Riding the SuRF to continuous land: precise FWER control for Gaussian random fields

Fabian J.E. Telschow[1], Samuel Davenport[2]

[1]Department of Mathematics, Humboldt Universität zu Berlin
[2]Division of Biostatistics, University of California, San Diego

December 17, 2023

Abstract

The Gaussian Kinematic Formula (GKF) is a powerful and computationally fast tool to perform inference on random fields. As such it became a well-established method for *voxelwise inference* in the analysis of neuroimaging data. However, it has been found that under realistic scenarios its use leads to conservative control of the familywise error rate (FWER). In this article we identify and resolve the main causes of conservativeness and lack of power in the traditional usage of the GKF and demonstrate in simulations that a modified framework allows for precise FWER control. In a nutshell we use the smoothing step from preprocessing to bridge the gap between continuous probabilistic theory and discrete data. As a byproduct we propose novel, simple Lipschitz-Killing curvature estimators and shed light upon the actual type of FWER control which standard voxelwise inference methodologies in neuroimaging provide when smoothing has been applied.

1 Introduction

Motivation: Random Field Theory in Neuroimaging In experiments in neuroscience using functional Magnetic Resonance Imaging (fMRI) of the brain, data typically consists of 3D-images representing the time dynamics of the blood-oxygen-level dependence (BOLD). After an extensive pre-processing pipeline, including corrections for head motion and respiratory effects the time-series for each subject is combined into subject level maps (3D-images, typically consisting of coefficients of a linear model) via a first-level analysis step, see e.g. Poldrack and Nichols [2011] and the references therein. These high resolution images containing hundreds of thousands of voxels and allow for the detection of the locations of differences in %BOLD in the brain across different tasks and subjects. A major practical challenge of such data is the low signal-to-noise ratio which is often increased by convolving each subject level map with a smoothing kernel - typically an isotropic 3D-Gaussian kernel with Full Width at Half Maximum (FWHM) ≈ 3 voxels [Woo et al., 2014]. Statistical analysis is then performed on the subject level maps in order to identify areas of activation. This is typically done using a mass-univariate approach based on a test statistic image $\tilde{T}$ obtained from the smoothed subject level maps in which the hypotheses tests based on $\tilde{T}$ at each voxel are combined using a multiple testing procedure.

A practical solution to the multiple comparison of the thousands of voxels in the brain for such tests in terms of controlling the family-wise error rate (FWER) was pioneered in Worsley

et al. [1992], Friston et al. [1994], Worsley et al. [1996] and is known as *Random Field Theory* (RFT) in the neuroimaging community. The key innovation was to use the Gaussian Kinematic Formula (GKF) Adler [1981], Adler and Taylor [2007], Taylor et al. [2006] to approximate the probability that the maximum of a Gaussian random field over a compact manifold $\mathcal{M} \subset \mathbb{R}^D$, $D \in \mathbb{N}$, with C^3 sample paths exceeds a given threshold. The accuracy of this approximation has been theoretically validated in [Taylor et al., 2005]. We will refer to the setting from these papers, where $\mathcal{M}$ denotes the brain or the surface of the brain, as traditional RFT. Under the assumption of stationarity of the data, traditional RFT became a standard tool in the analysis of fMRI data, Eklund et al. [2016], because of its low computational costs. In the late 1990s it even became the core inference methodology in standard software packages such as *Statistical Parametric Mapping*, [Friston et al.] and is still the backbone of peak and cluster inference Chumbley and Friston [2009], Friston et al. [1994]. However, in the early 2000s it was shown in Monte-Carlo-simulations that RFT for voxel-wise inference is conservative in typical experimental situations [Nichols and Hayasaka, 2003]. Therefore it has been superseded by time costly, yet accurate permutation approaches, among others, [Nichols and Holmes, 2002, Winkler et al., 2016a,b]. More recently, in the seminal article Eklund et al. [2016], it was demonstrated using resting state data combined with fake task designs that voxelwise RFT inference is conservative (see their Figure 1). Moreover they showed that cluster inference based on RFT often fails to control false positive rates. They argued that a lack of stationarity and sufficient smoothness of the data are among the causes for these findings. Although these observations play a role in the conservativeness of voxel-wise inference using RFT we identify the true cause of conservativeness and present an elegant solution for possibly non-stationary Gaussian data. Since analysis of real fMRI data comes with its own challenges, we point to our paper Davenport et al. [2022a], which validates our approach using a large resting-state dataset from the UK Biobank [Alfaro-Almagro et al., 2018], and additionally offers a methodology to cope with the non-Gaussianity of brain imaging data for finite sample sizes.

The cause of conservativeness The lack of sufficient smoothness – as was argued for example in [Nichols and Hayasaka, 2003]– is not the true underlying cause of conservativeness in voxel-wise inference using traditional RFT. It rather is a relic of a mismatch between data and probabilistic theory. In traditional RFT it is assumed that an underlying continuous random field is approximated well by the random field observed on the voxel lattice. We call this the *good lattice assumption* which has been expressed in the following statement from [Worsley et al., 1996]: *"[...] the search region was regarded as a region with a smooth boundary defined at every point in 3-D. In practice only voxel data are available, and this will be regarded as a continuous image sampled on a lattice of equally spaced points. Thus a voxel is treated as a point in 3-D with zero volume, although it is often displayed on computer screens and in publications as a volume centered at that point."* As we shall argue, this mixes two concepts: the *data* and the *atoms of a probabilistic model*.

(Raw) data is the information about nature extracted from measurement devices which without exception is currently discrete. In contrast, the atoms (or data objects[1]) of probabilistic models used to describe data and perform statistical inference can be functions, curves, images, shapes, trees or other complex mathematical objects [Wang and Marron, 2007, Marron and Alonso, 2014]. In our setting the data consists of the subject level maps obtained from the first level model, i.e., discrete images. We will denote the discrete set of voxels of the image "belonging" to the brain as $\mathcal{V}$. Thus, the test statistic $\tilde{T}$ used for inference in the traditional

---

[1]We favor the term atoms over data objects because this nomenclature better separates the two concepts.

RFT setup is a random field over $\mathcal{V}$. This is incompatible with the fact that the required atoms of RFT are random fields over a Whitney stratified manifold $\mathcal{M}$ of dimension $D > 0$ with twice differentiable sample paths. Historically, this mismatch has been argued away using the good lattice assumption, though in practice this is known to be reasonable if a high level amount of smoothing is applied, [Kiebel et al., 1999, Nichols and Hayasaka, 2003, Taylor et al., 2007].

In this article we resolve this inconsistency by acknowledging that the data is discrete. Worsley [2005] and Taylor et al. [2007] tried to move away from the assumption of continuous data providing a discrete parametric method to control the FWER, however this requires the strong and unrealistic assumption that the noise arises as Gaussian random fields with a squared exponential covariance function. Our approach is quite different. It is based on the key observation that the smoothing operation[2] which is used to increase the signal-to-noise ratio in the pre-processing of the data naturally transforms the subject level maps, i.e., the discrete data, into functions over a Whitney stratified manifold $\mathcal{M} \supset \mathcal{V}$. Thus, the smoothed subject level maps are samples of atoms from RFT and the test statistic $\tilde{T}$ can be interpreted as a function over $\mathcal{M} \supset \mathcal{V}$, too.

This allows us to resolve the two main problems caused by the failure of the good lattice assumption in a typical fMRI data analysis which are the origin of the observed conservativeness of the current usage of RFT. These are as follows.

(i) The maximum of any test-statistic $\tilde{T}$ defined over $\mathcal{M}$ and derived from the atoms of RFT is always greater than or equal to the maximum of $\tilde{T}$ evaluated on a subset $\mathcal{V} \subset \mathcal{M}$. This leads to conservative control of the FWER using RFT as

$$\mathrm{P}\Big(\max_{v \in \mathcal{V}} \tilde{T}(v) > u_\alpha \Big) \leq \mathrm{P}\Big(\max_{x \in \mathcal{M}} \tilde{T}(x) > u_\alpha \Big) \approx \alpha. \tag{1}$$

Here $u_\alpha \in \mathbb{R}$ is the α level threshold obtained from the GKF.

(ii) Continuous quantities are replaced by voxel-based, discrete counterparts.

Riding the SuRF to continuous land To provide solutions to problems (i) and (ii) we use the smoothing step to obtain a test-statistic defined over a Whitney stratified manifold and to compute the critical threshold for voxelwise inference using RFT. In what follows we express this in a more general setting than required for data analysis in neuroimaging because similar issues, in terms of the mismatch between data and atoms, arise in other areas of statistics and are seldom treated in a similar fashion to what we present here, see Chen et al. [2017] for an exception. They occur, for example, when evaluating the coverage rate of simultaneous confidence bands in functional data analysis or the coverage of confidence sets for excursions above a given level [Sommerfeld et al., 2018, Bowring et al., 2019, 2021].

Assume that the data X can be described as a realization of a random field over a finite, discrete set $\mathcal{V} \subset \mathbb{R}^D$. The data can be transformed into a function over a compact Whitney stratified manifold $\mathcal{M}$ of dimension $D > 1$ by applying a linear transformation encoded by a kernel function $K : \mathcal{V} \times \mathcal{M} \to \mathbb{R}$, i.e., the transformed data is given by

$$\tilde{X}(s) = \sum_{v \in \mathcal{V}} X(v) K(v, s) \tag{2}$$

which can be evaluated at all $s \in \mathcal{M}$. We call $\tilde{X}$ a *SUper Resolution Field* (SuRF) because $\tilde{X}$ can be evaluated at arbitrarily high resolution. Since SuRFs are a class of random fields,

---

[2]Typically a convolution with a Gaussian kernel.

for which many path properties (e.g., differentiability) are inherited directly from the kernel function K, we are able to show in Theorem 1 that the GKF holds for SuRFs under mild conditions on K and the data process X.

In order to perform inference using the GKF we must specify the domain $\mathcal{M}$. This has not been attempted rigorously in the neuroimaging literature so far because of the reliance on the good lattice assumption. Since the GKF cannot be applied to random fields over discrete sets this leads to a further mismatch between theory and practice in traditional RFT. In the SuRF setting there is no such mismatch yet we must specify the domain $\mathcal{M}$. For the purpose of our applications, i.e., neuroimaging, it is sensible to ensure that $\mathcal{M} \supset \mathcal{V}$. In particular we propose to use the Whitney stratified manifold composed of the union of D-dimensional hyperrectangles which are centered at the elements of $\mathcal{V}$, are aligned with the standard coordinate axis and have as edge lengths $\min \left\{|v_d - w_d| \mid v, w \in \mathcal{V} : v_d \neq w_d\right\}$ in the dth coordinate direction (Figure 1). We call this set the voxel manifold $\mathcal{M}_{\mathcal{V}}$ associated with $\mathcal{V}$ and define it formally in Section 3.3. This is in many ways the most natural way of defining a Whitney stratified
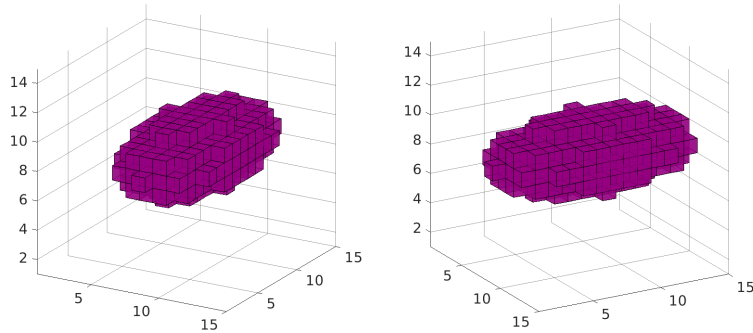

Figure 1: An example of a voxel manifold viewed from different angles.

manifold from the lattice $\mathcal{V}$ and was used in Worsley et al. [1996] to approximate the volumes of their underlying unknown search region. For us the main benefit is that the boundary of $\mathcal{M}_{\mathcal{V}}$ consists of the union of lower dimensional hyperrectangles parallel to the coordinate axes. This choice of domain simplifies the computation the Riemannian metric, induced on $\mathcal{M}_{\mathcal{V}}$ by $\tilde{X}$, and other derived geometric quantities required in the GKF, compare Section 3.4 and Proposition 7. Furthermore, it allows us to propose simple consistent estimators of the Lipschitz-Killing-Curvatures (LKCs) of SuRFs over the domain which are the main quantities required in order to apply the GKF.

The fact that $\mathcal{M}_{\mathcal{V}}$ is a Whitney stratified manifold of dimension greater than one means that the transformation $X \mapsto \tilde{X}$ transforms the discrete random field X into an atom of RFT provided that the functions $s \mapsto K(v, s)$ for all $v \in \mathcal{V}$ are sufficiently regular. Because the domain $\mathcal{M}_{\mathcal{V}}$ of the test statistic $\tilde{T}$ (which depends on observed SuRFs) is precisely defined, local maxima and excursions of $\tilde{T}$ above a given threshold can be obtained using numerical optimizers. Thus problem (i), the dominant factor in the conservativeness of traditional RFT, is resolved within our SuRF framework.

SuRFs also allow us to address problem (ii). In current software packages the LKCs are estimated from the random fields evaluated on $\mathcal{V}$ which are assumed to satisfy the good lattice assumption and often are required to be stationary (Forman et al. [1995], [Kiebel et al., 1999], Worsley et al. [1996]). These approaches use discrete approximations of the derivatives of $\tilde{X}$

4

instead of the true derivatives of $\tilde{X}$. Less widely used estimators drop the restrictive assumption of stationarity but still rely on discrete approximation of derivatives and are substantially more difficult to implement, these include the warping estimator [Taylor and Worsley, 2007], and the Hermite projection estimator (HPE) [Telschow et al.].[3] Here we will propose a SuRF based LKC estimator, which uses the explicitly available formulas for the LKCs up to $D = 3$ and does not rely on either stationarity or on discrete approximations of derivatives. For estimators such as the HPE, evaluating the SuRFs at a high resolution can reduce the effect of (ii). This comes at the price of large computation times as we shall show in our simulations, compare Table 1. In contrast, our newly proposed SuRF estimator has a low computational cost because it requires only a minimal resolution increase to obtain accurate estimates of the LKCs of SuRFs due to the fact that we can rely on the true derivatives of the SuRFs.

As a further contribution we shed some light on effect localization in the data from methods controlling FWER in the strong sense for SuRFs. Any statistical inference methodology relying on a test statistic derived from fields of the form (2) can only localize effects up to the support of the kernel functions K. Although important, the exact nature of the FWER control provided has, to the best of our knowledge, not been stated clearly in the literature. Even widely cited articles such as Nichols and Hayasaka [2003] refer only to methods having FWER control in the strong sense, yet do not clearly explain that the strong control is only with respect to the smoothed signal (the atoms) and not on the level of data.

Structure of the article: In Section 2 we define notation which is used throughout the article. Section 3 develops the theory of voxelwise inference using SuRFs. In particular, we provide a formal definition of a SuRF and give different sufficient conditions for it to satisfy the GKF in Section 3.1 and 3.2 respectively. Voxel manifolds are defined in Section 3.3 and theoretical forms for the LKCs of SuRFs defined on voxel manifolds as well as estimators thereof are provided in Sections 3.4 and 3.5 respectively. In the final section of the theory, Section 3.6, we explain in detail how voxelwise inference based on the GKF for SuRFs is carried out and provide a precise explanation of the type of FWER control that this implies. In Section 4 we report the results of our simulations of the performance of the SuRF LKC estimator and the FWER control based on SuRFs. More concretely, in Section 4.2 the results of the LKC estimates using the SuRF LKC estimator are reported under different models and are compared to the performances of other published LKC estimators. Simulations which demonstrate that the SuRF methodology avoids the conservativeness of traditional voxelwise RFT inference can be found in Section 4.3. In Section 5 we discuss our findings and review other potential applications of our SuRF methodology.

All proofs of the results in the main manuscript are available in Appendix D in the supplementary material. The supplementary also contains additional simulations and results. In particular Appendix A contains simulations further illustrating the performance of the SuRF LKC estimator for voxel manifolds of dimension $D \leq 3$. Furthermore, in Appendices B and C, we provide a general study of the LKCs of normalized random fields, which are helpful in proving our main results.

Code implementing the SuRF methodology is available in Matlab as part of the RFTtoolbox [Davenport and Telschow, 2023]. Scripts reproducing our simulation results are available at https://github.com/ftelschow/ConvolutionFieldsTheory.

---

[3]In theory the HPE does not depend on discrete derivatives. However the HPE does requires the critical values of residual fields and the current implementation of it uses discrete derivatives to do so.

2 Notation and Definitions

In this section we define notation which is used throughout the article. We assume that $v \in \mathbb{R}^D$ is a column vector, i.e., we identify $\mathbb{R}^D$ with $\mathbb{R}^{D \times 1}$. With $\mathcal{M}$ we denote a D-dimensional, compact $\mathcal{C}^2$-Whitney stratified manifold isometrically embedded into a D-dimensional manifold $\overline{\mathcal{M}}$ without boundary. Recall that a Whitney stratified manifold of dimension D is a space $\mathcal{M} = \bigcup_{d=1}^{D} \partial_d \mathcal{M}$ decomposed into strata where the stratum $\partial_d \mathcal{M} \subset \overline{\mathcal{M}}$ is a manifold of dimension d and all the strata are disjoint, i.e., $\partial_d \mathcal{M} \cap \partial_{d'} \mathcal{M} = \emptyset$ for all $d = d'$, compare [Adler and Taylor, 2007, Chapter 8] for more details. A $\mathcal{C}^2$-chart (which gives local $\mathcal{C}^2$-coordinates) around $s \in \overline{\mathcal{M}}$ is given by a tuple $\left(\overline{U}, \overline{\phi} \right)$ where $\overline{U} \subset \overline{\mathcal{M}}$ open, $s \in \overline{U}$ and $\overline{\phi} \in \mathcal{C}^2(\overline{U}, \overline{V})$ is a diffeomorphism onto an open set $\overline{V} \subset \mathbb{R}^D$. By the compactness of $\mathcal{M}$ there exists a set of finitely many charts $(\overline{U}_\alpha, \overline{\phi}_\alpha)_{\alpha \in \{1, \dots, P\}}$ of $\mathcal{M}$, $P \in \mathbb{N}$, such that $\mathcal{M} \subset \bigcup_{\alpha=1}^{P} \overline{U}_\alpha$. This union is the only relevant part of $\overline{\mathcal{M}}$ since we are only interested in properties of $\mathcal{M}$. The surrounding manifold $\overline{\mathcal{M}}$ is only introduced for technical requirements in the formulation of the GKF.

We denote with $f_\alpha = f \circ \phi_\alpha^{-1}$ the coordinate representation of f in the chart U_α and with ∇f_α the gradient of f_α, i.e., $\nabla f_\alpha(x) = \left(\frac{\partial f_\alpha}{\partial x_1}(x), \dots, \frac{\partial f_\alpha}{\partial x_D}(x) \right) \in \mathbb{R}^{1 \times D}$ for $x \in V_\alpha$, and with $\nabla^2 f_\alpha$ the Hessian of f_α, i.e., $\nabla^2 f_\alpha(x) \in \mathbb{R}^{D \times D}$ with d-d'th entry $\frac{\partial^2 f_\alpha}{\partial x_d \partial x_{d'}}(x)$ for $x \in V_\alpha$. If the gradient ∇ or the Hessian ∇^2 is applied to a function with two arguments, then it is always assumed to be with respect to the first argument. We will use s, s' for points in $\mathcal{M}$ or $\overline{\mathcal{M}}$ and x, y for points in local coordinates. For simplicity in notation, given $h \in C^1(\mathbb{R}^D)$ and a multi-index $\beta \in \mathbb{N}^d$, $d \leq D$, we introduce the notation

$$\partial_\beta h(x) = \frac{\partial^d h(x)}{\partial x_{\beta_1}, \dots, \partial x_{\beta_d}}(x), \quad x \in \mathbb{R}^D. \tag{3}$$

If $h \in C^1(\mathbb{R}^D \times \mathbb{R}^D)$, we sometimes write $\left(\partial_\beta^x h \right)(x, y)$ for $(\partial_\beta h(\cdot, y))(x)$ evaluated at $(x, y) \in \mathbb{R}^D \times \mathbb{R}^D$. Note that the ∂ also appears in our notation for the strata of a Whitney stratified manifold. This should cause no confusion since there ∂ is always applied to a topological space (a curly or capital letter) and not to a function (non-capital or small greek letter).

Moreover, given $D \in \mathbb{N}$ and a symmetric matrix $A \in \mathbb{R}^{D \times D}$ we define its *half-vectorization* as

$$\mathbb{V}(A) = (A_{11}, \dots, A_{D1}, A_{22}, \dots, A_{D2}, \dots, A_{D-1D-1}, A_{DD-1}, A_{DD})$$

and the set

$$a \cdot \mathbb{Z}^D = \left\{ x \in \mathbb{R}^D \mid x = (a_1 z_1, \dots, a_D z_N), z \in \mathbb{Z}^D \right\}$$

for $D \in \mathbb{N}$ and any vector $a \in \mathbb{R}^D$ which has positive entries. We also assume throughout the article that $\mathcal{V}$ is a discrete set.

3 Theory

In this section we define Super-Resolution Fields (SuRFs), introduce some of their basic properties and show that they satisfy the Gaussian kinematic formula under mild conditions. Moreover, we derive estimators of the LKCs of SuRFs defined over voxel manifolds.

3.1 Super-resolution fields

Definition 1. We call a map $K : \overline{\mathcal{M}} \times \mathcal{V} \to \mathbb{R}$ a *kernel*. We say it is continuous/differentiable, if $s \mapsto K(s, v)$ is continuous/differentiable for all $v \in \mathcal{V}$.

Remark 1. This definition of a kernel is more general than its standard definition in statistics, where it usually is defined as a function $k : \overline{\mathcal{M}} \to \mathbb{R}$ satisfying a normalization property. The function k can be seen as a kernel from $\overline{\mathcal{M}} \times \overline{\mathcal{M}} \to \mathbb{R}$ in the sense of Definition 1 by defining $K(s, s') = k(s - s')$.

Definition 2. Let X be a random field on $\mathcal{V}$ with values in $\mathbb{R}$ and covariance function $\mathfrak{c}(u, v) = \text{Cov}\,[X(u), X(v)]$ for $u, v \in \mathcal{V}$. For a kernel $K : \overline{\mathcal{M}} \times \mathcal{V} \to \mathbb{R}$ the associated random field over $\overline{\mathcal{M}}$,

$$\tilde{X}(s) = \sum_{v \in \mathcal{V}} K(s, v) X(v)\,, \quad s \in \overline{\mathcal{M}}\,, \tag{4}$$

is called a *Super-Resolution Field* (SuRF) associated to K and $\mathcal{V}$ which we shall denote by $(\tilde{X}, X, K, \mathcal{V})$ or $\tilde{X}$ for short if X, K and $\mathcal{V}$ are clear from the context.

For a SuRF $(\tilde{X}, X, K, \mathcal{V})$ we define the *normalized SuRF for a kernel K* by the SuRF obtained from X and $\mathcal{V}$ and the normalized kernel

$$\tilde{K} : \ \overline{\mathcal{M}} \times \mathcal{V} \to \mathbb{R}\,, \quad (s, v) \mapsto \frac{K(s, v)}{\sqrt{\sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} K(s, u) K(s, v) \mathfrak{c}(u, v)}}\,.$$

Remark 2. Let $\tilde{X}$ be a SuRF associated to a kernel K such that $\mathbb{E}[X(v)] < \infty$ for all $v \in \mathcal{V}$. Then by linearity

$$\mathbb{E}\big[\tilde{X}(s)\big] = \sum_{v \in \mathcal{V}} K(s, v) \mathbb{E}[X(v)]\,, \quad s \in \overline{\mathcal{M}}\,, \tag{5}$$

and if additionally $\mathbb{E}[X(v)^2] < \infty$ for all $v \in \mathcal{V}$ then

$$\text{Cov}\,\Big[\tilde{X}(s), \tilde{X}(s')\Big] = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} K(s, u) K(s', v) \mathfrak{c}(u, v) < \infty\,, \quad s, s' \in \overline{\mathcal{M}}\,. \tag{6}$$

The last formula explains the definition of the normalized SuRF, since it shows that any normalized SuRF satisfies $\text{Var}\big[\tilde{X}(s)\big] = 1$ for all $s \in \overline{\mathcal{M}}$.

Remark 3. Equation (6) is similar to an inner product, where $(\mathfrak{c}(u, v))_{u,v \in \mathcal{V}}$ is the representing "matrix". As such, we introduce, for all $s, s' \in \overline{\mathcal{M}}$, the simplifying abbreviations

$$\begin{aligned}
\langle K_s, \, K_{s'} \rangle &= \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} K(s, u) K(s', v) \mathfrak{c}(u, v)\,, \\
\|K_s\|^2 &= \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} K(s, u) K(s, v) \mathfrak{c}(u, v).
\end{aligned} \tag{7}$$

Remark 4. If $\mathcal{V} \subseteq \overline{\mathcal{M}}$ and $k : \overline{\mathcal{M}} \to \mathbb{R}$, then we call the SuRF obtained from the kernel given in Remark 1 a *convolution field*. Convolution fields appear naturally in many applications since convolving the observed data with a smoothing kernel is often a preprocessing step in signal processing or neuroimaging to improve the signal to noise ratio [Turin, 1960, Worsley et al., 2002].

SuRFs are a relatively nice class of random fields since all path properties are inherited directly from the kernel K. This is theoretically convenient since it is often possible to prove that a SuRF satisfies the assumptions of results such as the GKF by imposing conditions such as differentiability on the known kernel K. The next proposition is self-evident, but we state it here for convenience.

Proposition 1. *Let $(\tilde{X}, X, K, \mathcal{V})$ be a SuRF.*

1. If $K(\cdot, v)$ is continuous for all $v \in \mathcal{V}$ then $\tilde{X}$ has surely continuous sample paths.

2. If $K(\cdot, v)$ is k-times differentiable for all $v \in \mathcal{V}$, then $\tilde{X}$ has surely k-times differentiable sample paths.

3.2   Gaussian Kinematic Formula for SuRFs

In this section we pose assumptions on a kernel K and the discrete field X such that the associated normalized SuRF $\tilde{X}$ satisfies the assumptions of the Gaussian Kinematic formula [Adler and Taylor, 2007, Theorem 12.4.1, 12.4.2].

