Statistics for whole brain analysis & corrections for multiple comparisons

Dr Cyril Pernet, PhD

Neurobiology Research Unit
Rigshospitalet
Copenhagen University Hospital

November 2022


@CyrilPernet@mastodon.online

Motivation for whole channel/IC analyses

- **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

Motivation for whole channel/IC analyses

Processing specificity for human voice stimuli: electrophysiological evidence

Daniel A. Levy,[1] Roni Granot[2] and Shlomo Bentin[1,3,CA]

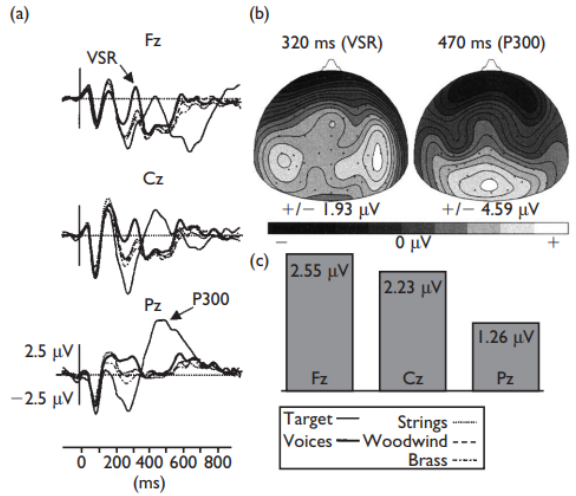


Fig. 1. (**a**) ERPs elicited by piano (target, voice and instrument non-target stimuli in Experiment 1. (**b**) Scalp distributions of voice-specific response (VSR) at 320 ms post-stimulus onset, and of P300 response to target (at 470 ms). (**c**) The anterior–posterior distribution of the difference between the VSR and ERPs elicited by all musical instruments along the sagittal line.

If we had done central channels only and not all scalp we would have had no effect

BMC Neuroscience

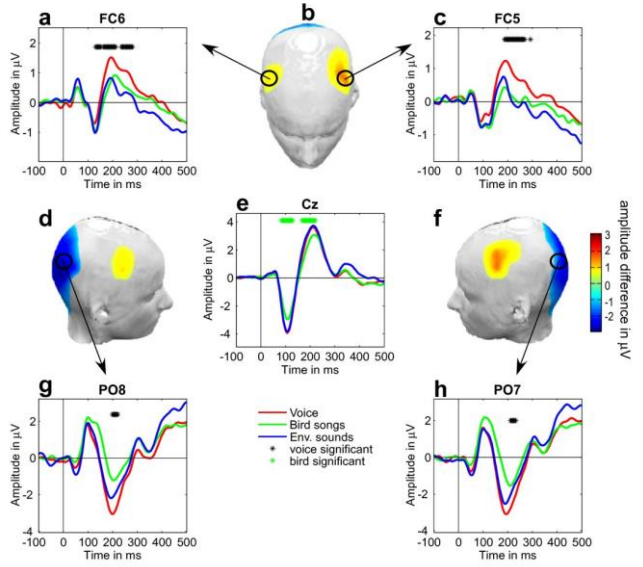
Home About Articles Submission Guidelines Join The Editorial Board

Research article | Open Access | Published: 20 October 2009

Electrophysiological evidence for an early processing of human voices

Ian Charest, Cyril R Pernet, Guillaume A Rousselet, Ileana Quiñones, Marianne Latinus, Sarah Fillion-Bilodeau, Jean-Pierre Chartrand & Pascal Belin

BMC Neuroscience **10**, Article number: 127 (2009) | Cite this article



Motivation for whole channel/IC analyses

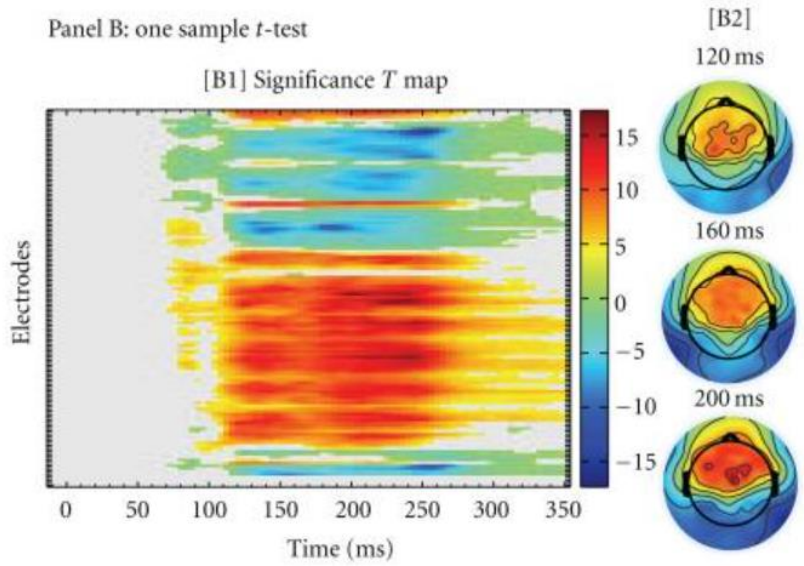
- Statistics on peak latencies and amplitudes? But several lines of evidence suggest that peaks mark the end of a process and therefore it is likely that most of the interesting effects lie in a component before a peak
- **Neurophysiology**: whether ERPs are due to additional signal or to phase resetting 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

Motivation for whole channel/IC analyses

- Like for MRI, we refer to whole brain analysis as a ‘mass univariate’ approach
- This is simply computing statistical test at every channel/IC/source * time/freq.

Pernet et al. "LIMO EEG: A Toolbox for Hierarchical LInear MOdeling of ElectroEncephaloGraphic Data", *Computational Intelligence and Neuroscience*, vol. 2011

