Localized Cluster Enhancement: TFCE revisited with valid error control

Samuel Davenport, Wouter Weeda, Thomas E. Nichols, Jelle Goeman

University of California, San Diego

June 4, 2026

False positive control in fMRI

False positives in fMRI

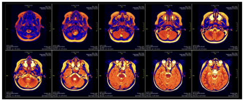
RESEARCH ARTICLE | STATISTICS |

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

Anders Eklund, Thomas E. Nichols, and Hans Knutsson Authors Info & Affiliations

A Bug in FMRI Software Could Invalidate 15 Years of Brain Research

HUMANS 06 July 2016 By BEC CREW Add ScienceAlert on Google



Kondor8/Shutterstock.com

There could be a very serious problem with the past 15 years of research into human brain activity, with a new study suggesting that a bug in fMRI software could invalidate the results of some 40,000 papers.

Eklund, Nichols & Knutsson (2016) showed that cluster-size inference - one of the most widely used fMRI FPR methods - had **false positive rates of around 30%** (at a nominal $\alpha = 5\%$).

Robust FWER control

Robust FWER control in Neuroimaging using Random Field Theory: Riding the SuRF to Continuous Land Part 2

Samuel Davenport, Armin Schwartzman, Thomas E. Nichols and Fabian J.E. Telschow

Part of my PhD work addressed some of these issues and provided robust control using parametric method.

The remainder of this talk concerns the analogous problem for **TFCE** another widely used method.

TFCE recap

Notation

- Let $\mathcal{B}$ denote the set of voxels/vertices making up the brain.
- Suppose that, for each $v \in \mathcal{B}$, X_v are the underlying data corresponding to voxel v. (Containing data from multiple subjects.)
- Suppose that at each v we have a test statistic $T_v = T(X_v)$.

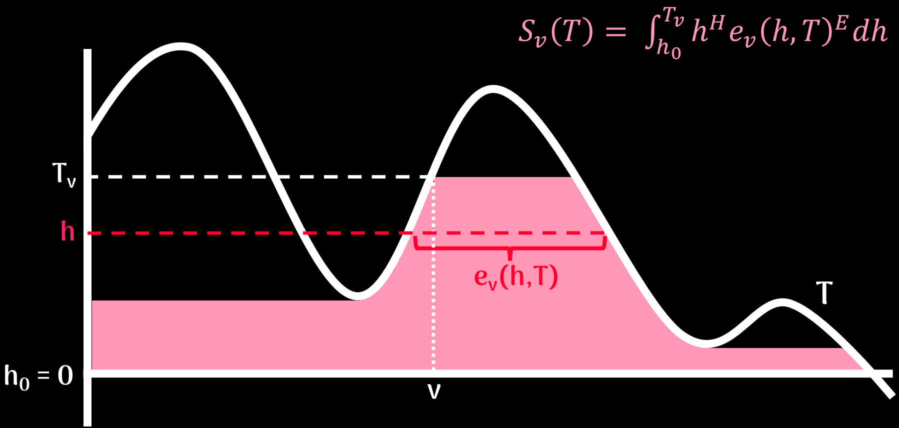
TFCE recap

- TFCE is a widely used method for identifying activation with over 6000 citations.
- At each voxel v, TFCE transforms T to

$$S_v(T) = \int_{h_0}^{T_v} h^H e_v(h, T)^E \mathrm{d}h.$$

We can also write $S_v(T) = S_v(\mathbf{X})$
- Here h is the height and $e_v(h, T)$ is the extent of the test statistic at the level h for the voxel v.

Understanding the TFCE integral



TFCE recap

$$S_v(T) = \int_{h_0}^{T_v} h^H e_v(h, T)^E \mathrm{d}h.$$

- In practice - $H = 2$, $E = 0.5$ and $h_0 = 0$ are the default parameters typically chosen.
- Permuted TFCE test-statistics: $S^*_{v,1}, \dots, S^*_{v,P}$ are calculated.
- Given $\alpha \in (0, 1)$, a TFCE threshold t^* is chosen based on the $(1-\alpha)\%$ quantile of the permutation distribution of $\max_v S^*_{v,p}$. Reject v such that $S_v(T) > t^*$.

TFCE claims

Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference

Stephen M. Smith[a,*] and Thomas E. Nichols[b,a,c]

