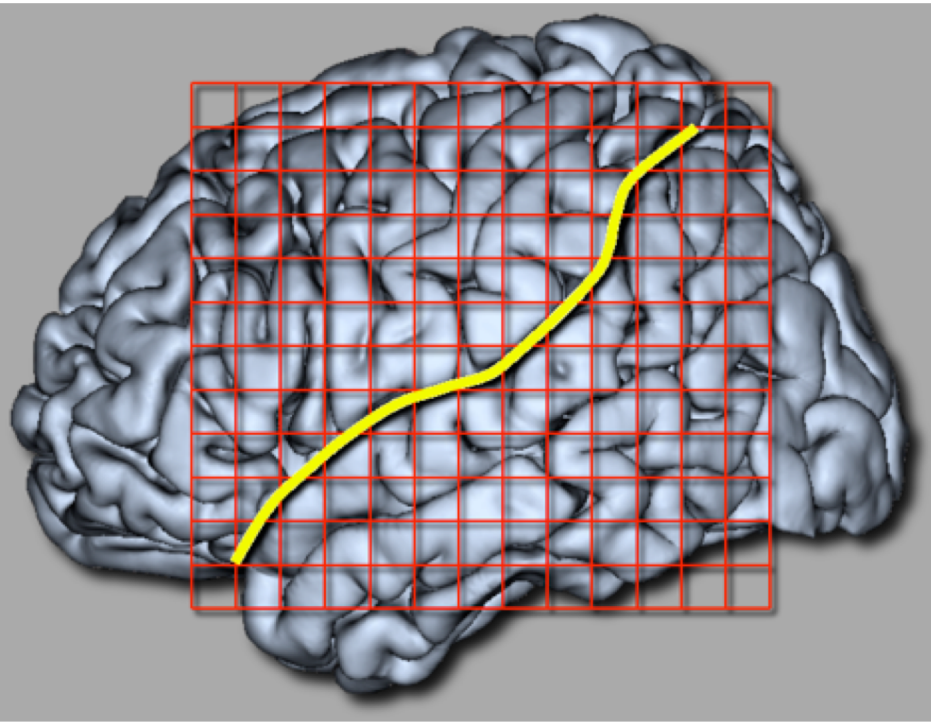
Threshold-free Cluster Enhancement
TFCE



Christian Gaser
Structural Brain Mapping Group

Kliniken für Neurologie und Psychiatrie | Universitätsklinikum Jena

Cluster failure

PNAS

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[b,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; [b]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; [c]Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

The most widely used task functional magnetic resonance imaging (fMRI) analyses use parametric statistical methods that depend on a variety of assumptions. In this work, we use real resting-state data and a total of 3 million random task group analyses to compute empirical familywise error rates for the fMRI software packages SPM, FSL, and AFNI, as well as a nonparametric permutation method. For a nominal familywise error rate of 5%, the parametric statistical methods are shown to be conservative for voxelwise inference and invalid for clusterwise inference. Our results suggest that the principal cause of the invalid cluster inferences is spatial autocorrelation functions that do not follow the assumed Gaussian shape. By comparison, the nonparametric permutation test is found to produce nominal results for voxelwise as well as clusterwise inference. These findings speak to the need of validating the statistical methods being used in the field of neuroimaging.

(FWE), the chance of one or more false positives, and empirically measure the FWE as the proportion of analyses that give rise to any significant results. Here, we consider both two-sample and one-sample designs. Because two groups of subjects are randomly drawn from a large group of healthy controls, the null hypothesis of no group difference in brain activation should be true. Moreover, because the resting-state fMRI data should contain no consistent shifts in blood oxygen level-dependent (BOLD) activity, for a single group of subjects the null hypothesis of mean zero activation should also be true. We evaluate FWE control for both voxelwise inference, where significance is individually assessed at each voxel, and clusterwise inference (19–21), where significance is assessed on clusters formed with an arbitrary threshold.

In brief, we find that all three packages have conservative voxelwise inference and invalid clusterwise inference, for both

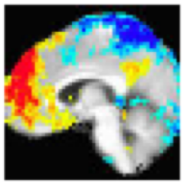
Cluster failure

New Study Raises Doubts About **fMRI** Neuroimaging Research
Mad In America - 30.08.2016
Cluster failure: Why fMRI inferences for spatial extent have inflated fal
positive rates. Proceedings of the National Academy of Sciences. ...

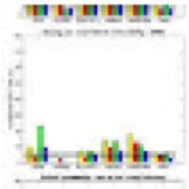
False-Positive fMRI Hits The Mainstream
Discover Magazine (blog) - 07.07.2016
Cluster failure: Why fMRI inferences for spatial extent have inflated fal
positive rates. Proceedings of the National Academy of Sciences of the ...
A software bug could render the last 15 years of brain research ...
ZME Science - 07.07.2016
Alle ansehen

Softwares for **fMRI** yield erroneous results
Science Daily - 27.06.2016
Cluster failure: why fMRI inferences for spatial extent have inflated false positive rates.
June 2016 DOI: 10.1073/pnas.1602413113 ...

JSMRM 2016/第44回日本磁気共鳴医学会大会が開催
インナビネット - 11.09.2016
初日午前には，第2会場において，社会ニュースでも話題になった，**fMRI**の
析の問題をテーマに ... に講演し，Eklundらの"**Cluster failure ; Why fMRI**
inferences for spatial extent have inflated false-positive rates"において
れ ...



Forscher sehen Qualitätsprobleme
ORF.at - 18.07.2016
Dem Gehirn beim Arbeiten zusehen - das verspricht die funktionelle
Magnetresonanztomografie (fMRT). Forscher schlugen kürzlich Alarm: 70 .



Software Flaw Casts Doubt on Past Task **fMRI** Studies
Alzforum - 25.08.2016
Task-based functional MRI **has** opened a window to the mind, ... as a 70 p
chance of producing at least one **false positive** result. ... The same set of
produced a 20-40 percent error **rate** when run with lenient parameters for
size ... and 5 percent or less using voxel-based **inference** (bottom).

Swedish Study Highlights Potential Flaws in **fMRI** Findings
Lexology (registration) - 11.07.2016
Anders Eklund, "**Cluster failure: Why fMRI inferences for spatial extent have inflated fal**
rates," PNAS, June 2016. For more than 15 ...

Introducing Sparks, Our New Science Podcast
FiveThirtyEight - 19.08.2016
It's called Sparks, and it's where FiveThirtyEight's science writers and I tal
science writing that sparked our interest (hey, titles **have** to ...

When Big Data is Bad Data
ZDNet - 15.07.2016
In the paper **Cluster failure: Why fMRI inferences for spatial extent ha**
false-positive rates researchers Anders Eklund and Hans ...

Cluster failure

Correction for Eklund et al., Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Extract | Full Text | Authors & Info | Metrics | Related Content | PDF

NEUROSCIENCE, STATISTICS Correction for "Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates," by Anders Eklund, Thomas E. Nichols, and Hans Knutsson, which appeared in issue 28, July 12, 2016, of *Proc Natl Acad Sci USA* (113:7900–7905; first published June 28, 2016; 10.1073/pnas.1602413113).

The authors note that on page 7900, in the Significance Statement, lines 9–11, "These results question the validity of some 40,000 fMRI studies and may have a large impact on the interpretation of neuroimaging results" should instead appear as "These results question the validity of a number of fMRI studies and may have a large impact on the interpretation of weakly significant neuroimaging results."