To begin with we restate the assumptions on a random field f defined over $\overline{\mathcal{M}}$ such that the GKF over $\mathcal{M}$ holds. Recall that f_α is the representation of f in the chart $(\overline{U}_\alpha, \overline{\phi}_\alpha)$ as introduced in Section 2. In this notation the GKF holds if for any chart $(\overline{U}_\alpha, \overline{\phi}_\alpha)$ in the atlas of $\overline{\mathcal{M}}$ we have that

(G1) f is a zero-mean, unit-variance Gaussian random field on $\overline{\mathcal{M}}$ with a.s. C^2-sample paths.

(G2) For all $\alpha \in \{1, \ldots, P\}$ the covariance matrix of the Gaussian vector $\left(\nabla f_\alpha(x), \mathbb{V}\left(\nabla^2 f_\alpha(x) \right) \right)$ is non-degenerate for all $x \in \overline{\phi}_\alpha(\overline{U}_\alpha) \cap \mathcal{M}$.

(G3) For all $\alpha \in \{1, \ldots, P\}$, there exist constants $\kappa, \gamma, \epsilon > 0$ such that for each $d, d' \in \{1, \ldots, D\}$,

$$\mathbb{E}\left[\left(\partial_{dd'} f_\alpha(x) - \partial_{dd'} f_\alpha(y) \right)^2 \right] \leq \kappa \left| \log \|x - y\| \right|^{-1-\gamma},$$

for all $x, y \in \overline{\phi}_\alpha(\overline{U}_\alpha) \cap \mathcal{M}$ for which $|x - y| < \epsilon$.

Remark 5. Conditions **(G1)**-**(G3)** imply that the sample paths of f are almost surely Morse functions, compare [Adler and Taylor, 2007, Corollary 11.3.2.]. Moreover, by Lemma 1 from Davenport and Telschow [2022] these conditions do not depend on the particular choice of the $\mathcal{C}^3$ charts (U_α, ϕ_α), $\alpha \in \{1, \ldots, P\}$, but rather hold for all $\mathcal{C}^3$ charts (V, φ) of $\overline{\mathcal{M}}$.

Applying Proposition 1, **(G1)** holds for any normalized SuRF derived from a zero-mean Gaussian random field X on $\mathcal{V}$ and a twice continuously differentiable kernel K. The next two propositions establish that Condition **(G3)** is satisfied for a SuRF with a $\mathcal{C}^3$-kernel.

Proposition 2. *Let $\gamma \in (0, 1]$ and for all $v \in \mathcal{V}$ let $K(\cdot, v)$ be γ-Hölder continuous with Hölder constants bounded above by $A > 0$ and $\mathbb{E}\left[X(v)^p\right] < \infty$, $p \in [1, \infty)$. Then $\tilde{X}$ has almost surely $\mathcal{L}^p$-Hölder continuous paths, i.e.,*

$$|\tilde{X}_\alpha(x) - \tilde{X}_\alpha(y)| \leq L\|x - y\|^\gamma \tag{8}$$

for the charts $(\overline{U}_\alpha, \overline{\phi}_\alpha)$, $\alpha \in \{1, \ldots, P\}$, in the atlas of $\overline{\mathcal{M}}$ covering $\mathcal{M}$, all $x, y \in \overline{\phi}(\overline{U}_\alpha) \cap \mathcal{M}$ and some random variable L with finite p-th moment.

Proposition 3. *Let $K(\cdot, v) \in \mathcal{C}^1(\overline{\mathcal{M}})$ and $\mathbb{E}\left[X(v)^2\right] < \infty$ for all $\nu \in \mathcal{V}$. Then there exists a constant $\kappa > 0$ such that*

$$\mathbb{E}\left[\left(\tilde{X}_\alpha(x) - \tilde{X}_\alpha(y) \right)^2 \right] \leq \kappa \left| \log \|x - y\| \right|^{-2} \tag{9}$$

for the charts $(\overline{U}_\alpha, \overline{\phi}_\alpha)$, $\alpha \in \{1, \ldots, P\}$, in the atlas of $\overline{\mathcal{M}}$ covering $\mathcal{M}$, all $x, y \in \overline{\phi}(\overline{U}_\alpha) \cap \mathcal{M}$ such that $0 < \|x - y\| < 1$.

Remark 6. Proposition 3 can be used to establish condition **(G3)** for a $\mathcal{C}^3$-kernel K because each second order partial derivative of a normalized SuRF is a continuous SuRF itself.

Definition 3. Given a set $\mathcal{W} \subset \mathbb{R}^D$, we say that functions $f_1, \ldots, f_J : \mathcal{V} \to \mathbb{R}$ are $\mathcal{W}$-linearly independent if given constants $a_1, \ldots, a_J \in \mathbb{R}$, the relation

$$\sum_{j=1}^{J} a_j f_j(v) = 0$$

holding for all $v \in \mathcal{W}$ implies that $a_j = 0$ for all $j \in \{1, \ldots, J\}$.

Given this definition we have the following proposition.

Proposition 4. *Let $(\tilde{X}, X, K, \mathcal{V})$ be a SuRF with $K(\cdot, v) \in C^2(\overline{\mathcal{M}})$ for all $v \in \mathcal{V}$ and $\tilde{Z}$ denote the corresponding normalized SuRF. For $\alpha \in \{1, \ldots, P\}$ and $x \in \overline{\phi}_\alpha(\overline{U}_\alpha) \cap \mathcal{M}$ define $K_\alpha(x, \cdot) : \mathcal{V} \to \mathbb{R}$ by $v \mapsto K(\overline{\phi}_\alpha^{-1}(x), v)$ and set $\mathcal{V}_x = \{v \in \mathcal{V} : K_\alpha(x, v) \neq 0\}$. Assume that $K_\alpha(x, \cdot)$, $\partial_d^x K_\alpha(x, \cdot)$ and $\partial_{d'd''}^x K_\alpha(x, \cdot)$ for $d, d', d'' \in \{1, \ldots, D\}$ with $1 \leq d' \leq d'' \leq D$ are $\mathcal{V}_x$-linearly independent and that the random vector $(X(v) : v \in \mathcal{V}_x)$ is non-degenerate. Then*

$$\left(\tilde{X}_\alpha(x), \nabla \tilde{X}_\alpha(x), \mathbb{V}(\nabla^2 \tilde{X}_\alpha(x)) \right) \ and \ \left(\tilde{Z}_\alpha(x), \nabla \tilde{Z}_\alpha(x), \mathbb{V}(\nabla^2 \tilde{Z}_\alpha(x)) \right)$$

are non-degenerate Gaussian random vectors.

A sufficient condition, for the non-degeneracy condition on X to be satisfied, is that the vector $(X(v) : v \in \mathcal{V})$ is non-degenerate. The latter may often be a reasonable assumption in applications.

The next proposition states a condition under which the kernels $K_\alpha(x, \cdot)$ are $\mathcal{V}_x$-linearly independent, for all $x \in \overline{\phi}_\alpha(\overline{U}_\alpha \cap \mathcal{M})$. It follows immediately from the fact that if $\mathbf{a} \in \mathbb{R}^D$ is orthogonal to all elements of a set, which span $\mathbb{R}^D$, then $\mathbf{a} = 0$.

Proposition 5. *Let $\tilde{D} = D + 1 + D(D+1)/2$ and fix $\alpha \in \{1 \ldots, P\}$. Define the vector valued functions indexed by $x \in \overline{\phi}_\alpha(\overline{U}_\alpha \cap \mathcal{M})$ by*

$$\mathbf{K}_{\alpha,x}(v) = \left(K_\alpha(x, v), \nabla K_\alpha(x, v), \mathbb{V}(\nabla^2 K_\alpha(x, v)) \right), \tag{10}$$

where the gradient and Hessian are taken w.r.t. $x \in \overline{\phi}_\alpha(\overline{U}_\alpha \cap \mathcal{M})$. If for each $x \in \overline{\phi}_\alpha(\overline{U}_\alpha \cap \mathcal{M})$ there exist $v_1, \ldots, v_{\tilde{D}} \in \mathcal{V}_x$ such that $\mathbf{K}_{\alpha,x}(v_1), \ldots, \mathbf{K}_{\alpha,x}(v_{\tilde{D}})$ are linearly independent, then $K_\alpha(x, \cdot)$, $\partial_d^x K_\alpha(x, \cdot)$ and $\partial_{d'd''}^x K_\alpha(x, \cdot)$ for $d, d', d'' \in \{1, \ldots, D\}$ with $1 \leq d' \leq d'' \leq D$ are $\mathcal{V}_s$ linearly independent.

Remark 7. Note that the results in Propositions 4 and 5 are stronger than **(G2)**. This is because we want to emphasize that SuRFs often also satisfy the assumptions of the expectation Metatheorem 11.2.1 from Adler and Taylor [2007], compare Corollary 11.2.2 for the Gaussian version. However, it is obvious from the proofs that **(G2)** follows from the less strict assumption $\partial_d^x K(x, \cdot)$ and $\partial_{d'd''}^x K(x, \cdot)$ for $d \in \{1, \ldots, D\}$ and $1 \leq d' \leq d'' \leq D$ being $\mathcal{V}_x$-linearly independent for all $x \in \overline{\phi}(\overline{U}_\alpha) \cap \mathcal{M}$ and all $\alpha \in \{1, \ldots, P\}$.

The linear independence condition on the kernels is satisfied for a wide-range of kernels. In particular we have the following.

Proposition 6. *Let $K(s,v) = e^{-(s-v)^T \Omega (s-v)/2}$, for some positive definite matrix $\Omega \in \mathbb{R}^{D \times D}$ be the D-dimensional Gaussian kernel. Assume that $\mathcal{V}$ is a D-dimensional lattice which contains an element v such that*

$$\left\{ v + \sum_{d=1}^{D} \lambda_d \gamma_d e_d : \gamma_d \in \{-1, 0, 1\} \right\} \subset \mathcal{V}$$

where $(e_d)_{1 \leq d \leq D}$ is the standard basis and $\lambda \in \mathbb{R}_{>0}^D$ is fixed. Then $K(s, \cdot)$, $\partial_d^s K(s, \cdot)$ and $\partial_{d'd''}^s K(s, \cdot)$ for $d \in \{1, \ldots, D\}$ and $1 \leq d' \leq d'' \leq D$ are $\mathcal{V}_s$-linearly independent for each $s \in \mathbb{R}^D$.

In practice we want to apply the GKF to functions of these SuRFs. Having established the conditions above we can now state the GKF for SuRFs which can be viewed as a corollary to Theorem 12.4.2 from Adler and Taylor [2007]. In order to state the result we define $\mathcal{A}_u(f) = \{s \in \mathcal{M} : f(s) \geq u\}$ to be the excursion set of a random field f above the threshold u on $\mathcal{M}$. Moreover, we write $\chi_f(u)$ to denote the Euler characteristic of the excursion set $\mathcal{A}_u(f)$.

Theorem 1. *Let $(\tilde{X}_1, X_1, K, \mathcal{V}), \ldots, (\tilde{X}_N, X_N, K, \mathcal{V}) \sim (\tilde{X}, X, K, \mathcal{V})$ be i.i.d. SuRFs and $F \in \mathcal{C}^2(\overline{\mathcal{M}})$. Assume that X is a Gaussian field on $\mathcal{V}$ with covariance function $\mathfrak{c}$ and that for all $v \in \mathcal{V}$ it holds that $K(\cdot, v) \in C^3(\overline{\mathcal{M}})$ and $\mathfrak{c}(v, v) > 0$. Furthermore, for all $\alpha \in \{1, \ldots, P\}$ and $x \in \overline{\phi}(\overline{U}_\alpha) \cap \mathcal{M}$ assume that the random vector $(X(v) : v \in \mathcal{V}_x)$ is non-degenerate for the $\mathcal{V}_x$ defined in Proposition 4 and that $\partial_d^x K_\alpha(x, \cdot)$ and $\partial_{d'd''}^x K_\alpha(x, \cdot)$ for $d \in \{1, \ldots, D\}$ and $1 \leq d' \leq d'' \leq D$ are $\mathcal{V}_x$-linearly independent. Define a random field T such that $T(s) = F(\tilde{X}_1(s)/\|K_s\|, \ldots, \tilde{X}_N(s)/\|K_s\|)$ for all $s \in \overline{\mathcal{M}}$. Then*

$$\mathbb{E}[\chi_T(u)] = \sum_{d=0}^{D} \mathcal{L}_d \rho_d^T(u), \quad u \in \mathbb{R}, \tag{11}$$

where $\mathcal{L}_0, \ldots, \mathcal{L}_D \in \mathbb{R}$ are the LKCs of $\overline{\mathcal{M}}$ endowed with the induced Riemannian metric from $\tilde{X}(s)/\|K_s\|$ and ρ_d^T's are functions depending solely on the marginal distribution of T.

3.3 The Voxel Manifold Corresponding to a SuRF

The transformation of a discrete random field X on a finite grid $\mathcal{V}$ into a SuRF using a kernel K leaves open the question of the domain $\mathcal{M} \subset \overline{\mathcal{M}}$ of the SuRF since in principle any compact Whitney stratified manifold $\mathcal{M} \subset \overline{\mathcal{M}}$ can be chosen. In the case that $\mathcal{V} \subset \mathbb{R}^D$ and $K(\cdot, v) : \mathbb{R}^D \to \mathbb{R}$, for each $v \in \mathcal{V}$, the voxel manifold associated with $\mathcal{V}$, which is defined below, is a practical choice for the domain of a SuRF.

A voxel manifold for a discrete set $\mathcal{V}$ is obtained by glueing together D-dimensional boxes centered at $v \in \mathcal{V}$ having edge lengths in each coordinate direction given by the minimal distance between the corresponding coordinates of two different $v_1, v_2 \in \mathcal{V}$. The main advantage of using a voxel manifold is that LKCs and other geometric properties can be calculated efficiently, since the d-th dimensional boundaries of a voxel manifold lie in hyperplanes parallel to the coordinate axes, see Figure 1.

Definition 4. Suppose that $\mathcal{V} \subset \mathbb{R}^D$ is a discrete set and define $\delta \in \mathbb{R}^D$ such that its d-th component is $\delta_d = \min\{|v_d - w_d| : v, w \in \mathcal{V}, v_d \neq w_d\}$. Moreover let

$$\mathcal{B}_v(\delta) = \left\{ x \in \mathbb{R}^D \;\Big|\; \max_{d \in \{1, \ldots, D\}} |x_d - v_d| - \delta_d/2 \leq 0 \right\}. \tag{12}$$

Then the *voxel manifold associated with $\mathcal{V}$* is the set $\mathcal{M}_\mathcal{V} = \bigcup_{v \in \mathcal{V}} \mathcal{B}_v(\delta)$.

In order that a SuRF associated to K and $\mathcal{V}$ satisfies the GKF it is necessary to assume that $\mathcal{B}_v(\delta) \subset \text{supp}\big(K(\cdot, v)\big)$ for each $v \in \mathcal{V}$. Otherwise **(G2)** cannot be satisfied since the field is zero on parts of $\mathcal{M}_{\mathcal{V}}$. This condition on the support, however, is satisfied for most sensible choices of K since the kernel is typically used to increase the signal-to-noise ratio through the averaging of observations at different $v \in \mathcal{V}$.

A voxel manifold is, by construction, a stratified space. This can be seen, for example, for $D = 3$ by defining the three dimensional stratum to be the union of all the open cubes $\text{int}(\mathcal{B}_v(\delta))$ with $v \in \mathcal{V}$, while the two dimensional, the one dimensional and the zero dimensional strata are the unions over all $v \in V$ of all faces, edges and corners of the cubes $\mathcal{B}_v(\delta)$ respectively. Since a voxel manifold is the union of polyhedra, it is a Whitney-stratified manifold [Adler and Taylor, 2007, p.187].

The basic geometric quantities in Riemannian geometry are the Riemmanian metric and the Christoffel symbols. Using Theorem 5 and 6 from Appendix B.1 these quantities for the Riemannian metric induced by a normalized SuRF on $\overline{\mathcal{M}}$ can be written in terms of the inner products introduced in (7).

Proposition 7. *For $\overline{\mathcal{M}} \subset \mathbb{R}^D$ the Riemannian metric $\mathbf{\Lambda}$ induced by a normalized SuRF expressed in the local coordinates induced by $\iota : \overline{\mathcal{M}} \hookrightarrow \mathbb{R}^D$ is given by*

$$\Lambda_{dd'}(x) = \frac{\langle \partial_d K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^2} - \frac{\langle \partial_d K_x, \, K_x \rangle \langle K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^4} \tag{13}$$

and the Christoffel symbols of the first kind are

$$\begin{aligned}
\Gamma_{kdd'}(x) = \;&\frac{\langle \partial_k \partial_d K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^2} - \frac{\langle \partial_k \partial_d K_x, \, K_x \rangle \langle K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^4} \\
&- \frac{\langle \partial_k K_x, \, K_x \rangle \langle \partial_d K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^4} - \frac{\langle \partial_d K_x, \, K_x \rangle \langle \partial_k K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^4} \\
&+ 2\frac{\langle \partial_k K_x, \, K_x \rangle \langle \partial_d K_x, \, K_x \rangle \langle K_x, \, \partial_{d'} K_x \rangle}{\|K_x\|^6}
\end{aligned} \tag{14}$$

The advantage of using voxel manifolds as the domain is that the numerical implementation of these quantities is feasible. In particular it is straightforward to locally construct, for a small open neighborhood $\mathcal{U} \subset \mathcal{M}$ and $x \in \mathcal{U}$ an orthonormal frame field on $\mathcal{U}$. I.e., an orthonormal basis of $\mathcal{T}_x \mathcal{M}_{\mathcal{V}}$, $x \in \mathcal{M}_{\mathcal{V}} \cap \mathcal{U}$, such that a subset of this basis is an orthonormal frame of $\mathcal{T}_x \partial_d \mathcal{M}_{\mathcal{V}}$, $d \in \{1, \ldots, D-1\}$, if $x \in \partial_d \mathcal{M}_{\mathcal{V}} \cap \mathcal{U}$.

More concretely, let $D = 3$ and $k, l, m \in \{1, 2, 3\}$ be such that $k < l$, $\{k, l, m\} = \{1, 2, 3\}$. Then using the Gram-Schmidt procedure on the Euclidean basis e_1, e_2, e_3 yields orthonormal vector fields with respect to the Riemannian metric induced by a SuRF defined at $x \in \mathcal{M}_{\mathcal{V}}$ by

$$U_x = \Lambda_{kk}^{-1/2}(x)e_k \,, \quad V_x = \frac{\Lambda_{kl}(x)}{\sqrt{c(x)\Lambda_{kk}(x)}}e_k - \sqrt{\frac{\Lambda_{kk}(x)}{c(x)}}e_l \,,$$
$$N_x = \frac{\mathbf{\Lambda}^{-1}(x)}{\sqrt{e_m^T \mathbf{\Lambda}^{-1}(x)e_m}}e_m \,. \tag{15}$$

Here $c(x) = \det\big(\mathbf{\Lambda}^I(x)\big)$ for $I = (k, l)$ and U_x, V_x are a basis of $\mathcal{T}_x \mathcal{F}_I$, where $\mathcal{F}_I$ is the subset of $\partial_2 \mathcal{M}_{\mathcal{V}}$ such that the coordinates with indices not contained in I are constant, and N_x is in the one dimensional vector space orthogonal to $\mathcal{T}_x \mathcal{F}_I$ with respect to $\mathbf{\Lambda}$.

3.4 LKCs of voxel manifolds

General formulas for the LKCs of Whitney stratified manifolds can be found in [Adler and Taylor, 2007, Theorem 12.4.2] and more concrete formulas for Whitney stratified manifolds of dimension at most 3 are given in Appendix B.2. Since our voxel manifolds are embedded into $\mathbb{R}^D$, $D \in \mathbb{N}$, and have a simple geometric structure, the highest two LKCs can be expressed as integrals of (sub-)determinants of the Riemannian metric $\mathbf{\Lambda}$, i.e.,

$$\mathcal{L}_D = \sum_{\nu \in \mathcal{V}} \int_{\mathcal{B}_\nu(\delta)} \sqrt{\det\left(\mathbf{\Lambda}(x)\right)}\, dx\,, \quad \mathcal{L}_{D-1} = \sum_{I:|I|=D-1} \int_{\mathcal{F}_I} \sqrt{\det\left(\mathbf{\Lambda}^I(x)\right)}\, dx^I\,.$$

Here I is any ordered subset of $\{1,\ldots,D\}$, $dx^I = dx_{I_1}\ldots dx_{I_{|I|}}$ and $\mathbf{\Lambda}^I(x)$ is the submatrix of $\mathbf{\Lambda}(x)$ consisting of the columns and rows given by the entries of I. Finally, $\mathcal{F}_I$ is the subset of $\partial_{|I|}\mathcal{M}_\mathcal{V}$ such that the coordinates with indices not contained in I are constant.

The LKCs $\mathcal{L}_1,\ldots,\mathcal{L}_{D-2}$ of a D-dimensional voxel manifold with $D > 2$ are substantially harder to express explicitly, compare [Adler and Taylor, 2007, Theorem 12.4.2]. An expression for $\mathcal{L}_1$ of a 3-dimensional voxel manifold is given in the following proposition.

Theorem 2. *Let $\mathcal{M}_\mathcal{V}$ be a 3-dimensional voxel manifold. Then*

$$\mathcal{L}_1 = \frac{1}{2\pi} \sum_{|I|=1} \int_{\mathcal{F}_I} \Theta(x)\sqrt{\det\left(\mathbf{\Lambda}^I(x)\right)} dx^I$$

$$+ \frac{1}{2\pi} \sum_{|I|=2} \int_{\mathcal{F}_I} \left[\left(U_{I_1}(x)^2 + V_{I_1}(x)^2\right) N^T(x) \begin{pmatrix} \Gamma_{I_1 I_1 1}(x) \\ \Gamma_{I_1 I_1 2}(x) \\ \Gamma_{I_1 I_1 3}(x) \end{pmatrix} \right.$$

$$\left. + \sum_{k=1}^{2} V_{I_k}(x) V_{I_2}(x) N^T(x) \begin{pmatrix} \Gamma_{I_k I_2 1}(x) \\ \Gamma_{I_k I_2 2}(x) \\ \Gamma_{I_k I_2 3}(x) \end{pmatrix} \right] \sqrt{\det\left(\mathbf{\Lambda}^I(x)\right)} dx^I$$

$$- \frac{1}{2\pi} \sum_{v \in \mathcal{V}} \int_{\mathcal{B}_v(\delta)} \mathrm{Tr}\left(R(x)\right) \sqrt{\det\left(\mathbf{\Lambda}(x)\right)} dx$$

where $\mathrm{Tr}\left(R(x)\right)$ is the trace of the Riemannian curvature tensor, compare (36) in the supplementary material, and

$$\Theta(x) = \begin{cases} \pi - \beta(x)\,, & \text{if } x \text{ belongs to a convex edge} \\ -2\beta(x)\,, & \text{if } x \text{ belongs to a double convex edge}\,. \\ \beta(x) - \pi\,, & \text{if } x \text{ belongs to a convex edge} \end{cases}$$

Here $\beta(x)$ is defined using the crossproduct $V_x \times N_x = \left(m_1^I(x), m_2^I(x), m_3^I(x)\right)$ of the elements of the ONB from (15) with $k = I$ by

$$\beta(x) = \arccos\left(\frac{m_2(x)m_3(x)}{\sqrt{m_2^2(x) + m_1^2(x)}\sqrt{m_3^2(x) + m_1^2(x)}} \right)\,.$$

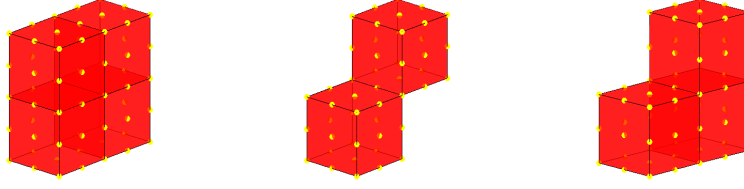*and the different types of edges are visualized in Figure 2.*

Figure 2: Visualization of the different types of edges appearing in a voxel manifold. In the left voxel manifold all edges are convex. The edge where the two cubes of the voxel manifold in the center are touching is a double convex edge and the same edge in the voxel manifold on the right is a concave edge since a third cube is added.

3.5 SuRF estimator for LKCs

We can use the theoretically derived formula (13) for the Riemannian metric $\mathbf{\Lambda}(x)$ to derive an estimate of the LKCs which does not rely on discrete derivatives since the latter introduce an additional source of numerical errors. The idea is to replace the population variances and covariances in (13) by their sample counterparts derived from the SuRF sample $\tilde{\mathbb{X}} = (\tilde{X}_1, \dots, \tilde{X}_N)$ obtained from an i.i.d. sample $\mathbb{X} = (X_1, \dots, X_N)$ of a discrete Gaussian random fields X observed on the lattice $\mathcal{V}$ and a kernel $K : \mathcal{M}_{\mathcal{V}} \times \mathcal{V} \to \mathbb{R}$. This results in the SuRF-Riemannian metric estimator

$$
\begin{aligned}
\hat{\Lambda}_{dd'}(x) = & \frac{\mathrm{Cov}\left[\partial_d \tilde{\mathbb{X}}(x),\, \partial_{d'}\tilde{\mathbb{X}}(x)\right]}{\mathrm{Var}\left[\tilde{\mathbb{X}}(x)\right]} \\
& - \frac{\mathrm{Cov}\left[\partial_d \tilde{\mathbb{X}}(x),\, \tilde{\mathbb{X}}(x)\right] \mathrm{Cov}\left[\tilde{\mathbb{X}}(x),\, \partial_{d'}\tilde{\mathbb{X}}(x)\right]}{\mathrm{Var}\left[\tilde{\mathbb{X}}(x)\right]^2},
\end{aligned}
\tag{16}
$$

where the variances and covariances are sample variances and covariances of the sample $\tilde{\mathbb{X}}$ and its derivatives. Note that derivatives of the form $\partial_d \tilde{\mathbb{X}}(x)$ can be explicitly computed from the derivatives of the kernel K, compare Proposition 1. Hence, our estimator does not rely on numerical derivatives. We denote with $\hat{\mathbf{\Lambda}}(x) \in \mathbb{R}^{D \times D}$, $x \in \mathcal{M}_{\mathcal{V}}$, the matrix with (d, d')-th entry $\hat{\Lambda}_{dd'}(x)$.

