



HAL
open science

On the geometry of excursion sets: theoretical and computational guarantees

Ryan Cotsakis

► **To cite this version:**

Ryan Cotsakis. On the geometry of excursion sets: theoretical and computational guarantees. Computational Geometry [cs.CG]. Université Côte d'Azur, 2024. English. NNT: 2024COAZ5007. tel-04599432

HAL Id: tel-04599432

<https://theses.hal.science/tel-04599432v1>

Submitted on 3 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

Sur la géométrie des ensembles d'excursion : garanties théoriques et computationnelles

Ryan COTSAKIS

Université Côte d'Azur; Laboratoire J.A. Dieudonné (LJAD)

Présentée en vue de l'obtention du grade
de docteur en mathématiques d'Université
Côte d'Azur.

Dirigée par : Elena DI BERNARDINO / Thomas
OPITZ

Soutenue le : 6 mai 2024

Devant le jury composé de :

Catherine AARON, MCF, Université Clermont-Auvergne

Pierre CALKA, PR, Université de Rouen

David COHEN-STEINER, CR, INRIA Sophia Antipolis

Elena DI BERNARDINO, PR, Université Côte d'Azur

Clément DOMBRY, PR, Université de Franche-Comté

Aurèlie FISCHER, MCF, Université Paris Cité (invitée)

Céline LACAUX, PR, Avignon Université

Clément LEVRARD, PR, Université de Rennes

Thomas OPITZ, DR, INRAE Avignon

PR: Professeur(e); DR: Directeur de recherche; CR: Chargé de recherche;
MCF: Maître de conférences

On the geometry of excursion sets: theoretical and computational guarantees

Ryan Cotsakis

Université Côte d'Azur; Laboratoire J.A. Dieudonné (LJAD)

Presented with the aim of obtaining a doctorate in mathematics from the Université Côte d'Azur.

Thesis supervisors: Elena Di Bernardino, Thomas Opitz

Date of defence: May 6, 2024

Jury members:

Catherine Aaron, MCF, Université Clermont-Auvergne

Pierre Calka, PR, Université de Rouen

David Cohen-Steiner, CR, INRIA Sophia Antipolis

Elena Di Bernardino, PR, Université Côte d'Azur

Clément Dombry, PR, Université de Franche-Comté

Aurèlie Fischer, MCF, Université Paris Cité (invited)

Céline Lacaux, PR, Avignon Université

Clément Levrard, PR, Université de Rennes

Thomas Opitz, DR, INRAE Avignon

PR: Professor; DR: Directeur de recherche; CR: Chargé de recherche;

MCF: Maître de conférences

For my parents

Sur la géométrie des ensembles d'excursion : garanties théoriques et computationnelles

Jury :

Rapporteurs

Pierre CALKA, PR, Université de Rouen

Clément DOMBRY, PR, Université de Franche-Comté

Examineurs

Catherine AARON, MCF, Université Clermont-Auvergne

David COHEN-STEINER, CR, INRIA Sophia Antipolis

Elena DI BERNARDINO, PR, Université Côte d'Azur

Céline LACAUX, PR, Avignon Université

Clément LEVRARD, PR, Université de Rennes

Thomas OPITZ, DR, INRAE Avignon

Invités

Aurèlie FISCHER, MCF, Université Paris Cité

PR: Professeur(e); DR: Directeur de recherche; CR: Chargé de recherche; MCF: Maître de conférences

On the geometry of excursion sets: theoretical and computational guarantees

Jury:

Thesis referees

Pierre Calka, PR, Université de Rouen

Clément Dombry, PR, Université de Franche-Comté

Examiners

Catherine Aaron, MCF, Université Clermont-Auvergne

David Cohen-Steiner, CR, INRIA Sophia Antipolis

Elena Di Bernardino, PR, Université Côte d'Azur

Céline Lacaux, PR, Avignon Université

Clément Levrard, PR, Université de Rennes

Thomas Opitz, DR, INRAE Avignon

Invited

Aurèlie Fischer, MCF, Université Paris Cité

PR: Professor; DR: Directeur de recherche; CR: Chargé de recherche; MCF: Maître de conférences

Sur la géométrie des ensembles d'excursion : garanties théoriques et computationnelles

Resumé

L'ensemble d'excursion $E_X(u)$ d'un champ aléatoire réel X sur \mathbb{R}^d à un niveau de seuil $u \in \mathbb{R}$ est le sous-ensemble du domaine \mathbb{R}^d où X dépasse u . Ainsi, l'ensemble d'excursion est aléatoire, et sa distribution à un niveau fixe u est déterminée par la distribution de X . Étant des sous-ensembles de \mathbb{R}^d , les ensembles d'excursions peuvent être étudiés en termes de leurs propriétés géométriques afin d'obtenir des informations spécifiques sur les propriétés de distribution des champs aléatoires sous-jacents.

Cette thèse examine :

- (a) comment les mesures géométriques d'un ensemble d'excursion peuvent être estimées à partir d'un échantillon discret de l'ensemble d'excursion, et
- (b) comment ces mesures peuvent être liées aux propriétés distributionnelles du champ aléatoire à partir duquel l'ensemble d'excursion a été obtenu.

Chacun de ces points est examiné en détail dans le Chapitre 1, qui fournit un aperçu global des résultats présentés tout au long du reste de ce manuscrit.

Les mesures géométriques que nous étudions pour les ensembles d'excursion et les sous-ensembles déterministes de \mathbb{R}^d en point (a) sont la mesure de la surface de dimension $(d - 1)$, le *reach*, et le rayon de r -convexité. Chacune de ces quantités peut être liée à la régularité du bord de l'ensemble, qui est généralement difficile à déduire à partir d'échantillons discrets de points. Pour résoudre ce problème, nous apportons les contributions suivantes au domaine de la géométrie computationnelle :

Dans le Chapitre 2, nous identifions le facteur de biais qui correspond aux algorithmes de comptage local pour calculer la mesure de la surface de dimension $(d - 1)$ des ensembles d'excursion sur une grande classe de pavages de \mathbb{R}^d . Le facteur de biais dépend uniquement de la dimension d et non de la géométrie précise du pavage.

Dans le Chapitre 3, nous introduisons un algorithme de comptage pseudo-local pour calculer le périmètre des ensembles d'excursion en deux dimensions. L'algorithme proposé est convergent multigrille (*multigrid convergent* en anglais) et comporte un hyperparamètre réglable pouvant être choisi automatiquement à partir d'informations accessibles.

Dans le Chapitre 4, nous introduisons le β -*reach* en tant que généralisation du *reach*, et l'utilisons pour prouver la cohérence d'un estimateur du *reach* des sous-ensembles fermés de \mathbb{R}^d . De même, nous définissons un estimateur cohérent du rayon de r -convexité des sous-ensembles fermés de \mathbb{R}^d . De nouvelles relations théoriques sont établies entre le *reach* et le rayon de r -convexité.

Nous étudions également comment ces mesures géométriques des ensembles d'excursion sont liées à la distribution du champ aléatoire et spécifiquement à son comportement extrémal. Dans le Chapitre 5, nous introduisons l'*extremal range* : une statistique géométrique locale qui caractérise l'étendue spatiale des dépassements de seuil à un niveau fixe $u \in \mathbb{R}$. La distribution de l'*extremal range* est entièrement déterminée par la distribution de l'ensemble d'excursion au niveau u . Nous montrons comment l'*extremal range* est liée distributionnellement aux volumes intrinsèques de l'ensemble d'excursion. De plus, le comportement limite de l'*extremal range* aux grands seuils est étudié en relation avec la notion de stabilité par rapport aux dépassements de seuil (*peaks-over-threshold stability*) du champ aléatoire sous-jacent. Enfin, la théorie est appliquée à des données climatiques disponibles sur une grille spatiale régulière pour mesurer le degré d'indépendance asymptotique présent et sa variation dans l'espace.

Des perspectives sur la manière dont ces résultats peuvent être améliorés et étendus sont fournies dans le Chapitre 6.

Mots-clés : β -reach, champ aléatoire, dépassement de seuil, *extremal range*, géométrie computationnelle, géométrie stochastique, inférence géométrique, r -convexité, *reach*, volume intrinsèque.

On the geometry of excursion sets: theoretical and computational guarantees

Abstract

The excursion set $E_X(u)$ of a real-valued random field X on \mathbb{R}^d at a threshold level $u \in \mathbb{R}$ is the subset of the domain \mathbb{R}^d on which X exceeds u . Thus, the excursion set is random, and its distribution at a fixed level u is determined by the distribution of X . Being subsets of \mathbb{R}^d , excursion sets can be studied in terms of their geometrical properties as a means of obtaining partial information about the distributional properties of the underlying random fields.

This thesis investigates

- (a) how the geometric measures of an excursion set can be inferred from a discrete sample of the excursion set, and
- (b) how these measures can be related back to the distributional properties of the random field from which the excursion set was obtained.

Each of these points are examined in detail in Chapter 1, which provides a broad overview of the results found throughout the remainder of this manuscript.

The geometric measures that we study for both excursion sets and deterministic subsets of \mathbb{R}^d when addressing point (a) are the $(d - 1)$ -dimensional surface area measure, the reach, and the radius of r -convexity. Each of these quantities can be related to the smoothness of the boundary of the set, which is often difficult to infer from discrete samples of points. To address this problem, we make the following contributions to the field of computational geometry:

In Chapter 2, we identify the bias factor in using local counting algorithms for computing the $(d - 1)$ -dimensional surface area of excursion sets over a large class of tessellations of \mathbb{R}^d . The bias factor is seen to depend only on the dimension d and not on the precise geometry of the tessellation.

