Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

On the identifiability of covariance from Betti curves and Euler characteristic of Gaussian random fields: information dimension bounds and inversion diagnostics

Kun Wei

conkty@hotmail.com

Huzhou College

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11739599/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

On the identifiability of covariance from Betti curves and Euler characteristic of Gaussian random fields: information dimension bounds and inversion diagnostics

Kun Wei[1*]

[1*]School of Electronic Information, Huzhou College, No. 3333, Chuangye Road, Huzhou, Zhejiang Province, China.

Corresponding author(s). E-mail(s): conkty@hotmail.com;

Abstract

Topological summaries—Betti curves and the Euler characteristic curve—are increasingly used to invert the covariance of stationary Gaussian random fields, on the untested assumption that curve shape determines the covariance parameters. We ask how much covariance information such statistics can carry.

The mean Euler characteristic curve depends exactly on the first two spectral moments. The mean Betti curves are additionally controlled by higher moments, but that control is an upper bound rather than a functional dependence, and its dimension holds only at a numerical tolerance. When an observable depends on the parameters only through the second moment, its Jacobian is rank one, so the ill-posedness of the inversion is structural rather than algorithmic.

Large ensembles of one-dimensional paths and two-dimensional simulations confirm this picture: fields with matched low-order moments have Euler curves coinciding to within noise, while their Betti curves separate with the fourth moment; matching the first three moments exactly and altering only a hidden higher moment leaves the curves inside the resampling floor, so two moments are insufficient and three suffice at that tolerance. Three diagnostics follow: differencing two curves to amplify the fourth-moment signal, a condition-number criterion for unidentifiability, and the curve variance, which lifts the identifiable dimension from finite to infinite.

Shape uniqueness therefore fails in general: distinct covariances can yield identical Euler curves, and topological curves recover at most a low-order moment functional of the covariance. Topological inversion is an inverse problem with explicit information bounds, not a routine algorithm.

Keywords: Betti curve; Euler characteristic; identifiability; Gaussian random field; spectral moment; inverse problem

1 Introduction

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a stationary, isotropic, ergodic Gaussian random field, fully specified by its covariance function $C(h)$ or, equivalently, by its power spectrum $S(\omega)$. For a threshold $u \in \mathbb{R}$, the excursion set $\{x : f(x) \geq u\}$ is topologically characterized by the Betti numbers $\beta_k(u)$ $(k = 0, \ldots, d-1)$ and the Euler characteristic $\chi(u)$.

The inverse problem of using these topological curves as observables to recover C has been

widely adopted in cosmology Heydenreich et al. (2021); Jalali Kanafi et al. (2024); Wang et al. (2025); Calles et al. (2025); Dark Energy Survey Collaboration (2025), in the analysis of amorphous materials and porous microstructures Minamitani et al. (2022, 2023, 2025), and in other fields, often combined with machine learning and simulation-based inference, and has become an important tool for parameter estimation. However, nearly all applications implicitly assume that the map $C \mapsto (\beta_0, \ldots, \beta_{d-1}, \chi)$ is injective on the parameter family under consideration.

The basic question this paper addresses is: **how much information does this map carry, and which functionals of the covariance can it determine uniquely?**

This paper reframes the problem from algorithmic inversion to foundational identifiability analysis. Our main results are explicit identifiability statements—definitions, control bounds, and tolerance-level verification—together with a structural explanation of inversion ill-posedness and three operational diagnostics. In brief: (i) χ depends exactly on the first two spectral moments (λ_0, λ_2), giving identification dimension 2; (ii) each β_k admits a three-moment control bound, while the three-dimensionality of the curve itself is a numerical statement at the pre-measured noise floor of 3.8% for β_0 and 8% for β_1; (iii) the Jacobian of mean Betti numbers with respect to any parameter whose influence is mediated solely through λ_2 is rank-one (a direct consequence of the chain rule), so ill-posedness is structural; (iv) the tolerance-level statement is two-sided: two moments are insufficient and three are sufficient for β_0; (v) variance statistics (Fano factors) inject infinite-dimensional information and break the mean-level degeneracy.

The remainder of this paper is organized as follows. Section 2 establishes the theoretical framework, including the dimension bounds and rank-one Jacobian structure. Section 3 positions this work within the recent literature. Section 4 describes the numerical methods. Section 5 presents the numerical verification of each claim. Section 6 tests robustness against lattice spacing, threshold density, and observation window. Section 7 discusses implications, limitations, and concludes.

2 Theoretical framework

2.1 Spectral moment convention

Let $\lambda_j = \int_{\mathbb{R}^d} |\omega|^j S(\omega)\, d\omega$ denote the j-th moment of the spectral measure, so that $\lambda_0 = \mathrm{Var}(f)$ and $\lambda_2 = \mathrm{Var}(\nabla f)/d$. This convention is used throughout; note that some sources define λ_4 with an extra factor of 2.

2.2 One dimension: Rice formula and unidentifiability

In one dimension, $\beta_0(u)$ counts upcrossings. The Rice formula Rice (1944) yields

$$\frac{\mathbb{E}[\beta_0(u)]}{|\Omega|} = \frac{1}{2\pi}\sqrt{\frac{\lambda_2}{\lambda_0}}\, e^{-u^2/(2\lambda_0)}. \tag{1}$$

The left-hand side is a density per unit length ($|\Omega|$ is the length of the one-dimensional domain), consistent with the per-unit-volume convention of eq. (2). The right-hand side depends only on (λ_0, λ_2). Therefore any transformation of one-dimensional Gaussian fields that preserves these two moments leaves the mean Betti curve invariant—a constructive recipe for unidentifiable families.

2.3 Two dimensions: information decoupling of χ and β

In two dimensions, the Gaussian Kinematic Formula (GKF) together with Morse theory Adler and Taylor (2007) implies that $\mathbb{E}[\chi(u)]$ also depends only on (λ_0, λ_2). In contrast, $\mathbb{E}[\beta_0(u)]$ and $\mathbb{E}[\beta_1(u)]$ depend additionally on λ_4. The structural reason is that the λ_4-dependent terms in $\chi = \beta_0 - \beta_1$ cancel in the alternating sum (a topological telescoping): χ sees strictly less spectral information than the individual Betti numbers.

2.4 Identifiability definitions and information dimension

Before stating the results we make two notions precise. Identifiability in the standard statistical sense means injectivity of the parameter map Rothenberg (1971); the weaker question of interest here is whether the observable enters through a small number of moments only. Let $\mathcal{P}$ be a parameter family and let $M_m(\theta) =$

$(\lambda_0, \lambda_2, \dots, \lambda_{2m-2})(\theta)$ be the *nested* sequence of even moments $(M_1 \subset M_2 \subset \cdots)$.

Definition 1 (Identifiability) *If $\theta \mapsto \mathcal{O}(\theta)$ is injective on $\mathcal{P}$, that is, $\mathcal{O}(\theta) = \mathcal{O}(\theta')$ implies $\theta = \theta'$, then $\mathcal{O}$ is* identifiable *on $\mathcal{P}$.*

Definition 2 (Identification dimension) *If there is a function Φ such that $\mathcal{O}(\theta) = \Phi(M_m(\theta))$ for all $\theta \in \mathcal{P}$, we say that $\mathcal{O}$ is* determined *by M_m; the smallest such m is the identification dimension $\dim_{\mathcal{P}}(\mathcal{O})$ relative to this nested sequence.*

Definition 3 (ε-determination) *Given $\varepsilon > 0$ and a norm $\|\cdot\|$, if $M_m(\theta) = M_m(\theta')$ always implies $\|\mathcal{O}(\theta) - \mathcal{O}(\theta')\| \le \varepsilon$, then $\mathcal{O}$ is determined by these m moments to tolerance ε; the smallest such m is written $\dim_\varepsilon(\mathcal{O})$.*

The three notions are related as follows: definition 1 is standard identifiability, definition 2 is exact functional dependence, and definition 3 is its testable weakening. Numerically we can only test definition 3, with ε the relative L_1 distance between curves, and ε itself fixed by an independently measured noise floor (section 5.8). The two propositions and the corollary below have distinct roles: the first gives an exact dimension, the second only a control bound, and the corollary translates the exact dimension into standard identifiability.

Proposition 1 (Two-dimensionality of the Euler characteristic) *Let f be a zero-mean, stationary, isotropic Gaussian field on $\mathbb{T}^d_L$ with $C \in C^4$, and let u be a regular level. Then*

$$\frac{\mathbb{E}[\chi(A_u)]}{|\Omega|} = (2\pi)^{-(d+1)/2} \Big(\frac{\lambda_2}{\lambda_0}\Big)^{d/2} \times H_{d-1}\Big(\frac{u}{\sqrt{\lambda_0}}\Big) e^{-u^2/(2\lambda_0)}, \quad (2)$$

whose right-hand side contains (λ_0, λ_2) only, and the map $(\lambda_0, \lambda_2) \mapsto$ curve is injective, so $\dim_{\mathcal{P}}(\chi) = 2$ ($\mathcal{P}$ allows a free λ_0; if $\sigma = 1$ is fixed, then $\dim_{\mathcal{P}}(\chi) = 1$ on that family).

Proof Set $g = f/\sqrt{\lambda_0}$; then g is zero-mean, unit-variance and isotropic with $\lambda_2(g) = \lambda_2/\lambda_0$. Applying GKF Adler and Taylor (2007) to isotropic Gaussian fields gives $\mathbb{E}[\chi(A_u)]/|\Omega| = \sum_{j=0}^{d} L_j(\Omega)\, \rho_j(u)$. On the torus $L_j = 0$ for $j < d$ and $L_d = |\Omega|$, while

$$\rho_j(u) = (2\pi)^{-(j+1)/2} \big(\lambda_2(g)\big)^{j/2} H_{j-1}(u)\, e^{-u^2/2} \quad (3)$$

contains $\lambda_2(g)$ only. Substituting g back yields eq. (2): for $d = 1$ it reduces to Rice's formula eq. (1), and for $d = 3$ it is the closed form used in section 5.5. It remains to show that the curve determines (λ_0, λ_2) uniquely. Since $H_{d-1}(u/\sqrt{\lambda_0})$ grows only polynomially in u, eq. (2) gives

$$\lim_{u \to +\infty} \frac{1}{u^2} \log \frac{\mathbb{E}[\chi(A_u)]}{|\Omega|} = -\frac{1}{2\lambda_0}, \quad (4)$$

so the tail decay rate identifies λ_0; once λ_0 is fixed, the amplitude at any regular level identifies $(\lambda_2/\lambda_0)^{d/2}$. Hence $(\lambda_0, \lambda_2) \mapsto$ curve is injective, its image is two-dimensional, and no one-dimensional scalar can reproduce it. (Equivalently, $\mathbb{E}[\chi]$ is real-analytic in u, so two parameter pairs that agree on any open interval agree everywhere; this argument does not rely on the numerically unobservable tail limit.) □

Corollary 2 (Identifiability boundary of χ) *proposition 1 converts the exact dimension into standard identifiability: if $\theta \mapsto (\lambda_0, \lambda_2)(\theta)$ is not injective on $\mathcal{P}$, i.e. there exist $\theta \neq \theta'$ with equal first two moments, then $\mathbb{E}[\chi]$ is not identifiable on $\mathcal{P}$ (definition 1): every equivalence class of $\mathcal{P}$ sharing (λ_0, λ_2) lies on one and the same level set of χ.*

Proof By proposition 1, $\mathbb{E}[\chi] = \Phi(\lambda_0, \lambda_2)$. If $(\lambda_0, \lambda_2)(\theta) = (\lambda_0, \lambda_2)(\theta')$, the two points share Φ, so $\mathbb{E}[\chi]$ coincides while $\theta \neq \theta'$, and injectivity fails. □

Proposition 3 (Three-moment control bound for Betti numbers) *Under the assumptions of proposition 1, for each $k = 0, \dots, d-1$,*

$$\mathbb{E}[\beta_k(A_u)] \le \Lambda_k(u; \lambda_0, \lambda_2, \lambda_4) := \mathbb{E}[c_k(A_u)], \quad (5)$$

where $c_k(A_u)$ is the number of critical points of index k in A_u.

