Accurate voxelwise FWER control in fMRI using Random Field Theory

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

University of Oxford

March 6, 2023

Voxelwise Inference using RFT

Smoothing

Suppose we have N subjects and that for each subject we observe a 3D image data X_n on a finite lattice $V \subset S$. In fMRI smoothing is done in order to increase the signal to noise ratio. I.e. for each subject n, X_n is smoothed with a kernel K to give

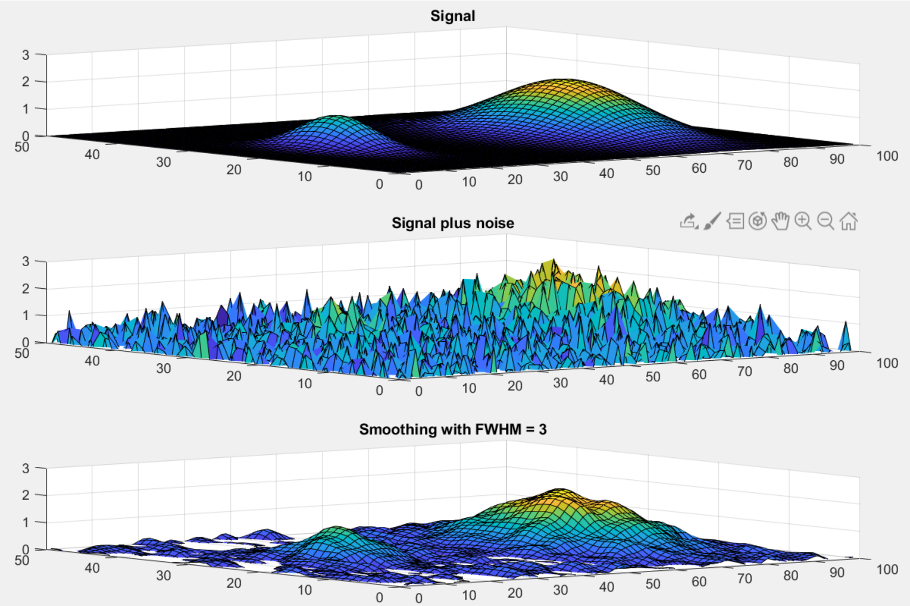
$$Y_n(v) = \sum_{l \in V} K(v-l) X_n(l)$$

at every voxel $v \in V$. An example of a typically used smoothing kernel is

$$K(x) = \frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2}x^T \Sigma^{-1} x}.$$

In fMRI it is typical to take $\Sigma = \sigma^2 I$ and to say that you're smoothing with FWHM $= 2\sqrt{2\log(2)}\sigma$.

Why you should smooth



Brain Imaging

Having smoothed, in brain imaging we have image data from each subject and at each voxel v we fit a linear model:

$$Y(v) = A\beta(v) + \epsilon$$

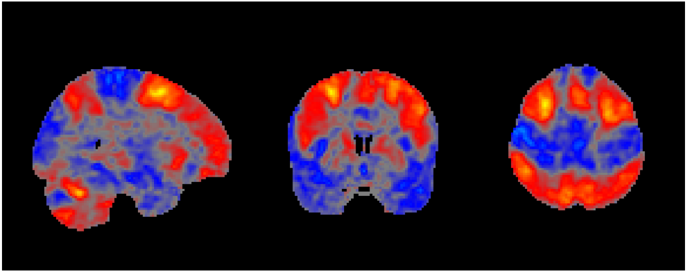
where $Y(v) = [Y_1(v), \dots, Y_N(v)]^T$ and A is some matrix. In this context we typically want to test whether $c^T\beta = 0$ for some contrast vector c. The simplest example of this is a one-sample t test. I.e. taking $A = [1, \dots, 1]^T$, $\beta \in \mathbb{R}$ we get

