Random Field Theory

Samuel Davenport

University of Oxford

May 22, 2023

1. Voxelwise RFT
2. Clusterwise Inference
3. Real Data Validation
4. Fixing the problems with voxelwise RFT inference

References38

Robert Adler - RFT

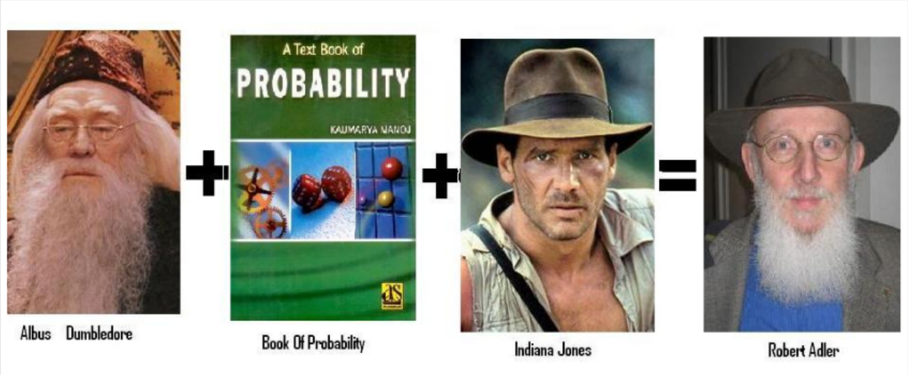


Figure 1: Robert Adler and others developed RFT in books such as the Geometry of Random Fields.

Neuroimagers - RFT

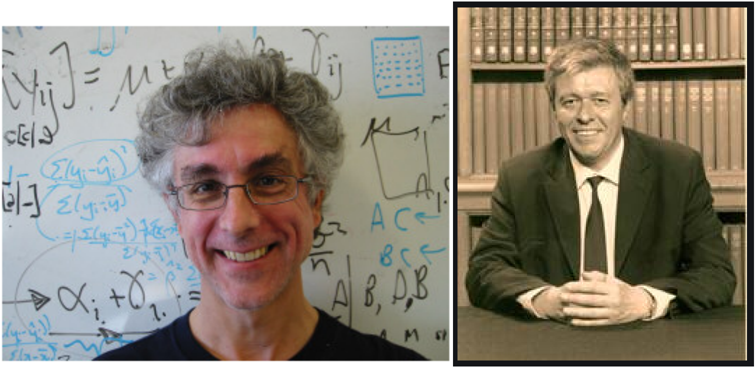


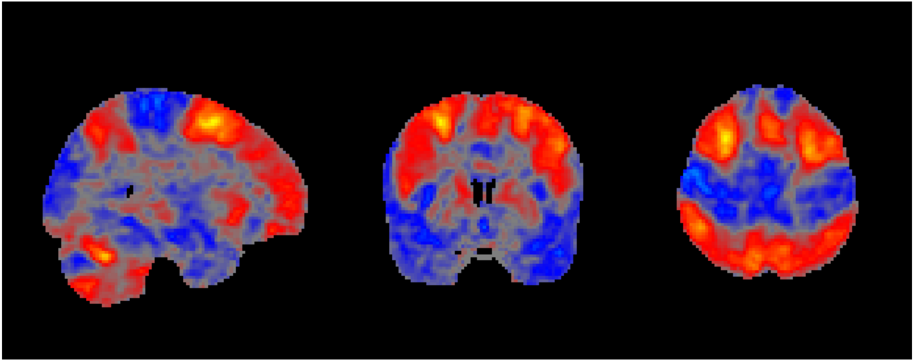
Figure 2: Keith Worsley (left) and Karl Friston (right) introduced RFT methods for determining the threshold.

Neuroimagers - Permutation

Figure 3: Thomas Nichols (left) and others developed non-parametric methods for performing inference. Recently Anders Eklund (right) and Tom showed that RFT (as currently implemented) does not control the false positive rate while permutation does.

