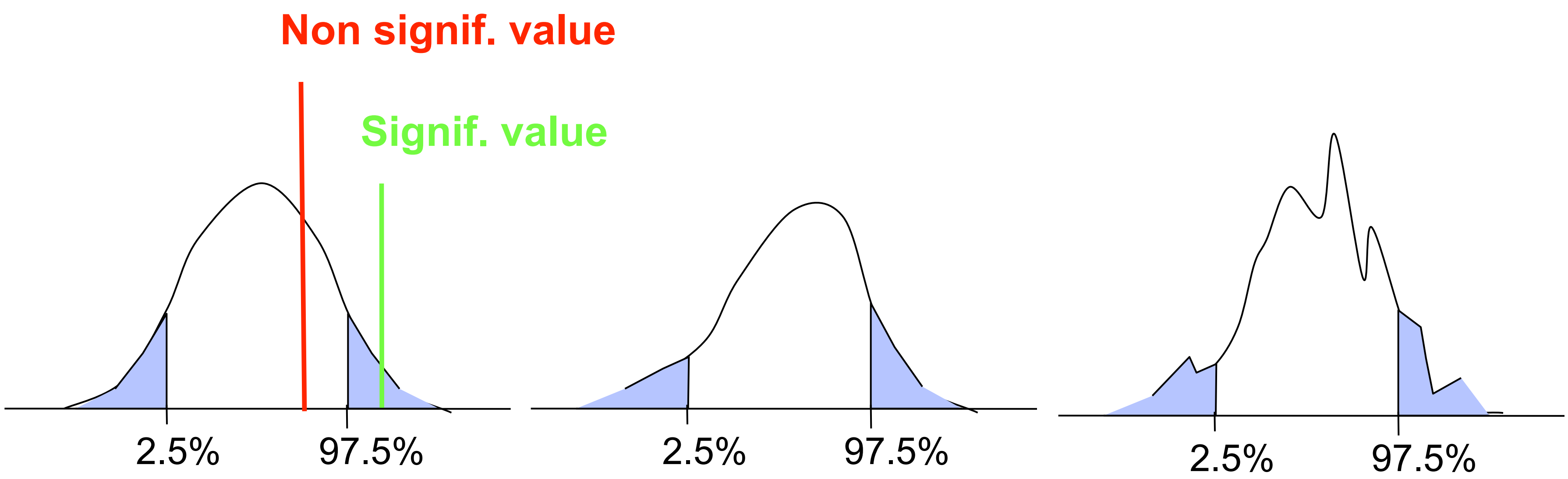
Correction for multiple comparisons: control type I error rate.

bootstrap philosophy

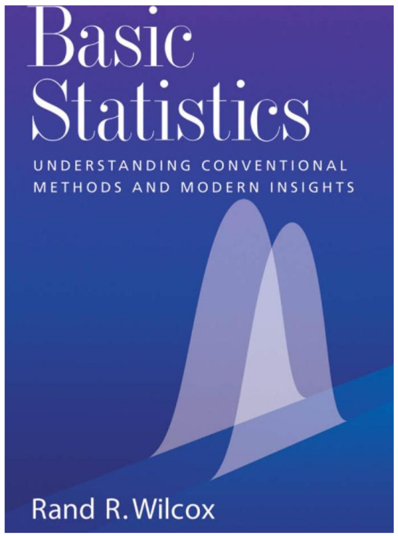


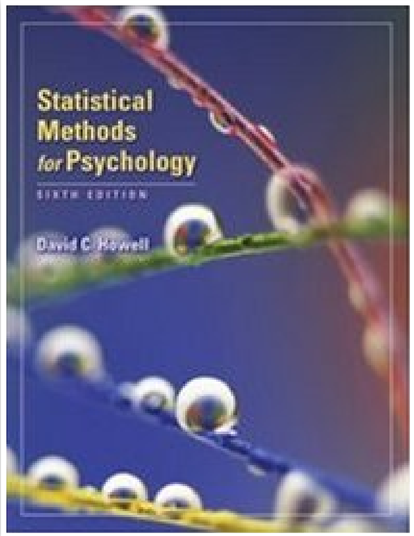
- Given that we have no other information about the population, the sample is our best single estimate of the population.

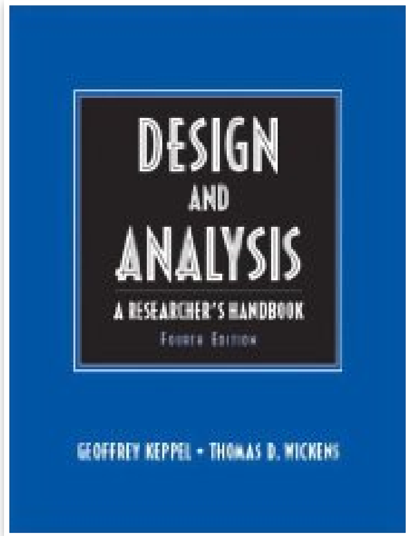
Distribution can take any shape

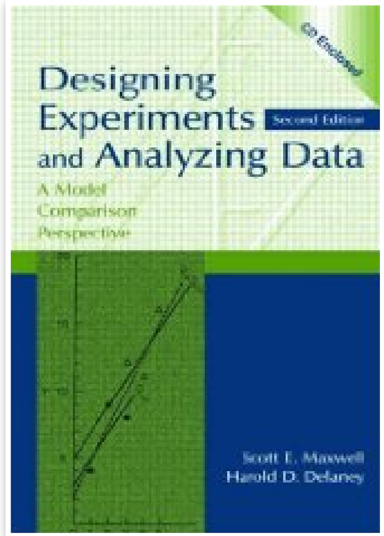


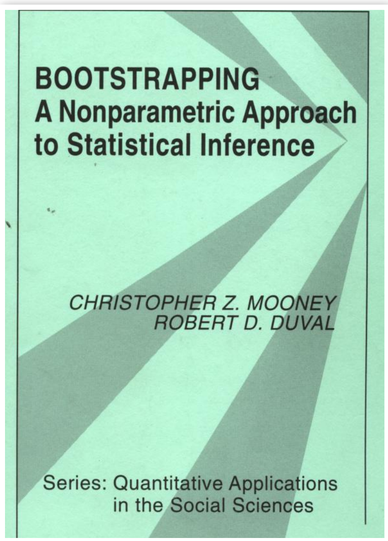
references

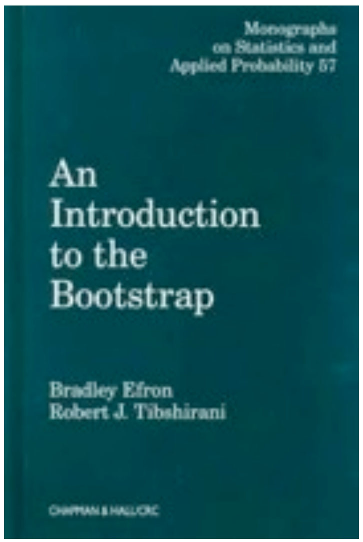


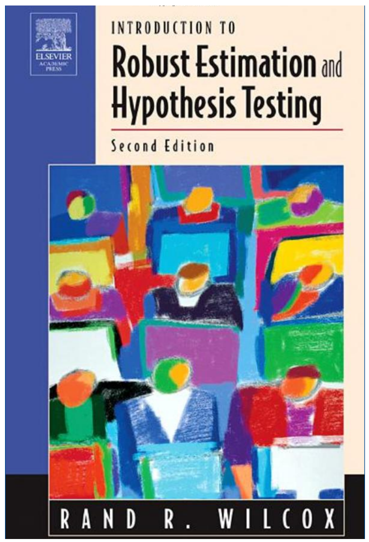


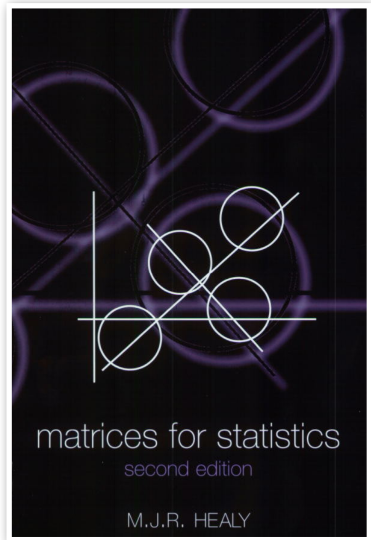












GLM on Cyril Pernet's webpage: http://www.sbirc.ed.ac.uk/lcl/

References

Delorme, A. 2006. Statistical methods. *Encyclopedia of Medical Device and Instrumentation*, vol 6, pp 240-264. Wiley interscience.

Genovese et al. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15: 870-878

Nichols & Hayasaka, 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, 12:419-446

Maris, 2004. Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41: 142-151

Maris et al. 2007. Nonparametric statistical testing of coherence differences. *Journal of Neuroscience Methods*, 163: 161-175

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing* (2nd Ed. ed.). New York, NY: Elsevier Academic Press.

Rousselet, G. A., & Pernet, C. R. (2011). Quantifying the time course of visual object processing using ERPs: it's time to up the game. Frontiers in Psychology 2(107)

Skewness & kurtosis:

DeCarlo, L.T., 1997. On the Meaning and Use of Kurtosis. *Psychological Methods*, 2(3), 292-307.

P value and significance test:

Cohen, J., 1994. The Earth Is Round (p < .05). *American Psychologist*, 49(12), 997-1003.

Goodman, S. N., 1999. Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, 130(12), 995-1004.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779-804.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6, 291-298.

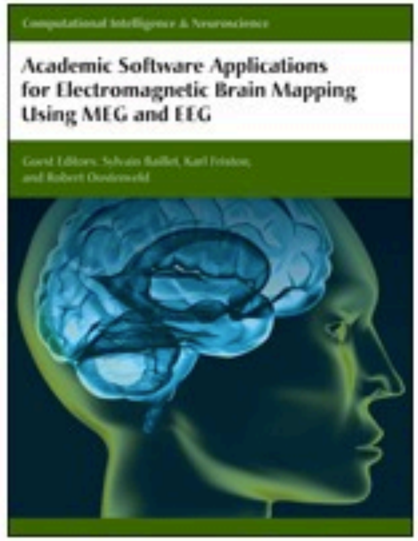
Bootstrap and robust statistics:

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: measures of central tendency. Psychol Methods, 8(3), 254-274.

Seco, Izquierdo, García, & Díez. (2006). A Comparison of the Bootstrap-F, Improved General Approximation, and Brown-Forsythe Multivariate Approaches in a Mixed Repeated Measures Design. *Educational and Psychological Measurement*, 66(1), 35-62.

Academic Software Applications for Electromagnetic Brain Mapping Using MEG and EEG

Guest Editors: Sylvain Baillet, Karl Friston, and Robert Oostenveld



- FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data, Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen
  Volume 2011 (2011), Article ID 156869, 9 pages
- Spatiotemporal Analysis of Multichannel EEG: CARTOOL, Denis Brunet, Micah M. Murray, and Christoph M. Michel
  Volume 2011 (2011), Article ID 813870, 15 pages
- Ragu: A Free Tool for the Analysis of EEG and MEG Event-Related Scalp Field Data Using Global Randomization Statistics, Thomas Koenig, Mara Kottlow, Maria Stein, and Lester Melie-García
  Volume 2011 (2011), Article ID 938925, 14 pages
- LIMO EEG: A Toolbox for Hierarchical LInear MOdeling of ElectroEncephaloGraphic Data, Cyril R. Pernet, Nicolas Chauveau, Carl Gaspar, and Guillaume A. Rousselet
  Volume 2011 (2011), Article ID 831409, 11 pages
- EEG and MEG Data Analysis in SPM8, Vladimir Litvak, Jérémie Mattout, Stefan Kiebel, Christophe

References

- **EEGLAB, SIFT, NFT, BCILAB, and ERICA**:
  http://www.hindawi.com/journals/cin/2011/130714/
- **FIELDTRIP**: http://www.hindawi.com/journals/cin/2011/156869/
- **SPM**: http://www.hindawi.com/journals/cin/2011/852961/
- **LIMO EEG**: http://www.hindawi.com/journals/cin/2011/831409/
- **Discriminant analysis**: Parra et al. (2005). Recipes for the linear analysis of EEG. Neuroimage, 28(2), 326-341.
  http://liinc.bme.columbia.edu/mainTemplate.htm?liinc_downloads.htm
- **PLS** (Partial Least Squares; Projection to Latent Structures)
  Krishnan et al. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. NeuroImage, 56, 455-475.
  http://www.rotman-baycrest.on.ca/index.php?section=84

References

- **Connectivity**:

  http://econnectome.umn.edu/

  Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. Neuroimage 52, 1059-1069.

