Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

UNIVERSAL WISER PUBLISHER

Research Article

# Improved Cluster-Wise Inference in Functional Magnetic Resonance Imaging via Geometric Redefinition of Cluster Size

**Huashuai Xu[1,2], Yuge Xing[3], Weiya Guo[1], Dongdong Zhou[4], Huanjie Li[1,2]***

[1] Women and Children's Hospital of Dalian University of Technology, Dalian, China
[2] School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, China
[3] Department of Pediatrics, Linyi People's Hospital, Linyi, China
[4] School of Computer Science and Technology, Dalian University of Technology, Dalian, China
E-mail: hj_li@dlut.edu.cn

**Received:** 22 August 2025; **Revised:** 16 September 2025; **Accepted:** 22 September 2025

**Abstract:** Accurate identification of significant activations in functional Magnetic Resonance Imaging (fMRI) is essential for reliable neuroimaging research. Cluster-wise inference based on Random Field Theory (RFT) has been widely adopted for over two decades due to its superior sensitivity compared with voxel-wise corrections. However, recent studies have revealed that conventional RFT-based approaches may yield inflated false-positive rates. This study proposes an improved cluster-wise inference method by redefining cluster size from a geometric perspective. Specifically, we calculate cluster size using intrinsic volume rather than voxel count, and we investigate the influence of expected cluster size estimation and voxel connectivity. Analyses use publicly available first-level images from 198 subjects and matched simulations; factors cross Cluster-Defining Threshold (CDT) $p = 0.001/0.01$, one- vs two-sample $t$ tests (two group sizes), and Euler Characteristic (EC) implementations (Statistical Parametric Mapping (SPM) default vs corrected), under 6/18-connectivity. On Gaussian-null simulations (32 configurations), the geometric definition achieved FWER at or near 0.05 across all settings and was consistently lower than voxel-counting. On real data (128 configurations spanning four paradigms), it reduced FWER relative to voxel-counting. Sensitivity was assessed only on simulations with five planted clusters over SNR$= 1-4$, where detection rates were comparable to the voxel-count approach. These findings highlight the value of geometric definitions in enhancing statistical inference for multi-voxel activation patterns in fMRI. Overall, redefining cluster extent by intrinsic volume improved family-wise error rate control while maintaining sensitivity, providing a simple drop-in change to cluster-wise RFT workflows.

***Keywords***: functional magnetic resonance imaging, cluster-wise inference, random field theory, Euler characteristic, intrinsic volume

**MSC:** 62M40, 60G60

## 1. Introduction

Functional Magnetic Resonance Imaging (fMRI) enables noninvasive mapping of brain activity at millimeter resolution, but the sheer number of voxels (often $> 10^5$) renders voxel-wise inference susceptible to severe multiple

DOI: https://doi.org/10.37256/cm.6620258354

comparisons. Because spatial correlations between neighboring voxels violate independence assumptions, purely voxel-wise Family-Wise Error (FWER) control tends to be overly conservative, leading to a loss of statistical power. Cluster-wise inference based on Random Field Theory (RFT) [1–5] has therefore become a standard solution: it leverages both a Cluster-Defining Threshold (CDT) and the spatial extent of supra-threshold clusters to test for image-wide significance, and it is implemented in major neuroimaging software such as SPM (https://www.fil.ion.ucl.ac.uk/spm/) and FMRIB Software Library (https://fsl.fmrib.ox.ac.uk) [6, 7]. Compared with Bonferroni-type corrections, RFT-based cluster inference generally achieves higher sensitivity by reducing the effective multiplicity through topological summaries of the excursion set [8, 9].

Despite its widespread adoption, practical pipelines can yield inflated False Positive Rates (FPRs). A landmark evaluation reported that the then-common implementations in SPM/FSL/Analysis of Functional NeuroImages (AFNI) could yield FWER substantially above nominal levels in resting-state data [10, 11]. Subsequent work traced potential contributors to (i) non-Gaussian and long-tailed Spatial Autocorrelation Functions (SACF) and nonstationarity [12–16]; (ii) resampling and interpolation choices during preprocessing [17]; and (iii) the use of "null" resting-state data potentially containing default-mode activity rather than true null signals [18, 19]. In response, several methodological refinements were proposed, including Monte Carlo-based Equitable Thresholding and Clustering (ETAC) [20], probabilistic Threshold-Free Cluster Enhancement (pTFCE) that projects cluster-level neighborhood information back to voxel space [21], and EC-based variants that re-express expected cluster counts [22]. While these approaches can improve robustness, a more foundational issue remains under-explored: how we define and measure cluster size in a framework (RFT) that is inherently continuous and geometric, yet is routinely applied to discrete voxel lattices.

From a mathematical standpoint, RFT characterizes the topology of excursion sets of smooth random fields over a continuous domain $S \subset \mathbb{R}^3$. Given a threshold $u$, the excursion set $E(u) = \{x \in S : Z(x) \geq u\}$ can be summarized by Lipschitz-Killing curvatures (intrinsic volumes), with the Euler Characteristic (EC) providing an approximation to the expected number of clusters at high thresholds. The expected size distribution of clusters is then derived under smoothness- and threshold-dependent asymptotics, with scale determined by RESolution ELements (RESELs). In this continuous theory, cluster "size" is a volume measure (an intrinsic, geometric quantity). Standard implementations estimate cluster extent by voxel counts on a discrete lattice, which can distort FWER control.

Two additional practical facets interact with this mismatch. First, the estimation of expected cluster size depends on EC densities and RESEL volumes; different "default" or "corrected" EC formula choices and smoothness estimates can shift the calibrated cluster-extent threshold, particularly when CDT is low or smoothness is modest. Second, voxel connectivity alters cluster topology on the lattice. In RFT's continuous formulation, a voxel is an infinitesimal point; adopting 6-connectivity is geometrically aligned with face-sharing adjacency, whereas 18/26-connectivity incorporate edge/vertex contacts that may artificially inflate connectivity in discrete space. Clarifying how these choices affect realized FWER is critical for principled cluster inference.

This paper addresses these issues by reinterpreting cluster size geometrically and aligning implementation with RFT's continuous-space assumptions. Our central idea is to redefine cluster size by intrinsic volume rather than by voxel count. Operationally, we compute cluster extent at the volume level and examine how this change interacts with (i) different EC density formulas used to obtain expected cluster counts and sizes, and (ii) alternative connectivity definitions. We evaluate performance using both simulated Gaussian nulls and real resting-state fMRI datasets under a comprehensive factorial design that varies smoothing, CDT, group sizes, and one- vs two-sample $t$-tests-capturing regimes where RFT approximations are known to be more or less reliable.

Our contributions are threefold:

Geometric redefinition of cluster size: we replace voxel-count extent with an intrinsic-volume-based measure, reducing model-implementation mismatch.

Systematic analysis of EC and connectivity: we quantify how EC formulations and 6/18-connectivity interact with CDT and realized FWER.

Extensive validation: simulations and real data demonstrate closer-to-nominal FWER with maintained sensitivity.

Paper organization. Section 2 formalizes the cluster-wise RFT framework, details the geometric redefinition of cluster size, and describes EC/RESEL computation and connectivity conventions, presents simulation and real-data designs and evaluation metrics. Section 3 reports FWER and sensitivity results under all parameter combinations. Section 4 discusses implications, limitations (e.g., residual inflation in real data due to assumption violations), and avenues for extending geometric definitions to other modalities and statistics. Section 5 concludes.