Brain Imaging

Bain Imaging

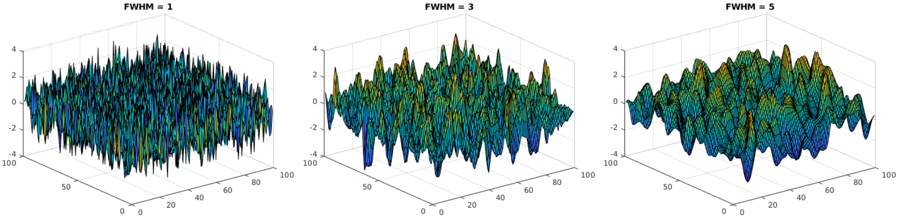


Definition

Given $D \in \mathbb{N}$ and $S \subset \mathbb{R}^D$, define an D-dimensional random field X to be a random function

$$Y : S \longrightarrow \mathbb{R}$$

we say that Y is a Gaussian random field if for all $k \in \mathbb{N}$, given $(t_1, \dots, t_k) \in S$, $(Y(t_1), \dots, Y(t_k))$ has a non-degenerate Gaussian distribution.

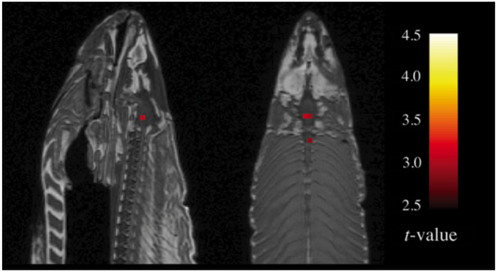


Voxelwise RFT

Multiple Testing

- Let $V \subset S$ be the set of voxels
- Take $Y(v)$ to be our test statistic at each $v \in V \subset T$ so we reject the null hypothesis that there is no activity at v if $\mathbb{P}(Y(v) > u) < \alpha$.

There are typically large number (around 200000) of voxels. In particular taking $\alpha = 0.05$ will mean around 10000 false discoveries! One amusing paper tried scanning a dead salmon to see what would be without without multiple testing correction.



Voxelwise Inference; Controlling the FWER

Definition

Suppose that $V_0 \subset V$ is the set of voxels that are null. Then we define the FWER (family wise error rate) to be the probability of at least one false discovery. I.e.

$$\mathbb{P}\left(\max_{v \in V_0} Y(v) > u\right)$$

and we seek to control this at a level α.

Note that for a fine enough lattice V,

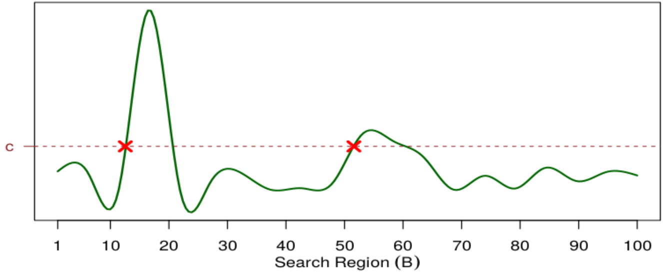
$$\mathbb{P}(\max_{v \in V_0} Y(v) > u) \approx \mathbb{P}(\max_{t \in S} Y(t) > u).$$

