LImear Modelling of EEG data

Multiple Comparison Correction

Cyril R. Pernet, University of Edinburgh
Guillaume, A. Rousselet, University of Glasgow

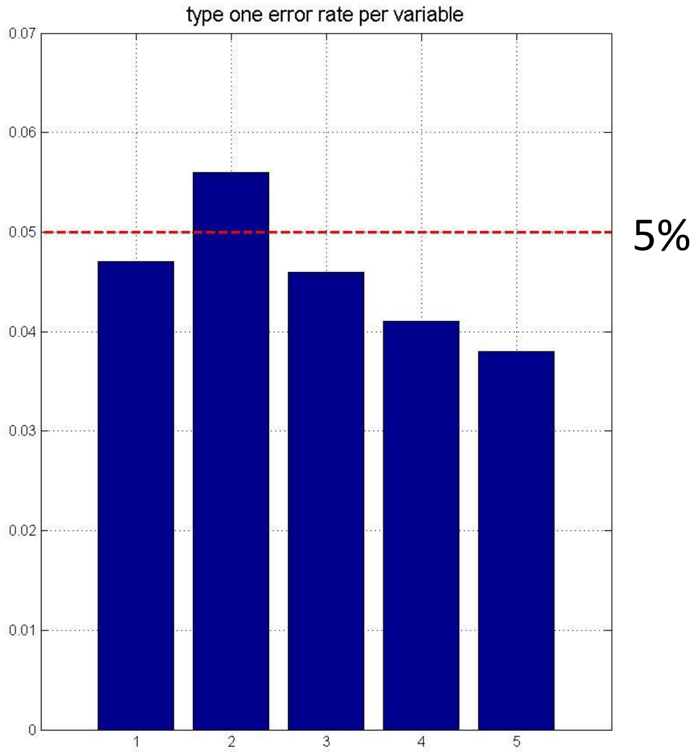
What is the problem?

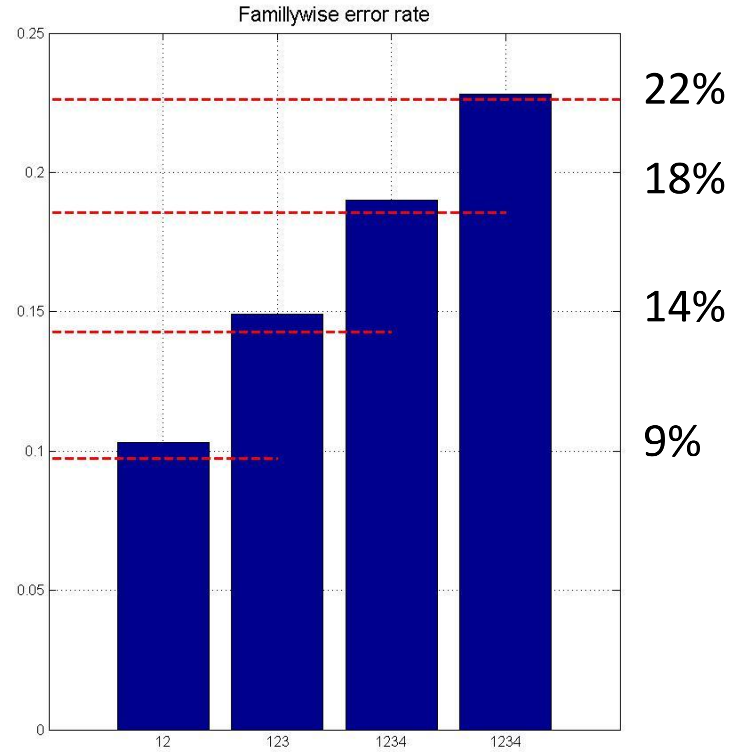
- When we do a statistical test, we set alpha, the probability to reject H0 (under H0) – this is also known as type I error rate

- 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

What is the problem?

