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

Bootstrapping and multiple comparison corrections in EEG

Cyril Pernet, PhD
Centre for Clinical Brain Sciences, University of Edinburgh

EEGLAB workshop - June 2019

@CyrilRPernet
cyril.pernet@ed.ac.uk

Overview

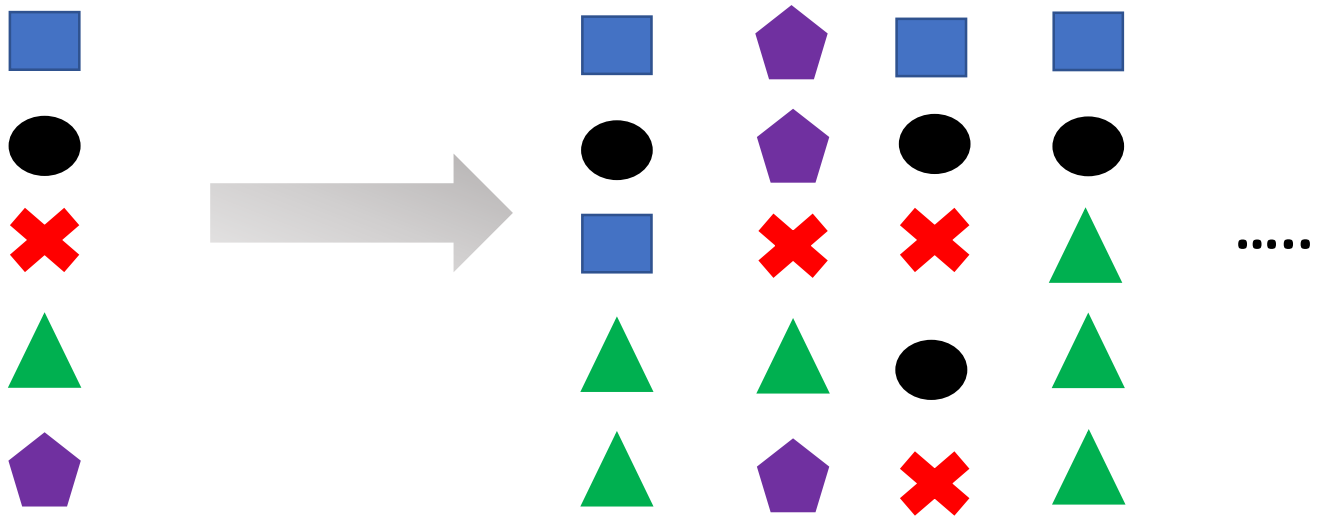
- The bootstrap idea (not robust, not assumption free)
- Highest Density Intervals for ERP (Bayesian bootstrap)
- Getting the null distribution using bootstrap (p-values)
- Multiple Comparison corrections in EEG
- Cluster mass and TFCE

Matlab code available to run as you listen !!

Bootstrap

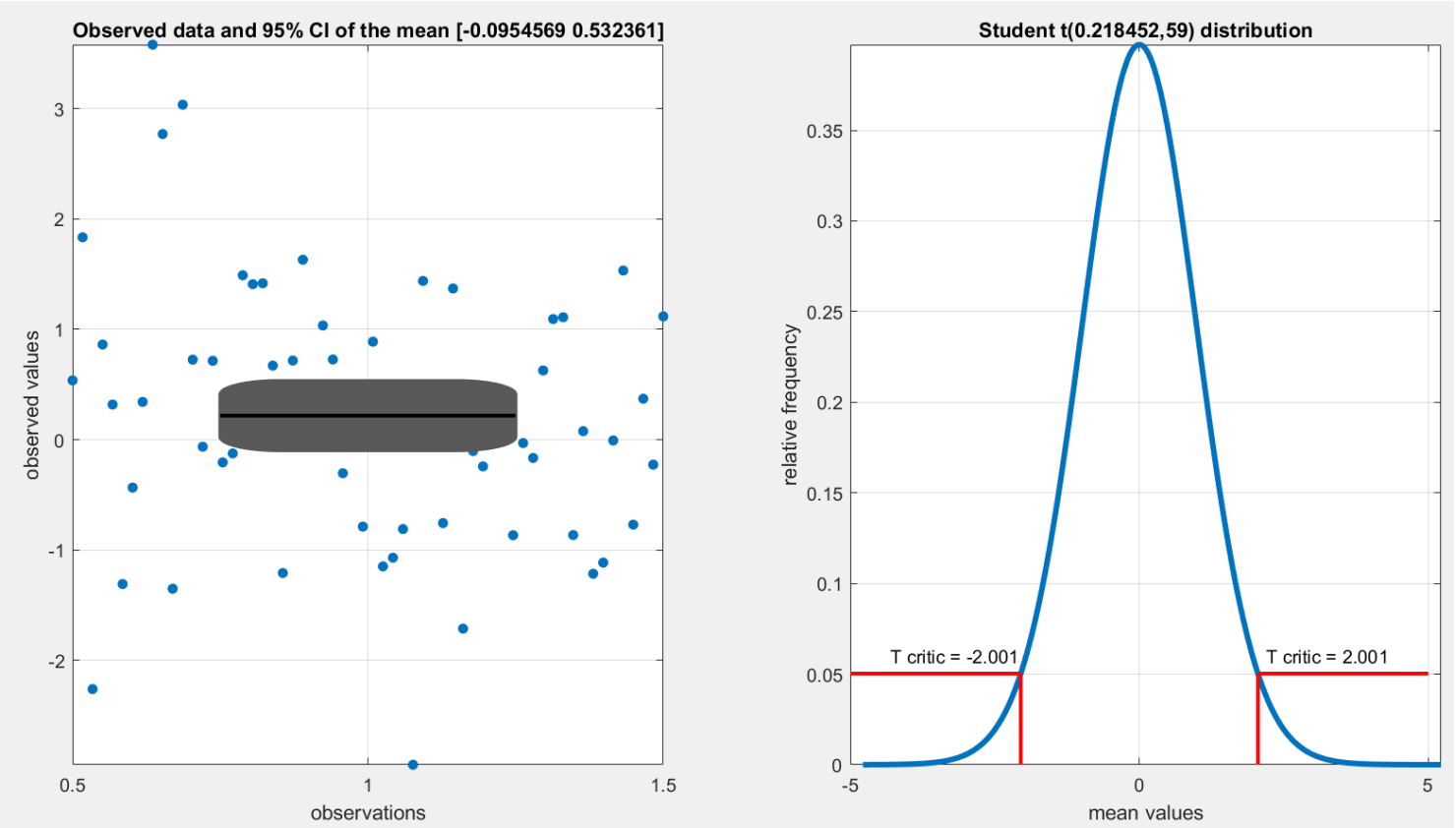
The bootstrap idea

- Make inferences using the data at hand only
- Avoid making assumptions about the underlying distribution, observation are coming from
- How: sample **with** replacement the data and compute



Let's compute a 95% Confidence interval

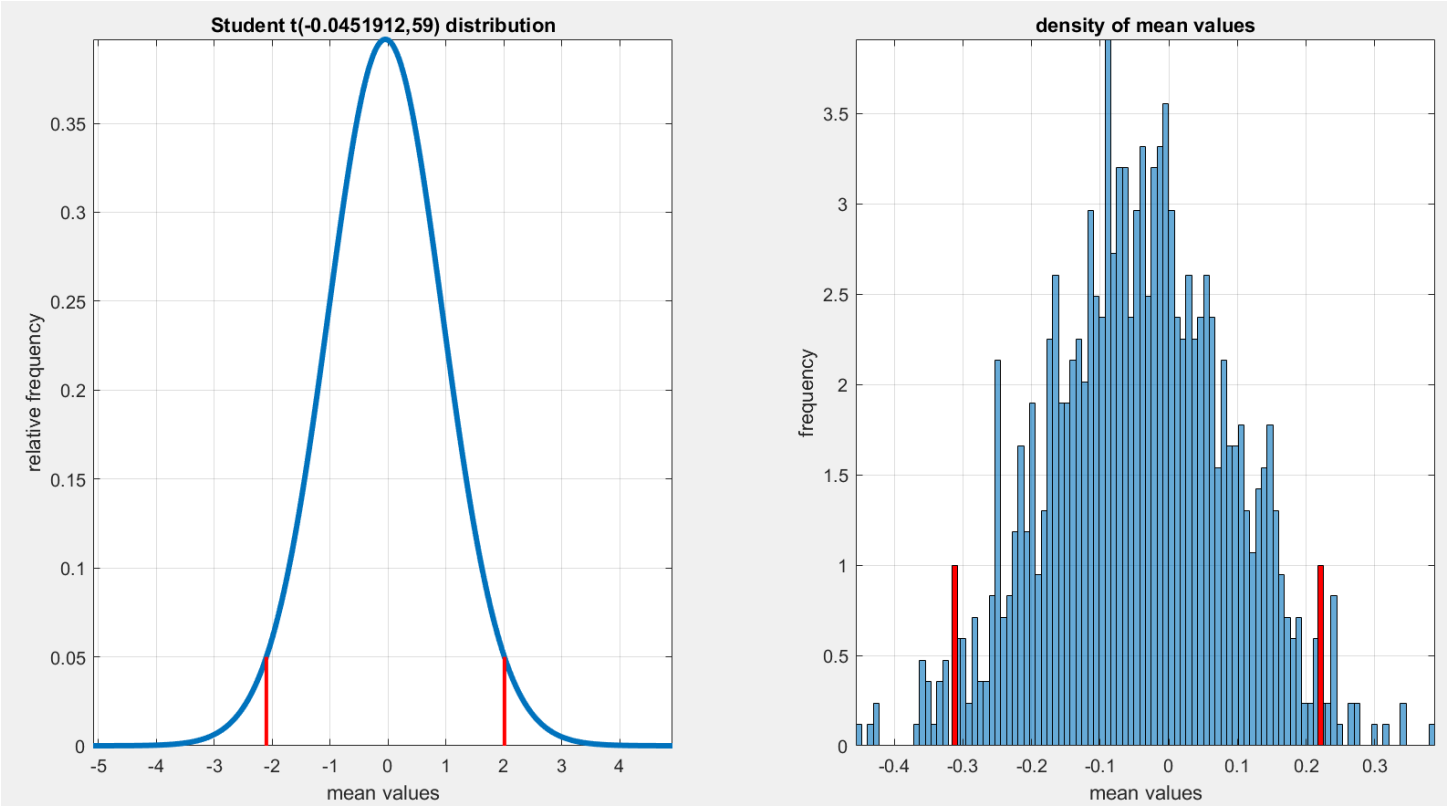
- CI = mean +/- Tcritic * standard error