Voxelwise RFT

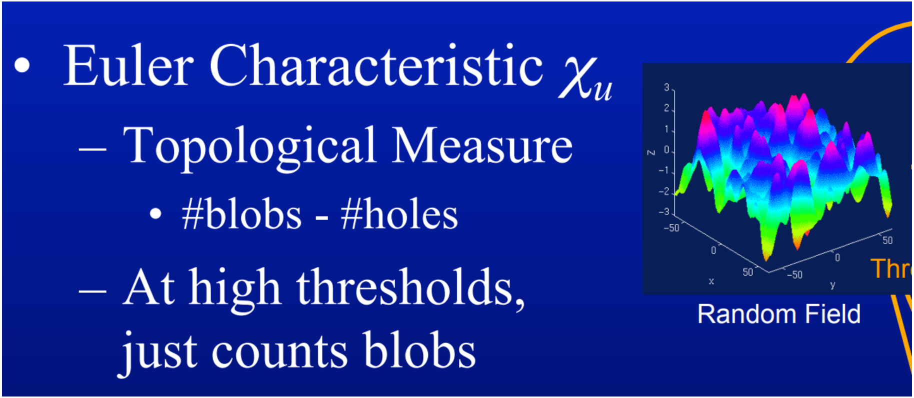
Let M_u be the number of local maxima of X above u then assuming that Y is twice differentiable,

$$\mathbb{P}\left(\sup_{t\in T} Y(t) > u\right) = \mathbb{P}(M_u \geq 1) \leq \mathbb{E}[M_u].$$

because Y exceeds u if and only if there is at least one local maxima above u. This is best seen by looking at a picture

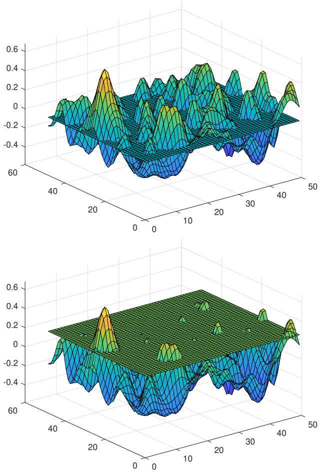


$\mathbb{E}[M]$ is difficult to estimate and requires us to be clever.
To do so we introduce a topological quantity called the Euler Characteristic χ_u which looks at the excursion set which in 2D calculates the number of blobs minus the number of holes.



The Euler Characteristic approximation

When there are no holes the Euler Char is the number of connected components i.e. clusters. At high thresholds it equals the number of local maxima.



Calculating the Expected Euler characteristic

Happily there is a closed form. In particular:

$$\mathbb{E}[\chi(\mathcal{A}_u)] = \sum_{d=0}^{D} \mathcal{L}_d \rho_d(u).$$

When $D = 3$,

$$\rho_0(u) = 1, \rho_1(u) = e^{-u^2}, \rho_2(u) = ue^{-u^2}, \rho_3(u) = u^2 e^{-u^2}.$$

and $\mathcal{L}_d$ are constants called the LKCs.

Estimating the LKCs

Happily $\mathcal{L}_D$ and $\mathcal{L}_{D-1}$ have closed forms. In particular,

$$\mathcal{L}_D = \int_S \det(\Lambda(t))^{1/2}\, ds$$

where $\Lambda(t) = \mathrm{cov}(\nabla(Y(t)/\sigma(t)))$. Note that if we assume stationarity,

$$\mathcal{L}_D = \det(\Lambda)^{1/2}|S|$$

we recover the stationary formula. Moreover

$$\mathcal{L}_0 = \chi(S)$$

i.e. the Euler characteristic of the domain. So if we can estimate Λ then we are able to easily estimate all the LKCs in 1D and 2D and all except $\mathcal{L}_1$ in 3D.

Using the Euler Characteristic

Definition

Given $u \in \mathbb{R}$, let $\mathcal{A}_u$ be the excursion set when the threshold is u ie

$$\mathcal{A}_u = \{t \in S : Y(t) \geq u\}.$$

Then for high thresholds, $M_u = \chi(\mathcal{A}_u)$. So

$$\mathbb{P}\left(\sup_{t \in S} Y(t) > u\right) = \mathbb{P}(M \geq 1) \leq \mathbb{E}[M_u] \approx \mathbb{E}[\chi(\mathcal{A}_u)].$$