$$\hat{\beta} = \frac{1}{N}\sum_{n=1}^{N} Y_n(v)$$

where N is the number of subjects. We test the hypothesis that $\beta = 0$ using

$$T_L(v) = \frac{\hat{\beta}\sqrt{N}}{\hat{\sigma}} = \frac{\frac{1}{\sqrt{N}}\sum_{n=1}^{N} Y_n(v)}{\left(\frac{1}{N-1}\sum_{n=1}^{N}\left(Y_n(v) - \frac{1}{N}\sum_{n=1}^{N} Y_n(v)\right)^2\right)^{1/2}}.$$

t fields



Definition

Given $S \subset \mathbb{R}^D$, and $N \in \mathbb{N}$ and Gaussian random fields $Y_1, \dots, Y_N$, define the t-field to be $T : S \to \mathbb{R}$,

$$T(s) = \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^{N} \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^{N} Y_n(s) \right)^2 \right)^{1/2}}.$$

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.

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

and we seek to control this at a level α.

Historically voxelwise RFT, as developed in (Worsley et al., 1996) has assumed that there is a smooth random field T such that $T \approx T_L$ and

$$\max_{l \in V} T_L(l) \approx \sup_{s \in S} T(s)$$

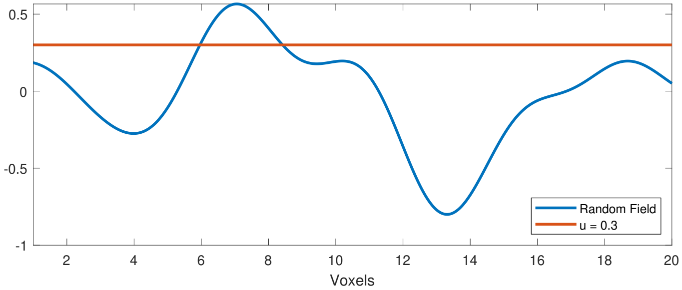
an assumption known as the **good lattice assumption**.

Voxelwise RFT

Let $M_u(T)$ be the number of local maxima of T above a threshold u then assuming that T is twice differentiable,

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

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



The Euler Characteristic approximation

$\mathbb{E}[M_u(T)]$ is difficult to estimate and requires us to be clever. To do so, given $u \in \mathbb{R}$, define the **excursion set** to be

$$\mathcal{A}_u(T) = \{s \in S : T(s) \geq u\}$$

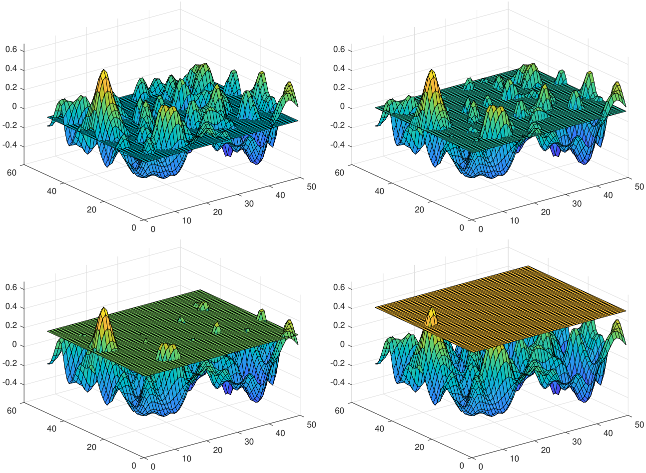
and let

$$\chi_u(T) = \chi(\mathcal{A}_u(T))$$

be the Euler characteristic of the excursion set. In 1D the Euler characteristic is the number of connected components. In 2D it's the number of connected components 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.



Gaussian Kinematic Formula

(Taylor et al., 2006) showed that the following Gaussian Kinematic Formula holds.

Theorem

Let $Y_1, \dots, Y_N$ be i.i.d D-dimensional unit variance Gaussian random fields and let $F : \mathbb{R}^N \to \mathbb{R}$. Let T be a random field such that

$$T(s) = F(Y_1(s), \dots, Y_n(s))$$

for all $s \in S$. Then, under certain regularity conditions, for all $u \in \mathbb{R}$,

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

where $\mathcal{L}_0, \dots, \mathcal{L}_D$ are constants and $\rho_d^F : \mathbb{R} \to \mathbb{R}$ are functions that depends on F and are easy to compute.