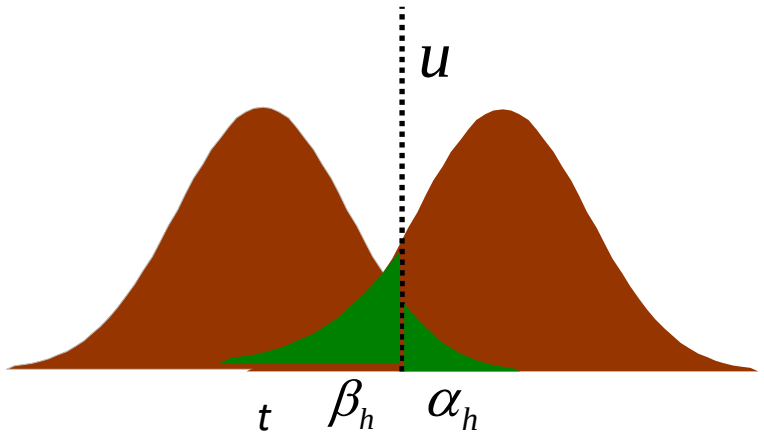


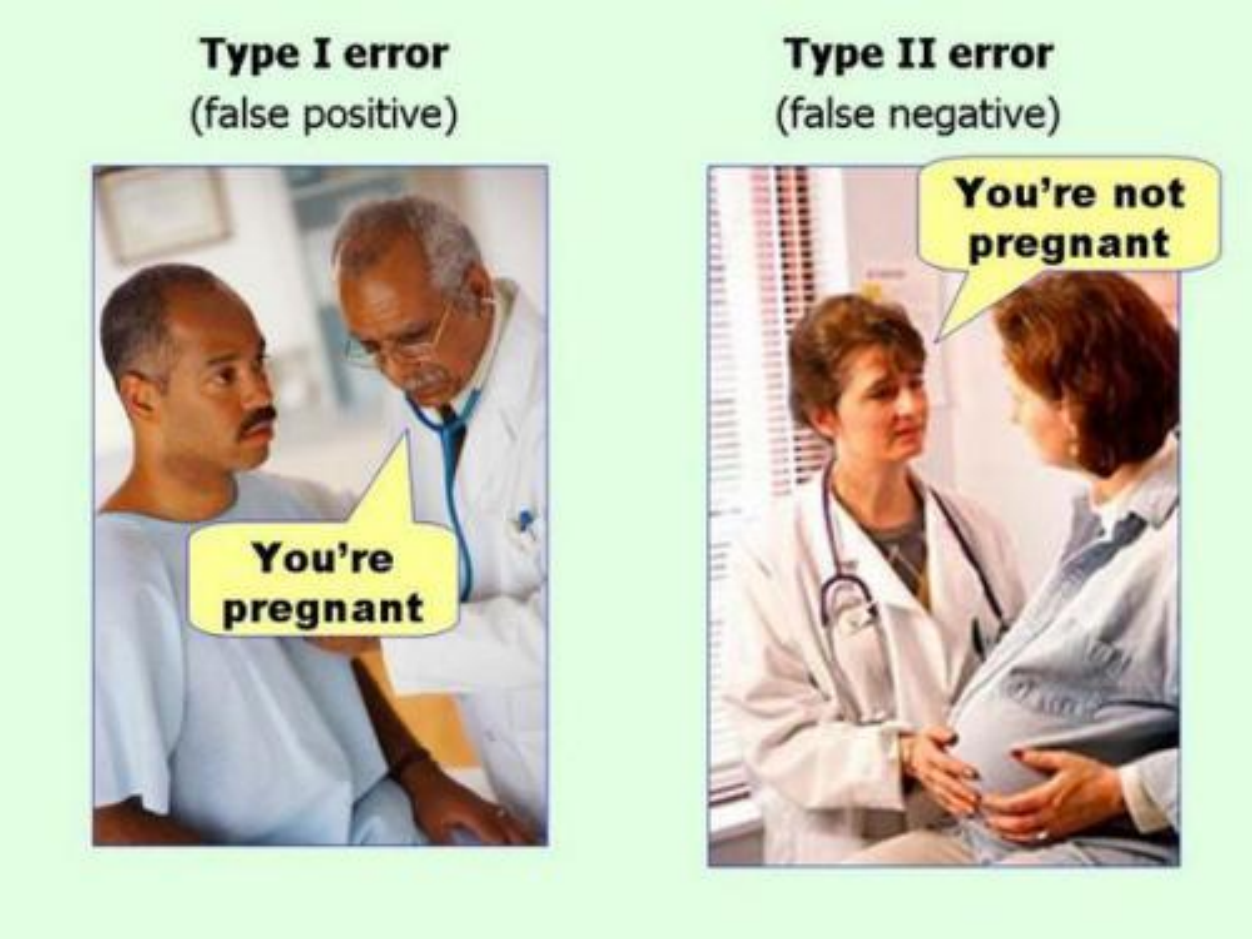
The problem with mass-univariate analyses

No free lunch

Pearson-Newman hypothesis testing

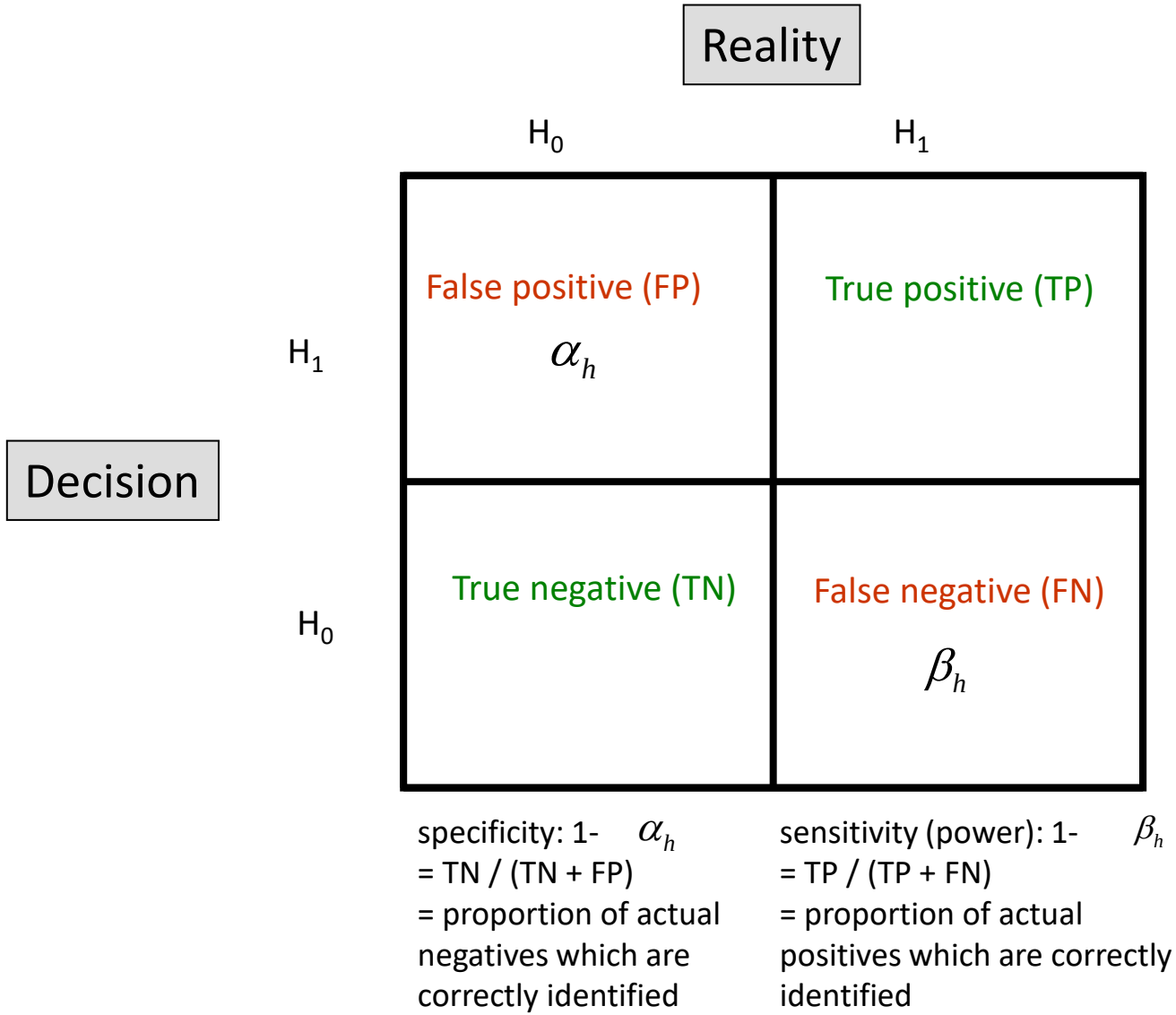
- H0: no effect
- H1: there is an effect