In Chapter 3, we introduce a pseudo-local counting algorithm for computing the perimeter of excursion sets in two-dimensions. The proposed algorithm is multigrid convergent, and features a tunable hyperparameter that can be chosen automatically from accessible information.

In Chapter 4, we introduce the β -reach as a generalization of the reach, and use it to prove the consistency of an estimator for the reach of closed subsets of \mathbb{R}^d . Similarly, we define a consistent estimator for the radius of r -convexity of closed subsets of \mathbb{R}^d . New theoretical relationships are established between the reach and the radius of r -convexity.

We also study how these geometric measures of excursion sets relate to the distribution of the random field, and specifically its extremal behavior. In Chapter 5, we introduce the *extremal range*: a local, geometric statistic that characterizes the spatial extent of threshold exceedances at a fixed

level threshold $u \in \mathbb{R}$. The distribution of the extremal range is completely determined by the distribution of the excursion set at the level u . We show how the extremal range is distributionally related to the intrinsic volumes of the excursion set. Moreover, the limiting behavior of the extremal range at large thresholds is studied in relation to the peaks-over-threshold stability of the underlying random field. Finally, the theory is applied to climate data available on a regular spatial grid to measure the degree of asymptotic independence present, and its variation throughout space.

Perspectives on how these results may be improved and expanded upon are provided in Chapter 6.

Keywords: β -reach, computational geometry, extremal range, intrinsic volume, geometric inference, r -convexity, random field, reach, stochastic geometry, threshold exceedance.

Preface

At the start of my PhD, the provisional title of this document, decided by my supervisors Elena Di Bernardino and Thomas Opitz, was *On the characterization of spatio-temporal dependence of extremes of high-resolution gridded data*. Our aim was to bring tools from stochastic geometry to the field of extreme value theory, by means of studying the geometry of the excursion sets of random fields, *i.e.*, the subregions of the index set on which the random field exceeds a certain threshold. Such tools might be of interest to scientists studying large spatial or spatio-temporal datasets, where the available information in space and/or time is dense enough to approximate these excursion sets. Thus, if our results were to be useful in practice, they should be *data-driven* solutions for extracting information about the extremal dependence structure of the random fields modelling the data.

To complete our data pipeline, we had to address two main barriers.

- The first would be to establish methods for extracting information from the discrete data about the continuous random fields that generated them. For this, we based our search on existing results in computational geometry such as [Gray \(1971\)](#); [Coeurjolly and Klette \(2002\)](#); [Meschenmoser and Spodarev \(2010\)](#), and others that study this problem specifically for random geometric objects, such as [Yukich \(2015\)](#); [Biermé and Desolneux \(2021\)](#). The culmination of our efforts adds to the repertoire of scientists and engineers several tools that can be used to extract geometric information from data that describes the excursion sets of random fields. These tools are novel, and are provided with rich theoretical justifications and guarantees. In particular, we provide new results for estimating the surface area, reach, and radius of r -convexity of Euclidean subsets; all of which are widely studied in the literature.
- The second barrier would be to study the geometry of the excursion sets of random fields at high thresholds, and connect this to the random field's extremal dependence structure. This would be greatly assisted by the seminal work of [Adler and Taylor \(2007\)](#), which provides a detailed analysis of the geometry of excursion sets of random fields at non-extreme thresholds. Combined with literature dealing with extreme threshold exceedances of random fields ([Kac and Slepian, 1959](#); [Ferreira and de Haan, 2014](#); [Dombry and Ribatet, 2015](#)), we are provided with a solid background for our analyses. Our newly introduced *extremal range* can be used to examine the asymptotic dependence structure of the random fields in question, completing the pipeline from discrete data to pertinent indicators describing extremal behavior.

At every step in bridging this gap, the geometry of excursion sets takes center stage, both in relation to the random fields that they represent, and the data by which they are represented. The remainder of this document details our study of these relationships.

Acknowledgements

I am proud to have been led by my thesis advisors Elena Di Bernardino and Thomas Opitz. Both Elena and Thomas have been attentive in providing valuable guidance that extends far beyond the scope of simply writing papers and getting published. My positive experience in the south of France—and probably much of the remainder of my career—is largely thanks to the kindness they have shown me.

A special thanks is dedicated to the referees, Pierre Calka and Clément Dombry. Their written reports convey with great clarity the time and consideration that they generously dedicated to reviewing my work. I am honoured to have received such thoughtfully written feedback from both of them.

To those already mentioned, and to each of the remaining jury members, Catherine Aaron, David Cohen-Steiner, Aurèlie Fischer, Céline Lacaux, and Clément Levrard, thank you for having volunteered to actively participate in my thesis defence.

Thank you to the members of my thesis committee, Nicolas Chenavier and Agnes Desolneux, for your insightful feedback, motivating questions, and helpful suggestions that have greatly improved this work.

I would like to thank the anonymous reviewers who provided feedback on our articles, which significantly improved their quality. The generous amount of time that has been dedicated to reviewing this work does not go unnoticed.

There have been, on several occasions, important informal discussions that helped shape the current version of this manuscript. Notably, I would like to acknowledge Eddie Aamari, Mariem Abaach, Andreu Arderiu, Alexis Boulin, Pablo Cañas Castellanos, Rory Cochrane, Aidan Cotsakis, Rahat Dhande, Céline Duval, Athanasios Vasileiadis, Mathijs Wintraecken, and Basil Wong.

I would like to express my deepest gratitude to my friends William Hammersley and Paul Mannix. Their encouragement, insightful conversations, and steadfast friendship have made this journey more manageable and a lot more enjoyable. I am equally indebted to the entire Valrose Ph.D. group for their continual support.

I am profoundly grateful to my partner, Charlotte, who accompanied me through the prosperity and hardship of a Ph.D. with enthusiasm and love.

The writing of this thesis wouldn't have been possible without the unwavering love from my parents. As much as it is mine, this is your accomplishment.

This thesis has been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. This work has been partially supported by the project ANR MISTIC (ANR-19-CE40-0005).

Contents

1	General introduction	1
1.1	Definitions and important notions	1
1.1.1	Geometric measures of excursion sets	3
1.1.2	A regularity condition for the existence of the intrinsic volumes	4
1.1.3	Stationarity and Isotropy	5
1.2	Excursion sets at extreme levels	6
1.2.1	The extremal range	8
1.2.2	Relationship with the intrinsic volumes	9
1.2.3	Application to French temperature data	11
1.3	The problem of discretization	12
1.3.1	The intrinsic volumes of discretized excursion sets	14
1.3.2	The Crofton formula	16
1.3.3	Honeycombs	18
1.3.4	The bias in computing the surface area with an LCA	21
1.3.5	A pseudo-local counting algorithm	25
1.4	The reach and related notions	28
1.4.1	The reach as a geometric quantity	28
1.4.2	Equivalent notions of the reach	29
1.4.3	Obtaining bounds on the reach of a set from a noiseless sample	30
1.4.4	r -convexity and its relation to the reach	31
1.5	Organization of the manuscript	32
2	Surface area and volume of excursion sets observed on point cloud based polytopic tessellations	35
2.1	Introduction	36
2.1.1	Motivations	36
2.1.2	Geometric measures and the Crofton formula	38
2.1.3	Approximated volume and surface area of excursion sets observed over a point cloud	40

2.2	Approximated surface area	44
2.3	Joint central limit theorem for the hypercubic lattice	55
2.3.1	Estimators for the hypercubic lattice	55
2.3.2	The dominant role of the L_1 -norm for the hypercubic lattice	55
2.3.3	Strong alpha mixing random fields	57
2.3.4	Joint Central Limit Theorem	58
2.4	Additional results and proofs associated to Section 2.2	60
2.4.1	Auxiliary lemmas on crossings	60
2.4.2	Proof of Proposition 2.1	64
2.5	Additional results and proofs associated to Section 2.3	65
2.5.1	Proof of Proposition 2.2	65
2.5.2	Auxiliary results for the proof of the joint central limit theorem	68
2.A	Supplementary results	70
2.A.1	Examples	70
2.A.2	Convergence of the bias factor	71
2.A.3	Further elements on the dimensional constant β_d	71
3	On the perimeter estimation of pixelated excursion sets of two-dimensional anisotropic random fields	72
3.1	Introduction	73
3.2	Definitions and Notation	75
3.2.1	Sampling locations on a regular grid	78
3.2.2	Definition of the estimators	79
3.3	Main Results	80
3.3.1	On a fixed domain with decreasing pixel width	81
3.3.2	On a growing domain with decreasing pixel width	82
3.4	Simulation studies	86
3.4.1	A proxy for the true perimeter	87
3.4.2	The anisotropic case	88
3.4.3	The isotropic case	90
3.4.4	Hyperparameter selection	92
3.4.5	Behaviour of the perimeter estimator as a function of the level u	93
3.5	Proofs	93
4	Computable bounds for the reach and r-convexity of subsets of \mathbb{R}^d	104
4.1	Introduction	104
4.2	Definitions and important notions	107
4.2.1	Set dilation, set erosion, and combinations of the two	107
4.2.2	The reach and related concepts	108