Note that for large enough u

$$\mathbb{P}(M_u \geq 1) \approx \mathbb{E}[M_u].$$

Clusterwise Inference

Notation

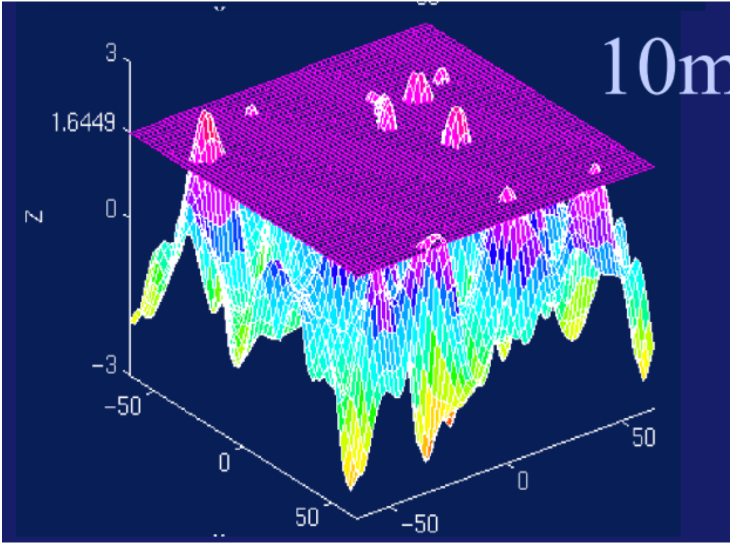
We typically take S to be compact (i.e the brain!) and given a threshold u

- let m be the number of clusters above u where the clusters are the connected components of the excursion set $\mathcal{A}_u$ and
- let $n_1, \dots, n_m$ be the number of voxels within each cluster.

note that $m, n_1, \dots, n_m$ are all random variables size the field X is random.

Note that the number of clusters is not in general equal to the number of local maxima but they are equal for large enough u. (I.e. $M \approx m$ for large enough u as you have one peak per cluster.)

Illustration in 2D



Assumptions of cluster inference

The use of clusterwise RFT in brain imaging makes a number of assumptions (which are not required for voxelwise RFT). These are listed below

- The random field Y is stationary.
- m the number of clusters above u has a Poisson distribution.
- The threshold u is reasonably high.
- The sizes of the clusters: $c_1, \dots, c_m$ are independent with common distribution denoted by c and are independent of m. (This is certainly not true but may be reasonable for high thresholds.)
- The lattice approximation to the random field is good enough. (Good Lattice Assumption.)

Distribution of the size of a cluster

Let $c_1, \ldots, c_m \sim c$ be the sizes of each of the m clusters. And let $c_{\max} = \max\{c_1, \ldots, c_m\}$ then (Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994),

Theorem

Suppose that the assumptions above hold, then

$$\mathbb{P}(c_{\max} \geq k) = 1 - e^{-\mathbb{E}m\mathbb{P}(c \geq k)}$$

Note that this is important as it allows us to perform FWER control on the cluster sizes. I.e if we take our test statistics to be $c_1, \ldots, c_m$ then we can choose k such that

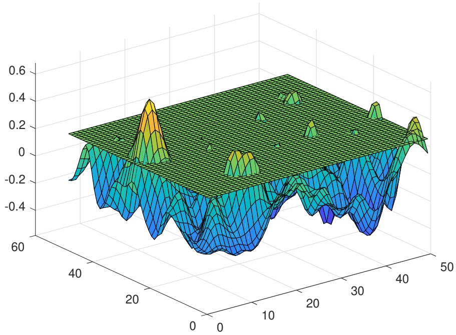
$$\mathbb{P}(c_{\max} \geq k) \leq \alpha$$

and reject cluster j if $c_j > k$.

Estimating the expected number of clusters

We can estimate the number of clusters above the threshold using the Euler characteristic. As at reasonable thresholds the Euler characteristic is the number of clusters.
As such,

$$\mathbb{E}(m) = \mathbb{E}(\#\text{of clusters}) \approx \mathbb{E}(\chi(\mathcal{A}_u)).$$



Real Data Validation

Testing using resting state data

Eklund 2016 validates RFT inference using resting state simulations.