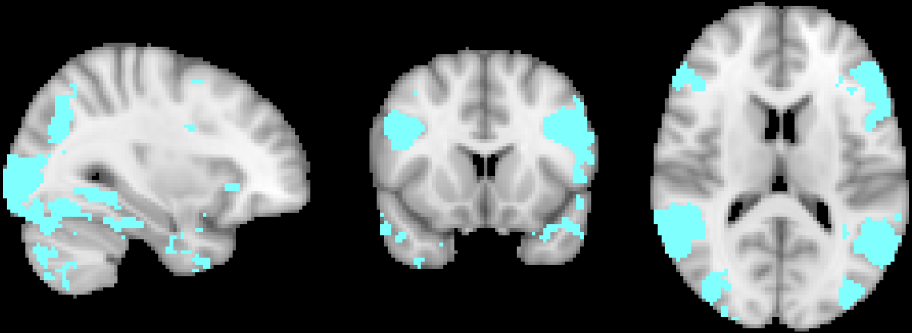
Many image enhancement and thresholding techniques make use of spatial neighbourhood information to boost belief in extended areas of signal. The most common such approach in neuroimaging is cluster-based thresholding, which is often more sensitive than voxel-wise thresholding. However, a limitation is the need to define the initial cluster-forming threshold. This threshold is arbitrary, and yet its exact choice can have a large impact on the results, particularly at the lower (e.g., t, z < 4) cluster-forming thresholds frequently used. Furthermore, the amount of spatial pre-smoothing is also arbitrary (given that the expected signal extent is very rarely known in advance of the analysis). In the light of such problems, we propose a new method which attempts to keep the sensitivity benefits of cluster-based thresholding (and indeed the general concept of "clusters" of signal), while avoiding (or at least minimising) these problems. The method takes a raw statistic image and

holding. For inference, the TFCE image can easily be turned into voxel-wise *p*-values (either uncorrected, or corrected for multiple comparisons across space) via permutation testing.

Unfortunately TFCE does not address these problems and in fact exacerbates them.

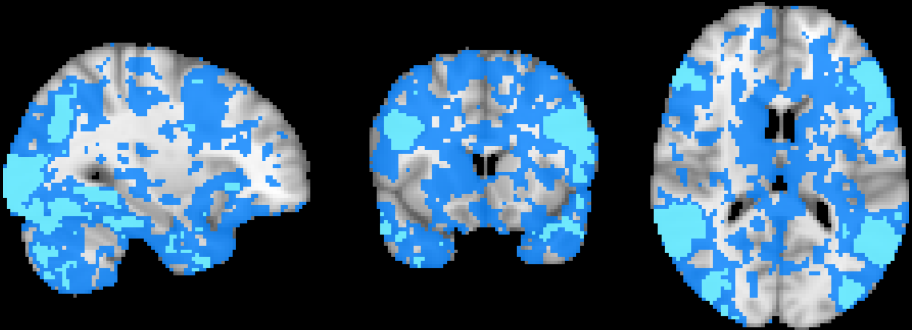
Thresholded TFCE

We apply TFCE to 20 subjects from the HCP to the primary social contrast.



However TFCE borrows information from across its support.

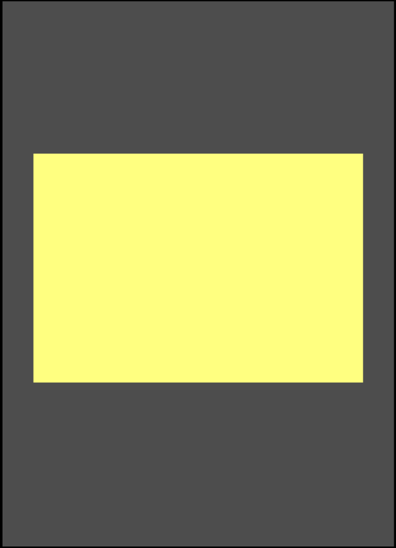
TFCE support



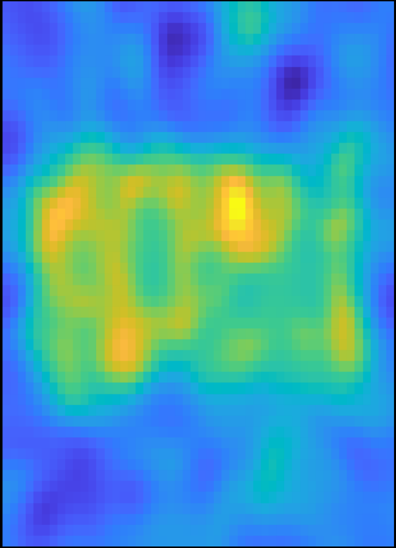
TFCE can be used to make a global statement but struggles to localize activation.

Simulation example (based on $n = 50$)

Signal

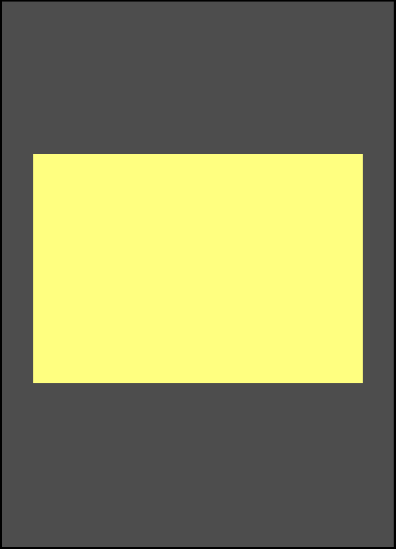


t-statistic

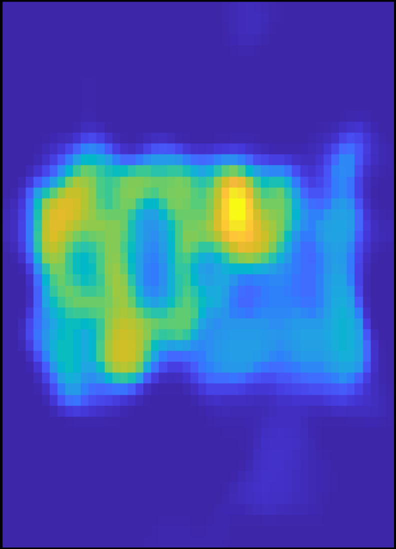


Simulation example (based on $n = 50$)