4.3	Methods for point cloud data	115
4.3.1	An upper bound for the r -convexity of a set	116
4.3.2	An upper bound for the reach of a set	119
4.3.3	The β -reach profiles of high-dimensional point clouds	124
4.4	Numerical studies	128
4.4.1	Aircraft data	128
4.4.2	Numerical convergence of reach and r -convexity bounds	130
4.5	Proofs and technical results	132
5	A local statistic for the spatial extent of extreme threshold exceedances	138
5.1	Introduction	139
5.2	The extremal range and relevant notations	141
5.3	Linking the extremal range and intrinsic volumes	142
5.4	The asymptotic behavior of the extremal range	144
5.4.1	Non-degenerate limit distributions of the extremal range	144
5.4.2	Connections with the tail dependence coefficient	147
5.4.3	A parametric model for the extremal range	148
5.5	Inference	149
5.5.1	Empirical CDF of the extremal range for stationary fields	149
5.5.2	Consistent estimation of the tail decay rate of the extremal range	150
5.5.3	Extrapolation for non-stationary data using median regression	151
5.6	Application to gridded temperature data	152
5.7	Conclusion	155
5.A	Technical definitions and examples	158
5.A.1	Positive reach	158
5.A.2	A case where asymptotic dependence is not captured by the extremal range	158
5.A.3	Regularly varying random fields	159
5.B	Proofs and auxiliary results	159
5.B.1	Proofs for Section 5.3	159
5.B.2	Proof of Proposition 5.2	162
5.B.3	Proofs and definitions for Section 5.4.1	163
5.B.4	Proof of Proposition 5.6	164
6	Perspectives	166

Chapter 1

General introduction

This manuscript is dedicated to the problem of extracting information about the random phenomena that underlay observations in spatial datasets. We take a geometric approach to this problem, and so our analysis bridges a number of concepts from several areas of mathematics, including stochastic geometry, integral geometry, computational geometry, and extreme value statistics. The principal purpose of this chapter is thus to provide the necessary tools such that the reader may appreciate the results despite a lack of expertise in one or more of these subjects. Throughout this introduction, the main results in each of the following chapters are summarized with the aim of building an intuition around the ideas that led to their development.

Each of the following chapters, namely Chapters 2 through 5, is an independent research article that is either published or submitted for publication in an international, peer-reviewed journal. Section 1.5 provides a brief summary of each article. In addition, a more detailed abstract can be found at the start of each chapter.

1.1 Definitions and important notions

Let us begin by introducing the objects at the center of our study.

Definition 1.1 (Random fields on \mathbb{R}^d). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *real-valued random field on \mathbb{R}^d* is a function $X : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ that is measurable with respect to the product sigma algebra. Throughout this document, the term *random field* is often used and the domain and range should be understood implicitly.

Definition 1.2 (Excursion sets of random fields). Let X be a real-valued random field on \mathbb{R}^d . Let $u : \mathbb{R}^d \rightarrow \mathbb{R}$ be a measurable threshold function that is allowed to vary over \mathbb{R}^d . The *excursion set of X* is the random set

$$E_X(u) := \{t \in \mathbb{R}^d : X(t) > u(t)\}.$$

This makes $E_X(u)$ a random variable with values in $\mathcal{B}(\mathbb{R}^d)$, the Borel subsets of \mathbb{R}^d .

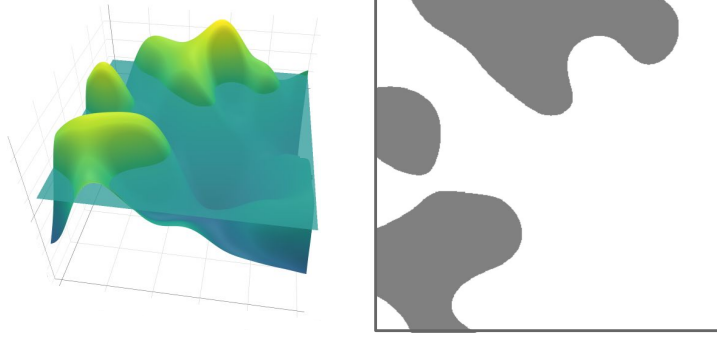


Figure 1.1: A random field (left) and its excursion set (right; shown in grey) at the chosen threshold.

Intuitively, a random field X is a random d -dimensional surface, and its excursion set is the set of locations at which X exceeds a deterministic d -dimensional surface u (see Figure 1.1). In general, u is allowed to vary in space, allowing for “adaptive” thresholds. Often, u is taken to be a constant function, and so in these cases, we make a slight abuse of notation and use u to denote the constant image in \mathbb{R} .

Excursion set theory has been applied in previous works to a wide variety of real-world phenomena, ranging from meteorology and wildlife studies (Frölicher et al., 2018; Lhotka and Kyselý, 2015; Eymard et al., 2000; Longuet-Higgins, 1957; Camp et al., 2022), cosmology (Gott et al., 1990; Gott et al., 2007, 2008; Jennings et al., 2013; Schmalzing and Górski, 1998; Casaponsa et al., 2016), brain imaging (Adler and Taylor, 2011, Section 5), geology (Roubin and Colliat, 2016), and other applications to computer vision applications (Bleau and Leon, 2000; Sezgin and Sankur, 2004).

There are several reasons for the ubiquity of excursion set theory in the study of various physical phenomena. We outline just a few of these reasons.

The first consideration is computer memory, and computational facility. The excursion set $E_X(u)$ of a random field X is sparse in information in the sense that it takes at most n bits of memory to store binary values at a finite set of sites $t_1, \dots, t_n \in \mathbb{R}^d$ corresponding to 1 if $t_i \in E_X(u)$ and 0 otherwise, for $i = 1, \dots, n$. Contrast this with storing a 32 bit floating-point value to approximate $X(t_i)$ for each $i \in 1, \dots, n$. Thus, for large databases or problems that demand several computations of the values at t_1, \dots, t_n , algorithms can be optimized to accept the excursion set as an input, and large gains in efficiency can be expected.

A second practical consideration is that theory regarding excursion sets is well adapted to data for which the entire range of observations is not accurately accessible. This can be due to the low reliability of very large or very small values in data, censoring of the data, or the grouping of sufficiently large or small observations. In such cases, the excursion set is observable for any threshold u in a certain range, while strategies using the full information about the random field might suffer from the censoring, or error in the extremes. Censoring can be a useful strategy to eliminate outliers, or to ignore extreme regions of data highly prone to error—we will see an example

of how discretization error may be large in the extremes in Chapter 5, and how excursion set analysis is used to combat this issue.

A third, crucial point is that the excursion set keeps important information related to the distribution of the underlying random field. There are several works that study how random fields are connected to their excursion sets (see Adler and Taylor, 2007, and references therein). Other authors have established statistical tests of Gaussianity (Spodarev, 2014; Di Bernardino et al., 2017), isotropy (see Definition 1.4; Cabaña (1987); Berzin (2021); Abaach et al. (2023)), and marginal symmetry Abaach et al. (2021) of random fields based entirely on their excursion sets. In Das et al. (2021), excursion sets are shown to be useful in the extrapolation of random fields outside of the domain of observation. Many works in extreme value theory are interested in random fields that exceed high thresholds (see Azaïs and Wschebor, 2009, and references cited), and so in this context, it is quite natural to study the excursion sets of these random fields for large thresholds u .

A final, significant advantage in working with excursion sets is that they open the door to a geometric and topological approach to studying random fields, making them amenable to the use of machine learning techniques. Many statistical procedures used to study spatial random fields exploit only information from bivariate distribution for pairs of locations but fail to capture and utilize key geometric properties that characterize stochastic spatial patterns more precisely and provide information about the parametric models being used. As we will discuss later (see Remark 1.5), there are certain distributional parameters that play a large role in the distribution of various geometric measures of the excursion set; see for example Bierné et al. (2019). A geometric approach to studying random fields allows for new applications of tools in existing stochastic geometry literature. This is done, for example, in Chapters 2 and 5 of this document.

1.1.1 Geometric measures of excursion sets

The geometric features of excursion sets provide concise, informative summaries of their corresponding random fields, and so a mathematical framework for studying these features is essential. The framework that lends itself nicely to our study of random fields is *integral geometry*. At its core, integral geometry is the study of measures on Euclidean subsets that are invariant under rigid motion, and integral transformations between them. Length, surface area, and volume are among the most significant examples of invariant geometric measures. Random sets are constructed by assigning probability measures to Euclidean subsets, giving rise to deep relationships between the fields of integral geometry, geometric probability, and stochastic geometry. These connections are the central focus of the seminal works Matheron (1975) and Santaló (1976) (see also Calka (2019) for an introduction to the subject and Kendall and Molchanov (2009) for a detailed treatment).

Here we outline a major result in integral geometry that provides us with a finite canonical basis of rigid motion invariant functions that we can use to measure subsets of \mathbb{R}^d . To state the result, we let \mathcal{R} denote the ring of polyconvex sets in \mathbb{R}^d , *i.e.*, the smallest set of Euclidean subsets containing the compact, convex sets that is also closed under finite unions. In addition, a function $\phi : \mathcal{R} \rightarrow \mathbb{R}$

is said to be *additive* if for any $A, B \in \mathcal{R}$, it holds that $\phi(A \cup B) = \phi(A) + \phi(B) - \phi(A \cap B)$. *Hadwiger’s characterization theorem* states that any additive, continuous, rigid motion invariant function $\phi : \mathcal{R} \rightarrow \mathbb{R}$ can be expressed as a linear combination of $d + 1$ canonical basis functions, known as the *intrinsic volumes*, $V_i : \mathcal{R} \rightarrow \mathbb{R}$, for $i = 0, \dots, d$.

While not all intrinsic volumes have an immediately intuitive interpretation, three of them (for $d \geq 2$) correspond to familiar geometric features. The first, V_0 , measures the *Euler characteristic* of the set, a topological invariant providing a notion of connectivity. The last two intrinsic volumes, V_d and V_{d-1} , measure the d -volume and half of the $(d-1)$ -dimensional surface area, respectively. This framework can be seen as a higher-dimensional analysis of the level crossings of one-dimensional processes, as in dimension $d = 1$, V_1 corresponds to the sojourn time of the process, and V_0 corresponds to the number of up-crossings—both notions having been extensively studied for several decades (Lévy, 1940; Rice, 1944).