Additionally, the authors note that on page 7904, left column, fifth full paragraph, lines 1–3, "It is not feasible to redo 40,000 fMRI studies, and lamentable archiving and data-sharing practices mean most could not be reanalyzed either" should instead appear as "Due to lamentable archiving and data-sharing practices, it is unlikely that problematic analyses can be redone."

These errors do not affect the conclusions of the article. The online version has been corrected.

PubMed Commons PubMed Commons home

1 comment

How to join PubMed Commons

Thomas E. Nichols 2016 Jul 14 1:36 p.m. 5 of 5 people found this helpful

This work has two unfortunate statements that can be misunderstood to mean that the findings apply to all "40,000" publications in the fMRI literature. The following two corrections will resolve this problem:

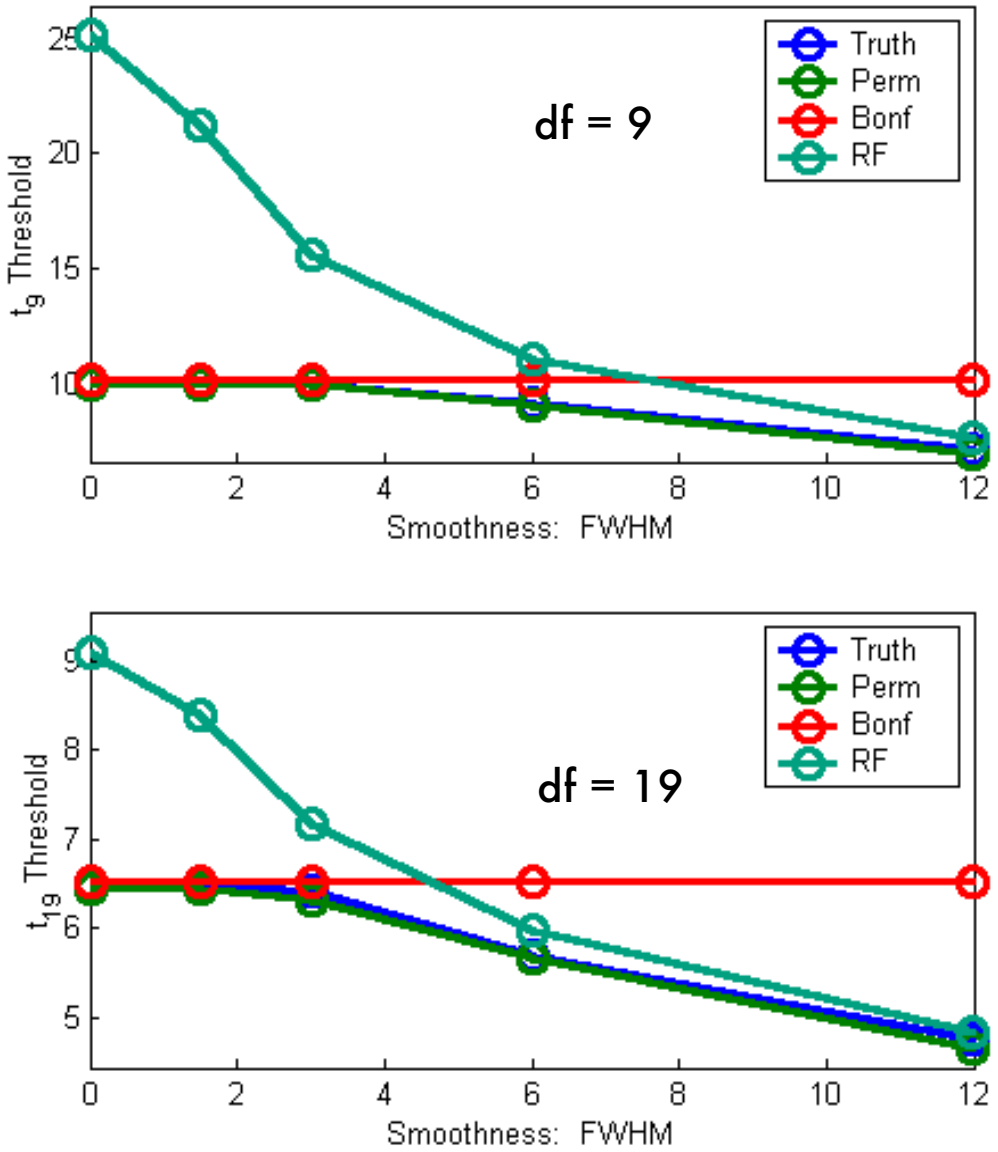
The last sentence of the Significance statement should read: "These results question the validity of a number of fMRI studies and may have a large impact on the interpretation of weakly significant neuroimaging results."

The first sentence after the heading "The future of fMRI" should have read: "Due to lamentable archiving and data-sharing practices it is unlikely that problematic analyses can be redone."

For more on this, see the blog entry "Bibliometrics of Cluster Inference" http://blogs.warwick.ac.uk/nichols/entry/bibliometrics_of_cluster/