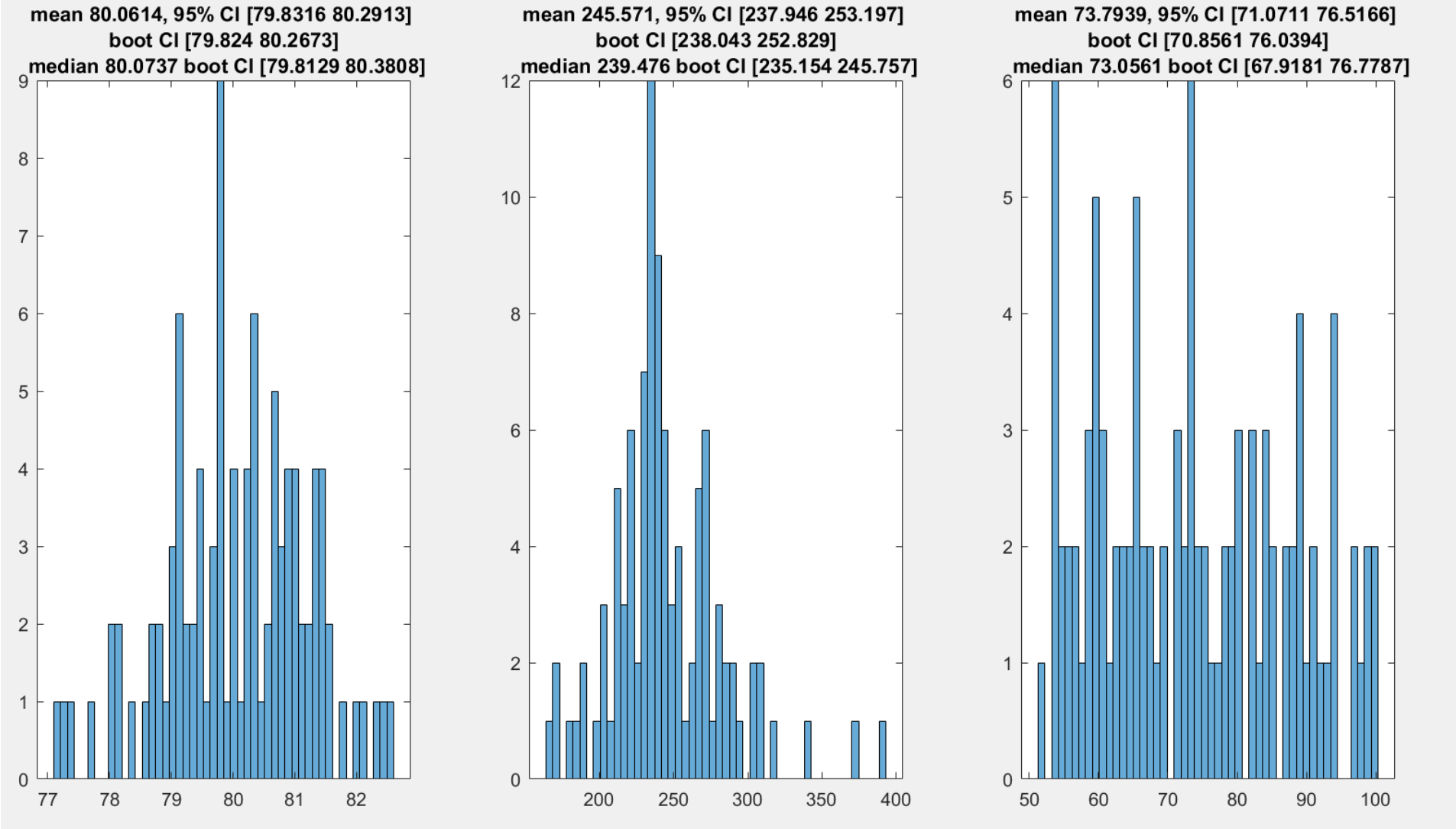
Assumes data are normally distributed and therefore the mean values of samples of size N come from Student t-distributions of N-1 df

Let’s compute a 95% Confidence interval

- Resample 1000 times observed data and plot the resampled means
- 95% CI is the 2.5/97.5 quantiles of this distribution

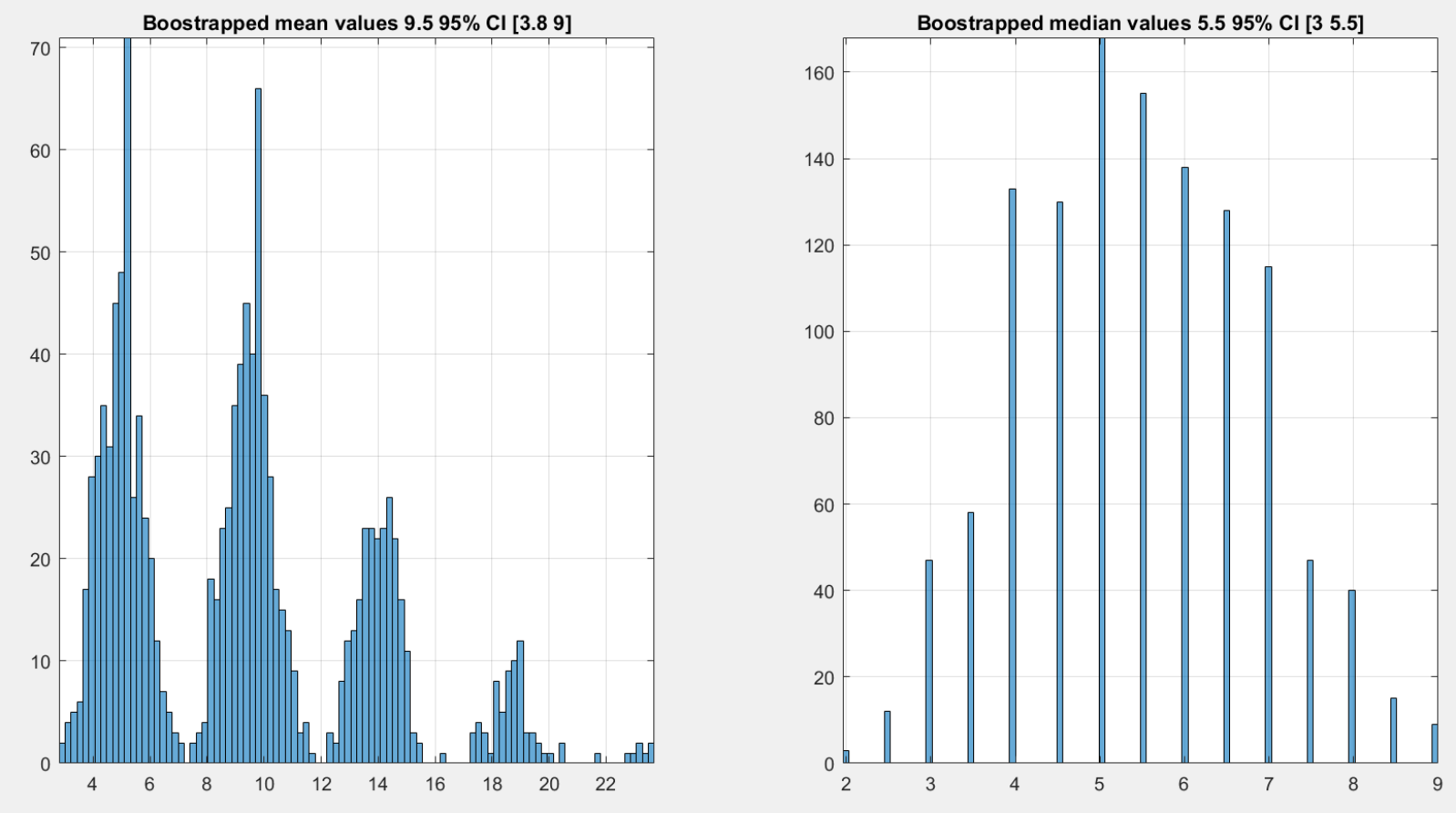


Let’s compare some 95% Confidence Intervals



Bootstrap does not bring robustness

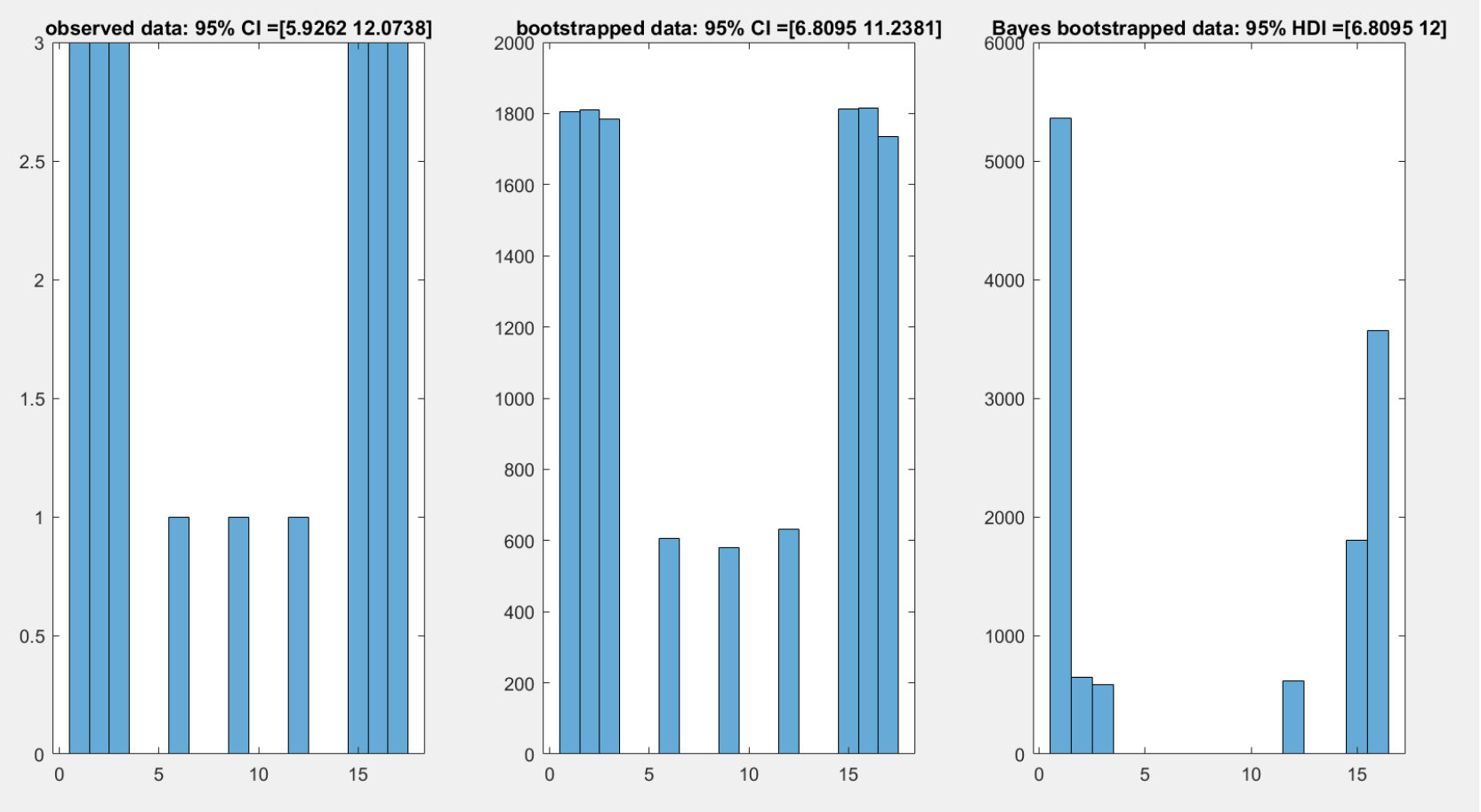
- Robust = resistant to outliers, e.g. [1 2 3 4 5 6 7 8 9 50]



The mean has a breakdown point of 0 while the median has a breakdown point of 50% (i.e. doesn't change up to 50% of outliers).