Proof Morse's inequality gives $\beta_k(A_u) \le c_k(A_u)$ pathwise Milnor (1963). Taking expectations and applying Kac–Rice,

$$\mathbb{E}[c_k(A_u)] = |\Omega| \int_u^\infty p_f(t)\, p_{\nabla f}(0) \times \mathbb{E}\big[|\det \nabla^2 f|\, \mathbf{1}\{\text{index} = k\} \mid f = t\big]\, dt. \quad (6)$$

Isotropy makes the joint law of $(f, \nabla f, \nabla^2 f)$ depend on $(\lambda_0, \lambda_2, \lambda_4)$ only: $\mathbb{E}[\nabla f \nabla f^{\mathsf{T}}] = \lambda_2 I$, the fourth-order tensor $\mathbb{E}[\nabla^2 f \otimes \nabla^2 f]$ is assembled from δ-tensors scaled by λ_4, and ∇f is independent of $(f, \nabla^2 f)$. Hence the integrand in eq. (6), and therefore Λ_k, contains the first three moments only. □

Remark 1 (A bound is not a functional dependence) eq. (5) is a control relation: $\mathbb{E}[\beta_k]$ is dominated by a functional of the first three moments, which does not make it a function of $(\lambda_0, \lambda_2, \lambda_4)$—any bounded quantity is dominated by a constant. Hence proposition 3 does not imply $\dim_{\mathcal{P}}(\beta_k) \leq 3$. Three-dimensionality is a *numerical* statement: section 5.8 establishes it at tolerance $\varepsilon = 3.8\%$ for β_0 (the 95th percentile of the same-parameter null) and 8% for β_1, and it is bounded from both sides—two moments are insufficient (section 5.3 separation) and three are sufficient (section 5.8 coincidence)—independently of eq. (5).

Remark 2 This three-dimensionality is attained on the Matérn family with $\nu > 2$. From $\lambda_2 = \sigma^2\nu/[\ell^2(\nu-1)]$ and the dimensionless ratio $\lambda_4\lambda_0/\lambda_2^2 = 4(\nu-1)/(\nu-2)$ (note that λ_4/λ_2^2 has dimension $1/\lambda_0$ and must be multiplied by $\lambda_0 = \sigma^2$), the map $(\sigma, \nu, \ell) \mapsto (\lambda_0, \lambda_2, \lambda_4)$ is invertible on this family (the last expression is strictly monotone in ν), so the three moments determine the three parameters; and section 5.8 shows that the full β_0 curve is determined by the three moments to within 3.8%. Together these give: within this family the β_0 curve determines (σ, ν, ℓ) uniquely (to numerical precision). For families with more than three parameters this uniqueness fails: a two-scale mixture and a pure Matérn field can match on the first three moments while having different covariances.

The dimension statements above all concern the mean curves. Variance is a different matter: for example, $\mathrm{Var}[\beta_0(u)]$ involves integrals over the entire range of $C(h)$, $C'(h)$, $C''(h)$; finitely many point constraints cannot determine the full curve. The gap between "mean-only" (finite dimension) and "mean plus variance" (infinite dimension) is therefore qualitative, not merely quantitative. This explains why Fano-factor-type second-order statistics can break the mean-level degeneracy.

2.5 Rank-one Jacobian structure

Suppose an observable $F(\theta)$ depends on parameters $\theta \in \mathbb{R}^p$ only through λ_2. By the chain rule,

$$J = \frac{\partial F}{\partial \lambda_2} \nabla \lambda_2^{\mathsf{T}} \tag{7}$$

is an outer product of rank at most one. Its null space is exactly the tangent space of the level set $\lambda_2 = \text{constant}$. Ill-posedness is therefore *structural*: no optimization algorithm, no matter how sophisticated, can recover directions in the null space. This is Rothenberg's rank condition Rothenberg (1971) in the present setting: a rank-deficient Jacobian of the observation map amounts to local non-identifiability.

2.6 Regularity–identifiability phase diagram

The convergence of spectral moments provides a regularity–identifiability link. The two-dimensional Matérn spectrum decays as $S(\omega) \sim |\omega|^{-2(\nu+1)}$, so $\lambda_j = \int |\omega|^j S$ converges iff $j - 2(\nu+1) < -2$, that is, iff $j < 2\nu$. Hence λ_4 is finite iff $\nu > 2$, and λ_6 iff $\nu > 3$. Thus fields with $\nu \in (2, 3]$ admit λ_4 but diverge at λ_6; claims attributing β_1 differences to λ_6 must be restricted to $\nu > 3$.

3 Related work

3.1 Asymptotic theory of random field topology

The large-sample theory of Betti numbers and Euler characteristic has been developed only recently: bootstrapping limits for persistent Betti numbers Roycraft et al. (2023), asymptotic normality Krebs and Polonik (2025), consistency and CLTs for multiparameter persistent Betti numbers Botnan and Hirsch (2023), smooth Gaussian fields and percolation Beliaev (2023), and a representation of Betti numbers/Euler characteristic as sums of discrete random variables Chingangbam (2025). These works concern the *distribution and fluctuations* of topological statistics; our question is complementary—what they *reveal* about the covariance.

3.2 Estimation of Euler characteristic curves and GKF

Telschow et al. Telschow et al. (2023) constructed consistent estimators for expected Euler characteristic curves of nonstationary smooth random fields. Meng et al. Meng et al. (2025) introduced the smooth Euler characteristic transform and established its statistical inference framework. GKF is the tool we use to prove that $\mathbb{E}[\chi]$ depends only on (λ_0, λ_2); neither line of work addresses identifiability of the covariance from these curves.

3.3 Topological inversion in cosmology

Betti curves and persistent homology have been used to constrain cosmological parameters: the early characterization of Betti numbers of Gaussian fields, which also reported that the curve shape varies with the spectral index Park et al. (2013)—the earliest quantitative benchmark in this line, and one of the few that treat both the smoothing scale and the spectral index as explicit control variables; later cosmological studies use the same topological observables but mostly as persistent homology of simulated catalogues rather than raw Betti curves Heydenreich et al. (2021); Tymchyshyn et al. (2024), and do not return to the fixed-smoothing, scan-index protocol; the same team's follow-up series, which gives analytic expectations for the Betti numbers, Euler characteristic and Minkowski functionals of 2D and 3D Gaussian fields Pranav et al. (2019a) and then treats critical points, excursion sets and persistent homology Pranav (2021); the non-Gaussian topology reported for CMB temperature fluctuations Pranav et al. (2019b); persistent Betti number constraints Heydenreich et al. (2021); neutrino-mass sensitivity Jalali Kanafi et al. (2024); Wang et al. (2025); machine-learning pipelines combining persistence images with PS/BS Calles et al. (2025); and the Dark Energy Survey Year 3 results with simulation-based inference Dark Energy Survey Collaboration (2025). All assume identifiability; we show that this assumption fails in general—for χ alone, and for any isotropic family with more than three parameters—and that on the three-parameter Matérn family it holds only through the Betti curves and only up to a $\lesssim 4.8\%$ spectral-shape residual (L_1 metric). For the Park et al. setting in particular, section 5.4 reproduces their parameter points and shows that the reported spectral variation of the Betti curves co-varies monotonically with the single scalar λ_4 within the scanned family ($R^2 = 0.92$; with R_s fixed, λ_4 and the spectral index are monotone reparameterisations, so spectral shape is not separately tested), while the invariance of the χ shape is the GKF statement itself.

3.4 Topological inverse problems in materials

Persistent homology has been used to build structure–property relationships: topological descriptors for thermal conductivity in amorphous silicon Minamitani et al. (2022), machine-learning potentials for amorphous structures Minamitani et al. (2023), and hierarchical structures in covalent amorphous solids Minamitani et al. (2025). These studies demonstrate predictive power; the uniqueness of their inverse maps, however, remains unexamined.

3.5 Relation to the above

Most of the applications above assume that the map from topological curves to covariance parameters is invertible. The addition made here is an explicit identifiability-dimension bound, a rank-one structural account of the ill-posedness, and three operational diagnostics (the κ threshold, the $\beta_1 - \chi$ difference, and Fano variance). We also note that the "shape varies with the spectrum" reported in the literature co-varies monotonically with the low-order spectral moments within the scanned power-law families, so spectral shape has not yet been isolated as an independent effect.

4 Numerical methods

4.1 Random field generation

Two-dimensional fields are generated by circulant embedding on $N \times N$ periodic grids, using the Matérn and power-law spectral families. One-dimensional paths are generated by spectral synthesis with 10^4 paths. The three-dimensional generator was validated separately: the one-dimensional line-crossing rate agrees with eq. (1) to a ratio of 0.98 at $u = 0$, and λ_4 inferred from $\mathrm{Var}(\Delta f)$ agrees with the closed form to within 7%, the expected second-order bias of the finite-difference operator, so the generator introduces no appreciable systematic error. The default configuration is $h = 1$, $\Delta u = 0.25$, $N = 512$ ($N = 128$ in three dimensions), with 120 fields per configuration (300 for the sufficiency experiment, to suppress the standard error at low thresholds); the baseline configurations are A: $(\nu, \ell) = (2.5, 28.87)$ and B: $(4.0, 25.82)$, both at $\lambda_2 = 0.002$. Robustness experiments additionally scan $h \in \{1, \ldots, 6\}$,

$\Delta u \in \{0.0625, \ldots, 0.5\}$ and $L \in \{128, \ldots, 512\}$. All experiments use fixed seeds; key results are cached for reproducibility.