Permalink Share

Validität GRF

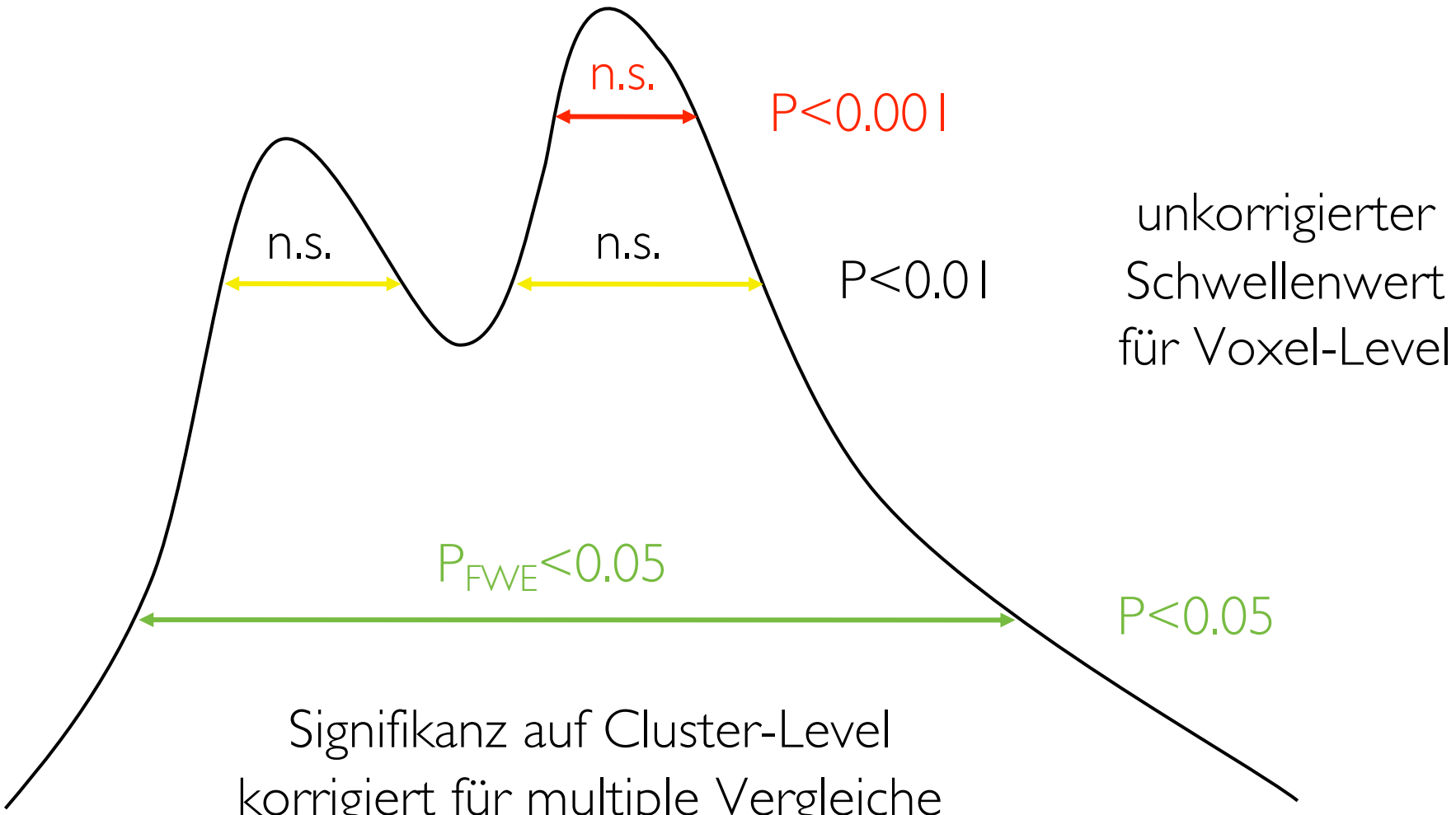


- GRF benötigt viele Freiheitsgrade und große FWHM, um valide zu sein
- nichtparametrische Tests als Alternative bei kleinem df