Bootstrap is not assumption free

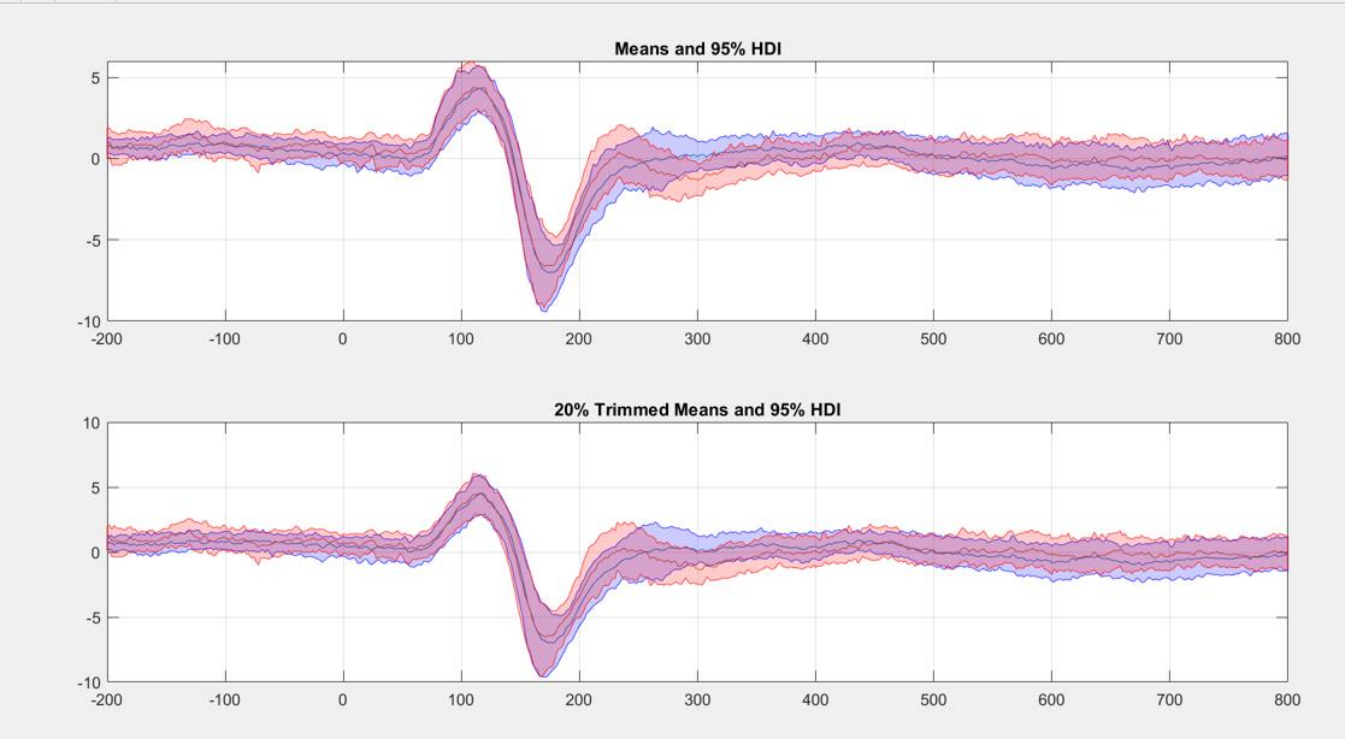
- Only observed value exist!



By making some assumptions about the data distribution (Bayesian priors) we can obtain intervals as if we sampled using missing values

Getting the mean ERP peak value

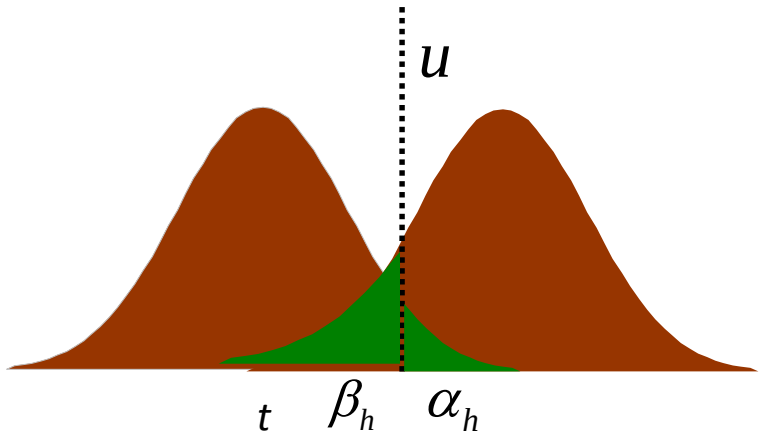
- Allows to get CI for any estimator (mean, trimmed mean, median)
- Non overlap of Bayesian CI (or HDI of the difference include 0) → you can accept the null (rather than cannot be rejected)

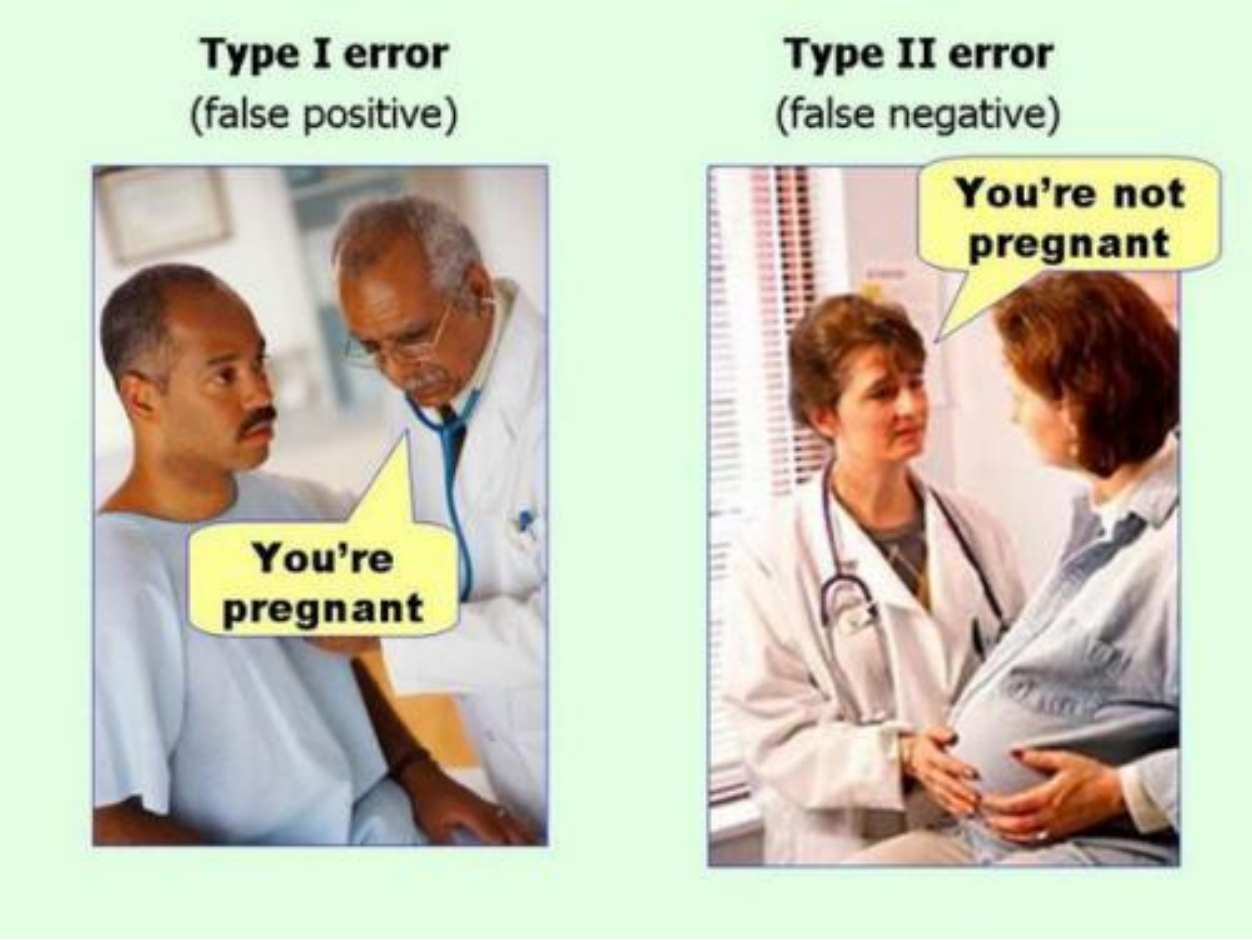


Multiple Comparisons Correction

Pearson-Newman hypothesis testing

- H0: no effect
- H1: there is an effect





https://chemicalstatistician.wordpress.com

Getting the null distribution (p-values)

- Is the trimmed mean different from 0?

1. Null hyp: data - trimmed mean
2. Resample the null
3. mean(trimmed mean> null means)
   or
   mean(t value > null t values)

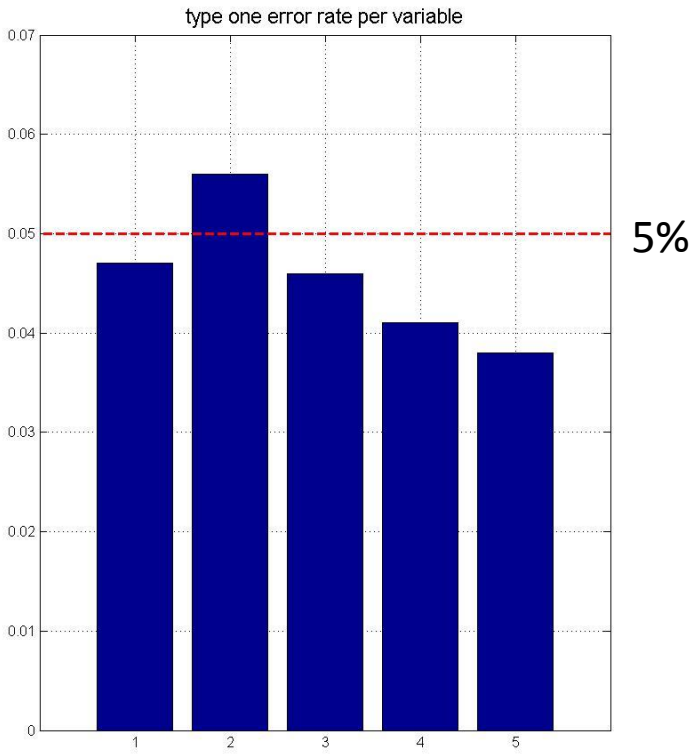
What is the problem?