Numerically it is impossible to evaluate $\hat{\mathbf{\Lambda}}(x)$ for all $x \in \mathcal{M}_{\mathcal{V}}$. Thus, in order to estimate the LKCs we evaluate $\hat{\mathbf{\Lambda}}(x)$ on a grid $\mathcal{M}_{\mathcal{V}}^{(r)} \subset \mathcal{M}_{\mathcal{V}}$ given by

$$
\mathcal{M}_{\mathcal{V}}^{(r)} = \bigcup_{v \in \mathcal{V}} \mathcal{B}_v(\delta) \cap \left(v + \tfrac{\delta}{r+1} \cdot \mathbb{Z}^D\right).
\tag{17}
$$

Here $r \in \left\{ 2r' + 1 \mid r' \in \mathbb{N} \right\}$, i.e., $r \geq 1$ and odd, is called the *added resolution*. The restriction to odd numbers is required to ensure that we sample the boundary of $\mathcal{M}_{\mathcal{V}}$.

13

The *SuRF-LKC estimators with added resolution r* for $\mathcal{L}_D$ and $\mathcal{L}_{D-1}$ are given by

$$\hat{\mathcal{L}}_D^{(r)} = \sum_{x \in \mathcal{M}_{\mathcal{V}}^{(r)}} \sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)} \prod_{d=1}^D \frac{\delta_d}{r+1} \, .$$

$$\hat{\mathcal{L}}_{D-1}^{(r)} = \sum_{|I|=D-1} \sum_{x \in \mathcal{F}_I^{(r)}} \sqrt{\det\left(\hat{\mathbf{\Lambda}}^I(x)\right)} \prod_{i \in I} \frac{\delta_i}{r+1} \, .$$

Here $\mathcal{F}_I^{(r)} = \mathcal{F}_I \cap \mathcal{M}_{\mathcal{V}}^{(r)}$. These formulas can be implemented without difficulties in a computer program and letting r tend to infinity shows that the numerical error in approximating the integral can be made arbitrary small.

It is in principle possible to obtain in the same fashion estimators $\hat{\mathcal{L}}_{D-d}^{(r)}$, $d \in \{1, \ldots D-2\}$ from the LKC formulas in [Adler and Taylor, 2007, Theorem 12.4.2] and our formulas of the geometric quantities induced by a SuRF on $\mathcal{M}_{\mathcal{V}}$, compare Corollary 7 and Appendix B.2. However, in practice it is tedious to implement these estimators. The computation time of the resulting estimators can be large, because for a 3-dimensional voxel manifold computation of the Christoffel symbols already requires 27 convolutions in addition to the 9 convolutions required to estimate the Riemannian metric. Computing the Riemannian curvature tensor would on top require another 36 convolutions.

A practical solution to this problem is to approximate the lower LKCs using their locally stationary counterparts. In the case $D = 3$ this means that only the first integral in Theorem 2 is implemented, which can be estimated by

$$\hat{\mathcal{L}}_1^{(r)} = \sum_{|I|=1} \sum_{v \in \mathcal{F}_I^{(r)}} \hat{\Theta}(v) \sqrt{\hat{\Lambda}^I(v)} \frac{\delta_I}{r+1} \, , \tag{18}$$

where $\hat{\Theta}$ is the plug-in estimate of Θ defined in Proposition 2 using the orthonormal frame U_x, V_x, N_x derived from $\hat{\mathbf{\Lambda}}$.

The following results show that our estimator for the highest two LKCs are unbiased and consistent.

Theorem 3. *Let K be the kernel of the SuRF. Assume that $K(\cdot, v) \in C^3$ for all $v \in \mathcal{V}$ and that $\mathbb{E}[X(v)] < \infty$ for all $v \in \mathcal{V}$. Then*

$$\lim_{r \to \infty} \mathbb{E}\left[\hat{\mathcal{L}}_d^{(r)}\right] = \mathbb{E}\left[\lim_{r \to \infty} \hat{\mathcal{L}}_d^{(r)}\right] = \mathcal{L}_d \, .$$

for $d \in \{D-1, D\}$.

Theorem 4. *Let $K(\cdot, v) \in C^3\left(\overline{\mathcal{M}}\right)$ for all $v \in \mathcal{V}$ and assume that $\mathbb{E}\left[X(v)^2\right] < \infty$. Then*

$$\lim_{N \to \infty} \lim_{r \to \infty} \hat{\mathcal{L}}_d^{(r)} = \lim_{r \to \infty} \lim_{N \to \infty} \hat{\mathcal{L}}_d^{(r)} = \mathcal{L}_d \, . \tag{19}$$

for $d \in \{D-1, D\}$.

Remark 8. We expect that a similar result can be derived for the complete estimator of $\mathcal{L}_1$, i.e., the plugin estimator resulting from Theorem 2 where also the $\Gamma_{kdd'}$'s from (14) and the Riemannian curvature (compare Appendix B.1), are estimated using the corresponding sample covariances. This result could be established along the same lines as the consistency in [Telschow et al., Section 3], but we leave this for future work since, currently, implementing this estimator seems infeasible.

3.6 FWER Control Using SuRFs

In this section we illustrate how to use the Gaussian Kinematic formula to provide familywise error rate (FWER) control. The idea of doing so dates back to Worsley et al. [1992]. Here we combine their approach with the SuRF framework in order to improve the power of RFT based voxelwise inference. In particular we shall consider the problem of detecting areas of non-zero signal μ in a signal plus noise model $X(v) = \mu(v) + \epsilon(v)$ for $v \in \mathcal{V}$ and a zero-mean random field ϵ, given an i.i.d. sample $X_1, \ldots, X_N \sim X$. The same approach can also be applied to the linear model and other probabilistic models for which the GKF holds, compare Worsley [1994], Taylor and Worsley [2007, 2008].

Assume the setting of Theorem 1 and take $F : \mathbb{R}^N \to \mathbb{R}$ such that for each $(a_1, \ldots, a_N) \in \mathbb{R}^N$,

$$F(a_1, \ldots, a_N) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_i \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(a_i - \frac{1}{N} \sum_{i=1}^{N} a_i \right)^2 \right)^{-1/2}.$$

Let $\tilde{X}_1, \ldots, \tilde{X}_N$ denote the sample of SuRFs with domain $\mathcal{M}_\mathcal{V}$ derived from random fields $X_1, \ldots, X_N$ with domain $\mathcal{V}$.[4] For $v \in \mathcal{V}$ let $\mu(v) = \mathbb{E}[X(v)]$ and for all $s \in \mathcal{M}_\mathcal{V}$, let

$$\tilde{\mu}(s) = \sum_{v \in \mathcal{V}} K(s, v) \mu(v).$$

Due to the invariance of F to scaling, for $s \in \mathcal{M}_\mathcal{V}$, we can write

$$\tilde{T}(s) := F\big(\tilde{X}_1(s)/\|K_s\|, \ldots, \tilde{X}_N(s)/\|K_s\|\big) = \frac{\sqrt{N}\hat{\mu}_N(s)}{\hat{\sigma}_N(s)} \tag{20}$$

where $\hat{\mu}_N(s) = \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_i(s)$ and $\hat{\sigma}_N(s) = \left(\frac{1}{N-1} \sum_{i=1}^{N} (\tilde{X}_i(s) - \hat{\mu}_N(s))^2 \right)^{1/2}$. Based on $\tilde{T}$ we want to construct a multiple hypothesis test for the family of hypotheses

$$\mathbf{H}_{0,s} : \ \tilde{\mu}(s) \leq 0 \quad vs. \quad \mathbf{H}_{1,s} : \ \tilde{\mu}(s) > 0 \tag{21}$$

over $\mathcal{M}_\mathcal{V}$ which controls the FWER in the strong sense at a pre-specified level $\alpha \in (0, 1)$. The two-sided hypothesis can be treated similarly.

Let the test reject $\mathbf{H}_{0,s}$ for $s \in \mathcal{M}_\mathcal{V}$ whenever $\tilde{T}(s) > u_\alpha$ and let $\mathcal{H}_0 = \big\{ s \in \mathcal{M}_\mathcal{V} \mid \tilde{\mu}(s) \leq 0 \big\} \subseteq \mathcal{M}_\mathcal{V}$ denote the set of true null hypotheses. Then the FWER at a threshold $u \in \mathbb{R}$ is given by

$$\mathrm{FWER}_{\tilde{T}}(u) = \mathbb{P}\left(\sup_{s \in \mathcal{H}_0} \tilde{T}(s) > u \right).$$

Defining $\tilde{T}_0 = \sqrt{N}\big(\hat{\mu}_N(s) - \tilde{\mu}(s)\big)/\hat{\sigma}_N(s)$, then any u_α satisfying

$$\mathbb{P}\big(\max_{s \in \mathcal{M}_\mathcal{V}} \tilde{T}_0(s) > u_\alpha \big) \leq \alpha$$

for some $\alpha \in (0, 1)$, controls the FWER in the strong sense at the level α because

$$\mathrm{FWER}_{\tilde{T}}(u_\alpha) \leq \mathbb{P}\left(\max_{s \in \mathcal{M}_\mathcal{V}} \tilde{T}_0(s) > u_\alpha \right) \leq \alpha$$

---

[4]Recall that for technical purposes, for the GKF to hold, the SuRFs need to actually be defined on a D-dimensional, compact manifold without boundary $\overline{\mathcal{M}_\mathcal{V}} \supset \mathcal{M}_\mathcal{V}$

for all $\mathcal{H}_0 \subseteq \mathcal{M}_\mathcal{V}$ and $\tilde{T}_0(s) = \tilde{T}(s)$ for all $s \in \mathcal{H}_0$.

To find such a threshold u_α we shall make use of the Euler characteristic approximation (Adler [1981], Adler and Taylor [2007], Taylor et al. [2005]) to the excursion probability. In particular, letting $M_u(\tilde{T}_0)$ be the number of local maxima of $\tilde{T}_0$ over $\mathcal{M}_\mathcal{V}$ that lie above the level u and $\chi_{\tilde{T}_0}(u)$ be the Euler characteristic of the excursion set $\{s \in \mathcal{M}_\mathcal{V} \mid \tilde{T}_0(s) > u\}$ we have that

$$\mathbb{P}\left(\max_{s \in \mathcal{M}_\mathcal{V}} \tilde{T}_0(s) > u\right) \leq \mathbb{E}\left[M_u(\tilde{T}_0)\right] \approx \mathbb{E}\left[\chi_{\tilde{T}_0}(u)\right] = \sum_{d=0}^{D} \mathcal{L}_d \rho_d^{\tilde{T}_0}(u). \tag{22}$$

Given estimates of the LKCs $\hat{\mathcal{L}}_0, \ldots, \hat{\mathcal{L}}_D$ we approximate the expected Euler characteristic via

$$\sum_{d=0}^{D} \hat{\mathcal{L}}_d \rho_d^{\tilde{T}_0}(u).$$

Here $\rho_d^{\tilde{T}_0}$ are the Euler characteristic densities of a (centered) t-field given for example in [Taylor and Worsley, 2007, p.915]. In order to control the FWER in the strong sense to a level $\alpha \in (0, 1)$, we can thus find the largest u_α such that $\sum_{d=0}^{D} \hat{\mathcal{L}}_d \rho_d^T(u_\alpha) = \alpha$. At high thresholds u_α the number of local maxima is either zero or one and so in fact $\mathbb{E}\left[\chi_{\tilde{T}_0}(u_\alpha)\right]$ is an extremely good approximation to $\mathbb{E}\left[M_{u_\alpha}(\tilde{T}_0)\right]$. Lower values of α yield higher thresholds u_α, thus in practice where it is typical to take $\alpha \leq 0.05$, we expect the approximation in (22) to be accurate.

Traditional RFT inference in neuroimaging [Worsley et al., 1992, 1996, Taylor and Worsley, 2007] uses the same framework but only evaluates the fields on the lattice $\mathcal{V}$ and uses the LKC estimators given in Forman et al. [1995], Kiebel et al. [1999] or Taylor and Worsley [2007] which are based on discrete derivatives. More precisely, for each $n \in \{1, \ldots, N\}$ it takes data X_n on a lattice $\mathcal{V}$ (corresponding to the centers of voxels making up the brain), smoothes it with a kernel K to obtain $\{\tilde{X}_n(v) : v \in \mathcal{V}\}$ and rejects all $v \in \mathcal{V}$ such that $\tilde{T}(v) > u_\alpha$ where u_α is obtained from the GKF approximation such that $\mathbb{P}\left(\max_{s \in \mathcal{M}} \tilde{T}_0(s) > u_\alpha\right) \approx \alpha$. Here the manifold $\mathcal{M} \subset \mathbb{R}^3$ represents for example the brain. $\mathcal{M}$ has never been defined precisely in the literature probably because it was assumed unnecessary by the good lattice assumption. By construction this approach leads to valid, but conservative inference since $\max_{v \in \mathcal{V}} \tilde{T}_0(s) \leq \max_{s \in \mathcal{M}} \tilde{T}_0(s)$ and u_α is chosen to approximate the tails of the distribution of $\max_{s \in \mathcal{M}} \tilde{T}_0(s)$ at level α and not the tails of the distribution of $\max_{v \in \mathcal{V}} \tilde{T}_0(v)$. The conservativeness has been observed in simulations in Hayasaka and Nichols [2003], Taylor et al. [2007] and in real data validations in Eklund et al. [2016], Davenport et al. [2023].

Our SuRF framework removes this conservativeness by specifying the underlying manifold $\mathcal{M}$ to be the known voxel manifold $\mathcal{M}_\mathcal{V}$. As the SuRF random fields $\tilde{X}_1, \ldots, \tilde{X}_N$ are defined over $\mathcal{M}_\mathcal{V}$ the application of the GKF within the SuRF framework is theoretically justified. In particular, this framework allows us to identify significant excursions of $\tilde{T}$ above u_α over the whole of $\mathcal{M}_\mathcal{V}$, not only over $\mathcal{V}$. Consequentially, our SuRF framework has strong FWER control over $\mathcal{M}_\mathcal{V}$ at level α up to the approximation in (22) and thus has a higher statistical power than traditional RFT.

Effect Localization In this type of analysis, as is commonly done in neuroimaging, we have applied smoothing to the data. This means that precise localisation of effects is only possible for the smoothed signal $\tilde{\mu}$. In the following we explore what can be said about the original signal μ when using these methods. Note that our arguments carry over to permutation tests

and other tests controlling the FWER in the strong sense with respect to the smoothed signal (as these tests are typically applied to the data after applying smoothing).

The key observation is that for any $s \in \mathcal{M}_\mathcal{V}$ it holds that

$$\tilde{\mu}(s) > 0 \iff \exists v \in \mathcal{V} \cap \operatorname{supp}\big(K(s, \cdot)\big) : \ \mu(v) > 0 \tag{23}$$

provided that $K(s, v) \geq 0$ for all $s \in \mathcal{M}_\mathcal{V}$ and all $v \in \mathcal{V}$. Assume we reject $\mathbf{H}_{0,s}$. By (23) we can conclude that there is at least one $v \in \mathcal{V} \cap \operatorname{supp}\big(K(s, \cdot)\big)$ such that $\mu(v) > 0$. Since we have strong control on the hypotheses (21) it holds that

$$\mathbb{P}\Big(\big\{ s' \in S \mid \mathbf{H}_{0,s'} \text{ is rejected} \big\} \subseteq \mathcal{M}_\mathcal{V} \setminus \mathcal{H}_0 \Big)$$
$$= 1 - \mathbb{P}\Big(\exists s \in \mathcal{H}_0 : \ \mathbf{H}_{0,s} \text{ is rejected} \Big)$$
$$\geq 1 - \alpha,$$

which implies that the probability of incorrectly claiming that somewhere $\mu(v) > 0$ on $\mathcal{V} \cap \operatorname{supp}\big(K(s, \cdot)\big)$ is at most α. Thus, the strong control of the multiple hypotheses (21) implies a weaker form of FWER control on the signal μ than controlling it in the strong sense. However, it is stronger than controlling the FWER in the weak sense as long as $\operatorname{supp}\big(K(s, \cdot)\big) \neq S$ for at least one $s \in S$.

In practice it is often difficult to get strong control with respect to the original signal μ since smoothing is needed to increase the low signal-to-noise ratio in neuroimaging experiments. Furthermore, it is only plausible to talk about the BOLD signal at a single voxel in the idealistic world of a probabilistic model, but not in a real fMRI data analysis because the analysed BOLD signal is already a distorted version of the observed data due to the extensive preprocessing pipeline which includes, among other aspects, motion correction and warping to a standardized brain.

Because the signal-to-noise ratio is increased by applying smoothing, the type of FWER control on the hypotheses $\mathbf{H}_{0,v} : \ \mu(v) \leq 0$ vs. $\mathbf{H}_{1,v} : \ \mu(v) > 0$ for $v \in \mathcal{V}$, which follows from strong FWER control on the hypotheses (21), satisfies a "Heisenberg's uncertainty principle": *The smaller the supports of $K(\cdot, s)$, $s \in S$, the better the test localizes effects, yet at the cost of losing power to detect them. On the other hand the higher the power to detect effects, i.e., the larger the support of $K(\cdot, s)$, $s \in S$, the less precise the test localizes effects.*

4 Simulations

In this section we first use Monte-Carlo simulations to compare the performance of the SuRF estimator of the LKCs from Section 3.5 to existing LKC estimators. The alternatives we consider are the Hermite projection estimator (HP) and its bootstrap improvement (bHP) from [Telschow et al.] as well as the LKC estimators developed for stationary processes from Kiebel et al. [1999] and Forman et al. [1995] which are used in established software packages such as SPM and FSL. We abbreviate the latter two estimators as *Kie* and *For* respectively. We do not compare to the warping estimator from Taylor and Worsley [2007] since the estimates, although computed differently, are almost identical to the estimates of the bHP estimator, see [Telschow et al.].

Our second set of Monte-Carlo simulations shows that our SuRF framework allows for accurate control of the FWER in mass-univariate testing, while the traditional approach currently used in major neuroimaging software packages is conservative. A further validation the SuRF

framework can be found in our accompanying article Davenport et al. [2022a] where voxelwise inference using SuRFs is applied to a null data sets created using resting state data from the UK Biobank.

4.1 Simulation Setup

We will generate our simulations by smoothing the Gaussian random field $\{X(v) : v \in \mathcal{V}\}$ where $X(v)$ are i.i.d. $\mathcal{N}(0,1)$ distributed and $\mathcal{V} \subset \mathbb{R}^D$. To transform X into a SuRF we smooth using the isotropic Gaussian kernel

$$K_f(x,v) = \left(\frac{4\log(2)}{\pi f^2}\right)^{D/2} e^{\frac{4\log(2)\|x-v\|^2}{f^2}}, \quad f \in \{1,2,,\ldots,6\}, \tag{24}$$

which we parametrize by its full width at half maximum (FWHM) f as is common in neuroimaging. We demonstrate the performance of different LKC estimators on a standardized almost stationary SuRF and a standardized non-stationary SuRF, which we define below.

The advantage of starting from the field X which is i.i.d. on the lattice is that we can determine the theoretical LKCs for the normalized almost stationary and the normalized non-stationary SuRF quickly with high precision on a computer since the double sum in the Riemannian metric induced by the SuRF reduces to a single sum due to the lack of correlation, compare Proposition 7 and (7).

For the almost stationary simulations, for $D \in \{1,2,3\}$, $a \geq 0$, we take $\mathcal{V} = \mathcal{V}_D^a \subset \mathbb{R}^D$, where $\mathcal{V}_1^a = [1-a, 100+a] \cap \mathbb{Z}$, $\mathcal{V}_2^a = [1-a, 20+a]^2 \cap \mathbb{Z}^2$ and $\mathcal{V}_3^a = [1-a, 20+a]^3 \cap \mathbb{Z}^3$. The almost stationary SuRF is given by $(\tilde{X}, X, K_f, \mathcal{V}_D^a)$ with $a = \sqrt{2}f/\sqrt{\log(2)}$ and $\tilde{X}$ where we choose $\mathcal{M}_{\mathcal{V}_D^0}$ to be the domain of $\tilde{X}$. In this setting we expand the initial lattice by a size a in each direction and then restrict $\tilde{X}$ to the voxel manifold $\mathcal{M}_{\mathcal{V}_D^0}$ to remove boundary effects in order to allow a comparison with other methods which assume stationarity. We call this SuRF almost stationary because it is a stationary process if the domain is restricted to $\mathcal{V}_D^0$, but it is not stationary on $\mathcal{M}_{\mathcal{V}_D^0}$. However, for f larger than ≈ 2.5 the LKCs of this field are almost identical to the LKCs of a stationary zero-mean, unit variance Gaussian random field defined on $\mathcal{M}_{\mathcal{V}_D^0}$ with the covariance function $c(s,s') = K_f(s,s')$ with $s,s' \in \mathcal{M}_{\mathcal{V}_D^0}$ as can be verified using the RFTtoolbox Davenport and Telschow [2023] with which the theoretical LKCs of this process can be approximated with arbitrary precision, compare Tables 2-4 in Appendix A.4. This is also supported by our observation in Davenport et al. [2023] that the SuRF LKC estimators yield precise smoothness estimates even at $f \approx 2.5$ where state-of-the-art methods in neuroimaging are well known to be biased. Note that the LKCs for stationary random fields, once standardized, are well known, see Worsley et al. [1996]. Moreover the example of smoothing i.i.d. white noise with a Gaussian kernel is the typical way that these types of methods have been validated in the literature [Hayasaka and Nichols, 2003, Taylor et al., 2007]. As such this is a important example on which to evaluate and compare the different LKC estimators, even though in practice we expect the data to be non-stationary and so this setting may not be realistic.

In order to show that our methods are robust to non-stationarity we consider simulations where we induce non-stationarity into the field via our choice of initial lattice and by deliberately

not correcting for boundary effects. To do so we let $\mathcal{V} = \mathcal{V}_D^0$, for $D \in \{1, 2, 3\}$ where

$$
\begin{aligned}
\mathcal{V}_1 =& ([1, 100] \cap \mathbb{Z}) \\
& \setminus \{2, 4, 8, 9, 11, 15, 20, 21, 22, 40, \ldots, 45, 60, 62, 64, 65, 98, \ldots, 100\} \\
\mathcal{V}_2 =& \{x \in \mathbb{R}^2 \mid x_1 \in \{1, 2, 19, 20\} \ \vee \ x_2 \in \{1, 2, 19, 20\}\} \cap [1, 20]^2 \\
\mathcal{V}_3 =& \{x \in \mathbb{R}^3 \mid x_1 \in \{1, 2, 19, 20\} \ \vee \ x_2 \in \{1, 2, 19, 20\} \ \vee \ x_3 \in \{1, 2, 19, 20\}\} \\
& \cap [1, 20]^3 \, .
\end{aligned}
$$

and take the voxel manifold $\mathcal{M}_{\mathcal{V}_D}$ to be the domain of the corresponding SuRF. In this example we actually take advantage of the boundary effects in order to obtain non-stationary random fields.

In each of the above example simulation settings, given a sample size $N \in \mathbb{N}$, we generate SuRFs $\tilde{X}_1, \ldots, \tilde{X}_N$ of SuRFs from $X_1, \ldots, X_N \sim X$. An illustration of individual sample fields in the case that $D = 2$ is illustrated in Figure 3.
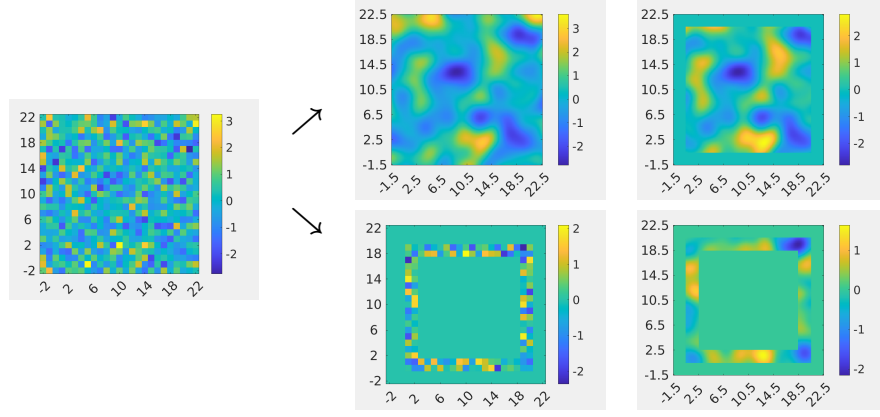

Figure 3: An example of the generation of our almost stationary and non-stationary SuRFs from the same white noise field on a grid using an isotropic Gaussian kernel with $f = 3$. For the generation of the almost stationary field the convolution with the kernel K_3 is applied before restricting the domain of the field in order to circumvent boundary effects (top row). The non-stationary field (bottom row) is instead generated by restricting to a narrow domain and then applying the convolution with the kernel K_3 without correcting for edge effects. The edge effects, in this case by design, induce non-stationarity into the field.

4.2   Results of the LKC estimation

We study the LKC estimates using the SuRF estimator from Section 3.5 for resolutions $r \in \{1, 3, 5\}$ and compare them to the estimates of the HP, bHP, Kiebel and Forman estimator obtained from evaluating samples of SuRFs on the resolution increased grids. Here we only report the results for $D = 2$. The results for $D \in \{1, 3\}$ can be found in Appendix A and are qualitatively the same as those for $D = 2$. In each simulation setting, given a sample size $N \in \mathbb{N}$, we run 1000 simulations. For each simulation we generate SuRFs $\tilde{X}_1, \ldots, \tilde{X}_N$ as described in Section 4.1 and use these fields to obtain estimates of the LKCs using the different methods.

Figure 4 shows boxplots of the resulting LKC estimates for fixed $f = 3$ and varying sample sizes $N \in \{20, 50, 100\}$ in the almost stationary setting. Figure 5 contains the same results

for the non-stationary setting. At all resolutions the SuRF estimator seems to be unbiased and has a lower variance than the other estimators. Only the bHP estimator is comparable efficient, however, it has a small bias at resolution $r = 1$. The HP estimator has a similar bias and a much larger variance, compare also Telschow et al.. The Kiebel and Forman estimators are biased for our non-stationary SuRF example and are even biased for the almost stationary SuRF for small values of f. This bias has been observed in the literature, see e.g., Kiebel et al. [1999].