4.2 Topological computation

Excursion sets use 8-connectivity on the lattice. The Euler characteristic χ is accumulated via a 2×2 local lookup table. Betti numbers are computed as β_0 (connected components), β_1 (cycles), with $\chi = \beta_0 - \beta_1$ as internal consistency check (negative β_1 triggers an implementation-error alert). Thresholds are sampled on a uniform grid.

4.3 Statistical criteria

Curve differences are assessed via per-threshold z-scores, median relative difference, and L_1 distance. Local identifiability is quantified via the singular-value spectrum and condition number κ of the Jacobian. Separating statistical power from bias is done by a L-scan at fixed field count N, combined with an N-scan at fixed L.

5 Results

5.1 One dimension: mean coincidence, variance separation

Under 10^4 paths, two field families matched on (λ_0, λ_2) produce coincident mean Betti curves (Fig. 1) with Rice ratios 0.999 and 1.008. A residual of about 1% shrinks monotonically with h (median relative difference 2.23% $\to$ 1.17% $\to$ 0.59% as h halves), following an $O(h)$ scaling consistent with the discretisation analysis of section 6; it is therefore attributed to the lattice rather than to sampling noise. Variance curves, by contrast, separate cleanly; Fano factors stabilize with sample size.

5.2 Variance: from finite to infinite dimension

The Fano factor $F(0)$ decreases monotonically across spectral parameters (0.897 $\to$ 0.575 $\to$ 0.431 $\to$ 0.349 $\to$ 0.314; Fig. 2). Because it is strictly monotone in ν, once (σ, ℓ) are fixed by the lower moments the observed $F(0)$ can be inverted directly for ν. By the delta method, $\mathrm{SE}[\hat{\nu}] \approx \mathrm{SE}[\hat{F}(0)]/|dF/d\nu|$; finite differences give $|dF/d\nu|$ falling from 1.07 at $\nu \in [1.2, 1.5]$ to 0.018 at $\nu \in [3, 5]$, so at fixed $F(0)$ precision the standard error of $\hat{\nu}$ is amplified by a factor of ~ 60—an estimation-theoretic explanation of the $\nu \gtrsim 3$ blind spot (section 7.3). Variance-based inversion covers the true value in 3/3 repetitions, and at $n = 5000$ paths $\mathrm{SE}[\hat{\nu}] = 0.092 < 0.1$, confirming the estimate that 5000 paths are needed to pin ν to ± 0.1. This is consistent in direction with the "infinite-dimensional variance" claim of section 2.4: the mean curves read out only finitely many moments, whereas the monotonicity of $F(0)$ supplies the one degree of freedom the mean level lacks.

5.3 Two dimensions: χ coincides, β separates with λ_4

Matérn fields matched on (λ_0, λ_2) have coincident χ curves ($L_1 \approx 3\%$), while β_0, β_1 separate monotonically with λ_4 (Fig. 3). To test whether this decoupling is specific to Matérn, we generate power-law spectra $P(k) = |k|^n e^{-k/k_c}$ for $n = 0, -1, -2, -3$ and tune k_c to match $r_2 = \lambda_2/\lambda_0$. All fields are normalised at generation to unit variance ($\lambda_0 = 1$; measured variances 0.986–1.008 for the two-dimensional fields), and the Matérn fields likewise take $\sigma^2 = 1$, so matching r_2 also matches (λ_0, λ_2); the 21.3% separation below cannot be attributed to a mismatch in λ_0. Cross-family pairwise relative L_1 distances (Fig. 4; 200 fields per configuration): χ 5.0%, β_0 21.3%, β_1 84.6%, i.e. the β_0 distance is more than four times that of χ and β_1 an order of magnitude above β_0, exactly what additional λ_4 dependence in β implies.

The 5.0% left in χ must be compared with its own noise floor: the null of section 5.8 comes from a 300-field half-sample (150 vs 150) and, rescaled by $\sqrt{N}$ to the present 200 vs 200 independent field sets, has a 95th percentile of about $4.4\% \times \sqrt{150/200} = 3.8\%$, so 5.0% is slightly above the statistical null. The excess of about 1.2 percentage points has the same sign and order as the discretisation bias at k_c up to 0.21 ($\ell/h \approx 5$; section 6) and is read as a systematic bias rather than a λ_4 dependence of χ; even if all of it were counted, the 21.3% of β_0 is still more than four times that of χ and the 84.6% of β_1 an order of magnitude larger, so the information decoupling is unaffected. The decoupling is therefore not an artifact of the Matérn family: the two-dimensionality of χ is guaranteed on any isotropic

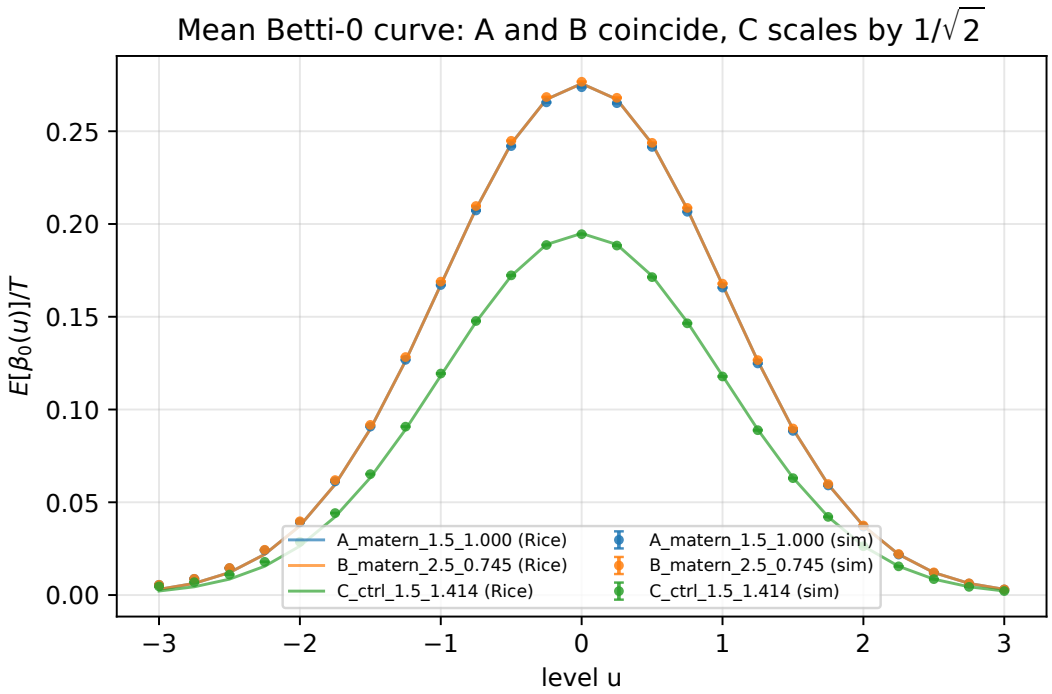

Fig. 1 One-dimensional mean Betti curves: matched (λ_0, λ_2) families coincide; Rice ratios 0.999, 1.008

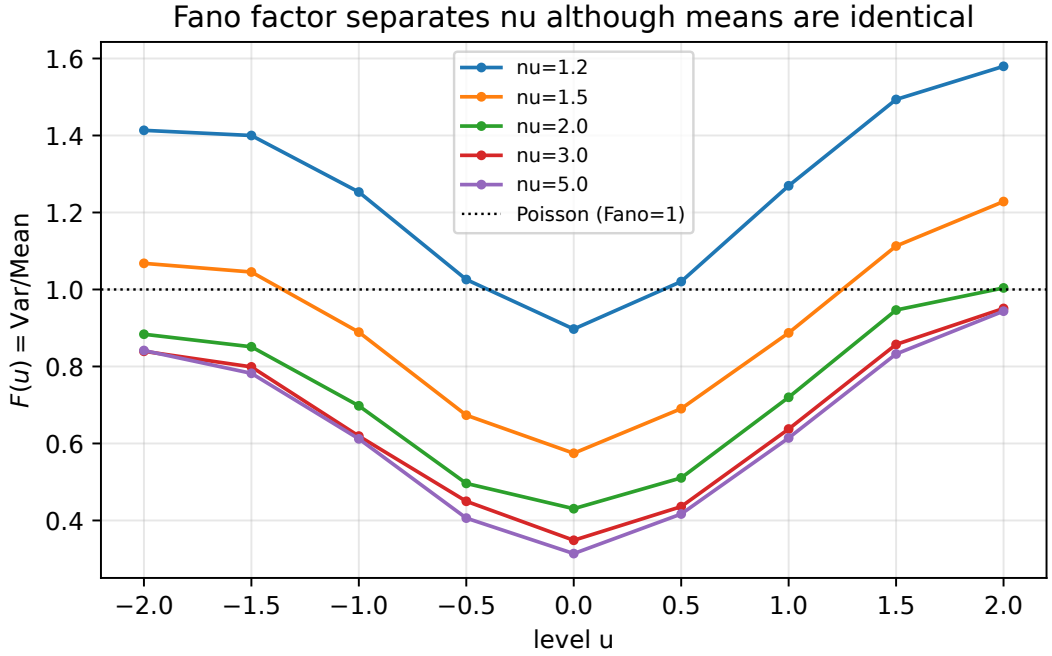

Fig. 2 Fano factor curves: strictly monotone ordering across spectral parameters breaks the mean-level degeneracy

Gaussian field by proposition 1, while the λ_4 separation of β appears consistently in both the Matérn and the power-law families. This is also consistent with the long-standing observation that Betti curves vary with spectral slope: within the scanned family that variation co-varies monotonically with low-order spectral moments, although spectral shape is not separately tested here.

5.4 Quantitative reproduction of the Park parameter points

This explanation can be compared directly with the observables reported in the existing literature. Park et al. (see Section 3) reported for a fixed Gaussian smoothing and a varying power-law index that (i) the shape of the χ (genus) curve is the same for all Gaussian fields, whereas (ii) the shape of the Betti curves depends on the spectral index n and becomes broader as n decreases Park et al. (2013). We reproduce their protocol on a two-dimensional torus: $P(k) = |k|^n e^{-k^2 R_s^2/2}$ with $R_s = 8$ grid units (fixed smoothing), $n = -2.5, \ldots, +1$ (eight values, including the $n = -1, 0, +1$ they used), $N = 512$ and 100 fields per configuration, and we compare peak-normalised curves (table 1 and fig. 5).