https://chemicalstatistician.wordpress.com

Types of error



From J Chumbley's slide 2015

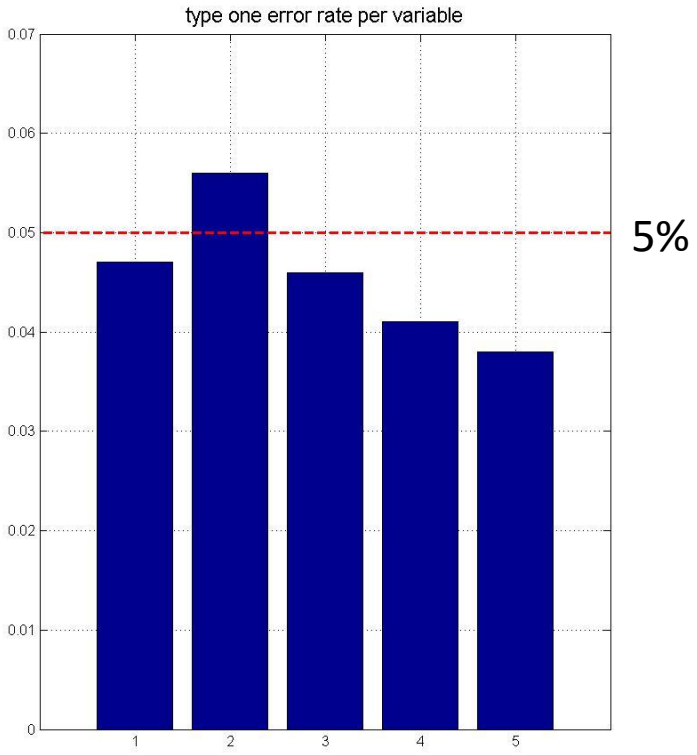
What is the problem?

- 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

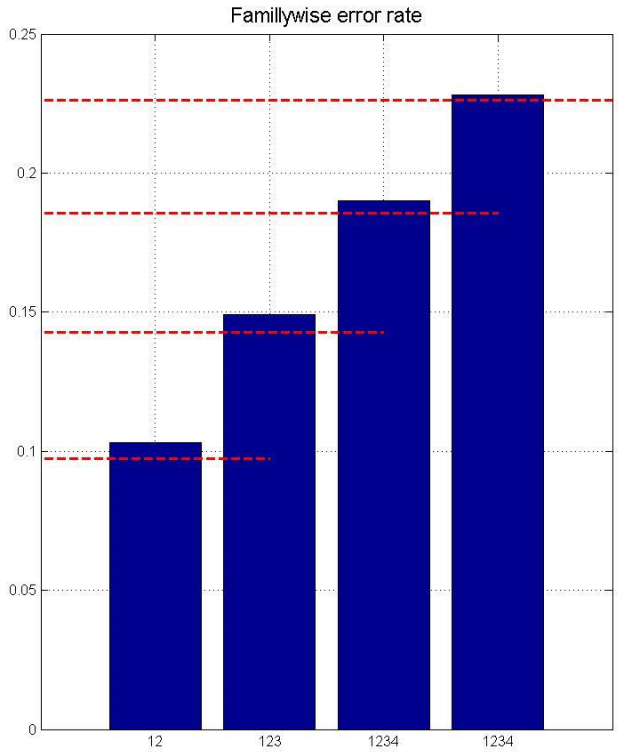
- for alpha =5/100, if we do 2 tests we should get about 1-(1-5/100)^2 ~ 9% false positives, if we do 126 electrodes * 150 time frames tests, we should get about 1-(1-5/100)^18900 ~ 100% false positives! i.e. you can’t be certain of any of the statistical results you observe

What is the problem?

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



5%



22%

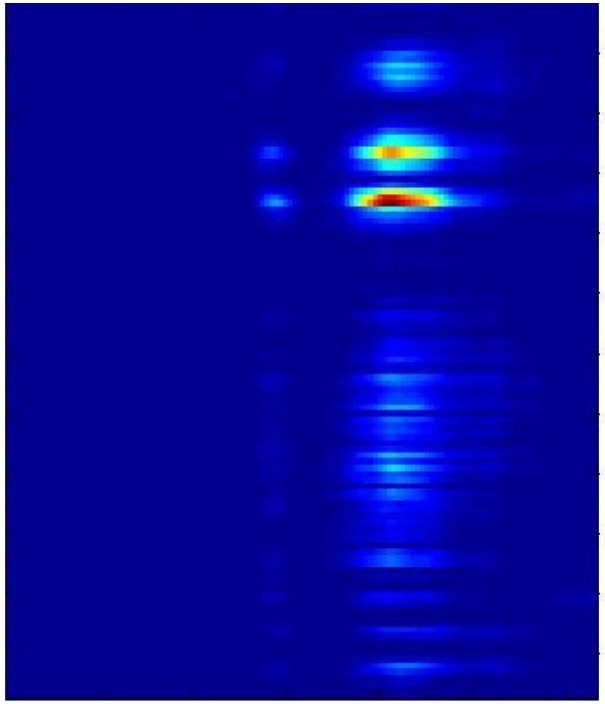
18%

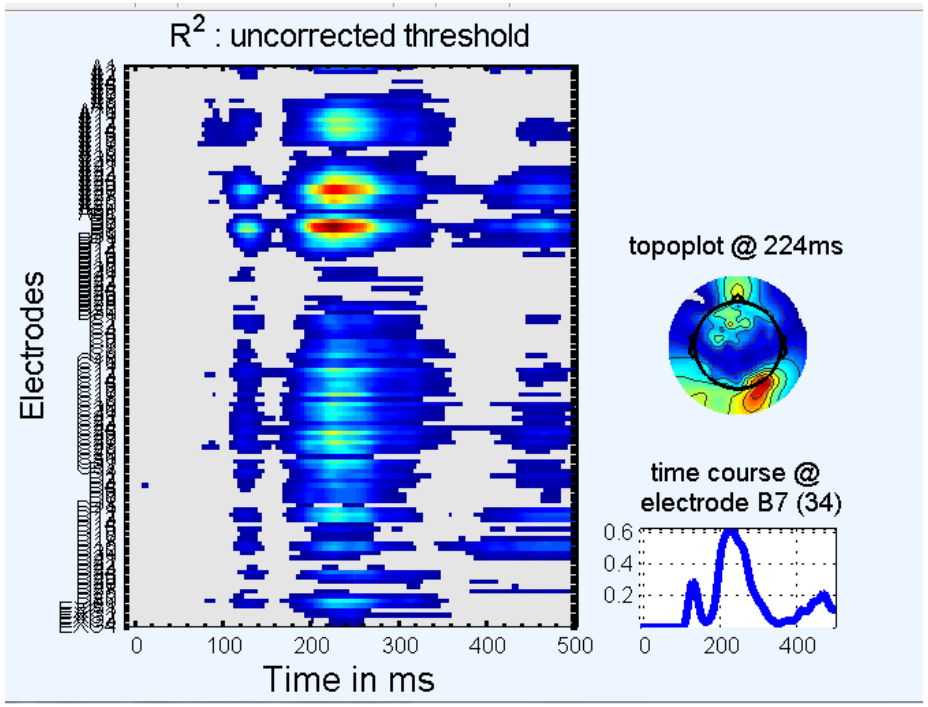
14%

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?