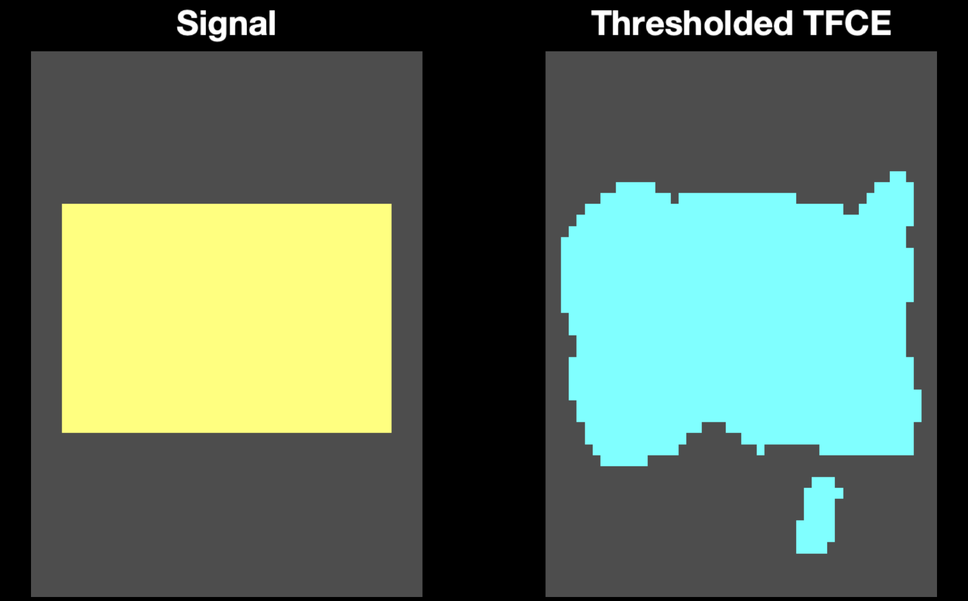
Signal



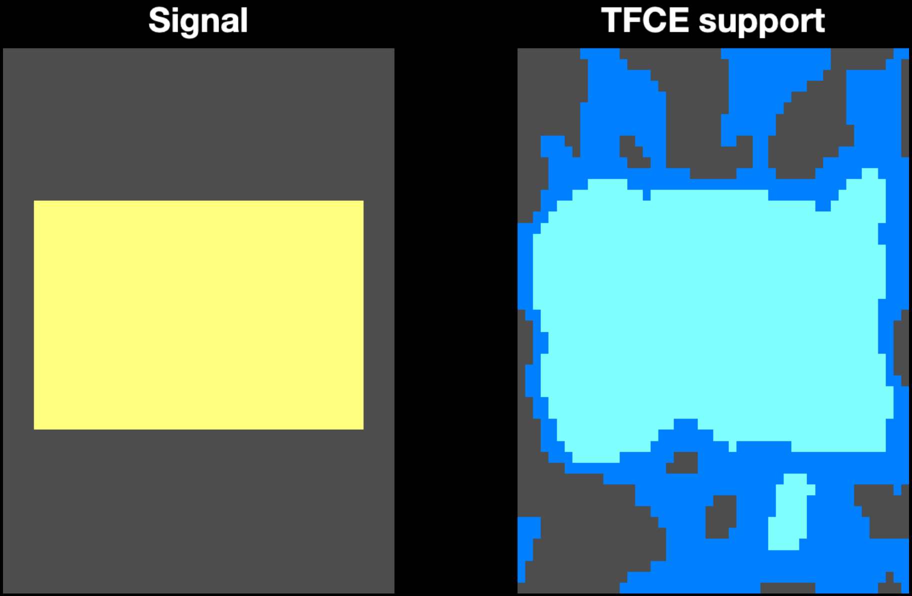
tfce-statistic



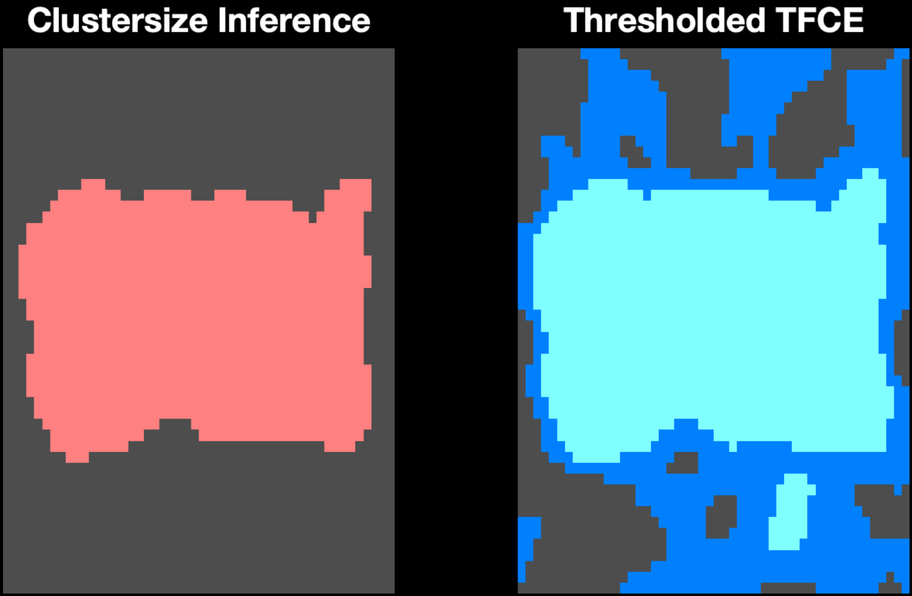
Simulation example (based on $n = 50$)



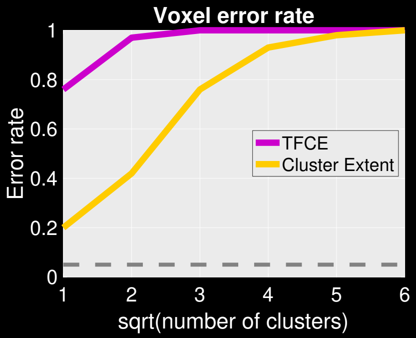
Simulation example (based on $n = 50$)



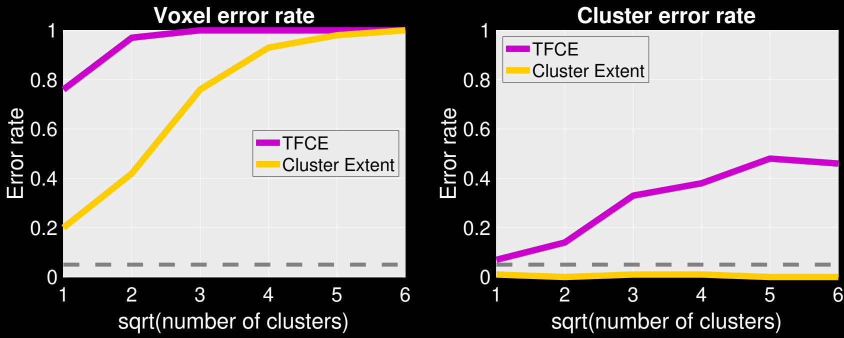
Simulation example (based on $n = 50$)



TFCE can have inflated error rates



TFCE can have inflated error rates



Classifying fMRI inference methods

1. Voxel: Every highlighted voxel is active
Voxelwise inference

2. Cluster: Every highlighted cluster contains at least one active voxel
Clustersize inference

3. Global: There is some voxel active somewhere in the brain.
TFCE

Localized Cluster Enhancement

Generalized TFCE statistic

$$S_v = \int_{h_0}^{\infty} f(h) g\big(e_v(h)\big) \mathrm{d}h. \tag{1}$$

- This reduces to the TFCE-statistic if we take $f(x) = x^H$ and $g(x) = x^E$.
- We recover **cluster-size inference** by taking f a Dirac δ-function shifted to h_0, and g the identity.
- **Cluster-mass inference** can be obtained by taking $f(x) = 1$ and g the identity.