The outcome agrees with the literature and adds a quantitative decomposition that the literature did not attempt. First, at $n = -1, 0, +1$ the mean absolute deviation of the normalised χ curve from the $n = 0$ baseline is 2.4%, 0 and 0.8%, while over the same spectra the half-width of the normalised β_1 curve falls from 0.83 to 0.67 (a decrease of 19%): under the same spectral variation the shape of χ barely moves while that of β changes markedly, which is precisely what propositions 1 and 3 predict. Second, regressing the β_1 half-width over all eight indices on $h = \lambda_4\lambda_0/\lambda_2^2$ gives $W = 0.073\,h + 0.59$ ($R^2 = 0.918$; h spans 1.67–8.43 and W spans 0.67–1.14): the reported broadening as n decreases is carried by the single scalar λ_4 within this one-parameter family. But R_s is fixed and only n varies, so n, r_2, h and W are monotone reparameterisations of one another, and $R^2 = 0.918$ cannot distinguish "W determined by λ_4" from "W determined by the spectral index n": within the scanned power-law family W co-varies monotonically with h, and spectral shape itself is not separately tested. The conclusion supported here is therefore weaker than "spectral shape is not an extra degree of freedom".

This is a reproduction at the level of shape, not a numerical replica of the original values: the reference study used three-dimensional 256^3 fields with $R_G = 5$ pixel smoothing and centred on $n = -1, 0, +1$, whereas we use two-dimensional 512^2 fields with $R_s = 8$, so only normalised shapes and trends are compared. For very red spectra ($n \le -2$) the mean absolute residual of the normalised χ curve grows to 7.8%–12% (at $n = -2.5$ the pointwise maximum deviation is 33.7%–41.9% over four seeds, i.e. systematic rather than sampling noise); these fields have the fewest resolvable structures (≈ 6.6 connected components per field at $u = 0$ versus 25.1 at $n = +1$). The residual runs in the same direction as the $O(h^2)$ discretisation bias of section 6 and does not constitute evidence for a λ_4 dependence of χ; it enters none of our conclusions.

Table 1 Two-dimensional reproduction of the parameter points of Park et al. Park et al. (2013) ($N = 512$, 100 fields per configuration). $\Delta\tilde{\chi}$ is the mean absolute deviation of the normalised χ curve from the $n = 0$ baseline; $W(\beta_1)$ is the half-width of the normalised β_1 curve in units of u

n	$r_2 = \lambda_2/\lambda_0$	$h = \lambda_4\lambda_0/\lambda_2^2$	$\Delta\tilde{\chi}$	$W(\beta_1)$
−2.5	0.0028	8.43	12.1%	1.14
−2.0	0.0056	5.60	7.8%	1.08
−1.5	0.0102	3.84	4.9%	0.93
−1.0	0.0164	2.85	2.4%	0.83
−0.5	0.0237	2.31	1.3%	0.76
0	0.0313	2.00	0	0.72
+0.5	0.0391	1.80	0.7%	0.68
+1.0	0.0469	1.67	0.8%	0.67

5.5 Numerical check in three dimensions

The bound (5) is dimension-independent, but the voxel-based topological estimator is more fragile in $d = 3$, so we check this case separately. We use $N = 128$ and 120 independent fields. Configurations A ($\nu = 2.5$, $\ell = 16.67$) and B ($\nu = 6$, $\ell = 14.14$) exactly match $(\lambda_0, \lambda_2) = (1, 6\times10^{-3})$ while differing in λ_4 by a factor 2.4 ($\lambda_4^{(B)}/\lambda_4^{(A)} = 0.42$).

The results agree with propositions 1 and 3 (Fig. 6): the two χ curves coincide at all 15 thresholds, with a median relative difference of 2.5%,

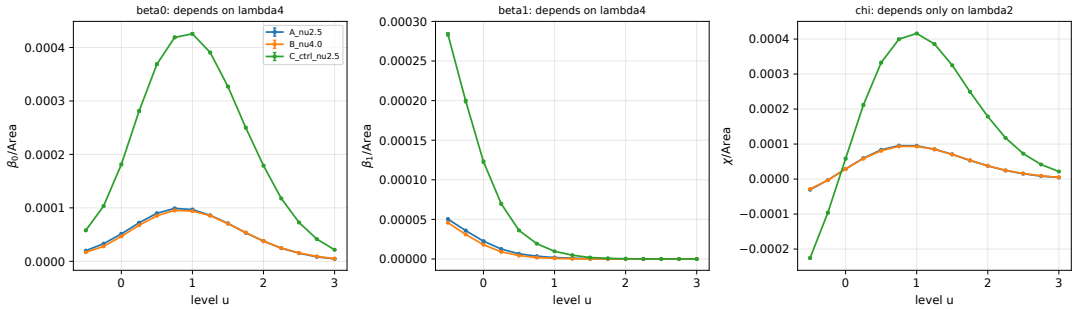

Fig. 3 Two-dimensional χ vs. Betti curves: matched (λ_0, λ_2) makes χ coincide while β_0, β_1 separate with λ_4

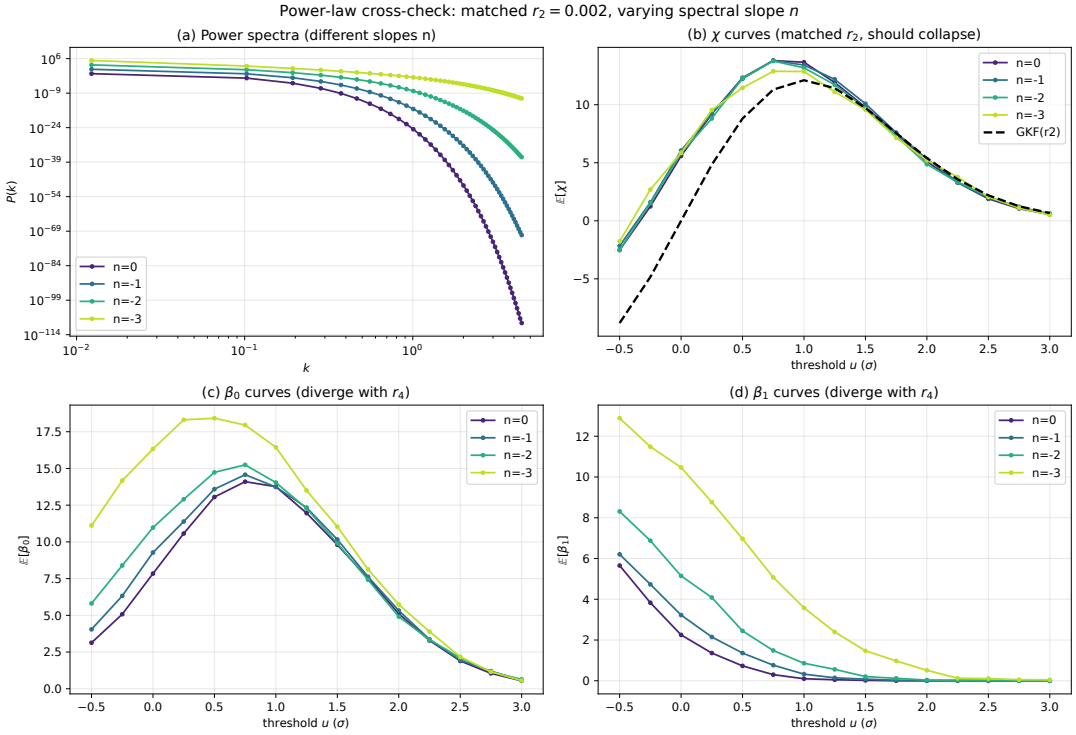

Fig. 4 Power-law cross-check: pairwise L_1 distances χ 5.0%, β_0 21.3%, β_1 84.6% across four spectral families with matched r_2

$\max|z| = 1.64$ and no threshold beyond 2σ; β_0, in contrast, separates by λ_4, with a median relative difference of 13.7%, $\max|z| = 6.4$ and 60% of thresholds beyond 2σ. In three dimensions, too, χ is blind to λ_4 whereas β_0 is not.

The Euler characteristic of the voxel complex ($V-E+F-C$, six-connected foreground) carries a threshold-dependent bias at $\ell/h \approx 17$: its *absolute value* $|\mathrm{MC} - \chi_{\mathrm{closed}}|$ is largest near the percolation threshold ($u \approx 0.8$; Fig. 6(c)), whereas the *ratio* $\mathrm{MC}/\chi_{\mathrm{closed}}$ is largest at the positive peak of χ ($u \approx 1.6$), where the closed form is about half the MC value; only for $u \gtrsim 2.5$ does the ratio fall back to 1.3–1.4. A direct numerical Kac–Rice quadrature of $\mathbb{E}[\chi(A_u)]/|\Omega| = (2\pi)^{-2}\lambda_2^{3/2}(u^2-1)e^{-u^2/2}$ over the six-dimensional conditional Hessian law agrees with the closed form pointwise to within 1% (zero at $u = 0.975$), so the offset is attributable to the voxel estimator rather than to the formula. Since both configurations share the estimator, it appears as a common mode in Fig. 6(c) (the two

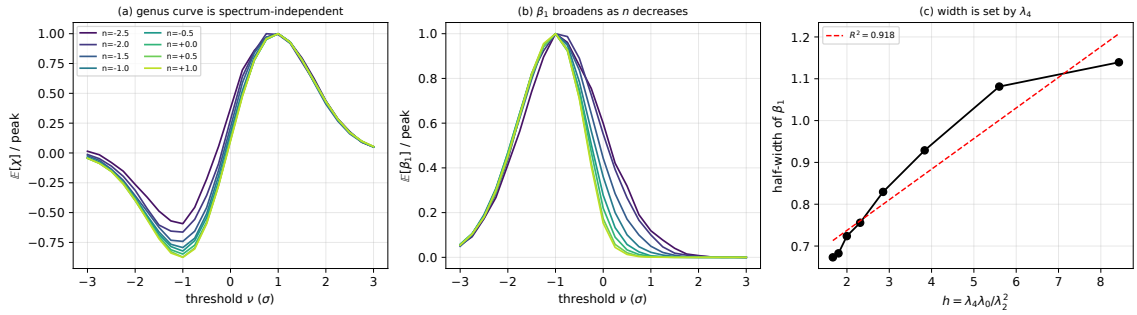

Fig. 5 Two-dimensional reproduction of the Park et al. parameter points ($R_s = 8$, $N = 512$, 100 fields per configuration). (a) After peak normalisation the χ curves nearly coincide across spectral indices; (b) the normalised β_1 curves broaden systematically as n decreases; (c) the β_1 half-width is linear in $h = \lambda_4\lambda_0/\lambda_2^2$, $R^2 = 0.918$

deviation curves differ by a median of 4%) and does not affect the two relative criteria above.