Family Wise Error rate solution

- 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*

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

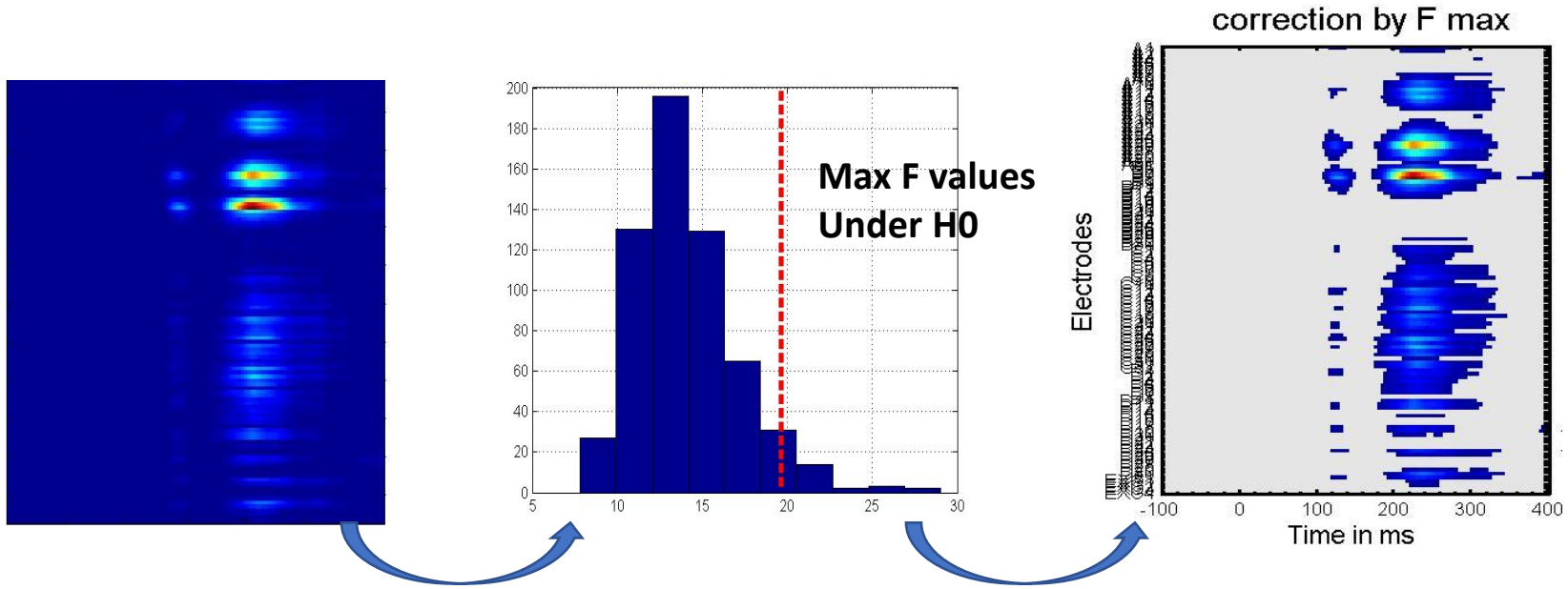
$$\leq \sum P(T_i \geq u|H0)$$ *Boole's inequality*