- and **peak height inference** with f the identity and $g(x)$ the indicator of $x > 0$.

Exchangeability Assumption

Let $\mathcal{I} \subseteq \mathcal{B}$ be the collection of all inactive voxels, and Π be the group of permutations considered.

Assumption

For every $\pi \in \Pi$, we have $(\mathbf{X}_v)_{v \in \mathcal{I}} \stackrel{d}{=} (\mathbf{X}_v \circ \pi)_{v \in \mathcal{I}}$.

Moreover let $H_{\mathcal{B}}$ be the null hypothesis that all voxels in $\mathcal{B}$ are inactive.

Localized Cluster Enhancement

LCE embeds TFCE within a **closed testing** procedure, yielding simultaneously valid inference over all regions $\mathcal{R} \subset \mathcal{B}$. I.e. let $\mathbf{X}_\mathcal{R} = (X_v 1[v \in \mathcal{R}])_{v \in \mathcal{B}}$, then

$$\text{LCE}(\mathcal{R}) = \frac{1}{P} \sum_{p=1}^{P} 1 \left[\max_{v \in \mathcal{R}} S_v(\mathbf{X}_\mathcal{R}) \leq \max_{v \in \mathcal{B}} S_v(\mathbf{X} \circ \pi_p) \right]. \qquad (2)$$

Reject $H_\mathcal{R}$ for a region $\mathcal{R} \subseteq \mathcal{B}$ if $\text{LCE}(\mathcal{R}) \leq \alpha$.