# 2. Materials and methods

## 2.1 *Notation and preliminaries*

Let $S \subset \mathbb{R}^3$ be the search domain (brain mask). Let $Z(x)$ be a smooth statistical random field on $S$ (e.g., a $t$- or $z$-field after first-level modeling and group analysis). For a Cluster-Defining Threshold (CDT) $u \in R$, the excursion set is

$$E(u) = \{x \in S : Z(x) > u\} \tag{1}$$

We denote the intrinsic volumes (Lipschitz-Killing curvatures) of S by $\{\mu_0(S),\ \mu_1(S),\ \mu_2(S),\ \mu_3(S)\}$ and the RESEL measures by $\{R_0,\ R_1,\ R_2,\ R_3\}$, with $R_d$ being the smoothness-adjusted counterpart of $\mu_d$. Spatial smoothness is parameterized by Full Width at Half Maximum (FWHM) $(f_{x1},\ f_{x2},\ f_{x3})$. The RESEL volumes $R_d(S)$ (smoothness-adjusted intrinsic volumes) are used in Euler Characteristic (EC) expansions; in 3D, $R_3(S) = \mu_3(S)/(f_{x1}f_{x2}f_{x3})$. The Euler Characteristic (EC) of a set $A \subset \mathbb{R}^3$ is denoted $\chi\ (A)$. The excursion-set volume $V_u = \lambda\ (E(u))$ (or $R_3(E(u))$ in RESEL units).

**Topology.** For any $A\ \subset\ \mathbb{R}^3$, the Euler characteristic

$$\chi(A) = \#(\text{connected components}) - \#(\text{handles}) + \#(\text{ voids}) \tag{2}$$

At sufficiently high threshold $u$ (i.e., the high-u regime where excursion sets consist of well-separated peaks; operationally in this paper $u$ corresponds to Z-thresholds of 2.3 and 3.1 for $p = 0.01$ and $p = 0.001$, respectively), $\chi\ (Eu)$ approximates the number of clusters in $E(u)$ [2, 9, 23–26].

Throughout, we write:

$E(C_u)$ for the expected number of clusters in $E(u)$;

$V_u$ for volume of the excursion set in $E(u)$;

$K_u$ for the observed cluster size at CDT $u$;

$\alpha$ for the nominal family-wise error rate (FWER), set to 0.05 unless stated otherwise.

## 2.2 *Classical RFT cluster-wise inference*

We summarize the identities that calibrate cluster-extent inference under Gaussian Random Field (GRF) theory.

**(A) Expected excursion-set volume.** With $F_X$ distribution function of the respective field,

$$E(\lambda(E_(u))) = \lambda(S)(1 - F_X(u)),\ \ E(R_3(E(u))) = R_3(S)(1 - F_X(u)) \tag{3}$$

Equation (3) states that the expected excursion volume equals the search-region size multiplied by the pointwise exceedance probability. The excursion set $E(u)$ contains locations in $S$ where the field exceeds $u$; viewing its volume as an indicator integral, taking expectation amounts to averaging the "above-threshold" probability over $S$. Under stationarity, this probability is the same everywhere, so the expectation factorizes into "region size × exceedance probability." The RESEL version is identical in logic: replace ordinary volume by the RESEL measure, which yields the same factorization when the RESEL density is spatially constant. Intuitively, the average fraction of $S$ above $u$ equals the exceedance fraction; it vanishes at very high $u$ and tends to the full search volume at very low $u$.

**(B) Expected EC ($\approx$ expected cluster count).** The expected Euler characteristic of the excursion set decomposes as a sum of intrinsic-volume (RESEL) measures $R_d(S)$ multiplied by EC densities $\rho_d(u)$ that depend on the threshold and field family (Z/T/F).

$$E(\chi(E(u))) = \sum_{d=0}^{3} R_d(S) * \rho_d(u), \quad E\left(C_u\right) = E(\chi(E(u))) \tag{4}$$

Importantly, the identification $E\left(C_u\right) = E(\chi(E(u)))$ holds only when $u$ is sufficiently high (the "high-$u$" regime); outside this regime, EC is not a reliable surrogate for the expected cluster count [2, 27–30]. For a Z-field in 3D,

$$\begin{aligned}
\rho_0(u) &= \int_u^{\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \\
\rho_1(u) &= \frac{\sqrt{4\ln 2}}{2\pi} e^{-\frac{u^2}{2}} \\
\rho_2(u) &= \frac{4\ln 2}{(2\pi)^{3/2}} u e^{-u^2/2} \\
\rho_3(u) &= \frac{(4\ln 2)^{\frac{3}{2}}}{(2\pi)^2} \left(u^2 - 1\right) e^{-\frac{u^2}{2}}
\end{aligned} \tag{5}$$

Authoritative derivations and tables can be found in: Adler and Taylor [1, 31]. For T- and F-fields, the corresponding EC density formulas are given in Worsley [32–35]. In SPM, the default evaluation of $E(\chi(E(u)))$ retains only the highest-order term $R_3(S) * \rho_3(u)$; accordingly, we report results for both the **default** (3D-term only) and the **corrected** (full $\sum_{d=0}^{3} R_d(S) * \rho_d(u)$) EC implementations and compare their impact.

**(C) Expected cluster size.** Let $K_u$ denote the size of a cluster given CDT = $u$. Under the standard approximation,

$$E\left(K_u\right) = \frac{E\left(V_u\right)}{E\left(C_u\right)} \tag{6}$$

When excursion-set volume is measured in Lebesgue units, set $Vu = \lambda\ (E(u))$; this yields the expected cluster volume in Lebesgue volume. When volume is measured in third-order RESEL units, set $Vu = R_3(E(u))$; this yields the expected cluster volume in RESELs, i.e., a smoothness-normalized measure comparable across resolutions.

**(D) Cluster-size tail (Friston form [36]).** In dimension $D = 3$,

$$P\left(K_u \geq k\right) = \exp\left(-\beta k^{\frac{2}{3}}\right), \quad \beta = \left(\frac{\Gamma\left(\frac{D}{2}+1\right)}{E\left(K_u\right)}\right)^{2/D} \tag{7}$$

It models the survival probability of cluster size as a simple exponential in a power of the extent, with the scale parameter determined by smoothness and threshold (through the expected cluster size). This approximation is most accurate in the high-u regime where clusters are well separated and peak theory applies [36].

**(E) Exceedance counts and FWER.** Let $C_{k,\ u}$ be the number of clusters in $E(u)$ with size $\geq k$. Using Poisson clumping,

$$C_{k,\ u} \sim \text{Possion}\left(\lambda_{k,\ u}\right), \quad \lambda_{k,\ u} = E\left(C_u\right) P\left(K_u \geq k\right) \tag{8}$$

Hence, the family-wise error rate is then the chance of observing at least one such cluster, yielding the familiar $1-\exp(-\lambda)$ form.

$$P\left(K_{\max}(u) \geq k\right) = 1 - \exp\left(-\lambda_{k,\ u}\right) \tag{9}$$

Intuitively, $\lambda$ is the expected number of qualifying clusters under the null; larger $\lambda$ implies higher FWER. Given target $\alpha$, the cluster-extent threshold $k_\alpha$ solves

$$1 - \exp\left(-E\left(C_u\right) P\left(K_u \geq k_\alpha\right)\right) = \alpha \tag{10}$$

and clusters are declared significant when $K_u \geq k_\alpha$.

## 2.3 *Geometric redefinition of cluster size*

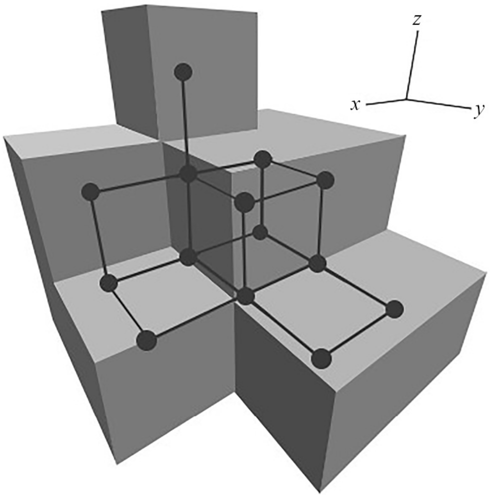


**Figure 1.** Example of volume calculation (adapted from Evans [4])

Random Field Theory (RFT) is formulated on a continuous domain, where the size of an excursion set is a Lebesgue volume and, more generally, is characterized by intrinsic volumes. A voxel is regarded as a sample location rather than a physical cube with nonzero measure, so voxel count is not equivalent to geometric volume. This distinction is illustrated in Figure 1, where a didactic configuration contains 14 voxels yet its intrinsic volume equals 1; the example shows that sparse or filamentary arrangements can substantially overstate "extent" when measured by voxel number while contributing very little geometric volume [4]. Because the calibration in Section 2.2 is written explicitly in terms of intrinsic/RESEL volumes, the statistic used for cluster-extent inference should live on that same geometric scale.