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

Maximum statistics and resampling

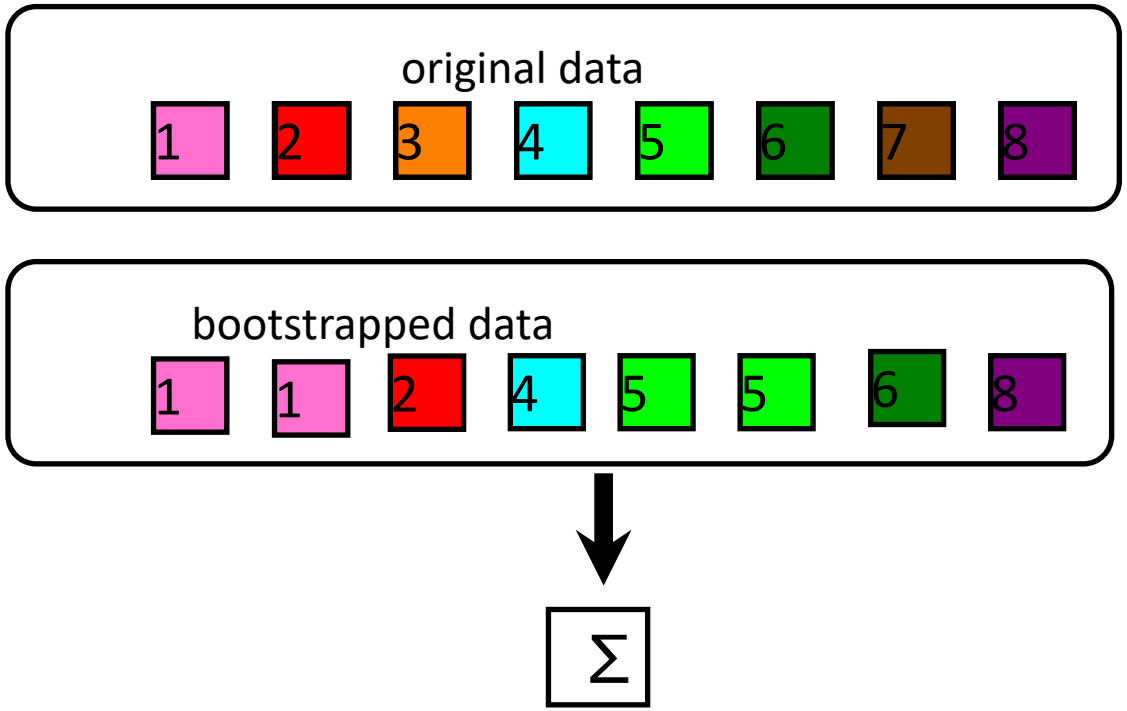
Maximum Statistics based on resampling

- 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



General recipe

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



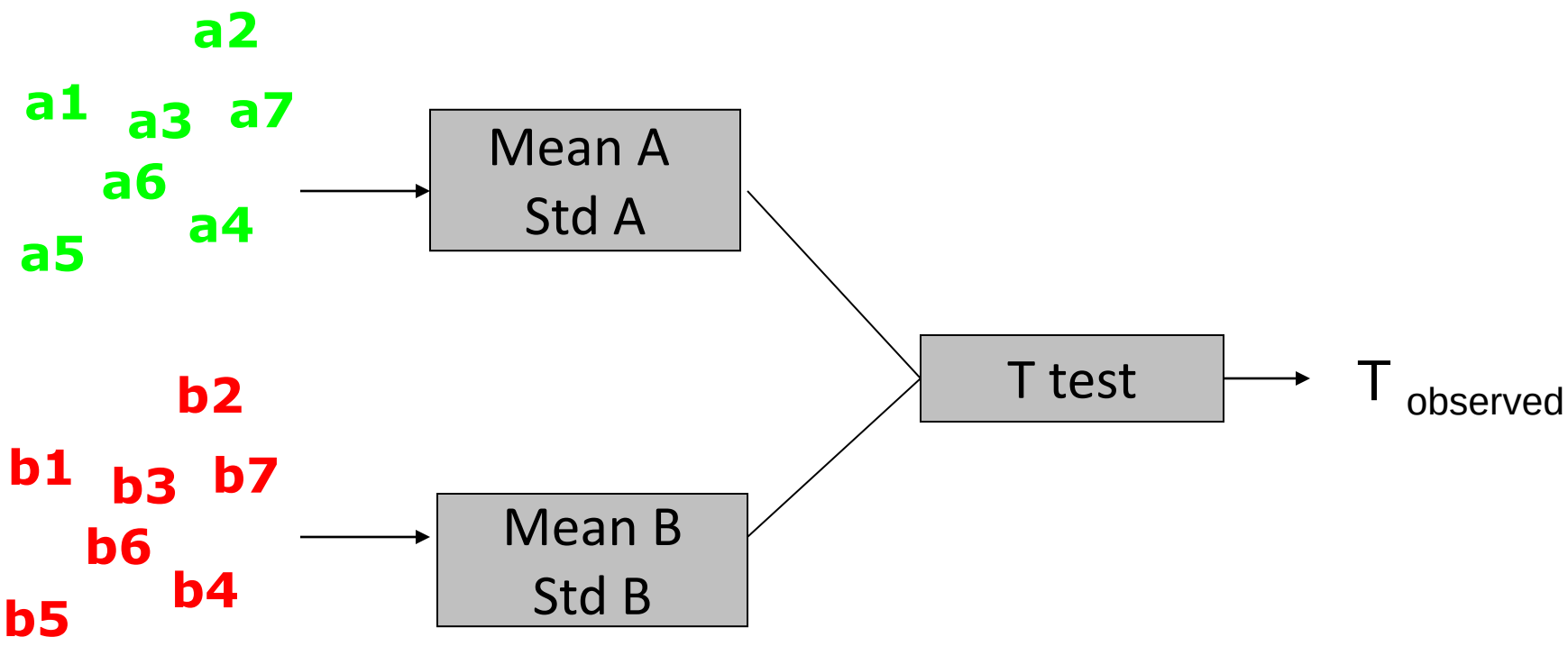
(2) compute estimate
e.g. sum, trimmed mean