Characterizing LCE

Corollary

Let t^* *be the TFCE threshold. Then LCE rejects* $\mathcal{R} \subseteq \mathcal{B}$ *if and only if* $\max_{v\in\mathcal{R}} S_v(\mathbf{X}_{\mathcal{R}}) > t^*$.

LCE shares the TFCE threshold - no extra permutations needed beyond those for TFCE itself.

Simultaneous validity of LCE

Theorem

Under Exchangeability, given $\alpha \in (0,1)$,

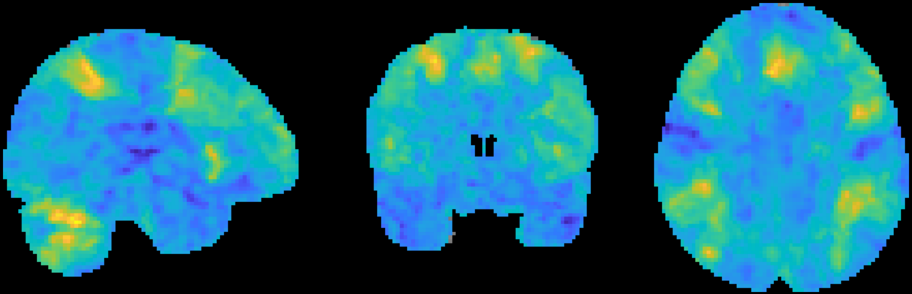
$$\mathbb{P}\big(\exists\, \mathcal{R} \subseteq \mathcal{B} : H_{\mathcal{R}} \textit{ is true and } LCE(\mathcal{R}) \leq \alpha\big) \leq \alpha.$$

Equivalently $\mathbb{P}\left(LCE(\mathcal{R}) \geq \alpha \;\textit{ for all } \mathcal{R} \subseteq \mathcal{I}\right) \geq 1 - \alpha.$

The guarantee is simultaneous over *all* regions $\mathcal{R}$ - so they can be **data-driven**.

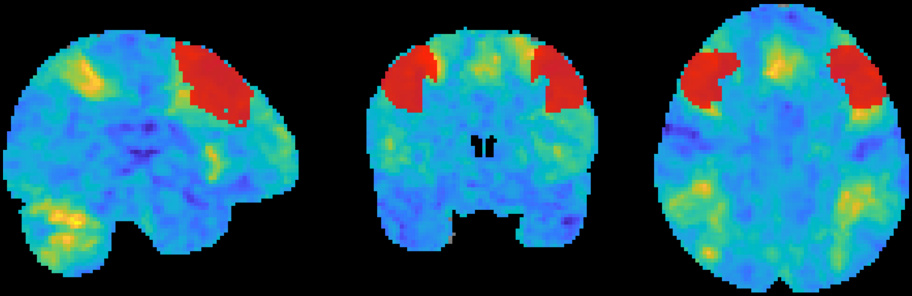
Localized Cluster Enhancement illustration

The t-statistic based on 20 subjects (for the HCP Gambling task)



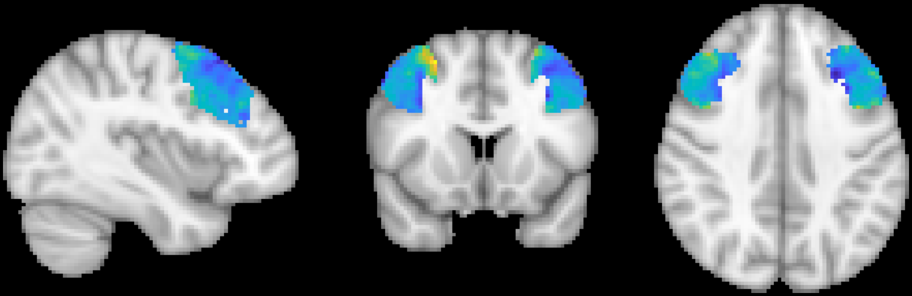
Localized Cluster Enhancement illustration

Apply a mask of the Middle Frontal Gyrus



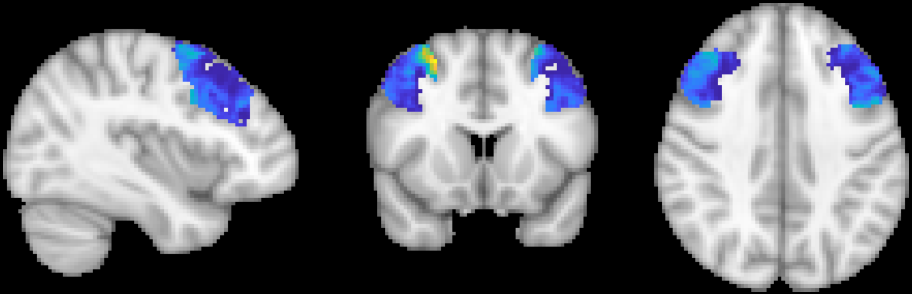
Localized Cluster Enhancement illustration

Localized Cluster Enhancement illustration



Localized Cluster Enhancement illustration

Apply TFCE transformation to obtain $\max_{v \in \mathcal{R}} S_v(\mathbf{X}_{\mathcal{R}})$



Reject if $\max_{v \in \mathcal{R}} S_v(\mathbf{X}_{\mathcal{R}}) > t^*$.