  Edward T. Bullmore, Danielle S. Bassett. Brain graphs: Graphical models of the human brain connectome. Annu Rev Clin Psychol, 2011, 7:113-140.

  http://www.indiana.edu/~cortex/connectivity.html

Goal	Dataset		
	Binomial or Discrete	Continuous measurement (from a normal distribution)	Continuous measurement, Rank, or Score (from non-normal distribution)
Example of data sample	List of patients recovering or not after a treatment	Readings of heart pressure from several patients	Ranking of several treatment efficiency by one expert
Describe one data sample	Proportions	Mean, SD	Median
Compare one data sample to a hypothetical distribution	χ^2 or binomial test	One-sample t test	Sign test or Wilcoxon test
Compare two paired samples	Sign test	Paired t test	Sign test or Wilcoxon test
Compare two unpaired samples	χ^2 square Fisher's exact test	Unpaired t test	Mann-Whitney test
Compare three or more unmatched samples	χ^2 test	One-way ANOVA	Kruskal-Wallis test
Compare three or more matched samples	Cochrane Q test	Repeated-measures ANOVA	Friedman test
Quantify association between two paired samples	Contingency coefficients	Pearson correlation	Spearman correlation

Matlab Statistics toolbox; Parra & Sajda plugin

EEGLAB FIELDTRIP LIMO EEG

Matlab Statistics toolbox

Delorme, A. (2006) Statistical methods. *Encyclopedia of Medical Device and Instrumentation*, vol 6, pp 240-264. Wiley interscience.

statcond function in EEGLAB

```
[stats, df, pvals, surrog] = statcond( data, 'key','val'... );
```

```
% Optional inputs:
%   'paired'    = ['on'|'off'] pair the data array {default: 'on' unless
%                 the last dimension of data array is of different lengths}.
%   'mode'      = ['perm'|'bootstrap'|'param'] mode for computing the p-values:
%                  'param' = parametric testing (standard ANOVA or t-test);
%                  'perm' = non-parametric testing using surrogate data
%                  'bootstrap' = non-parametric bootstrap
%                  made by permuting the input data {default: 'param'}
%   'naccu'     = [integer] Number of surrogate data copies to use in 'perm'
%                  or 'bootstrap' mode estimation (see above) {default: 200}.
%   'verbose'   = ['on'|'off'] print info on the command line {default: 'on'}.
%   'variance'  = ['homegenous'|'inhomogenous'] this option is exclusively
%                 for parametric statistics using unpaired t-test. It allows
%                 to compute a more accurate value for the degree of freedom
%                 using the formula for inhomogenous variance (see
%                 ttest2_cell function). Default is 'homegenous'.
```

- constraint: only the last dimension may differ across the n conditions.

statcond function in EEGLAB

```
Outputs:
  stats      = F- or T-value array of the same size as input data without
               the last dimension. A T value is returned only when the data
               includes exactly two conditions.
  df         = degrees of freedom, a (2,1) vector, when F-values are returned
  pvals      = array of p-values. Same size as input data without the last
               data dimension. All returned p-values are two-tailed.
  surrog     = surrogate data array (same size as input data with the last
               dimension filled with a number ('naccu') of surrogate data sets.

Important note: When a two-way ANOVA is performed, outputs are cell arrays
                with three elements: output(1) = column effects;
                output(2) = row effects; output(3) = interactions
                between rows and columns.
```

statcond function in EEGLAB

```
a = { rand(1,10) rand(1,10)+0.5 }; % pseudo 'paired' data vectors

[t df pvals] = statcond(a , 'mode', 'perm'); % perform paired t-test
pvals = 5.2807e-04 % standard t-test probability value

% Note: for different rand() outputs, results will differ.
[t df pvals surog] = statcond(a, 'mode', 'perm', 'naccu', 2000);
pvals = 0.0065 % nonparametric t-test using 2000 permuted data sets

a = { rand(2,11) rand(2,10) rand(2,12)+0.5 };
[F df pvals] = statcond(a , 'mode', 'perm'); % perform an unpaired ANOVA

pvals =
    0.00025 % p-values for difference between columns
    0.00002 % for each data row
```

statcond function in EEGLAB

```
a = { rand(3,4,10) rand(3,4,10) rand(3,4,10); ...
      rand(3,4,10) rand(3,4,10) rand(3,4,10)+0.5 };

% pseudo (2,3)-condition data array, each entry containing
% ten (3,4) data matrices
[F df pvals] = statcond(a , 'mode', 'perm');
                              % paired 2-way ANOVA

% Output:
pvals{1} % a (3,4) matrix of p-values; effects across columns
pvals{2} % a (3,4) matrix of p-values; effects across rows
pvals{3} % a (3,4) matrix of p-values; interaction effects across
   rows and columns
```

Exercise

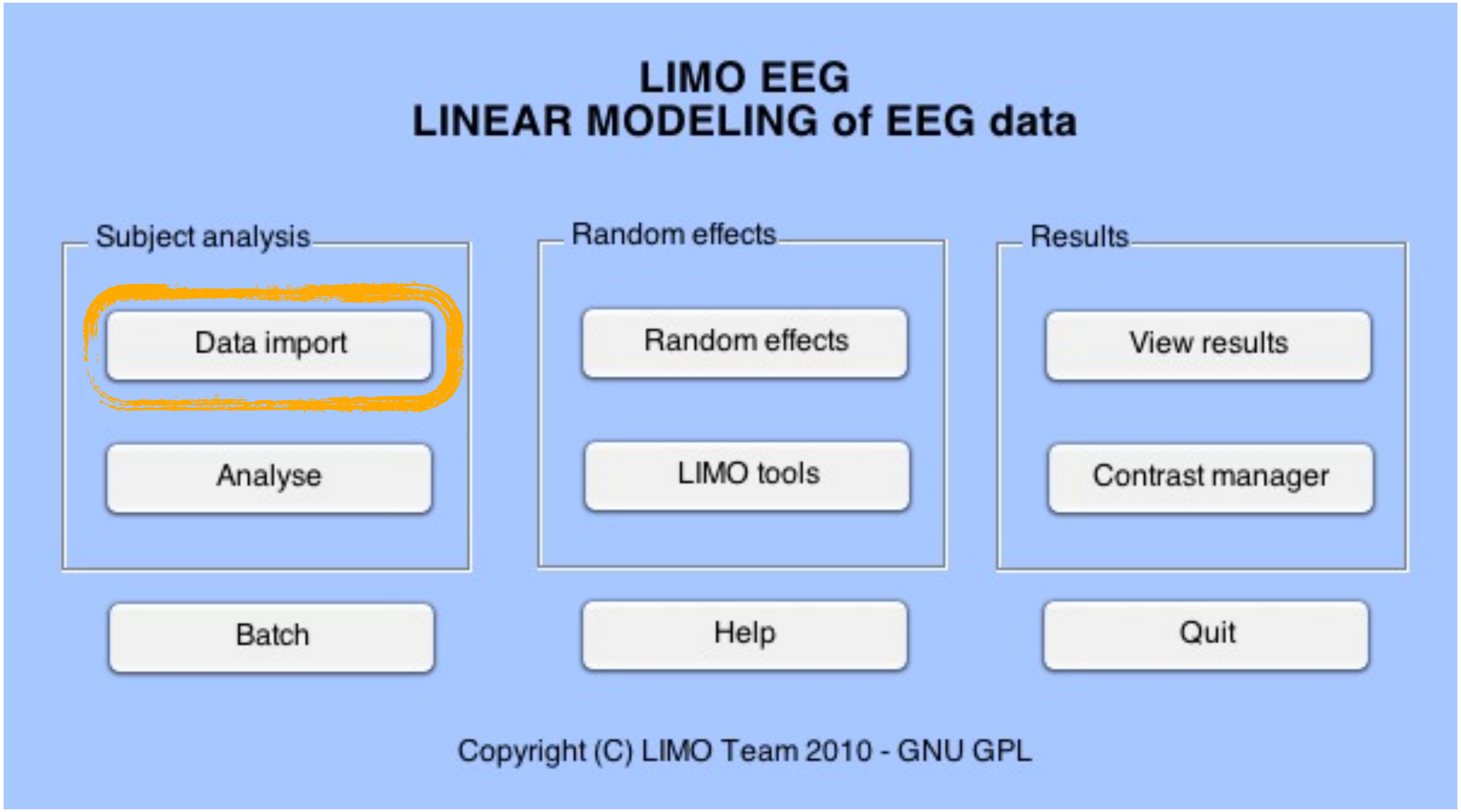
- Experiment with the statcond function
  - Create 2 random vectors of values.
  - Add “signal” to one of the variable.
  - Use statcond and compare permutation and parametric results.
  - Repeat 100 times and plot the histogram of p-values for each technique on the same figure.

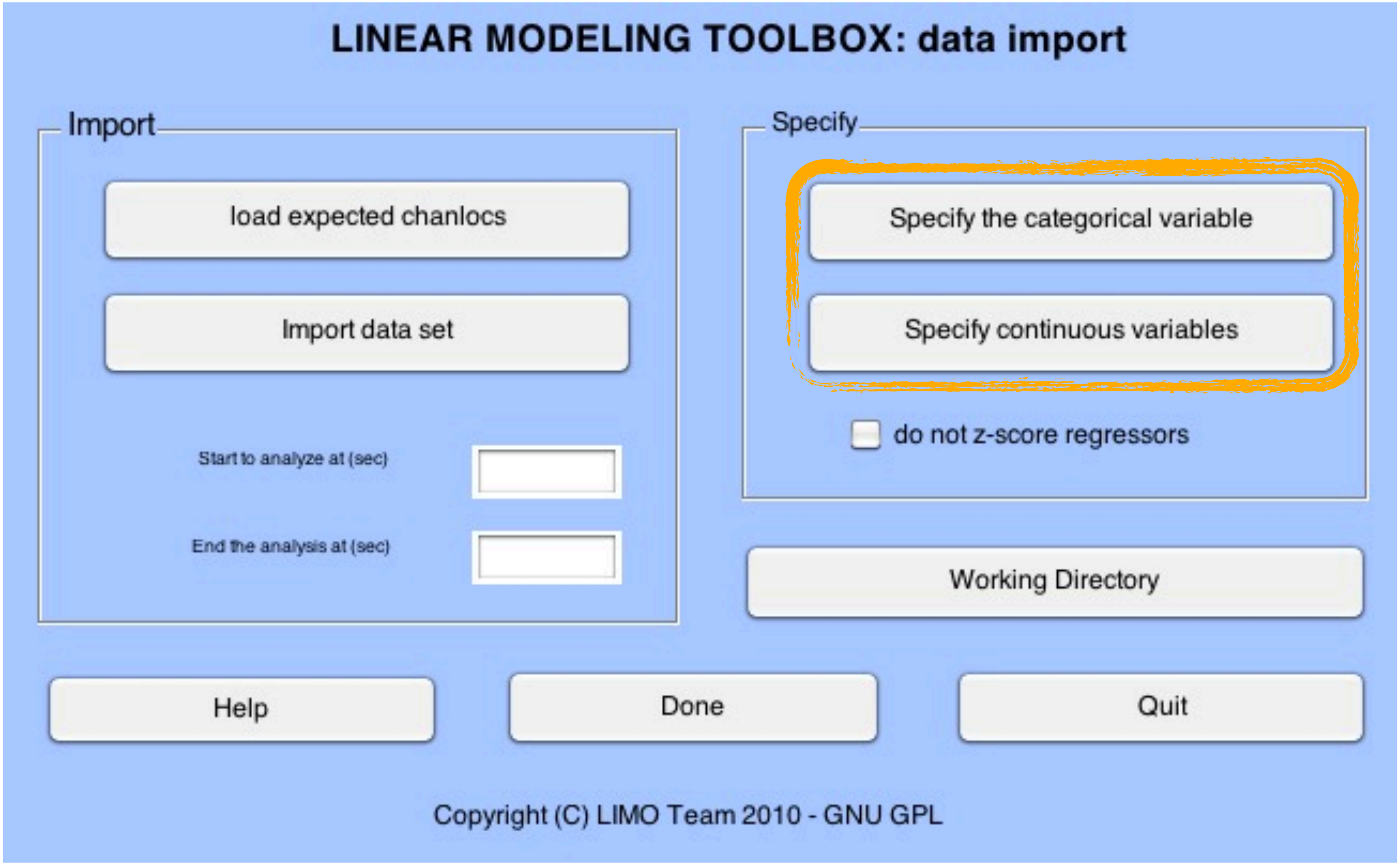
LIMO EEG

- Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical LInear Modeling of EletroEncephaloGraphic data. Computational Intelligence and Neuroscience, Article ID 831409.
- LIMO EEG is freely available @