(3) repeat (1) & (2) b times

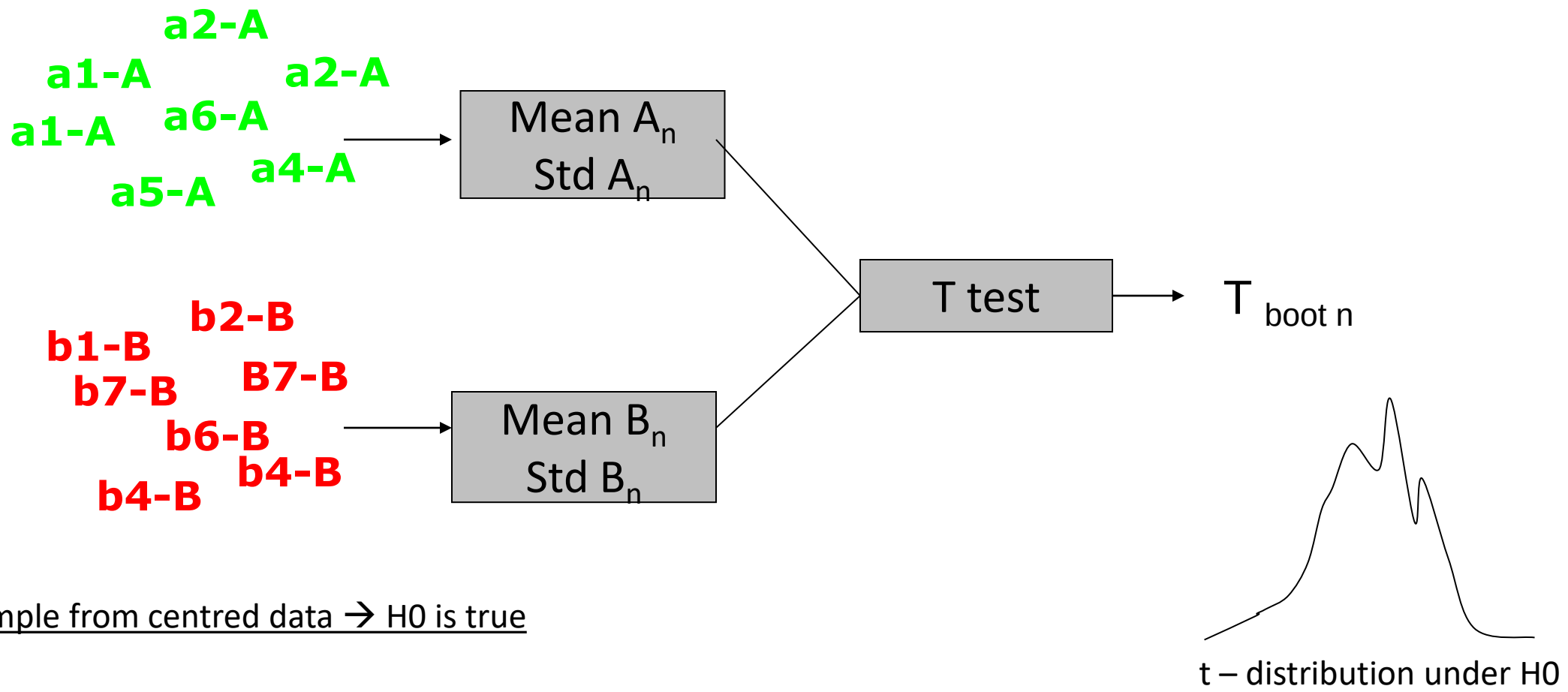
Σ_1 Σ_2 Σ_3 Σ_4 Σ_5 Σ_6 . . . Σ_b

(4) get bias, std, confidence interval, p-value

Application to a 2 samples t-test: Bootstrap under H0



Application to a 2 samples t-test: Bootstrap under H0

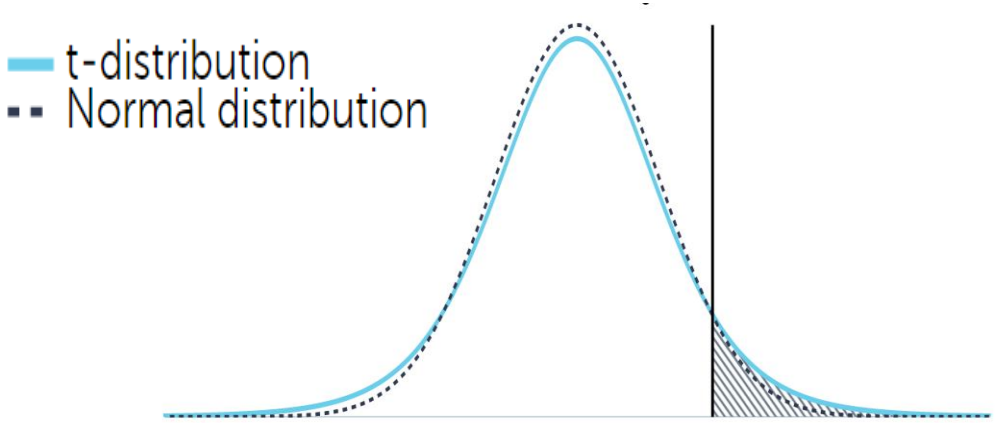


Resample from centred data → H0 is true

Application to a 2 samples t-test: Bootstrap under H0

What is the p value of the sample

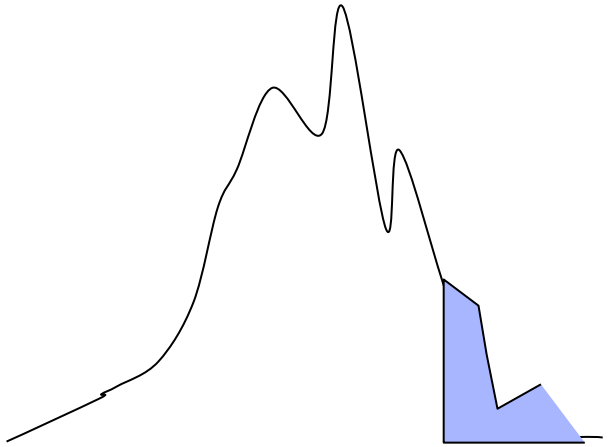
p(Obs≥t|H0) → cumulative probability



area under the curve for T obs = p value
Significance = point of T critical

What is the p value of the sample

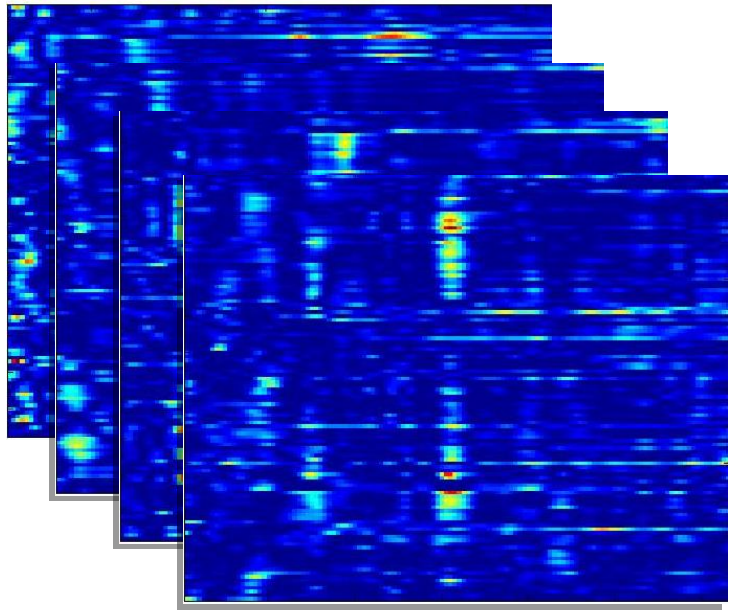
p(Obs≥t|H0) → cumulative probability



area under the curve for T obs = p value
Significance = percentile of the empirical t distribution
→ Theoretical T assumes data normality, we don't

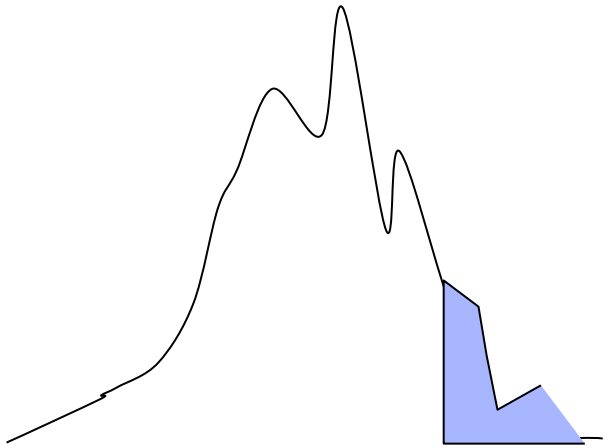
Application to a 2 samples t-test: Bootstrap under H0

For each H0, we now have maps



What is the p value of the sample

$p(Obs \geq t \mid \max H0)$



Remember FWER is the prob to make at least 1 error.
If we control for the biggest stat value, we control them all

Clustering

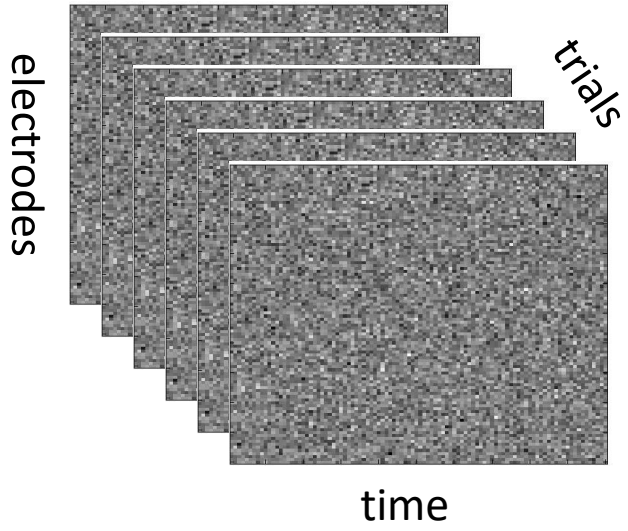
Solutions for imaging data

- 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)