GKF - tFields

In particular, given Gaussian random fields $Y_1, \dots, Y_N$ (some $N \in \mathbb{N}$), we can write the one-sample t-field as:

$$\begin{aligned} T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^N Y_n(s) \right)^2 \right)^{1/2}} \\ &= F(Y_1(s), \dots, Y_N(s)) \end{aligned}$$

where $F : \mathbb{R}^N \to \mathbb{R}$ sends $y = (y_1, \dots, y_N)$ to

$$F(y) = \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N y_n}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(y_n - \frac{1}{N} \sum_{n=1}^N y_n \right)^2 \right)^{1/2}}.$$

GKF - tFields

So for T-fields, we have

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

where $\rho_d^F(u)$ has a closed form. Note that for all $s \in S$

$$\begin{aligned} T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(Y_n(s) - \frac{1}{N}\sum_{n=1}^N Y_n(s)\right)^2\right)^{1/2}} \\ &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N \frac{Y_n(s)}{\sigma(s)}}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(\frac{Y_n(s)}{\sigma(s)} - \frac{1}{N}\sum_{n=1}^N \frac{Y_n(s)}{\sigma(s)}\right)^2\right)^{1/2}} \end{aligned}$$

where $\sigma^2(s) = \mathrm{var}(Y_1(s))$, so in this case the assumption that the fields are unit variance does not matter.

Calculating the voxelwise threshold

Given estimates of the LKCs: $\hat{\mathcal{L}}_0, \dots, \hat{\mathcal{L}}_D$ and $\alpha > 0$ we can calculate an α level threshold, u_α such that

$$\sum_{d=0}^{D} \hat{\mathcal{L}}_d \rho_F(u_\alpha) = \alpha$$

as this will control the FWER to a level α since

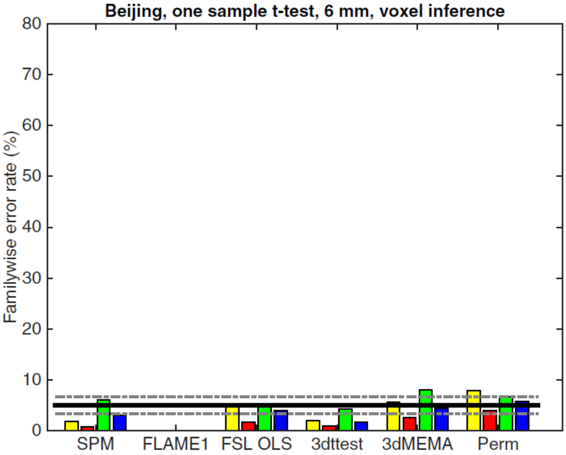
$$\mathbb{P}\left(\max_{s \in S} T(s) > u_\alpha\right) \leq \mathbb{E}[M_{u_\alpha}(T)] \leq \mathbb{E}[\chi(\mathcal{A}_{u_\alpha}(T))] \approx \alpha. \tag{1}$$

Two-tailed thresholds can be obtained similarly e.g. as:

$$\mathbb{P}\left(\max_{s \in S} |T(s)| > u_{\alpha/2}\right) \leq \mathbb{P}\left(\max_{s \in S} T(s) > u_{\alpha/2}\right) + \mathbb{P}\left(\min_{s \in S} T(s) < -u_{\alpha/2}\right).$$

Performance of Traditional RFT

In 2016 (Eklund, Nichols, & Knutsson, 2016) showed that clusterwise inference had massively inflated false positive rates. However they actually showed that the opposite held true for voxelwise inference.



Good Lattice Assumption

Historically voxelwise RFT has assumed that there is a smooth random field T such that $T \approx T_L$ and

$$\max_{l \in V} T_L(l) \approx \sup_{s \in S} T(s)$$

an assumption known as the **good lattice assumption**. With high enough smoothing this is not a problem. However at smoothing levels typically used in fMRI this fails. In particular given T_L, suppose that a random field T exists such that for all v, $T(v) = T_L(v)$ then

$$\max_{l \in V} T_L(l) = \max_{l \in V} T(l) < \sup_{s \in S} T(s).$$

Thus for any threshold u

$$\mathbb{P}(\max_{l \in V} T_L(l) > u) < \mathbb{P}(\max_{l \in V} T(l) > u)$$

so choosing thresholds for T_L based on T leads to conservativeness.