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

>> limo_eeg


LIMO EEG
LINEAR MODELING of EEG data
Subject analysis
Data import
Analyse
Random effects
Random effects
LIMO tools
Results
View results
Contrast manager
Batch
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

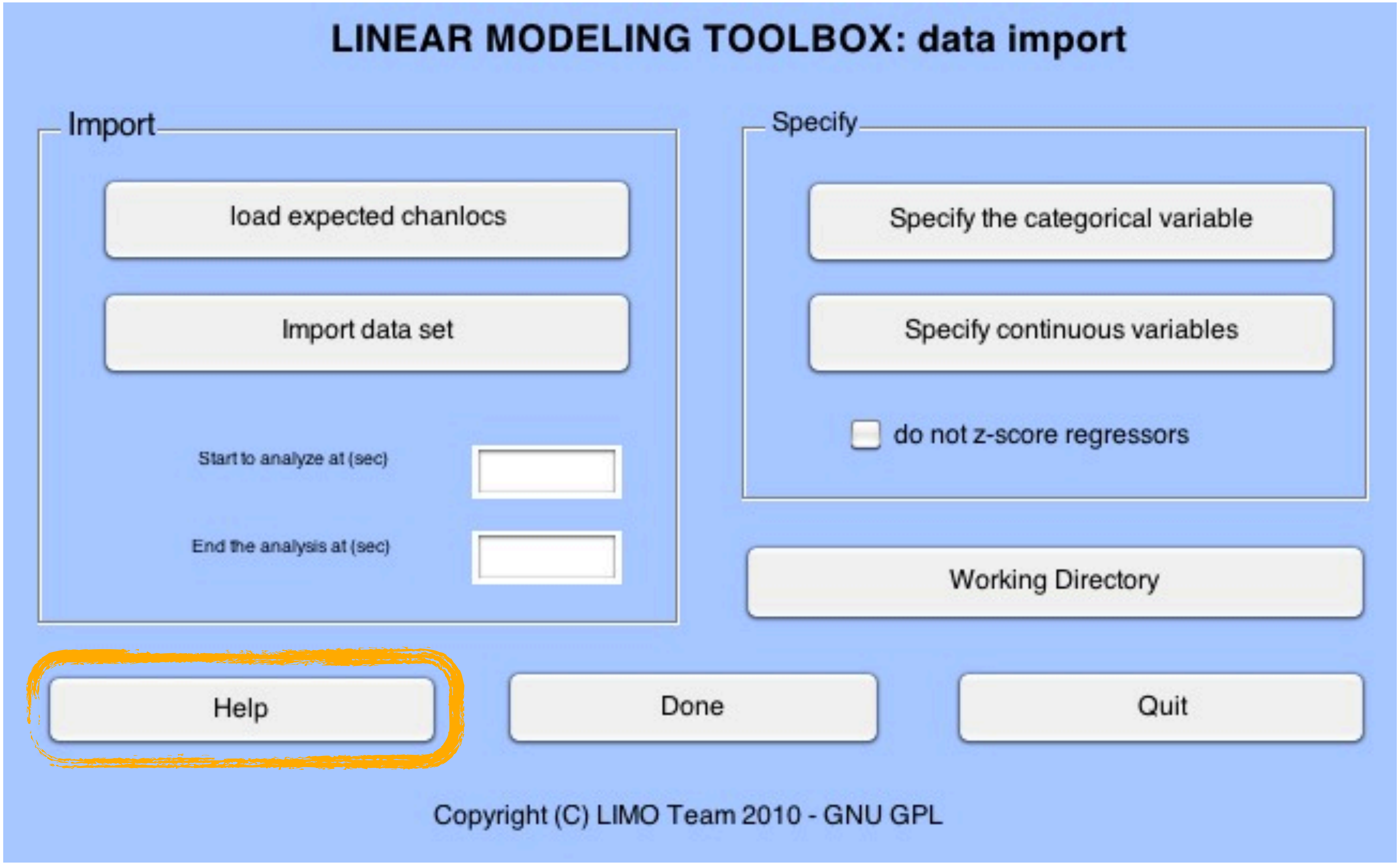

LINEAR MODELING TOOLBOX: data import
Import
load expected chanlocs
Import data set
Start to analyze at (sec)
End the analysis at (sec)
Specify
Specify the categorical variable
Specify continuous variables
do not z-score regressors
Working Directory
Help
Done
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

categorical_variable.txt

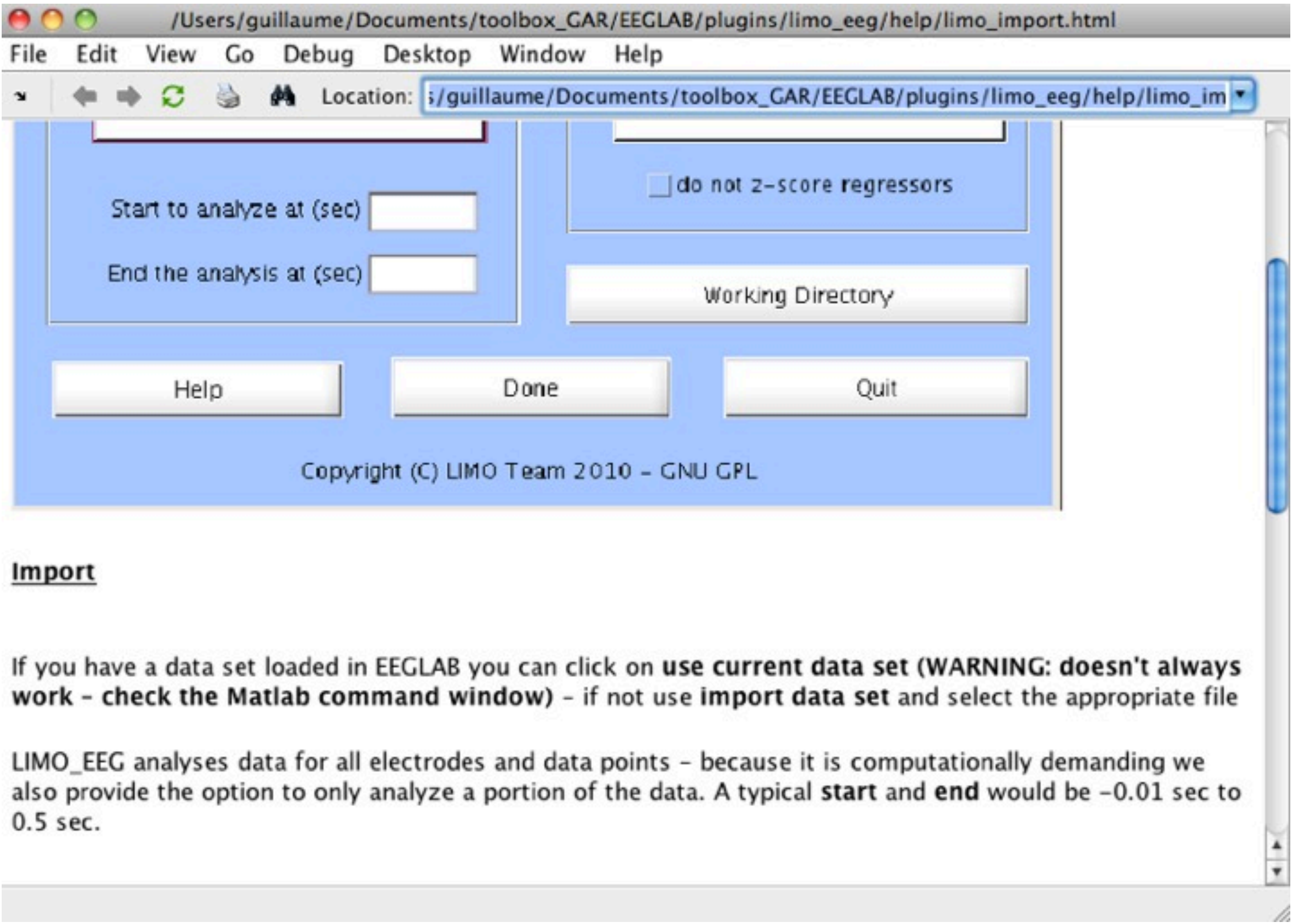
```
2.0000000e+00
1.0000000e+00
2.0000000e+00
1.0000000e+00
1.0000000e+00
2.0000000e+00
1.0000000e+00
2.0000000e+00
1.0000000e+00
2.0000000e+00
2.0000000e+00
2.0000000e+00
1.0000000e+00
2.0000000e+00
2.0000000e+00
2.0000000e+00
1.0000000e+00
1.0000000e+00
2.0000000e+00
2.0000000e+00
1.0000000e+00
1.0000000e+00
1.0000000e+00
1.0000000e+00
2.0000000e+00
1.0000000e+00
2.0000000e+00
1.0000000e+00
1.0000000e+00
1.0000000e+00
```

continuous_variables.txt

```
6.0000000e-01
5.5000000e-01
3.0000000e-01
1.0000000e-01
1.5000000e-01
1.0000000e-01
8.5000000e-01
8.0000000e-01
2.5000000e-01
6.5000000e-01
8.5000000e-01
4.0000000e-01
2.5000000e-01
3.5000000e-01
6.0000000e-01
4.5000000e-01
3.0000000e-01
2.0000000e-01
8.0000000e-01
5.0000000e-02
0.0000000e+00
1.0000000e-01
4.5000000e-01
6.0000000e-01
1.5000000e-01
0.0000000e+00
8.5000000e-01
7.0000000e-01
1.5000000e-01
3.5000000e-01
```



LINEAR MODELING TOOLBOX: data import
Import
load expected chanlocs
Import data set
Start to analyze at (sec)
End the analysis at (sec)
Specify
Specify the categorical variable
Specify continuous variables
do not z-score regressors
Working Directory
Help
Done
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