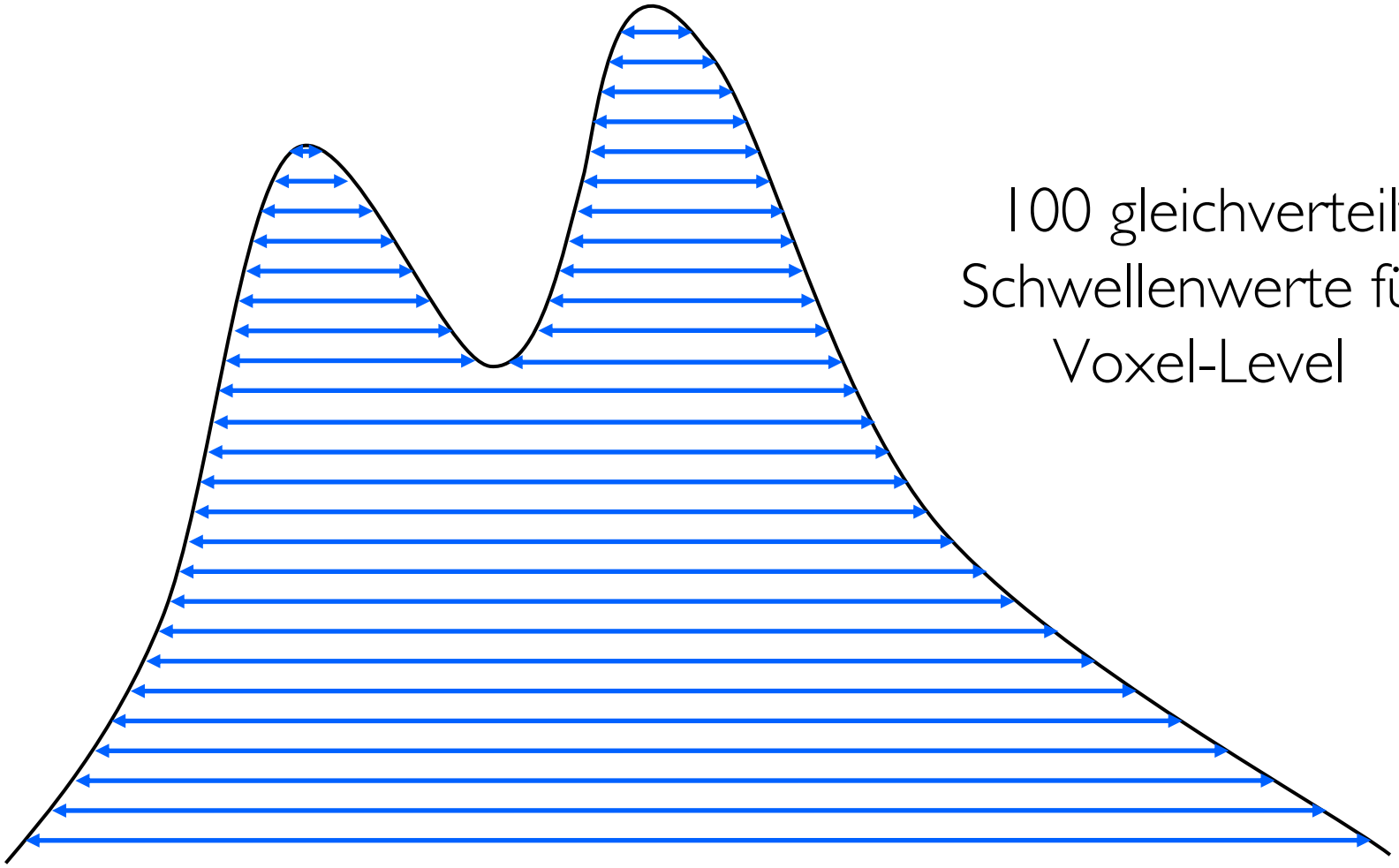
Nichols & Hayasaka

Stat Methods Med Res 2003

Abhängigkeit der Schwellenwerte



Threshold-free Cluster Enhancement

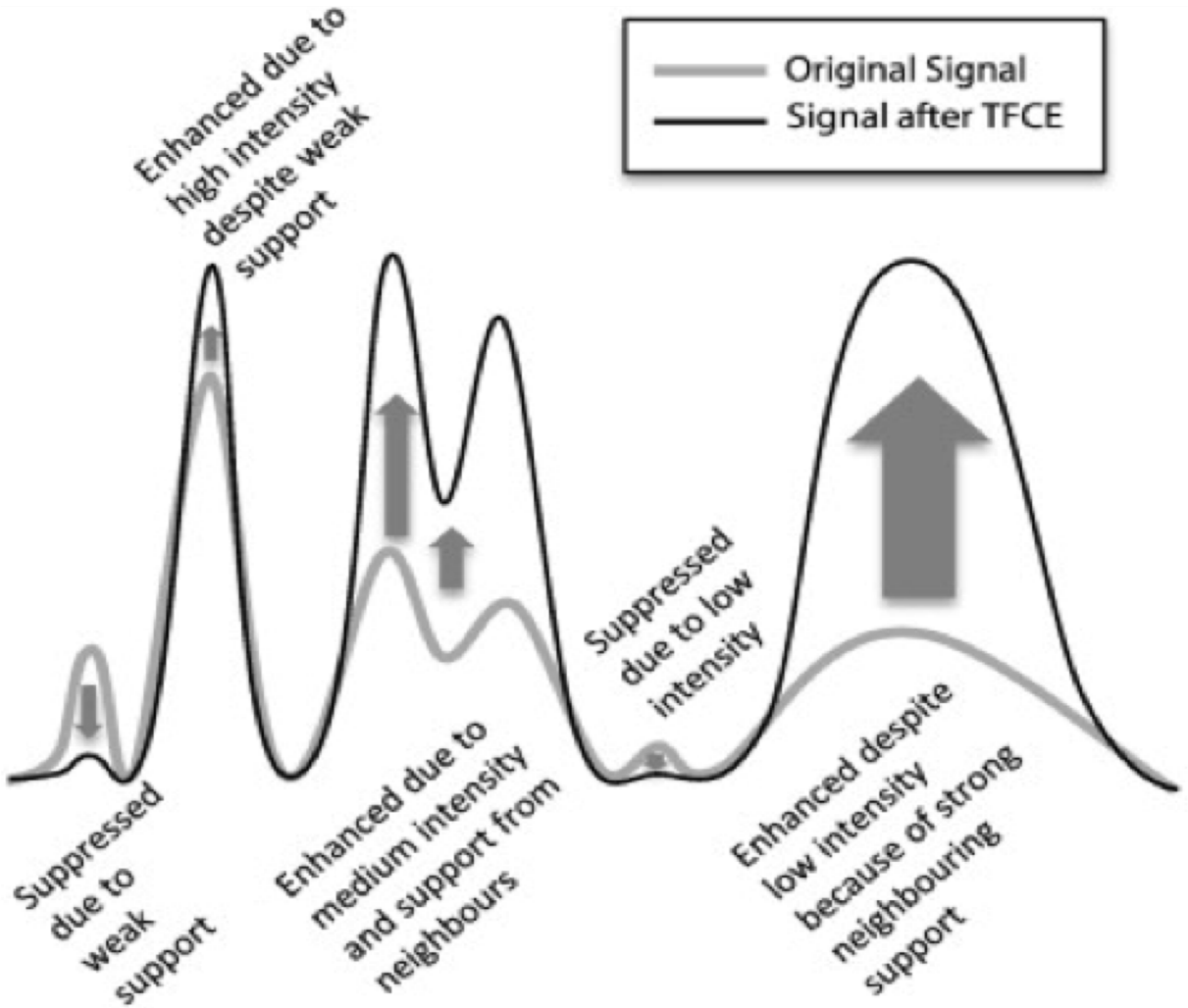


100 gleichverteilte Schwellenwerte für Voxel-Level

Summiere alle Clustergrößen für jeden Schwellenwert

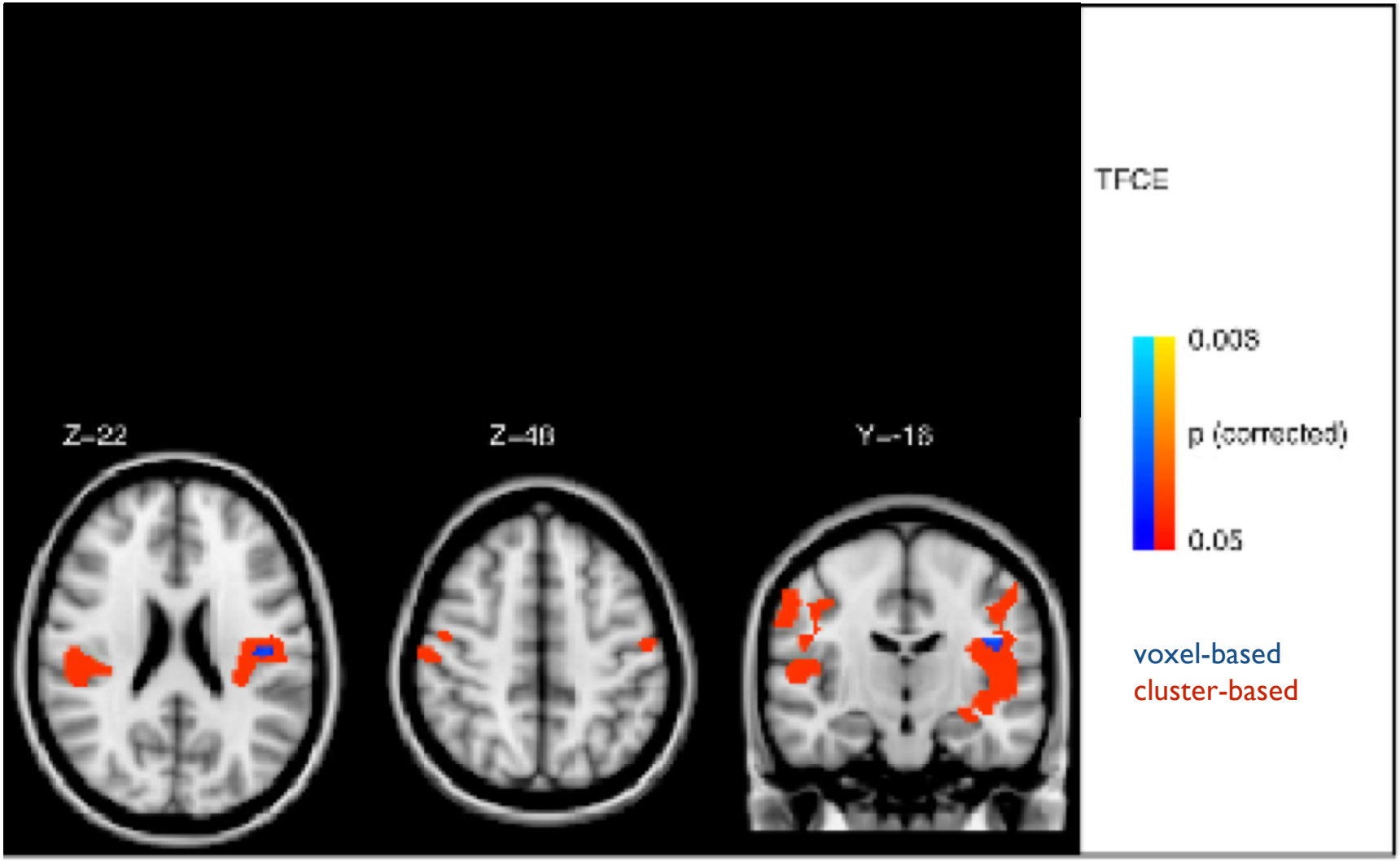
Smith & Nichols
Neuroimage 2009

Threshold-free Cluster Enhancement



Mensen & Khatami
Neuroimage 2013

TFCE vs. Voxel/Cluster-Statistik

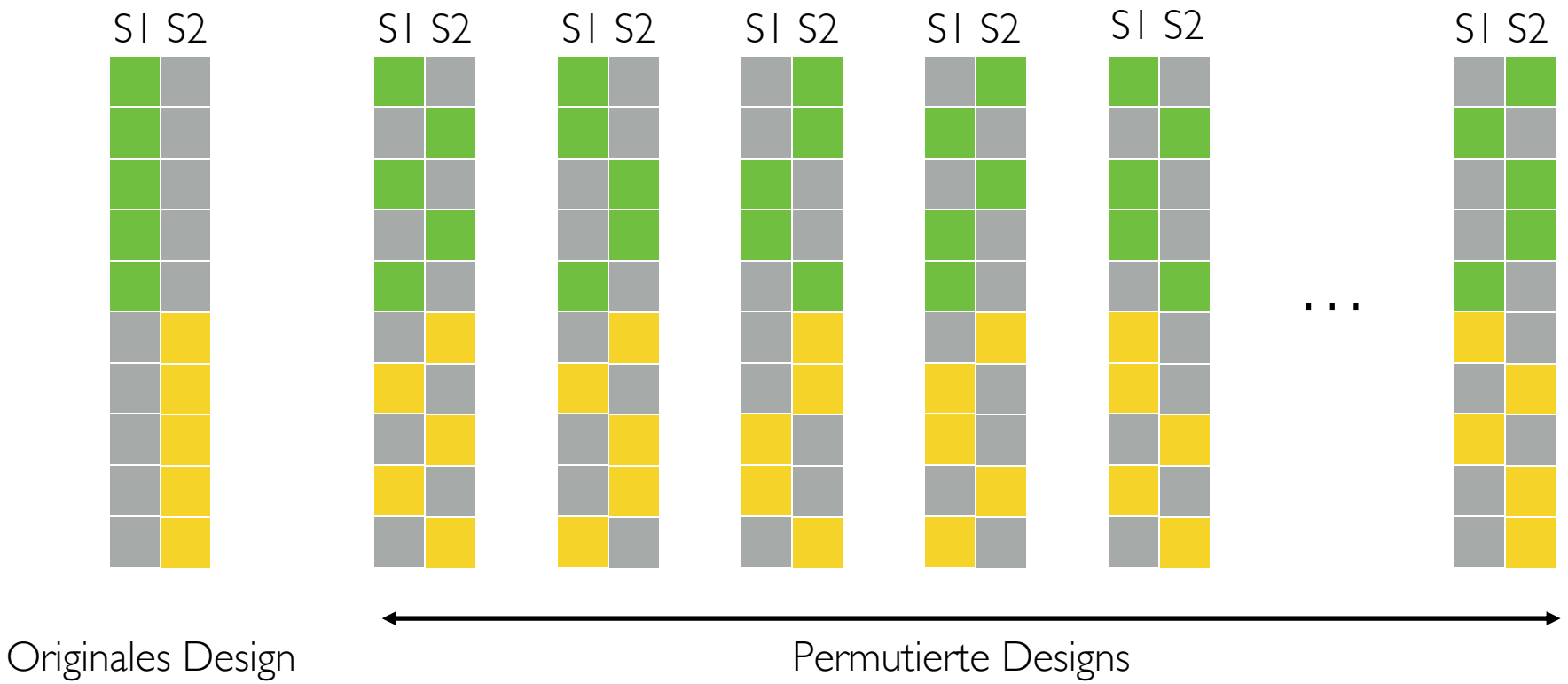