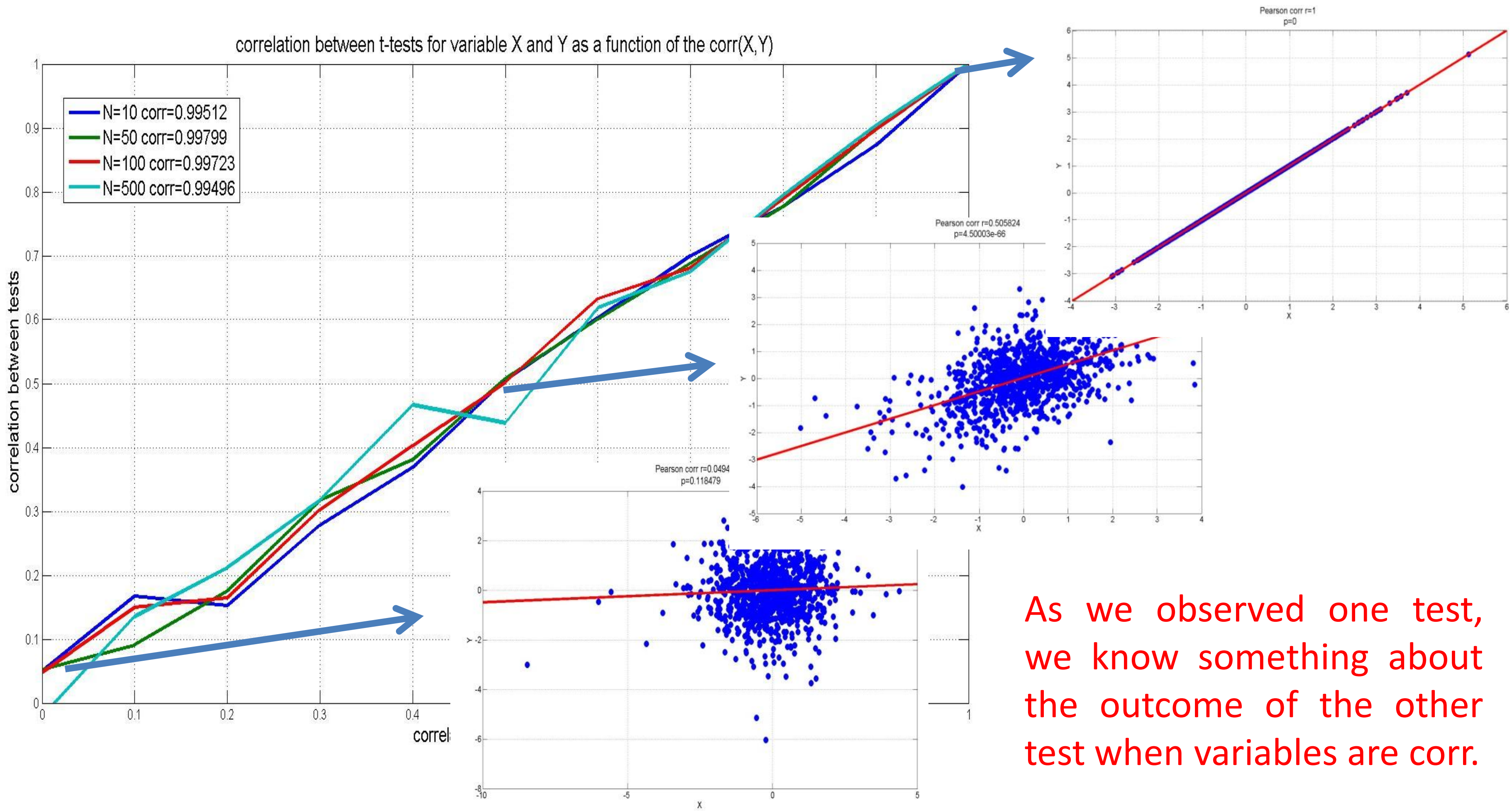
- Illustration with 5 independent variables from N(0,1)
- Repeat 1000 times and measures type 1 error rate





What is the problem?

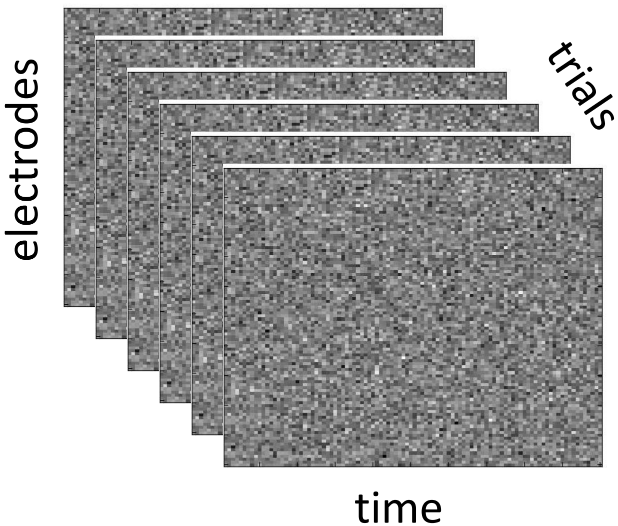
- Illustration with 2 variables with Pearson‘s r=0:1



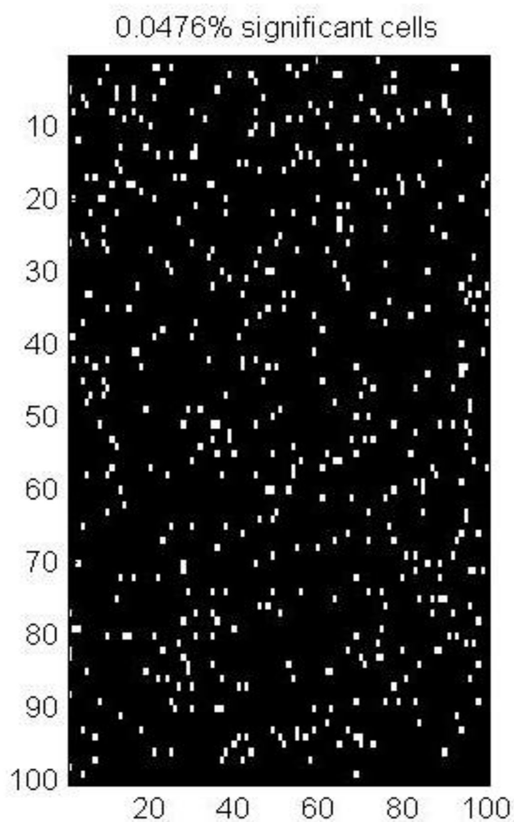
As we observed one test, we know something about the outcome of the other test when variables are corr.

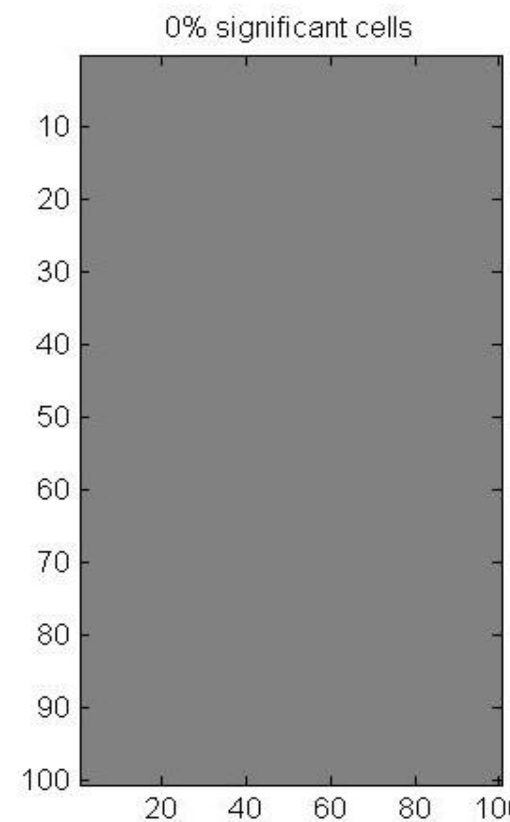
What is the solution?