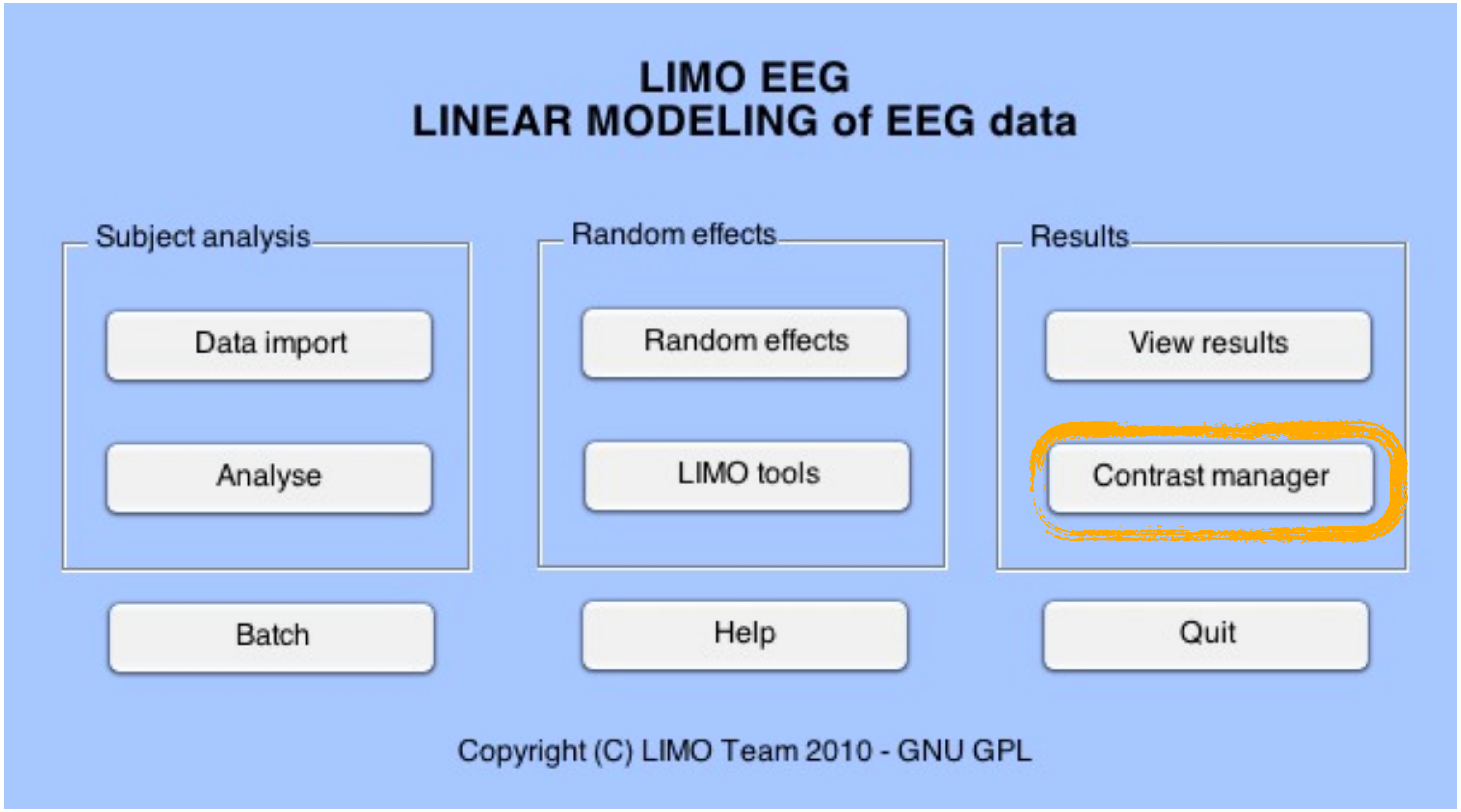


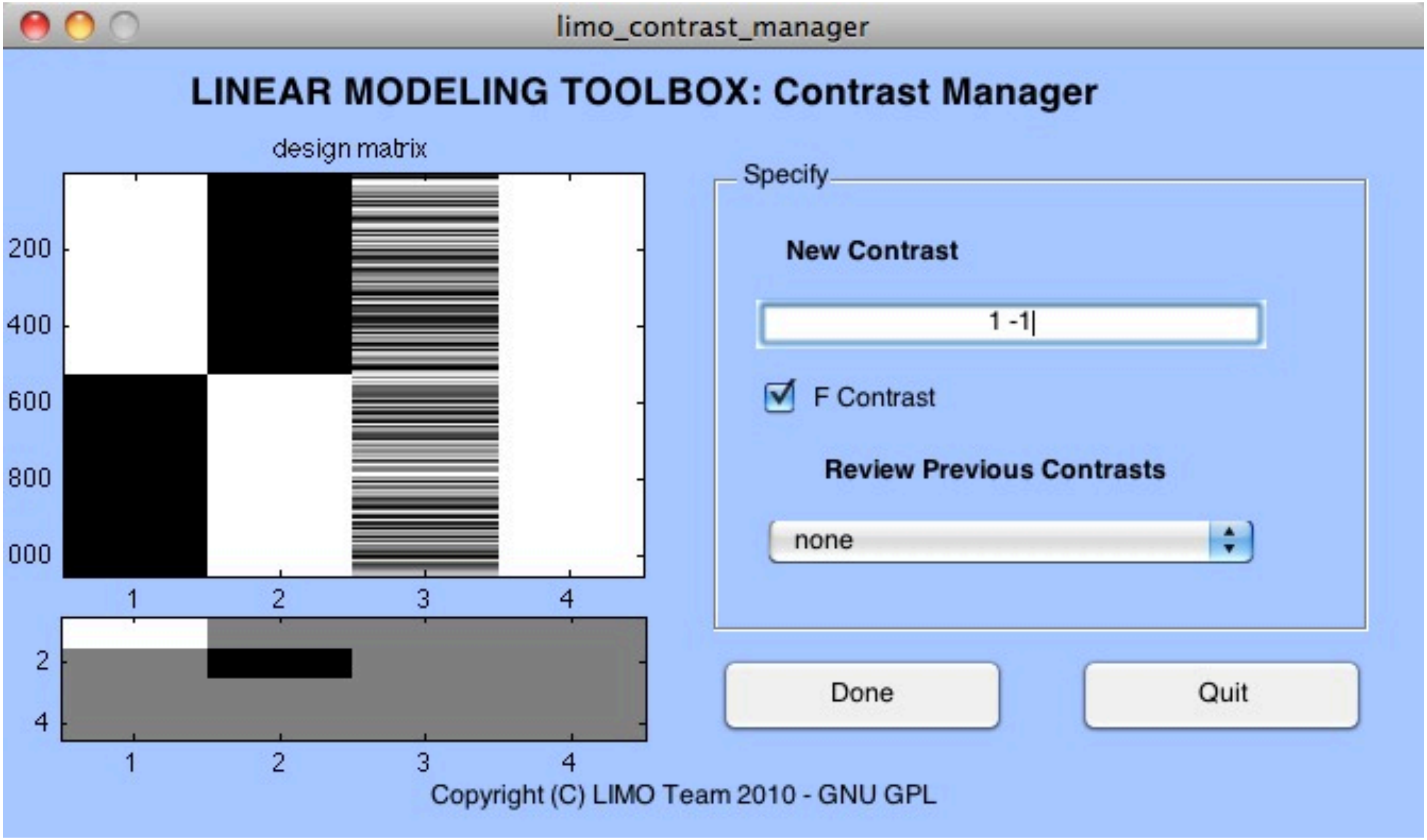
Import

If you have a data set loaded in EEGLAB you can click on **use current data set (WARNING: doesn't always work - check the Matlab command window)** - if not use **import data set** and select the appropriate file

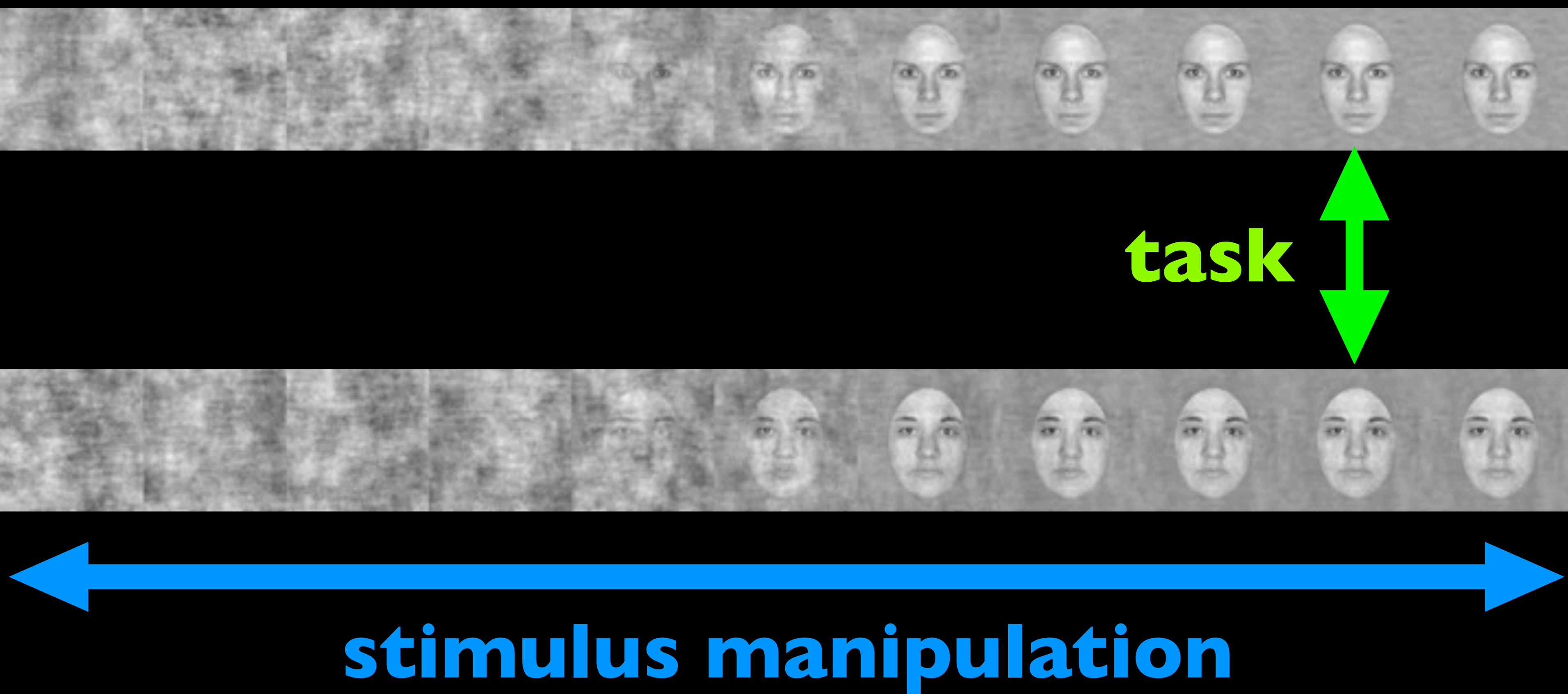
LIMO_EEG analyses data for all electrodes and data points - because it is computationally demanding we also provide the option to only analyze a portion of the data. A typical **start** and **end** would be -0.01 sec to 0.5 sec.

>> limo_eeg

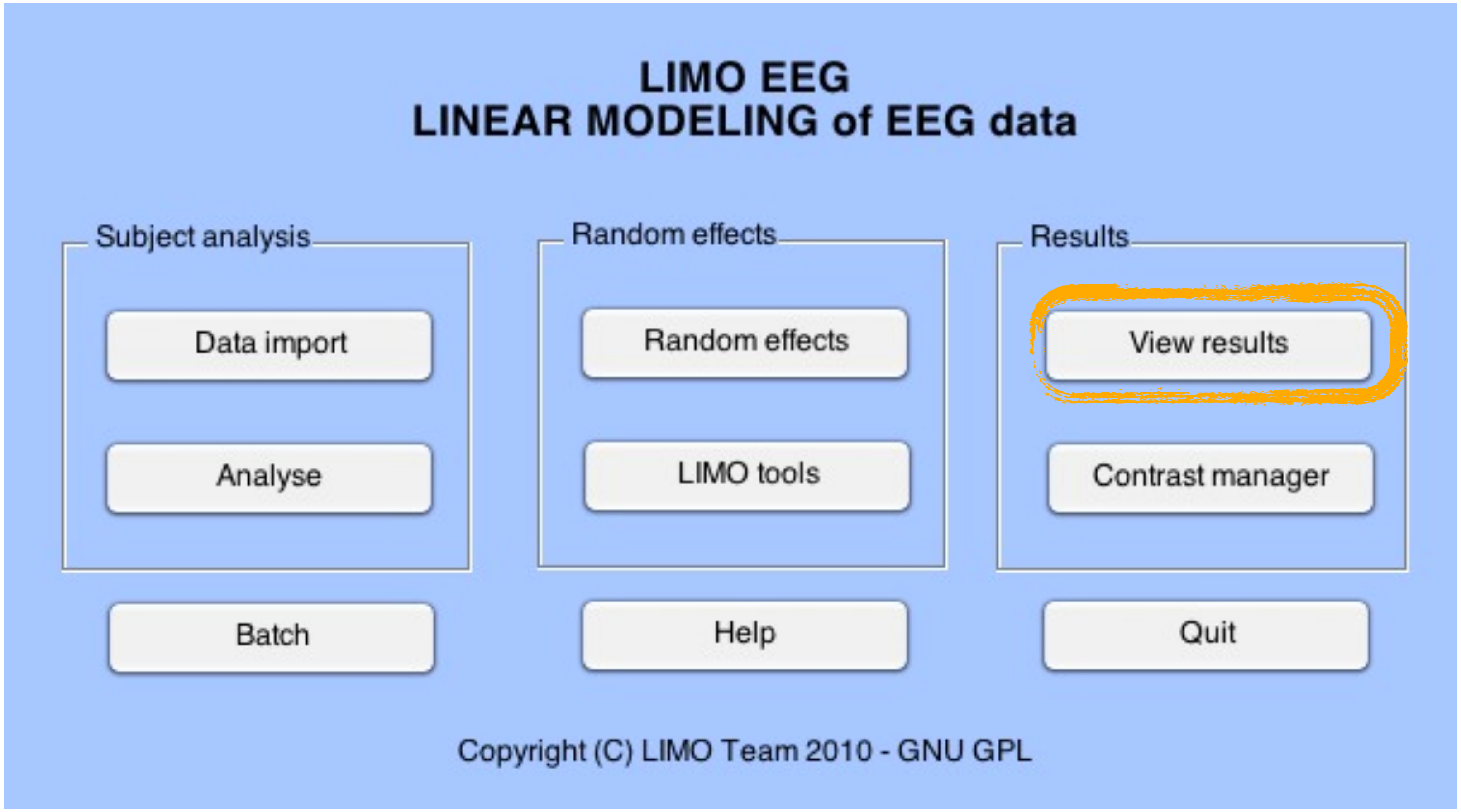

LIMO EEG
LINEAR MODELING of EEG data
Subject analysis
Data import
Analyse
Random effects
Random effects
LIMO tools
Results
View results
Contrast manager
Batch
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