- Data from 198 subjects
- They fit fake block designs to the time series to simulate the noise
- They then resampled from this data with repetition to create null datasets.

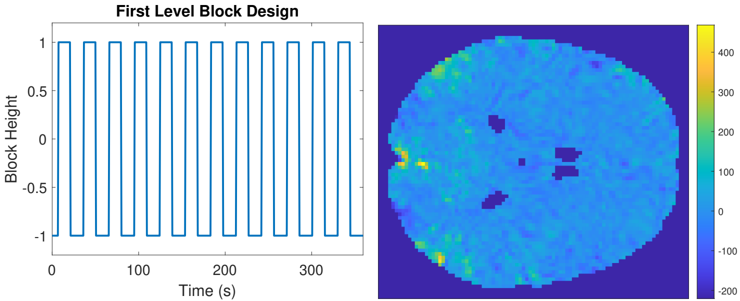


Figure 4: Example block design

Cluster Failure - Clusterwise

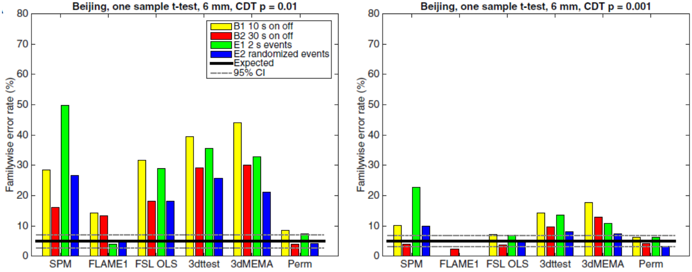


Figure 5: Results from (Eklund et al., 2016), preprocessing by software

Note to form their threshold

- p = 0.01 corresponds to a CDT of $u = 1 - \Phi^{-1}(0.01) \approx 2.32$
- p = 0.001 corresponds to a CDT of $u = 1 - \Phi^{-1}(0.001) \approx 3.09$

Assumptions Breakdown - Smoothness

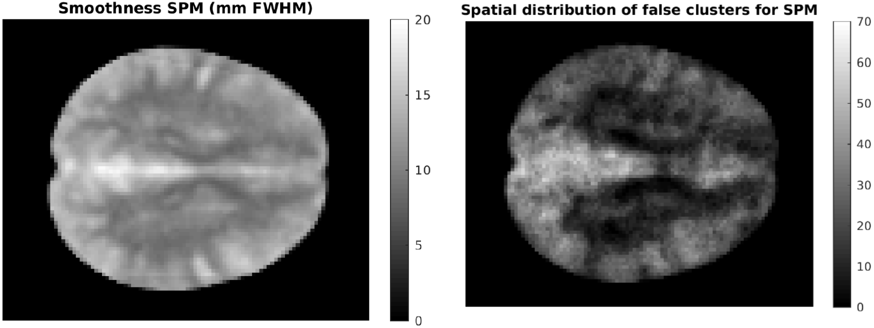


Figure 6: Smoothness is assumed to be the same everywhere (by stationarity) but in fact it's not. This leads to false clusters being detected in smooth regions since the smoother the region the greater the expected clustersize. Clearly a big problem.

General Notes

- Big impact paper, published 5 years ago and already cited 3500 times.
- Essentially what these results show is that applying RFT when the assumptions don't hold won't work. Which shouldn't be a big surprise.
- Permutation is not immune to issues e.g. Eklund 2018
- RFT methods need to be improved in order to use in practice.

This is an extremely important paper.