On a cubic lattice, the smallest configuration with strictly positive 3D volume is a $2 \times 2 \times 2$ block ("cubelet"). This mirrors standard discretizations of the 3D intrinsic volume $\mu_3$, which count elementary cubes (and, analogously, faces/edges for $\mu_2/\mu_1$) and then rescale by the physical voxel size. Two consequences follow: (i) isolated voxels and thin appendages that do not complete a $2 \times 2 \times 2$ block contribute no geometric volume; and (ii) among clusters with identical voxel counts, more cube-like shapes occupy more cubelets and thus have larger intrinsic volume.

We measure cluster extent on the geometric scale by counting the number of fully contained $2 \times 2 \times 2$ cubelets within each 6-connected cluster. Concretely, for a cluster $C \subset E(u)$, let Volume 1 denote the count of disjoint cubelets that lie entirely inside $C$, this count is used as the volume-level cluster size (expressed in cubelet units). For comparison only (see Results), we also report the conventional voxel-count extent, but all RFT calibrations are performed on the geometric scale. To obtain a shape-neutral upper envelope used in our ablations, we map a cluster's voxel count $N$ to the maximum cubelet count attainable by any cluster with $N$ voxels, denoted Volume 2. The complete mapping $N$ to Volume 2 for our lattice is provided in Table 1 and is used directly at analysis time; no run-time optimization is required. For each voxel count $N$, we derive the listed value by compactly arranging the $N$ voxels on a 3D lattice (same 6/18-connectivity as used in the analyses) to maximize the number of fully contained $2 \times 2 \times 2$ cubelets. Practically, voxels are placed as tightly as possible: a largest near-cubic 3D block, then an adjacent 2D slab from the remainder, then a 1D strip. The resulting maximum cubelet count is taken as the volume-based extent for that $N$. As illustrated by Figure 1, clusters with identical voxel counts can differ markedly in cubelet occupancy, and hence in geometric volume, which motivates working on this scale for extent inference. For example, a compact $3 \times 3 \times 3$ block ($N = 27$) admits at most $(3-1) \times (3-1) \times (3-1) = 8$ fully contained $2 \times 2 \times 2$ cubelets, whereas a thinner $3 \times 3 \times 2$ block ($N = 18$) admits at most $(3-1) \times (3-1) \times (2-1) = 4$. For the same voxel count, alternative (less compact) arrangements do not attain this maximal value; they yield fewer fully contained $2 \times 2 \times 2$ cubelets and thus a smaller geometric extent.

**Table 1.** Mapping from voxel count N to maximum cubelet count Volume 2

| #voxels | Volume 2 |
|---|---|
| 1-7 | 0 |
| 8-11 | 1 |
| 12-15 | 2 |
| 16-17 | 3 |
| 18-21 | 4 |
| 22-23 | 5 |
| ... | ... |
| 497-498 | 330 |
| 499 | 331 |

In summary, we evaluate three cluster-size definitions on the same (6-connectivity) lattice: (i) VoxelCount = $N$ (baseline), (ii) Volume 1 = the number of disjoint $2 \times 2 \times 2$ cubelets fully contained in $C$, and (iii) Volume 2 = the maximum cubelet count attainable by any cluster with $N$ voxels (lookup in Table 1). For each thresholded map (CDT $u$), we compute the map-wise maximum $K_{max}(u)$ under the chosen definition and plug it into the Section 2.2 calibration (cluster-size tail and Poisson clumping) to obtain the FWER-controlling threshold $k_\alpha$; a cluster is significant when its chosen size measure is larger than $k_\alpha$.

## 2.4 *Data and statistical analysis*

### 2.4.1 *Real fMRI data*

We used the publicly available first-level statistical images released by Eklund et al. [10], which allows direct group-level analyses on preprocessed maps. The dataset is publicly hosted; a persistent access link is provided here for convenience: (https://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html). The sample comprises 198 healthy participants (age 18-26 years; mean 21.16, SD 1.83) from the Beijing dataset. Data were acquired with TR = 2 s, yielding 225 time points per subject on a grid of $64 \times 64 \times 33$ voxels with voxel size $3.125 \times 3.125 \times 3.6$ mm$^3$. First-level preprocessing included spatial normalization to a brain template, motion correction, and spatial smoothing at 4, 6, 8, and 10 mm Full Width at Half Maximum (FWHM). A standard General Linear Model (GLM) was fit to the preprocessed fMRI time series using four commonly employed regressors that emulate null-like conditions for cluster-inference benchmarking: B1 (10-s on/off), B2 (30-s on/off), E1 (2-s activation, 6-s rest), and E2 (1-4-s activation, 3-6-s rest; randomized). Further acquisition and preprocessing details are provided in Eklund et al. [10].

### 2.4.2 *Simulated data*

To mirror the structure of the real dataset while providing controlled null conditions, we generated simulated volumes on the same lattice. The only change relative to the real data is that the original values were replaced by Gaussian noise. All simulated (null and signal-present) volumes were generated in MATLAB (MathWorks) using the ‘randn’ function to draw standard-normal noise on the same lattice as the empirical images; when applicable, volumes were smoothed with a 3D Gaussian kernel to match the target smoothness used in the analyses. In signal-present runs, a compact cluster-shaped mean effect was added at a known location; all other analysis settings matched those of the null simulations.

### 2.4.3 *Group-level analyses and comparisons*

Group inference was performed with one-sample $t$ tests (group activation) and two-sample $t$ tests (group differences). We compared empirical false-positive rates across:

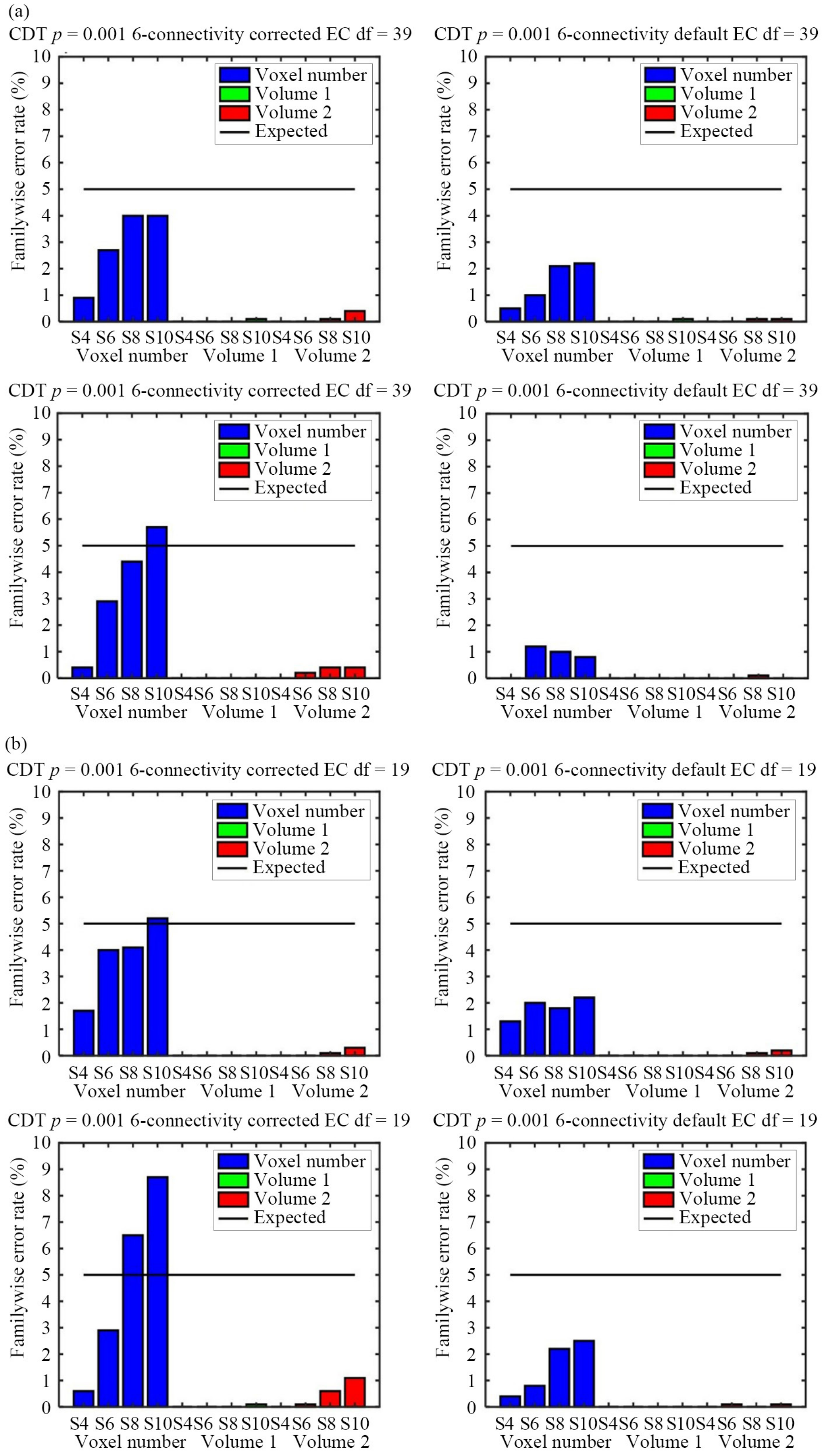
- Cluster-size definitions: voxel-count (in voxels) versus two volume-level definitions (denoted Volume 1 and Volume 2 for formal definitions);
- Connectivity: 6-connectivity versus 18-connectivity for cluster labeling;
- Euler characteristic implementation: corrected EC versus default EC.