- FWER is the probability of making one or more Type I errors (false positive) in a family of tests, under H0
- Assuming tests are independents from each other, the family-wise error rate FWER = 1 - (1 - alpha)^n

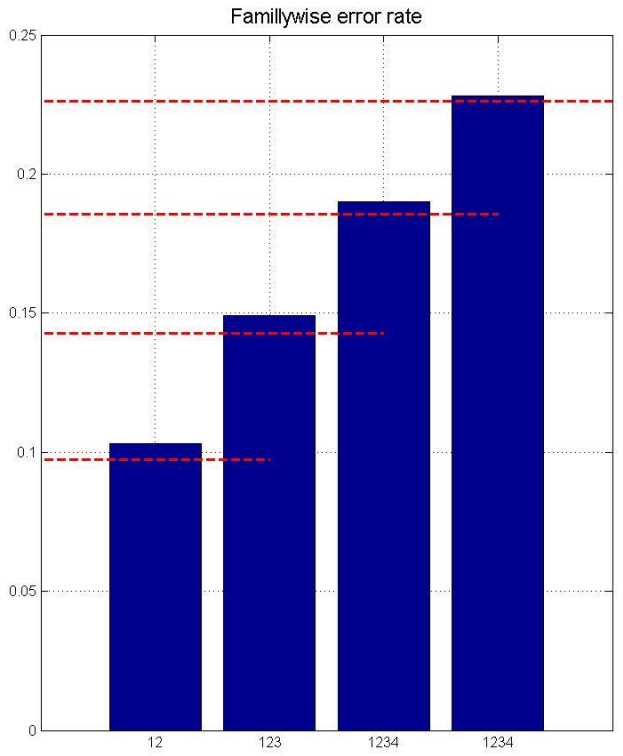
- for alpha =5/100, if we do 2 tests we should get about 1-(1-5/100)^2 ~ 9% false positives, if we do 126 electrodes * 150 time frames tests, we should get about 1-(1-5/100)^18900 ~ 100% false positives! i.e. you can’t be certain of any of the statistical results you observe

What is the problem?

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



5%



22%

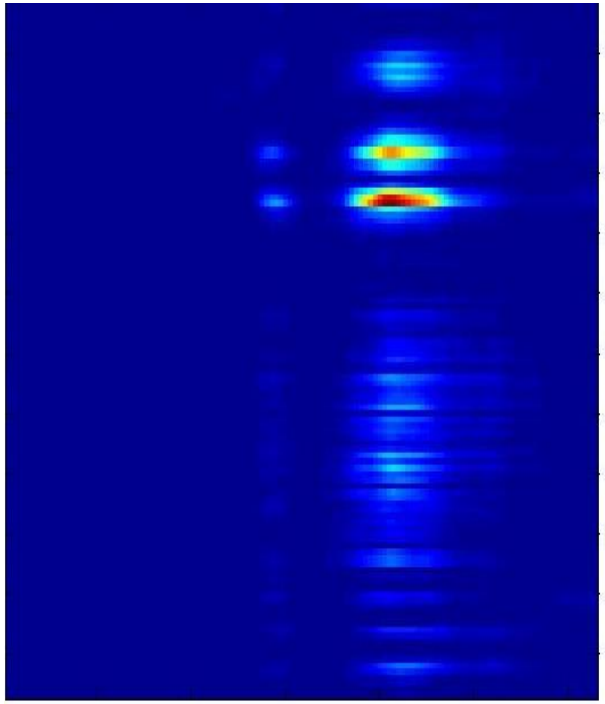
18%

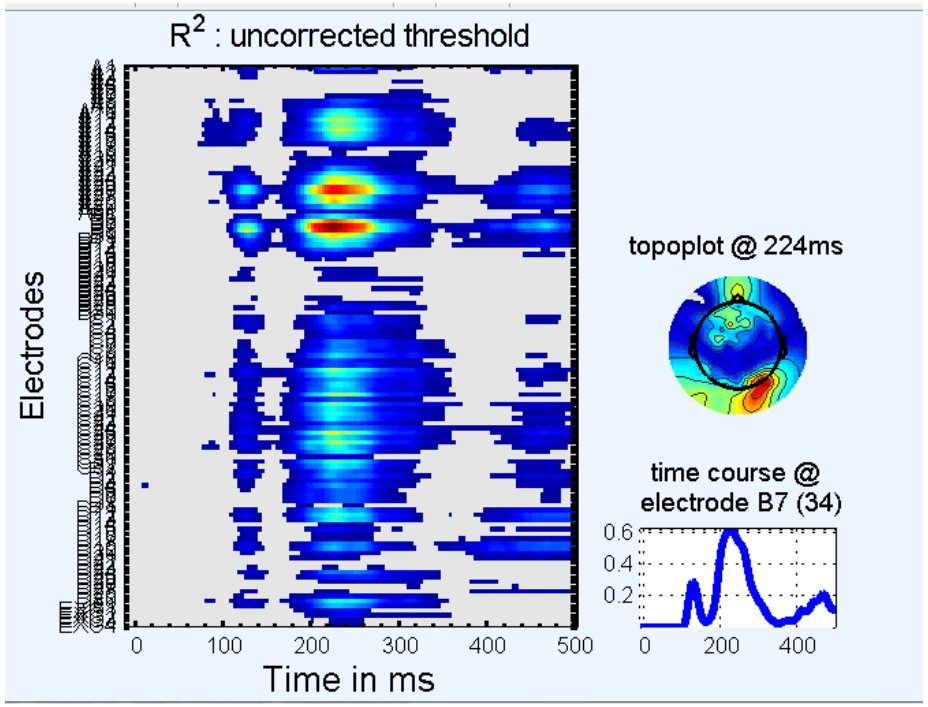
14%

9%

What is the problem?

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

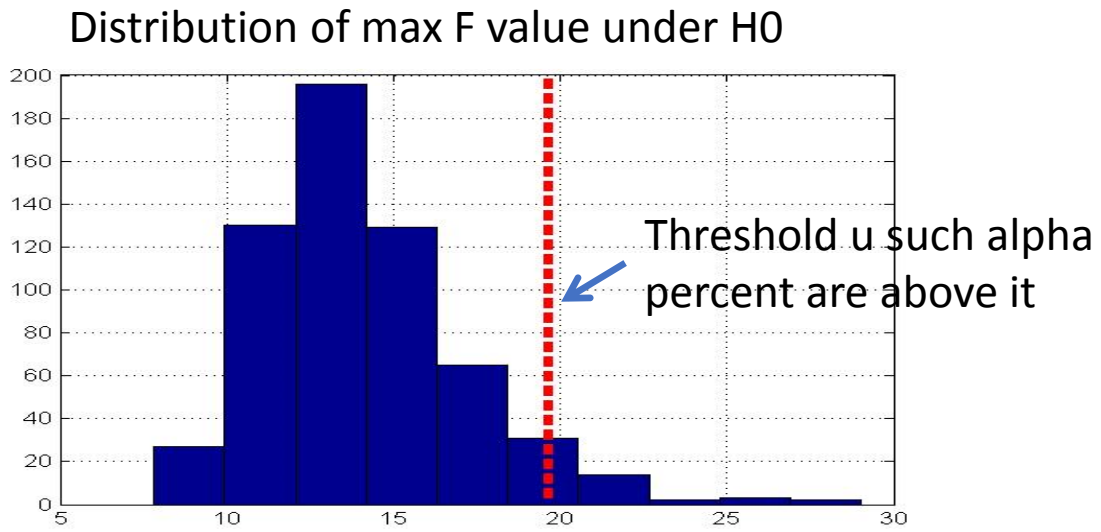




we know there are false positives – which ones is it?

FWER

- Since the type 1 FWER is the prob that any stats > u, then it 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.



Bonferroni Correction

Bonferroni correction allows to keep the FWER at 5% by simply dividing alpha by the number of tests – it find the threshold u

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

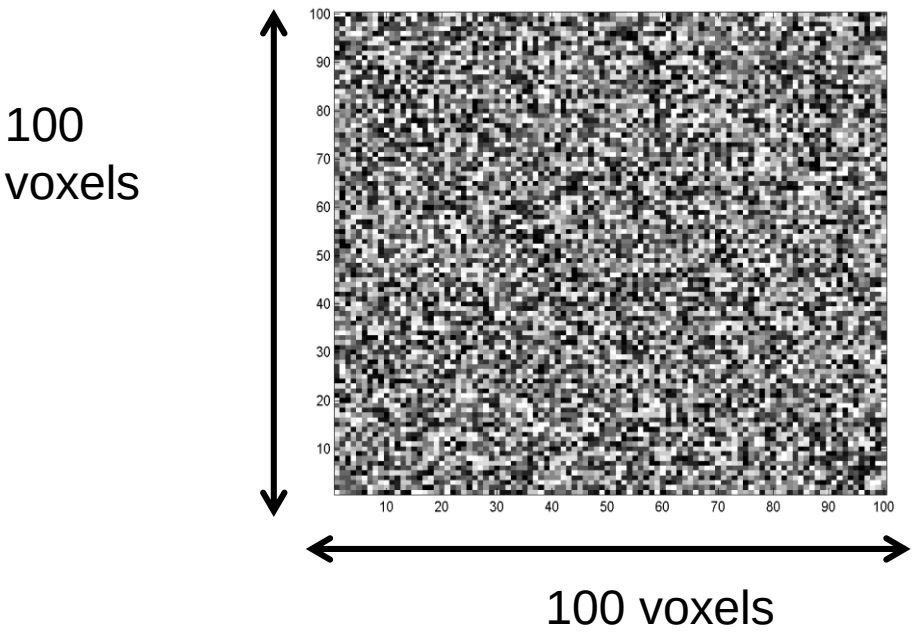
$$\text{FWER} = P(\cup_{i \in V}\{T_i \geq u\}|H_0) \leq \propto$$

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

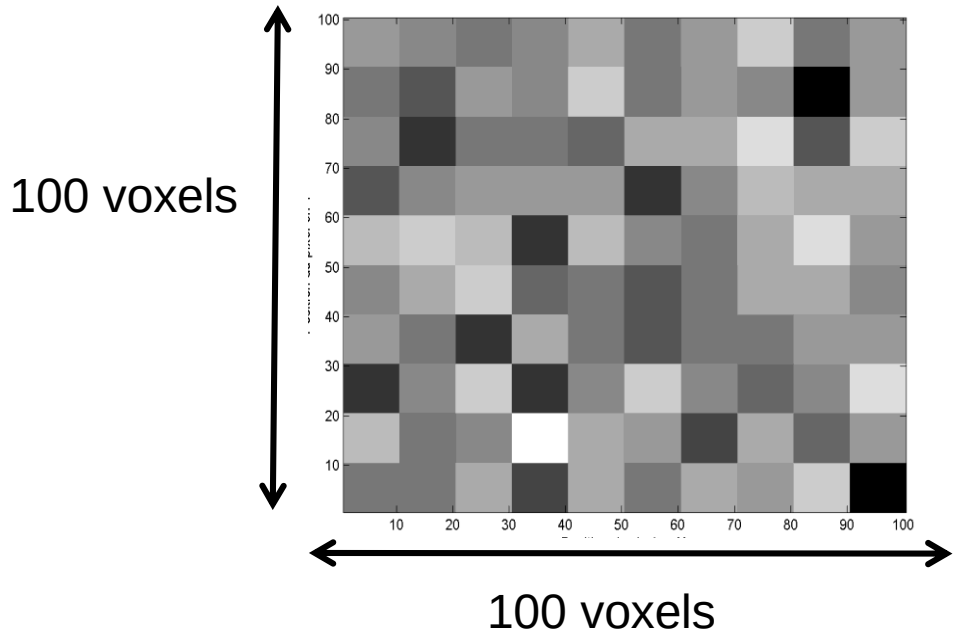
$$\leq \sum_i \frac{\propto}{m} = \propto$$