Assumptions of traditional RFT in SPM

- Good Lattice Assumption (i.e smoothness)
- Stationarity (needed for LKC calculation)
- Gaussianity (questionable validity in fMRI)
- Accuracy of the Euler characteristic approximation (requires high thresholds)

We will show that the good lattice assumption and stationarity can be dropped. We shall show that Gaussianity of the underlying fields can in practice be dropped by applying a transformation that accelerates the convergence of the Central Limit Theorem.

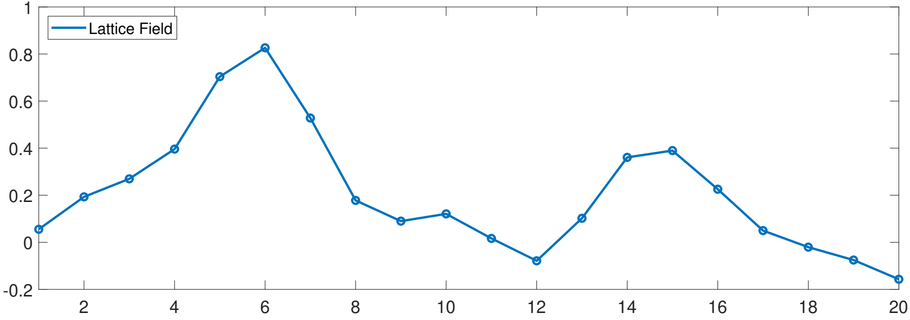
Super resolution random fields (SuRFs)

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.

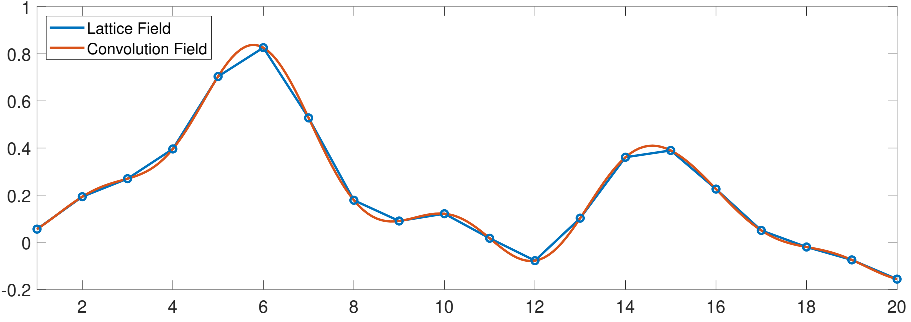


Convolution Random Fields

Definition

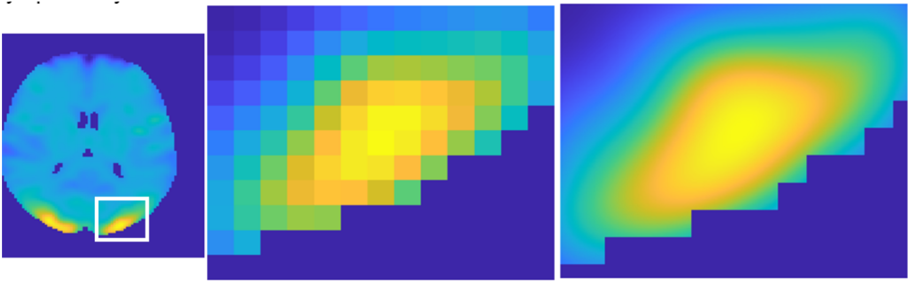
Given random data X on a lattice $L \subset \mathbb{R}^D$ for $s \in \mathbb{R}^D$ and some kernel K, define the SuRF $Y : \mathbb{R}^D \to \mathbb{R}$, s.t. for all $s \in S$,

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



Convolution Fields in Brain Imaging

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



Convolution t-fields

Given convolution random fields $Y_1, \dots, Y_N$ for each subject the convolution t-field is just what you'd expect i.e.

$$\begin{aligned} T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^{N} \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^{N} Y_n(s) \right)^2 \right)^{1/2}} \\ &= F(Y_1(s), \dots, Y_N(s)) \end{aligned}$$

We can similarly define convolution F-fields and more complicated fields.