For each experimental condition (CDT $p = 0.01/0.001$; one- vs two-sample $t$ tests with two group sizes; 6/18-connectivity; default vs corrected EC), we perform 1,000 independent runs using the analysis pipeline described above. Under the null setting, the realized FWER is computed as the proportion of runs (out of 1,000) in which at least one activation cluster is detected by the cluster-wise procedure (i.e., a supra-threshold cluster that survives the extent criterion for that condition). Under the signal-present setting, sensitivity is summarized by the number of activation clusters detected per run; in Section Results we report the corresponding count statistics aggregated over the 1,000 runs for each condition. The same definitions are applied consistently to the public first-level maps and to the simulations, ensuring that the graphical summaries in Section Results reflect identically computed quantities. All analyses were conducted in MATLAB R2018b using SPM12.

# 3. Results

This section reports two complementary performance summaries, defined in Methods 2.4.3: (i) realized family-wise error under the null and (ii) detection sensitivity under signal. Estimates are obtained across the full experimental grid-cluster-defining threshold (0.001, 0.01), EC implementation (default vs corrected), and voxel connectivity (6 vs 18)-using an identical analysis pipeline for both simulations and the public first-level fMRI maps described in 2.4.1. Taken together, these results characterize error control and detection capability under each factor combination and assess the impact of the proposed geometric (volume-based) extent relative to the voxel-count baseline.

### 3.1 *Simulated gaussian nulls (FWER)*



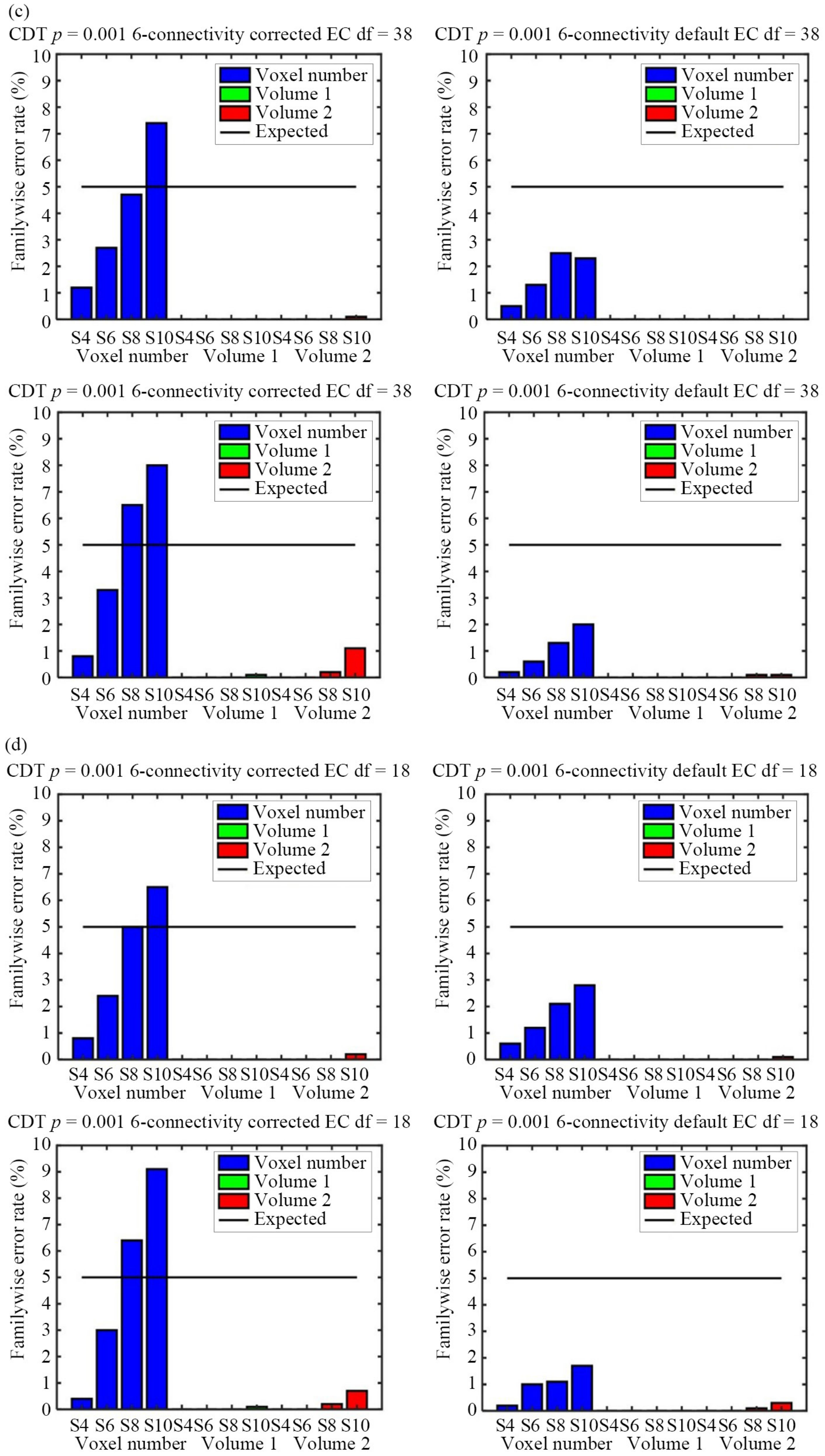
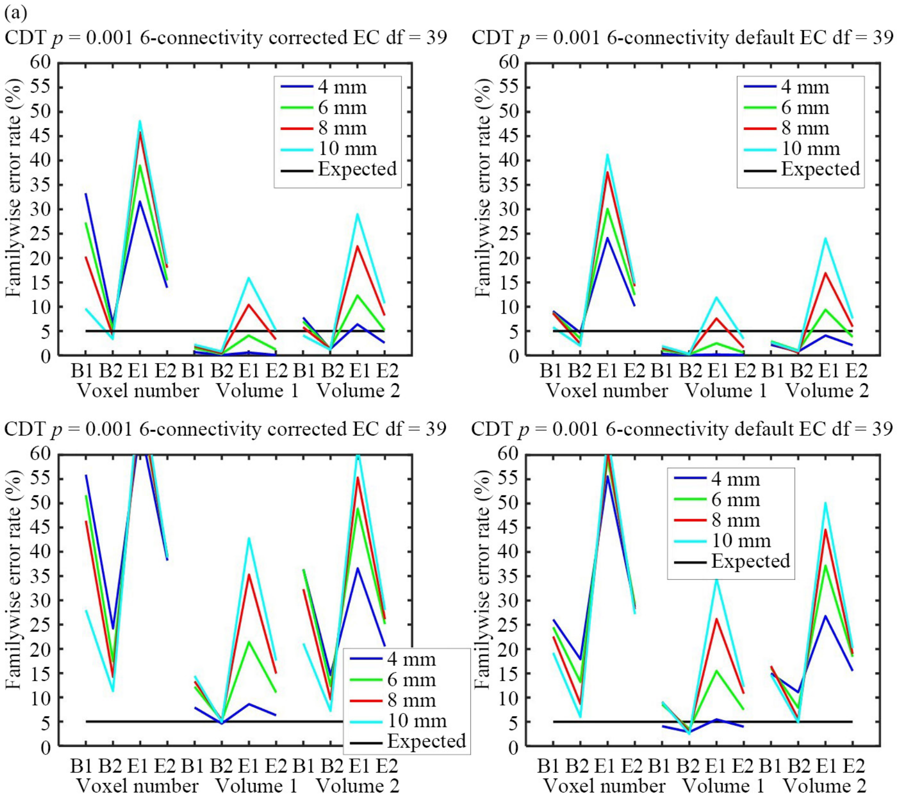