Remark 1.1. We use the term “surface area” generally to refer to the $(d-1)$ -dimensional surface area of objects embedded in \mathbb{R}^d , for $d \geq 3$. More precisely we are referring to the $(d-1)$ -dimensional Lebesgue measure of the boundary. In general, we write \mathcal{L}_k to denote the k -dimensional Lebesgue measure for $k = 0, \dots, d$.

The intrinsic volumes often take center stage in literature concerned with the excursion sets of random fields. They are used, for example, for parametric inference in Bierné et al. (2019); Di Bernardino et al. (2022). A test for Gaussianity based on the Euler characteristic, V_0 , is developed in Di Bernardino et al. (2017). Tests for isotropy based on the surface area, or $2 \times V_{d-1}$, are developed in Berzin (2021); Cabaña (1987); Fournier (2018). The authors of Abaach et al. (2021) implement a test of marginal symmetry based on the statistics of V_{d-1} . In Di Bernardino and Duval (2022), the authors use the excursion volume, V_d , to quantify how a small additive perturbation on the Gaussian random field impacts the intrinsic volumes of the associated perturbed excursion sets. These works are just some of the many that motivate the study of intrinsic volumes, and in particular, their statistical properties when used to measure smooth excursion sets.

The books Adler and Taylor (2007) and Azaïs and Wschebor (2009) provide a detailed, theoretical treatment of the statistics of the intrinsic volumes of smooth excursion sets. The literature on this subject is quite vast. The renormalized Euler characteristic of excursions of Gaussian random fields is shown to be asymptotically normal in Estrade and León (2016) as larger domains are considered. Later, Müller (2017) and Kratz and Vadlamani (2018) show that all of the intrinsic volumes are asymptotically normal in this sense. A comprehensive summary of previous limit theorems of intrinsic volumes of excursion sets is provided in Spodarev (2014).

1.1.2 A regularity condition for the existence of the intrinsic volumes

As we have previously noted, the intrinsic volumes are defined for Euclidean subsets in the polyconvex ring. However, it is quite restrictive to consider random fields whose excursion sets are almost

surely polyconvex. The intrinsic volumes can be defined for a more general subclass of Euclidean sets, namely, the class \mathcal{U}_{PR} .

A set $B \subseteq \mathbb{R}^d$ belongs to \mathcal{U}_{PR} if B can be expressed as a locally finite union of sets with *positive reach*. A set $A \subseteq \mathbb{R}^d$ has positive reach if there exists an $r_0 > 0$ such that all points within a distance r_0 of A have a unique nearest point in A . The supremum of the values of r_0 for which this holds is known as the *reach* of A , denoted $\text{reach}(A)$ (see Definition 4.2). The class \mathcal{U}_{PR} is thus quite large in that it contains all sets with positive reach and thus all compact C^2 -smooth submanifolds of \mathbb{R}^d (Thäle, 2008, Proposition 14). Additionally, since a set A is convex if and only if $\text{reach}(A) = \infty$ (Federer, 1959), the polyconvex ring \mathcal{R} is included in \mathcal{U}_{PR} .

We will not enter into all of the details, but the intrinsic volumes $V_i : \mathcal{R} \rightarrow \mathbb{R}$ have well defined extensions to the more general class \mathcal{U}_{PR} , for which we write $V_i : \mathcal{U}_{PR} \rightarrow \mathbb{R}$; we refer the interested reader to Zähle (1987); Rataj and Zähle (2003); Thäle (2008). It is sometimes assumed in excursion set literature that the excursion sets belong to the class \mathcal{U}_{PR} (Biermé et al., 2019; Cotsakis et al., 2023b), or that the underlying random field is sufficiently smooth such that the excursion set has positive reach almost surely (Biermé and Desolneux, 2020; Cotsakis et al., 2023c; Qiao and Polonik, 2018).

The reach of a set quantifies the regularity of its boundary, and the extent to which different features of the set can be distinguished from each other. In Section 1.4, we elaborate on the reach as a geometric summary for subsets of \mathbb{R}^d , and explore related quantities.

1.1.3 Stationarity and Isotropy

The geometric properties of a subset of \mathbb{R}^d that we study are those that are invariant under isometries. Isometries can include translations, rotations, and combinations of the two. Two common assumptions on the distribution of random fields reflecting this invariance are those of *stationarity* and *isotropy*.

Definition 1.3 (Stationarity). A random field X is said to be *stationary* if, for any finite collection of points in the domain, the joint distribution of X evaluated at these points is invariant to translations of each point by the same vector. The excursion set $E_X(u)$ is said to be stationary if the binary random field $\mathbb{1}_{\{X \geq u\}}$ is stationary.

Remark that if X is stationary, then $E_X(u)$ is stationary for all $u \in \mathbb{R}$. The same line of reasoning holds for isotropy, which is defined as follows.

Definition 1.4 (Isotropy). A random field X is said to be *isotropic* if it is distributionally invariant to rotations about the origin. That is, for any finite collection of points in the domain, the joint distribution of X evaluated at these points is invariant to transforming each point by the same unitary operation. The excursion set $E_X(u)$ is said to be isotropic if the binary random field $\mathbb{1}_{\{X \geq u\}}$ is isotropic.

This a stronger definition of isotropy than the one given by [Adler and Taylor \(2007, Equation \(5.7.1\)\)](#), provided that the random field has finite second moments. The strong isotropy of [Definition 1.4](#) is required to guarantee that the excursion sets be isotropic.

Although these assumptions have clear geometric interpretations, they are too restrictive to apply to many real datasets. For example, consider a meteorological dataset over a region that is large enough to include mountains, plains, forest, and ocean. It would be unreasonable to expect such a dataset to exhibit either stationarity or isotropy, since the distribution of most weather phenomena will vary depending on the topography.

In the chapters that follow, we explore how the assumptions of stationarity and isotropy can be used to construct theoretical and computational guarantees, and what can be done when these assumptions are relaxed.

Now that we are equipped with the geometric measures that we have discussed so far, we have the necessary tools to study random fields via the geometry of their excursion sets. The intrinsic volumes serve as the basis of all additive, continuous, geometric properties of excursion sets. Positive reach provides sufficient conditions for well-posedness of the excursion sets while acting as its own informative summary statistic. These statistics are global in the sense that they accumulate information from the entire excursion set and reduce that information to a single scalar number. In contrast, local geometric statistics can be defined that vary spatially throughout the set. In the next section, we introduce a local geometric statistic for studying excursion sets, and use it to analyze the extremal behavior of the underlying field. Moreover, as we will see, this local statistic has strong links with the intrinsic volumes for excursion sets that have stationary (*i.e.*, translation invariant) distributions.

Section 1.1: Summary

- *Random fields* can be studied through the geometry of their *excursion sets*.
- *Stationarity* and *isotropy* are common assumptions on random fields that reflect a distributional invariance to rigid transformations.
- The *intrinsic volumes* are well defined for sets in the class \mathcal{U}_{PR} , and provide a natural basis for all continuous, additive, rigid motion invariant measures.

1.2 Excursion sets at extreme levels

A topic that we will explore in detail is how excursion set geometry can be used to infer the extremal behavior of random fields. In particular, we are interested in characterizing the extremal dependence structure, which focuses on how very large values of the random field are related across the domain \mathbb{R}^d . In real datasets, these extremes may represent record-breaking temperatures,

devastating floods, or extreme stock market fluctuations, and so understanding their structure is of utmost importance. Extreme value theory (EVT), dating back to Fisher and Tippett (1928), offers tools to model the probability of observing high (possibly unprecedented) quantiles of random variables; see Coles (2001); de Haan and Ferreira (2006) for comprehensive introductions to the subject. These concepts were generalized to random fields in the seminal work of Pickands (1975), in which the notion of max-stable random fields were introduced.

In environmental applications, there may be devastating consequences due to the occurrence of simultaneous or compounding extreme events (Dombry et al., 2018; AghaKouchak et al., 2020; Boulin et al., 2023). It is important in many applications to understand how far-reaching extreme events are, and how this range might change as different levels of “extreme” are considered. Existing approaches can be quite costly computationally, especially performing analyses on very large datasets defined over a dense regular grid. These are precisely the settings in which solutions based on excursion sets thrive. The standard asymptotic models of random fields used in spatial EVT typically exhibit what is known as *peaks-over-threshold (POT) stability* (Ferreira and de Haan, 2014; Dombry and Ribatet, 2015; Thibaud and Opitz, 2015), a notion related to asymptotic dependence which means that the distribution of the random field is “stable” when conditioned to exceed a high threshold at at least one point in the domain. In particular, this implies stability in the distribution of the excursion sets at high thresholds, and so the distribution of the size or extent of extreme events (as measured by the intrinsic volumes for example) is also stable in these models. The rigorous notion of *functional regular variation* is used to characterize random fields for which POT stability arises asymptotically as the threshold level tends to infinity.

Despite the popularity of POT stable models, there is strong empirical evidence that many environmental processes do not possess this property (Tawn et al., 2018; Huser and Wadsworth, 2022). Some flexible models for the extremes random fields have been developed that adapt to data that does not exhibit POT stability, and a few other models are able to flexibly adapt to both situations of asymptotic dependence and independence (Huser et al., 2017; Huser and Wadsworth, 2022; Zhang et al., 2022).