SuRFS fix the Good Lattice Problem

- Historically RFT inference has only worked at high smoothness levels.
- Using SuRFs to do inference means that RFT works at any applied smoothness because the theory is valid for continuous random fields.
- In order to do inference can consider the maximum of the SuRF rather than the field on the lattice.

LKCs under non-stationarity

Because fMRI data is non-stationary we need to be able to estimate the LKCs uner non-stationarity. There has been some work on this though not much progress until recently. One thing that you can take advantage of is closed forms of the higher LKCs, in particular in D dimensions:

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

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. $\mathcal{L}_{D-1}$ also has a closed form in any dimension and

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

i.e. the Euler characteristic of the domain.

Other LKCs

However these nice closed forms do not exist for the other LKCs! (A problem in 3D for instance.)
However recalling the GKF,

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

If T is a 3D Gaussian field then (up to multiplicative constants)

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

So $\mathcal{L}_3$ dominates at high u. As such when doing voxelwise inference, it is not a problem. And the same holds when considering T and F fields.

FWER simulation results

Simulation settings

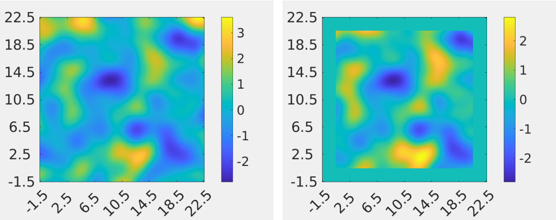


Figure 2: Stationary simulation setting

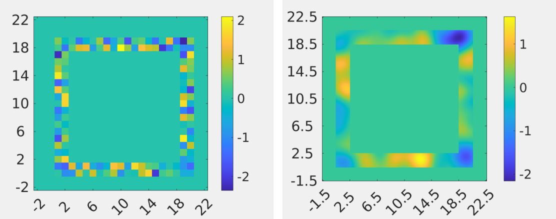


Figure 3: Non-Stationary simulation setting

Stationary Simulations

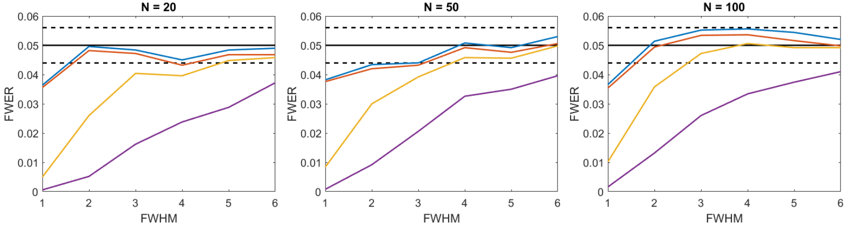


Figure 4: Stationary box simulation: FWER control. Blue: Expected Euler characteristic, Red: Convolution field converage, Yellow: resolution one lattice, Purple: Traditional RFT - i.e. evaluation on the original lattice.

Non-Stationary Simulations

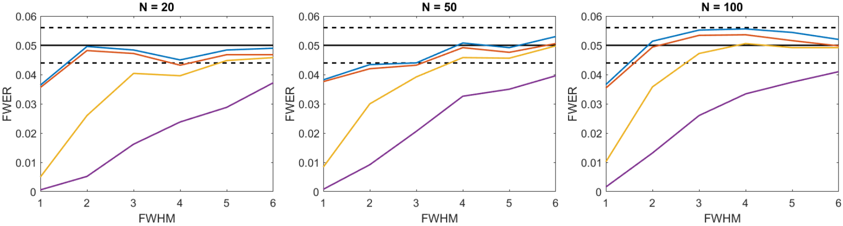
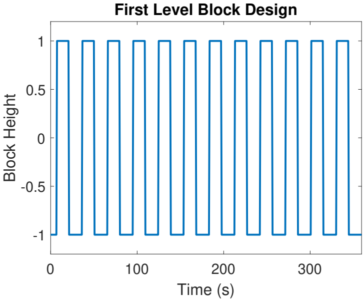