In Figure 6 and 7 we illustrate how the performance of the estimates of $\mathcal{L}_2$ depends on the smoothing bandwidth f. The corresponding plots for the estimates of $\mathcal{L}_1$ can be found in Figures 13 and 14 in Appendix A. The SuRF estimator is the only approach which correctly estimates the LKCs even for f smaller or equal than 3 which is a typical FWHM of the Gaussian kernel used in neuroimaging. For the SuRF estimator only for $f = 1$ is a resolution increase of at least $r = 3$ necessary to have unbiased estimates. Remarkably, already for $f = 2$ a resolution increase of $r = 1$ seems to be sufficient for unbiased estimation of the LKCs using the SuRF estimator. Another important observation is that the SuRF estimator is 10 times faster to compute than its only reliable competitor the bHP estimator, see Table 1. In this simulation we compared the average computation time of the SuRF, the bHPE and the Kiebel estimator on the stationary box example with bandwidth $f = 3$ and sample size $N = 100$ at different added resolution.
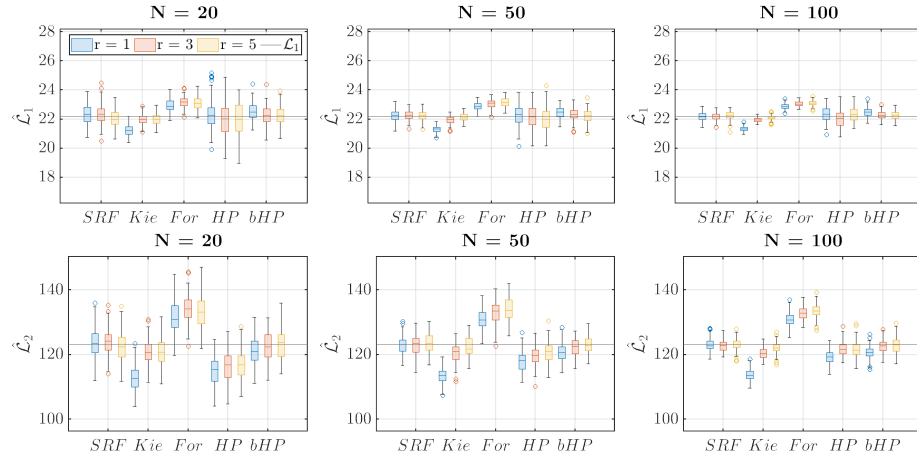

Figure 4: 2D Simulation results for estimation of the LKCs of the almost stationary SuRF described in Section 4.1. The smoothing bandwidth is fixed to $f = 3$.
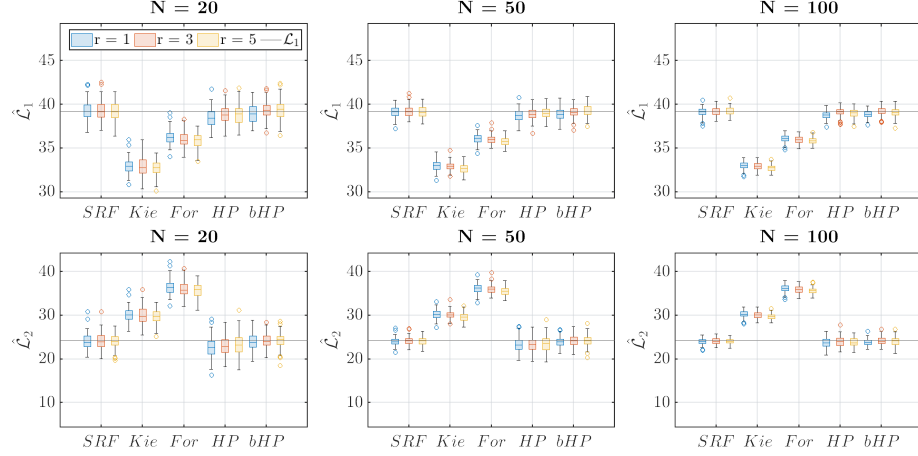
Figure 5: 2D Simulation results for estimation of the LKCs of the non-stationary SuRF described in Section 4.1. The smoothing bandwidth is fixed to $f = 3$.
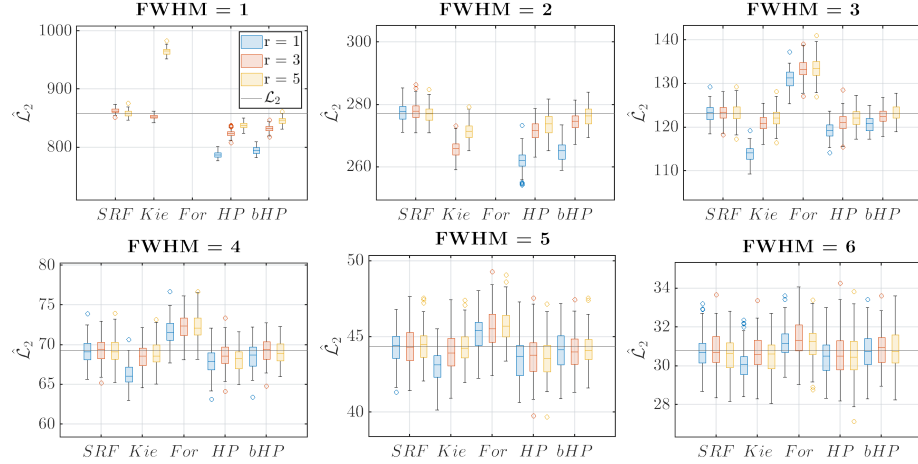

Figure 6: 2D Simulation results for estimation of the LKCs of SuRFs derived from the stationary box example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
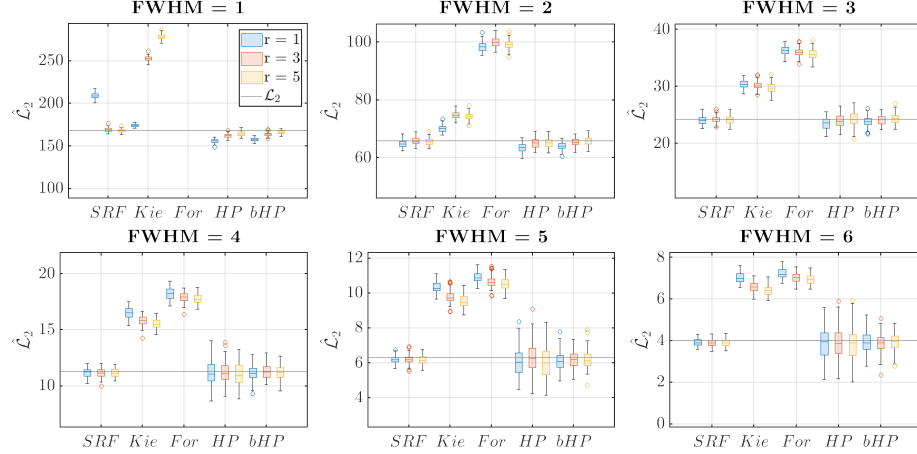
Figure 7: 2D Simulation results for the estimation of the LKCs of SuRFs in the non-stationary setting. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The sample size $N = 100$ is fixed.

	SuRF			bHPE			Kiebel		
r	1	3	5	1	3	5	1	3	5
$D = 1$	0.06	0.05	0.05	7.01	13.35	16.13	4	0.03	0.04
$D = 2$	0.28	0.66	1.66	19.96	30.67	46.67	0.11	0.42	0.97
$D = 3$	18.22	80.45	239.30	196.44	-	-	8.24	47.90	151.75

Table 1: Simulation time of different LKC estimators using the stationary box example with $N = 100$ and FWHM = 3. The values are the average computation time in seconds of one hundred runs of the estimators. We do not report the time for resolution increases beyond $r = 1$ of the bHPE for $D = 3$ because they are unrealistically long.

4.3   Results of the FWER simulation

In this section we illustrate the improvements that are provided by the SuRF framework, in terms of FWER with respect to traditional RFT [Worsley et al., 1996]. To do so we calculate the FWER that results from using RFT (as described in Section 3.6) to threshold one-sample t-fields obtained from Gaussian random fields generated as described in Section 4.1. In each simulation setting, for $1 \leq b \leq B$ and $N \in \{20, 50, 100\}$ we obtain t-fields $T_{N,b,r}$ of resolution $r \in \{0, 1, \infty\}$. Here $r = 0$ corresponds to the traditional RFT approach, i.e., $T_{N,b,0}$ is the test statistic described in Section 3.6 evaluated on the lattice $\mathcal{V}$ on which the original data is observed. Similarly, the case $r = 1$ corresponds to $T_{N,b,1}$ being the test statistic evaluated on $\mathcal{M}_{\mathcal{V}}^{(1)}$, i.e., $\mathcal{V} \cap \mathcal{M}_{\mathcal{V}}$ with an added resolution of 1, compare (17). The case $r = \infty$ corresponds to the use of a SuRF, i.e., $T_{N,b,\infty}$ is the test statistic on $\mathcal{M}_{\mathcal{V}}$. In order to evaluate the FWER in the latter case we use numerical optimization methods (in particular sequential quadratic programming, Nocedal and Wright [2006]), initialized at the largest peaks of $T_{N,b,1}$ to find the global maximum of $T_{N,b,\infty}$ over $\mathcal{M}_{\mathcal{V}}$.

In each simulation we estimate the LKCs as described in Section 3.6 from using the sampling $\mathcal{M}_{\mathcal{V}}^{(1)}$ of $\mathcal{M}_{\mathcal{V}}$. Adding only a resolution of 1 to the lattice $\mathcal{V}$ for the estimation of the LKCs is justified by the simulation results of Section 4.2. In order to control the FWER at a level

$\alpha = 0.05$, we then use the estimated LKCs to obtain an α-level threshold $\hat{u}_{N,b}$ as described in Section 3.6. This allows us to estimate the FWER, for each $r \in \{0, 1, \infty\}$, by evaluating

$$\frac{1}{B} \sum_{b=1}^{B} 1 \left[\sup_{s \in \mathcal{M}_{\mathcal{V}}^{(r)}} T_{N,b,r}(s) > \hat{u}_{N,b} \right] ,$$

where $\mathcal{M}_{\mathcal{V}}^{(0)} = \mathcal{V}$ and $\mathcal{M}_{\mathcal{V}}^{(\infty)} = \mathcal{M}_{\mathcal{V}}$. The results in 2D and 3D are shown in Figures 8 and 9 respectively.

In each simulation we also count the number of local maxima. Let $l_{N,b,r}$ be the number of local maxima of $T_{N,b,r}$ on $\mathcal{M}_{\mathcal{V}}^{(r)}$ which exceed the level $\hat{u}_{N,b}$. The number of local maxima above the threshold is almost identical to the Euler characteristic of the excursion set (as the threshold has been chosen so that the EEC is 0.05). As such we can test whether our EEC estimate is unbiased by calculating

$$\widehat{\mathrm{EEC}} = \frac{1}{B} \sum_{b=1}^{B} l_{N,b,r}.$$

This quantity is plotted in yellow in the figures.

These results reproduce the well-known observation that the traditional ($r = 0$) approach is conservative, while our proposed the SuRF approach controls the FWER at the nominal rate in most settings. The difference is particularly observable at low smoothness levels, but even at FWHM $= 6$, evaluations of the random field on the original lattice are conservative. This effect is slightly more pronounced in 3D. Moving from the original ($r = 0$) lattice to the resolution 1 lattice already reduces the conservativeness which demonstrates that the main cause of the conservativeness is the mismatch between the discreteness of the data and modeling it as a random field over a manifold.

As can be seen from the yellow curve in the plots the expected Euler characteristic is typically well estimated. Only at the lowest considered smoothness level the EEC is underestimated, causing a slight conservativeness in the SuRF approach at FWHM $= 1$. This conservativeness could also be a consequence of the poor performance at FWHM $= 1$ of the SuRF LKC estimator with added resolution of 1, compare for example the first plot in Figures 6 and 7.
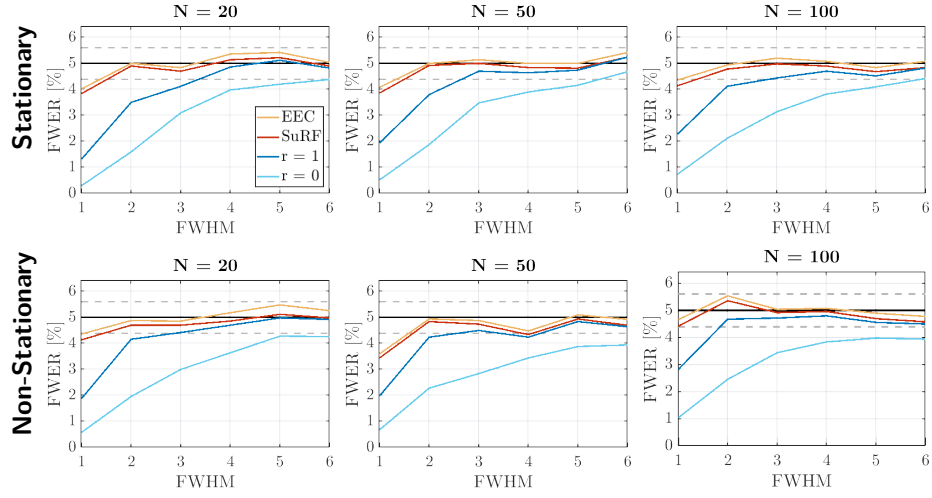
Figure 8: FWER results for the almost-stationary (top row) and non-stationary (bottom row) settings for the 2 dimensional domain. The error rate is controlled for all FWER controlling methods with the SuRF error rate being the least conservative. The EEC is accurately estimated within the SuRF framework for FWHM ≥ 2.
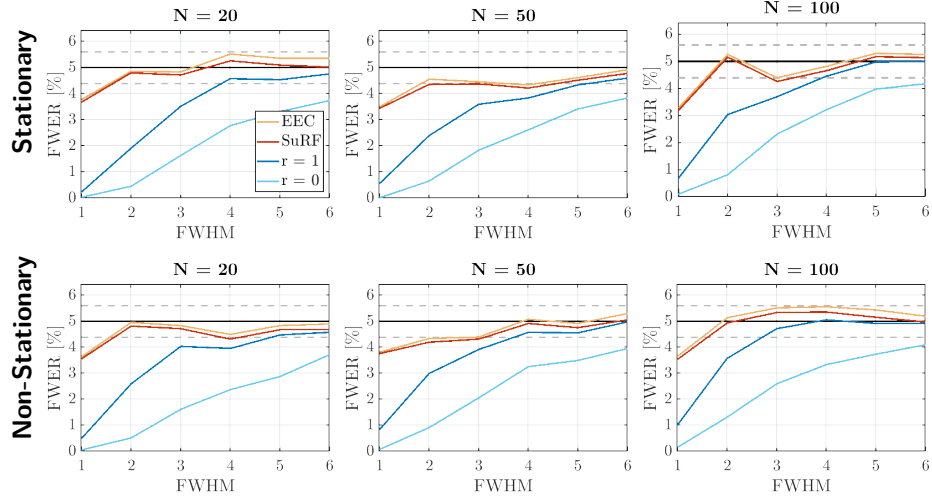

Figure 9: FWER results for the almost-stationary (top row) and non-stationary (bottom row) settings for the 3 dimensional domain. The description of the compared methods can be found in the caption of Figure 8.

5 Discussion

In this article we have provided a solution to the long-standing problem that voxelwise inference using RFT in neuroimaging is conservative (Nichols and Hayasaka [2003], Taylor et al. [2007]). This is important because the conservativeness of traditional RFT leads to a loss of power. The philosophical concept that we have sought to emphasize and employ to resolve this issue is the careful distinction between data of an experiment and atoms of a probabilistic theory (or data objects, Marron and Alonso [2014]). Doing so allowed us to identify the main causes of conservativeness in voxelwise inference using RFT in neuroimaging: (i) the test statistic is only evaluated on a voxel lattice defined by the data although the atoms of RFT are random fields over a Whitney-stratified manifold such as the brain and (ii) continuously available quantities are replaced by discrete counterparts. Our solution was to interpret the kernel smoother used in preprocessing neuroimaging data as a ferryman transferring discrete data observed on a voxel lattice into random fields over a Whitney stratified manifold - the atoms of RFT. Since the maximum of the SuRF test-statistic is strictly higher than the maximum on the lattice, using our proposed framework controls the FWER precisely at the pre-specified significance level and leads to an improvement in the statistical power to detect significant effects relative to traditional RFT.

Although we have restricted our SuRF framework to Gaussian related fields, this assumption can be relaxed. As demonstrated in Telschow and Schwartzman [2022] the GKF also approximates the expected Euler characteristic of the excursion sets of the test statistic $\tilde{T}$ well, if it is asymptotically a Gaussian random field[5], compare Nardi et al. [2008] for a similar approach. Hence for large enough sample sizes the presented methodology is applicable even under non-Gaussianity. In particular the estimates of the LKCs using the SuRF framework are consistent so long as derivatives of the sample covariance of the residuals converge uniformly to the derivatives of the covariance function of the limiting field, compare Telschow et al. for these type of arguments. Unfortunately, the sample size required to deal with the inherent non-Gaussianity in real fMRI experiments seems relatively large as we demonstrate in our accompanying work, Davenport et al. [2022a], which provides a solution to this issue.

In this work we validated our improvements using a range of simulations. In these we showed that, using the SuRF framework, the LKCs of SuRFs can be precisely estimated at the levels of smoothing typically encountered in applications and that voxelwise inference using RFT precisely controls the family-wise error rate at a prespecified level. In particular we identified (i) as the main cause of the observed conservativeness of traditional RFT approaches. Further validation of our methods using a large resting state data set from the UK Biobank can be found in our accompanying applied paper (Davenport et al. [2022a]).

The distinction between data and atoms also allowed us to shed some light on the precise localization of effects within frameworks that apply smoothing to the data before the statistical analysis is carried out. In such scenarios it is difficult to control the FWER in the strong sense for statistical hypotheses tests on the unsmoothed data. Nevertheless, as explained in Section 3.6 strong FWER control of the smoothed signal implies a weaker form of FWER control with respect to the unsmoothed signal which enables some degree of localization of the significant effects which depends on the support of the smoothing kernels.

The importance of our findings for neuroimaging is twofold. Firstly, our GKF based method is in general computationally faster than resampling based inference methods such as permutation tests which control the FWER directly over the grid $\mathcal{V}$ and therefore yield FWER tests

---

[5]$\tilde{T}$ satisfies a central limit theorem.

at level α at the cost of a high computational burden. Secondly, solving the conservativeness problem that has long caused power problems for voxelwise inference using RFT is a first step towards identifying and solving the problems of false positive rates in cluster inference (Eklund et al. [2016]) because the latter also relies on the GKF and applies continuous theory to smoothed fields evaluated on the voxel lattice [Chumbley and Friston, 2009].

The estimator for $\mathcal{L}_1$ in 3D obtainable from Theorem 2 is difficult to implement. In order to approximate it we used a local stationary approximation in our 3D simulations relying only on the first integral which is the only part which appears if the random field is stationary. For voxelwise inference using RFT this has a minimal effect, even when the fields are highly non-stationary because the estimation of u_α from the expected Euler characteristic is primarily driven by the values of $\hat{\mathcal{L}}_2$ and $\hat{\mathcal{L}}_3$. This can be seen from the fact that for typical values $u_\alpha \approx 4.2$ in neuroimaging (e.g., Supplementary material of Telschow et al.) it holds that $\rho_3^{\tilde{T}}(u_\alpha)/\rho_1^{\tilde{T}}(u_\alpha) \approx 2.6$ and $\rho_2^{\tilde{T}}(u_\alpha)/\rho_1^{\tilde{T}}(u_\alpha) \approx 1.6$ for sample size 50 and that typically $\mathcal{L}_1 \ll \mathcal{L}_3$, compare our simulation section and for real fMRI data the Supplementary material of Telschow et al.. The latter can be seen theoretically in the case of a stationary random field with square covariance function with bandwidth parameter h over a convex Whitney stratified manifold $\mathcal{M}$. Here $\mathcal{L}_3$ is the volume of $\mathcal{M}$ divided by h^3, $\mathcal{L}_2$ is half the surface area of $\mathcal{M}$ divided by h^2 and $\mathcal{L}_1$ is twice the diameter of $\mathcal{M}$ divided by h [Worsley et al., 2004, Table 2]. Note that the FWHM from their Table 2 needs to be transformed into a bandwidth by multiplication with $\sqrt{4\log(2)}/2\pi$ in order to transform resels (a concept introduced in Worsley et al. [1996] designed primarily for isotropic, stationary processes with a squared exponential covariance function) into LKCs. Additionally, our simulations from this article and in Davenport et al. [2022a] demonstrate that, although we use the local stationary approximation of $\mathcal{L}_1$, we are able to accurately control the false positive rate.

For clustersize inference the precise value of $\mathcal{L}_1$ may be more relevant as it has a greater effect at lower thresholds. As such future work should focus on obtaining an easier to evaluate form of $\mathcal{L}_1$ for three dimensional voxel manifold within the SuRF framework. Alternatively a modified version of the bHPE, which uses critical values of the SuRF, could be used to estimate $\mathcal{L}_1$.

Last but not least, we want to emphasize that observations (i) and (ii) are likely be relevant whenever continuous probabilistic theory is applied to discrete data. As such our proposed methodology based on SuRFs is not restricted to voxelwise inference in brain imaging. In fact, the ideas presented apply equally when considering the coverage of confidence bands or the family-wise error rate of tests in functional data analysis where smoothing is applied for example via basis function representations or kernel smoothers. *[SD: What do you mean by this: shouldn't it be especially not even?]*However, even for rough functional processes (i) needs to be carefully taken into account in simulations validating theory. Here the coverage or the FWER needs to be checked on sufficiently finely sampled instances of the processes or alternatively within the considered model itself. Otherwise it is likely that the reported coverage or type I error will be overestimated.

There are a number of further scenarios in which the SuRF methodology is applicable. A concrete example of this in practice is Davenport et al. [2022b]. Here confidence regions for peaks of the signal of random fields over bounded open domains of $\mathbb{R}^D$ are developed and convolution fields are used to localize peaks of activation in fMRI and MEG. In fact, MEG is a natural domain in which our theory applies, because the power spectrum can be written as a convolution field, see the Supplementary material of Davenport et al. [2022b] for details. A further potential application of the SuRF paradigm is to provide coverage probability excursion

(CoPE) sets which have the nominal coverage rate. CoPE sets provide confidence sets for the excursion above a value $c \in \mathbb{R}$ of a real-valued target function defined over a domain in $\mathbb{R}^D$, $D > 0$, from noisy data, [Sommerfeld et al., 2018, Bowring et al., 2019, 2021]. These works found in simulations that the empirical coverage obtained by the CoPE sets is typically larger than the specified coverage, but converges to the correct coverage probability as the domain is sufficiently densely sampled. This effect was theoretically explained in Theorem 1.b) from [Sommerfeld et al., 2018] and is well illustrated in Figure 3 of Bowring et al. [2019]. Using the SuRF paradigm in this setting would fix this overcoverage problem and allow for more precise CoPE sets that attain the nominal rate of coverage.

Acknowledgments

References

Russell Poldrack and Thomas E. Nichols. *Handbook of functional MRI data analysis*, volume 4. 2011. ISBN 9780521517669. URL http://webcat.warwick.ac.uk/record=b2542955{~}S15.

Coong-Wan Woo, Anjali Krishnan, and Tor D. Wager. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. pages 412–419, 2014. doi: 10.1016/j.neuroimage.2013.12.058.Cluster-extent.

Keith J Worsley, Alan C Evans, Sean Marrett, and P Neelin. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918, 1992.

Karl Friston, Keith Worsley, R Frackowiak, J Mazziotta, and A Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220, 1994.

Keith J Worsley, Sean Marrett, Peter Neelin, Alain C Vandal, Karl J Friston, and Alan C Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, 1996.

Robert J. Adler. *The Geometry of Random Fields.* 1981. ISBN 9780898716931.

Robert J Adler and Jonathan E Taylor. *Random fields and geometry.* Springer Science & Business Media, 2007.

Jonathan E Taylor et al. A gaussian kinematic formula. *The Annals of Probability*, 34(1):122–158, 2006.

Jonathan Taylor, Akimichi Takemura, and Robert J. Adler. Validity of the expected Euler characteristic heuristic. *Annals of Probability*, 33(4):1362–1396, 2005. ISSN 00911798. doi: 10.1214/009117905000000099.

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

K.J. Friston, John T Ashburner, Stefan J Kiebel, Thomas E. Nichols, and William D Penny. *Statistical Parametric Mapping*. ISBN 9780123725608.

Justin R Chumbley and Karl J Friston. False discovery rate revisited: Fdr and topological inference using gaussian random fields. *Neuroimage*, 44(1):62–70, 2009.

Thomas E. Nichols and Satoru Hayasaka. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12(5):419–446, 2003. ISSN 0962-2802. doi: 10.1191/0962280203sm341ra.

Thomas E Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

Anderson M. Winkler, Matthew A. Webster, Jonathan C. Brooks, Irene Tracey, Stephen M. Smith, and Thomas E. Nichols. Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping*, 37(4):1486–1511, 2016a. ISSN 10970193. doi: 10.1002/hbm.23115.

Anderson M. Winkler, Gerard R. Ridgway, Gwenaëlle Douaud, Thomas E. Nichols, and Stephen M. Smith. Faster permutation inference in brain imaging. *NeuroImage*, 141:502–516, 2016b. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.05.068.

Samuel Davenport, Fabian JE Telschow, Thomas E Nichols, and Armin Schwarzman. Accurate voxelwise fwer control in fmri using random field theory. *arXiv preprint arXiv:2201.01591*, 2022a.

Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper LR Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N Sotiropoulos, Saad Jbabdi, Moises Hernandez-Fernandez, Emmanuel Vallee, et al. Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, 166:400–424, 2018.

Haonan Wang and JS Marron. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.

J Steve Marron and Andrés M Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.

Stefan J. Kiebel, Jean-Baptiste Poline, Karl J. Friston, Andrew P. Holmes, and Keith J. Worsley. Robust Smoothness Estimation in Statistical Parametric Maps Using Standardized Residuals from the General Linear Model. *NeuroImage*, 10(6):756–766, 1999. ISSN 10538119. doi: 10.1006/nimg.1999.0508.

J. E. Taylor, K. J. Worsley, and F. Gosselin. Maxima of discretely sampled random fields, with an application to 'bubbles'. *Biometrika*, 94(1):1–18, 2007.

K. J. Worsley. An improved theoretical P value for SPMs based on discrete local maxima. *NeuroImage*, 28(4):1056–1062, 2005. ISSN 10538119. doi: 10.1016/j.neuroimage.2005.06.053.

Yen-Chi Chen, Christopher R Genovese, and Larry Wasserman. Density level sets: Asymptotics, inference, and visualization. *Journal of the American Statistical Association*, 112(520):1684–1696, 2017.

Max Sommerfeld, Stephan Sain, and Armin Schwartzman. Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate. *Journal of the American Statistical Association*, 1459:0–0, 2018. ISSN 0162-1459. doi: 10.1080/01621459.2017.1341838.