Cluster Failure - Voxelwise

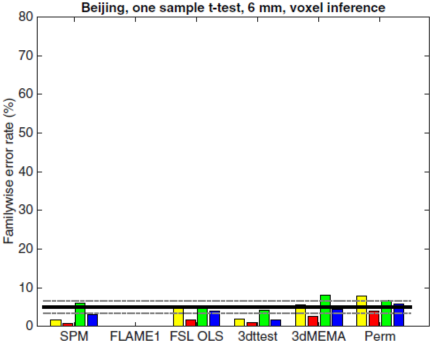


Figure 7: Results from (Eklund et al., 2016), preprocessing by software

Failure of the good lattice assumption

The good lattice assumption states that the continuous random field well approximates the field on a lattice.
The trouble is that if $V \subset S$ then

$$\sup_{s \in V} Y(s) < \sup_{s \in S} Y(s)$$

So for $u \in \mathbb{R}$,

$$\mathbb{P}\left(\sup_{s \in V} Y(s) > u\right) < \mathbb{P}\left(\sup_{s \in S} Y(s) > u\right)$$

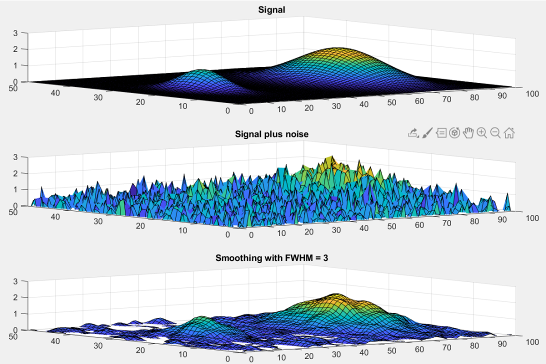
When the field is smooth,

$$\mathbb{P}\left(\sup_{s \in V} Y(s) > u\right) \approx \mathbb{P}\left(\sup_{s \in S} Y(s) > u\right)$$

Fixing the problems with voxelwise RFT inference

Why you should smooth

In order to fix this, we first note that smoothing the data is essential in fMRI.



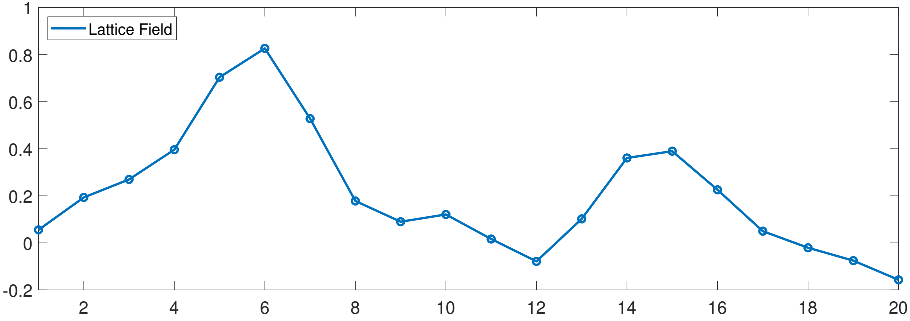
In order for the good lattice assumption to be satisfied RFT has historically required a high level of applied smoothing.

Lattice smoothing

To understand how smoothing works in fMRI, let $X(l)$ be random at every point l of a lattice L. Then smoothing X with a kernel K gives

$$Y(v) = \sum_{l \in L} K(v-l)X(l)$$

at every voxel $v \in L$. Y is plotted below. Y are the fields that have typically been used in fMRI but these are not continuous random fields!

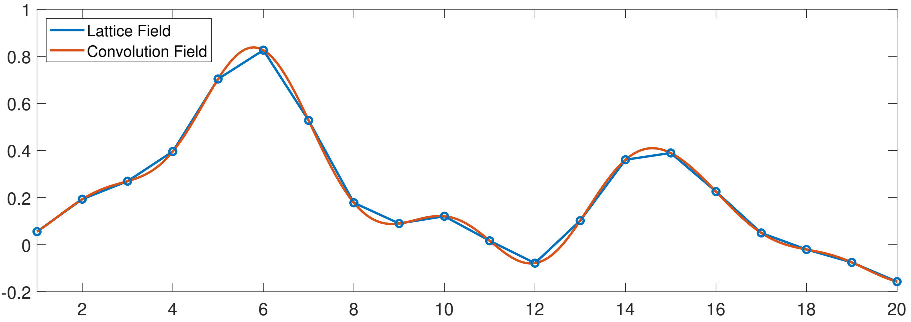


Super resolution fields (SuRFs)