Understanding TFCE via an LCE lense

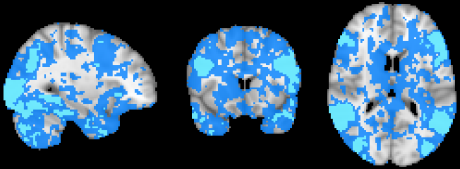
Characterizing the error control of TFCE

Corollary

Let $\mathcal{C}_1, \ldots, \mathcal{C}_m$ *denote the TFCE-significant clusters and let* $supp_{h_0}(\mathcal{C})$ *be the* h_0*-superthreshold connected component containing* $\mathcal{C}$*. Then*

$$\mathbb{P}\left(supp_{h_0}(\mathcal{C}_i) \subseteq \mathcal{I} \text{ for some } 1 \leq i \leq m\right) \leq \alpha.$$

So we can guarantee there are active voxels around a given TFCE cluster. Moreover, h_0 effectively acts as a cluster-defining threshold - just like in clustersize inference.



Every TFCE detection ⇒ an LCE detection

Strengthening the support corollary:

Theorem

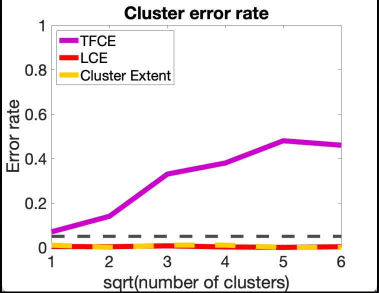
Let $\mathcal{R}$ be a TFCE-significant cluster. Then any h_0-superthreshold cluster $\mathcal{R}' \supseteq \mathcal{R}$ is LCE-significant.

Corollary

The support $supp_{h_0}(\mathcal{R})$ of any TFCE-significant cluster is LCE-significant. In particular, if the TFCE global test rejects then so does the LCE global test.

Upshot: LCE never loses to TFCE - the support regions TFCE actually controls are recovered as LCE rejections, with the added benefit of valid local interpretation.

Empirical error control (simulations)



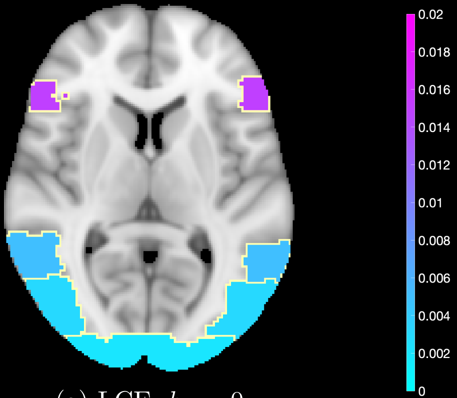
LCE controls the clusterwise error at the nominal $\alpha = 0.05$ (right, red) while TFCE does not.

Advantages of Localized Cluster Enhancement

- If $\max_{v \in \mathcal{R}} S_{v,\mathcal{R}}(T) > t^*$, then LCE claims that there is at least one active voxel within the region $\mathcal{R}$.
- Provably valid simultaneously over all regions $\mathcal{R}$, including data-driven ones.
- **Controls clusterwise error rates** - unlike TFCE.
- Can be applied to pre-defined atlas regions or to clusters identified by TFCE itself.
- Minimal extra computation on top of the TFCE permutation distribution.

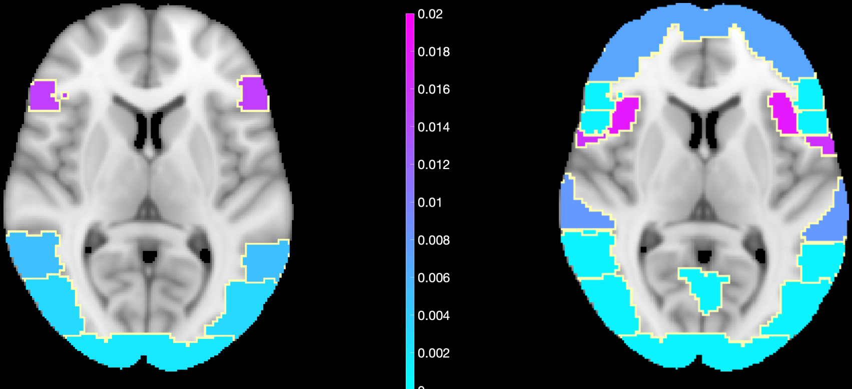
Regional inference on real data

Regional activation (HCP - Social)