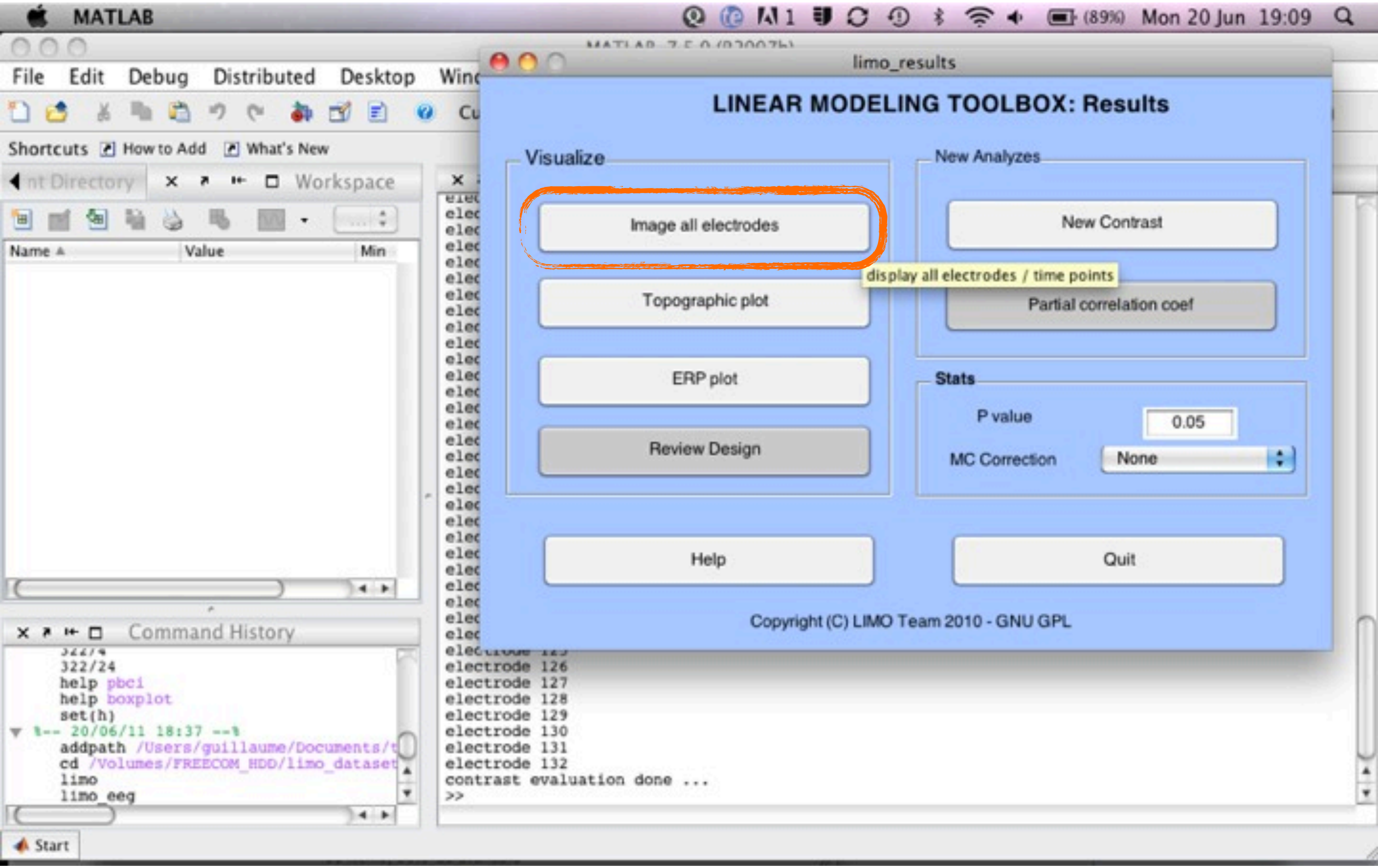

limo_contrast_manager
LINEAR MODELING TOOLBOX: Contrast Manager
design matrix
200
400
600
800
000
1
2
3
4
2
4
1
2
3
4
Specify
New Contrast
1 -1
F Contrast
Review Previous Contrasts
none
Done
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

Tracking EEG sensitivity to phase noise using a parametric design

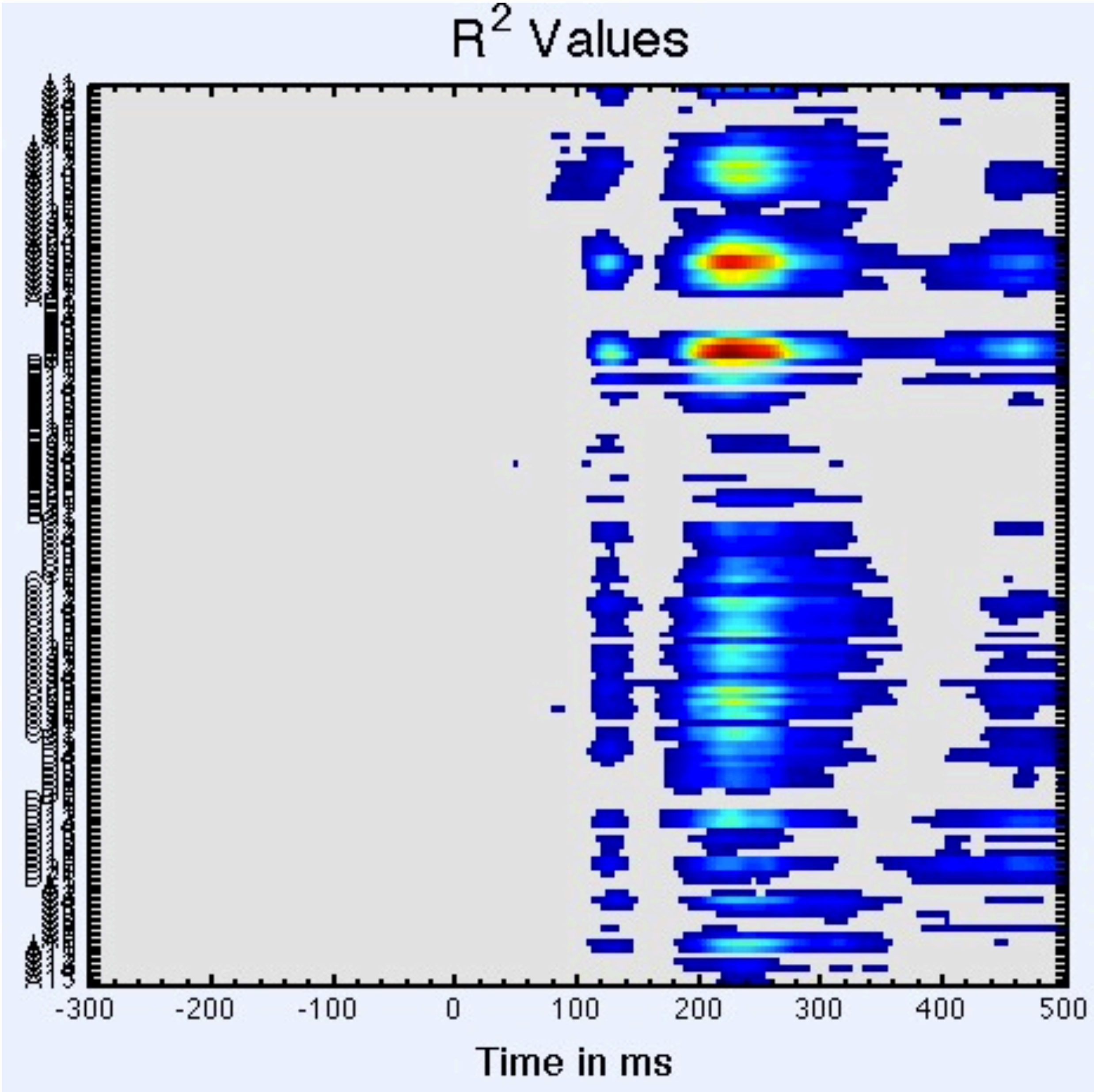


>> limo_eeg

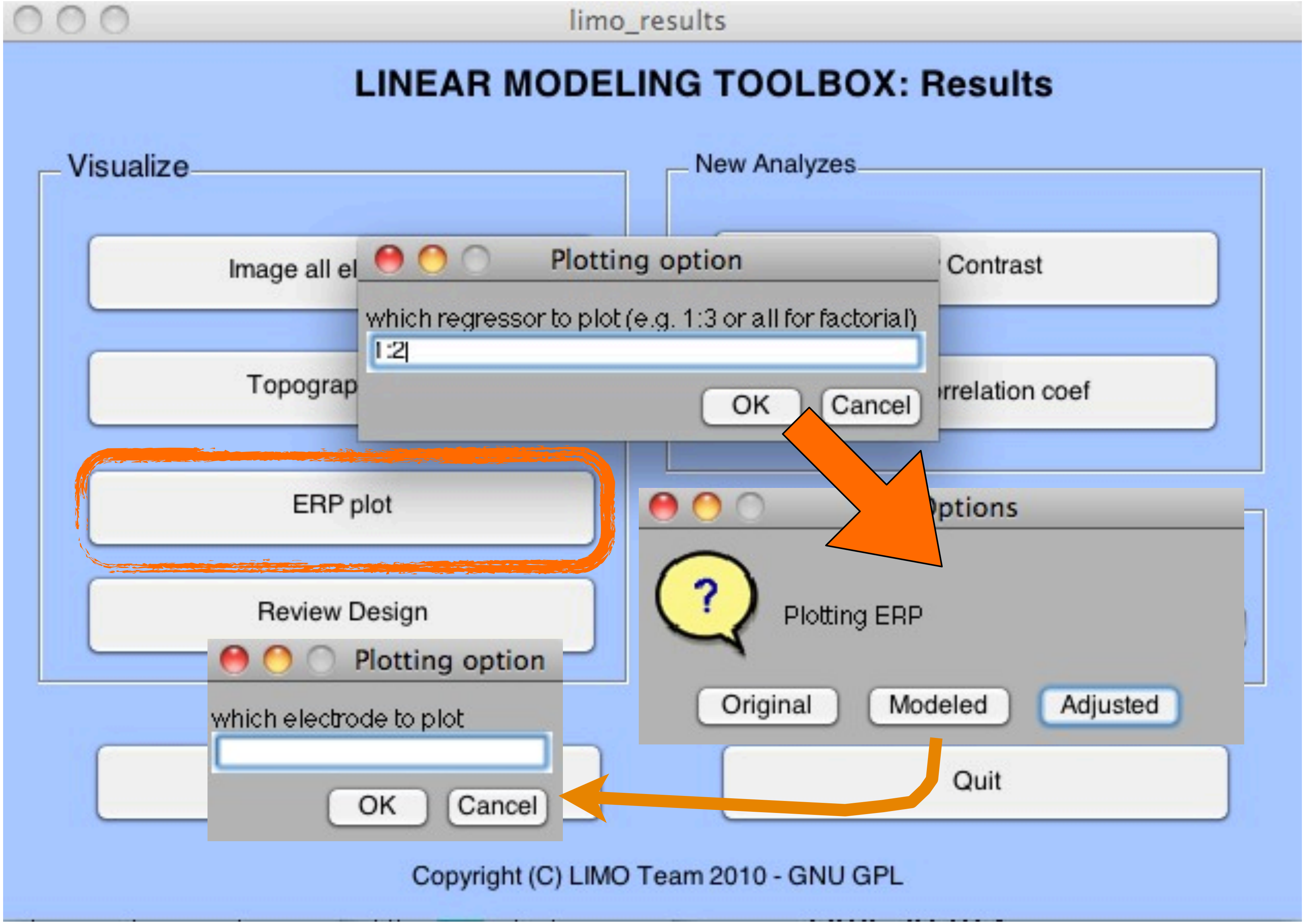

LIMO EEG
LINEAR MODELING of EEG data
Subject analysis
Data import
Analyse
Random effects
Random effects
LIMO tools
Results
View results
Contrast manager
Batch
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL