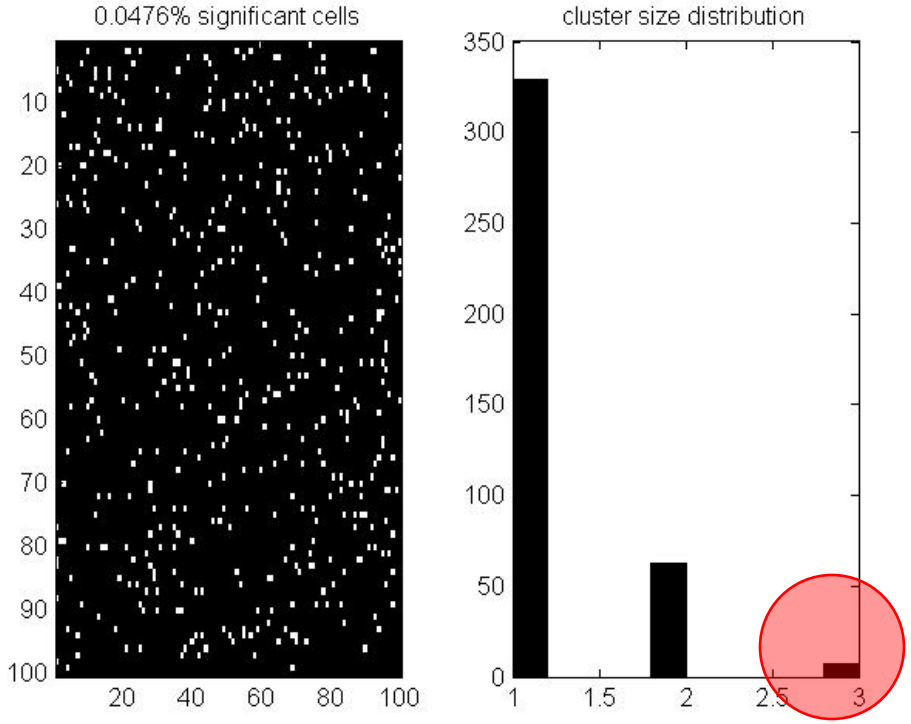
Chumbley et al., 2009 NeuroImage 44

Let’s analyse clusters

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

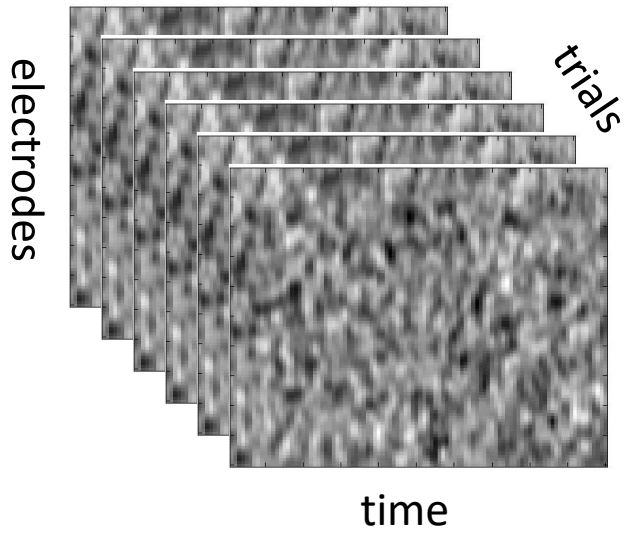


One sample t test > 0 ?

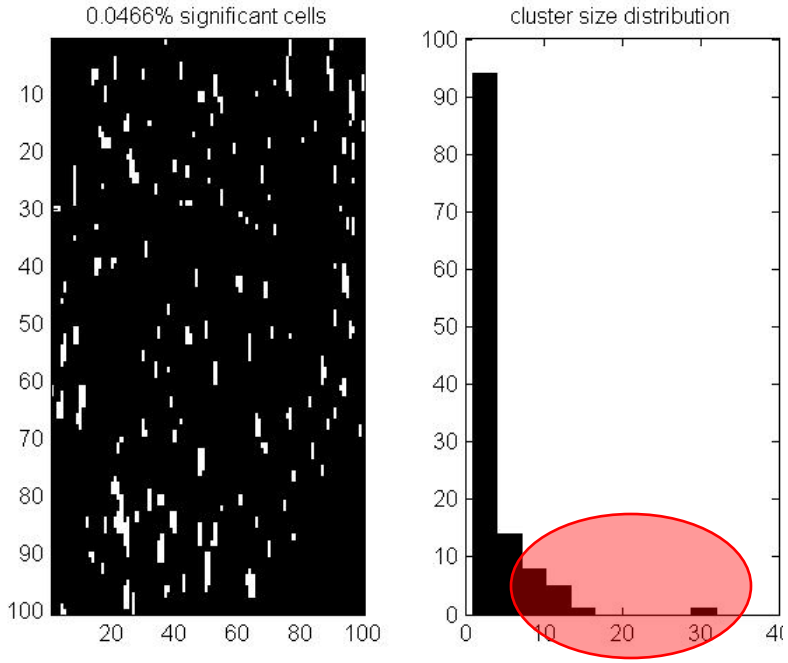


Let’s analyse clusters

- 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 ?