(a) LCE, $h_0 = 0$

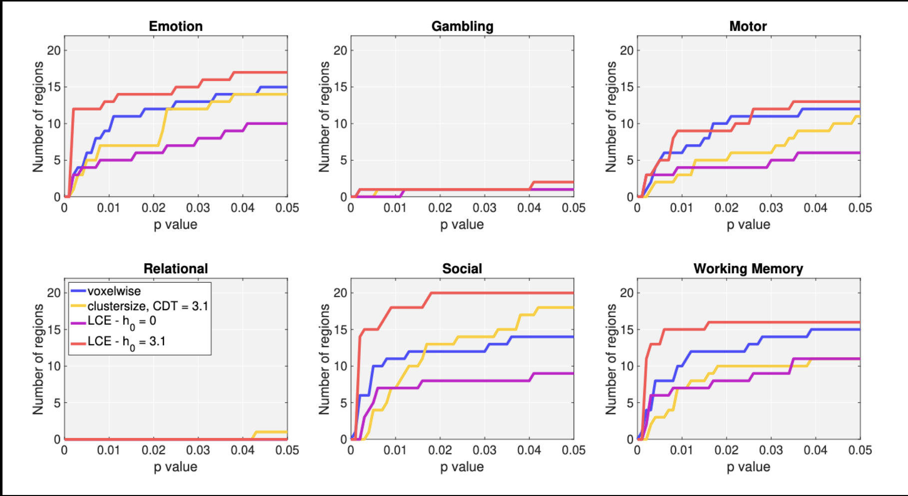
Regional activation (HCP - Social)



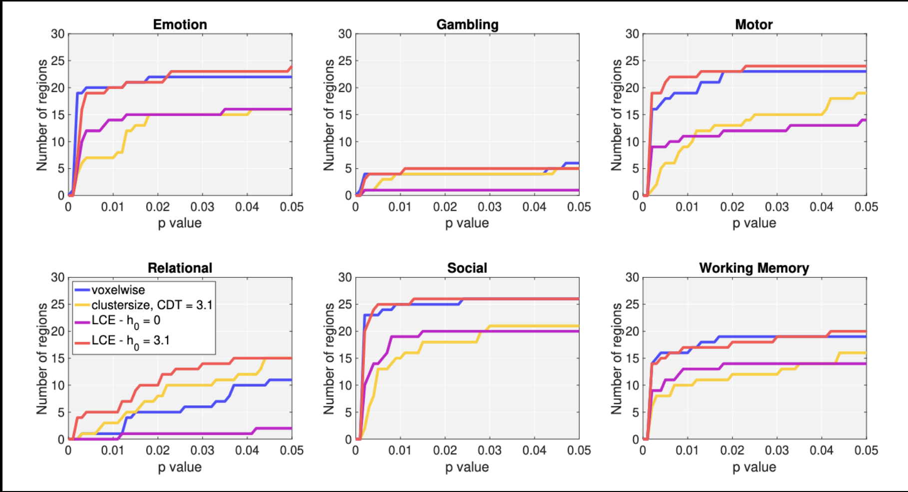
(a) LCE, $h_0 = 0$ (b) LCE, $h_0 = 3.1$

LCE regions at $\alpha = 0.05$: $h_0 = 0$ (magenta) vs. $h_0 = 3.1$ (light blue).

Power on 20 subjects from the HCP



Power on 40 subjects from the HCP



Conclusions

Conclusions

- TFCE has inflated voxel *and* clusterwise error rates - it only provides weak FWER control over h_0-supports.

- **Localized Cluster Enhancement** provably controls regional/clusterwise error rates simultaneously over all (data-driven) regions.

- Every TFCE detection is also an LCE detection - so LCE is a strict improvement.

- TFCE is not threshold free: h_0 acts as a cluster-defining threshold and the default $h_0 = 0$ is empirically suboptimal.

- We recommend $h_0 = 3.1$ for LCE, matching standard cluster size inference - confirmed on HCP (6 tasks, $n \in \{20, 40, 80\}$).

Thanks

- Slides for this talk are available on my website:
`sjdavenport.github.io/talks`

- Code to implement LCE and a tutorial on TFCE are available in the StatBrainz MATLAB package available at:
`sjdavenport.github.io/software`

LCE as an Algorithm

Algorithm 1 Localized Cluster Enhancement

Require: Data $\mathbf{X} = (X_v)_{v\in\mathcal{B}}$, real-valued functions f, g, a threshold $h_0 \in \mathbb{R}$, a set of n regions of interest $\{\mathcal{R}_i\}_{i=1}^n \subseteq \mathcal{B}$, a number of permutations P, and a collection of permutations Π.