MATLAB
Mon 20 Jun 19:09
File Edit Debug Distributed Desktop
Shortcuts How to Add What's New
Workspace
Name Value Min
Command History
322/24
help pbci
help boxplot
set(h)
addpath /Users/guillaume/Documents/t
cd /Volumes/FREECOM_HDD/limo_dataset
limo
limo_eeg
electrode 126
electrode 127
electrode 128
electrode 129
electrode 130
electrode 131
electrode 132
contrast evaluation done ...
>>
limo_results
LINEAR MODELING TOOLBOX: Results
Visualize
Image all electrodes
display all electrodes / time points
Topographic plot
ERP plot
Review Design
New Analyzes
New Contrast
Partial correlation coef
Stats
P value
0.05
MC Correction
None
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL
Start

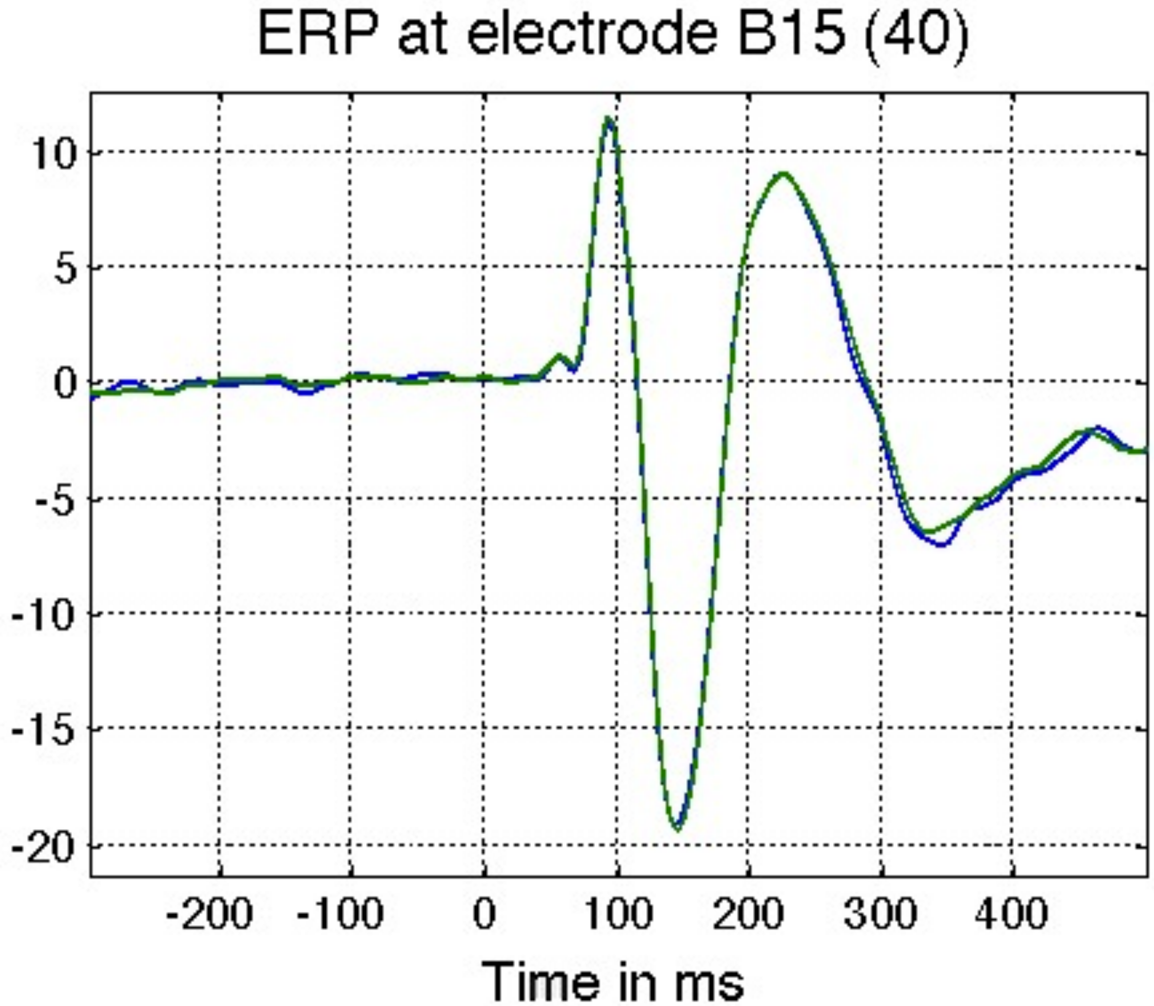


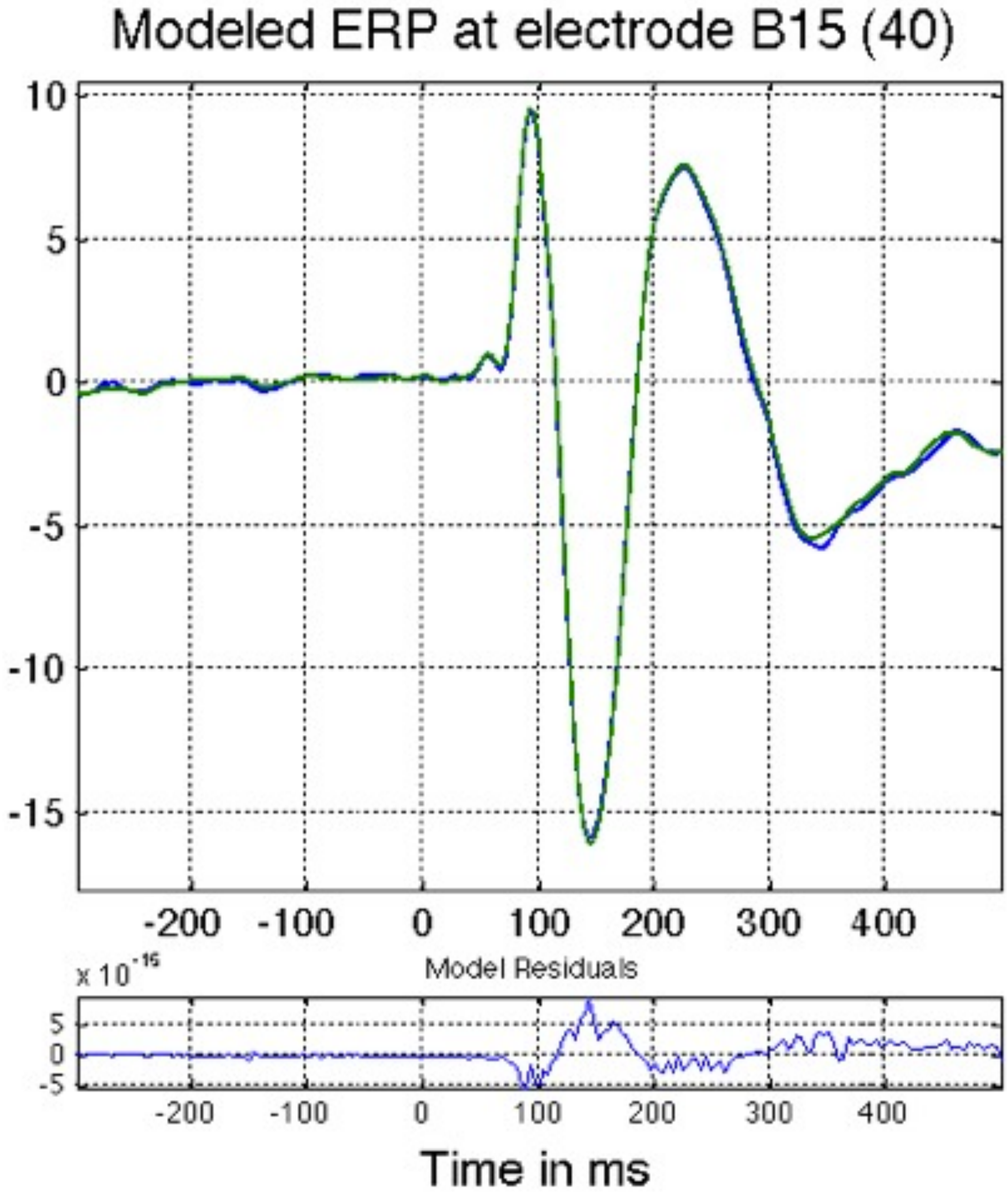
R² Values
-300
-200
-100
0
100
200
300
400
500
Time in ms