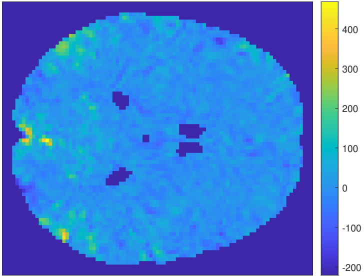
Figure 5: FWER control. Blue: Expected Euler characteristic, Red: Convolution field converage, Yellow: resolution one lattice, Purple: Traditional RFT - i.e. evaluation on the original lattice.

Resting State Validation

Data processing

We processed data from 7000 subjects from the UK biobank. Each subject has a time series of 490 images. Combine these into one contrast image using a block design at each voxel.





The results is 7000 contrast images (one for each subject). Which have mean zero by construction. Importantly we randomized the blocks. Not doing so can lead to incorrect inference.

Bootstrap validation

We followed (Eklund et al., 2016) and randomly drew 1000 subsets (of size $N = 10, 20$ and 50) from the data to test the methods.

I.e each N for $j = 1, \dots, 1000$, we sampled images $X_1^{(j)}, \dots, X_N^{(j)}$ from the 7000 images, smoothed them and computed the one-sample t-statistic T_j from them

Then we can estimate the SuRF FWER as:

$$\frac{1}{1000} \sum_{j=1}^{1000} 1[\max_{s \in S} T_j(s) > u_\alpha]$$

and the lattice FWER as:

$$\frac{1}{1000} \sum_{j=1}^{1000} 1[\max_{v \in V} T_j(v) > u_\alpha]$$

We use 7000 images instead of the between 100-200 samples used in (Eklund et al., 2016) meaning that we don't suffer from the same level of bias due to dependence between the draws.

FWER control on the original data

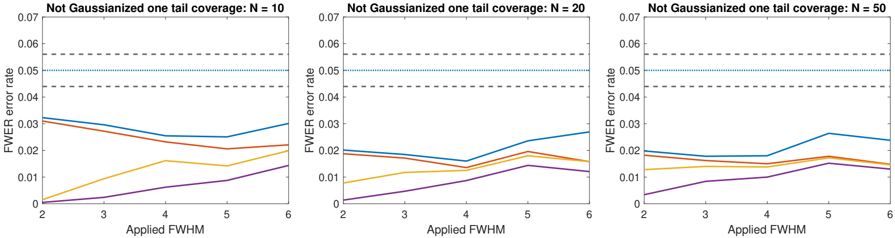


Figure 6: FWER control. Blue: Expected Euler characteristic, Red: Convolution field converage, Yellow: resolution one lattice, Purple: Traditional RFT - i.e. evaluation on the original lattice.

But this doesn't work that well...!!!

Resting state LKCs

We are estimating $\mathbb{E}[\chi(\mathcal{A}_u(T))]$ using $\sum_{d=0}^{D} \hat{\mathcal{L}}_d \rho_d^F(u)$.
For comparison we shall also compare to two stationary LKC estimation methods.

- Kiebel which averages lattice based estimates of $\hat{\Lambda}$ to estimate Λ
- Forman which assumes a stationary Gaussian covariance function.

But we can also estimate the true EEC distribution using the resting state data. I.e compute the curves

$$u \mapsto \frac{1}{1000} \sum_{j=1}^{1000} \chi(\mathcal{A}_u(T_j)).$$

Importantly this is the truth!