Recent works such as Heffernan and Tawn (2004); Heffernan and Resnick (2007); Wadsworth and Tawn (2022) consider a conditional approach to extreme value theory, where a random vector or a random field is conditioned to exceed a high threshold at a given index or location, and its distribution is studied in the limit as the threshold tends to infinity. This approach is backed by the theory in Dombry and Ribatet (2015), where the limiting distribution is given for a certain class of asymptotically dependent random fields. For a random field X on \mathbb{R}^d , a fixed location $s \in \mathbb{R}^d$, and a constant threshold $u \in \mathbb{R}$, the conditioning event $\{\omega \in \Omega : X(s) \text{ exceeds the threshold } u\}$ is equivalent to $\{\omega \in \Omega : s \in E_X(u)\}$. Hence, excursion sets are a natural object to study in the context of conditional spatial extreme value theory. The relevance of excursion sets in meteorological data and spatial statistics (Sommerfeld et al., 2018; Bolin and Lindgren, 2015) justifies our simplification of performing the following analysis using $d = 2$ for the dimension of the geographical ambient

space.

1.2.1 The extremal range

In Chapter 5, we create a method for measuring the “rate” at which asymptotic dependence is lost in asymptotically independent models, when dependence is known to diminish to 0 at high thresholds. Our method is based entirely on excursion set data, and so in no way do we consider only pairwise observations, thus avoiding a large degree of computational complexity that would arise from studying all possible pairs of locations. In studying the asymptotic dependence properties via excursion sets, our analysis lends itself to a natural indicator that can be used in climate extremes. To be specific about our strategy, we place ourselves in the conditional extreme value theory framework, and study the excursion set, conditioned on containing a certain location in \mathbb{R}^2 . Despite the motivating use case for asymptotically independent random fields, asymptotic independence is not a necessary assumption; that is, our methods can be applied to both asymptotically dependent and independent random fields. To achieve these results, we introduce the *extremal range*, a local summary statistic of the conditioned excursion set, which is defined as follows.

Definition 1.5 (Extremal range). Let $E_X(u)$ denote the excursion set of a random field X on \mathbb{R}^2 . Let $\tilde{R}^{(u)} : \mathbb{R}^2 \times \Omega \rightarrow \mathbb{R} \cup \{\infty\}$ be a random field defined by

$$\tilde{R}^{(u)}(s) := \inf\{\|t - s\| : t \notin E_X(u)\}, \quad (1.1)$$

for $s \in \mathbb{R}^2$, where $\|\cdot\|$ denotes the Euclidean norm. Let the *extremal range* at s at the level u be defined as the conditional random variable,

$$R_s^{(u)} := \tilde{R}^{(u)}(s) \mid \tilde{R}^{(u)}(s) > 0.$$

Remark 1.2. If $E_X(u)$ is almost surely an open set, or if $s \in \mathbb{R}^2$ is almost surely not on the boundary of $E_X(u)$, then the conditioning events $\{\tilde{R}^{(u)}(s) > 0\}$, $\{s \in E_X(u)\}$, and $\{X(s) > u(s)\}$ are equal up to a set of probability 0.

Remark 1.3. The extremal range at the threshold level u is well-defined even in the case where the excursion set $E_X(u)$ is non-stationary and non-isotropic (see Definitions 1.3 and 1.4). Therefore, the distribution of $R_s^{(u)}$ can depend on s .

The extremal range at $s \in \mathbb{R}^2$ can be interpreted as the radius of the largest ball, centered at s , such that exceedance $X(t) > u(t)$ occurs for every t in the ball. The extremal range provides a direct measure of the extent of extreme events at a prescribed level of rarity. Since we are interested in the rarest of events, we study its distributional properties for large threshold functions u . This, in turn, sheds light on the extremal dependence structure of random fields. The extremal range is a real-valued, conditional random variable, which is more tractable to work with than the conditioned

excursion sets themselves. Intuitively, if the distribution of $R_s^{(u)}$ is skewed towards larger values, it means that there is a large spatial dependence between values of the random field X at locations near s at the level u . Similarly, if $R_s^{(u)}$ is distributed near 0, then we expect little spatial dependence of threshold exceedances at the level u .

One can track the evolution of the distribution of the extremal range as the threshold function increases according to the marginal distributions of X . For some $p \in (0, 1)$, the adaptive threshold function $u = u_p$ is taken to be the map of p -quantiles of the margins. If the extremal range at s at the level u_p does not tend to 0 in probability as $p \rightarrow 1$, then the random field is *asymptotically dependent* in the traditional sense; see Proposition 5.4. Equivalently, the extremal range tends to 0 in probability for *asymptotically independent* models.

Remark 1.4. To clarify an important point, we are more interested in the extremal range at extreme quantiles of the random field rather than extreme quantiles of the extremal range itself. In fact, in much of our analysis of the extremal range, we take interest in its median at high quantiles of the marginal distribution of the random field.

The problem that is intrinsic to EVT arises, in that there are very few observations of the random field X exceeding very high thresholds. To address this issue, the extremal range at a location s can be studied at *moderately* large quantiles of the random field at s , and its behavior can be extrapolated to larger, possibly unobserved thresholds.

1.2.2 Relationship with the intrinsic volumes

Despite the extremal range being a local statistic, its distribution can be related to the those of the global intrinsic volumes, for stationary excursion sets satisfying a reach-type condition almost surely (see Assumption 5.1). Hence, the evolution of the intrinsic volumes at high thresholds sheds light on the evolution of the extremal range, which in turn relates to the asymptotic dependence structure of the random field. We highlight some of the main components of this argument, while postponing many of the details to Chapter 5.

Firstly, we must consider the problem of unboundedness. If a stationary excursion set is defined on all of \mathbb{R}^2 , then its excursion sets may be unbounded, and their intrinsic volumes may not be finite. To overcome this issue, we introduce a compact, convex subset $T \subset \mathbb{R}^2$ that contains a neighborhood of the origin, and consider the intrinsic volumes $V_i(E_X(u) \cap T)$, for $i \in \{1, 2\}$. These correspond to the area and half of the perimeter length of the truncated excursion set (including the portion of the boundary of T contained in $E_X(u)$) for $i = 2$ and $i = 1$ respectively. The excursion set is assumed to be stationary in distribution, so in the case of the surface area measure, the expected value of $V_2(E_X(u) \cap T)$ is proportional to the Lebesgue measure of T . With that said, the *density of V_2* defined as

$$V_2^*(u) := \frac{\mathbb{E}[V_2(E_X(u) \cap T)]}{V_2(T)},$$

and is constant for any choice of T . This is not true for the analogous density of V_1 , since the artificial truncation boundary increases the perimeter length of the excursion set. We would like to keep the portion of the boundary of $E_X(u)$ in T and eliminate the portion of the boundary of T in $E_X(u)$. The common strategy is to let T grow to infinity (Schneider and Weil, 2008; Kratz and Vadlamani, 2018; Biermé et al., 2019). For a real number $r \in (0, \infty)$, define $rT := \{r \times t : t \in T\}$ to be T simply scaled by a factor of r . The perimeter of rT scales like r while the surface area scales like r^2 , and so the ratio of perimeter to area tends to 0 (this is true for a more general class of growing domains; see Bulinski et al. (2012, Section 2.3)). This ensures that the limiting density of V_1

$$V_1^*(u) := \lim_{r \rightarrow \infty} \frac{\mathbb{E}[V_1(E_X(u) \cap rT)]}{V_2(rT)}$$

does not suffer from border effects asymptotically. These densities can be calculated directly from the so-called *kinematic formula* (Adler and Taylor, 2007, Theorem 15.9.5), which provides the expected value of each intrinsic volume of the excursion set when intersected with a compact domain. In the case of a Gaussian random field, this expression simplifies to the concise expressions in (Biermé et al., 2019, Equation (11)).

Remark 1.5. Interestingly, for a stationary, Gaussian-type random field X (Di Bernardino et al., 2022) that is sufficiently regular (satisfying Definition 2.2 in Biermé et al. (2019)), the density $V_1^*(u)$ for constant u depends on the correlation function of the random field only through its second spectral moment. Moreover, the density $V_2^*(u)$ is independent of the correlation function and depends only on the marginal distribution at the threshold u , since $V_2^*(u) = \mathbb{P}(X(0) > u)$.

We stated earlier that the rate at which $R_0^{(u_p)} \rightarrow 0$ in asymptotically independent models as $p \rightarrow 1$ is related to the intrinsic volumes. More precisely, if there is a non-degenerate limiting distribution with a well-behaved density near 0, then the quantiles of the extremal range at the level u must scale like $V_2^*(u_p)/V_1^*(u_p)$ as $p \rightarrow 1$ (see Corollary 5.1).

The extremal range can be studied at several, moderately large thresholds, and its behavior at extreme thresholds can be extrapolated to higher thresholds according to an appropriate model. To define such a model, we look to two common examples of random fields.

- For stationary, Gaussian random fields, the ratio $V_2^*(u)/V_1^*(u)$ can be shown to scale asymptotically like the threshold $1/u$ (up to a constant multiple) as $u \rightarrow \infty$. Hence, by previous arguments, we expect the quantiles of the extremal range to decay at the same rate, which is proven in Proposition 5.2. We have thus a model for the asymptotic behavior of the quantiles of the extremal range for Gaussian random fields.
- For random fields exhibiting POT stability, the quantiles of the extremal range stabilize and converge to constants (see Proposition 5.3 for the limiting distribution).

In Chapter 5, we define a parametric model (see Equation (5.10)) that interpolates between these

two behaviors. This allows for the extrapolation of the distributional properties of the extremal range for arbitrarily large threshold exceedances.

There are two main advantages to this approach. The first is that it provides a measure of the degree to which data is asymptotically independent. Indeed, by fitting the model at moderately large thresholds, one gains insight on the rate at which the extremal range tends to 0 in probability as the threshold is increased. Such an approach can be applied to non-stationary data to measure how the degree of asymptotic dependence varies locally. Likewise, one can conduct local, statistical tests for asymptotic dependence of the random field based on the model parameters that best describe real data. The other significant advantage is that we gain a global view of the spatial extent of extreme events, and how this varies in both space and the severity of the extreme event.