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



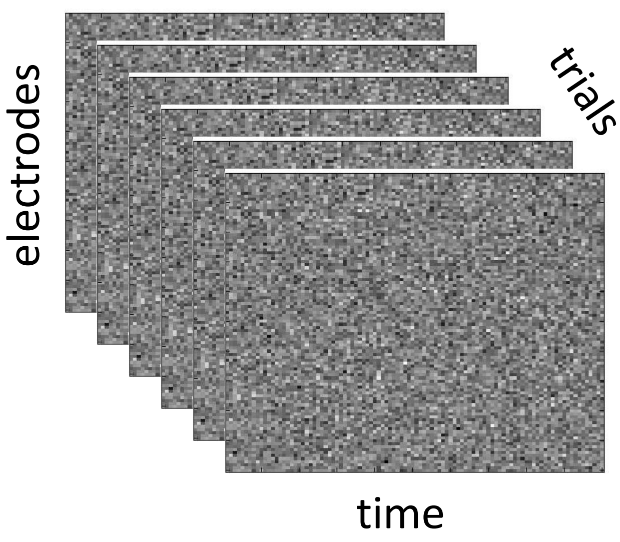
One sample t test > 0 ?



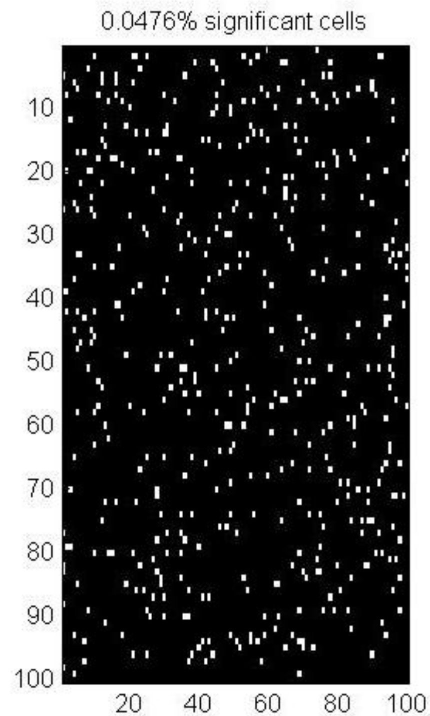


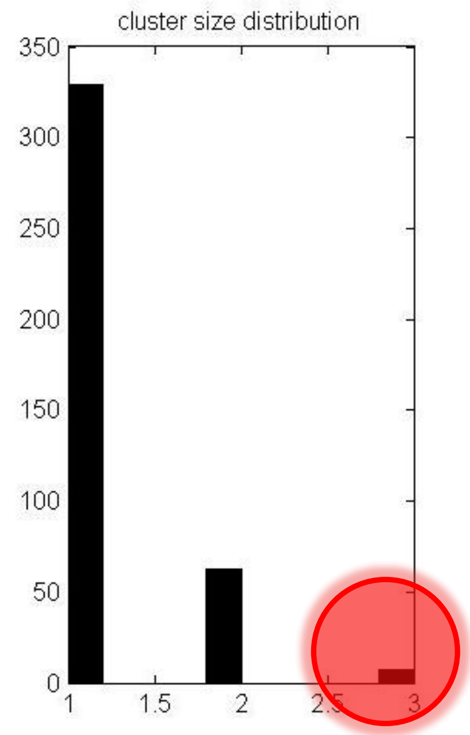
What is the solution?

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



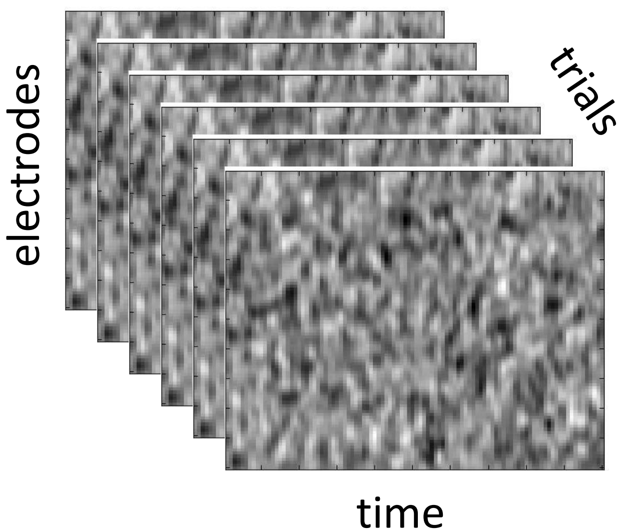
One sample t test > 0 ?