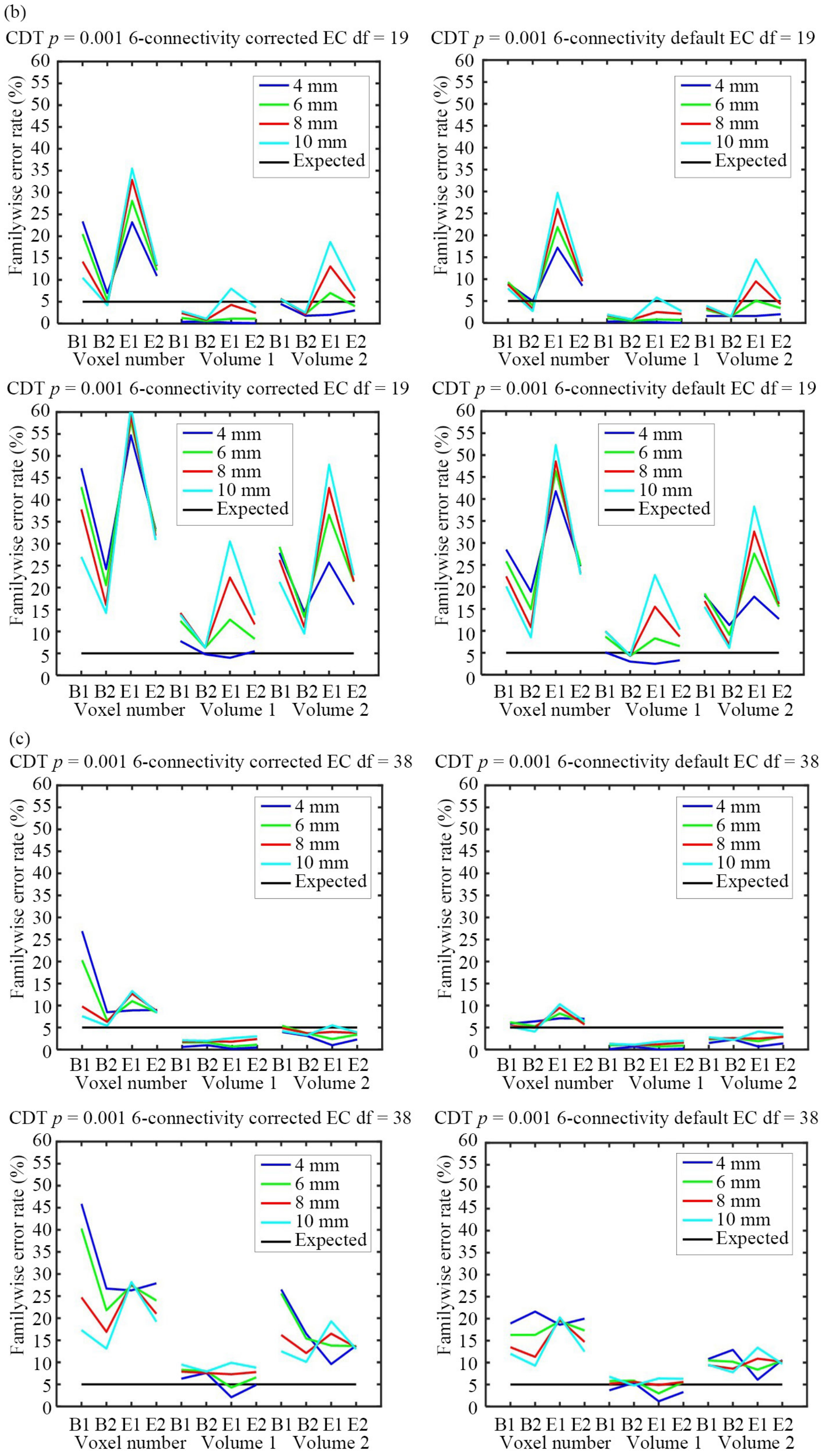
**Figure 2.** Family-wise errors on simulated data for (a-b) one-sample $t$ tests (group sizes 40 and 20); (c-d) two-sample $t$ tests (per-group size = 20, 10); From left to right: (i-ii) CDT $p = 0.001$, corrected EC and default EC respectively; (iii-iv) CDT $p = 0.01$, corrected EC and default EC respectively. The $x$-axis denotes smoothing levels (FWHM). Each bar reports the proportion of 1,000 null replicates with ≥ 1 supra-threshold cluster; a horizontal reference line at 0.05 is shown to facilitate comparison

All results in Figure 2 are reported with 6-connectivity; 18-connectivity produced the same outcomes and is therefore omitted for brevity. Across the simulated settings formed by crossing smoothing (4/6/8/10 mm FWHM), Cluster-Defining Thresholds (CDT $p = 0.001$ and 0.01), test types (one-sample and two-sample $t$), group sizes (two levels per design), and EC implementations (corrected vs. default), the volume-based cluster-size definitions (Volume 1 and Volume 2) yielded Family-Wise Error Rates (FWER) at or close to the nominal level throughout, whereas the voxel-count definition showed higher FWER under matched configurations. Specifically, at matched design, CDT $p = 0.001$ yields lower FWER than CDT $p = 0.01$. Differences between one- and two-sample $t$-tests are small across panels. Within each test type, smaller group sizes exhibit higher FWER than larger group sizes. Comparing EC implementations under the same settings, default EC tends to produce lower FWER than corrected EC. Finally, FWER increases with smoothing (from 4 to 10 mm FWHM).

## 3.2 *Real resting-state data (FWER)*

Figure 3 reports 6-connectivity results for the real dataset; 18-connectivity again matched these outcomes and is not shown. Using the same figure layout as Figure 2 for direct comparison, volume-based cluster size produced lower FWER than the voxel-count baseline across conditions spanning the four first-level paradigms (B1, B2, E1, E2), smoothing levels (4/6/8/10 mm), CDTs ($p = 0.001/0.01$), test types (one- and two-sample $t$), and group sizes (two levels). On the real dataset, patterns largely mirror the simulations. At matched design and smoothing, CDT $p = 0.001$ yields lower FWER than CDT $p = 0.01$; smaller group sizes show higher FWER; FWER increases with smoothing (4 to 10 mm FWHM); and default EC tends to yield lower FWER than corrected EC. Differences from the simulations are also evident: one-sample $t$-tests exhibit substantially higher FWER than two-sample tests under comparable settings, and across paradigms E1 shows the largest FWER overall (exceeding B1, B2, and E2 when other factors are matched).