In Chapter 5, we suggest using the median of the extremal range to summarize its distribution, which serves as a local indicator for the spatial dependence that has a physical intuition. That is, if there is an extreme event at a site s , then there is a 50% chance that there are extreme events at all other locations within a distance equal to the median extremal range.

1.2.3 Application to French temperature data

We apply our analysis to real temperature data, which generally varies smoothly across space (Perkins et al., 2012). The SAFRAN reanalysis dataset that we use (Vidal et al., 2010) includes twenty-eight years of daily data on a dense 8×8 km grid that spans over continental France. Figure 5.2 depicts the temperature map on June 1st on four different years. Rather than plotting the absolute temperature, Figure 5.2 shows the temperature on uniform margins, so that the indicator between 0 and 1 can be interpreted as the *severity* relative to normal summer temperatures at the same location. Each color in the plot can be thought of the excursion set at the indicated quantile.

Importantly, stationary margins do not imply that the dependence structure of the random field is stationary, and so our analysis relating the extremal range to global intrinsic volume densities does not hold per se. Nonetheless, the extremal range is well defined locally, and we expect its distribution to vary over space.

The extrapolation model: There are several advantages to the model in Equation (5.10) to extrapolate the quantiles of the extremal range at high thresholds. We list the main ones below.

- The model applies to both asymptotically independent and asymptotically dependent models. Gaussian random fields can be modelled as such with $\theta_s = 1/2$, for all s , and random fields that exhibit POT stability are modelled by $\theta_s = 0$. In this way, the parameter θ_s provides a notion of the degree to which the random field is asymptotically independent at s , and it allows interpolating between the common modeling frameworks of classical geostatistics (Gaussian fields) and extreme value theory (asymptotically stable fields).
- The data need not be assumed stationary nor isotropic.

- The model is easy to implement using quantile regression on some positive observations of the extremal range for several moderate values of the threshold u . Many locations can be considered simultaneously using what is known as *generalized additive quantile regression* with the location s and threshold u as covariates.

We perform the strategy using generalized additive median regression to estimate the extrapolation model for the SAFRAN reanalysis data. The same analysis is performed for simulated data from one of the Regional Climate Models routinely used for studies of climate change impact in France, provided by the DRIAS web service (<http://www.drias-climat.fr/>) developed with the support of the French weather service. The resulting model fits are illustrated in Figures 5.3 and Figure 5.4. The resulting model parameter θ_s is plotted in Figure 5.5. Our analysis reveals significant distributional differences between the extremes of the simulated and reanalysis datasets, indicating that this specific Regional-Climat-Model has biases in the spatial extent of extreme weather events.

Section 1.2: Summary of contributions

- The *extremal range* at a point s and a threshold u is defined as the radius of the largest ball centered at s that is completely contained in the excursion set $E_X(u)$, conditioned on this radius being positive (see Definition 1.5).
- In many cases, the quantiles of the extremal range evolve like the ratio of the excursion set area to the excursion set perimeter as the threshold u is increased. The precise relationship is elucidated in Theorems 5.1 and 5.2.
- The locality of the extremal range allows for the use on non-stationary, and non-isotropic random fields.
- The varying degree of asymptotic independence in real French temperature data is measured over space, as shown in Figure 5.5.

1.3 The problem of discretization

In the development of the analysis in the previous section, we have avoided the treatment of a very important consideration. That is, in applications, one does not have access to the exact excursion set, but rather finite data representations. For example, a realization of a random field X on a square grid in \mathbb{R}^2 might be stored in computer memory as a square matrix of floating-point values. The excursion set of the same field can be represented by a single n -bit integer, where n corresponds to the number of elements in the matrix. Extracting the geometric features of the underlying

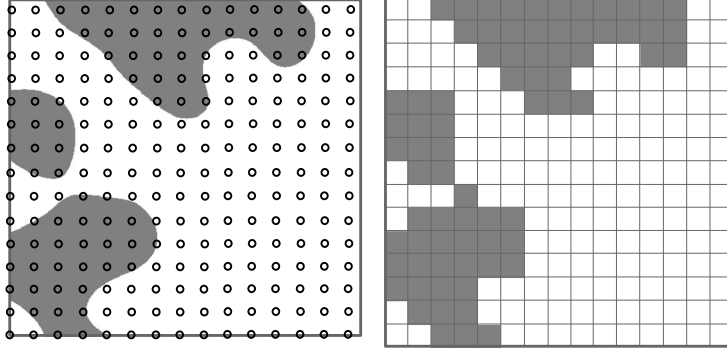


Figure 1.2: An example of Setting 1.1 applied to the excursion set in Figure 1.1. The points $(\xi_i)_{i \in I}$ are arranged in a square lattice. With $A = E_X(u)$, the binary, pixelated image on the right depicts the value of Z_i for each ξ_i , grey corresponding to $Z_i = 1$ and white corresponding to $Z_i = 0$.

excursion set from these data representations is not always straightforward. The importance in understanding excursion set geometry based on discrete data motivates the following problem: *how can one identify and measure the global, geometric features of a mapping on \mathbb{R}^d based on a finite set of local observations?* The answer to this question, of course, depends on a multitude of factors. It depends on

- the nature of the mapping (is the image set all of \mathbb{R} , as in the case of a random field, or is it $\{0, 1\}$, as in the case of an excursion set? Is the mapping random or deterministic?);
- the geometric features of interest;
- the structure of the locations of the observations (are they spaced evenly on a regular grid, or irregularly positioned? Are the observation locations independent of the mapping?);
- and sometimes the quality (is there noise in the observation, or in the locations themselves?).

For these reasons, it is important that we be very clear about which of these problems we aim to tackle.

Setting 1.1. Let $A \subseteq \mathbb{R}^d$. For an index set $I \subseteq \mathbb{N}^+$, let $(\xi_i)_{i \in I}$ be a countable sequence of points in \mathbb{R}^d . To each point, we assign a label in $\{0, 1\}$, such that the label associated to the point ξ_i is given by

$$Z_i := \begin{cases} 1, & \text{if } \xi_i \in A, \\ 0, & \text{if } \xi_i \in A^c, \end{cases}$$

for $i \in I$. Indeed, we can express the set of points paired with their labels as the disjoint union

$$\Xi := \{(\xi_i, Z_i) : i \in I\} = \Xi_{A^c} \sqcup \Xi_A,$$

where

$$\Xi_{A^c} := \{\xi_i : \xi_i \in A^c\} \quad \text{and} \quad \Xi_A := \{\xi_i : \xi_i \in A\}.$$

In general, the set Ξ is known, and one aims to infer geometric information about the set A .

The state of the art in solutions for problems that fall into our general setting belong to the field of computational geometry, which aims to design and analyze algorithms for solving geometric problems. These problems typically involve the manipulation and analysis of geometric objects, such as points, lines, polygons, and other shapes. There are a number of papers that deal with the general problem of geometric inference of shapes from point cloud data. These works are largely classified under the subject of computational geometry (see, *e.g.*, [de Berg et al. \(2008\)](#)). Topological data analysis, pioneered by [Edelsbrunner et al. \(2002\)](#), is a related discipline that focuses more on characterizing the topology of the set under study (see, *e.g.*, [Chazal and Michel \(2021\)](#) for a concise introduction to the subject).

To relate our Setting 1.1 to the discussion of excursion sets, we suppose that for a fixed set of locations ξ_1, \dots, ξ_n , the excursion set $E_X(u)$ (which takes the place of A) is “observed” at each of the points. That is, each ξ_i for $i = 1, \dots, n$ is known to be either inside the excursion set, or outside. This is well illustrated by the example in Figure 1.2. The global, geometric features that we aim to identify from the marked points Ξ are the intrinsic volumes, and the reach. Setting 1.1 does not specify whether A and the points $(\xi_i)_{i \in I}$ are random or deterministic; neither does it demand that the points ξ_i follow any particular structure. The flexibility of Setting 1.1 provides a framework for the study of several related problems.

1.3.1 The intrinsic volumes of discretized excursion sets

A notable reference that identifies methods for computing all of the intrinsic volumes of sets approximated by regular grids is [Klenk et al. \(2006\)](#). This work is criticised and improved upon in [Kiderlen \(2006\)](#) and [Meschenmoser and Spodarev \(2010\)](#). In the aforementioned works, the set in question is deterministic, whereas in [Mrkvička and Rataj \(2008\)](#), the authors design an algorithm for computing all of the intrinsic volumes for stationary random sets from binary digital images of them. Their strategy uses the estimator in [Rataj \(2006\)](#) to compute the d -volume of sets dilated by various radii, and approximates these measurements by a polynomial, which by the Steiner formula ([Federer, 1959](#), Theorem 5.6) is related to the intrinsic volumes. Such a strategy falls into the category of *global shape recognition algorithms* which aim to identify and analyze entire shapes or objects within the image by considering global information about the image as a whole. Contrast this to *local counting algorithms* (LCAs) that operate on individual pixels or small local neighborhoods and accumulate information locally ([Gray, 1971](#); [Miller, 1999](#)). The advantage of these algorithms is their simplicity, which allows for efficient implementation in parallel architecture ([Lindblad, 2005](#)). In [Kampf \(2014\)](#), the authors show that the only intrinsic volumes that can be accurately inferred from LCAs (see, *e.g.*, [Svane \(2014b\)](#), Definition 2.7) on pixelated images are V_d and V_{d-1} .