5.6 β_1 as a λ_4 probe

The difference $\beta_1 - \chi$ extracts the λ_4 signal: for the same A/B pair, the change in λ_4 leaves a $\sim 4\%$ curve difference in β_0 and a $\sim 32\%$ difference in β_1. Whether this factor is due to λ_4 alone has to be checked separately: the reference configuration A has $\nu = 2.5$, whose two-dimensional λ_6 diverges (section 2.6), so the A/B contrast mixes the change in λ_4 with the transition from a divergent to a finite tail moment. We therefore fixed λ_6 exactly in a restricted family in which every component has $\nu > 3$, and varied only λ_4 (Fig. 7): the reference is a pure Matérn field with $\nu = 3.2$ ($\lambda_4/\lambda_2^2 = 7.33$, $\lambda_6/\lambda_2^3 = 484.0$), and four controls are two-scale mixtures of $\nu = 3.2$ and $\nu = 12$ that exactly match $(\lambda_0, \lambda_2, \lambda_6)$ while taking $\lambda_4/\lambda_2^2 = 5.6, 9.0, 11.0, 12.5$ (300 fields per configuration).

The mixing weights are far from uniform: the $\nu = 3.2$ weight is 0.98/0.95/0.92 for C90/C110/C125 but only 0.011 for C56, whose $\nu = 3.2$ component nevertheless carries 4.4 times the total λ_2. C56 is therefore a mixture dominated by $\nu = 12$ in weight but with its higher moments supplied by the short-scale component: its $(\lambda_4/\lambda_2^2, \lambda_6/\lambda_2^3) = (5.60, 483.9)$ differs both from pure $\nu = 12$ $((4.4, 32.3))$ and from pure $\nu = 3.2$ $((7.33, 484.0))$.

table 2 lists the moments and the $u = 0$ values configuration by configuration: ordered by λ_4/λ_2^2, $\beta_1(0)$ rises monotonically from 3.81 to 8.10 (corresponding to $5.6 \to 7.33 \to 9.0 \to 11.0 \to 12.5$), while $\chi(0)$ stays between 7.5 and 8.1. The two endpoint mixtures 5.6 and 12.5 differ by a median relative difference of 58.8%, no less than the 32% of A/B; over the same runs the median relative difference of χ stays at 1.0%–2.2% (null control 1.3%), confirming that the (λ_0, λ_2) matching is effective.

Every component has $\nu_1 = 3.2 < 4$, so λ_8/λ_2^4 diverges in all five configurations and λ_8 is not an admissible explanatory variable; the variation comes only from λ_4 and from the two-scale composition, and the latter co-varies monotonically with λ_4 along the constraint manifold, so this design cannot fully separate the two. What can be established is that λ_4 alone suffices to produce an effect no smaller than the observed A/B difference, so *there is no evidence* that the divergence of λ_6 is needed to explain the 32%, but the two contributions cannot be separated. For inversion the estimator is useful only when the λ_4 difference is large: taking the largest residual 4.8% of section 5.8 as a conservative upper bound on the *spectral-shape residual* of the β_1 curve (L_1 metric; median 2.7%) and converting through the measured response slope in the same metric—a 16.7% mismatch in λ_4 produces 5.0% in L_1, i.e. about 0.30% of curve difference per 1% of mismatch—gives about 16% (about 9% with the median 2.7%), conservatively rounded up to 20%. Since that slope is only about 1σ significant at 300 fields, 20% is an order-of-magnitude estimate, not a precision guarantee. β_1 is therefore a coarse detector of λ_4, not a precise estimator.

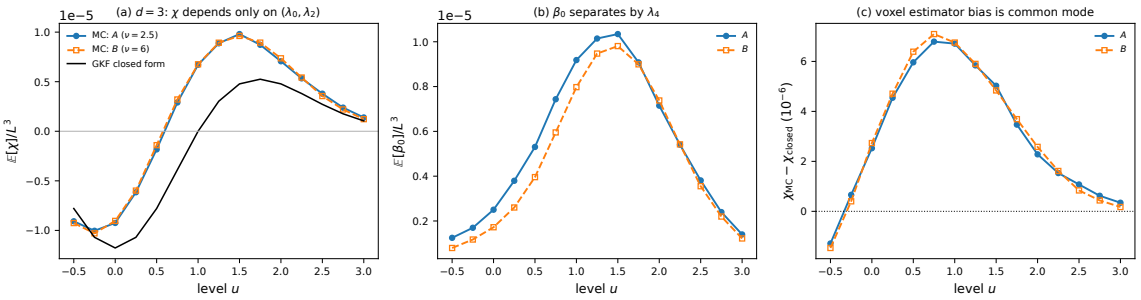

Fig. 6 Numerical check in $d = 3$ ($N = 128$, 120 fields). (a) The χ curves of two fields matched in (λ_0, λ_2) coincide and share the shape of the GKF closed form. (b) β_0 separates according to λ_4. (c) The deviations $\chi_{\mathrm{MC}} - \chi_{\mathrm{closed}}$ of the two fields nearly coincide, showing that the normalisation offset is a common mode of the voxel estimator rather than a property of the fields

Table 2 Configuration-by-configuration check of the decontamination family ($N = 512$, 300 fields per configuration), ordered by λ_4/λ_2^2. Every component has $\nu > 3$ and λ_6/λ_2^3 is exactly matched; since $\nu_1 = 3.2 < 4$, λ_8/λ_2^4 diverges in all five configurations. The last three columns are per-field counts at $u = 0$. The median relative differences use only the levels $u \le 0.75$, whereas the 32% of the A/B contrast uses all levels; the two are compared in order of magnitude only

Configuration	λ_4/λ_2^2	λ_6/λ_2^3	λ_8/λ_2^4	$\beta_0(0)$	$\beta_1(0)$	$\chi(0)$
C56 ($\nu = 12$-dominated, $w = 0.011$)	5.60	483.9	∞	11.87	3.81	8.06
Reference (pure $\nu = 3.2$)	7.33	484.0	∞	12.87	5.14	7.73
C90	9.00	484.0	∞	13.34	5.68	7.66
C110	11.00	484.1	∞	14.56	6.70	7.86
C125	12.50	483.9	∞	15.57	8.10	7.47

5.7 Jacobian diagnosis and Fisher equivalence

For mean-level observables the second singular value is $\sigma_2 \sim 10^{-13}$, the condition number $\kappa \gtrsim 10^{13}$, and the cosine of the angle between the degeneracy direction and $\nabla\lambda_2$ is 1.0000 (Fig. 8). Adding Fano factors lifts σ_2 and increases the information measure by $\sim 10^{21}$ in magnitude. Under Gaussian observation, Fisher information $F = J^{\mathsf{T}}\Sigma^{-1}J$, so $\det F = \det(\Sigma^{-1})(\sigma_1\sigma_2)^2$, hence $\det F = 0 \iff \sigma_2 = 0$: the SVD rank-loss criterion is fully equivalent to vanishing Fisher determinant, with SVD additionally supplying κ and the null-space direction. We propose $\kappa > 10^8$ as a practical threshold above which mean curves alone should not be used for multi-parameter inversion. The threshold lies between the measured $\kappa \sim 10^{11}$–10^{13} and the double-precision noise floor $\sigma_2/\sigma_1 \sim 10^{-12}$, leaving about two orders of magnitude of margin, so it is not specific to a particular configuration. It is a screening criterion, not a statistical test: for $\nu \gtrsim 3$ even genuine parameter values can fall below the threshold, so it must be used together with the regularity check.

5.8 Sufficiency: three moments determine the curve at a tolerance

By definition 3, three-dimensionality holds at tolerance $\varepsilon_0 = 3.8\%$ for β_0 and 8% for β_1. The tolerance is the 95th percentile of a *same-parameter, different-sample* control, fixed before the matching experiment is run: each 300-field configuration is split at random into two halves of 150 fields, the half-sample distance is computed with the same L_1 definition, and the procedure is repeated 400 times. This gives a null median of 2.1% and a 95th percentile of 3.8% for β_0 (χ: 2.6%, 4.4%; β_1, restricted to $u \le 0.75$: 3.9%, 8.0%). The tolerance is thus fixed by the null alone and does not depend on the matched configurations. The largest residuals of the thirteen configurations, 2.53%, 3.00% and 4.79%, all lie below the corresponding 95th percentile. In terms of margins, β_0 has 0.16 percentage points left (about 6% of the tolerance), χ

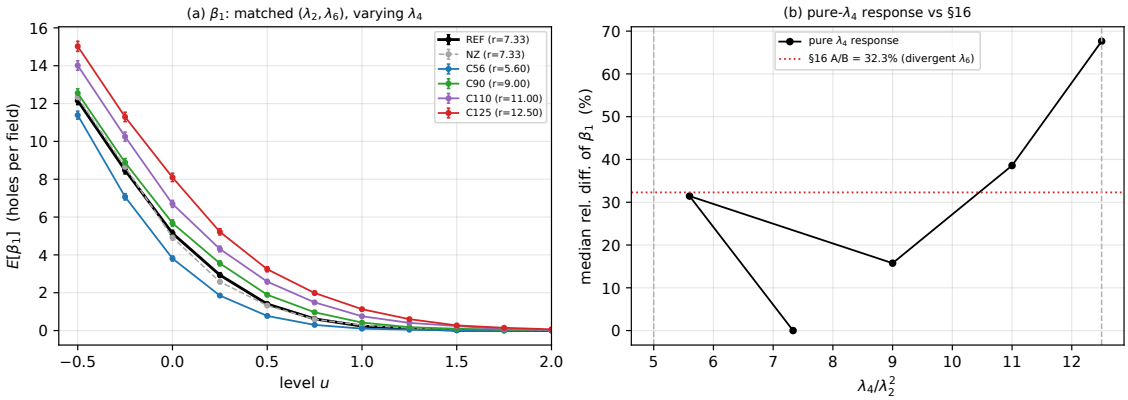

Fig. 7 β_1 curves of the decontamination family ($N = 512$, 300 fields per configuration). (a) With $(\lambda_0, \lambda_2, \lambda_6)$ exactly matched, $\mathbb{E}[\beta_1(u)]$ (holes per field) rises monotonically as λ_4/λ_2^2 increases from 5.60 to 12.50. (b) The median relative difference of each β_1 curve with respect to the reference field ($\lambda_4/\lambda_2^2 = 7.33$) grows with λ_4, reaching 38% and 67% at 11.0 and 12.5, above the 32.3% of the A/B contrast (red dashed line), and is 31% at 5.60, of the same order as A/B