Bonferroni Correction

- 10000 Z-scores ; alpha = 5%
- alpha corrected = .000005
- z-score = 4.42

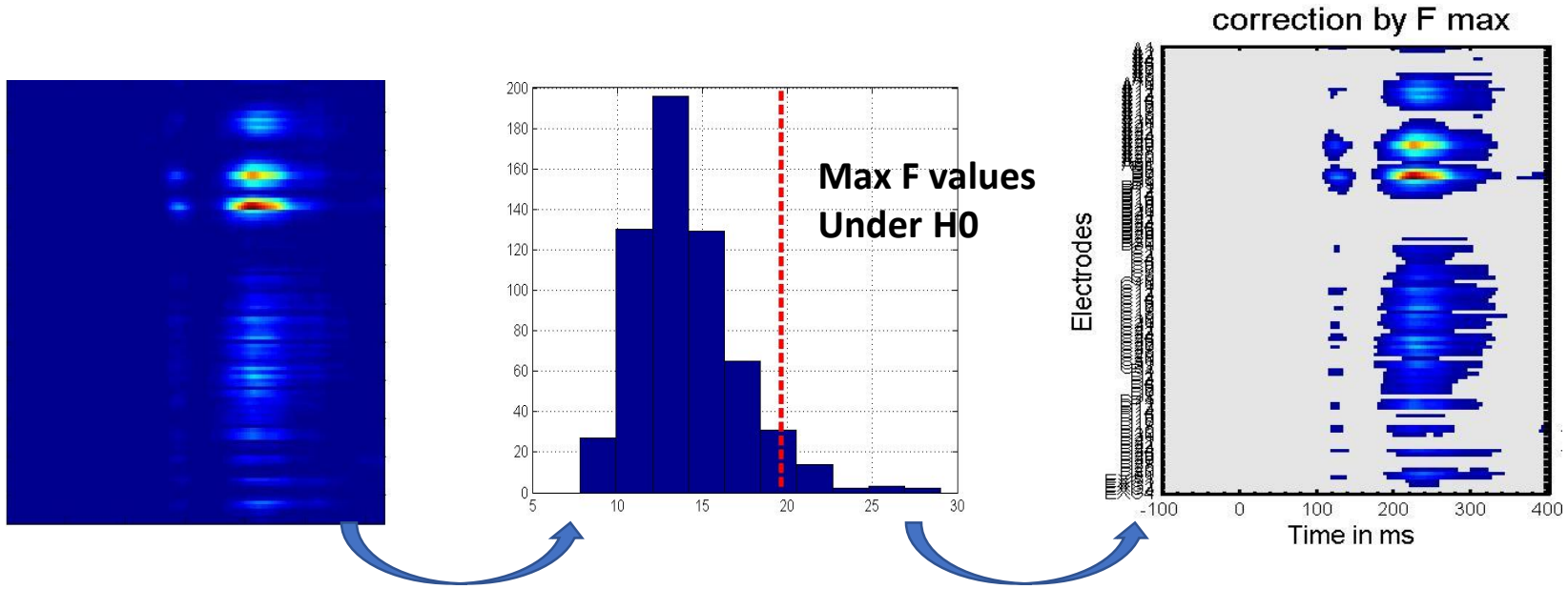


- 2D homogeneous smoothing – 100 independent observations
- alpha corrected = .0005
- z-score = 3.29



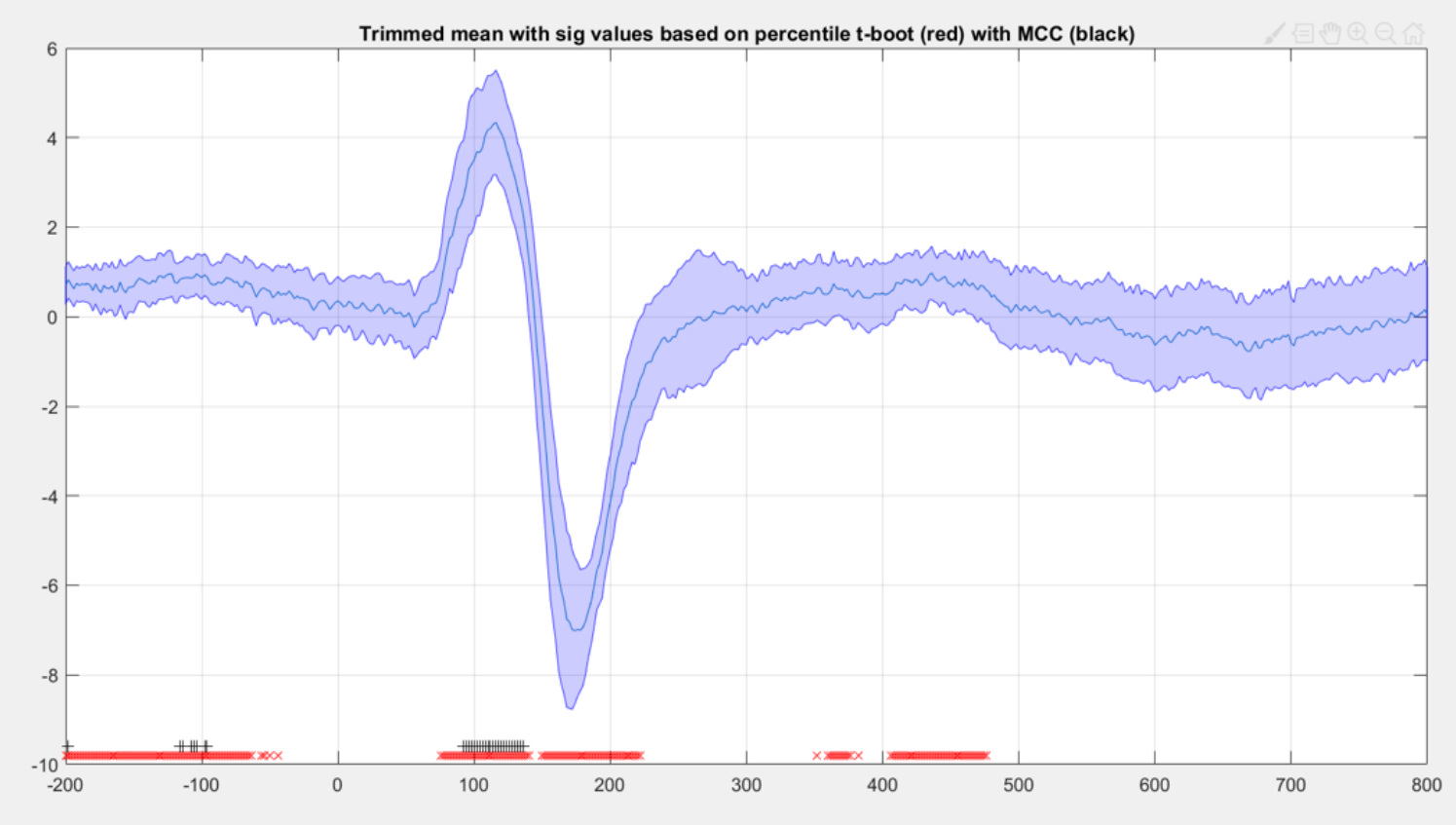
Maximum Statistics based on resampling

- Estimate the distribution of max under H0 (bootstrap/permutation) and simply threshold the observed results a threshold u like Bonferroni
- Accounts inherently for smoothness but still assumes all tests are independent



Maximum Statistics based on resampling

- Estimate the distribution of max under H0 (bootstrap/permutation) and simply threshold the observed results a threshold u like Bonferroni



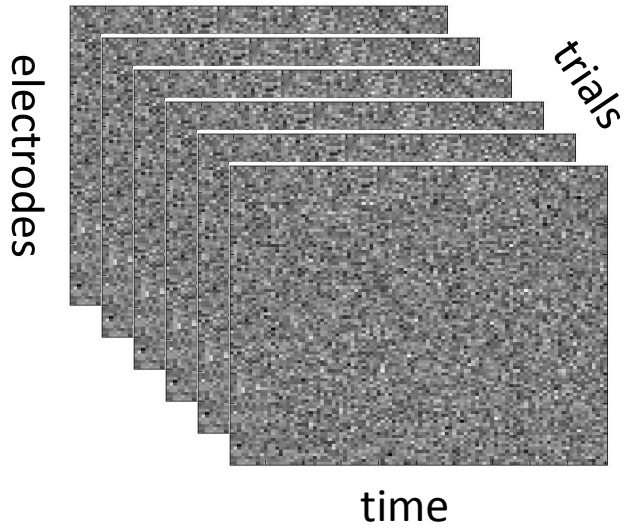
Solutions for imaging data

- An important feature of neuroimaging data is that we have a family of stat values that has topological features (Bonferroni for instance consider tests as independent) and we can thus considering data as a smooth lattice, i.e. based our inference on clusters
- fMRI/PET are projection methods of data points onto the whole space – MEEG forms continuous functions in time and are smooth by the scalp (space)
- Neural activity propagate locally through intrinsic/lateral connections and is distributed via extrinsic connections / Hemodynamic correlates are initiated by diffusing signals (e.g. NO)