Smith & Nichols
Neuroimage 2009

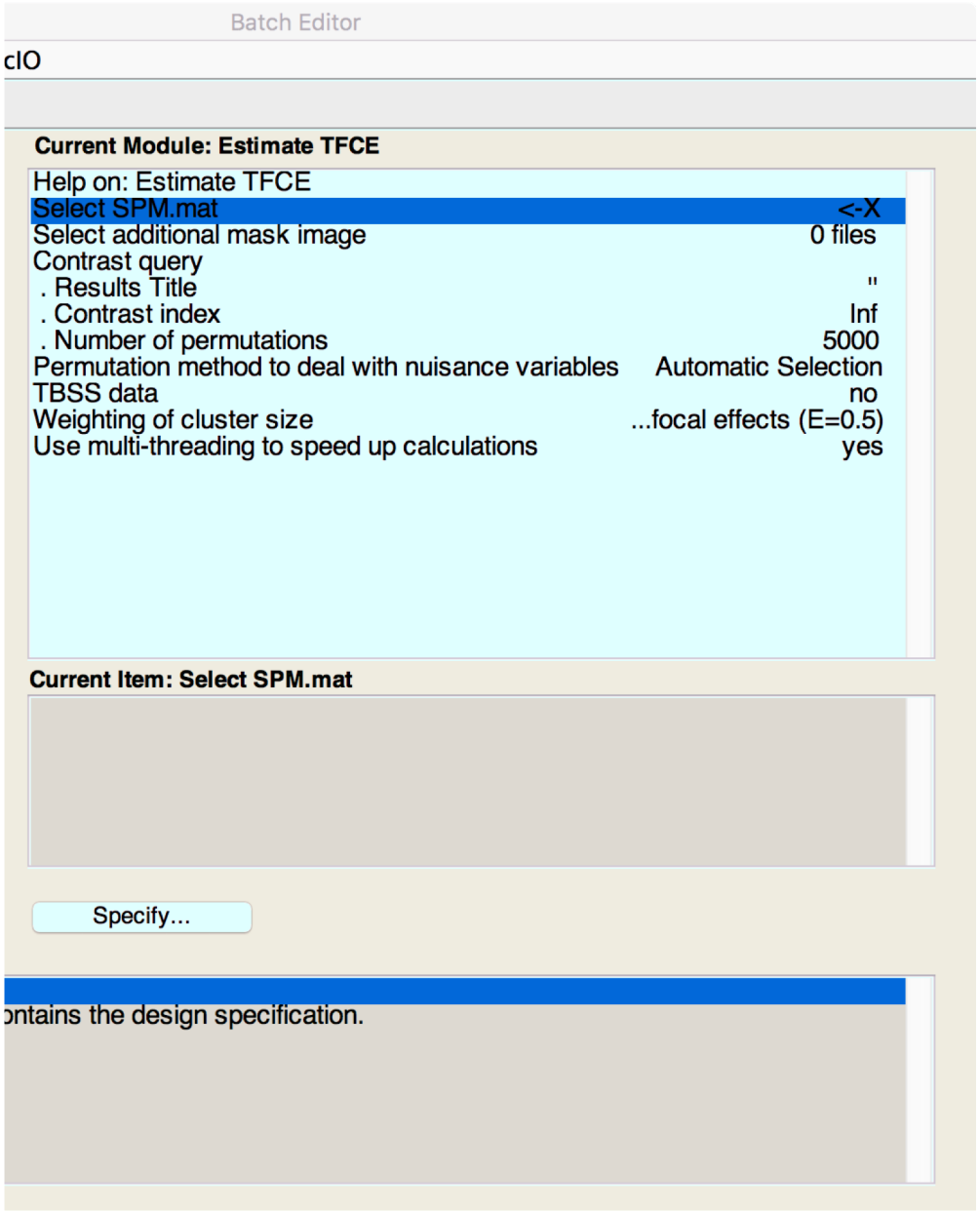
Permutation

Beispiel

- 2-Sample T-Test



TFCE Toolbox



tfce

Threshold-free Cluster Enhancement

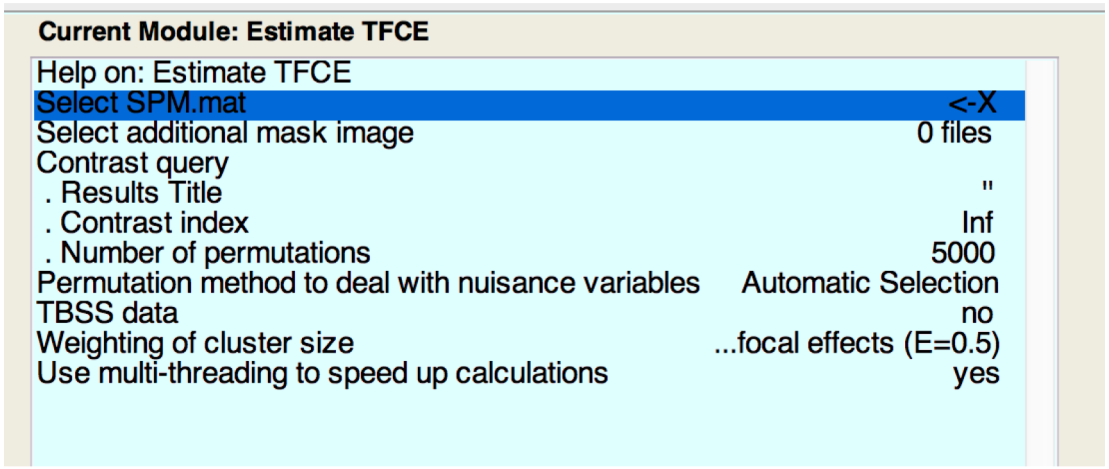
http://dbm.neuro.uni-jena.de/tfce

- Benötigt nur bestehendes SPM-Design aus 2nd Level (SPM.mat)
- Alle statistische Modelle mit differentiellen Kontrasten und Korrelationen möglich (auch Interaktionen)
- Für jeden Kontrast muss Permutation gerechnet werden
- Multiprozessor-Unterstützung