(b)
CDT p = 0.001 6-connectivity corrected EC df = 19
CDT p = 0.001 6-connectivity default EC df = 19
CDT p = 0.001 6-connectivity corrected EC df = 19
CDT p = 0.001 6-connectivity default EC df = 19
(c)
CDT p = 0.001 6-connectivity corrected EC df = 38
CDT p = 0.001 6-connectivity default EC df = 38
CDT p = 0.001 6-connectivity corrected EC df = 38
CDT p = 0.001 6-connectivity default EC df = 38
Familywise error rate (%)
0
5
10
15
20
25
30
35
40
45
50
55
60
B1 B2 E1 E2 B1 B2 E1 E2 B1 B2 E1 E2
Voxel number
Volume 1
Volume 2
4 mm
6 mm
8 mm
10 mm
Expected

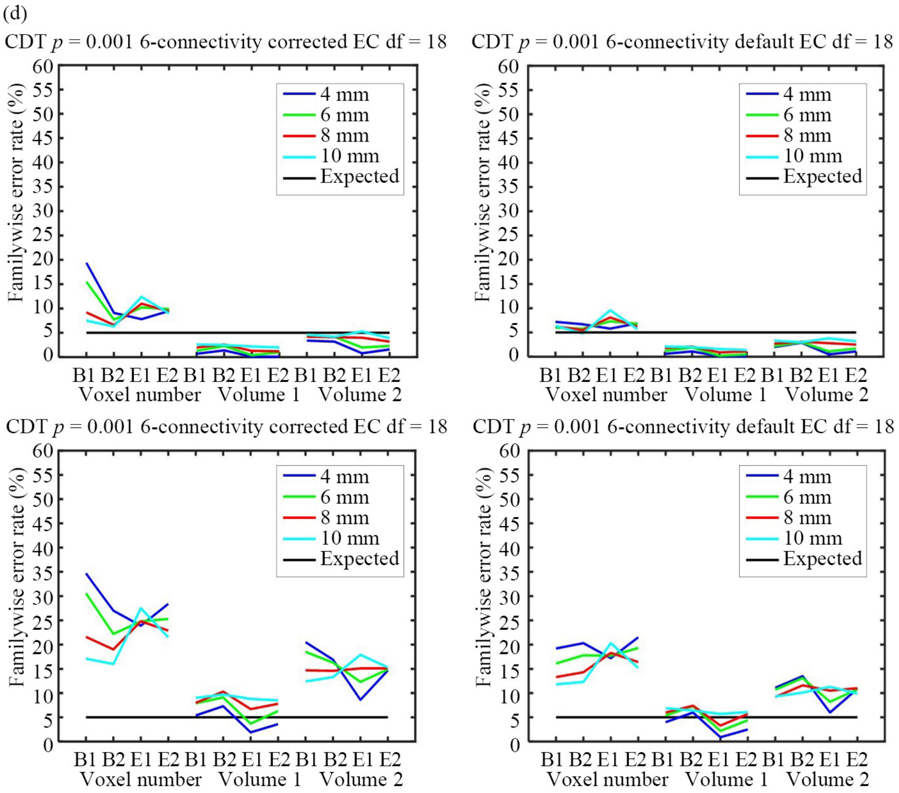
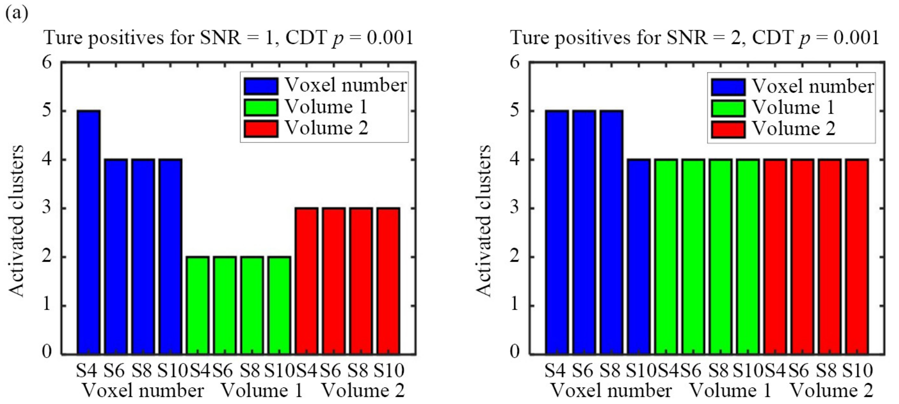


**Figure 3.** Family-wise errors on real resting-state data for (a-b) one-sample $t$ tests (group sizes 40 and 20); (c-d) two-sample $t$ tests (per-group size = 20, 10); From left to right: (i-ii) CDT $p = 0.001$, corrected EC and default EC respectively; (iii-iv) CDT $p = 0.01$, corrected EC and default EC respectively. The $x$-axis denotes first-level paradigms (B1, B2, E1, E2). Each bar reports the proportion of 1,000 null replicates with ≥ 1 supra-threshold cluster; a horizontal reference line at 0.05 is shown to facilitate comparison

## 3.3 *Sensitivity in signal-added simulations*

Sensitivity analyses with five planted clusters (nominal sizes: 12, 21, 6, 599, 176 voxels) at SNR$= 1 - 4$ and smoothing = 4/6/8/10 mm are summarized in Figure 4 using 6-connectivity (the 18-connectivity results were the same and are omitted). For one-sample group analyses (group size 40), panel A presents CDT $p = 0.001$ and panel B presents CDT $p = 0.01$, with columns indexing SNR and rows indexing smoothing. Across all SNRs and smoothing levels, the true-positive detection rates of the volume-based definitions were comparable to those of the voxel-count definition, and the relative ordering of methods remained unchanged across SNR columns within each panel of Figure 4.