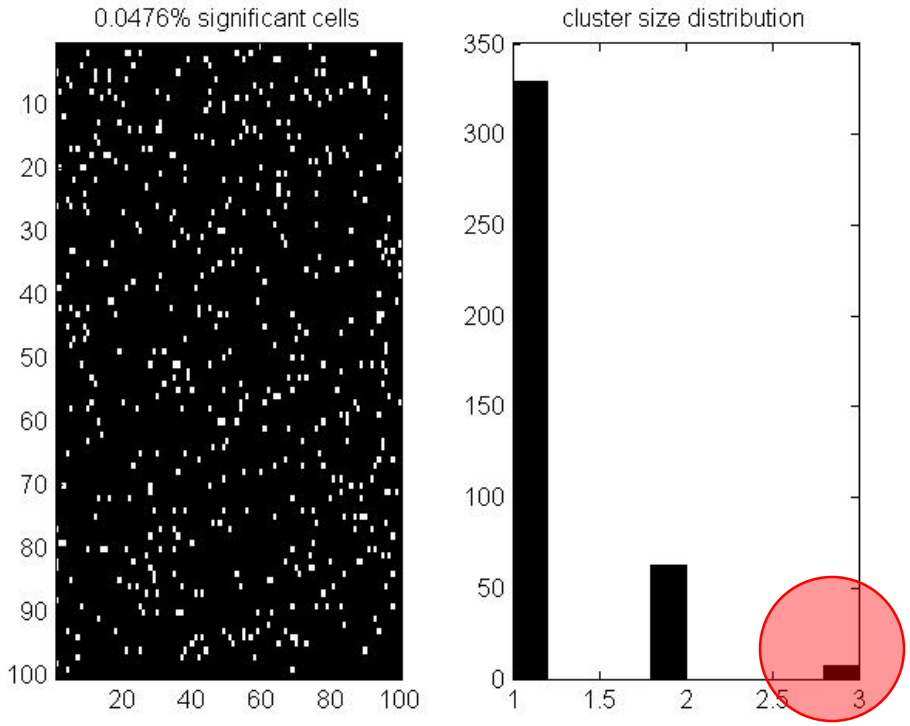
Chumbley et al., 2009 NeuroImage 44

Let’s analyse clusters

- Instead of the max, we consider clusters as it is much less likely that statistics are significant in groups

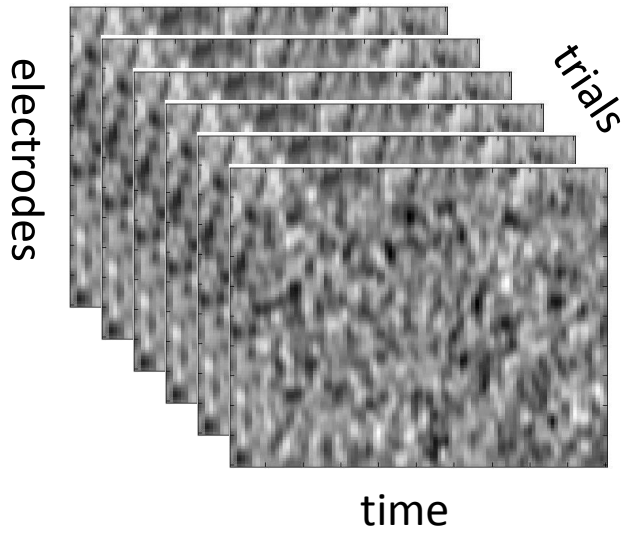


One sample t test > 0 ?

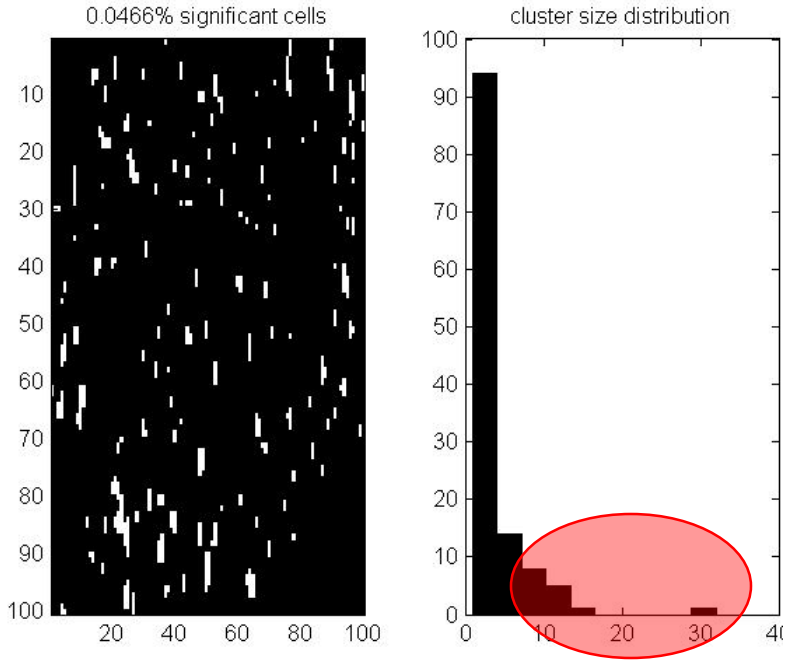


Let’s analyse clusters

- Instead of the max, we consider clusters as it is much less likely that statistics are significant in groups because data are smooth in space and time!



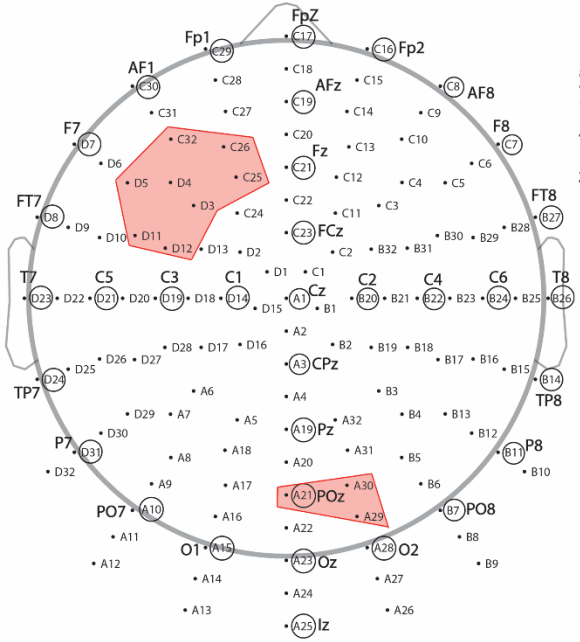
One sample t test > 0 ?

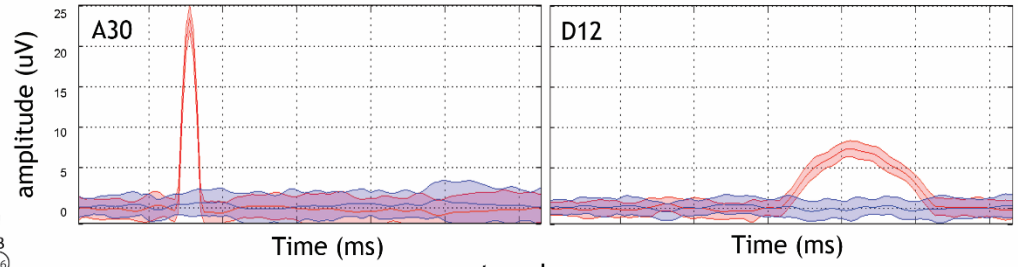


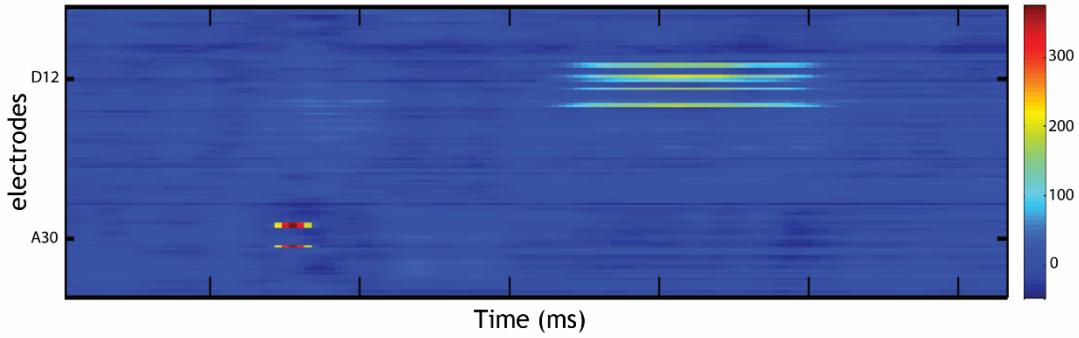
The clustering solution

- 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.

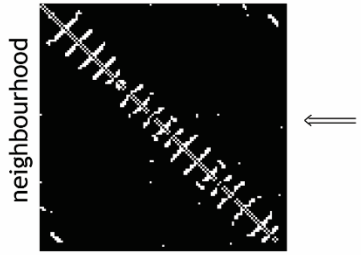
- To use clustering we need to consider cluster statistics rather than individual statistics
- Cluster statistics depend on (i) the cluster size, which depends on the data at hand (how correlated data are in space and in time/frequency), and (ii) the strength of the signal (how strong are the t, F values in a cluster) or (iii) a combination of both.





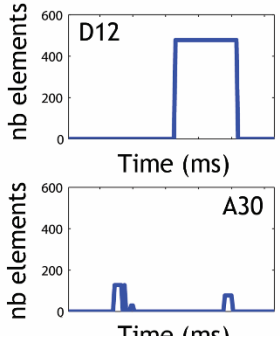


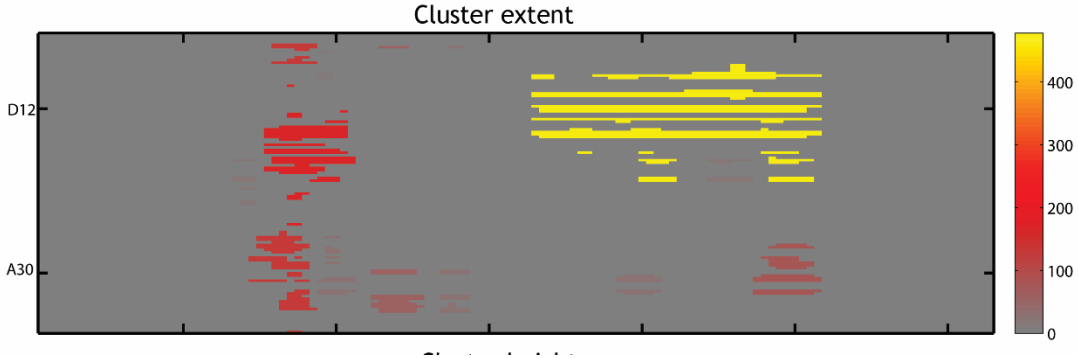
Spatial - Temporal clustering

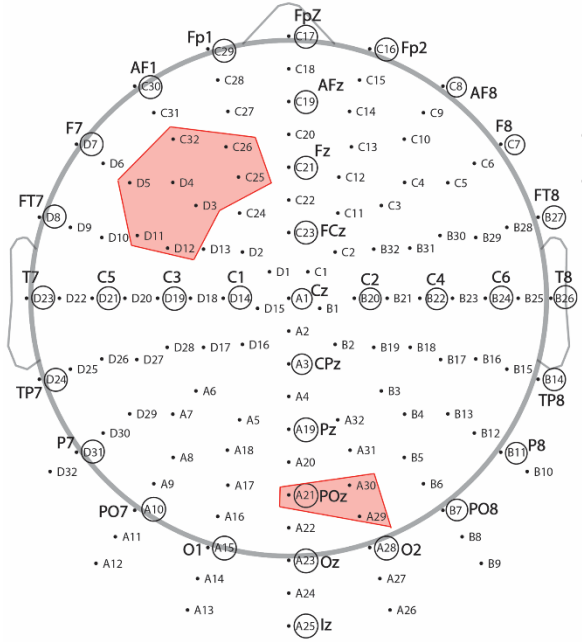


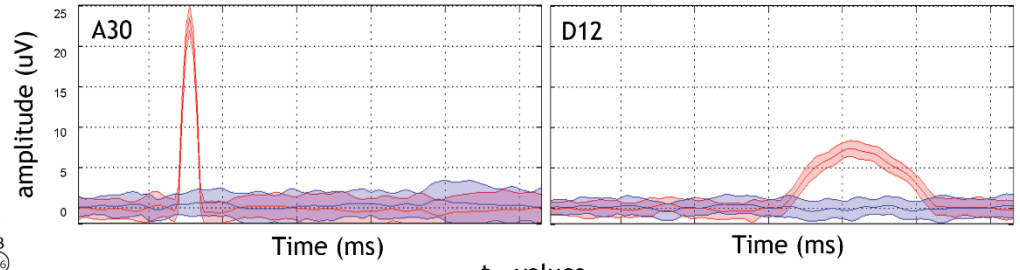
maximum extent
= number of
electrodes and
time points