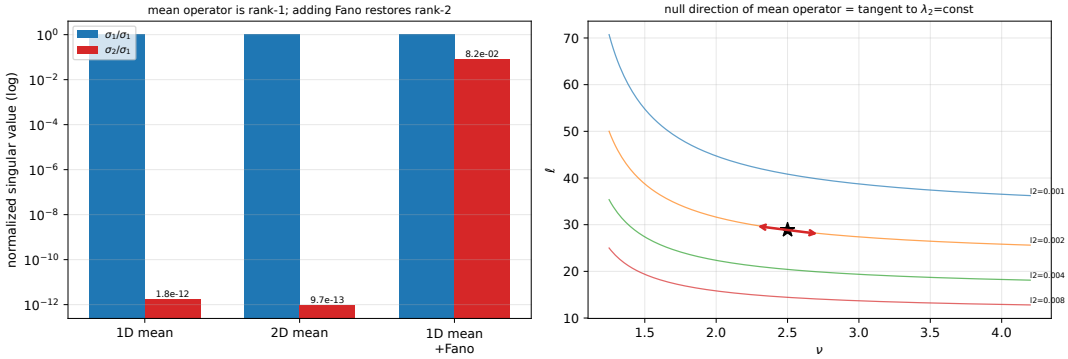

Fig. 8 Jacobian diagnosis: mean observables have $\sigma_2 \sim 10^{-13}$, $\kappa \gtrsim 10^{13}$, and a degeneracy direction perfectly aligned with $\nabla\lambda_2$ ($\cos = 1.0000$)

has 0.11 (about 3.5%) and is nearly at the line, and β_1 has 0.87 (about 15%).[1]

Two moments are insufficient: in section 5.3 the β_0 curves of fields matched in (λ_0, λ_2) separate according to λ_4 ($L_1 = 21.3\%$ across spectral families, about twelve times the sample-size-matched null median $2.1\% \times \sqrt{150/200} \approx 1.8\%$), so the dimension exceeds 2.

Three moments are sufficient on the scanned family: two families of two-scale mixtures—family A with $(\nu_1, \nu_2) = (3.2, 8)$ and family B with $(3.5, 10)$—are matched to the reference field R (a pure Matérn field with $\nu = 4$) exactly on $(\lambda_0, \lambda_2, \lambda_4)$ at thirteen mixing weights ($N = 512$, 300 fields each; Fig. 9), while the hidden moment λ_6/λ_2^3 is pushed to 1.31–3.19 times its reference value (λ_6 itself grows by 31%–219%). In the same $L_1 = \sum|a - b|/\sum|b|$ convention as the noise floor: β_0 gives 0.77%–2.53% (median 1.24%, percentiles 21%–81% of the null), χ gives 0.74%–3.00% (1.64%), and β_1 gives 1.72%–4.79% (2.68%). None of the thirteen exceeds the preregistered ε_0 (3.8%, 4.4%, 8.0%), and the β_0 residuals show no monotone trend in λ_6. Recomputing the null of the reference field itself by

[1] A half-sample (150 vs 150) difference has standard deviation $\sigma\sqrt{2/150}$, whereas "one 300-field configuration against the 300-field reference" has $\sigma\sqrt{2/300}$; ε_0 is therefore systematically larger by about $\sqrt{2}$, giving about β_0 2.7%, χ 3.1% and β_1 5.7% after correction, and the largest residuals remain below these corrected values. Both the residual and the floor scale as $N^{-1/2}$, so increasing the field count lowers them proportionally without materially changing the ratio; comparing the largest of thirteen residuals with the 95th percentile of a single comparison is stricter than a per-configuration test.

400 random equal splits gives (median/95th percentile) β_0 2.07%/3.50%, χ 2.78%/4.67% and β_1 4.22%/8.23%, consistent with the pre-registered floors. As a power control, a pure Matérn field with λ_4 lowered by 16.7% (around the $\nu = 3.2$ reference used earlier) gives β_0 $L_1 = 2.81\%$ (half-sample 3.14%, 88th percentile), and its separability shows up in the paired per-level test: 27% of levels have $|z| > 2$, against at most 7% for the thirteen matched configurations. Hence *no separable pair was found on the scanned family*, and β_0 is determined by the first three moments at $\varepsilon_0 = 3.8\%$.

This is an empirical statement rather than a proof for all θ, θ': definition 3 quantifies over all pairs, whereas we scanned thirteen configurations (two ν-pairs) around a single reference field, and the full cross-product grid (several references $\times$ several weights) remains to be run (section 7.3). L_1 alone does not resolve a 17% mismatch in λ_4 at 300 fields (the control sits at the 88th percentile); the resolving power comes from the paired per-level z test. Any tolerance between 3.8% and 21.3% leaves the conclusions unchanged.

The tolerance for β_1 is wider. After exactly matching the first three moments, the thirteen configurations give β_1 residuals on $u \leq 0.75$ (the same window as the null) of 1.7%–4.8% (median 2.7%), below the null median of 4.2% and still short of the 95th percentile of 8.2%; β_1 is therefore likewise determined by the first three moments at $\varepsilon = 8\%$. A tighter tolerance cannot be claimed: the 5.0% (L_1) produced in β_1 by a 16.7% mismatch in λ_4 (median relative difference 14.3%) is of the same order as the upper end of these residuals and is not resolvable at this sample size. This residual is measured *after* matching λ_4, whereas the 58.8% of section 5.6 is the response to *changing* λ_4: the two measure different objects, the latter being the sensitivity of β_1 to λ_4. Fields with $\nu = 2.5$ have divergent λ_6 (per section 2.6), so their β_1 differences cannot be attributed to λ_6.

Accordingly, eq. (5) is a control bound, and three-dimensionality holds in the tolerance sense: 3.8% (the 95th percentile of the null) for β_0 and 8% for β_1. Both presuppose that λ_4 is finite ($\nu > 2$); for $\nu \leq 2$ the bound and the tolerance test fail simultaneously.

5.9 Numerical check of upper bounds

This subsection checks only whether eq. (5) is an *effective* control; identification dimensions are dealt with in section 5.8. If Λ_k contains the first three moments only, the zero-free-parameter Kac–Rice prediction $\mathbb{E}[c_0(u)]$ from eq. (6) should differ from the Monte Carlo count by a threshold-independent factor. For configuration A ($\nu = 2.5$, $\lambda_2 = 0.002$, $\lambda_4 = 4.8 \times 10^{-5}$; $N = 512$, 120 fields, 15 thresholds) that ratio is 0.933–1.073 over 8 thresholds ($u = -0.5, \ldots, 3.0$), constant at 1.06 overall; when downsampling to $\ell/h = 14.4$ and 7.2 it becomes 0.98 and 0.87, crossing unity monotonically, as expected for a discretisation count, so the 6% offset is attributed to discretisation rather than to a failure of the three-moment control. The Morse inequality $\beta_0 \leq c_0$ holds separately at every threshold for every field. The bound thus has the same shape as the quantity it controls, up to a threshold-independent multiplicative factor, confirming that the three moments do suffice for the bound (Fig. 10).

6 Robustness

6.1 Lattice spacing h

In two dimensions the discretisation bias of χ is multiplicative and $O(h^2)$ (Fig. 11): fixing $L = 384$ with 400 fields and scanning $h \in \{1, 2, 3, 4, 6\}$ gives $\mathbb{E}[\hat{\chi}_h] = \mathbb{E}[\chi](1 - c\,h^2)$ with $c\ell^2 \approx 1$–1.7 ($R^2 = 0.46$–0.95 for $u \in [0.5, 2]$). The relative bias is below 1% at $h = 1$ and about 7% at $h = 6$; pushing it below 1% requires $\ell/h \gtrsim 13$, which explains why the A/B χ discrepancy falls from 2.3% to 0.66% when N rises from 192 to 512. In one dimension the bias decreases linearly in h (2.23% $\to$ 1.17% $\to$ 0.59%; section 5.1). Because the two-dimensional bias is multiplicative, both configurations are scaled equally at a given h, so ratio-type criteria and the χ/β decoupling are unaffected.

6.2 Threshold sampling density Δu

The influence splits into three layers. Single-threshold values of β_0, β_1, χ depend only on the binary image at that threshold and are therefore strictly independent of Δu: on nested grids with $\Delta u \in \{0.5, 0.25, 0.125, 0.0625\}$, all 24 pairwise

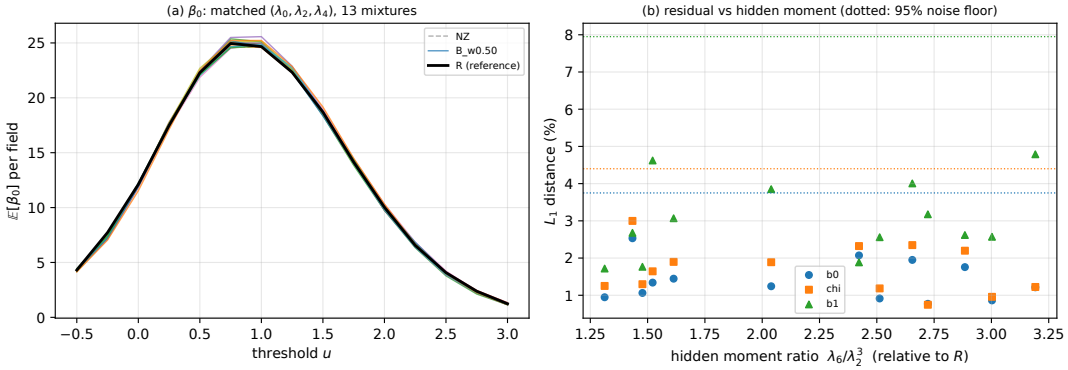

Fig. 9 Sufficiency experiment: thirteen two-scale mixtures (two ν-pairs, with the hidden moment λ_6/λ_2^3 at 1.31–3.19 times its reference value) exactly matching $(\lambda_0, \lambda_2, \lambda_4)$. (a) The full β_0 curves coincide with the reference field R; (b) L_1 residuals of β_0, χ and β_1 (same convention as the noise floor) against the hidden moment, with the 95% floors dotted (3.8%, 4.4%, 8.0%). None of the thirteen configurations crosses its floor