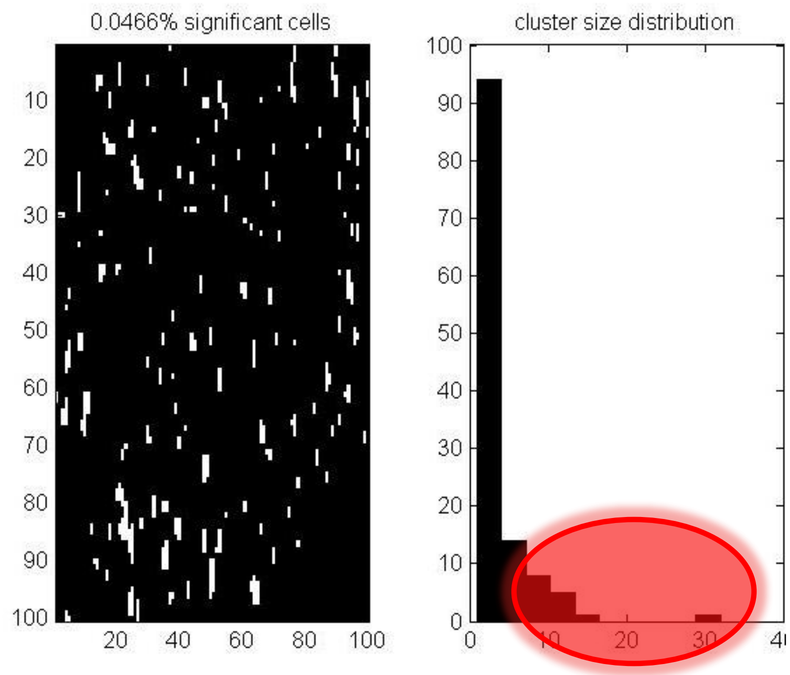


What is the solution?

- 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 ?



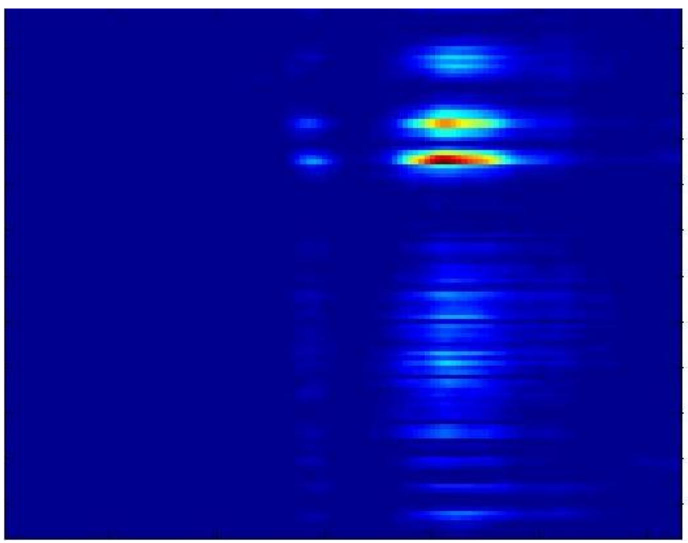
What is the solution?

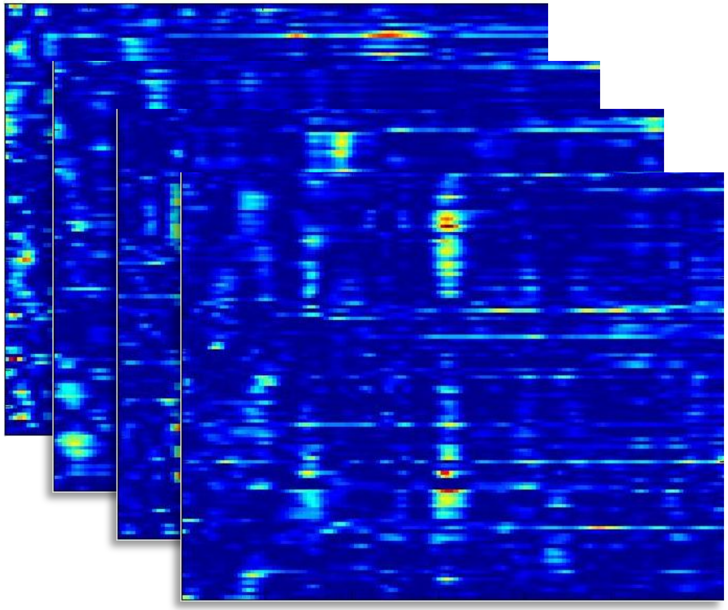
- Clustering is a good option because it accounts for topological features in the data. Techniques like Bonferroni, FDR, max(stats) control the FWER but independently of the correlation between tests.

- 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 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.

What is the 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 – by chance some significant results are obtained.