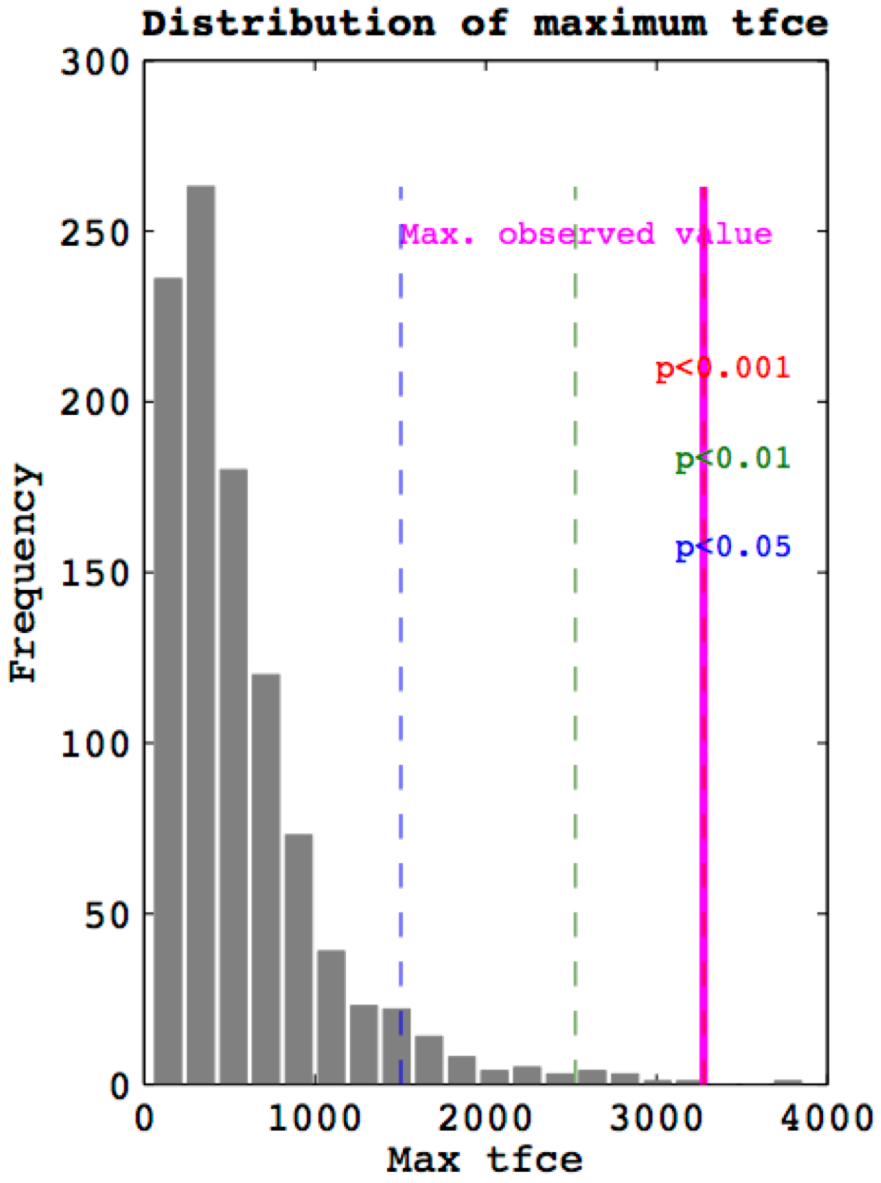
TFCE Parameter

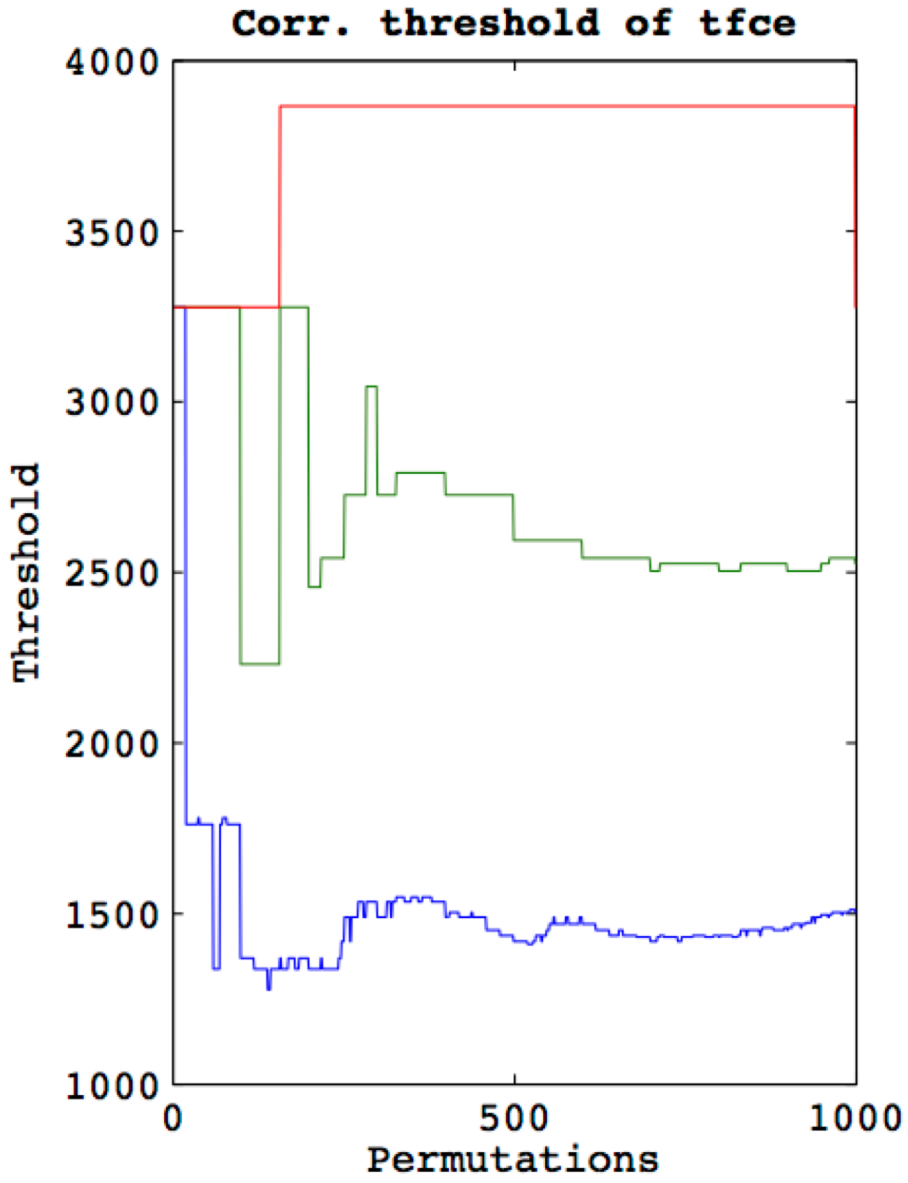
Parameter

- Zusätzliche Maske (SVC)
- Anzahl der Permutationen
- Umgang mit Nuisance-Variablen
- Wichtung Clustergröße