Expected Euler characteristic curve - original data

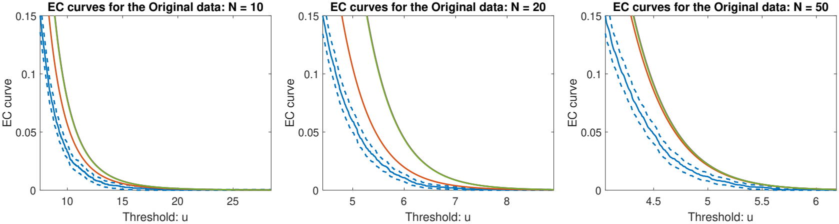
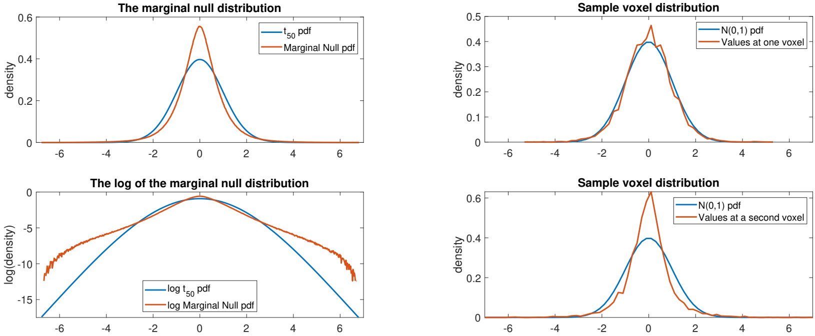


Figure 7: Blue: resting estimate EEC + 95% uncertainty, Red: SuRF LKC approximation. Green: Stationary LKC estimates (Kiebel + Forman).

Why doesn't it work?

Well the crucial and really only assumption left (as smoothness is no longer necessary nor is stationarity) is Gaussianity.

fMRI data is non-Gaussian



Transforming the data

Currently:

$$\frac{\sqrt{N}\hat{\mu}(Y_1(v), \dots, Y_N(v))}{\hat{\sigma}(Y_1(v), \dots, Y_N(v))}.$$

instead we compute

$$\frac{\sqrt{N}\hat{\mu}(f(Y_1(v)), \dots, f(Y_N(v)))}{\hat{\sigma}(f(Y_1(v)), \dots, f(Y_N(v)))}.$$

where we choose f to improve Gaussianity.
If we knew the marginal CDF of the data:Ψ_v under the null at a given voxel v we could then transform our data to

$$Y_n'(v) = \Phi^{-1}\Psi_v(Y_n(v))$$

to ensure that the data was marginally Gaussian under the null.

Gaussianization transformation

More formally, at each voxel v we standardize and demean the underlying (pre-smoothing) fields X_n . This yields standardized fields:

$$X_n^{S,D} = \frac{X_n - \hat{\mu}}{\hat{\sigma}}. \tag{2}$$

Going back to the original data we standardize it (without demeaning) to yield:

$$X_n^S = \frac{X_n}{\hat{\sigma}}$$

and for each voxel v and subject n we compare $X_n^S(v)$ to the null distribution to obtain a quantile

$$q_n(v) = \frac{1}{N|\mathcal{V}|} \sum_{n=1}^{N} \sum_{v' \in \mathcal{V}} 1[X_n^S(v) \leq X_n^{S,D}(v')].$$

Gaussianized SuRFs

The Gaussianized fields, for each voxel v and subject n, are then given by

$$X_n^G(v) = \Phi^{-1}(q_n(v))$$

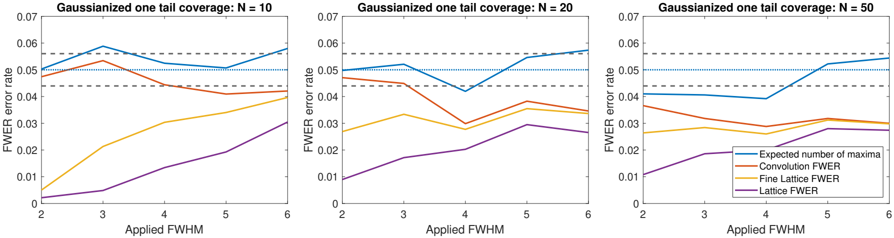
and from these we can calculate the Gaussianized SuRFs