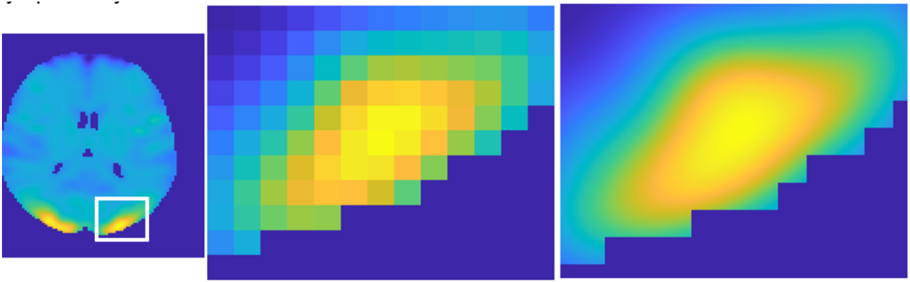
Definition

Given random data X on a lattice $L \subset \mathbb{R}^D$ for $s \in \mathbb{R}^D$ and some kernel K, define the SuRF $Y : \mathbb{R}^D \to \mathbb{R}$, s.t. for all $s \in S$,

$$Y(s) := (K \star X)(s) = \sum_{l \in L} K(s-l)X(l).$$



Taking slices through a 3D SuRF generated from brain imaging data, you get the following images!



Non-Stationary Simulations

We generate 3D non-stationary random fields and test the FWER using the SuRF approach.

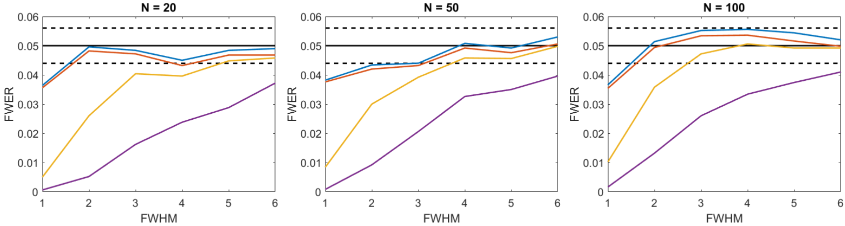


Figure 8: FWER control. Blue: Expected Euler characteristic, Red: SuRF coverage, Yellow: resolution one lattice, Purple: Traditional RFT - i.e. evaluation on the original lattice.

This results in a big power improvement from using SuRFs.

Conclusion

- Existing software (SPM, FSL, AFNI etc) only has LKC implementations under stationarity but the framework is more general.
- Using SuRFs accurately and quickly controls the FWER at the right level and allows you to drop the good lattice assumption. In particular they allow RFT inference to work at any level of applied smoothness.
- I haven't discussed a further key assumption that fails in fMRI namely Gaussianity! In fact fMRI data is non-Gaussian and further fixes are required. See e.g. my thesis for further details: `sjdavenport.github.io/research/`.
- I interested, relevant papers: (Telschow et al, 2023) and (Davenport, Telschow, Schwarzman, & Nichols, 2023) will both soon be available on arxiv.

Bibliography

Davenport, S., Telschow, F., Schwarzman, A., & Nichols, T. E. (2023). Accurate voxelwise FWER control in fMRI using Random Field Theory.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, *113*(28), 7900–7905.

Friston, K., Worsley, K. J., Frackowiak, R., Mazziotta, J., & Evans, A. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 214–220.

Telschow et al, F. (2023). Riding the SuRF to continuous land: precise FWER control for gaussian random fields. *Preprint.*