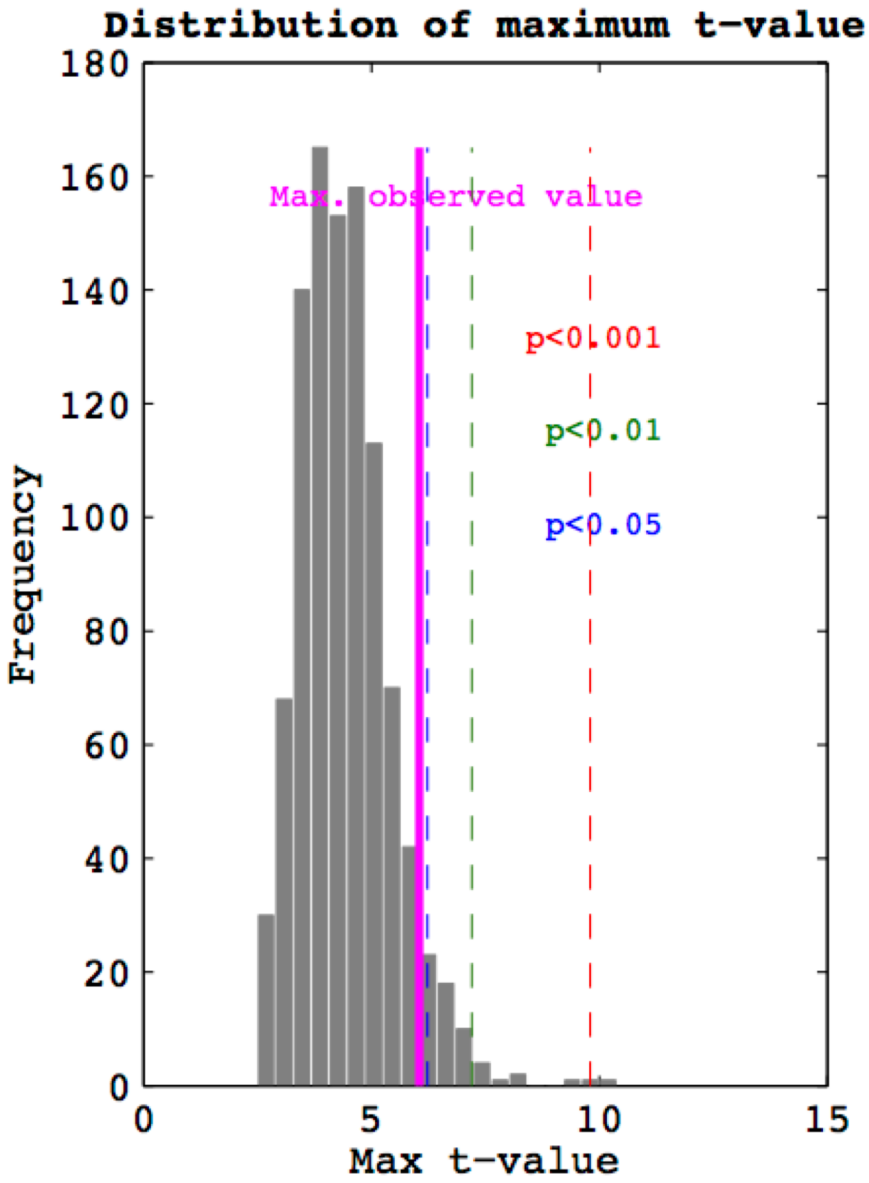


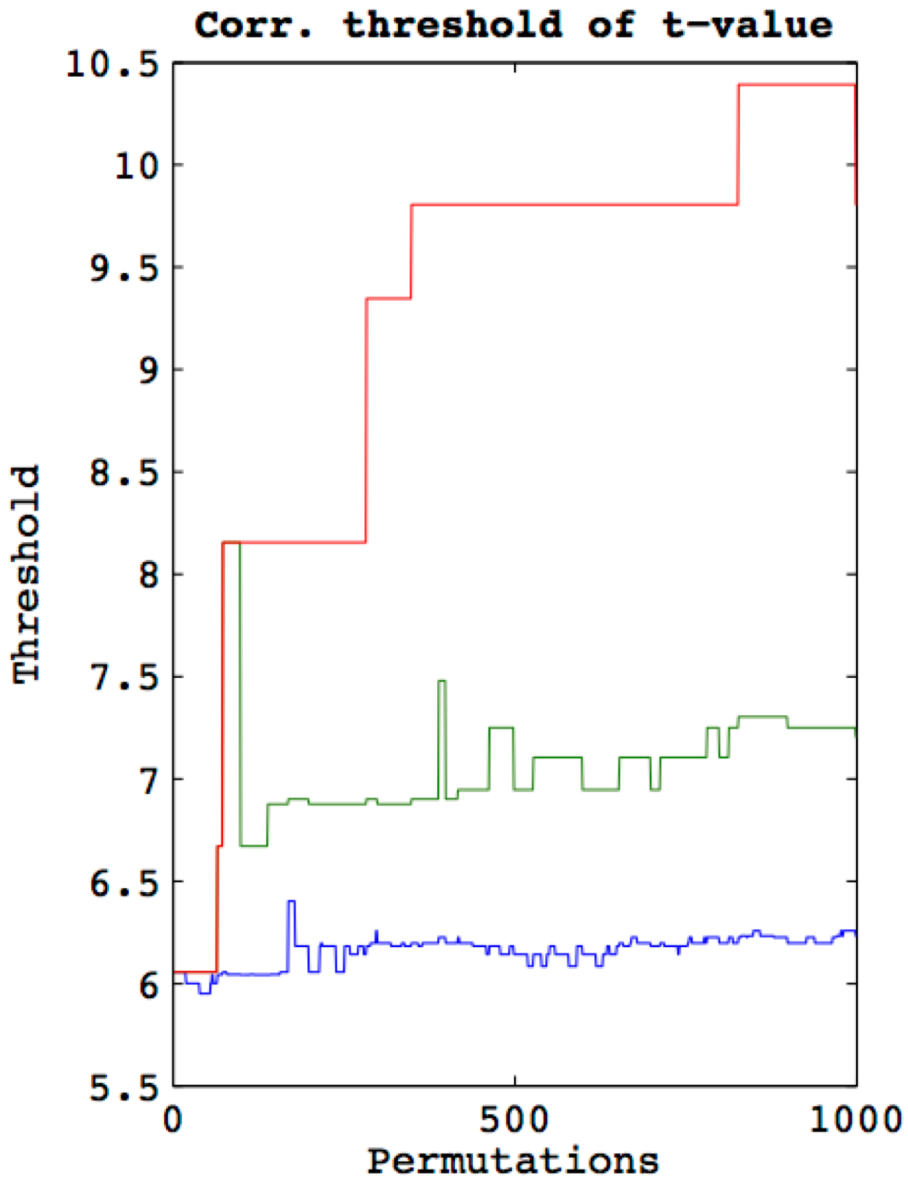
TFCE-Toolbox





TFCE-Toolbox





Eigenschaften

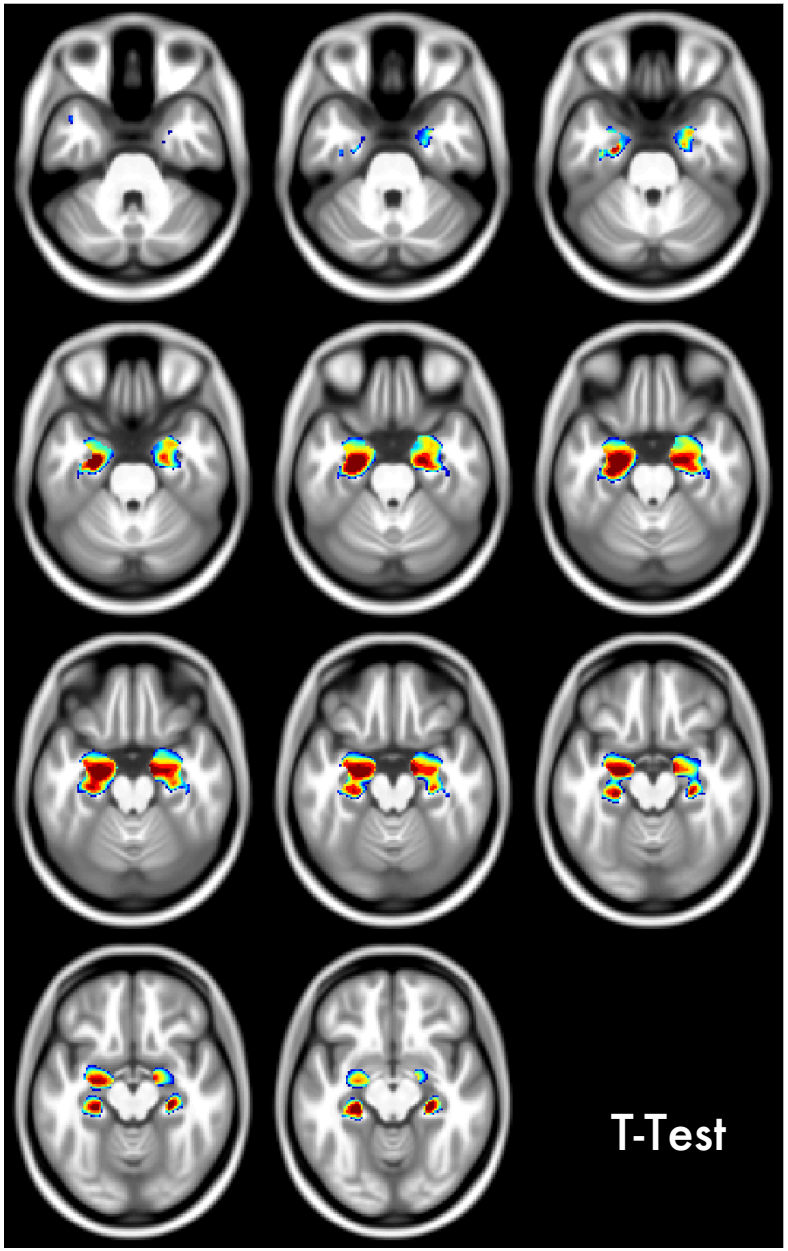
Vorteile

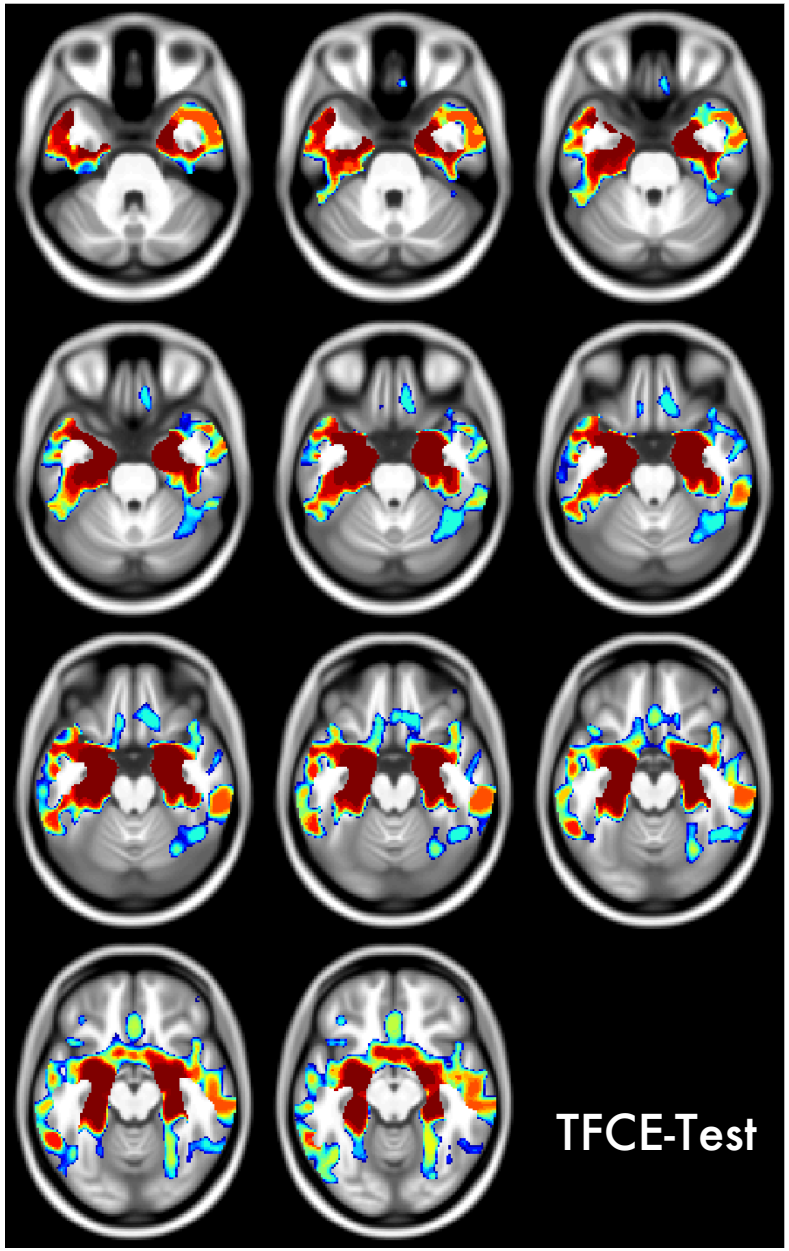
- Schwellenwert-unabhängig („threshold-free“)
- sensitiv sowohl für fokale hohe Effekte und ausgedehnte niedrige Effekte („cluster enhancement“)
- nicht-parametrischer Test: keine Annahme über Verteilung
- keine Probleme mit lokal unterschiedlicher Glätte (non-stationarity cluster inference) bei VBM
- unkorrigierte Werte oder Korrektur für multiple Vergleiche mit FWE oder FDR möglich
- tendenziell kleinere Filterbreiten (FWHM) möglich

Nachteile

- Verteilung muss bestimmt werden, da für TFCE nicht bekannt
- Permutationstest ist relativ zeitaufwendig (# Permutationen 5000-10000)
- Stichprobengröße (n<700 ?) begrenzt

Beispiel: 25 AD<25 CTL (P_{FWE}<0.01)





Beispiel: 25 AD<25 CTL (P_{FWE}<0.01)