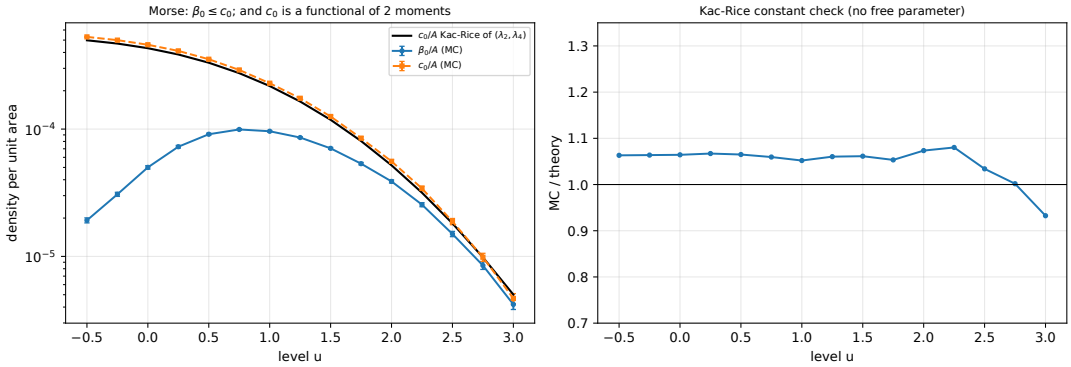

Fig. 10 Numerical check of the bound: the Kac–Rice three-moment bound matches the Monte Carlo count up to a threshold-independent factor, and the Morse inequality holds pathwise

comparisons over 8 configuration–grid combinations give a difference of exactly 0. Only cross-threshold operations carry error, and that error is $O(\Delta u^2)$: the relative bias of $\int \beta_0$ falls from 0.77%–1.32% at $\Delta u = 0.5$ to below 0.025% at 0.125; but $\beta_1 = \beta_0 - \chi$ is a signed cancellation, so the same absolute error is amplified and $\int \beta_1$ reaches 13% at $\Delta u = 0.5$. Fitted amplitudes are the most sensitive, dropping to 1.3%–5.3% only at $\Delta u = 0.125$. The diagnostic κ does not vary with the grid at all ($\kappa = 4.98$–6.36×10^{11}, $\cos = 1.000000$), because rank and null space are fixed by the structural outer product. Accordingly, threshold-level comparisons and integral statistics may use $\Delta u \le 0.25$, whereas fitting and inversion require $\Delta u \le 0.125$; the separation criteria do not flip at any of the four densities.

6.3 Observation window L

Window effects scale as $O(1/L)$ (Fig. 12). Cross-scanning window size and field count ($L \in \{128, 192, 256, 384, 512\}$, $N \in \{60, 120, 240, 480\}$, 20 combinations with nested subsamples) separates bias from power: the *median relative difference* changes only with L, varying by at most 2.5 percentage points across the four values of N at fixed L and showing no monotone trend (β_0 at $L = 512$: 5.6/3.9/3.6/3.8%); the *detection rate* rises with N, $\max|z|$ growing roughly as $\sqrt{N}$ but at an empirical rate about 0.8 times the ideal one, because the reference curve itself moves with N. Over all 20 combinations χ attains a detection rate of only 0–33% with $\max|z| \le 2.87$ and no monotone dependence on N, confirming that the coincidence of the χ curves is a structural null

result rather than a power deficit. The quantitative operating regime is therefore $L \geq 256$ with $N \geq 240$; at $L = 128$ the $O(1/L)$ bias is not negligible and only qualitative discrimination is possible, and only for $N \geq 480$. Ratio-type criteria remain window-invariant.

7 Discussion

7.1 Summary of identification dimensions

Table 3 collects the results by observable. The second column is an *identification dimension* ($\mathbb{E}[\chi]$ exact, the β rows in the tolerance sense): it states how many parameter degrees of freedom the observable can determine, not how accurately they can be estimated. The third column records whether that bound is *tight*, that is, whether removing one spectral moment would separate the curves. The two columns rest on different kinds of evidence: the dimension and tightness of $\mathbb{E}[\chi]$ are both analytic (by proposition 1, the tail decay rate recovers λ_0 and the amplitude recovers λ_2/λ_0, so the dimension cannot drop to one); the dimensions of $\mathbb{E}[\beta_0]$ and $\mathbb{E}[\beta_1]$ are numerical statements in the tolerance sense (section 5.8) and are *not* derived from the bound of proposition 3 (see remark 1).

What remains is a stratified precision: β_0 is three-dimensional and tight at the 3.8% noise floor; β_1 holds only to 8%, corresponding to a λ_4 parameter resolution of about 20%, and that estimate is limited by the significance of the response slope; and the infinite dimensionality of variance/Fano statistics is an argument rather than a theorem, usable as a directional route to break the degeneracy but not as a dimensional guarantee.

One boundary should be kept in mind: for $\nu \leq 2$, λ_4 diverges, the right-hand side of eq. (5) is no longer finite, and the usable mean-level information falls back to the two dimensions carried by χ—precisely the boundary drawn by the phase diagram in section 2.6.

7.2 Implications for inversion practice

The results above organize into an ordered procedure (successively in section 5.7, section 5.6 and section 5.2): diagnose before inverting—compute the condition number of the mean operator before fitting, since $\kappa > 10^8$ means the null space already dominates and a multi-parameter fit will merely slide along the level set of λ_2; choose the observable according to the target moment—use χ when only (λ_0, λ_2) are needed, and the $\beta_1 - \chi$ difference when λ_4 is needed, whose resolution is about 20%, below which a difference is indistinguishable from spectral-shape effects and should not be an inversion target; escalate the information source once the parameter count exceeds the three-dimensional bound, where variance or higher-order statistics are the only route and Fano factors are the minimal injection, at the price of a marked increase in sample size (about 5000 paths to pin ν to ± 0.1 in one dimension). The procedure has one known blind spot: for $\nu \gtrsim 3$ the β_1 curve saturates and κ gives no warning near the true value (section 7.3), so the κ criterion must be used together with a regularity check.

7.3 Limitations

The scope and known boundaries of our conclusions are as follows.

- **Setting.** Isotropic, Gaussian, torus-excursion fields; numerical validation is carried out in two and three dimensions, while the $d \geq 4$ results remain theoretical.
- **Low-regularity regime, $\nu \leq 2$ (two dimensions).** λ_4 diverges and the Morse/Kac–Rice framework breaks down entirely, so the finite-dimensional bounds no longer apply (section 2.6).
- **High-regularity regime, $\nu \gtrsim 3$ (two dimensions).** The β_1 curve saturates and inverting λ_4 from β_1 incurs a systematic bias of order $+0.5$; κ cannot flag this near the true parameter (section 5.6).
- **Tolerance for β_1.** The noise floor is 8% (against 3.8% for β_0) and the parameter resolution about 20%; the upper end of the residuals is of the same order as the control with a 16.7% mismatch in λ_4, so no tighter tolerance can be claimed (section 5.8).
- **Coverage of the matched family.** The thirteen configurations share one reference field R and are positively correlated, so "none exceeds the floor" is not the outcome of thirteen independent tests; the universal quantification of definition 3 is only partially tested, and a grid

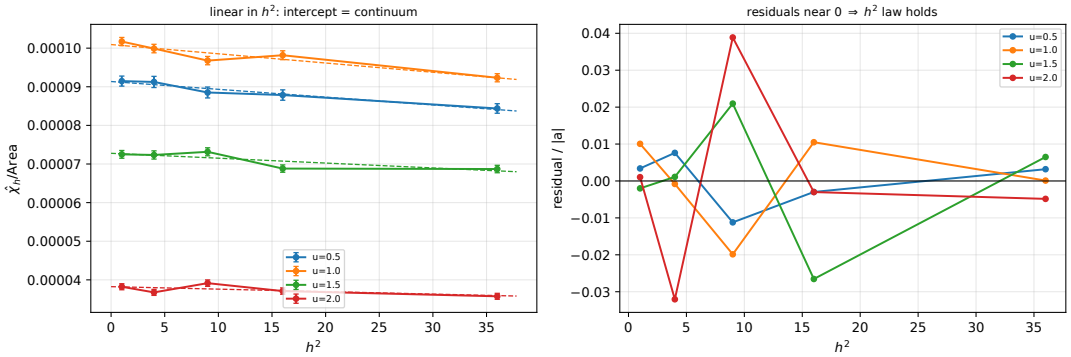

Fig. 11 Lattice robustness: the discretisation bias of χ is multiplicative $O(h^2)$ and identical for both configurations, so ratio-type criteria are unaffected

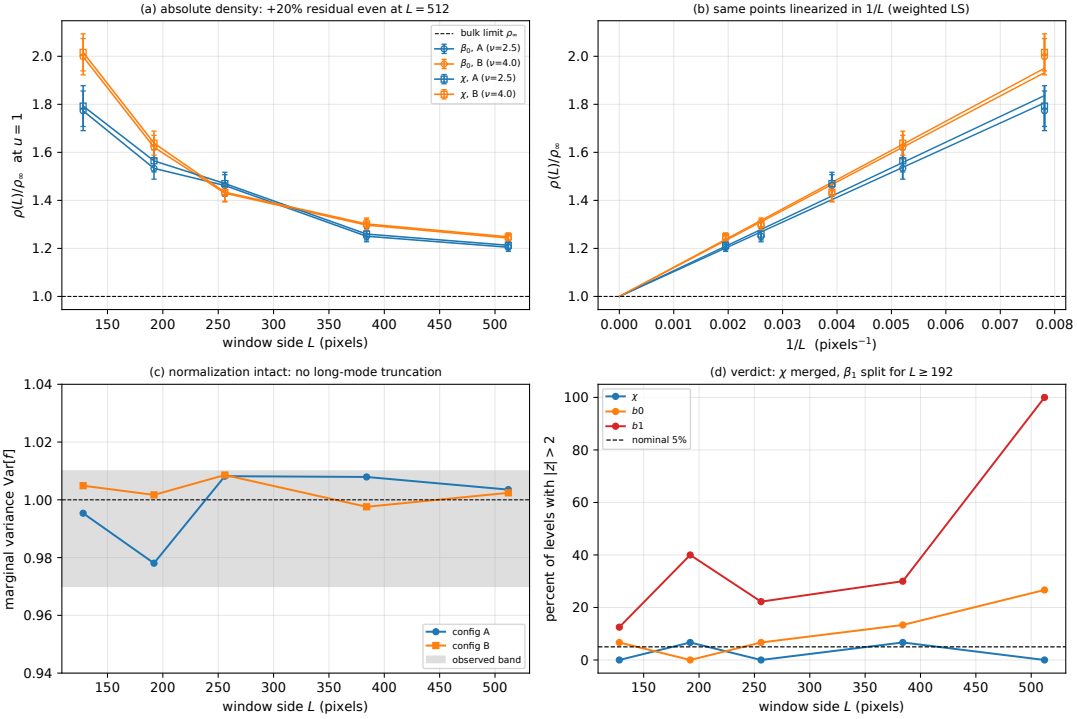

Fig. 12 Window robustness: window effects scale as $O(1/L)$; median relative difference of β_0 is constant across both L and N, so detection-rate increase is purely a power effect

Table 3 Identification dimensions of mean and variance observables

Observable	Dimension	Tightness / tolerance
$\mathbb{E}[\chi]$	2 (λ_0, λ_2)	Tight (analytical)
$\mathbb{E}[\beta_0]$	3 $(\varepsilon = 3.8\%)$	Tight (ε-sense, numerical)
$\mathbb{E}[\beta_1]$	3 $(\varepsilon = 8\%)$	Weakly tight (residuals 1.7–4.8%; resolution $\gtrsim 20\%$)
Variance / Fano	Infinite	—

"Tight" means the dimension cannot be reduced further and the right-hand entry is the supporting evidence. The dimensions of the two β rows are tolerance-based (definition 3), with the tolerance taken as the 95th percentile of the same-parameter null (section 5.8; both the β_1 residuals and the null are restricted to $u \leq 0.75$), and are fixed from both sides—two moments insufficient, three sufficient—rather than derived from eq. (5); "—" marks a row supported by an argument rather than a numerical test.

over several references and weights is left for future work (section 5.8).