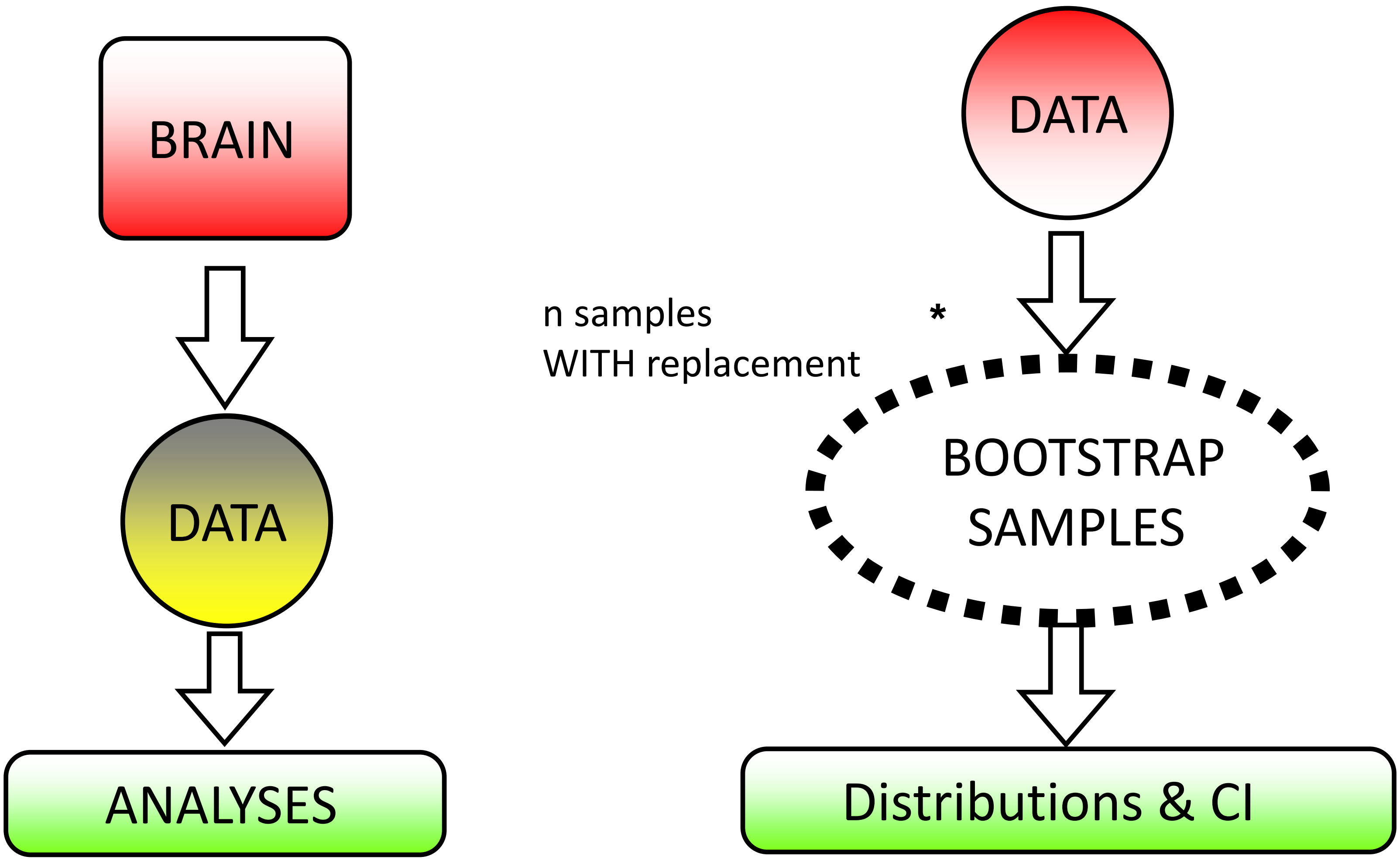


Observed F values → F values under H0

Bootstrap: central idea

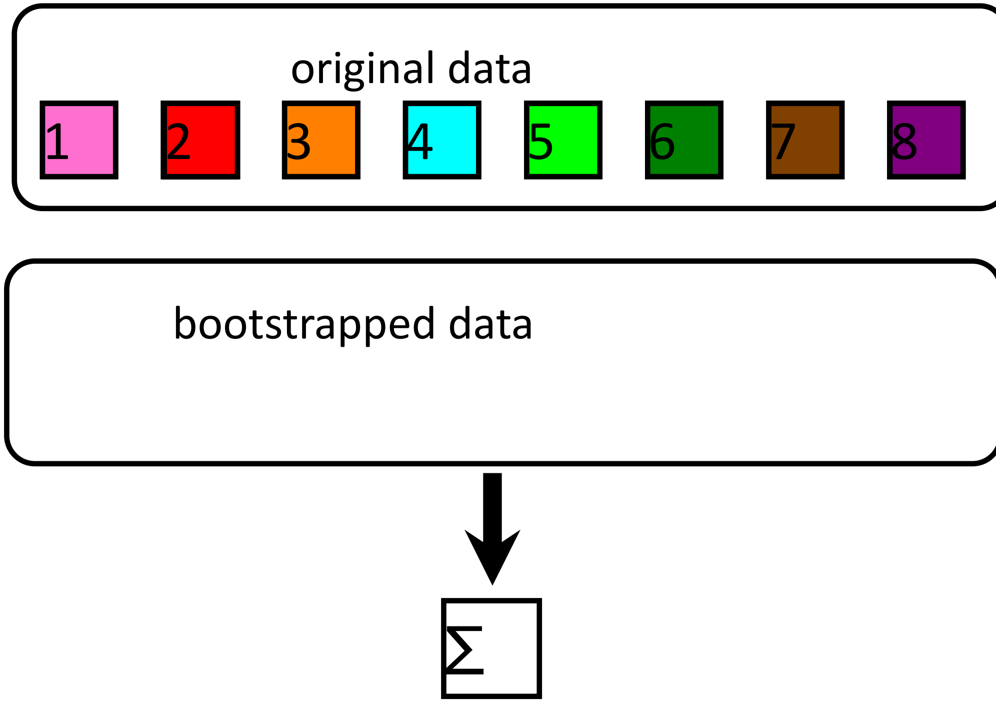
- “The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates.” Efron & Tibshirani, 1993
- “The central idea is that it may sometimes be better to draw conclusions about the characteristics of a population strictly from the sample at hand, rather than by making perhaps unrealistic assumptions about the population.” Mooney & Duval, 1993

Bootstrap philosophy



Percentile bootstrap: general recipe

(1) sample WITH replacement n observations (under H1 for CI of an estimate, under H0 for the null distribution)



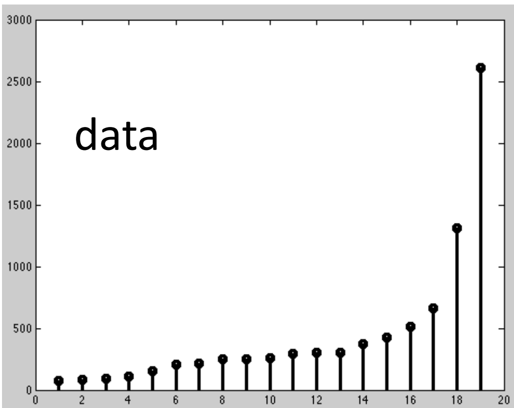
(2) compute estimate
e.g. sum, trimmed mean