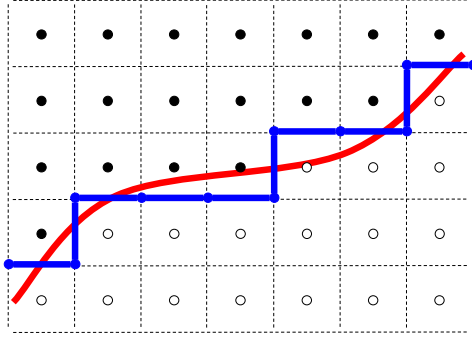


Figure 1.3: An example of a partitioning curve (in red) that defines the boundary of the set $A \subset \mathbb{R}^2$, and separates the points in Ξ_A (in black) from those in Ξ_{A^c} (in white). A local, edge-counting algorithm yields the length of the blue path along the edges, which can be used as a crude estimate of the length of the red curve.

For the moment, restricting our focus to local counting algorithms allows us to identify some interesting characteristics of the problem. When estimating on a regular grid, we say that an algorithm is *multigrid convergent* if the estimation error can be made arbitrarily small by choosing the scale of the grid to be sufficiently small. For example, if each ξ_i for $i \in I$ is assigned to a grid cell (or voxel in the case of a hyper-cubic lattice, as in the case of Figure 1.2), then counting the number of elements in Ξ_A and multiplying by the d -volume of each grid cell provides an estimate for $\mathcal{L}_d(A)$, the d -volume of A . Moreover, this local counting algorithm defines a Riemann sum and so is multigrid convergent. Estimating the intrinsic volume corresponding to the surface area with a local counting algorithm is more involved.

Let us consider the specific case of a square grid in \mathbb{R}^2 and place the ξ_i 's at the center of each cell. Then, the natural estimator of the perimeter of A is the length of the boundary of the union of cells that correspond to points in Ξ_A (see Figure 1.3). It is shown in [Biermé and Desolneux \(2021\)](#) that when A is the excursion set $E_X(u)$ of an isotropic random field (see Definition 1.4 for isotropy), this local counting algorithm is not multigrid convergent. The authors show that the expectation of the estimator converges to $4/\pi$ times the true perimeter of A , and that this same bias factor shows up when considering a hexagonal tiling in \mathbb{R}^2 . This is consistent with [Miller \(1999\)](#), who found the same bias factor when measuring (uniform-)randomly oriented line segments in \mathbb{R}^2 for both square and hexagonal tilings. However, it is noted in [Miller \(1999\)](#) that if the same problem is considered in three dimensions on a cubic lattice, the bias factor is increased to $3/2$.

A major contribution of this thesis is that we unify the results concerning the bias factor of local counting algorithms for computing the surface area over a general class of tilings in arbitrary dimension d . Before presenting the result, its implications, and a sketch of its proof in Section 1.3.4, we start the discussion with several important notions in integral geometry.

1.3.2 The Crofton formula

In the following, we take $M \subset \mathbb{R}^d$ to denote a compact, C^1 -smooth, $(d-1)$ -dimensional, orientable manifold. The orientability of M ensures that a consistent *outward-pointing* normal vector exists at each point in M . The unit $(d-1)$ -sphere centered at the origin in \mathbb{R}^d is denoted S^{d-1} .

Often in practice, data is defined over grids, and so although there may not be a bias in the orientation of the object under consideration, there may be a directional bias in how it is observed. Thus, it may be useful to consider a measure of how much surface area is oriented in a certain direction of interest.

For example, let μ_M be a measure on S^{d-1} such that for any Borel subset G of S^{d-1} ,

$$\mu_M(G) := \mathcal{L}_{d-1}(\{s \in M : n_s \in G\}), \quad (1.2)$$

where n_s is the unit outward-pointing normal vector of M at s . In this way, μ_M provides a selective surface area measure of M , and $\mu_M(S^{d-1}) = \mathcal{L}_{d-1}(M)$.

It may be tempting to employ the measure μ_M in practice when measuring surface area in a directional sense, but in most cases, it is intractable to work with. For instance, if M has a planar subset with non-zero \mathcal{L}_{d-1} measure, then μ_M has an atom at the orientation of the planar subset. If instead M is the sphere S^{d-1} , then μ_M has no atoms and admits a constant density function.

A more useful approach is to consider the function $g_M : S^{d-1} \rightarrow [0, \infty)$, defined by

$$g_M(v) := \int_{S^{d-1}} |\langle v, s \rangle| \mu_M(ds) = \int_M |\langle v, n_s \rangle| \mathcal{L}_{d-1}(ds), \quad v \in S^{d-1}. \quad (1.3)$$

Intuitively, $g_M(v)$ is the \mathcal{L}_{d-1} measure of M once projected onto $(v)^\perp$, the $(d-1)$ -dimensional subspace of \mathbb{R}^d normal to v . Alternatively, remark that $g_M(v)$ is invariant to stretching or compressing M in the direction of v by a non-zero scalar. These intuitions are formalized by the equation

$$g_M(v) = \int_{(v)^\perp} \mathcal{L}_0(\text{proj}_{(v)^\perp}^{-1}(s) \cap M) ds, \quad (1.4)$$

where $\text{proj}_{(v)^\perp}^{-1}(s) = \{s + \lambda v : \lambda \in \mathbb{R}\}$, and \mathcal{L}_0 is the counting measure. Loosely speaking, the quantity $g_M(v)$ gives some notion of the amount of surface area of M biased in the direction v .

The fundamental property of $g_M(\cdot)$ that makes it useful in practice is that its average value over S^{d-1} is proportional to $\mathcal{L}_{d-1}(M)$, and the proportionality constant only depends on the dimension d . That is,

$$\int_{S^{d-1}} g_M(s) ds = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})} \mathcal{L}_{d-1}(M). \quad (1.5)$$

This is a special case of the *Crofton formula*. In this particular form, it relates an object's surface area to the average number of times it is crossed by lines of varying orientations. In terms of the implications of this equation in stochastic geometry, this collection of lines can be interpreted as a

single *random* line, and the expected number of crossings is proportional to the amount of surface area.

Proof of Equation (1.5). Fix $v \in S^{d-1}$. Starting from Equation (1.3),

$$\begin{aligned} \int_{S^{d-1}} g_M(s) ds &= \int_{S^{d-1}} \int_{S^{d-1}} |\langle s, s' \rangle| \mu_M(ds') ds \\ &= \int_{S^{d-1}} \int_{S^{d-1}} |\langle s, s' \rangle| ds \mu_M(ds') \\ &= \int_{S^{d-1}} |\langle s, v \rangle| ds \int_{S^{d-1}} \mu_M(ds') \end{aligned}$$

In this final expression, the latter integral equates to $\mu_M(S^{d-1}) = \mathcal{L}_{d-1}(M)$. The former integral can be computed explicitly to yield (1.5). \square

The full Crofton formula relates not only the \mathcal{L}_{d-1} measure of M and the \mathcal{L}_0 measure of crossings with lines, but more generally it relates any intrinsic volume of an object to any other lower dimensional intrinsic volume of crossings of that object with linear subspaces of appropriately chosen dimension. The interested reader is directed to [Schneider and Weil \(2008, Theorem 5.1.1\)](#) and Chapter 2 of this document.

We suggested earlier that the function g_M might help in understanding the bias when directional biases are introduced. This phenomena is recognized in [Svane and Vedel Jensen \(2017\)](#), and so the authors examine how the Crofton formula may be used to average directional biases over all directions to measure the intrinsic volumes of a set, with applications to local stereology in mind. In many applications, discretization only allows finitely many directions to be considered. In this setting, [Kiderlen and Meschenmoser \(2011\)](#) determines the error in surface area estimates due to this discretization procedure.

In the following, we show how the Crofton formula may be used to explicitly calculate the bias in surface area measurements using LCAs, where the directions of measurement are locally finite. We start by defining the density of g_M in the same way we defined the densities for the intrinsic volumes in Section 1.2.2.

Definition 1.6. Suppose that $M \subset \mathbb{R}^d$ is a random set distributed according to a stationary probability measure. Moreover, suppose that M is almost surely a C^1 -smooth $(d-1)$ -dimensional orientable manifold as before. Let $T \subset \mathbb{R}^d$ be a compact, convex domain. Then for $v \in S^{d-1}$, define

$$g_M^*(v) := \frac{\mathbb{E}[g_{M \cap T}(v)]}{\mathcal{L}_d(T)}.$$

Remark that by stationarity, g_M^* does not depend on T .

A useful interpretation of g_M^* is given by the following lemma.

Lemma 1.1. *Let M be as in Definition 1.6, and $v \in S^{d-1}$. Then $g_M^*(v)$ is the expected number of times M intersects a line segment of length 1, oriented parallel with v .*

Sketch of proof. Write $g_{M \cap T}(v)$ as in Equation (1.4). After taking the expectation, apply Fubini's theorem, and remark that the resulting integral is proportional to the d -volume of T . \square

If M satisfies Definition 1.6 and partitions \mathbb{R}^d into two sets A and A^c , then we can enter into the framework of Setting 1.1. We can then ask, for example, what is the probability that two nearby points ξ_i and ξ_j do not belong to the same partition, *i.e.*, $Z_i \neq Z_j$? If we let $q := \|\xi_j - \xi_i\|$ and $v := (\xi_j - \xi_i)/q$, then under some technical regularity assumptions on M (see Theorem 2.1), it can be shown that

$$\mathbb{P}(Z_i \neq Z_j) = q \times g_M^*(v) + o(q). \quad (1.6)$$

This is very relevant to our discussion about local counting algorithms, because this is precisely the probability of counting the edge between ξ_i and ξ_j when approximating the surface area of M from Ξ .