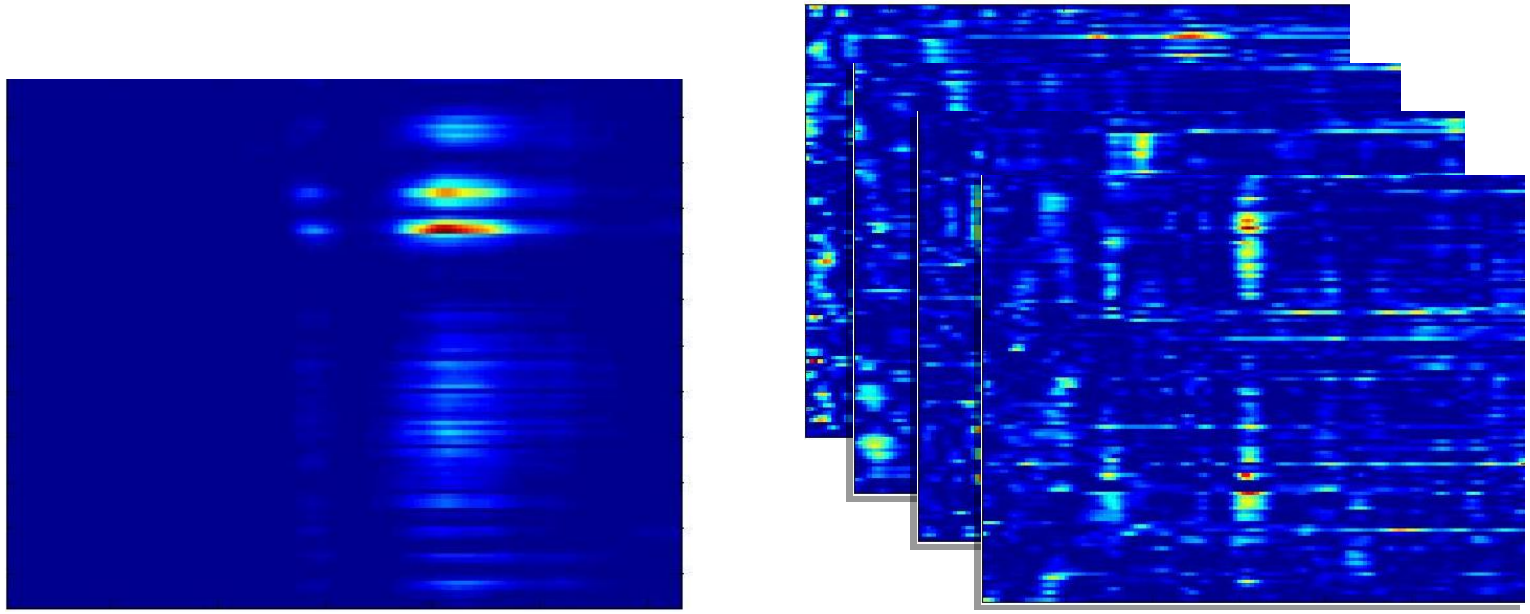
The clustering solution

- Clustering is 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

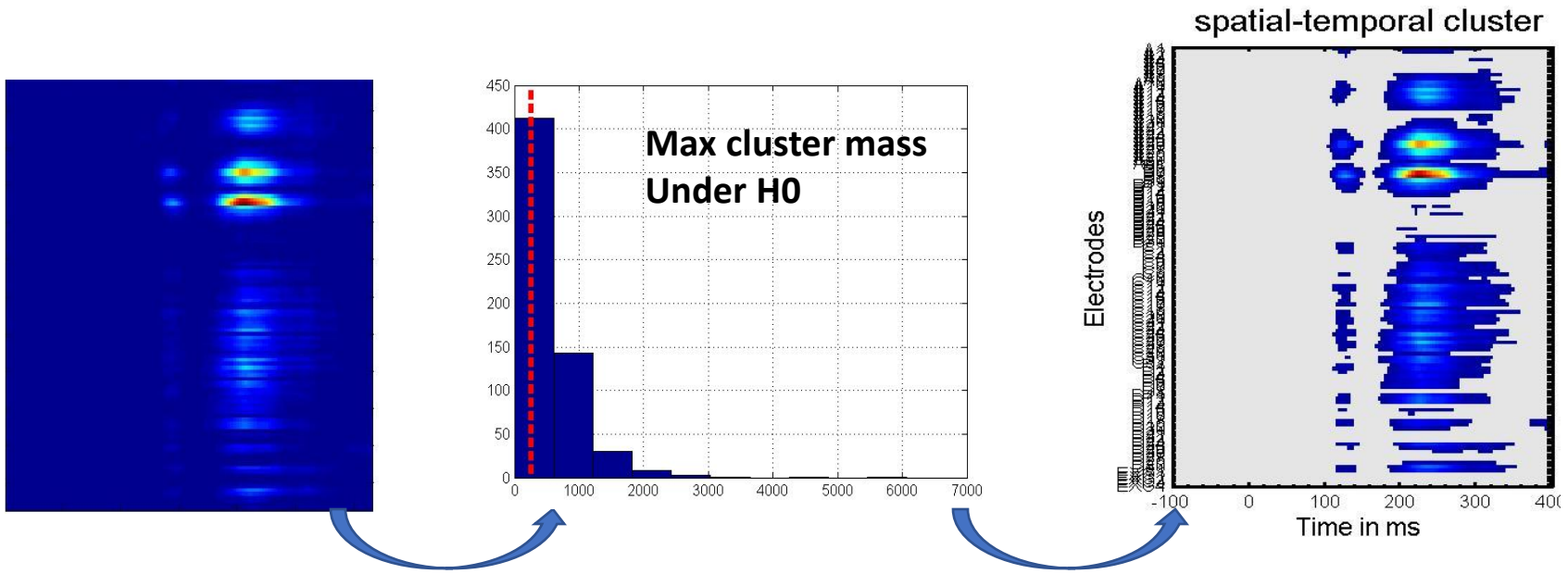
- In LIMO EEG, we bootstrap the data under H0: center the data or break the link between the design matrix and the data and then resample and test. This way we can find u for a single bin, the the whole space, or for clusters.



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 !

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

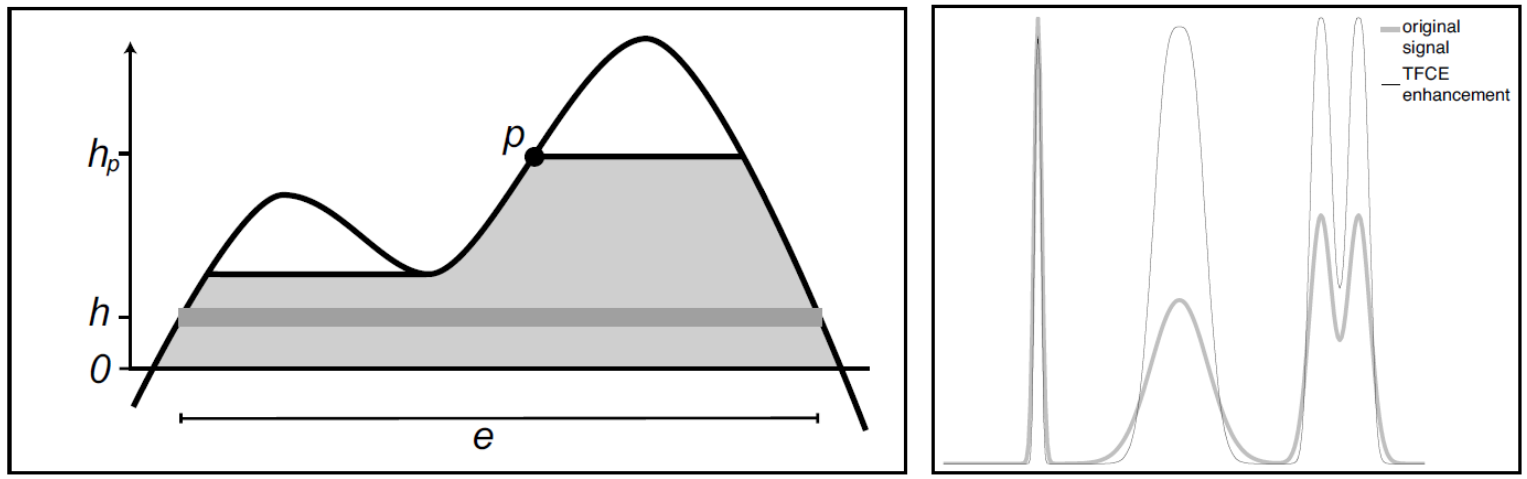
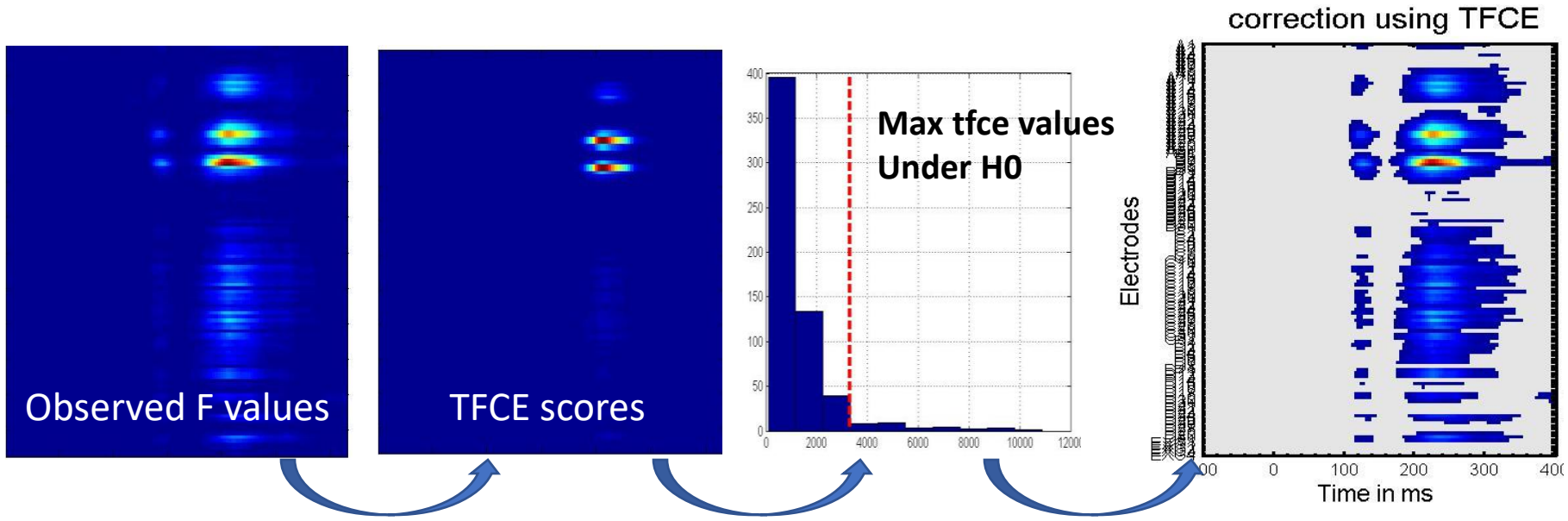


Figure 1: Illustration of the TFCE approach. Left: The TFCE score at voxel p is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark grey band) within the area of "support" of p (light grey). The score for each section is a simple function of its height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

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