(3) repeat (1) & (2) b times
(4) sort the b estimates*
(5) get 1-alpha confidence interval

$\Sigma_1 \;\; \Sigma_2 \;\; \Sigma_3 \;\; \Sigma_4 \;\; \Sigma_5 \;\; \Sigma_6 \;\; \ldots \;\; \Sigma_b$

Percentile bootstrap estimate of mean

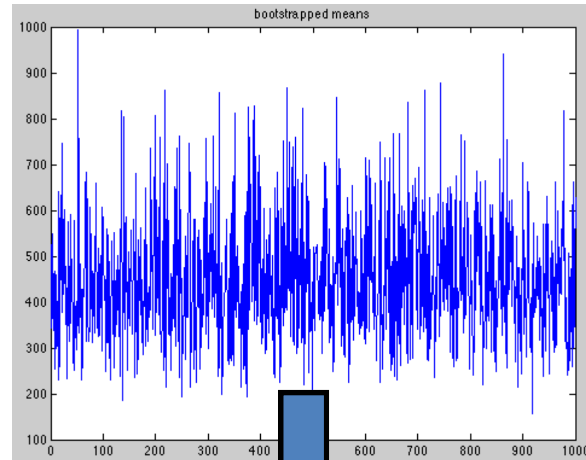
% self-awarness data, Wilcox, 2005, p58



Sample with
replacement b times

compute estimate

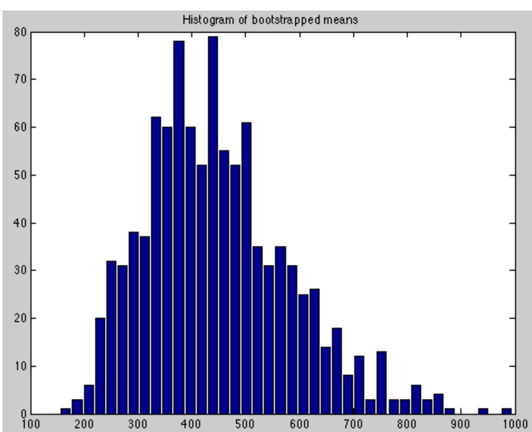
Bootstrapped estimates

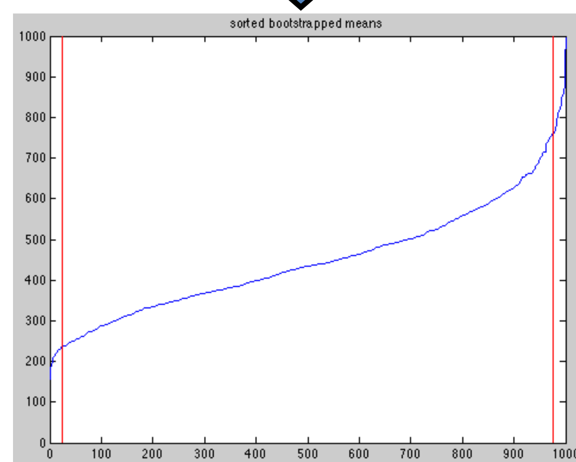


get PDF

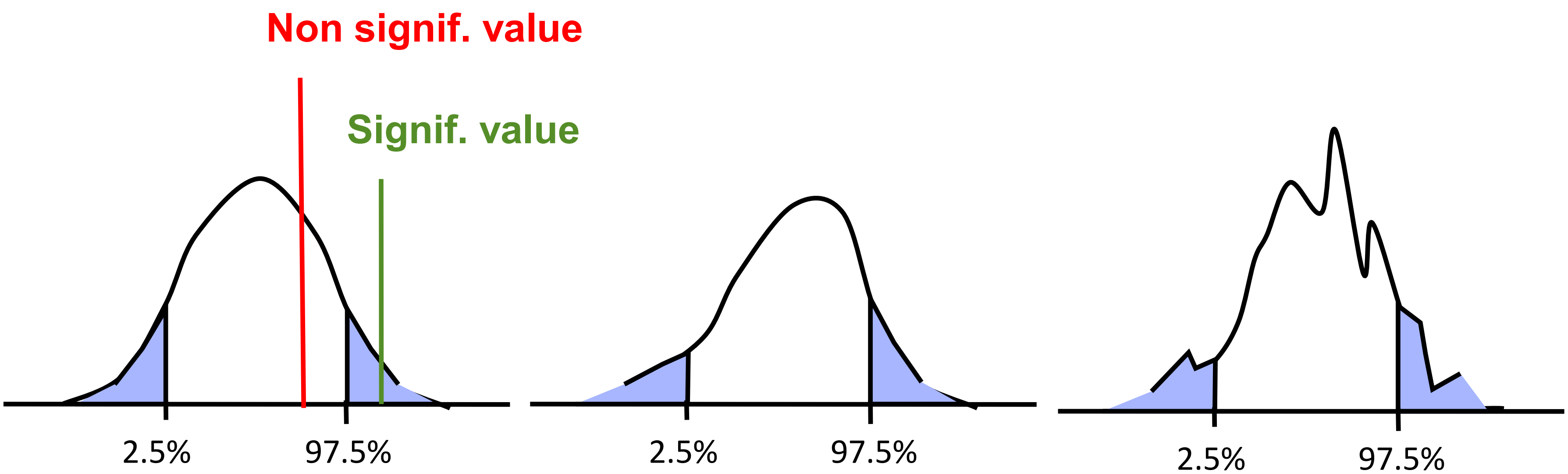
Sort & get CI

Distribution of bootstrapped estimates of the mean





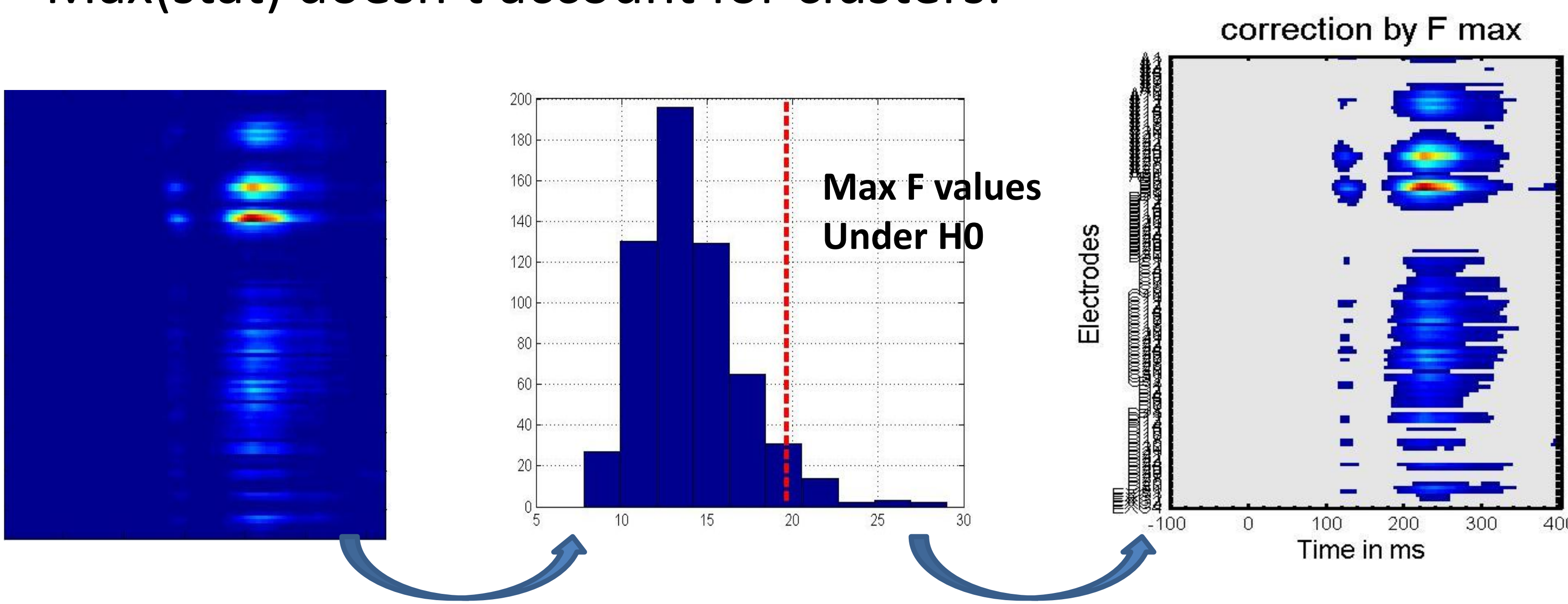
Distributions can take any shape