Alexander Bowring, Fabian Telschow, Armin Schwartzman, and Thomas E Nichols. Spatial confidence sets for raw effect size images. *NeuroImage*, 203:116187, 2019.

Alexander Bowring, Fabian JE Telschow, Armin Schwartzman, and Thomas E Nichols. Confidence sets for cohen'sd effect size images. *NeuroImage*, 226:117477, 2021.

Steven D. Forman, Jonathan D. Cohen, Mark Fitzgerald, William F. Eddy, Mark A. Mintun, and Douglas C. Noll. Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold. *Magnetic Resonance in Medicine*, 33(5):636–647, 1995. ISSN 15222594. doi: 10.1002/mrm.1910330508.

Jonathan E Taylor and K. J. Worsley. Detecting Sparse Signals in Random Fields, With an Application to Brain Mapping. *Journal of the American Statistical Association*, 102(479):913–928, 2007. ISSN 0162-1459. doi: 10.1198/016214507000000815.

Fabian Telschow, Dan Cheng, and Pratyush Pranav.

Samuel Davenport and Fabian Telschow. RFTtoolbox, 2023. URL https://github.com/sjdavenport/RFTtoolbox.

George Turin. An introduction to matched filters. *IRE transactions on Information theory*, 6(3):311–329, 1960.

K. J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A General Statistical Analysis for fMRI Data. *NeuroImage*, 15(1):1–15, 2002. ISSN 10538119. doi: 10.1006/nimg.2001.0933. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811901909334.

Samuel Davenport and Fabian JE Telschow. On the finiteness of the second moment of the number of critical points of gaussian random fields. *arXiv preprint arXiv:2201.01591*, 2022.

Keith J Worsley. Local maxima and the expected euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, 26(1):13–42, 1994.

J. E. Taylor and K. J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, 36(1):1–27, 2008. ISSN 00905364. doi: 10.1214/009053607000000406.

Satoru Hayasaka and Thomas E. Nichols. Validating cluster size inference: Random field and permutation methods. *NeuroImage*, 20(4):2343–2356, 2003. ISSN 10538119. doi: 10.1016/j.neuroimage.2003.08.003.

Samuel Davenport, Armin Schwarzman, Thomas E. Nichols, and Fabian Telschow. Accurate voxelwise fwer control in fmri using random field theory. 2023.

Jorge Nocedal and Stephen J Wright. Quadratic programming. *Numerical optimization*, pages 448–492, 2006.

Fabian JE Telschow and Armin Schwartzman. Simultaneous confidence bands for functional data using the Gaussian kinematic formula. *Journal of Statistical Planning and Inference*, 216:70–94, 2022.

Yuval Nardi, David O Siegmund, and Benjamin Yakir. The distribution of maxima of approximately gaussian random fields. 2008.

K. J. Worsley, Jonathan E. Taylor, Francesco Tomaiuolo, and Jason Lerch. Unified univariate and multivariate random field theory. *NeuroImage*, 23(SUPPL. 1):189–195, 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.07.026.

Samuel Davenport, Thomas E Nichols, and Armin Schwarzman. Confidence regions for the location of peaks of a smooth random field. *arXiv preprint arXiv:2208.00251*, 2022b.

Jan Magnus and Heinz Neudecker. The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods*, pages 422–449, 1980.

Kantilal Varichand Mardia, John T Kent, and John M Bibby. Multivariate analysis. *Probability and mathematical statistics*, 1979.

Appendix

A   Additional Simulation Results and Tables
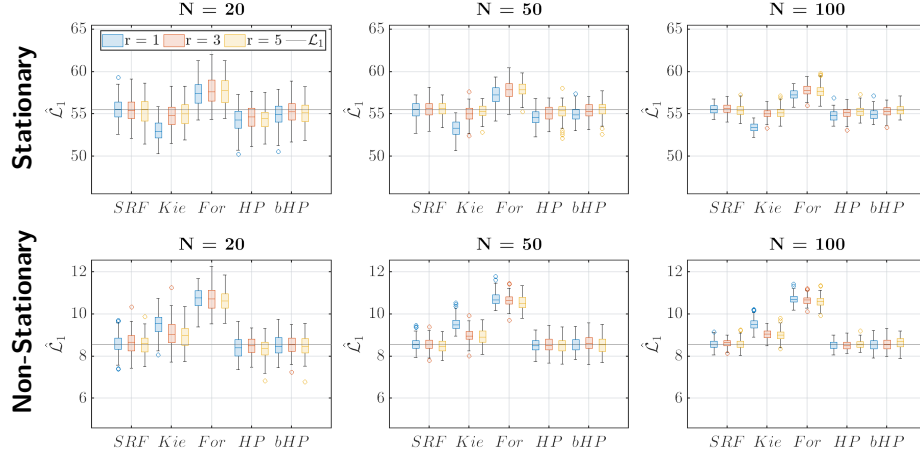
A.1   LKC estimation for $D = 1$

Figure 10: 1D Simulation results of estimation of the LKCs of the two SuRFs derived from the stationary box example and the non-stationary sphere example. The FWHM is fixed to 3.
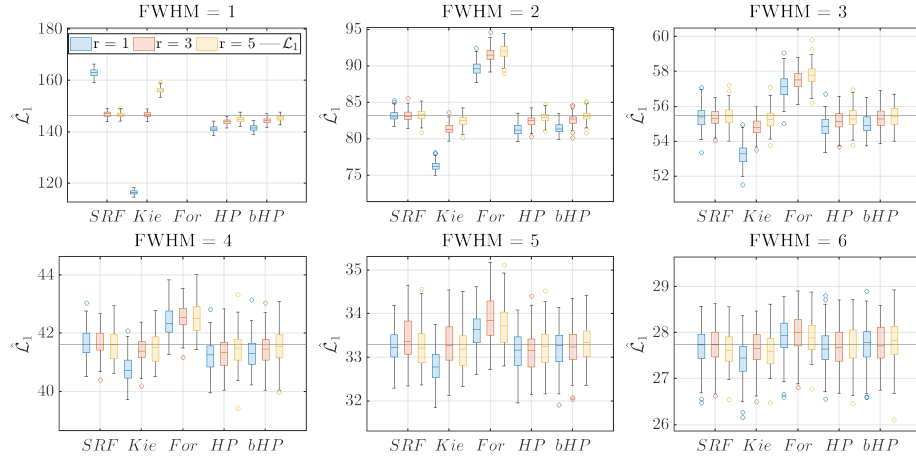

Figure 11:   1D Simulation results of estimation of the LKCs of SuRFs derived from the stationary box example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.

Figure 12: 1D Simulation results of estimation of the LKCs of SuRFs derived from the non-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.

A.2 LKC estimation for $D = 2$

Figure 13: 2D Simulation results for estimation of the LKCs of the almost stationary SuRF described in Section 4.1. The figures show the dependence of the LKC estimation on the smoothing bandwidth f while $N = 100$ is fixed.
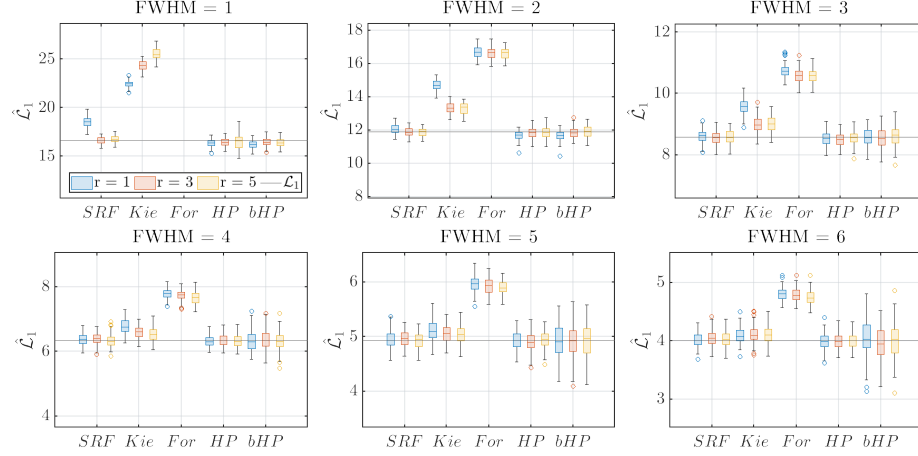
Figure 14: 2D Simulation results of estimation of the LKCs of SuRFs derived from the non-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
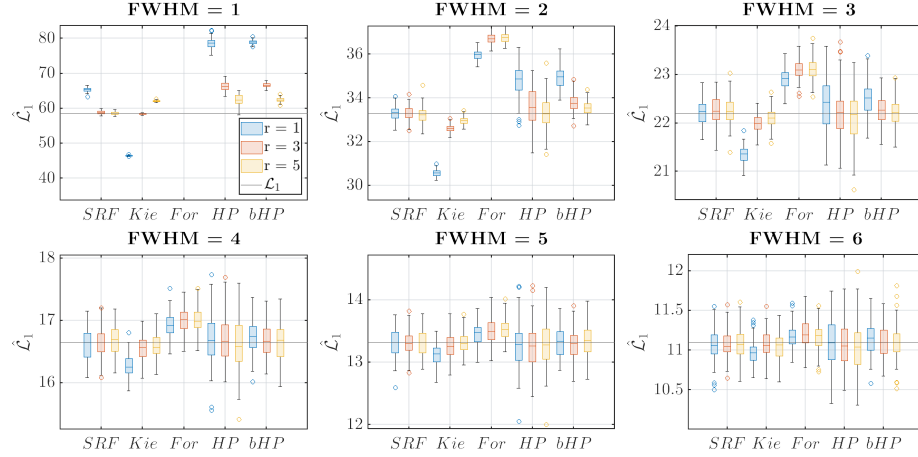
A.3 LKC estimation for $D = 3$

Figure 15:   3D Simulation results of estimation of the LKCs of the SuRFs derived from the stationary box example. The FWHM is fixed to 3.

Figure 16: 3D Simulation results of estimation of the LKCs of the two SuRFs derived from the non-stationary sphere example. The FWHM is fixed to 3. Note that the theoretical value for $\mathcal{L}_1$ is the theoretical value for the locally stationary $\mathcal{L}_1$. The true value for $\mathcal{L}_1$ is currently infeasible to obtain.

Figure 17: 3D Simulation results of estimation of the LKCs of SuRFs derived from the almost stationary box example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
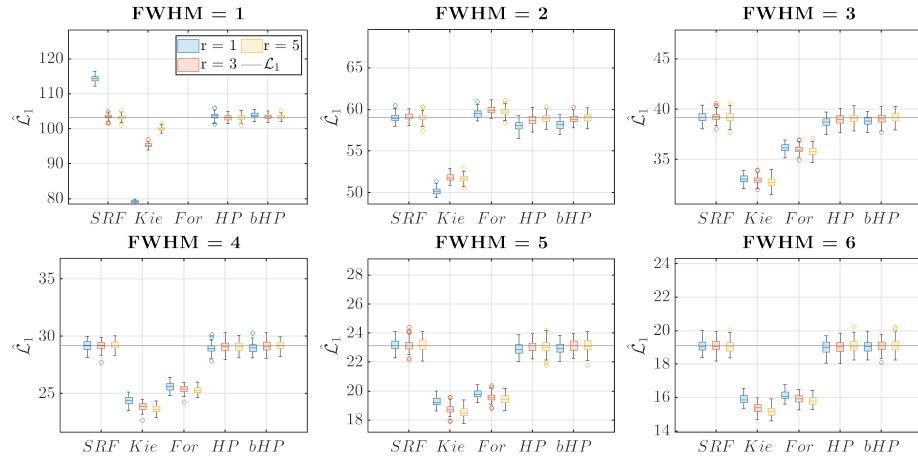
Figure 18: 3D Simulation results of estimation of the LKCs of SuRFs derived from the almost stationary box example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.

Figure 19: 3D Simulation results of estimation of the LKCs of SuRFs derived from the almost-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
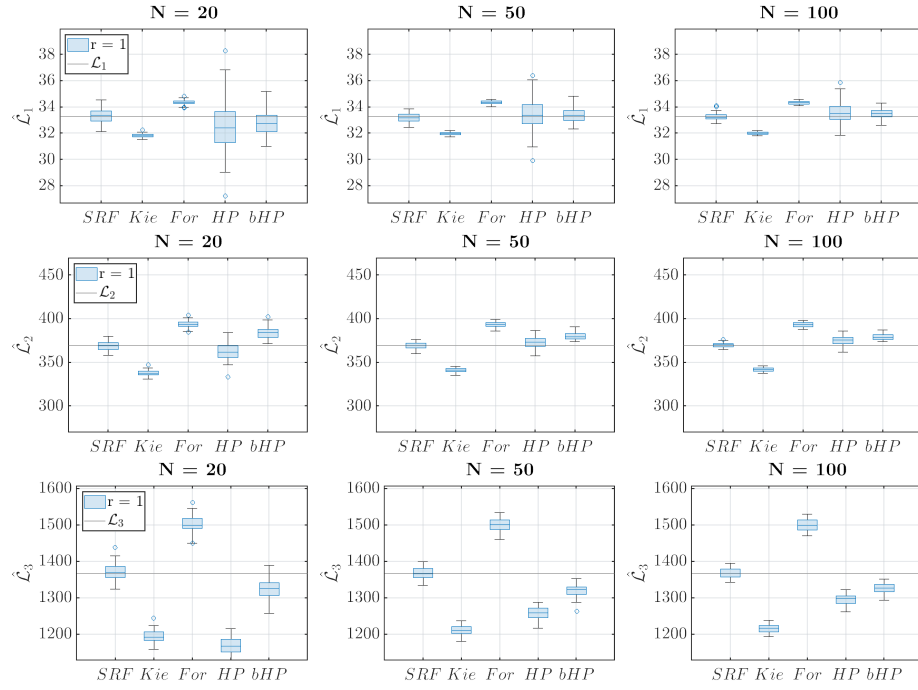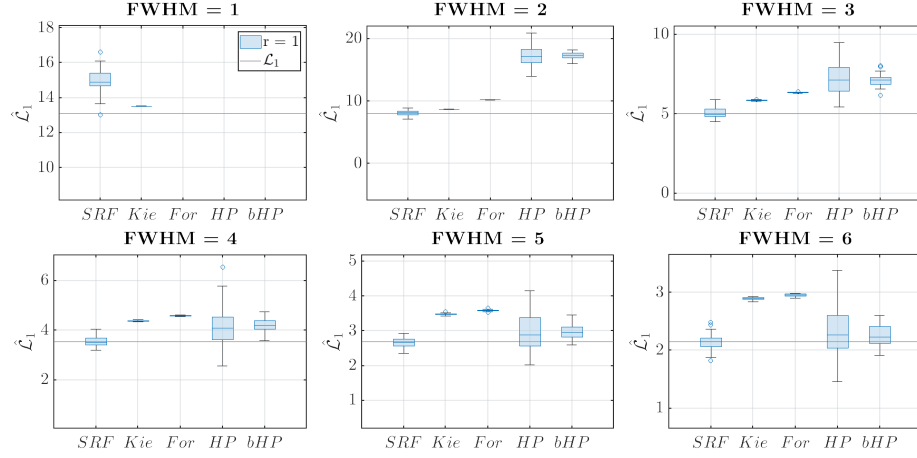
Figure 20: 3D Simulation results of estimation of the LKCs of SuRFs derived from the non-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed. Note that the theoretical value for $\mathcal{L}_1$ is the theoretical value for the locally stationary $\mathcal{L}_1$. The true value for $\mathcal{L}_1$ is currently infeasible to obtain.
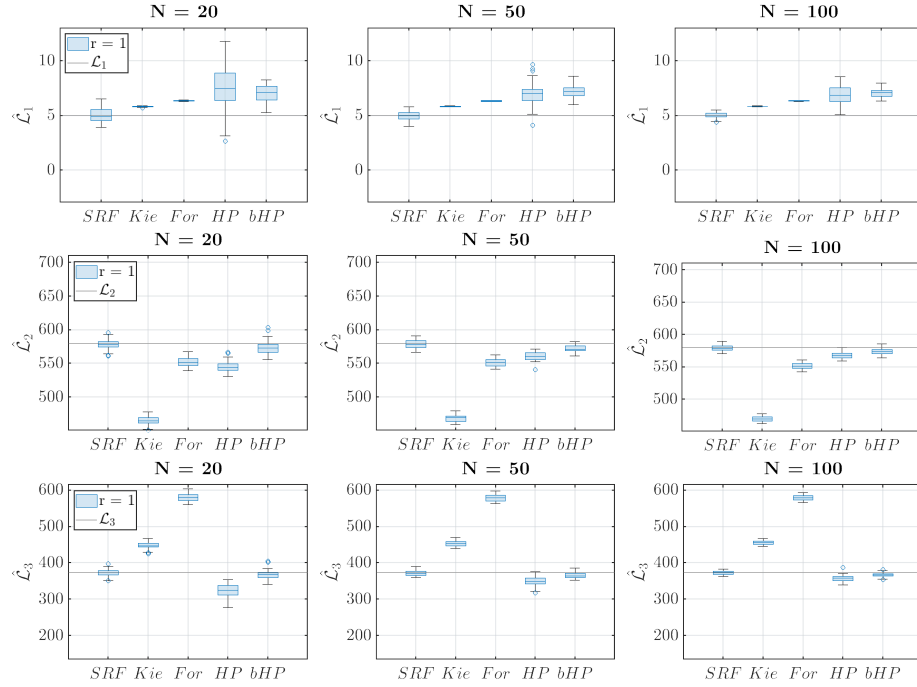
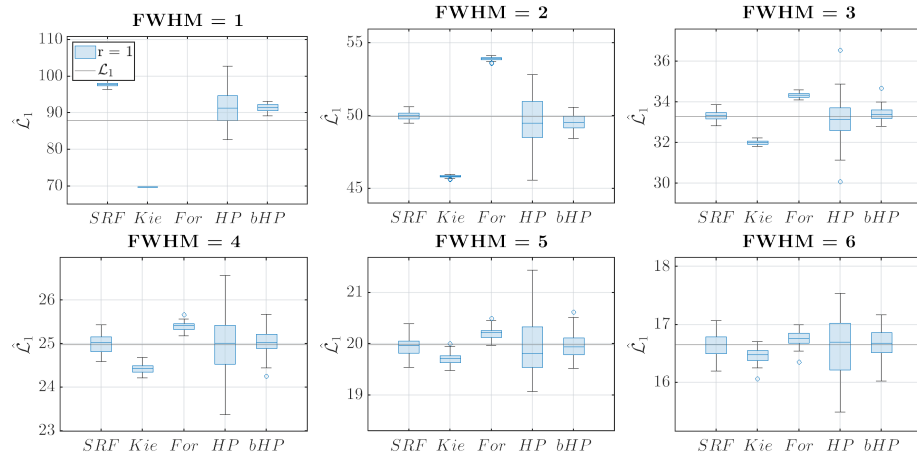

Figure 21: 3D Simulation results of estimation of the LKCs of SuRFs derived from the non-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
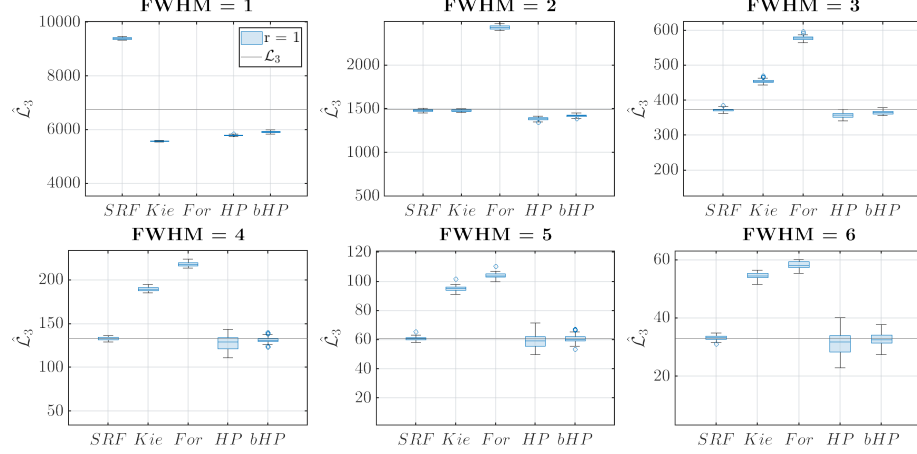
Figure 22: 3D Simulation results of estimation of the LKCs of SuRFs derived from the non-stationary sphere example. The results show the dependence of the LKC estimation on the FWHM used in the smoothing kernel. The $N = 100$ is fixed.
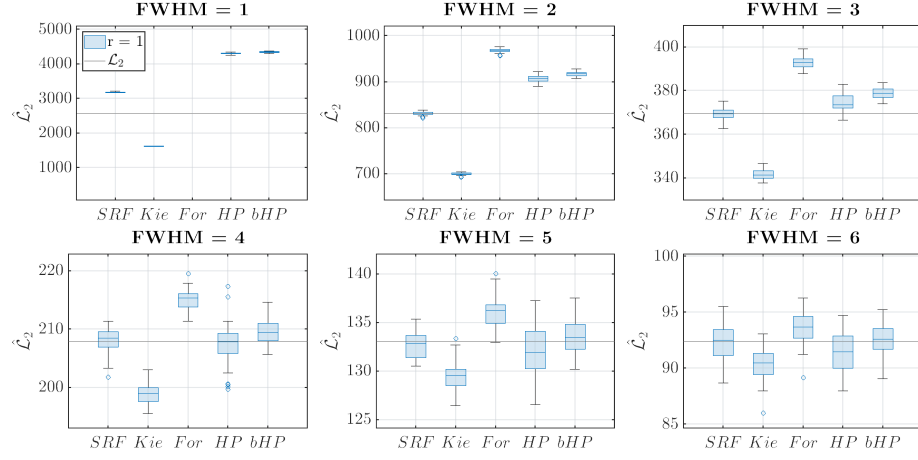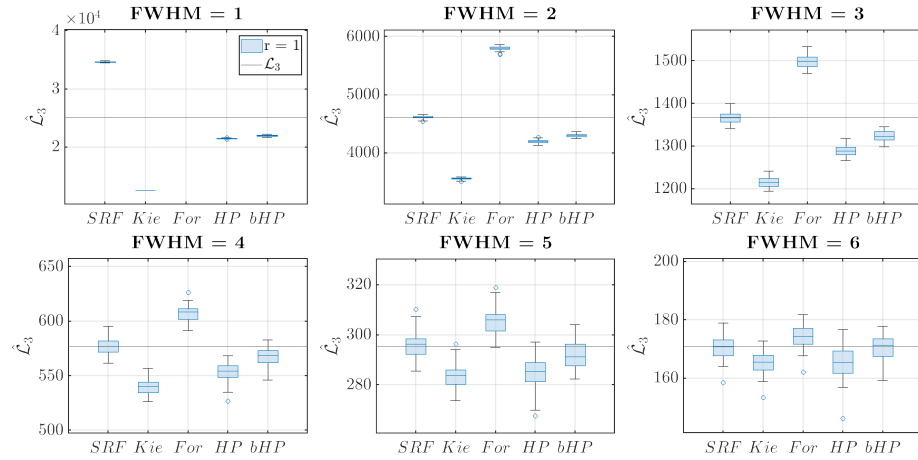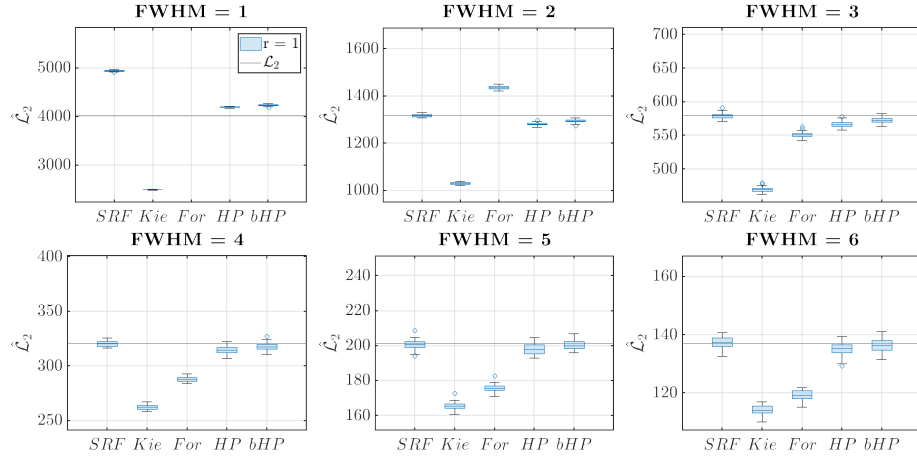
A.4 Comparison of the LKCs of the Almost Stationary Box Example and its Stationary Counterpart

Here we give an example of approximating the theoretical LKCs (up to an arbitrary resolution increase to approximate the integrals) of a SuRF derived from a random vector $(X(v) : v \in V)$ satisfying $X(v_1)$ being independent of $X(v_2)$ for all $v_1 \neq v_2 \in \mathcal{V}$. This can be achieved using the function $LKC\_wncfield\_theory()$ from the RFTtoolbox. We apply it to demonstrate that the theoretical LKCs from the almost stationary box example used in our simulations which has covariance function $\tilde{\mathfrak{c}}_f$ depending on $f > 0$ has essentially the same LKCs for f larger than ≈ 2.5 as the zero-mean Gaussian field with covariance function $\mathfrak{c}_f(x, y) \exp\big(-4\log(2)|x - y|^2/f^2\big)$. The LKCs of the latter random field over a domain $S \subset \mathbb{R}^D$, $D \in \{1, 2, 3\}$, are given by

$$D = 1: \quad \mathcal{L}_1 = \frac{\text{vol}(S)}{\sqrt{4\log(2)}f}$$

$$D = 2: \quad \mathcal{L}_1 = \frac{\text{vol}(\partial S)}{2\sqrt{4\log(2)}f}, \quad \mathcal{L}_2 = \frac{\text{vol}(S)}{4\log(2)f^2}$$

$$D = 3: \quad \mathcal{L}_1 = \frac{2\text{Diameter}(S)}{\sqrt{4\log(2)}f}, \quad \mathcal{L}_2 = \frac{\text{vol}(\partial S)}{2 \cdot 4\log(2)f^2}, \quad \mathcal{L}_3 = \frac{\text{vol}(S)}{f^3\big(4\log(2)\big)^{3/2}},$$

compare for example Telschow et al. and in particular Table 2 from Worsley et al. [2004] for $D = 3$. To approximate the theoretical LKCs we use an added resolution of 11 for $D \in \{1, 2\}$ and for the sake of computation time use an added resolution of 7 for $D = 3$. The results are presented in Tables 2-4.