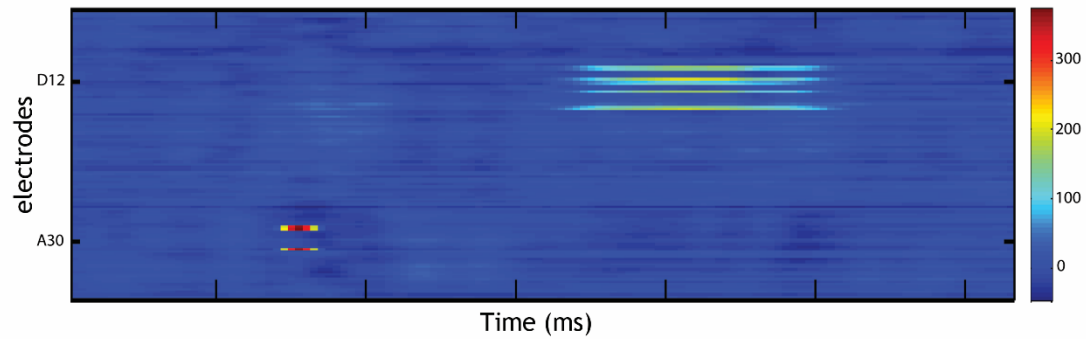
cluster 1 = 478
cluster 2 = 127



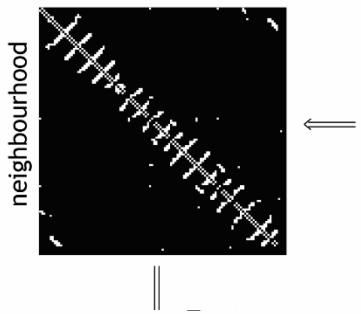


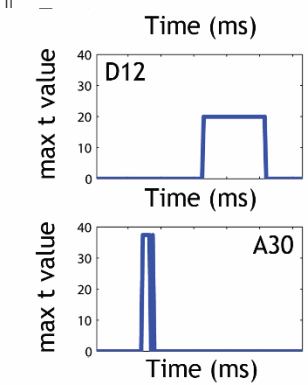






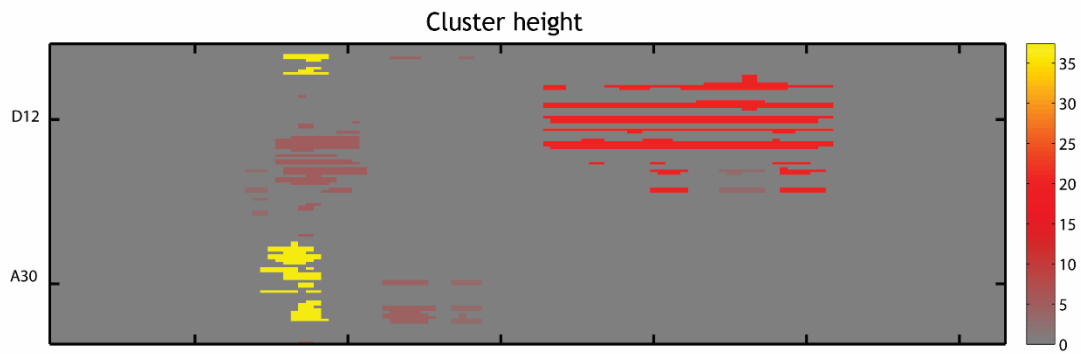
Spatial - Temporal clustering

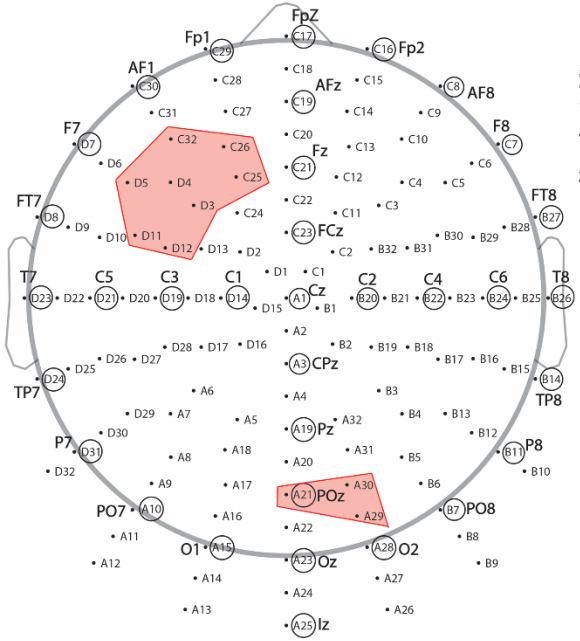


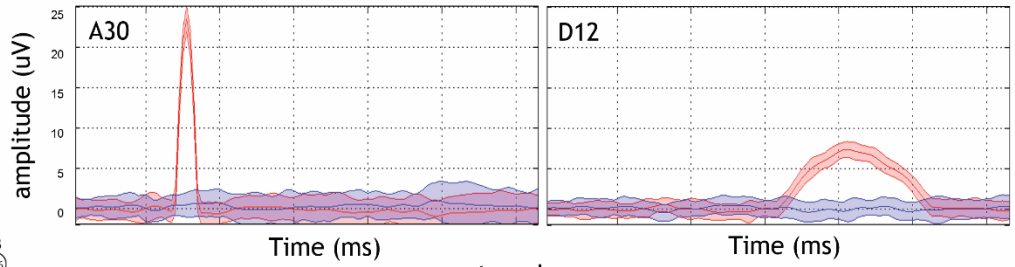


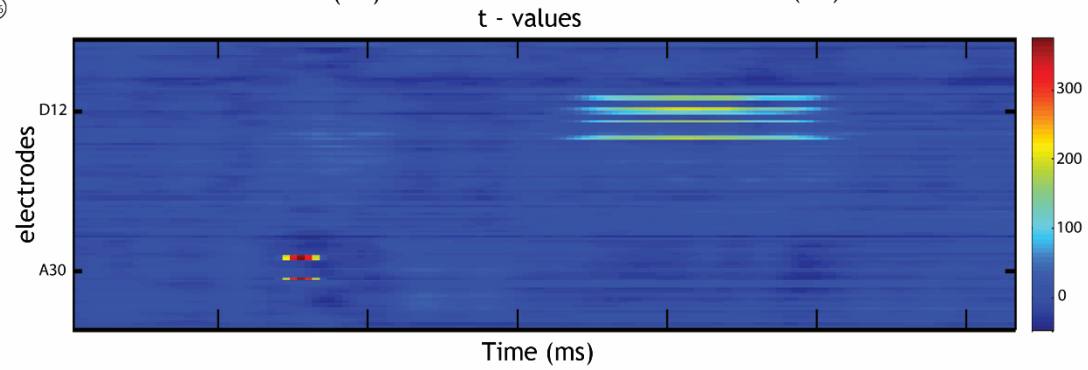
maximum height
within a cluster
of electrodes
and time points