The percentile bootstrap method allows the bootstrap estimate of the sampling distribution to conform to any shape the data suggest, taking into account the variance and the skewness of the sample. This can be the distribution of T/F values, the distribution of cluster size, height or mass or the distribution of TFCE scores.

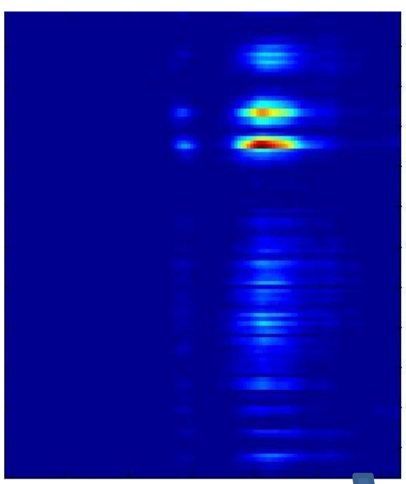
MCC using max across the whole space

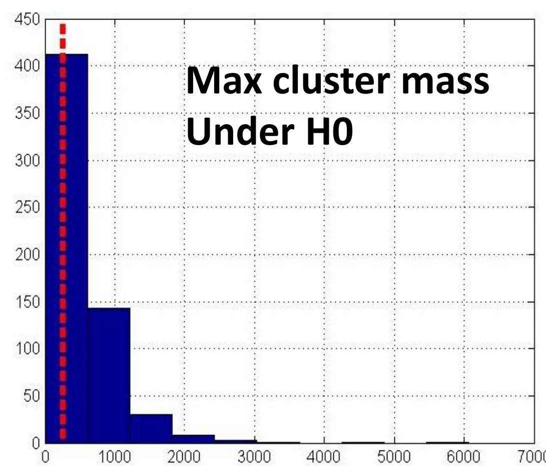
- Max(stat): for each bootstrap record the max(t) or max(F) to build the distribution of max under H0. Then threshold the observed results using this distribution. Because the max value is obtained across all electrodes and time frames, it corrects to thresholding data through this whole space. Max(stat) doesn’t account for clusters.

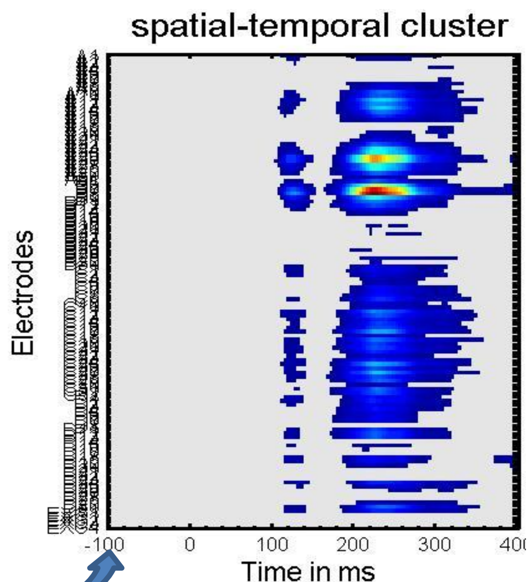


MCC using (max of) clusters

- 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.