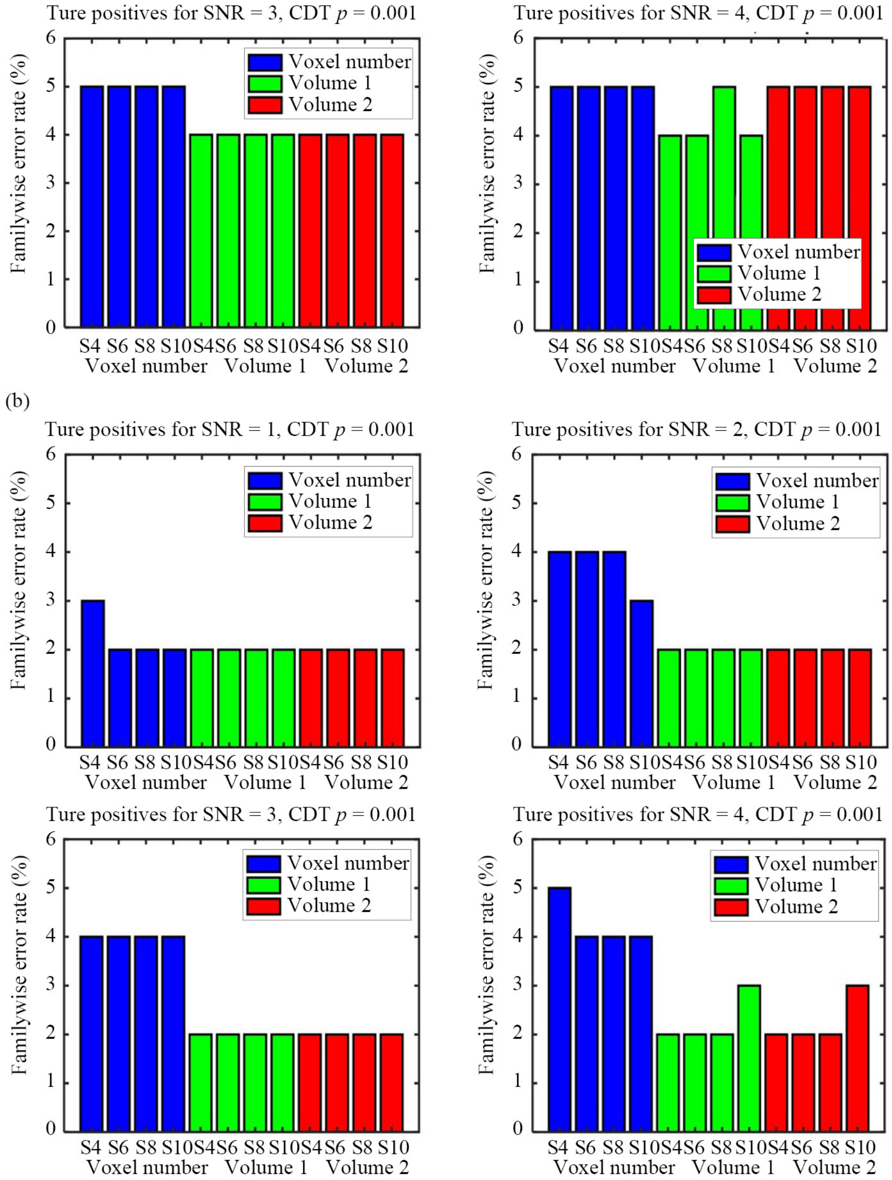


**Figure 4.** Sensitivity on signal-added simulations (five planted clusters) for (a) CDT $p = 0.001$ and (b) CDT $p = 0.01$. From left to right within each panel: SNR = 1, 2, 3, 4. The $x$-axis denotes smoothing levels (FWHM)

## 4. Discussion

This work examined cluster-wise inference in fMRI by redefining cluster size at the geometric (volume) level and benchmarking it against the conventional voxel-count extent across simulated Gaussian nulls, real resting-state group analyses, and signal-added simulations. Three empirical patterns are prominent. First, on simulated nulls, the volume-based definitions kept FWER at (or very close to) nominal across all factor combinations, whereas voxel-count extent systematically inflated FWER. Second, on real data, volume-based definitions reduced FWER relative to voxel-counting but residual inflation remained-most notably for one-sample tests-highlighting modeling gaps between ideal RFT assumptions and real fMRI. Third, sensitivity on signal-added simulations was comparable between volume-based and voxel-count definitions, indicating no practical loss of detection power when moving to a geometric definition. Below we discuss the methodological implications of these patterns.

**Why redefining cluster size matters.** RFT is formulated in continuous space: excursion-set geometry is summarized via intrinsic volumes and EC densities, and cluster "size" is a volume in $R^3$. The routine practice of measuring extent by voxel count introduces a discretization mismatch: voxel cardinality conflates sampling resolution with geometry and is sensitive to lattice idiosyncrasies. The observed FWER improvements with the volume-based definitions are consistent with the idea that aligning the observed statistic (cluster size) with the theoretical reference (expected cluster volume) reduces calibration error when mapping Eq. (expected EC) and cluster-size tails to a cluster-extent threshold. In short, the proposed definition addresses a modeling inconsistency rather than tuning an implementation detail.

**CDT and smoothness.** Across both simulated and real data, CDT $p = 0.001$ produced lower FWER than CDT $p = 0.01$, a pattern expected from RFT's high-threshold asymptotics: the Poisson clumping and EC-based approximations are sharper when excursion sets comprise isolated peaks rather than complex, percolating topologies. The small but visible differences between corrected and default EC at lower smoothness suggest that EC-density choices mainly matter when smoothness is modest, precisely where asymptotic assumptions are most strained. Notably, the geometric redefinition improved FWER across the entire CDT-by-smoothness grid without needing to optimize thresholds ad hoc.

**Connectivity invariance.** Because the geometric definition operates at the volume level, it is less sensitive to how connectivity is defined on the voxel lattice. Empirically, the 6- vs. 18-connectivity results were indistinguishable, so we reported 6-connectivity only. This invariance is desirable: it implies that statistical conclusions depend on excursion-set geometry rather than on arbitrary lattice conventions (e.g., face vs. edge adjacency).

**One-sample vs. two-sample behavior and paradigm effects.** On real resting-state data, one-sample tests exhibited higher FWER than two-sample tests, and the E1 paradigm yielded higher FWER than B1/B2/E2 under matched settings. While our study did not attempt to adjudicate causes, two non-exclusive explanations are plausible and consistent with prior discussions of resting-state "nulls": (i) one-sample designs can aggregate consistent physiological or processing-related structure across subjects, creating apparent effects that violate the GRF null; and (ii) certain event-like paradigms (e.g., E1) may align more strongly with common temporal fluctuations, increasing excursion-set connectivity at lenient CDTs or lower smoothness. The key point for practice is empirical: the volume-based definition attenuated these effects yet did not eliminate them in real data, reaffirming that geometric alignment helps but does not wholly resolve model-data mismatches.

**Sensitivity and practical trade-offs.** In signal-added simulations, detection rates for the volume-based definitions were on par with voxel-counting across SNRs, smoothing levels, and CDTs. Thus, improved FWER control did not come at an observable cost to power in these scenarios. Together with the connectivity invariance and CDT robustness, this supports the practical viability of using geometric cluster size in routine group studies.

**Methodological implications.** For users reliant on RFT-based cluster inference, two takeaways are immediate. First, compute and threshold clusters at the volume level to keep the test statistic on the same geometric scale as the RFT reference, thereby improving nominal FWER in conditions where assumptions are approximately met. Second, prefer higher CDTs and ensure adequate smoothness to stabilize EC-based calibrations; when smoothness is modest, the choice between corrected vs. default EC can slightly shift thresholds, but the geometric definition remains beneficial.

**Limitations.** Our inference still inherits RFT's classical approximations: the expected-clusters $\approx$ expected-EC equivalence at high thresholds; the Friston parametric form [36] for cluster-size tails; the Poisson model for the number of clusters exceeding a size threshold; and the independence assumption underpinning $E\left[K_u\right] = E\left[V_u\right] / E\left[C_u\right]$. Real fMRI violates stationarity and exact Gaussianity to varying degrees, and smoothness estimation remains an additional source of uncertainty. The residual FWER inflation for one-sample tests on real data underscores these points and motivates complementary nonparametric checks when feasible.

**Future directions.** Three avenues follow naturally from our findings. (i) Model-based calibration: regress empirically observed EC curves on the Gaussian kinematic formula (Lipschitz-Killing regression) to reduce reliance on explicit smoothness estimation and potentially improve calibration in low-smoothness regimes. (ii) Nonstationarity diagnostics: combine the geometric cluster-size definition with localized smoothness maps or robust variance estimators to mitigate spatial heterogeneity. (iii) Hybrid inference: integrate the geometric definition with modern resampling (e.g., permutation) or with alternative cluster statistics (e.g., TFCE variants) to probe robustness beyond the strict RFT setting.

In summary, redefining cluster size at the geometric volume level brings the observed statistic into principled alignment with RFT's continuous-space theory. This alignment yielded near-nominal FWER on simulated nulls, lower FWER on real resting-state analyses, no discernible loss of sensitivity, and insensitivity to connectivity choices. These properties argue for adopting geometric cluster size as a drop-in refinement of RFT-based cluster inference, while recognizing that full nominal control on real data will continue to depend on how closely analysis conditions approximate the assumptions of the underlying random-field model.

## 5. Conclusions

We evaluated a geometric (volume-level) definition of cluster size for RFT-based fMRI inference against the conventional voxel-count extent across simulated Gaussian nulls, real resting-state data, and signal-added simulations. The volume-based definitions achieved near-nominal FWER in simulations, reduced FWER relative to voxel-counting on real data, and maintained comparable detection performance. Results were insensitive to connectivity choice (6 vs. 18), so only 6-connectivity is reported. Across factors, CDT $p = 0.001$ yielded lower FWER than CDT $p = 0.01$, and differences between corrected and default EC were small and mainly visible at lower smoothness.