cluster 1 = 19.7
cluster 2 = 37.4

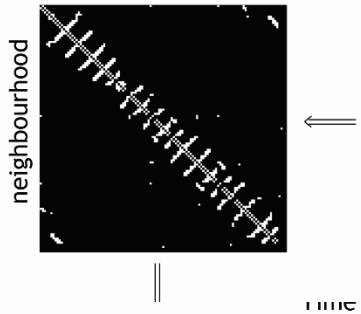






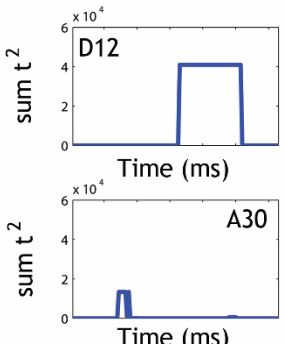


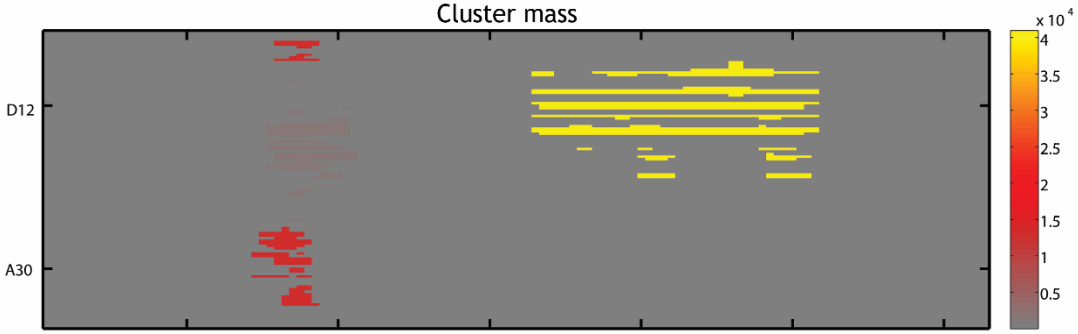
Spatial - Temporal clustering



mass (sum t^2) of values within a cluster of electrodes and time points

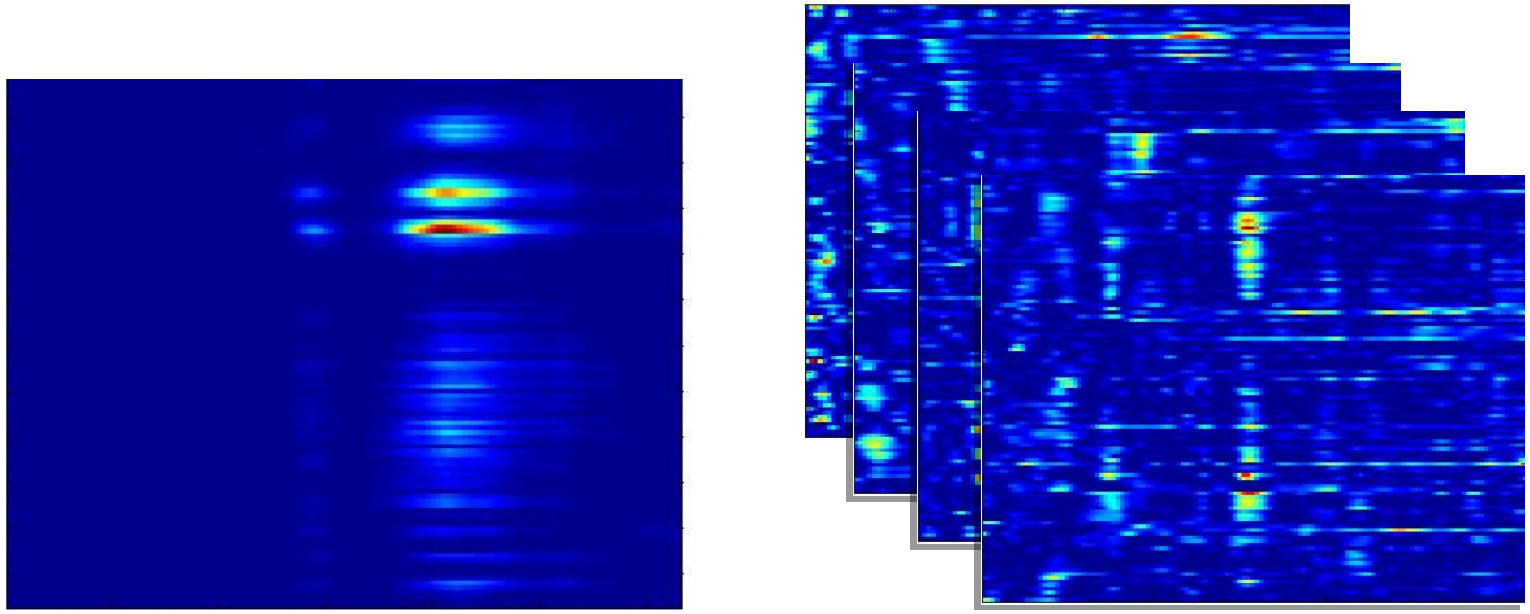
cluster 1 = 40984
cluster 2 = 13386





The clustering solution

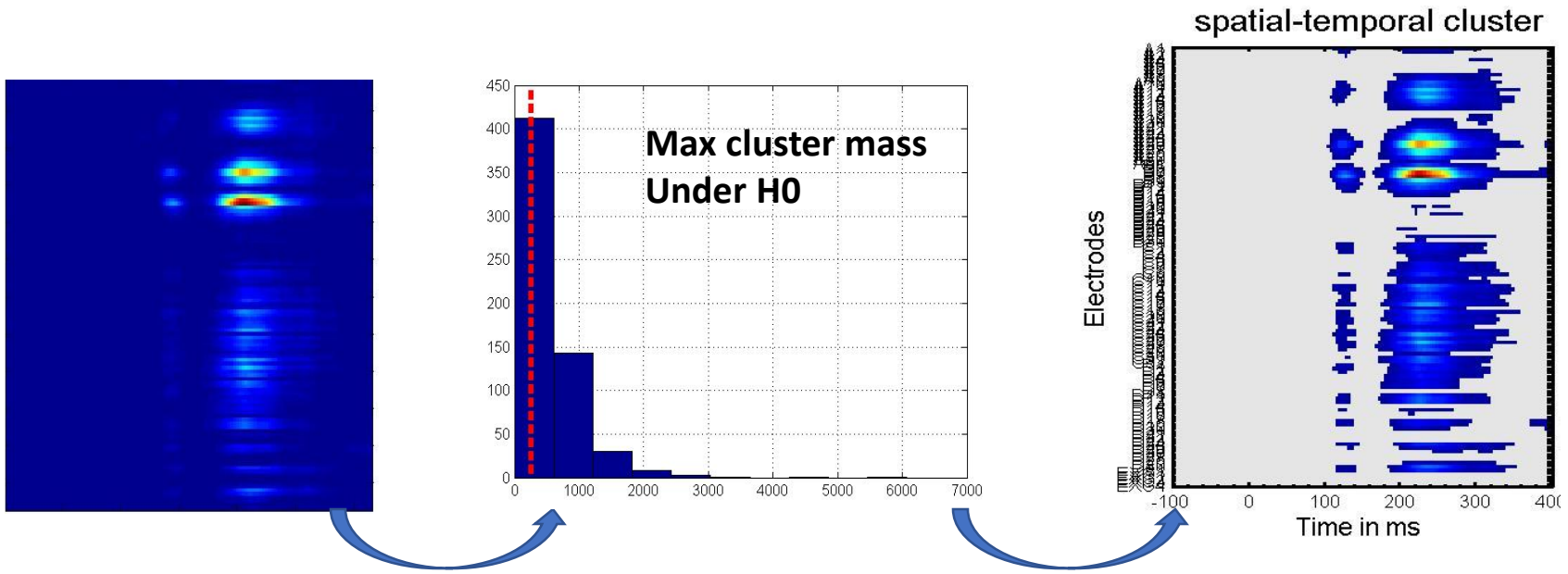
- In LIMO EEG, we bootstrap the data under H0: center the data or break the link between the design matrix and the data and then resample and test. This way we can find u for a single bin, the the whole space, or for clusters.



Observed F values → F values under H0

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 !

TFCE for MEEG

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.

Smith & Nichols 2009 NeuroImage 44

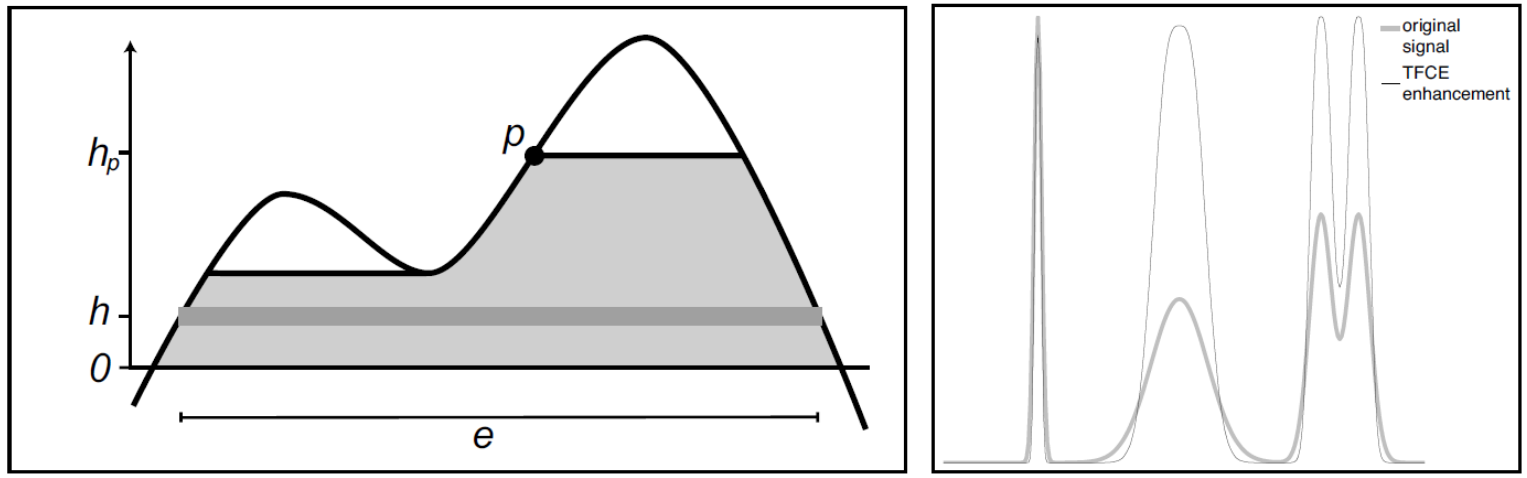
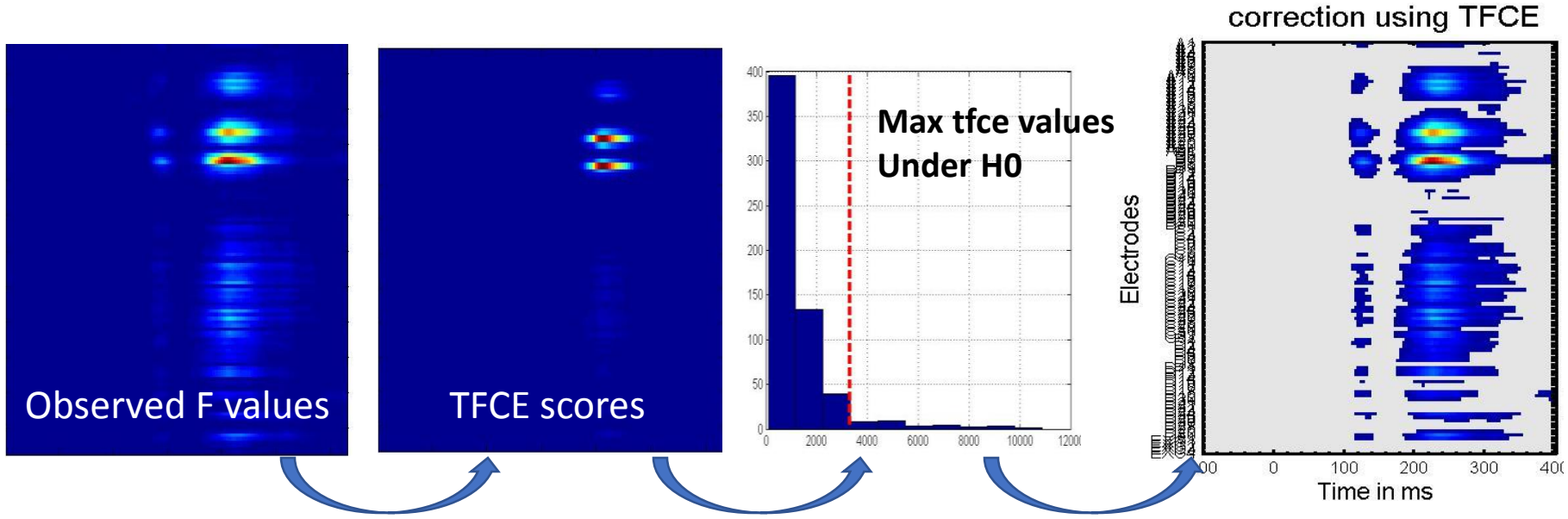


Figure 1: Illustration of the TFCE approach. Left: The TFCE score at voxel p is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark grey band) within the area of "support" of p (light grey). The score for each section is a simple function of its height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

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

MCC summary

- Simulation work show that overall permutation / bootstrap / cluster-mass / TFCE control well the type 1 FWER.
- a minimum of 800 iterations are necessary to obtain stable results
- for low critical family-wise error rates (e.g. p = 1%), permutations can be too liberal;
- For within subject bootstrap, a min of 50 trials per condition is requested at the risk to be too conservative

Conclusions

- When performing multiple tests, statistical correction MUST be applied.

- All techniques provide a FWER at the specified level but not all techniques have the same power.
- Spatial-temporal clustering and TFCE seem to provide good estimates, with TFCE giving higher spatio-temporal inference resolution, but at the cost of long computing time.

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