$D=1$	f	1	1.5	2	2.5	3	3.5	4
$\mathcal{L}_1$	$\tilde{\mathfrak{c}}_f$	146.52	110.41	83.25	66.60	55.50	47.57	41.63
	$\mathfrak{c}_f$	166.51	111.01	83.26	66.60	55.50	47.57	41.63

Table 2: Comparison of the theoretical LKCs for different smoothing bandwidths between the almost stationary box example for $D = 1$, which has the covariance function $\tilde{\mathfrak{c}}_f$, and the zero-mean stationary Gaussian field having covariance $\tilde{\mathfrak{c}}_f$ over the domains specified in Section 4.1 of the main manuscript.

$D=2$	f	1	1.5	2	2.5	3	3.5	4
$\mathcal{L}_1$	$\tilde{\mathfrak{c}}_f$	58.61	44.16	33.30	26.64	22.20	19.03	16.65
	$\mathfrak{c}_f$	66.60	44.40	33.30	26.64	22.20	19.03	16.65
$\mathcal{L}_2$	$\tilde{\mathfrak{c}}_f$	858.72	487.59	277.24	177.45	123.23	90.53	69.31
	$\mathfrak{c}_f$	1109.00	492.90	277.26	177.45	123.23	90.53	69.31

Table 3: Comparison of the theoretical LKCs for different smoothing bandwidths between the almost stationary box example for $D = 2$, which has the covariance function $\tilde{\mathfrak{c}}_f$, and the zero-mean stationary Gaussian field having covariance $\tilde{\mathfrak{c}}_f$ over the domains specified in Section 4.1 of the main manuscript.

$D=3$	f	1	1.5	2	2.5	3	3.5	4
$\mathcal{L}_1$	$\tilde{\mathfrak{c}}_f$	87.91	66.24	49.95	39.96	33.30	28.54	24.98
	$\mathfrak{c}_f$	99.91	66.60	49.95	39.96	33.30	28.54	24.98
$\mathcal{L}_2$	$\tilde{\mathfrak{c}}_f$	2576.13	1462.77	831.72	532.34	369.68	271.60	207.94
	$\mathfrak{c}_f$	3327.11	1478.71	831.78	532.34	369.68	271.6	207.94
$\mathcal{L}_3$	$\tilde{\mathfrak{c}}_f$	25163.37	10766.66	4616.20	2363.73	1367.90	861.42	577.08
	$\mathfrak{c}_f$	36933.30	10943.20	4616.66	2363.73	1367.90	861.42	577.08

Table 4: Comparison of the theoretical LKCs for different smoothing bandwidths between the almost stationary box example for $D = 3$, which has the covariance function $\tilde{\mathfrak{c}}_f$, and the zero-mean stationary Gaussian field having covariance $\tilde{\mathfrak{c}}_f$ over the domains specified in Section 4.1 of the main manuscript.

B LKCs induced by a normalized field

B.1 Induced Riemannian metric of a normalized random field

The most important quantity for the Gaussian Kinematic formula is the Riemannian metric induced by a random field. It is the backbone of the GKF for Gaussian related fields developed in Taylor et al. [2006].

In this section f denotes a zero-mean random field with almost surely continuously, differentiable sample paths over the domain $\overline{\mathcal{M}}$ and we call the random field $f/\sqrt{\mathrm{Var}[f]}$ the normalized field derived from f. Recall that a vector field $V \in \mathcal{T}\overline{\mathcal{M}}$ can be interpreted as a first order differential operator, i.e., for all $h \in C^1(\overline{\mathcal{M}})$ the expression $Vh : \overline{\mathcal{M}} \to \mathbb{R}, \ s \mapsto (Vh)(s)$ defines a function in $C(\overline{\mathcal{M}})$. This can be made precise by taking a local chart $\varphi : \overline{\mathcal{M}} \supseteq \mathcal{U} \to \varphi(\mathcal{U}) = \mathcal{W} \subset \mathbb{R}^D$ of $\overline{\mathcal{M}}$ with inverse $\psi = \varphi^{-1}$. The vector field V in local coordinates $(\mathcal{U}, \varphi)$ can be written as

$$V = \sum_{d=1}^{D} V_d \partial_d, \quad V_1, \ldots, V_D \in C(\mathcal{W}) \tag{25}$$

where ∂_d is the vector field on $\mathcal{W}$ defined by

$$\partial_d(h \circ \psi)(x) = \frac{\partial h \circ \psi}{\partial x_d}(x) = \partial_d^x h(\psi(x)), \quad d = 1, ..., D \tag{26}$$

where $h \in C(\mathcal{U})$ and $x \in \varphi(\mathcal{U})$.

Definition 5. Let $V, W \in \mathcal{T}\overline{\mathcal{M}}$ be differentiable vector fields and f unit-variance random field over a manifold $\overline{\mathcal{M}}$ with almost surely differentiable sample paths. Then

$$\bar{\Lambda}_s(V, W) = \mathrm{Cov}\left[Vf(s), \, Wf(s)\right], \quad s \in \overline{\mathcal{M}}, \tag{27}$$

is called the *induced Riemannian metric of f* on $\overline{\mathcal{M}}$. In local coordinates at a point $z \in \mathcal{W}$ it is represented by

$$\begin{aligned}
\bar{\Lambda}_z(V, W) &= \sum_{d,d'=1}^{D} V_d W_{d'} \partial_d^x \partial_{d'}^y \mathrm{Cov}\left[f(\psi(x)), \, f(\psi(y))\right]\Big|_{(x,y)=(z,z)} \\
&= \sum_{d,d'=1}^{D} V_d W_{d'} \mathrm{Cov}\left[\partial_d^x f(\psi(x))\big|_{x=z}, \, \partial_{d'}^y f(\psi(y))\big|_{y=z}\right].
\end{aligned} \tag{28}$$

Here $V_1, \ldots, V_D, W_1, \ldots, W_D \in C(\mathcal{W})$ are the coordinate coefficients representing the vector fields V and W.

Remark 9. Assumption **(G2)** from the main manuscript ensures that this is a Riemannian metric on $\overline{\mathcal{M}}$.

Since the vector fields ∂_d, $d = 1, \ldots, D$, form a basis of $\mathcal{T}\mathcal{U}$ the Riemannian metric induced by the random field f can be written in local coordinates as the $D \times D$ matrix having components

$$\bar{\Lambda}_{dd'}(z) = \mathrm{Cov}\left[\partial_d^x f(\psi(x))\big|_{x=z}, \, \partial_{d'}^y f(\psi(y))\big|_{y=z}\right] \tag{29}$$

For simplicity in what follows, we establish the following alternative notations suppressing the dependencies on f:

$$\mathrm{Cov}\left[\partial_d^x f(x), \, f(y)\right] = \langle \partial_d^x, \, 1_y \rangle, \quad \mathrm{Cov}\left[\partial_d^x f(x), \, \partial_{d'}^y f(y)\right] = \langle \partial_d^x, \, \partial_{d'}^y \rangle. \tag{30}$$

Similarly, $\|1_x\|^2 = \mathrm{Var}[f(x)]$ and $\|\partial_d^x\|^2 = \mathrm{Var}[\partial_d^x f(x)]$.

Theorem 5. *The Riemannian metric on $\overline{\mathcal{M}}$ induced by a normalized random field is given in local coordinates by*

$$\bar{\Lambda}_{dd'}(x) = \frac{\langle \partial_d^x, \partial_{d'}^x \rangle}{\|1_x\|} - \frac{\langle \partial_d^x, 1_x \rangle \langle \partial_{d'}^x, 1_x \rangle}{\|1_x\|^2} . \tag{31}$$

The computation of Lipshitz-Killing curvatures requires the shape operator and the Riemannian curvature. Therefore, the next theorem computes the Christoffel symbols of the first kind, which can be used to express these quantities in local coordinates. This is because most fundamental geometric quantities such as the covariant derivative and the Riemannian curvature are functions of the Christoffel symbols and their derivatives.

Theorem 6. *The Christoffel symbols $\bar{\Gamma}_{kdd'}$, $k, d, d' \in \{1, \ldots, D\}$ of the first kind of the induced Riemannian metric by a normalized field are given by*

$$\begin{aligned}
\bar{\Gamma}_{kdd'}(x) = {}& \frac{\langle \partial_k^x \partial_d^x, \partial_{d'}^x \rangle}{\|1_x\|} - \frac{\langle \partial_k^x \partial_d^x, 1_x \rangle \langle \partial_{d'}^x, 1_x \rangle}{\|1_x\|^2} - \frac{\langle \partial_k^x, \partial_{d'}^x \rangle \langle \partial_d^x, 1_x \rangle}{\|1_x\|^2} \\
& - \frac{\langle \partial_k^x, 1_x \rangle \langle \partial_d^x, \partial_{d'}^x \rangle}{\|1_x\|^2} + 2\frac{\langle \partial_k^x, 1_x \rangle \langle \partial_d^x, f_x \rangle \langle \partial_{d'}^x, 1_x \rangle}{\|1_x\|^3} .
\end{aligned} \tag{32}$$

In terms of Christoffel symbols the covariant derivative $\bar{\nabla}$ on $\overline{\mathcal{M}}$ is expressed in the local chart φ by

$$\bar{\nabla}_{\partial_d} \partial_{d'} = \sum_{h=1}^{D} \sum_{h'=1}^{D} \bar{\Lambda}^{hh'} \bar{\Gamma}_{dd'h'} \partial_h = \bar{\Lambda}^{-1} (\bar{\Gamma}_{dd'1}, ..., \bar{\Gamma}_{dd'D})^T .$$

Here $\bar{\Lambda}^{dd'}$ denotes the (d, d')-entry of the inverse of $\bar{\Lambda}$ in the coordinates $(\mathcal{U}, \varphi)$. This formula can be used to extend the covariant derivative to any vector field, i.e.,

$$\bar{\nabla}_V W = \sum_{d=1}^{D} \sum_{d'=1}^{D} V_d \bar{\nabla}_{\partial_d} W_{d'} \partial_{d'} = \sum_{d=1}^{D} \sum_{d'=1}^{D} V_d \Big(\partial_d(W_{d'}) \partial_{d'} + W_{d'} \bar{\nabla}_{\partial_d} \partial_{d'} \Big) . \tag{33}$$

In particular, if all $\Gamma_{dd'd''} = 0$, $d, d', d'' \in \{1, \ldots, D\}$, then

$$\bar{\nabla}_V W = \sum_{d=1}^{D} \sum_{d'=1}^{D} V_d \partial_d(W_{d'}) \partial_{d'} . \tag{34}$$

This happens, if $\bar{\Lambda}(s) = \bar{\Lambda}(s')$ for all $s, s' \in \overline{\mathcal{M}}$.

The last geometric quantity required to compute LKCs is the Riemannian curvature tensor $\bar{R}$. The curvature tensor in local coordinates can be written as the tensor having entries

$$\bar{R}^{\varphi}_{lkdd'} = \partial_l \bar{\Gamma}_{kdd'} - \partial_k \Gamma_{ldd'} + \sum_{m,n=1}^{D} \big(\bar{\Gamma}_{ldm} \bar{\Lambda}^{mn} \bar{\Gamma}_{kd'n} - \bar{\Gamma}_{kdm} \bar{\Lambda}^{mn} \bar{\Gamma}_{ld'n} \big) \tag{35}$$

for $i, j, k, l \in \{1, \ldots, D\}$, compare [Adler and Taylor, 2007, eq. (7.7.4)]. By now, we derived almost all quantities to state the Riemannian curvature tensor in local coordinates. The missing quantities are the derivatives of the Christoffel symbols which can be found in the next Lemma.

Lemma 1. *The difference of the derivatives of Christoffel symbols in the Riemannian curvature tensor* (35) *of a normalized field can be expressed as:*

$$
\partial_l \bar{\Gamma}_{kdd'}(x) - \partial_k \bar{\Gamma}_{ldd'}(x)
$$
$$
= \frac{\langle \partial_k^x \partial_d^x, \partial_l^x \partial_{d'}^x \rangle - \langle \partial_l^x \partial_d^x, \partial_k^x \partial_{d'}^x \rangle}{\|1_x\|}
$$
$$
- \|1_x\|^{-2} \Big(\quad \langle \partial_l^x, 1_x \rangle \langle \partial_k^x \partial_d^x, \partial_{d'}^x \rangle - \langle \partial_k^x, 1_x \rangle \langle \partial_l^x \partial_d^x, \partial_{d'}^x \rangle
$$
$$
+ \langle \partial_k^x \partial_d^x, \partial_l^x \rangle \langle \partial_{d'}^x, 1_x \rangle - \langle \partial_l^x \partial_d^x, \partial_k^x \rangle \langle \partial_{d'}^x, 1_x \rangle
$$
$$
+ \langle \partial_k^x \partial_d^x, 1_x \rangle \langle \partial_l^x \partial_{d'}^x, 1_x \rangle - \langle \partial_l^x \partial_d^x, 1_x \rangle \langle \partial_k^x \partial_{d'}^x, 1_x \rangle
$$
$$
+ \langle \partial_k^x, 1_x \rangle \langle \partial_l^x \partial_{d'}^x, \partial_d^x \rangle - \langle \partial_l^x, 1_x \rangle \langle \partial_k^x \partial_{d'}^x, \partial_d^x \rangle
$$
$$
+ \langle \partial_l^x, \partial_d^x \rangle \langle \partial_k^x, \partial_{d'}^x \rangle - \langle \partial_k^x, \partial_d^x \rangle \langle \partial_l^x, \partial_{d'}^x \rangle
$$
$$
+ \langle \partial_l^x, \partial_d^x \rangle \langle \partial_k^x, \partial_{d'}^x \rangle - \langle \partial_k^x, \partial_d^x \rangle \langle \partial_l^x, \partial_{d'}^x \rangle
$$
$$
+ \langle \partial_k^x, \partial_l^x \partial_{d'}^x \rangle \langle \partial_d^x, 1_x \rangle - \langle \partial_l^x, \partial_k^x \partial_{d'}^x \rangle \langle \partial_d^x, 1_x \rangle \Big)
$$
$$
+ 2\|1_x\|^{-3} \Big(\quad \langle \partial_l^x, 1_x \rangle \langle \partial_k^x \partial_d^x, 1_x \rangle \langle \partial_{d'}^x, 1_x \rangle - \langle \partial_k^x, 1_x \rangle \langle \partial_l^x \partial_d^x, 1_x \rangle \langle \partial_{d'}^x, 1_x \rangle
$$
$$
+ \langle \partial_k^x, 1_x \rangle \langle \partial_d^x, \partial_l^x \rangle \langle \partial_{d'}^x, 1_x \rangle - \langle \partial_l^x, 1_x \rangle \langle \partial_d^x, \partial_k^x \rangle \langle \partial_{d'}^x, 1_x \rangle
$$
$$
+ \langle \partial_k^x, 1_x \rangle \langle \partial_d^x, 1_x \rangle \langle \partial_l^x \partial_{d'}^x, 1_x \rangle - \langle \partial_l^x, 1_x \rangle \langle \partial_d^x, \rangle \langle \partial_k^x \partial_{d'}^x, 1_x \rangle
$$
$$
+ \langle \partial_l^x, 1_x \rangle \langle \partial_d^x, 1_x \rangle \langle \partial_k^x, \partial_{d'}^x \rangle - \langle \partial_k^x, 1_x \rangle \langle \partial_d^x, 1_x \rangle \langle \partial_l^x, \partial_{d'}^x \rangle \Big)
$$

Proof. Simple, but lengthy computation. □

To compute the first LKC $\mathcal{L}_1$ for a 3-dimensional manifold the trace of the Riemannian tensor is needed which in the coordinates $(\mathcal{U}, \varphi)$ can be expressed in terms of the entries of the inverse of the square root of the Riemannian metric and the Riemannian tensor as

$$
\operatorname{tr}(\bar{R}) = \sum_{i,j,k,l=1}^{3} \bar{R}_{ijkl}^{\varphi} \Bigg(\frac{\bar{\Lambda}_{1i}^{-1/2} \bar{\Lambda}_{1j}^{-1/2} \bar{\Lambda}_{1k}^{-1/2} \bar{\Lambda}_{1l}^{-1/2}}{2}
$$
$$
+ \frac{\bar{\Lambda}_{2i}^{-1/2} \bar{\Lambda}_{2j}^{-1/2} \bar{\Lambda}_{2k}^{-1/2} \bar{\Lambda}_{2l}^{-1/2}}{2}
$$
$$
+ \frac{\bar{\Lambda}_{3i}^{-1/2} \bar{\Lambda}_{3j}^{-1/2} \bar{\Lambda}_{3k}^{-1/2} \bar{\Lambda}_{3l}^{-1/2}}{2} \tag{36}
$$
$$
+ \bar{\Lambda}_{1i}^{-1/2} \bar{\Lambda}_{2j}^{-1/2} \bar{\Lambda}_{1k}^{-1/2} \bar{\Lambda}_{2l}^{-1/2}
$$
$$
+ \bar{\Lambda}_{1i}^{-1/2} \bar{\Lambda}_{3j}^{-1/2} \bar{\Lambda}_{1k}^{-1/2} \bar{\Lambda}_{3l}^{-1/2}
$$
$$
+ \bar{\Lambda}_{2i}^{-1/2} \bar{\Lambda}_{3j}^{-1/2} \bar{\Lambda}_{2k}^{-1/2} \bar{\Lambda}_{3l}^{-1/2} \Bigg).
$$

B.2   Definition of LKCs

Lipshits Killing Curvatures $\mathcal{L}_1, \ldots, \mathcal{L}_D$ are the intrinsic volumes of a compact D-dimensional Whitney stratified manifold $(\mathcal{M}, \Lambda)$ isometrically embedded into $(\overline{\mathcal{M}}, \bar{\Lambda})$. Here Λ and $\bar{\Lambda}$ denote Riemannian metrics of $\mathcal{M}$ and $\overline{\mathcal{M}}$. They are related by $\bar{\Lambda} \circ \iota = \Lambda$ where $\iota : \mathcal{M} \to \overline{\mathcal{M}}$ is the embedding and hence $\mathcal{M}$ is isometrically embedded into $\overline{\mathcal{M}}$. In this section we make the

formula for the LKCs [Adler and Taylor, 2007, Definition 10.7.2] for the LKCs of an at most 3-dimensional Whitney stratified manifold more explicit. From the definition of the LKCs it is easy to deduce

$$\mathcal{L}_{D-1} = \mathrm{vol}_{D-1}\big(\partial_{D-1}\mathcal{M}\big) \quad \text{and} \quad \mathcal{L}_D = \mathrm{vol}_D\big(\partial_D\mathcal{M}\big). \tag{37}$$

Here $\partial_d\mathcal{M}$ denotes the d-dimensional stratum of $\mathcal{M}$ and the volume is the volume of the Riemannian manifold $(\partial_d\mathcal{M}, \Lambda|_{\partial_d\mathcal{M}})$ where $\Lambda|_{\partial_d\mathcal{M}}$ is the restriction of Λ to $\partial_d\mathcal{M}$.

For a 3-dimensional manifold it remains to compute $\mathcal{L}_1$, which we derive from [Adler and Taylor, 2007, Definition 10.7.2] and some geometric computations in the next proposition.

Proposition 8. *Let $(\overline{\mathcal{M}}, \bar{\Lambda})$ be a closed Riemannian manifold of dimension 3 and $\mathcal{M} \subset \overline{\mathcal{M}}$ be a compact Whitney stratified manifold of dimension 3 isometrically embedded into $\overline{\mathcal{M}}$. Then*

$$\begin{aligned}
\mathcal{L}_1 = \;& \frac{1}{2\pi} \int_{\partial_1\mathcal{M}} \int_{\mathbb{S}\left(\mathcal{T}_s\partial_1\mathcal{M}^\perp\right)} \alpha(v)\mathcal{H}_1(dv)\mathcal{H}_1(ds) \\
& + \frac{1}{2\pi} \int_{\partial_2\mathcal{M}} \bar{\Lambda}\Big(\bar{\nabla}_{U_s}U_s + \bar{\nabla}_{V_s}V_s, N_s\Big)\mathcal{H}_2(ds) \\
& - \frac{1}{2\pi} \int_{\partial_3\mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s\partial_3\mathcal{M}}(\bar{R})\mathcal{H}_3(ds)
\end{aligned} \tag{38}$$

Here $\mathcal{H}_1(dv)$ is the volume form induced on the sphere

$$\mathbb{S}\big(\mathcal{T}_s\partial_1\mathcal{M}^\perp\big) = \big\{v \in \mathcal{T}_s\overline{\mathcal{M}} \mid \bar{\Lambda}(v, v) = 1 \;\wedge\; \bar{\Lambda}(v, w) = 0 \text{ for } w \in \mathcal{T}_s\partial_1\mathcal{M}\big\}$$

by $\bar{\Lambda}$ and $\mathcal{H}_d(ds)$ the volume form of $\partial_d\mathcal{M}$. Moreover, $\alpha(v)$ denotes the normal Morse index given in [Adler and Taylor, 2007, Scn 9.2.1] and U, V, N is a piecewise differentiable vector field on $\partial_2\mathcal{M}$ such that U_s, V_s form an orthonormal basis for $\mathcal{T}_s\partial_2\mathcal{M}$ for all $s \in \partial_2\mathcal{M}$ and N is outward pointing normal vector field.

Remark 10. In the special case that the metric $\bar{\Lambda}$ is constant, it holds that

$$\begin{aligned}
\mathcal{L}_1 = \;& \frac{1}{2\pi} \int_{\partial_1\mathcal{M}} \int_{\mathbb{S}\left(\mathcal{T}_s\partial_1\mathcal{M}^\perp\right)} \alpha(v)\mathcal{H}_1(dv)\mathcal{H}_1(ds) \\
& + \frac{1}{2\pi} \int_{\partial_2\mathcal{M}} \bar{\Lambda}\Big(\bar{\nabla}_{U_s}U_s + \bar{\nabla}_{V_s}V_s, N_s\Big)\mathcal{H}_2(ds),
\end{aligned}$$

since the curvature tensor $\bar{R}$ vanishes. If $\mathcal{M} = \mathcal{M}_\mathcal{V}$ is a voxel manifold than even

$$\mathcal{L}_1 = \frac{1}{2\pi} \int_{\partial_1\mathcal{M}} \int_{\mathbb{S}\left(\mathcal{T}_s\partial_1\mathcal{M}^\perp\right)} \alpha(v)\mathcal{H}_1(dv)\mathcal{H}_1(ds),$$

as $\nabla_{U_s}U_s = \nabla_{V_s}V_s = 0$ on $\partial_2\mathcal{M}$ by (34) and $\bar{\Lambda}$ being constant.

C   Proofs of Results in the Appendix

C.1   Proof of Theorem 5

Proof. Interchanging expectation and derivatives yields

$$
\begin{aligned}
\bar{\Lambda}_{dd'}(z) &= \left. \partial_d^x \partial_{d'}^y \frac{\mathrm{Cov}\,[f_x,\, f_y]}{\|1_x\|\|1_y\|} \right|_{(x,y)=(z,z)} \\
&= \frac{\langle \partial_d^x,\, \partial_{d'}^y \rangle}{\|1_x\|\|1_y\|} - \frac{\langle \partial_d^x,\, 1_y \rangle \langle 1_y,\, \partial_{d'}^y \rangle}{\|1_x\|\|1_y\|^3} \\
&\quad \left. - \frac{\langle \partial_d^x,\, 1_x \rangle \langle 1_x,\, \partial_{d'}^y \rangle}{\|1_x\|^3 \|1_y\|} + \frac{\langle 1_x,\, 1_y \rangle \langle \partial_d^x,\, 1_x \rangle \langle 1_y,\, \partial_{d'}^y \rangle}{\|1_x\|^3 \|1_y\|^3} \right|_{(x,y)=(z,z)} \\
&= \frac{\langle \partial_d^z,\, \partial_{d'}^z \rangle}{\|1_z\|^2} - \frac{\langle \partial_d^z,\, 1_z \rangle \langle 1_z,\, \partial_{d'}^z \rangle}{\|1_z\|^4}
\end{aligned}
\tag{39}
$$

□

C.2   Proof of Theorem 6

Proof. From the definition of the Christoffel symbols one can derive that

$$
\Gamma_{kdd'}(z) = \left. \frac{\partial^3 \tilde{\mathfrak{c}}(x,y)}{\partial_{x_k} \partial_{x_d} \partial_{y_{d'}}} \right|_{(x,y)=(z,z)},
\tag{40}
$$

see also [Adler and Taylor, 2007, eq. 12.2.17]. Here $\tilde{\mathfrak{c}}(x,y) = \frac{\mathrm{Cov}[f_x, f_y]}{\sqrt{\mathrm{Cov}[f_x, f_x]}\sqrt{\mathrm{Cov}[f_y, f_y]}}$. Thus, by simply taking another derivate of (39), we obtain

$$
\begin{aligned}
\Gamma_{kdd'}(z) &= \partial_k^x \partial_d^x \partial_{d'}^y \frac{\mathrm{Cov}\left[f_x, f_y\right]}{\sqrt{\|1_x\|\|1_y\|}}\Bigg|_{(x,y)=(z,z)} \\
&= \frac{\langle \partial_k^x \partial_d^x, \partial_{d'}^x \rangle}{\|1_x\|^2} - \frac{\langle \partial_k^x, 1_x \rangle \langle \partial_d^x, \partial_{d'}^x \rangle}{\|1_x\|^4} - \partial_k^x \frac{\langle \partial_d^x, 1_y \rangle \langle 1_y, \partial_{d'}^y \rangle}{\|1_x\|\|1_y\|^3} \\
&\quad - \partial_k^x \frac{\langle \partial_d^x, 1_x \rangle \langle 1_x, \partial_{d'}^y \rangle}{\|1_x\|^3 \|1_y\|} + \partial_k^x \frac{\langle 1_x, 1_y \rangle \langle \partial_d^x, 1_x \rangle \langle 1_y, \partial_{d'}^y \rangle}{\|1_x\|^3 \|1_y\|^3}\Bigg|_{(x,y)=(z,z)} \\
&= \frac{\langle \partial_k^z \partial_d^z, \partial_{d'}^z \rangle}{\|1_z\|^2} - \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, \partial_{d'}^z \rangle}{\|1_z\|^4} - \frac{\langle \partial_k^z \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^4} \\
&\quad + \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^6} - \frac{\langle \partial_k^z \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle + \langle \partial_d^z, \partial_k^z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^4} \\
&\quad - \frac{\langle \partial_d^z, 1_z \rangle \langle \partial_k^z, \partial_{d'}^z \rangle}{\|1_z\|^4} + 3 \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^6} \\
&\quad + \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^6} + \frac{\langle \partial_k^z \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^4} + \frac{\langle \partial_d^z, \partial_k^z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^4} \\
&\quad - 3 \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^6} \\
&= \frac{\langle \partial_k^z \partial_d^z, \partial_{d'}^z \rangle}{\|1_z\|^2} - \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, \partial_{d'}^z \rangle}{\|1_z\|^4} - \frac{\langle \partial_k^z \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^4} \\
&\quad - \frac{\langle \partial_d^z, 1_z \rangle \langle \partial_k^z, \partial_{d'}^z \rangle}{\|1_z\|^4} + 2 \frac{\langle \partial_k^z, 1_z \rangle \langle \partial_d^z, 1_z \rangle \langle 1_z, \partial_{d'}^z \rangle}{\|1_z\|^6}
\end{aligned}
$$

$\square$

C.3 Proof of Proposition 8

Proof. Using $D - d' = D'$ we have that the LKCs of a Whitney stratified manifold $\mathcal{M}$ are defined by

$$
\begin{aligned}
\mathcal{L}_d = \sum_{d'=d}^{D} \frac{1}{(2\pi)^{\frac{d'-d}{2}}} \sum_{l=0}^{\lfloor \frac{d'-d}{2} \rfloor} \frac{(-1)^l C(D', d-d'-2l)}{l!(d-d'-2l)!} \\
\times \int_{\partial_{d'}\mathcal{M}} \int_{\mathbb{S}(\mathcal{T}_s \partial_{d'}\mathcal{M}^\perp)} \mathrm{Tr}^{\mathcal{T}_s \partial_{d'}\mathcal{M}} \left(R^l S_{\nu_{D'}}^{d'-d-2l} \right) \alpha(\nu_{D'}) \mathcal{H}_{-1}(d\nu_{D'}) \mathcal{H}_{d'}(ds),
\end{aligned}
\tag{41}
$$

compare [Adler and Taylor, 2007, Definition 10.7.2]. This formula requires further explanations. The constant $C(m, i)$ is defined in [Adler and Taylor, 2007, eq. (10.5.1), p.233], i.e.,

$$
C(m, i) = \begin{cases} \frac{(2\pi)^{\frac{i}{2}}}{s_{m+i}}, & m + i > 0, \\ 1, & m = 0 \end{cases} \quad \text{with} \quad s_m = \frac{2\pi^{m/2}}{\Gamma(m/2)},
\tag{42}
$$

which implies $C(m, 0) = \Gamma(m/2)/2/\pi^{m/2}$. Moreover, $\mathcal{H}_{D-d-1}$ is the volume form on $\mathbb{S}(\mathcal{T}_s \partial_d \mathcal{M}^\perp)$ and $\mathcal{H}_d$ the volume form of $\partial_d \mathcal{M}$. R denotes the Riemannian curvature tensor of the different

strata on $\mathcal{M}$ depending on the strata the integral integrates over. In particular, note that $\bar{R} = R$ for $\partial_3 \mathcal{M}$.