Scientifically, our contribution is to align the discrete extent statistic with the continuous-space assumptions of RFT by defining cluster size in intrinsic volume units via counts of fully contained $2 \times 2 \times 2$ cubelets, with a deterministic voxel-count to volume-extent mapping (Table 1). This makes results comparable across resolutions/smoothing, and the change is drop-in for SPM-style pipelines (no alteration of thresholding logic). Computational overhead is minimal, and the method applies uniformly to the Z/T/F families through the EC-density framework. Practically, users seeking stricter error control may prefer CDT $p = 0.001$ and higher smoothness when feasible, while noting that residual inflation on real data-most evident for one-sample tests-points to the value of improved smoothness/nonstationarity handling and complementary nonparametric checks as future work.

## Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (No. DUT24YG115).

## Conflict of interest

The authors declare no conflicts of interest.

## References

[1] Gray LF, Adler RJ. The geometry of random fields. *Journal of the American Statistical Association*. 1982; 77(380): 937. Available from: https://doi.org/10.2307/2287334.

[2] Worsley KJ. Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *The Annals of Statistics*. 1995; 23(2): 640-669. Available from: https://doi.org/10.1214/aos/1176324540.

[3] Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*. 1992; 12(6): 900-918. Available from: https://doi.org/10.1038/jcbfm.1992.127.

[4] Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*. 1996; 4(1): 58-73.

[5] Adler RJ. On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*. 2000; 10(1): 1-74.

[6] Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*. 2009; 45: S173-S186. Available from: https://doi.org/10.1016/j.neuroimage.2008.10.055.

[7] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004; 23: S208-S219. Available from: https://doi.org/10.1016/j.neuroimage.2004.07.051.

[8] Andreella A, Vesely A, Weeda W, Goeman J. Selective inference for fMRI cluster-wise analysis, issues, and recommendations for critical vector selection: A comment on Blain et al. *Imaging Neuroscience*. 2024; 2: 1-7. Available from: https://doi.org/10.1162/imag_a_00198.

[9] Ostwald D, Schneider S, Bruckner R, Horvath L. Random field theory-based *p*-values: A review of the SPM implementation. *arXiv:1808.04075*. 2021. Available from: http://arxiv.org/abs/1808.04075.

[10] Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(28): 7900-7905. Available from: https://doi.org/10.1073/pnas.1602413113.

[11] Eklund A, Knutsson H, Nichols TE. Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. *Human Brain Mapping*. 2018; 40(7): 2017-2032. Available from: https://doi.org/10.1002/hbm.24350.

[12] Gopinath K, Krishnamurthy V, Sathian K. Accounting for non-Gaussian sources of spatial correlation in parametric functional magnetic resonance imaging paradigms I: Revisiting cluster-based inferences. *Brain Connectivity*. 2018; 8(1): 1-9. Available from: https://doi.org/10.1089/brain.2017.0521.

[13] Gopinath K, Krishnamurthy V, Lacey S, Sathian K. Accounting for non-Gaussian sources of spatial correlation in parametric functional magnetic resonance imaging paradigms II: A method to obtain first-level analysis residuals with uniform and Gaussian spatial autocorrelation function and independent and identically distributed time-series. *Brain Connectivity*. 2017; 8(1): 10-21. Available from: https://doi.org/10.1089/brain.2017.0522.

[14] Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. fMRI clustering in AFNI: False positive rates redux. *Brain Connectivity*. 2017; 7(3): 152-171. Available from: https://doi.org/10.1089/brain.2016.0475.

[15] Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences*. 2017; 114(17): E3370-E3371. Available from: https://doi.org/10.1073/pnas.1614961114.

[16] Eklund A, Knutsson H, Nichols TE. Reply to Chen et al.: Parametric methods for cluster inference perform worse for two-sided *t*-tests. *Human Brain Mapping*. 2018; 40(5): 1689-1691. Available from: https://doi.org/10.1002/hbm.24465.

[17] Mueller K, Lepsien J, Möller HE, Lohmann G. Commentary: Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Frontiers in Human Neuroscience*. 2017; 113: 7900-7905. Available from: https://doi.org/10.3389/fnhum.2017.00345.

[18] Slotnick SD. Resting-state fMRI data reflects default network activity rather than null data: A defense of commonly employed methods to correct for multiple comparisons. *Cognitive Neuroscience*. 2017; 8(3): 141-143. Available from: https://doi.org/10.1080/17588928.2016.1273892.

[19] Slotnick SD. Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. *Cognitive Neuroscience*. 2017; 8(3): 150-155. Available from: https://doi.org/10.1080/17588928.2017.1319350.

[20] Cox RW. Equitable thresholding and clustering (ETAC): A novel method for fMRI clustering in AFNI. *Brain Connectivity*. 2019; 9(7): 529-538. Available from: https://doi.org/10.1089/brain.2019.0666.

[21] Spisák T, Spisák Z, Zunhammer M, Bingel U, Smith S, Nichols T, et al. Probabilistic TFCE: A generalized combination of cluster size and voxel intensity to increase statistical power. *NeuroImage*. 2019; 185: 12-26. Available from: https://doi.org/10.1016/j.neuroimage.2018.09.078.

[22] Bansal R, Peterson BS. Cluster-level statistical inference in fMRI datasets: The unexpected behavior of random fields in high dimensions. *Magnetic Resonance Imaging*. 2018; 49: 101-115. Available from: https://doi.org/10.1016/j.mri.2018.01.004.

[23] Taylor J, Takemura A, Adler RJ. Validity of the expected Euler characteristic heuristic. *Annals of Probability*. 2005; 33(4): 1362-1396. Available from: https://doi.org/10.1214/009117905000000099.

[24] Li H, Nickerson LD, Xiong J, Zou Q, Fan Y, Ma Y, et al. A high performance 3D cluster-based test of unsmoothed fMRI data. *NeuroImage*. 2014; 98: 537-546. Available from: https://doi.org/10.1016/j.neuroimage.2014.05.015.

[25] Li H, Nickerson LD, Zhao X, Nichols TE, Gao JH. A voxelation-corrected non-stationary 3D cluster-size test based on random field theory. *NeuroImage*. 2015; 118: 676-682. Available from: https://doi.org/10.1016/j.neuroimage.2015.05.094.

[26] Maren AJ. The 2-D cluster variation method: Topography illustrations and their enthalpy parameter correlations. *Entropy*. 2021; 23(3): 319. Available from: https://doi.org/10.3390/e23030319.

[27] Maullin-Sapey T, Schwartzman A, Nichols TE. Spatial confidence regions for combinations of excursion sets in image analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2024; 86(1): 177-193. Available from: https://doi.org/10.1093/jrsssb/qkad104.

[28] Ledberg A, Åkerman S, Roland PE. Estimation of the probabilities of 3D clusters in functional brain images. *NeuroImage*. 1998; 8(2): 113-128. Available from: https://doi.org/10.1006/nimg.1998.0336.

[29] Vandekar SN, Satterthwaite TD, Xia CH, Ruparel K, Gur RC, Gur RE, et al. Robust spatial extent inference with a semiparametric bootstrap joint testing procedure. *Biometrics*. 2018: 75(4): 1145-1155. Available from: https://doi.org/10.1111/biom.13114.

[30] Schwartzman A, Telschow F. Peak $p$-values and false discovery rate inference in neuroimaging. *NeuroImage*. 2019; 197: 402-413. Available from: https://doi.org/10.1016/j.neuroimage.2019.04.041.

[31] Adler RJ, Bartz K, Kou SC, Monod A. Estimating thresholding levels for random fields via Euler characteristics. *arXiv:1704.08562*. 2017. Available from: http://arxiv.org/abs/1704.08562.

[32] Worsley KJ. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. San Diego: Academic Press; 2007.

[33] Worsley KJ. Testing for signals with unknown location and scale in a chi-square random field, with an application to fMRI. *Advances in Applied Probability*. 2001; 793: 1-29.

[34] Worsley KJ. Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage*. 1996; 9(1): 15-26. Available from: https://doi.org/10.1006/nimg.1996.0248.

[35] Worsley KJ. Local maxima and the expected Euler characteristic of excursion sets of $\chi^2$, $F$ and $t$ fields. *Advances in Applied Probability*. 1994; 26(1): 13-42. Available from: https://doi.org/10.2307/1427576.

[36] Evans AC. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*. 1994; 1(3): 210-220. Available from: https://doi.org/10.1002/hbm.460010306.