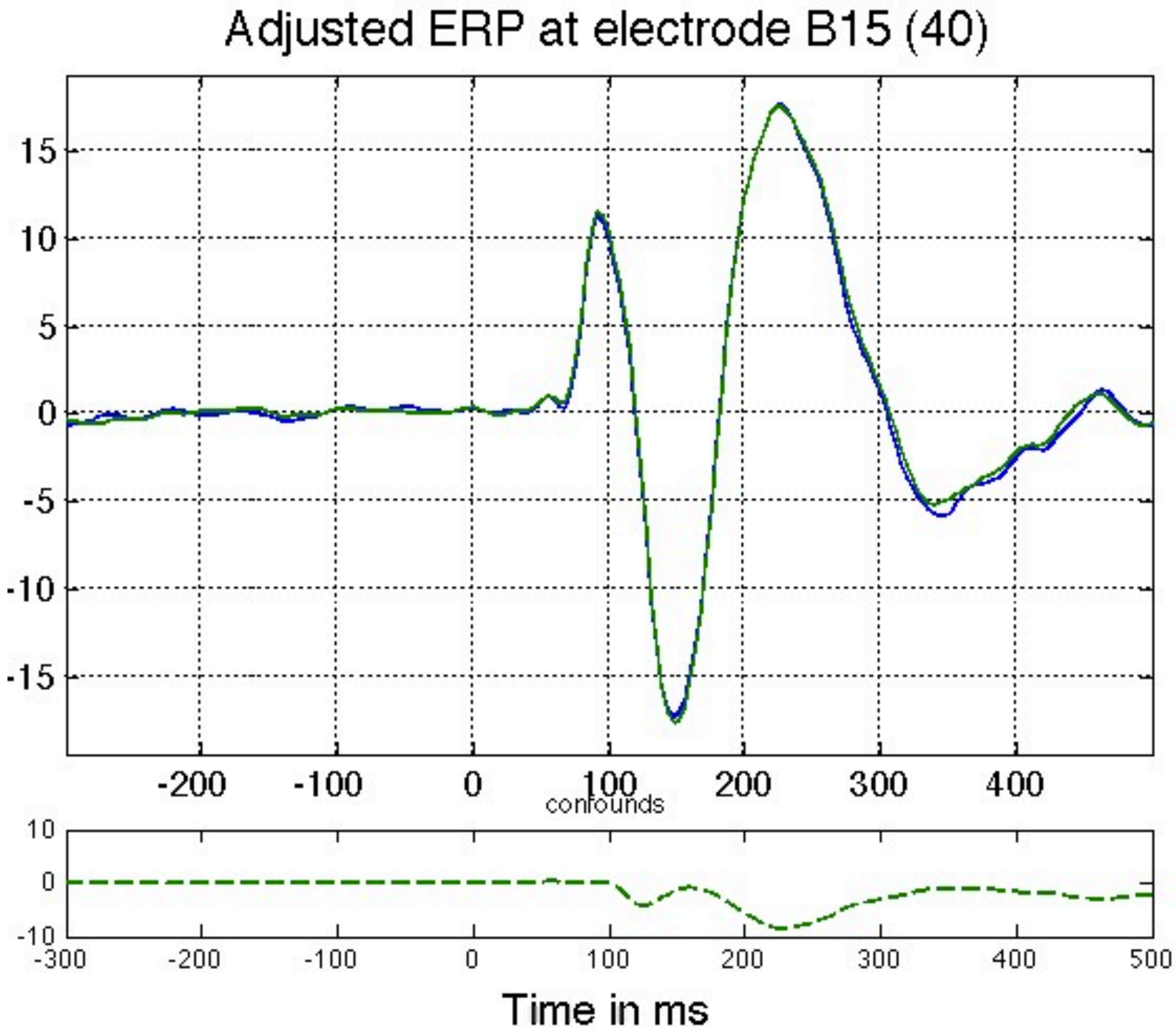


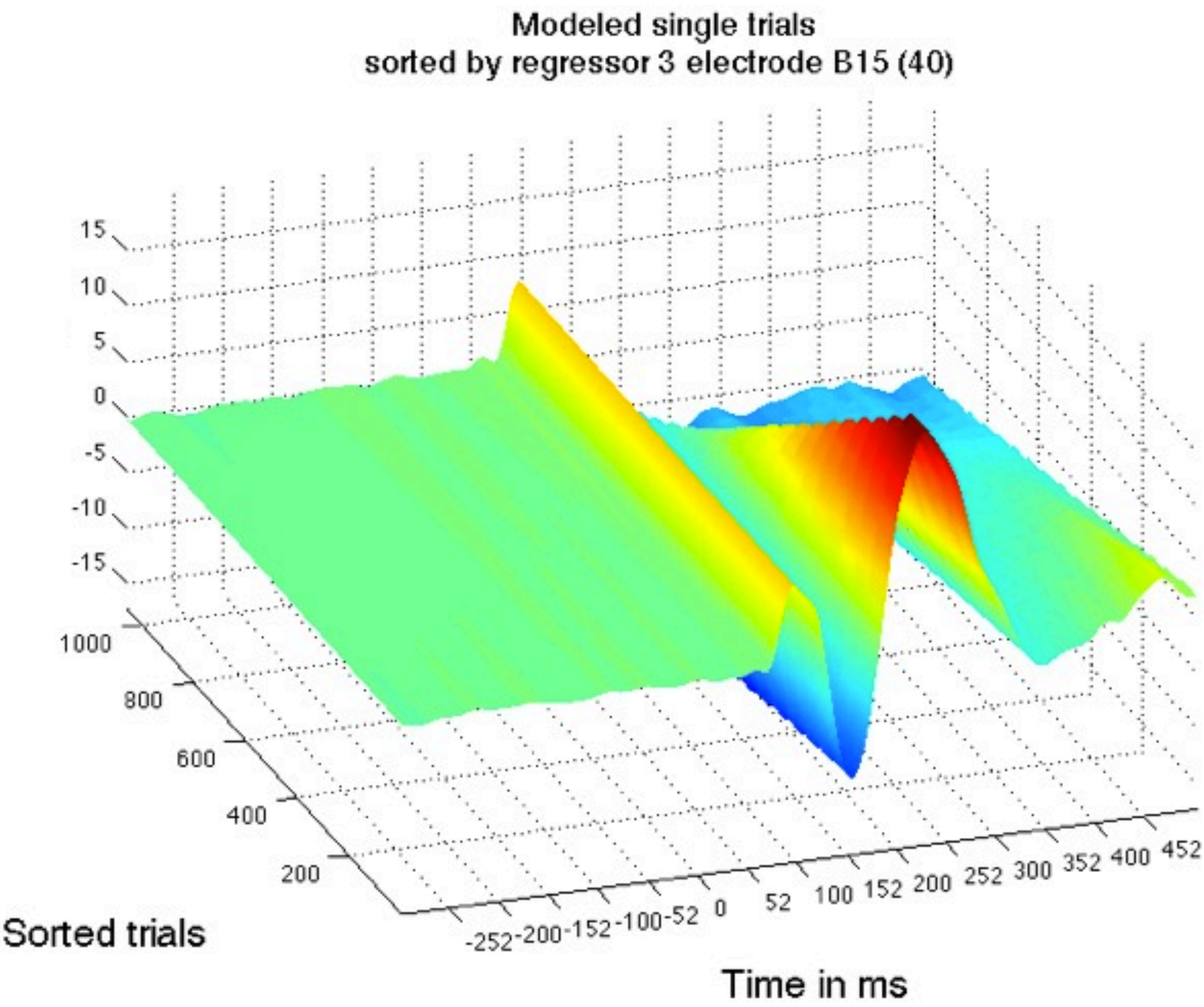
'ok' with no electrode => plot at electrode with max R^2