Compute the original test-statistics $\{T_v\}_{v\in\mathcal{B}}$.
Compute the transformed statistics $S_v(\mathbf{X}) = \int_{h_0}^{\infty} f(h)g\big(e_v(h, \mathbf{X})\big)\mathrm{d}h$, for each $v \in \mathcal{B}$.
Draw permutations $\pi_1, \dots, \pi_P$ independently from Π.
for $p = 1$ to P **do**
    Compute the permuted transformed map: $S_v(\mathbf{X}\circ\pi_p) = \int_{h_0}^{\infty} f(h)g\big(e_v(h, \mathbf{X}\circ\pi_p)\big)\mathrm{d}h$, for $v \in \mathcal{B}$.
end for
for $i = 1$ to n **do**
    Mask the data to the region $\mathcal{R}_i$, to compute $\mathbf{X}_{\mathcal{R}_i} = (X_v 1[v \in \mathcal{R}_i])_{v\in\mathcal{B}}$.
    Compute the masked transformed map $S_v(\mathbf{X}_{\mathcal{R}_i}) = \int_{h_0}^{\infty} f(h)g\big(e_v(h, \mathbf{X}_{\mathcal{R}i})\big)\mathrm{d}h$.
    Calculate $\mathrm{LCE}(\mathcal{R}_i) = \frac{1}{P}\sum_{p=1}^{P} 1\left[\max_{v\in\mathcal{R}_i} S_v(\mathbf{X}_{\mathcal{R}_i}) \leq \max_{v\in\mathcal{B}} S_v(\mathbf{X}\circ\pi_p)\right]$.
end for
return $\{\mathrm{LCE}(\mathcal{R}_i)\}_{i=1}^n$: the set of LCE-adjusted p-values.
given $\alpha \in (0,1)$, reject $H_{\mathcal{R}_i}$ for each $1 \leq i \leq n$ such that $\mathrm{LCE}(\mathcal{R}_i) \leq \alpha$.

Voxelwise inference via LCE

Voxelwise statements using LCE

Given $v \in \mathcal{B}$, taking $\mathcal{R} = \{v\}$:

$$S_v(\mathbf{X}_{\{v\}}) = \int_{h_0}^{T_v} f(h)g(1)\, dh, \tag{3}$$

since $e_v(h) = 1[h_0 \leq h \leq T_v]$.
In the default TFCE setting, $f(h) = h^H$ and $g(1) = 1$, we have

$$S_v(\mathbf{X}_{\{v\}}) = \int_{h_0}^{T_v} h^H\, dh = \frac{1}{H+1}(T_v^{H+1} - h_0^{H+1}).$$

As such $T_v \geq (t^*(H+1) + h_0^{H+1})^{\frac{1}{H+1}}$ implies v is voxelwise significant with strong FWER control.

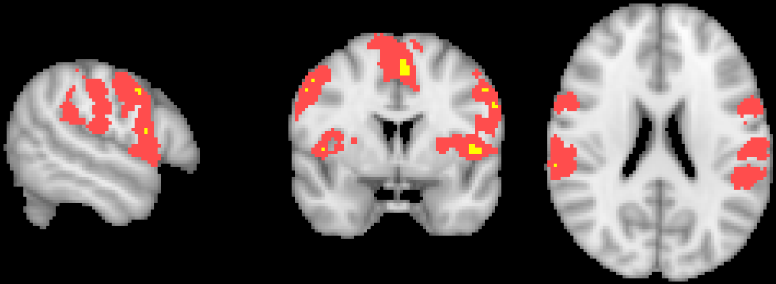
Voxelwise statements using LCE - cont

$$S_v(\mathbf{X}_{\{v\}}) = \frac{1}{H+1}(T_v^{H+1} - h_0^{H+1}).$$

With $H = 2$ and $h_0 = 0$, $S_v(\mathbf{X}_{\{v\}}) = T_v^3/3$. So we may call all voxels such that $T_v > (3t^*)^{1/3}$ voxelwise significant - with strong FWER control.

In practice this can be very high so is mainly of theoretical interest for $h_0 = 0$. It is more useful at higher h_0.

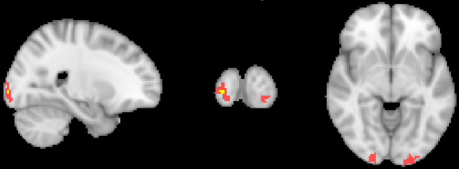
Voxelwise inference on the HCP (Motor, $n = 80$)



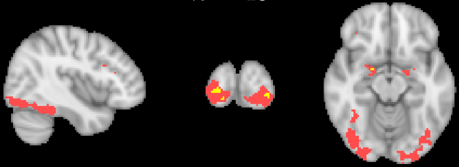
LCE-derived voxelwise thresholds compare favourably to standard voxelwise inference while remaining jointly valid with the regional LCE statements.

Voxelwise inference, HCP Emotion

$n = 20$

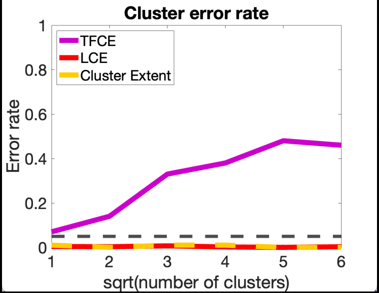


$n = 40$



$n = 80$

Empirical error control (simulations)



At the voxel level (left) all three methods inflate - though there is a version of LCE