$$Y_n^G(s) = \sum_{v' \in \mathcal{V}} K(s - v')X_n^G(v') \tag{3}$$

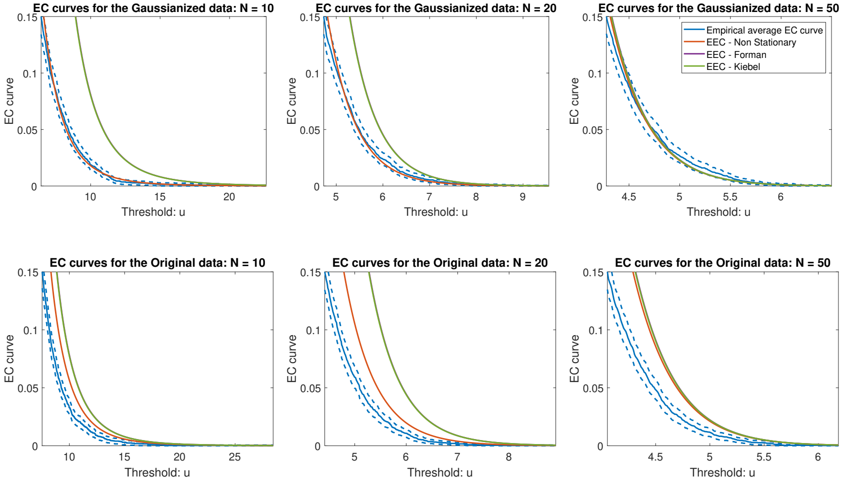
and generate corresponding t fields in order to perform FWER inference.

Results

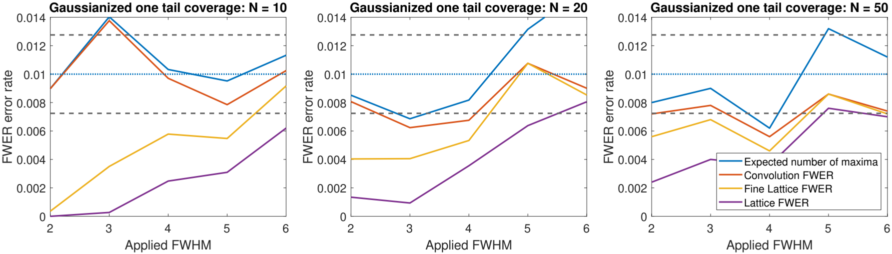
FWER control on the Gaussianized data



Expected Euler characteristic curve



Controlling at $\alpha = 0.01$



Conclusion

- Existing software (SPM, FSL, AFNI etc) only has LKC implementations under stationarity but the framework is more general.
- Using convolution fields accurately and quickly controls the FWER at the right level and allows you to drop the good lattice assumption.
- fMRI data is non-Gaussian and using a transformation can accelerate convergence of the CLT.

Papers

- This talk summarizes the work in two papers: (Telschow, Davenport, & Schwartzman, 2023) and (Davenport, Telschow, Schwarzman, & Nichols, 2023). Both will soon be available on arxiv.
- If you would like to read more about it, more details are available in my thesis found on my website, see: `sjdavenport.github.io/research/`.
- Software in MATLAB to perform RFT inference is available in the RFTtoolbox (Davenport & Telschow, 2023).
- Slides available at `sjdavenport.github.io/talks`.
- Checkout my work with Pierre, (Davenport, Thirion, & Neuvial, 2022) which provides control of the False Discovery Proportion in the linear model using the non-parametric bootstrap.

Bibliography

Davenport, S., & Telschow, F. (2023). *RFTtoolbox.* Retrieved from `https://github.com/sjdavenport/RFTtoolbox`

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

Davenport, S., Thirion, B., & Neuvial, P. (2022). FDP control in multivariate linear models using the bootstrap. *arXiv preprint arXiv:2208.13724.*

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.

Taylor, J. E., et al. (2006). A Gaussian Kinematic Formula. *The Annals of Probability*, *34*(1), 122–158.

Telschow, F., Davenport, S., & Schwartzman, A. (2023). Riding the SuRF to continuous land: precise FWER control for gaussian random fields. *Preprint.*

LKC estimation results

We run 2D simulations, of white noise smoothed with a Gaussian Kernel. Kiebel and Forman are designed to estimate the LKCs under stationarity but they are biased. HPE and bHPE are unbiased but have a higher variance.