From this $\mathcal{L}_1$ can be simplified as follows:

$$
\begin{aligned}
\mathcal{L}_1 &= C(3,0) \int_{\partial_1 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_1 \mathcal{M}^\perp)} \mathrm{Tr}^{\mathcal{T}_s \partial_1 \mathcal{M}}\left(R^0 S_{\nu_2}^0\right) \alpha(\nu_2) \mathcal{H}_1(d\nu_2) \mathcal{H}_1(ds) \\
&\quad + \frac{C(1,1)}{\sqrt{2\pi}} \int_{\partial_2 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_2 \mathcal{M}^\perp)} \mathrm{Tr}^{\mathcal{T}_s \partial_2 \mathcal{M}}\left(R^0 S_{\nu_1}^1\right) \alpha(\nu_1) \mathcal{H}_0(d\nu_1) \mathcal{H}_2(ds) \\
&\quad + \frac{C(0,2)}{4\pi} \int_{\partial_3 \mathcal{M}} \int_{S(\mathbb{O})} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}\left(R^0 S_{\nu_0}^2\right) \alpha(\nu_0) \mathcal{H}_{-1}(d\nu_0) \mathcal{H}_3(ds) \\
&\quad - \frac{C(0,0)}{2\pi} \int_{\partial_3 \mathcal{M}} \int_{S(\mathbb{O})} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}\left(R^1 S_{\nu_0}^0\right) \alpha(\nu_0) \mathcal{H}_{-1}(d\nu_0) \mathcal{H}_3(ds) \\
&= \frac{1}{2\pi} \int_{\partial_1 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_1 \mathcal{M}^\perp)} \alpha(\nu_2) \mathcal{H}_1(d\nu_2) \mathcal{H}_1(ds) \\
&\quad + \frac{1}{2\pi} \int_{\partial_2 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_2 \mathcal{M}^\perp)} \mathrm{Tr}^{\mathcal{T}_s \partial_2 \mathcal{M}}\left(1 \cdot S_{\nu_1}^1\right) \alpha(\nu_1) \mathcal{H}_0(d\nu_1) \mathcal{H}_2(ds) \\
&\quad + \frac{1}{4\pi} \int_{\partial_3 \mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}\left(1 \cdot S_{\mathbb{O}}^2\right) \mathcal{H}_3(ds) \\
&\quad - \frac{1}{2\pi} \int_{\partial_3 \mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}\left(R^1 S_{\mathbb{O}}^0\right) \mathcal{H}_3(ds) \\
&= \frac{1}{2\pi} \int_{\partial_1 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_1 \mathcal{M}^\perp)} \alpha(\nu_2) \mathcal{H}_1(d\nu_2) \mathcal{H}_1(ds) \\
&\quad + \frac{1}{2\pi} \int_{\partial_2 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_2 \mathcal{M}^\perp)} \mathrm{Tr}^{\mathcal{T}_s \partial_2 \mathcal{M}}\left(S_{\nu_1}^1\right) \alpha(\nu_1) \mathcal{H}_0(d\nu_1) \mathcal{H}_2(ds) \\
&\quad - \frac{1}{2\pi} \int_{\partial_3 \mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}(\bar{R}) \mathcal{H}_3(ds) \\
&= \frac{1}{2\pi} \int_{\partial_1 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_1 \mathcal{M}^\perp)} \alpha(\nu_2) \mathcal{H}_1(d\nu_2) \mathcal{H}_1(ds) \\
&\quad + \frac{1}{2\pi} \int_{\partial_2 \mathcal{M}} g\left(\bar{\nabla}_{e_1(s)} e_1(s), \tilde{\nu}(s)\right) + g\left(\bar{\nabla}_{e_2(s)} e_2(s), \tilde{\nu}(s)\right) \mathcal{H}_2(ds) \\
&\quad - \frac{1}{2\pi} \int_{\partial_3 \mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}(\bar{R}) \mathcal{H}_3(ds) \\
&= \frac{1}{2\pi} \int_{\partial_1 \mathcal{M}} \int_{S(\mathcal{T}_s \partial_1 \mathcal{M}^\perp)} \alpha(\nu_2) \mathcal{H}_1(d\nu_2) \mathcal{H}_1(ds) \\
&\quad + \frac{1}{2\pi} \int_{\partial_2 \mathcal{M}} g\left(\bar{\nabla}_{e_1(s)} e_1(s) + \bar{\nabla}_{e_2(s)} e_2(s), \tilde{\nu}(s)\right) \mathcal{H}_2(ds) \\
&\quad - \frac{1}{2\pi} \int_{\partial_3 \mathcal{M}} \mathrm{Tr}^{\mathcal{T}_s \partial_3 \mathcal{M}}(\bar{R}) \mathcal{H}_3(ds)
\end{aligned}
$$

Here $\tilde{\nu}(s)$ is the inward pointing normal at x in $\partial_2 \mathcal{M}$ and $e_1(s), e_2(s)$ an orthonormal basis of $\mathcal{T}_s \partial_2 \mathcal{M}$ and used Remark (10.5.2) [Adler and Taylor, 2007, p.233], i.e.,

$$
S_{\mathbb{O}}^j = \begin{cases} 1, & j = 0, \\ 0, & \text{otherwise} \end{cases}. \tag{43}
$$

$\square$

46

D Proofs of the Results in the Main Manuscript

D.1 Proof of Proposition 2

Proof. Define $q = p/(p-1)$ if $p > 1$ and $q = \infty$ if $p = 1$. Using the triangle inequality and Hölder's inequality yields for the charts $(\overline{U}_\alpha, \overline{\varphi}_\alpha)$, $\alpha \in \{1, \ldots, P\}$, in the atlas of $\overline{\mathcal{M}}$ covering $\mathcal{M}$, all $x, y \in \overline{\varphi}(\overline{U}_\alpha) \cap \mathcal{M}$ that

$$
\begin{aligned}
|\tilde{X}_\alpha(x) - \tilde{X}_\alpha(y)| &= \left| \sum_{v \in \mathcal{V}} \left(K\big(\overline{\varphi}_\alpha^{-1}(x), v\big) - K\big(\overline{\varphi}_\alpha^{-1}(y), v\big) \right) X(v) \right| \\
&\leq \sqrt[q]{\sum_{v \in \mathcal{V}} \left| K\big(\overline{\varphi}_\alpha^{-1}(x), v\big) - K\big(\overline{\varphi}_\alpha^{-1}(y), v\big) \right|^q} \sqrt[p]{\sum_{v \in \mathcal{V}} X(v)^p} \\
&\leq \sqrt[q]{\sum_{v \in \mathcal{V}} A^q} \sqrt[p]{\sum_{v \in \mathcal{V}} X(v)^p} \|x - y\|^\gamma \\
&\leq L \|x - y\|^\gamma .
\end{aligned}
$$

Here A bounds the Hölder constants of $K\big(\overline{\varphi}_\alpha^{-1}(\cdot), v\big)$, for all $\alpha \in \{1, \ldots, P\}$ and all $v \in \mathcal{V}$ from above, and $L = |\mathcal{V}| A \sqrt[p]{\sum_{v \in \mathcal{V}} X(v)^p}$. If $p = 1$ then the statement with the q-th root is the maximum over $\mathcal{V}$ instead of the q-norm. The result follows as by assumption $\mathbb{E}[L^p]$ is finite. $\square$

D.2 Proof of Proposition 3

Proof. The functions $K\big(\overline{\varphi}_\alpha^{-1}(\cdot), v\big)$ are Lipshitz continuous for each $v \in \mathcal{V}$ since they are $\mathcal{C}^1$ and $\overline{\mathcal{M}}$ is compact. Because $\mathcal{V}$ is finite and $\alpha \in \{1, \ldots, P\}$, there exists an $M > 0$ that bounds all the Lipschitz constants of the functions $K\big(\overline{\varphi}_\alpha^{-1}(\cdot), v\big)$. Thus, applying Proposition 2 with $p = 2$ and $\gamma = 1$ yields for the charts $(\overline{U}_\alpha, \overline{\varphi}_\alpha)$, $\alpha \in \{1, \ldots, P\}$, in the atlas of $\overline{\mathcal{M}}$ covering $\mathcal{M}$, all $x, y \in \overline{\varphi}(\overline{U}_\alpha) \cap \mathcal{M}$ such that $0 < \|x - y\| < 1$ that

$$
\mathbb{E}\left[\left(\tilde{X}_\alpha(x) - \tilde{X}_\alpha(y) \right)^2 \right] \leq \mathbb{E}\left[L^2 \|x - y\|^2 \right] = \mathbb{E}\left[L^2 \right] \|x - y\|^2 .
$$

The claim follows since $x^2 \leq (\log |x|)^{-2}$ for $0 < x < 1$. $\square$

D.3 Proof of Proposition 4

Proof. As the property is local, we can w.l.o.g. assume that $\mathcal{M}$ is a compact domain in $\mathbb{R}^D$ and drop the chart notation for simplicity. Given $x \in \mathcal{M}$, suppose that there exist sets of real constants a, a_i, a_{jk}, c $(1 \leq i \leq D, 1 \leq j \leq k \leq D)$ such that

$$
a\tilde{X}(s) + \sum_{i=1}^D a_i \tilde{X}_i(s) + \sum_{1 \leq j \leq k \leq D} a_{jk} \tilde{X}_{jk}(s) = c,
$$

which implies that

$$
\begin{aligned}
a \sum_{v \in \mathcal{V}_s} K(s, v) X(v) &+ \sum_{i=1}^D a_i \sum_{v \in \mathcal{V}_s} \partial_i^s K(s, v) X(v) \\
&+ \sum_{1 \leq j \leq k \leq D} a_{jk} \sum_{v \in \mathcal{V}_s} \partial_{jk}^s K(s, v) X(v) = c .
\end{aligned}
$$

Non-degeneracy of $(X(v) : v \in \mathcal{V}_s)$ then implies that for all $v \in \mathcal{V}_s$

$$aK(s,v) + \sum_{i=1}^{D} a_i \partial_i^s K(s,v) + \sum_{1 \leq j \leq k \leq D} a_{jk} \partial_{jk}^s K(s,v) = 0,$$

which by the linear independence constraint implies that the constants are all zero. This proves non-degeneracy of $\big(Y(s), \nabla Y(s), (\nabla^2 Y(s))\big)$.

For the normalized field $\tilde{X}/\sqrt{\mathrm{Var}[\tilde{X}]} = \tilde{X}/\sigma$, we note that

$$\nabla \frac{X}{\sigma} = \frac{\nabla X}{\sigma} - \frac{X \nabla \sigma}{\sigma^2} = \frac{\nabla X}{\sigma} - \frac{\nabla \sigma}{\sigma}\left(\frac{X}{\sigma}\right)$$

and

$$\nabla^2 \frac{X}{\sigma} = \frac{\nabla^2 X}{\sigma} - \frac{2(\nabla X)^T (\nabla \sigma)}{\sigma^2} - \frac{X \nabla^2 \sigma}{\sigma^2} + \frac{2(\nabla \sigma)^T (\nabla \sigma) X}{\sigma^3}.$$

Hence using an invertible matrix we can transform $\big(X(s), \nabla X(s), \mathbb{V}(\nabla^2 X(s))\big)$ into $\big(Z(s), \nabla Z(s), \mathbb{V}(\nabla^2 Z(s))\big)$. Thus, $\big(Z(s), \nabla Z(s), \mathbb{V}(\nabla^2 Z(s))\big)$ is non-degenerate by Lemma A.2 from Davenport and Telschow [2022]. □

D.4 Proof of Proposition 6

D.4.1 Establishing non-degeneracy of the isotropic kernel and its derivatives under linear transformations

Lemma 2. *Suppose that $\mathcal{V}$ satisfies the conditions of Proposition 6 and let $K^* = e^{-\|s-v\|^T/2}$ be the D-dimensional isotropic Gaussian kernel. Then given constants c, a_i, a_{jk} (for $1 \leq i \leq D$ and $1 \leq j \leq k \leq D$), $s \in \mathbb{R}^D$ and an invertible symmetric matrix $\Omega' \in \mathbb{R}^{D \times D}$ such that*

$$cK^*(\Omega' s, \Omega' v) + \sum_{j=1}^{D} a_j \partial_j^x K^*(\Omega' s, \Omega' v) \tag{44}$$
$$+ \sum_{1 \leq j \leq k \leq D} a_{jk} \partial_{jk}^x K^*(\Omega' s, \Omega' v) = 0,$$

for all $v \in \mathcal{V}$ then $c = a_i = a_{jk} = 0$ for $1 \leq i \leq D$ and $1 \leq j \leq k \leq D$.

Proof. For all $v \in \mathcal{V}$, letting $s^* = \Omega' s$ and dividing (44) by $e^{-\|s-v\|^2/2}$, it follows that

$$c + \sum_{j=1}^{D} a_j \left(s_j^* - \sum_{l=1}^{D} \Omega'_{jl} v_l\right)$$
$$+ \sum_{1 \leq j \leq k \leq D} a_{jk} \left(\left(s_j^* - \sum_{l=1}^{D} \Omega'_{jl} v_l\right)\left(s_k^* - \sum_{l=1}^{D} \Omega'_{kl} v_l\right) - \delta_{jk}\right) = 0.$$

In particular,

$$c + \sum_{j=1}^{D} a_j \left(s_j^* - \sum_{l=1}^{D} \Omega'_{jl} v_l\right) \tag{45}$$
$$+ \sum_{1 \leq j,k \leq D} a'_{jk} \left(\left(s_j^* - \sum_{l=1}^{D} \Omega'_{jl} v_l\right)\left(s_k^* - \sum_{l=1}^{D} \Omega'_{kl} v_l\right) - \delta_{jk}\right) = 0.$$

where $a'_{jk} = a_{jk}/2$, $j \neq k$ and $a'_{jj} = a_{jj}$.

For $i \in \{1, \ldots, D\}$, fixing $(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_D)$, we can view (45) as a quadratic in v_i. As such the only way that it can have more than two distinct solutions is if where the coefficient of v_i^2 is zero, i.e.,

$$\sum_{1 \leq j,k \leq D} a'_{jk} \Omega'_{ji} \Omega'_{ki} = (\Omega' A' \Omega')_{ii} = 0. \tag{46}$$

Similarly the coefficient of v_i must be zero, i.e.,

$$\sum_{j=1}^{D} a_j \Omega'_{ji} + \sum_{j,k} \Omega'_{ji} \left(s_k^* - \sum_{m \neq i} \Omega'_{km} v_m \right) + \sum_{j,k} \left(s_k^* - \sum_{l \neq i} \Omega'_{jl} v_l \right) \Omega'_{ki} = 0. \tag{47}$$

Now allowing v_n to vary for some $n \neq i$, by the same logic, the coefficient of v_n in (47) is equal to zero, i.e.,

$$\sum_{j,k} a'_{jk} \Omega'_{ji} \Omega'_{kn} + \sum_{j,k} a'_{jk} \Omega'_{jn} \Omega'_{ki} = (\Omega' A' \Omega')_{in} + (\Omega' A' \Omega')_{ni} = 2(\Omega' A' \Omega')_{in} = 0.$$

As such $(\Omega' A' \Omega')_{in} = (\Omega' A' \Omega')_{ni} = 0$ for all $i \neq n$. Combining this with (46), it follows that $\Omega' A' \Omega' = 0$. In particular $A' = 0$ as Ω' is invertible. Thus, the remaining linear equation in v_i from (45) can only have more than one solution if

$$\sum_{j} a_j \Omega'_{ji} = 0.$$

Therefore $(\Omega' a)_i = 0$, where $a = (a_1, \ldots, a_D)^T$. Since this holds for all i and Ω' is invertible, we obtain $a = 0$. Finally this implies that $c = 0$. $\qquad \square$

D.4.2 Establishing Proposition 6

Proof. We can write $K(s, v) = K^*(\Sigma^{-1/2} s, \Sigma^{-1/2} v)$ where $K^* = e^{-\|s-v\|^2/2}$ is the isotropic Gaussian kernel. Arguing as in proof of Lemma 1 of Davenport and Telschow [2022] (taking $\phi(s) = \Sigma^{-1/2} s$ and φ to be the identity in their notation), for each $s \in S$ we have

$$\mathbb{V}\big(\nabla^2 K(s, v)\big) = L\big(\Sigma^{1/2} \otimes \Sigma^{1/2}\big) R \mathbb{V}\Big(\nabla^2 K^*\big(\Sigma^{-1/2} s, \Sigma^{-1/2} v\big)\Big)$$

where ∇^2 as usual always denotes the Hessian with respect to the first argument and $L \in \mathbb{R}^{D(D+1)/2 \times D^2}$ is the elimination matrix and $R \in \mathbb{R}^{D^2 \times D(D+1)/2}$ is the duplication matrix, the precise definitions of which can be found in Magnus and Neudecker [1980]. The matrix $L\big(\Sigma^{1/2} \otimes \Sigma^{1/2}\big) R$ is invertible by Lemma 4.4.iv of Magnus and Neudecker [1980], and the fact that $\Sigma^{1/2}$ is invertible.

Moreover $\nabla K\big(\Sigma^{-1/2} s, v\big) = \Sigma^{-1/2} \nabla K^*\big(\Sigma^{-1/2} s, \Sigma^{-1/2} v\big)$. As such there is an invertible linear transformation between the vector

$$\Big(K(s, v), \nabla K(s, v), \mathbb{V}\big(\nabla^2 K(s, v)\big)\Big)$$

and the vector

$$\Big(K^*\big(\Sigma^{-1/2} s, \Sigma^{-1/2} v\big), \nabla K^*\big(\Sigma^{-1/2} s, \Sigma^{-1/2} v\big), \mathbb{V}\big(\nabla^2 K^*\big(\Sigma^{-1/2} s, \Sigma^{-1/2} v\big)\big)\Big).$$

In particular if there exists constants c, a_d, a_{kl} such that $1 \leq d \leq D$ and $1 \leq k \leq l \leq D$ (with at least one of them being non-zero) such that

$$cK(s,v) + \sum_{d=1}^{D} a_d \partial_d^s K(s,v) + \sum_{1 \leq k \leq l \leq D} a_{kl} \partial_{kl}^s K(s,v) = 0, \tag{48}$$

then there existing corresponding constants (c^*, a_d^*, a_{kl}^*) (with at least one of them being non-zero) such that

$$\begin{aligned} c^* K^*\big(\Sigma^{-1/2}s, \Sigma^{-1/2}v\big) + \sum_{d=1}^{D} a_d^* \partial_d^s K^*\big(\Sigma^{-1/2}s, \Sigma^{-1/2}v\big) \\ + \sum_{1 \leq k \leq l \leq D} a_{kl}^* \partial_{kl}^s K^*\big(\Sigma^{-1/2}s, \Sigma^{-1/2}v\big) = 0. \end{aligned} \tag{49}$$

Applying Lemma 2 yields a contradiction and thus establishes the result. $\qquad \square$

D.5   Proof of Proposition 7

Proof. Note that for a SuRF $(\tilde{X}, X, K, \mathcal{V})$ we obtain the following identity

$$\langle \partial_d K_s, K_{s'} \rangle = \mathrm{Cov}\left[\partial_d^s \tilde{X}(s), \tilde{X}(s') \right] = \langle \partial_d^s, 1_{s'} \rangle. \tag{50}$$

Hence the result for a normalized SuRF is a Corollary of Theorem 5. $\qquad \square$

D.6   Proof of Theorem 1

Proof. In order to apply Theorem 12.4.2 from Adler and Taylor [2007] we need to prove that the assumptions **(G1)-(G3)** hold. The assumption that $\tilde{X}_\alpha$ is Gaussian with almost surely C^2-sample paths clearly holds by the assumption that X is a Gaussian field on $\mathcal{V}$ and $K(\cdot, v) \in C^3(\overline{\mathcal{M}})$ for all $v \in \mathcal{V}$. The non-degeneracy condition follows from Proposition 4. The last assumption that there is an $\epsilon > 0$ such that

$$\mathbb{E}\left[\big(\partial_{dd'} \tilde{X}_\alpha(x) - \partial_{dd'} \tilde{X}_\alpha(y) \big)^2 \right] \leq K \big| \log \|x - y\| \big|^{-(1+\gamma)}$$

for some $K > 0$, all $d, d' \in \{1, ..., D\}$ and for the charts $(\overline{U}_\alpha, \overline{\varphi}_\alpha)$, $\alpha \in \{1, \ldots, P\}$, in the atlas of $\overline{\mathcal{M}}$ covering $\mathcal{M}$, all $x, y \in \overline{\varphi}(\overline{U}_\alpha) \cap \mathcal{M}$ such that $|x - y| < \epsilon$ is established in Proposition 3. $\quad \square$

D.7   Proof of Theorem 2

Computation of $\Theta(x) = \int_{\mathbb{S}\left(\mathcal{T}_x \partial_1 \mathcal{M}_\mathcal{V}^{\perp}\right)} \alpha(\nu) \mathcal{H}_1(d\nu)$ for Voxel Manifolds.   In order to compute $\Theta(x)$ we need to introduce the normal Morse index $\alpha(\nu)$. We specialize here to the case of $\mathcal{M}_\mathcal{V}$ being a voxel manifold embedded into $\mathbb{R}^3$, yet the exact same concept is defined for any Whitney stratified manifold, compare [Adler and Taylor, 2007, Scn 9.2.1].

For any $x \in \mathcal{M}_\mathcal{V}$ and any direction $\nu \in \mathbb{S}(\mathcal{T}_x \mathcal{M}_\mathcal{V})$ the normal Morse index is one minus the local Euler characteristic of the intersection of $\mathcal{M}_\mathcal{V}$, the δ-ball centered at x and the affine plane $\{\lambda \in \mathbb{R}^3 : \lambda^T \nu + x + \epsilon = 0\}$ for $\epsilon > 0$. If δ is sufficiently small, then this Euler characteristic does not depend on ϵ provided that ϵ is small enough. Although this definition sounds complicated at first, it can be easily computed for all x and ν for a voxel manifold. Note that $\mathcal{M}_\mathcal{V}$ or more

precisely an open neighbourhood of it might be endowed with a different Riemannian metric than the standard Riemannian metric on $\mathbb{R}^3$, which in our case is the induced Riemannian metric $\boldsymbol{\Lambda}$ by a unit-variance random field f. In this case $\nu \in \mathbb{S}(\mathcal{T}_x\mathbb{R}^3)$ are vectors $\nu \in \mathbb{R}^3$ such that $\nu^T \boldsymbol{\Lambda}(x)\nu = 1$.

In the case that $x \in \partial_3\mathcal{M}_\mathcal{V}$ and $\nu \in \mathbb{S}(\mathcal{T}_x\mathbb{R}^3)$ it is obvious that $\alpha(\nu) = 0$ because the intersection of the affine plain defined by ν and the δ-ball is always homeomorphic to a filled disk, which has Euler characteristic 1. Similarly, if $x \in \partial_2\mathcal{M}_\mathcal{V}$, then $\alpha(\nu) = 0$ for all $\nu \in \mathbb{S}(\mathcal{T}_x\mathbb{R}^3) \setminus \{\nu_{out}\}$. Here ν_{out} is the unique outside pointing normal (w.r.t. the Euclidean metric!) at $x \in \partial_2\mathcal{M}_\mathcal{V}$ and it holds that $\alpha(\nu_{out}) = 1$ as the plain defined by ν_{out} for a small enough δ-ball is parallel to $\partial_2\mathcal{M}_\mathcal{V}$ and therefore its intersection with $\mathcal{M}_\mathcal{V}$ is empty. The important cases for us and the interesting cases happen at edges of the voxel manifold, i.e., for $x \in \partial_1\mathcal{M}_\mathcal{V}$. Here the behavior of $\alpha(\nu)$ can be classified by the three types of possible edges: the convex, the double convex and the concave edge. These cases are shown in Figure 2 in the main manuscript.