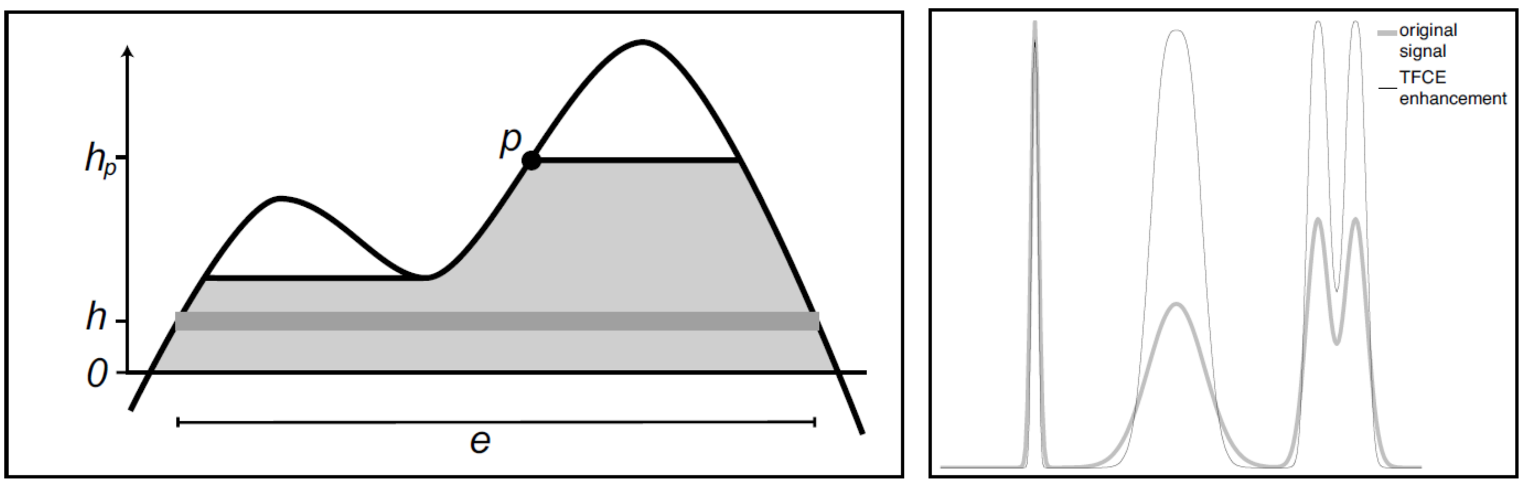
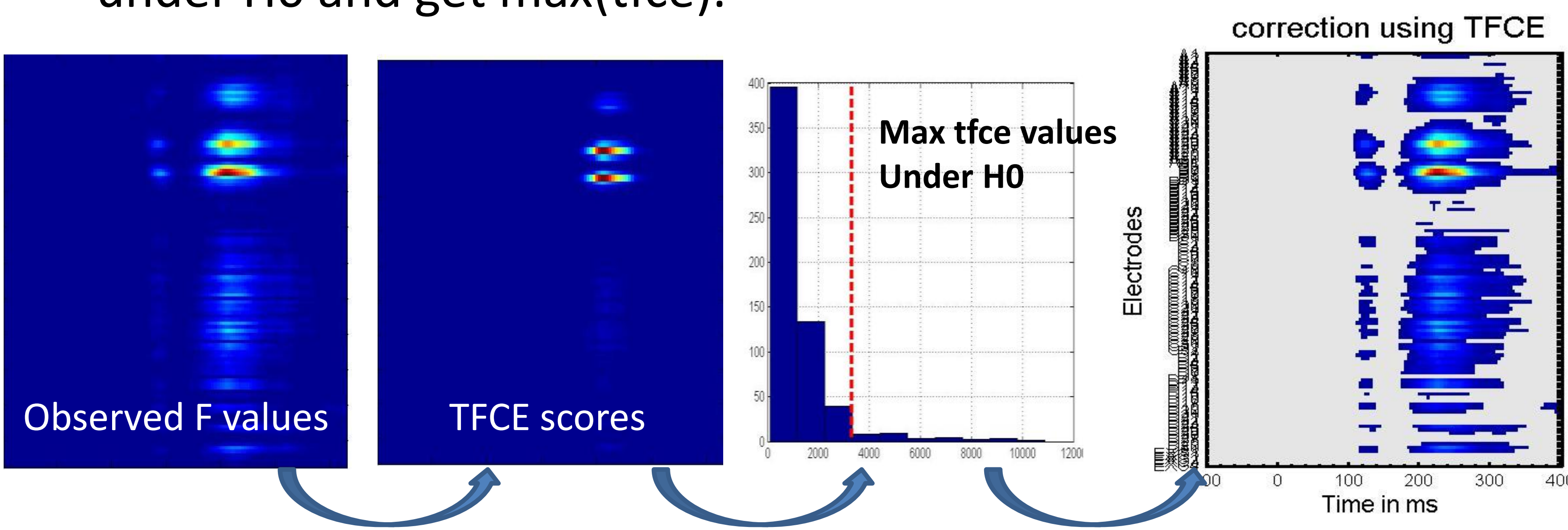


Figure 1: Illustration of the TFCE approach. Left: The TFCE score at voxel p is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark grey band) within the area of "support" of p (light grey). The score for each section is a simple function of its height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

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

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.