- **Residual of the decontaminated control.** Pinning $(\lambda_0, \lambda_2, \lambda_6)$ while varying only λ_4 forces a two-scale mixture, so λ_8 and higher moments still co-vary with λ_4; moment-by-moment isolation was not carried out (section 5.6).
- **No estimation theory.** Consistency and asymptotic distributions of $\hat{\nu}$ and $\hat{\lambda}_4$ are not established, and neither $\kappa > 10^8$ nor the "5000 paths" figure is a confidence statement.
- **Window and discretization.** The 5×4 $L \times N$ scan uses a single seed stream per $(L, \text{parameter})$ pair, so fluctuations across N are not separated from seed-to-seed fluctuations (section 6).
- **Anisotropy and non-Gaussianity.** Under anisotropy the fourth-order term becomes a full tensor: the bound stays finite but is no longer a single scalar. Non-Gaussian fields require new theoretical tools and lie beyond the scope of this paper.

8 Conclusion

This paper reframes "topology-based covariance inversion" as an inverse problem with explicit information boundaries. Theoretically, χ has identification dimension exactly 2 (a GKF corollary) and β_k admits a three-moment control bound (Morse–Kac–Rice), but that bound is not a functional dependence; three-dimensionality is established numerically from both sides—two moments insufficient, three sufficient at the 3.8% noise floor for β_0 and 8% for β_1, with no separable pair found among the scanned matched configurations. The rank-one origin of the ill-posedness is a direct consequence of the chain rule, and its value lies in the three operational diagnostics: the κ threshold, the $\beta_1 - \chi$ difference, and Fano-factor variance injection. The results show that the hypothesis "curve shape uniquely determines the covariance parameters" fails in general: in one dimension distinct covariances produce identical mean Betti curves, in two dimensions identical χ curves, and topological curves recover at most a three-moment functional of the covariance; uniqueness holds only within an exactly three-parameter isotropic family (such as Matérn with $\nu > 2$), only through the Betti curves, and only to within a tolerance. These conclusions impose direct applicability constraints on existing practices that fit covariance parameters from Betti curves, and provide a theoretical basis for breaking the mean-level degeneracy through second-order statistics.

Supplementary information. Not applicable.

Acknowledgements. The author thanks the anonymous referees for helpful comments and acknowledges the support of Huzhou College.

Statements and Declarations

Competing interests

The author has no competing interests to declare that are relevant to the content of this article.

Funding

No funds, grants, or other support was received.

Data availability

The datasets generated during and analysed during the current study are available in the project repository at https://gitee.com/scipub/mctad.git.

Code availability

The 18 analysis scripts used in this paper are available in the same repository at https://gitee.com/scipub/mctad.git; further exploratory code, which produces none of the numbers reported here, is included as well.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

The sole author designed the study, wrote the code, analysed the results, and drafted the manuscript.

Appendix A Reproducibility checklist

The 18 scripts referenced in this paper live in `src/` and are numbered 01–18; three helper modules and some exploratory code that produces none of the reported numbers are included as well. Each entry gives the file name and the subsection where its output is used:

- `01_one_dim_counterexample.py` (§5.1)
- `02_variance_inversion.py` (§5.2)
- `03_two_dim_validation.py` (§5.3)
- `04_lambda4_inversion.py` (§5.6)
- `05_lattice_bias.py` (§6)
- `06_jacobian_diagnostic.py` (§5.7)
- `07_upper_bound_check.py` (§5.9)
- `08_threshold_density_robustness.py` (§6)
- `09_window_size_robustness.py` (§6)
- `10_field_count_scan.py` (§6)
- `11_sufficiency_lambda4.py` (§5.8)
- `12_powerlaw_crosscheck.py` (§5.3)
- `13_joint_window_field_scan.py` (§6)
- `14_three_dim_upper_bound.py` (§5.5)
- `15_noise_floor_matched_family.py` (§5.8)
- `16_b3_decontamination.py` (§5.6)
- `17_matched_family_grid.py` (§5.8)
- `18_park_parameter_points.py` (§5.4)

References

Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Monographs in Mathematics. Springer, New York (2007)

Beliaev, D.: Smooth Gaussian fields and percolation. Probability Surveys **20**, 897–937 (2023) https://doi.org/10.1214/23-PS56

Botnan, M.B., Hirsch, C.: On the consistency and asymptotic normality of multiparameter persistent Betti numbers. Journal of Applied and Computational Topology (2023) https://doi.org/10.1007/s41468-023-00127-4

Chingangbam, P.: On the statistical nature of Betti numbers and Euler characteristic of smooth random fields. arXiv preprint arXiv:2507.06255 (2025)

Calles, J.X.E.A., Yip, J.H.T., Contardo, G., Noreña, J., Rouhiainen, A., Shiu, G.: Cosmology with persistent homology: parameter inference via machine learning. Journal of Cosmology and Astroparticle Physics (2025) https://doi.org/10.1088/1475-7516/2025/06/025 . arXiv:2412.15405

Dark Energy Survey Collaboration: Dark energy survey year 3 results: $w$cdm cosmology from simulation-based inference with persistent homology on the sphere. Monthly Notices of the Royal Astronomical Society **545**(3), 2152 (2025) https://doi.org/10.1093/mnras/staf2152

Heydenreich, J., Brinckmann, S., *et al.*: Persistent Betti numbers for cosmological parameter constraints. Astronomy & Astrophysics (2021) https://doi.org/10.1051/0004-6361/202140736

Jalali Kanafi, M.H., Ansarifard, S., Movahed, S.M.S.: Imprint of massive neutrinos on persistent homology of large-scale structure. Monthly Notices of the Royal Astronomical Society **535**(1), 657–678 (2024) https://doi.org/10.1093/mnras/stae229

Krebs, J., Polonik, W.: On the asymptotic normality of persistent Betti numbers. Advances in Applied Probability (2025) https://doi.org/10.1017/apr.2024.65

Milnor, J.: Morse Theory. Annals of Mathematics Studies, vol. 51. Princeton University Press, Princeton, NJ (1963)

Minamitani, E., Nakamura, T., Obayashi, I., Mizuno, H.: Persistent homology elucidates hierarchical structures responsible for mechanical properties in covalent amorphous solids. Nature Communications (2025) https://doi.org/10.1038/s41467-025-58432-6

Minamitani, E., Obayashi, I., Shimizu, K., Watanabe, S.: Persistent homology-based descriptor for machine-learning potential of amorphous structures. The Journal of Chemical Physics **159**, 084101 (2023) https://doi.org/10.1063/5.0161399

Minamitani, E., Shiga, T., Kashiwagi, M., Obayashi, I.: Topological descriptor of thermal

conductivity in amorphous Si. The Journal of Chemical Physics **156**, 244502 (2022) https://doi.org/10.1063/5.0092875

Meng, K., Wang, J., Crawford, L., Eloyan, A.: Randomness of shapes and statistical inference on shapes via the smooth Euler characteristic transform. Journal of the American Statistical Association **120**(549) (2025) https://doi.org/10.1080/01621459.2024.2353947

Pranav, P., Adler, R.J., Buchert, T., Edelsbrunner, H., Jones, B.J.T., Schwartzman, A., Wagner, H., Weygaert, R.: Unexpected topology of the temperature fluctuations in the cosmic microwave background. Astronomy & Astrophysics **627**, 163 (2019) https://doi.org/10.1051/0004-6361/201834623

Park, C., Pranav, P., Chingangbam, P., Weygaert, R., Jones, B., *et al.*: Betti numbers of Gaussian fields. Journal of the Korean Astronomical Society **46**, 125–131 (2013) https://doi.org/10.5140/JKAS.2013.46.3.125

Pranav, P.: Topology and geometry of Gaussian random fields II: on critical points, excursion sets, and persistent homology. Monthly Notices of the Royal Astronomical Society (2021) https://doi.org/10.1093/mnras/stab2637 . arXiv:2109.08721

Pranav, P., Weygaert, R., Vegter, G., Jones, B.J.T., Adler, R.J., Feldbrugge, J., Park, C., Buchert, T., Kerber, M.: Topology and geometry of Gaussian random fields I: on Betti numbers, Euler characteristic and Minkowski functionals. Monthly Notices of the Royal Astronomical Society **485**(3), 4167–4208 (2019) https://doi.org/10.1093/mnras/stz541

Rice, S.O.: Mathematical analysis of random noise. Bell System Technical Journal **23**–**24** (1944) https://doi.org/10.1002/j.1538-7305.1944.tb03745.x

Roycraft, B., Krebs, J., Polonik, W.: Bootstrapping persistent Betti numbers and other stabilizing statistics. The Annals of Statistics **51**(4), 1484–1509 (2023) https://doi.org/10.1214/23-AOS2277

Rothenberg, T.J.: Identification in parametric models. Econometrica **39**(3), 577–591 (1971) https://doi.org/10.2307/1913267

Telschow, F.J.E., Cheng, D., Pranav, P., Schwartzman, A.: Estimation of expected Euler characteristic curves of nonstationary smooth random fields. The Annals of Statistics **51**(5), 2272–2297 (2023) https://doi.org/10.1214/23-AOS2337

Tymchyshyn, V., Tsizh, M., Vazza, F., Baldi, M.: On the evolution of Betti curves in the Cosmic web. Journal of Physical Studies **28**(4), 4903 (2024)

Wang, J., Elbers, W., Frenk, C.S., Cole, S., Yang, X., McCarthy, I.G., Weygaert, R.: Revealing the neutrino mass through persistent homology of the cosmic web. Monthly Notices of the Royal Astronomical Society **549**(2), 937 (2025) https://doi.org/10.1093/mnras/stag937