The behavior of α for directions $\nu \in \mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp)$ is demonstrated in Figure 23 within the hyperplane $x + \mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp$ with $\mathcal{M}_\mathcal{V}$. Here we show the two possible intersection scenarios of the hyperplane (bold green line) orthogonal to the direction $-\nu \in \mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp)$ (green arrow) with a small δ-ball (dotted black line) and the voxel manifold $\mathcal{M}_\mathcal{V}$ from the definition of α. In particular, it can be seen that $\alpha(\nu)$ is constant, if $-\nu$ is inside the normal cone $\mathcal{N}_x\mathcal{M}_\mathcal{V}$ (for a definition see [Adler and Taylor, 2007, p.189]) and constant on $\mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp) \setminus \mathcal{N}_x\mathcal{M}_\mathcal{V}$ indpendent on the type of edge to which x belongs. The geometric embedding of the the intersection of the (geometric) normal cone $x + \mathcal{N}_x\mathcal{M}_\mathcal{V}$ with the hyperplane $x + \mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp$ is represented by the red shaded areas. From this we deduce that on a convex edge the Euler characteristic of the intersection of the green hyperplane with the δ-ball and $\mathcal{M}_\mathcal{V}$ is 0, if $-\nu \in \mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp)$, as the intersection is empty and 1 otherwise because the intersection is homeomorphic to a disk. Similar, it holds that the Euler characteristic of the intersection is 2 for x on a double convex or concave edge, if $-\nu \in \mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp)$, as the intersection is homeomorphic to the disjoint union of two disks and 1 else because the intersection is homeomorphic to a disk. Therefore we obtain that $\alpha(\nu)$ for $\nu \in \mathbb{S}(\mathcal{T}_x\partial_1\mathcal{M}_\mathcal{V}^\perp)$ is given by

$$\alpha(\nu) = \begin{cases} 1, & \text{if } x \text{ on convex edge and } -\nu \in (\mathcal{N}_x\mathcal{M}_\mathcal{V})^\circ \\ -1, & \text{if } x \text{ on a double convex edge and } -\nu \in (\mathcal{N}_x\mathcal{M}_\mathcal{V})^\circ \\ -1, & \text{if } x \text{ on concave edge and } -\nu \in (\mathcal{N}_x\mathcal{M}_\mathcal{V})^\circ \\ 0, & \text{else} \end{cases} . \tag{51}$$

Using this we can compute the function $\Theta(x)$.

Lemma 3. *Let $\mathcal{M}_\mathcal{V}$ be a voxel manifold, $x \in \partial_1\mathcal{M}_\mathcal{V}$ and U_x, V_x, W_x denote an ON frame for $\mathcal{T}_x\overline{\mathcal{M}_\mathcal{V}}$. Define $M = V_x \times W_x = \big(m_1(x), m_2(x), m_3(x)\big)^T$. Define*

$$\beta(x) = \arccos\left(\frac{m_2(x)m_3(x)}{\sqrt{m_2^2(x) + m_1^2(x)}\sqrt{m_3^2(x) + m_1^2(x)}}\right).$$

Then we obtain

$$\Theta(x) = \begin{cases} \pi - \beta(x), & \text{if } x \text{ belongs to a convex edge} \\ -2\beta(x), & \text{if } x \text{ belongs to a double convex edge} \\ \beta(x) - \pi, & \text{if } x \text{ belongs to a convex edge} \end{cases} .$$
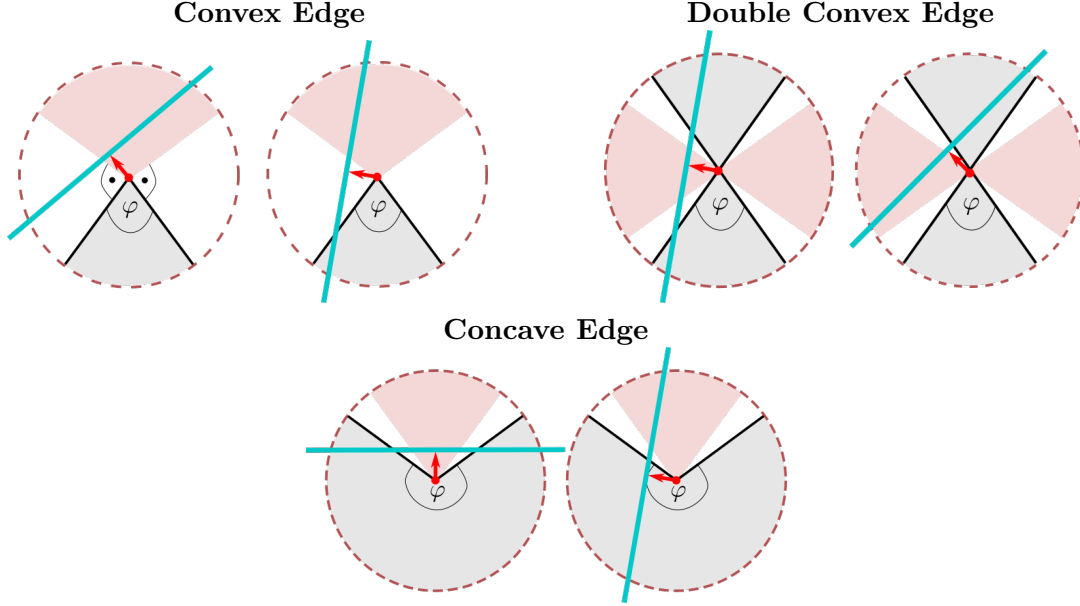
Figure 23: The two different scenarios of intersections of the hyperplane at the three different types of edges appearing in a voxel manifold $\mathcal{M}_\mathcal{V}$ illustrated within the hyperplane $x + \mathcal{T}_x \partial_1 \mathcal{M}_\mathcal{V}^\perp$. The dashed red line is the boundary of an δ-ball centered at x. The grey shaded area belongs to $\mathcal{M}_\mathcal{V}$ and the red shaded area is the (geometric) normal cone $x + \mathcal{N}_x \mathcal{M}_\mathcal{V}$. Note that the boundary of the normal cone is orthogonal to the boundary of $\mathcal{M}_\mathcal{V}$ The red arrow is the unit direction $-\nu \in \mathbb{S}(\mathcal{T}_x \partial_1 \mathcal{M}_\mathcal{V}^\perp)$ scaled by ϵ and the turquoise line is the hyperplane defined by $-\nu$.

Proof. Let $\beta(x)$ be the minimum of φ and $2\pi - \varphi$ where ϕ is the opening angle within $\mathcal{M}_\mathcal{V}$ from Figure 23. This yields

$$\int_{\mathbb{S}(\mathcal{T}_x \partial_1 \mathcal{M}_\mathcal{V}^\perp)} \alpha(\nu_2) \mathcal{H}_1(d\nu_2) = \int_0^{2\pi} \mathbb{1}_{-\sin(t)E_1 - \cos(t)E_2 \in (\mathcal{N}_x \mathcal{M}_\mathcal{V})^\circ} dt$$

$$= \begin{cases} \pi - \beta(x), & \text{if } x \text{ on a convex edge} \\ -2\beta(x), & \text{if } x \text{ on a double convex edge} \\ \beta(x) - \pi, & \text{if } x \text{ on a convex edge} \end{cases},$$

because the characteristic function is only 1, if $-\nu$ belongs to the normal cone and hence the integral is equal to the opening angle of the normal cone (red shaded area in Fig. 23).

It remains to compute the angle $\beta(x)$ which is obtained by computing the angle between the intersection of the affine plane $x + \mathcal{T}_x \partial_1 \mathcal{M}_\mathcal{V}^\perp$ and the boundary of $\mathcal{M}_\mathcal{V}$ at x. We only treat the case of $x \in \mathcal{M}_\mathcal{V}$ lying on a convex edge. Double convex and concave edges follow analogously. Since the tangent space and $\mathbb{R}^3$ can be identified we assume w.l.o.g. that $x = 0$ and the voxel is given by the set $\{y \in \mathbb{R}^3 : y_3 \leq 0, y_2 \leq 0\}$. (note we extend the edge infinitely, which does not make a difference in this argument) Its boundary is given by the set $A \cup B = \{y \in \mathbb{R}^3 : y_3 = 0, y_2 \leq 0\} \cup \{y \in \mathbb{R}^3 : y_3 \leq 0, y_2 = 0\}$. The edge to which x belongs is given by $E = \{y \in \mathbb{R}^3 : y_3 = 0, y_2 = 0\}$. An orthonormal basis at x with U_x spanning $\mathcal{T}_x E$ is given in Proposition 15. The plane in which the unit circle $\mathbb{S}^1(\mathcal{T}_x E^\perp)$ lies is given by the linear span of V_x, W_x, which we denote by $\mathcal{F} = \{y \in \mathbb{R}^3 : m_1(x)y_1 + m_2(x)y_2 + m_3(x)y_3 = 0\}$ for

some $m_1(x), m_2(x), m_3(x) \in \mathbb{R}$. The intersection $\mathcal{F} \cap A$ and $\mathcal{F} \cap B$ are given by

$$
\begin{aligned}
A \cap \mathcal{F} &= \left\{ y \in \mathbb{R}^3 : \ m_1(x)y_1 + m_2(x)y_2 = 0 \ \wedge \ y_2 \leq 0 \ \wedge \ y_3 = 0 \right\} \\
B \cap \mathcal{F} &= \left\{ y \in \mathbb{R}^3 : \ m_1(x)y_1 + m_3(x)y_3 = 0 \ \wedge \ y_3 \leq 0 \ \wedge \ y_2 = 0 \right\}
\end{aligned}
\tag{52}
$$

By construction $m_1(x) \neq 0$, since otherwise $\mathcal{F} \cap A = \mathcal{F} \cap B = E$ and hence V, W cannot be both orthogonal to E which contradicts the assumption that U_x, V_x, W_x form an orthonormal basis for $\mathcal{T}_x \overline{\mathcal{M}_\mathcal{V}}$. Thus, if $m_1(x) \neq 0$ we have that $(m_2(x)/m_1(x), -1, 0)$ and $(m_3(x)/m_1(x), 0, -1)$ are vectors in the intersection, which we can identify with tangent directions along E. Thus,

$$
\begin{aligned}
\varphi &= \arccos \left(\frac{\frac{m_2(x)m_3(x)}{m_1^2(x)}}{\sqrt{\frac{m_2^2(x)}{m_1^2(x)} + 1}\sqrt{\frac{m_3^2(x)}{m_1^2(x)} + 1}} \right) \\
&= \arccos \left(\frac{m_2(x)m_3(x)}{\sqrt{m_2^2(x) + m_1^2(x)}\sqrt{m_3^2(x) + m_1^2(x)}} \right).
\end{aligned}
\tag{53}
$$

The same formula holds true if $m_1(x) = 0$ and $m_2(x) \neq 0$ and $m_3(x) \neq 0$. $\qquad\square$

Computation of $\int_{\partial_2 M} \mathbf{\Lambda}_x \left(\overline{\nabla}_{U_x} U_x + \overline{\nabla}_{V_x} V_x, N_x \right) \mathcal{H}_2(dx)$ **for Voxel Manifolds** Assume w.l.o.g. that $\mathcal{T}_x \mathcal{M}_\mathcal{V}$ for $x \in \partial_2 \mathcal{M}_\mathcal{V}$ is contained in the x_1-x_2-plane. An orthonormal frame is given by U_x, V_x, W_x from (15) and by construction $N_x = \pm W_x$, where the sign depends on whether W_x is inward or outward pointing. Using the coordinate representations $U_x = \sum_{d=1}^{3} U_d(x)\partial_d$ and $V_x = \sum_{d=1}^{3} V_d(x)\partial_d$, linearity and product rule for the covariant derivative and formula (33), we obtain

$$
\begin{aligned}
\mathbf{\Lambda}_x \left(\overline{\nabla}_{U_x} U_x, N_x \right) &= N_x^T \mathbf{\Lambda}(x) U_1(x) \begin{pmatrix} \partial_1 U_1(x) \\ 0 \\ 0 \end{pmatrix} + U_1^2(x) N_x^T \begin{pmatrix} \Gamma_{111}(x) \\ \Gamma_{112}(x) \\ \Gamma_{113}(x) \end{pmatrix} \\
&= U_1^2(x) N_x^T \begin{pmatrix} \Gamma_{111}(x) \\ \Gamma_{112}(x) \\ \Gamma_{113}(x) \end{pmatrix}.
\end{aligned}
$$

Here the second equality is due to the fact that $N_x \propto \mathbf{\Lambda}^{-1} E_3$, if represented as a vector. Similarly it holds that

$$
\begin{aligned}
\mathbf{\Lambda}_x \left(\overline{\nabla}_{V_x} V_x, N_x \right) &= V_1^2(x) N_x^T \begin{pmatrix} \Gamma_{111}(x) \\ \Gamma_{112}(x) \\ \Gamma_{113}(x) \end{pmatrix} + V_2^2(x) N_x^T \begin{pmatrix} \Gamma_{221}(x) \\ \Gamma_{222}(x) \\ \Gamma_{223}(x) \end{pmatrix} \\
&\quad + V_1(x) V_2(x) N_x^T \begin{pmatrix} \Gamma_{121}(x) \\ \Gamma_{122}(x) \\ \Gamma_{123}(x) \end{pmatrix}
\end{aligned}
$$

Summarizing this yields the following proposition about the trace of the shape operator along $\partial_2 \mathcal{M}_\mathcal{V}$, i.e., $\overline{g} \left(\overline{\nabla}_{U_x} U_x + \overline{\nabla}_{V_x} V_x, N_x \right)$.

Proposition 9. *Let $\mathcal{M}_\mathcal{V}$ be a voxel manifold and assume that $x \in \partial_2 \mathcal{M}_\mathcal{V}$ such that $\mathcal{T}_x \partial_2 \mathcal{M}_\mathcal{V}$ is spanned by E_k, E_l. Then*

$$\mathbf{\Lambda}_x\left(\overline{\nabla}_{U_x} U_x + \overline{\nabla}_{V_x} V_x, N_x\right)$$

$$= \left(U_k^2(x) + V_k^2\right)\nu^T \begin{pmatrix} \Gamma_{kk1}(x) \\ \Gamma_{kk2}(x) \\ \Gamma_{kk3}(x) \end{pmatrix} + V_l^2(x)\nu^T \begin{pmatrix} \Gamma_{ll1}(x) \\ \Gamma_{ll2}(x) \\ \Gamma_{ll3}(x) \end{pmatrix}$$

$$+ V_k(x)V_l(x)\nu^T \begin{pmatrix} \Gamma_{kl1}(x) \\ \Gamma_{kl2}(x) \\ \Gamma_{kl3}(x) \end{pmatrix}$$

D.8 Proof of Theorem 3

We prove this result in more generality as the restriction to SuRFs is not necessary. The proof extends that of Taylor and Worsley [2007] which established the result for the trivial linear model. To do so assume that $Y_1, \ldots, Y_N \sim Y$ is an iid sample of Gaussian random fields on a voxel manifold $\mathcal{M}_\mathcal{V}$. We restrict here to voxel manifolds to avoid working in a chart, however, the proof easily generalizes. This sample we represent as the vector $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$ and therefore the corresponding estimator generalizing (16) is, for $x \in \mathcal{M}_\mathcal{V}$,

$$\hat{\Lambda}_{dd'}(x) = \frac{\mathrm{Cov}\left[\partial_d \mathbf{Y}(x), \partial_{d'} \mathbf{Y}(x)\right]}{\mathrm{Var}\left[\mathbf{Y}(x)\right]}$$

$$= \mathrm{Cov}\left[\partial_d\left(\frac{\mathbf{Y}(x)}{\sqrt{\mathrm{Var}\left[\mathbf{Y}(x)\right]}}\right), \partial_{d'}\left(\frac{\mathbf{Y}(x)}{\sqrt{\mathrm{Var}\left[\mathbf{Y}(x)\right]}}\right)\right] \tag{54}$$

$$- \frac{\mathrm{Cov}\left[\partial_d \mathbf{Y}(x), \mathbf{Y}(x)\right]\mathrm{Cov}\left[\mathbf{Y}(x), \partial_{d'} \mathbf{Y}(x)\right]}{\mathrm{Var}\left[\mathbf{Y}(x)\right]^2}.$$

Here as in the main manuscript the operation $\mathrm{Var}\left[\,\cdot\,\right]$ and $\mathrm{Cov}\left[\cdot, \cdot\right]$ denote the sample variance and sample covariance respectively. We define the vector of normalized residuals to be $\mathbf{R}(x) = \frac{\mathbf{HY}(x)}{\|\mathbf{HY}(x)\|}$, $x \in \mathcal{M}_\mathcal{V}$, where $\|\cdot\|$ denotes the Euclidean norm and $\mathbf{H} = I_{N \times N} - \mathbb{1}\mathbb{1}^T$ with $\mathbb{1}^T = (1, \ldots, 1) \in \mathbb{R}^N$, is the centering matrix. Note that $\mathbf{R}$ does not depend on the (unknown) variance $\mathrm{Var}[Y]$ as for the sample $\tilde{\mathbf{Y}} = \mathbf{Y}/\sqrt{\mathrm{Var}[Y]}$ we have that $\tilde{\mathbf{R}}(x) = \mathbf{R}(x)$ for all $x \in \mathcal{M}_\mathcal{V}$. Thus, using

$$\nabla \mathbf{X}(x) = \begin{pmatrix} \frac{\partial X_1}{\partial x_1}(x) & \cdots & \frac{\partial X_1}{\partial x_D}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial X_D}{\partial x_1}(x) & \cdots & \frac{\partial X_D}{\partial x_D}(x) \end{pmatrix}$$

for $\mathbf{X} = (X_1, \ldots, X_D)^T \in C(\mathcal{M}_\mathcal{V}, \mathbb{R}^D)$, we can rewrite (54) in matrix terms as

$$\hat{\mathbf{\Lambda}} = (\nabla \mathbf{R})^T \nabla \mathbf{R}.$$

Recall that the true underlying $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda} = \mathbb{E}\left[\left(\nabla\left(\frac{Y - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}\right)\right)^T \nabla\left(\frac{Y - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}\right)\right] = \mathbb{E}\left[(\nabla \tilde{Y}^T)\nabla \tilde{Y}\right],$$

where $\tilde{Y} = \frac{Y}{\sqrt{\mathrm{Var}[Y]}}$. Using this we obtain the following lemma which the proof of which follows the corresponding proof in Taylor and Worsley [2007].

Lemma 4. *Under the assumption and notation described above, we have that*

$$\mathbb{E}\left[\sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)}\right] = \sqrt{\det\left(\mathbf{\Lambda}(x)\right)} \tag{55}$$

Proof. $\mathbf{H}$ is a projection matrix and so $\mathbf{HH} = \mathbf{H}$ and $\mathbf{H}^T = \mathbf{H}$. We can expand $\nabla\mathbf{R}$ in terms of $\tilde{\mathbf{Y}}$ as follows:

$$\nabla\mathbf{R} = \frac{\nabla\mathbf{H}\tilde{\mathbf{Y}}}{\|\mathbf{H}\tilde{\mathbf{Y}}\|} - \frac{\mathbf{H}\tilde{\mathbf{Y}}\nabla\left((\mathbf{H}\tilde{\mathbf{Y}})^T\mathbf{H}\tilde{\mathbf{Y}}\right)}{2\|\mathbf{H}\tilde{\mathbf{Y}}\|^3} = \left(I_{N\times N} - \frac{\mathbf{H}\tilde{\mathbf{Y}}(\mathbf{H}\tilde{\mathbf{Y}})^T}{\|\mathbf{H}\tilde{\mathbf{Y}}\|^2}\right)\frac{\mathbf{H}\nabla\tilde{\mathbf{Y}}}{\|\mathbf{H}\tilde{\mathbf{Y}}\|}.$$

Now note that

$$\mathbf{H}\left(I_{N\times N} - \frac{\mathbf{H}\tilde{\mathbf{Y}}(\mathbf{H}\tilde{\mathbf{Y}})^T}{\|\mathbf{H}\tilde{\mathbf{Y}}\|^2}\right)\mathbf{H} = \mathbf{H} - \frac{\mathbf{H}\tilde{\mathbf{Y}}(\mathbf{H}\tilde{\mathbf{Y}})^T}{\|\mathbf{H}\tilde{\mathbf{Y}}\|^2}$$

and therefore it is idempotent with

$$\text{tr}\left(\mathbf{H} - \frac{\mathbf{H}\tilde{\mathbf{Y}}(\mathbf{H}\tilde{\mathbf{Y}})^T}{\|\mathbf{H}\tilde{\mathbf{Y}}\|^2}\right) = \text{tr}\left(\mathbf{H}\right) - 1$$

Thus, applying Cochran's Theorem [Mardia et al., 1979, Theorem 3.4.4], it follows that for all $x \in S$,

$$\left(\nabla\mathbf{R}(x)\right)^T\nabla\mathbf{R}(x) \mid \tilde{\mathbf{Y}}(x) \sim Wish_D\left(\mathbf{\Lambda}\|\mathbf{H}\tilde{\mathbf{Y}}\|^{-2}, \text{tr}\left(\mathbf{H}\right) - 1\right).$$

As such by [Mardia et al., 1979, Corollary 3.4.1.2]

$$\mathbf{W}(x) \sim \|\mathbf{H}\tilde{\mathbf{Y}}\|^2\mathbf{\Lambda}(x)^{-1/2}\left(\nabla\mathbf{R}(x)\right)^T\nabla\mathbf{R}(x)\mathbf{\Lambda}(x)^{-1/2} \mid \tilde{\mathbf{Y}}(x)$$
$$\sim Wish_D\left(I_{N\times N}, \text{tr}(\mathbf{H}) - 1\right)$$

Taking determinants and rearranging it follows that, unconditionally,

$$\sqrt{\det\left(\left(\nabla\mathbf{R}(x)\right)^T\nabla\mathbf{R}(x)\right)} \sim \sqrt{\det\left(\mathbf{\Lambda}(x)\right)\det\left(\mathbf{W}(x)\right)V(x)^{-D}},$$

where $\mathbf{W}(x)$ is independent of $V(x) = \|\mathbf{H}\tilde{\mathbf{Y}}\|^2 \sim \chi^2_{\text{tr}(\mathbf{H})}$. From the independence, the formula for the moments of χ^2-distributions and Theorem 3.4.8 from Mardia et al. [1979] we obtain

$$\mathbb{E}\left[\sqrt{\det\left(\left(\nabla\mathbf{R}(x)\right)^T\nabla\mathbf{R}(x)\right)}\right] = \sqrt{\det\left(\mathbf{\Lambda}(x)\right)}\mathbb{E}\left[\sqrt{\det\left(\mathbf{W}(x)\right)}\right]$$
$$\mathbb{E}\left[V(x)^{-\frac{D}{2}}\right] = \det\left(\mathbf{\Lambda}(x)\right)^{1/2}.$$

$\square$

Remark 11. Lemma 4 also holds for estimates of $\mathbf{\Lambda}$ in a linear model, compare Taylor and Worsley [2007]. This follows by a similar proof as, in that setting, the residuals $\mathbf{R}$ which are used to calculate $\hat{\mathbf{\Lambda}}$ are obtained by $\mathbf{R} = \mathbf{PY}$, where $\mathbf{P}$ is idempotent.

Proof of Theorem 3. In what follows we establish the results for $\hat{\mathcal{L}}_D^{(r)}$. The proof for $\hat{\mathcal{L}}_{D-1}^{(r)}$ is identical as each $\mathbf{\Lambda}^I$ is a $(D-1) \times (D-1)$-submatrix of $\mathbf{\Lambda}$.

Using Lemma 4 and the approximation of integrals of continuous functions by Riemann sums yields

$$
\begin{aligned}
\lim_{r \to \infty} \mathbb{E}\left[\hat{\mathcal{L}}_D^{(r)}\right] &= \lim_{r \to \infty} \mathbb{E}\left[\sum_{x \in \mathcal{M}_\mathcal{V}^{(r)}} \sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)} \prod_{d=1}^{D} \frac{\delta_d}{r+1}\right] \\
&= \lim_{r \to \infty} \sum_{x \in \mathcal{M}_\mathcal{V}^{(r)}} \mathbb{E}\left[\sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)}\right] \prod_{d=1}^{D} \frac{\delta_d}{r+1} \\
&= \lim_{r \to \infty} \sum_{x \in \mathcal{M}_\mathcal{V}^{(r)}} \sqrt{\det\left(\mathbf{\Lambda}(x)\right)} \prod_{d=1}^{D} \frac{\delta_d}{r+1} = \mathcal{L}_D
\end{aligned}
$$

Similarly, applying Fubini's theorem which is applicable as $\mathcal{M}_\mathcal{V}$ is compact, we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\lim_{r \to \infty} \hat{\mathcal{L}}_D^{(r)}\right] &= \mathbb{E}\left[\int_{\mathcal{M}_\mathcal{V}} \sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)} \prod_{d=1}^{D} dx_d\right] \\
&= \int_{\mathcal{M}_\mathcal{V}} \mathbb{E}\left[\sqrt{\det\left(\hat{\mathbf{\Lambda}}(x)\right)}\right] \prod_{d=1}^{D} dx_d \\
&= \int_{\mathcal{M}_\mathcal{V}} \sqrt{\det\left(\mathbf{\Lambda}(x)\right)} \prod_{d=1}^{D} dx_d = \mathcal{L}_D
\end{aligned}
$$

$\square$

D.9   Proof of Theorem 4

Proof. By the assumptions and Proposition 2 the SuRF has almost surely L^2-Hölder continuous paths and therefore the assumptions of Lemma 11 from Telschow and Schwartzman [2022] are satisfied. This means that $\hat{\mathbf{\Lambda}}$ converges uniformly almost surely to $\mathbf{\Lambda}$ over all $x \in \overline{\mathcal{M}_\mathcal{V}}$. Since the Riemann sum converges to the integral the $\lim_{r \to \infty} \lim_{N \to \infty}$ statement follows immediately. On the other hand the $\lim_{N \to \infty} \lim_{r \to \infty}$ statement is a special case of Theorem 3 from Telschow et al. since for $\mathcal{L}_2$ and $\mathcal{L}_3$ from their condition **(R)** only the uniform almost sure convergence of $\hat{\mathbf{\Lambda}}$ to $\mathbf{\Lambda}$ is required. $\square$