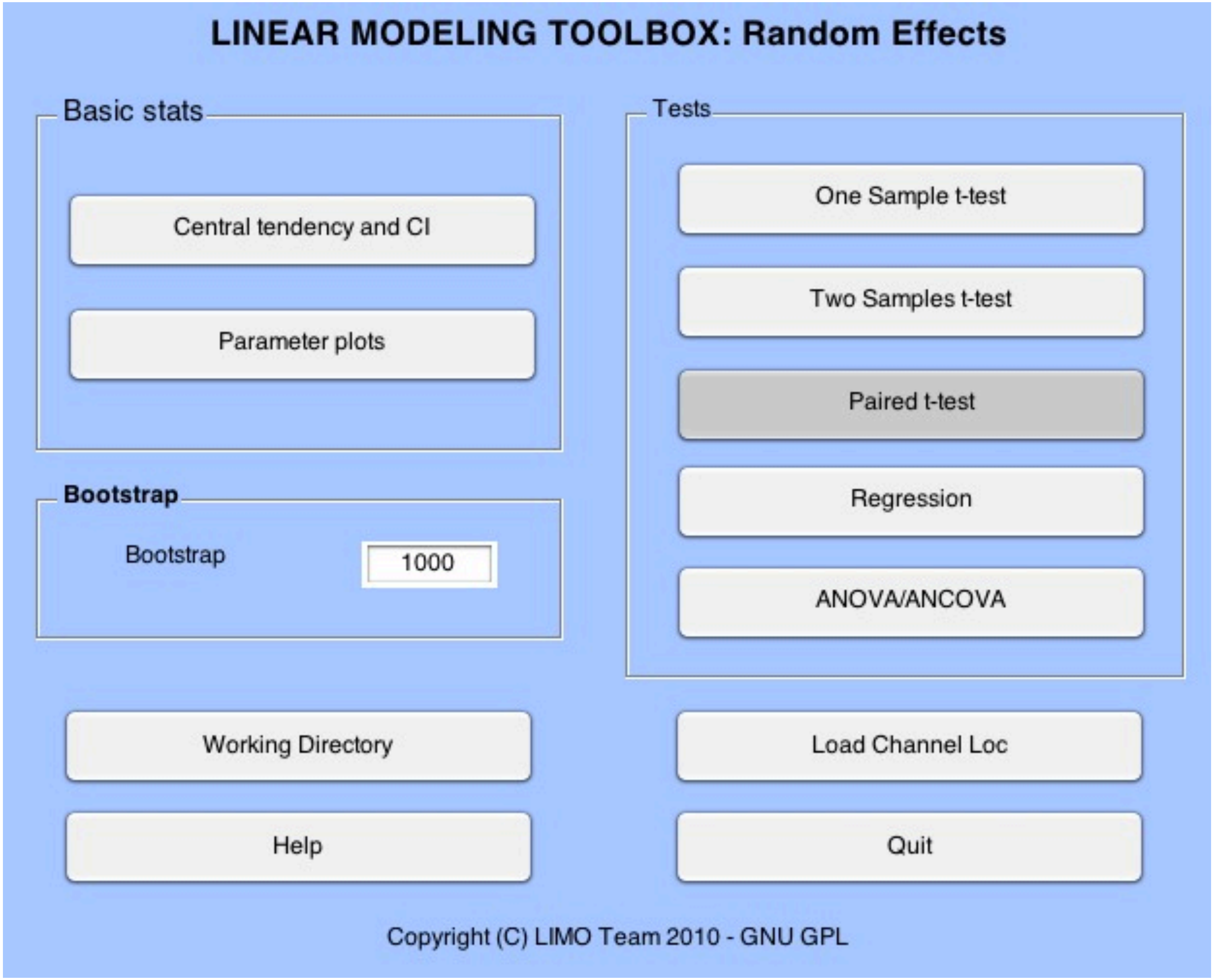
original (raw) data

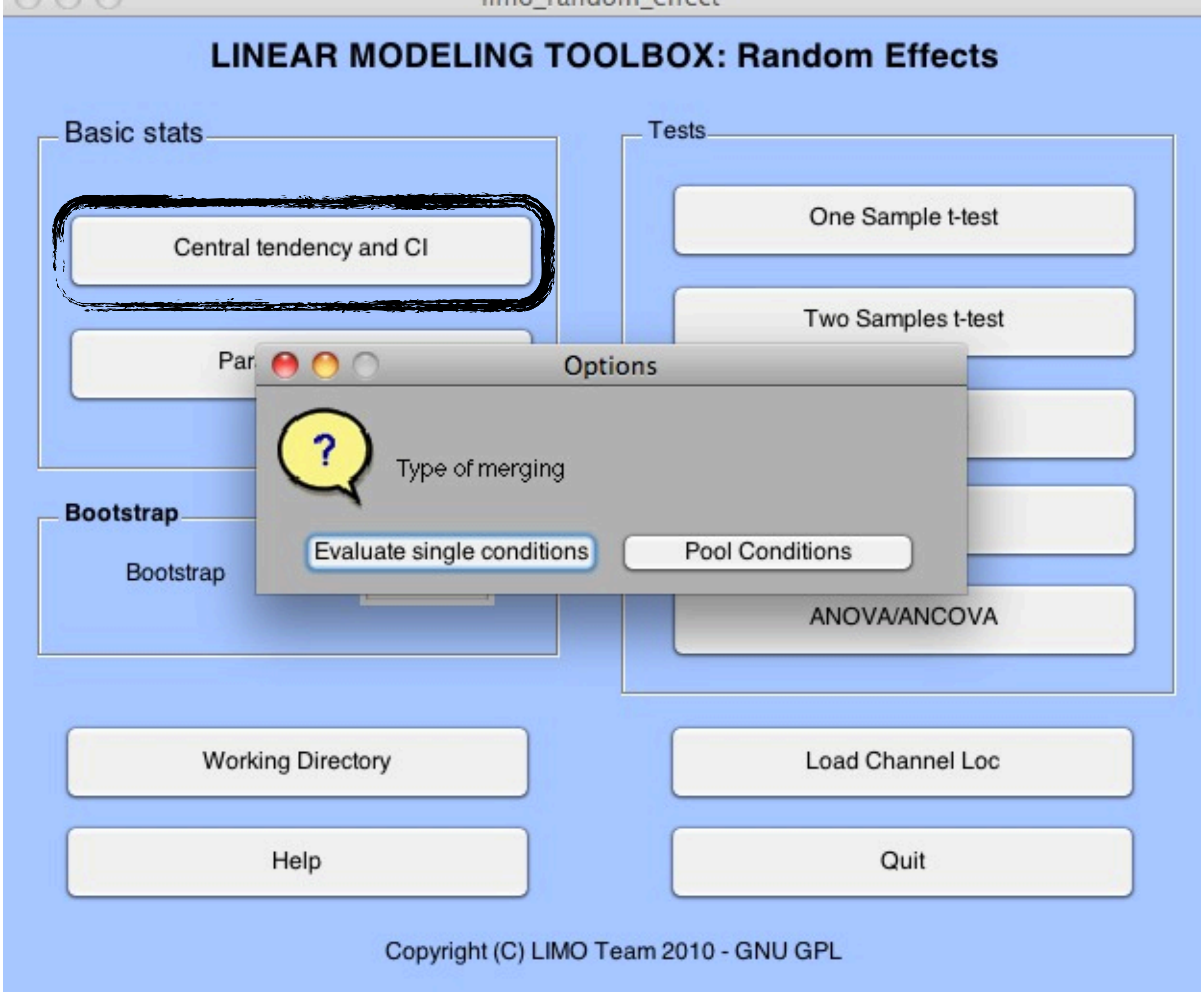


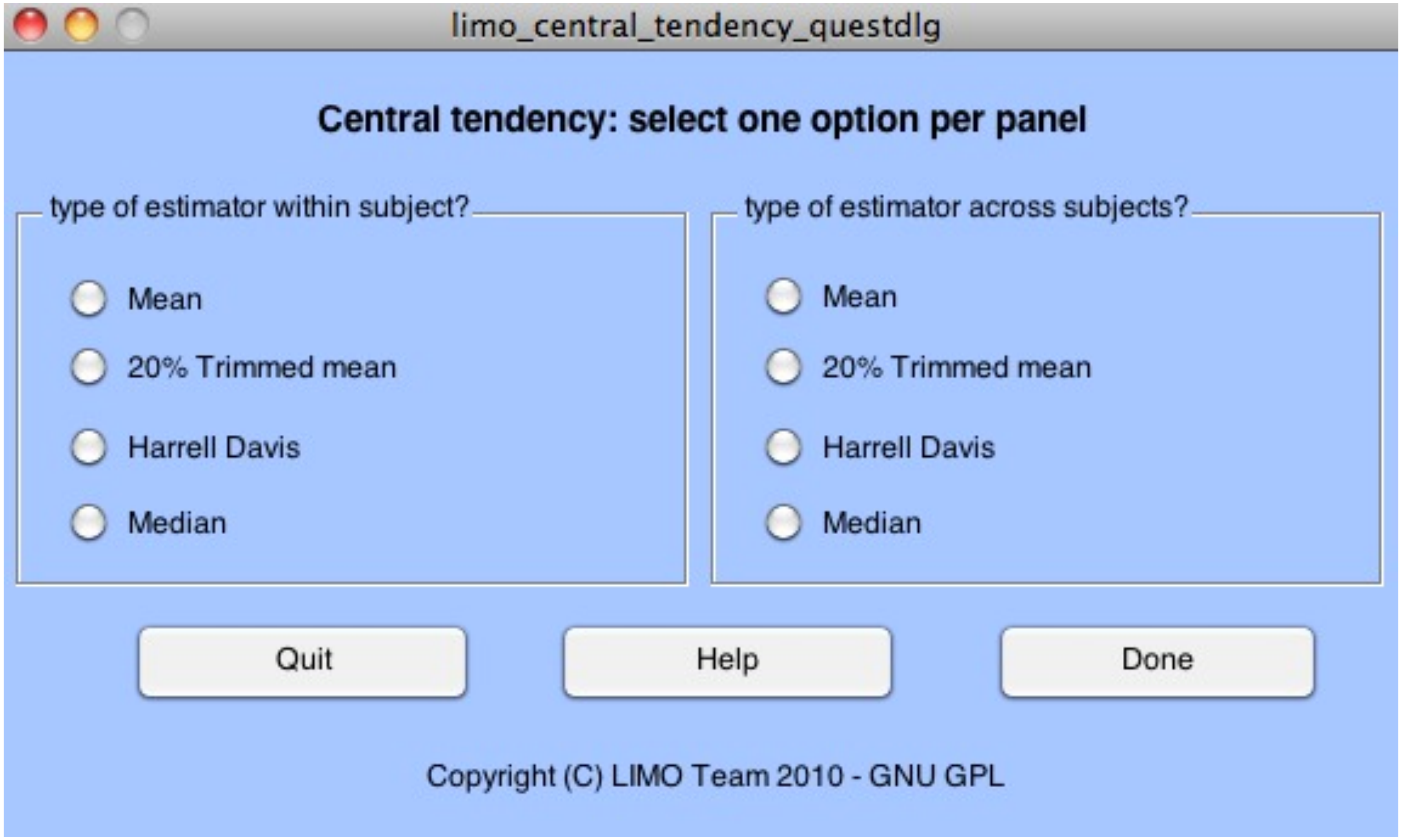

Modeled ERP at electrode B15 (40)
10
5
0
-5
-10
-15
-200 -100 0 100 200 300 400
x 10^-16
Model Residuals
5
0
-5
-200 -100 0 100 200 300 400
Time in ms


Adjusted ERP at electrode B15 (40)
15
10
5
0
-5
-10
-15
-200
-100
0
100
200
300
400
confounds
10
0
-10
-300
-200
-100
0
100
200
300
400
500
Time in ms

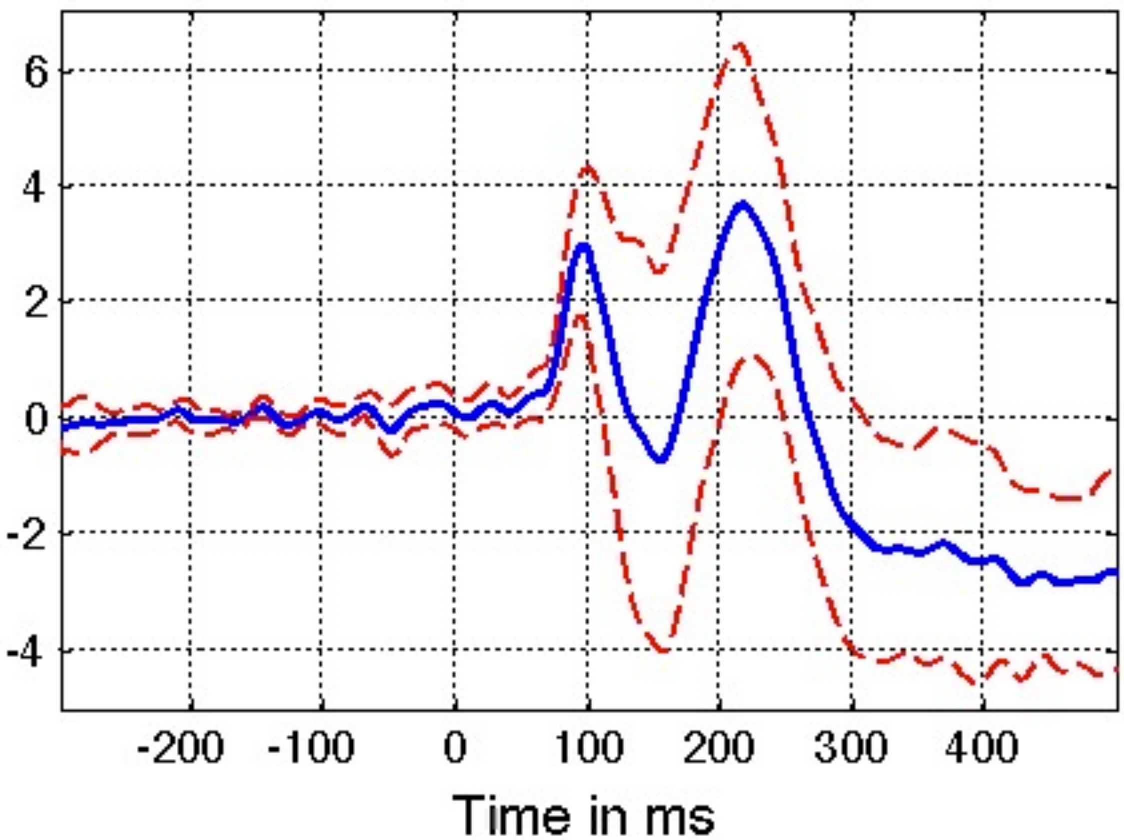

Modeled single trials
sorted by regressor 3 electrode B15 (40)
15
10
5
0
-5
-10
-15
1000
800
600
400
200
Sorted trials
-252 -200 -152 -100 -52 0 52 100 152 200 252 300 352 400 452
Time in ms

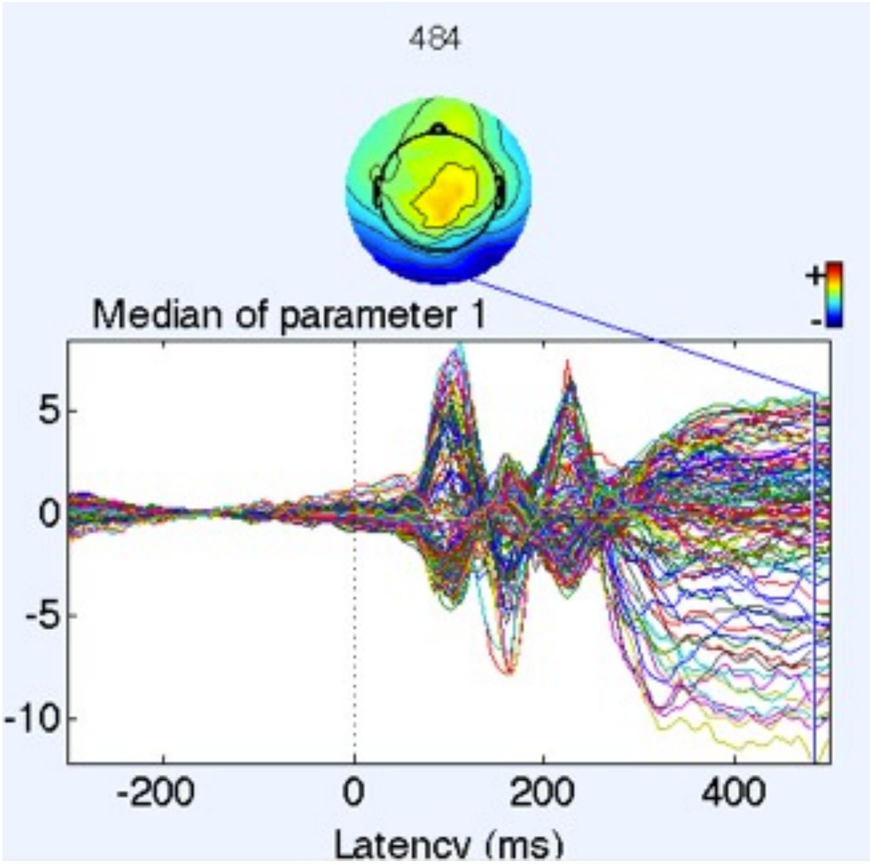

LINEAR MODELING TOOLBOX: Random Effects
Basic stats
Central tendency and CI
Parameter plots
Bootstrap
Bootstrap
1000
Tests
One Sample t-test
Two Samples t-test
Paired t-test
Regression
ANOVA/ANCOVA
Working Directory
Load Channel Loc
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL

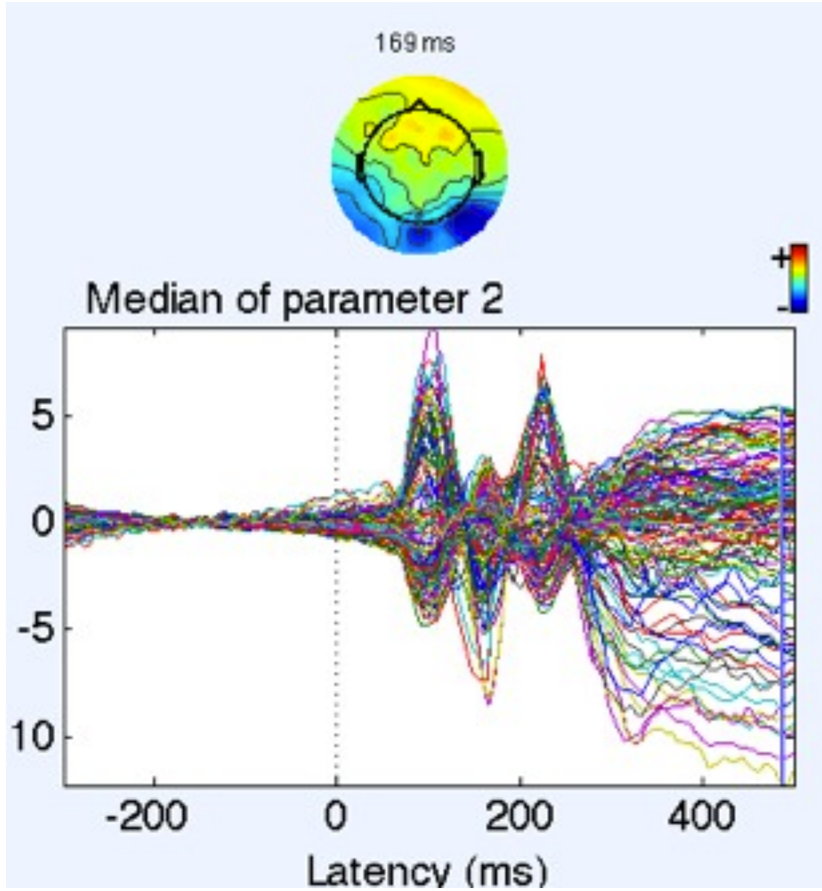

LINEAR MODELING TOOLBOX: Random Effects
Basic stats
Central tendency and CI
Par
Tests
One Sample t-test
Two Samples t-test
Options
Type of merging
Evaluate single conditions
Pool Conditions
Bootstrap
Bootstrap
ANOVA/ANCOVA
Working Directory
Load Channel Loc
Help
Quit
Copyright (C) LIMO Team 2010 - GNU GPL


limo_central_tendency_questdlg
Central tendency: select one option per panel
type of estimator within subject?
Mean
20% Trimmed mean
Harrell Davis
Median
type of estimator across subjects?
Mean
20% Trimmed mean
Harrell Davis
Median
Quit
Help
Done
Copyright (C) LIMO Team 2010 - GNU GPL

Mean parameter value and 95% CI at electrode B8 (40)




484
Median of parameter 1
+
-
5
0
-5
-10
-200
0
200
400
Latency (ms)


169 ms
Median of parameter 2
+
-
5
0
-5
10
-200
0
200
400
Latency (ms)

questions?