There is one additional assumption that we will impose on the distribution of the random manifold M . That is, if M is *isotropic*, then for any $v \in S^{d-1}$, and compact, convex $T \subset \mathbb{R}^d$, one has

$$\begin{aligned} g_M^*(v) &= \frac{\int_{S^{d-1}} g_M^*(s) ds}{\mathcal{L}_{d-1}(S^{d-1})} \\ &= \frac{\int_{S^{d-1}} \mathbb{E}[g_{M \cap T}(s)] ds}{\mathcal{L}_{d-1}(S^{d-1}) \mathcal{L}_d(T)} \\ &= \frac{\mathbb{E}[\int_{S^{d-1}} g_{M \cap T}(s) ds]}{\mathcal{L}_{d-1}(S^{d-1}) \mathcal{L}_d(T)} \\ &= \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2}) \mathcal{L}_{d-1}(S^{d-1})} \frac{\mathbb{E}[\mathcal{L}_{d-1}(M \cap T)]}{\mathcal{L}_d(T)} \\ &= \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi} \Gamma(\frac{d+1}{2})} \mathcal{L}_{d-1}^*(M), \end{aligned} \quad (1.7)$$

where we have defined $\mathcal{L}_{d-1}^*(M) := \mathbb{E}[\mathcal{L}_{d-1}(M \cap T)]/\mathcal{L}_d(T)$. The intuition behind this result is that, given that we have no information about the position and orientation of M , the expected number of times M crosses a line segment is proportional to the average ‘‘amount of M ’’ per unit volume. Moreover, we have determined this proportionality constant.

1.3.3 Honeycombs

Here, we introduce a generalization of the types of tilings that are frequently used to perform local counting algorithms to approximate the surface area of $(d-1)$ -dimensional manifolds. We have seen in Section 1.3.1 that several authors have considered the performance of these algorithms on square and hexagonal lattices. Regular polygonal tessellations are instances of a larger, well-studied class

of tessellations known as *Voronoi tessellations*. The definition of a Voronoi tessellation is simple enough that we provide it here for completeness.

Definition 1.7 (Voronoi tessellation). For a countable set of points $(\xi_i)_{i \in I}$ in \mathbb{R}^d , the corresponding *Voronoi tessellation* is a sequence of *Voronoi regions* $(R_i)_{i \in I}$, where R_i is the set of points in \mathbb{R}^d that are at least as close to ξ_i than any other point ξ_j , for $i \neq j \in I$. That is,

$$R_i := \{x \in \mathbb{R}^d : \|x - \xi_i\| \leq \|x - \xi_j\|, \forall j \in I\}, \quad i \in I.$$

Remark 1.6. The hypercubic voxel tessellation in \mathbb{R}^d and the hexagonal tiling in \mathbb{R}^2 are specific examples of Voronoi tessellations. It suffices to take the centers of each cell as the seeds $(\xi_i)_{i \in \mathbb{N}}$.

Suppose that two seeds ξ_i and ξ_j for $i \neq j$ are *neighbors* in a Voronoi tessellation, meaning that the intersection $R_i \cap R_j$ is a non-empty, $(d - 1)$ -dimensional planar surface. In this case, we refer to the intersection $R_i \cap R_j$ as the *facet* separating ξ_i and ξ_j . The facet is easily seen to be normal to the vector spanned between ξ_i and ξ_j , passing through their midpoint. In what follows we drop this last requirement regarding midpoints to define a more general class of tessellations on \mathbb{R}^d .

Definition 1.8. We say that a tessellation of compact, convex polytopes is a *point referenceable d -honeycomb* if there exists a sequence of points $(\xi_i)_{i \in I}$ in \mathbb{R}^d , such that

- Each polytope in the tessellation contains exactly one point in the sequence
- Any line segment contained in the intersection of two distinct polytopes is normal to the vector between the two corresponding points in the sequence $(\xi_i)_{i \in I}$.

If such a sequence $(\xi_i)_{i \in I}$ exists, then the sequence of tuples of polytopes paired with their respective points is called a *point-referenced d -honeycomb*.

Any Voronoi tessellation in \mathbb{R}^d is thus a point-referenced d -honeycomb; and since there is a unique Voronoi tessellation for every sequence of seeds $(\xi_i)_{i \in I}$, it is always possible to construct at least one point-referenced d -honeycomb with $(\xi_i)_{i \in I}$ as the sequence of seeds.

Proposition 1.1. *All edge-to-edge tilings of regular polygons are point-referenceable 2-honeycombs.*

Proof. The sequence of polygon centers satisfies Definition 1.8. □

Figure 1.4 depicts an edge-to-edge tiling of regular polygons. It, and its dual (the Voronoi tessellation of its vertices), are both point-referenceable 2-tessellations.

An important property of point-referenced d -honeycombs is that the distances between reference points and the surface area of the facets can be related to the d -volume occupied by the cells. Referring to the point-referenced 2-honeycomb in Figure 1.5, the area of the shaded region is equal to half of the length of the dashed blue line times the length of the solid blue edge. This is because the shaded region is an *orthodiagonal quadrilateral*. The analogous statement in \mathbb{R}^d is that the

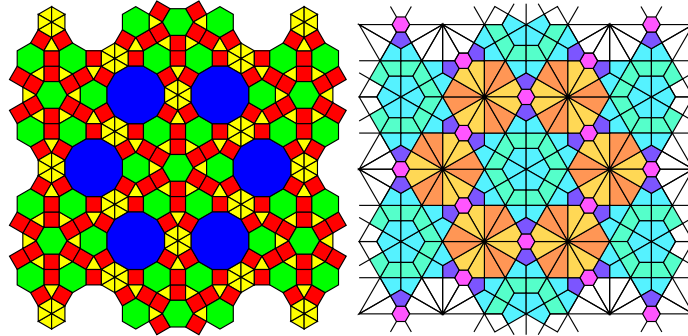


Figure 1.4: On the left is an edge-to-edge tiling of regular polygons. On the right is the Voronoi tessellation of its set of vertices. Both tessellations are point-referenceable 2-honeycombs. Retrieved from [Princeton \(2023a,b\)](#) under the Creative Commons licence.

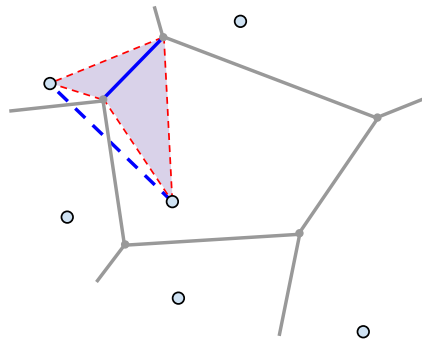


Figure 1.5: The volume of the shaded region is $1/d$ times the length of the dashed blue line times the surface area of the solid blue line. Complements Figure 2.4.

shaded region is the distance between the two chosen reference points times the surface area of the facet between them, divided by d . If we repeat the process for other pairs of neighboring cells in the point-referenced d -honeycomb, the resulting shaded regions themselves form a tessellation. So, by summing over distinct pairs of reference points contained in some compact, convex set $T \subset \mathbb{R}^d$, the higher-dimensional orthodiagonal quadrilaterals approximate T . This is formalized by the following proposition.

Proposition 1.2. *Let $(R_i)_{i \in \mathbb{N}}$ be a sequence of convex polytopes forming a point-referenced d -honeycomb over \mathbb{R}^d with corresponding reference points $(\xi_i)_{i \in \mathbb{N}}$. Suppose that there exists some finite constant that is larger than the diameter of each R_i . Let $T \subset \mathbb{R}^d$ be a compact, convex set containing a neighborhood of the origin, and let $rT := \{rt : t \in T\}$ for $r \in \mathbb{R}^+$. Define the index set $I_{rT} \subset \mathbb{N}$ of “observed points” by*

$$I_{rT} := \{i \in \mathbb{N} : \xi_i \in rT\}.$$

Then, as $r \rightarrow \infty$, one has

$$\sum_{\substack{i, j \in I_{rT} \\ i \neq j}} \|\xi_j - \xi_i\| \times \mathcal{L}_{d-1}(R_i \cap R_j) = d\mathcal{L}_d(rT) + \mathcal{O}(r^{d-1}). \quad (1.8)$$

The proof of Proposition 1.2 is contained in the proof of Theorem 2.2.

1.3.4 The bias in computing the surface area with an LCA

We are now ready to state our main result that unifies the analyses of local counting algorithms on grids of various geometries. The general geometry that we consider is the point-referenced d -honeycomb in Proposition 1.2. The asymptotic Equation (1.8) relates the fixed point-referenced d -honeycomb $\{(R_i, \xi_i) : i \in \mathbb{N}\}$ to a growing domain rT . An equivalent interpretation is that the domain of observation T stays fixed, while the point-referenced d -honeycomb $\{(R'_i, \xi'_i) : i \in \mathbb{N}\}$, with $\xi'_i := \xi_i/r$ and $R'_i := \{t/r : t \in R_i\}$, scales towards the origin as $r \rightarrow \infty$. Rewriting (1.8) accordingly, one obtains

$$\sum_{\substack{i, j \in I_{rT} \\ i \neq j}} \|\xi'_j - \xi'_i\| \times \mathcal{L}_{d-1}(R'_i \cap R'_j) = d\mathcal{L}_d(T) + \mathcal{O}(r^{-1}). \quad (1.9)$$

Now, with the geometry established, we return to the setting of Definition 1.6, where M is a random, stationary $(d-1)$ -dimensional manifold. Moreover, we assume that M partitions \mathbb{R}^d into two sets A and A^c , so that we are in Setting 1.1. Our aim is to identify the bias in using a local counting algorithm to compute the \mathcal{L}_{d-1} measure of $M \cap T$ from the labelled reference points

$(\xi'_i)_{i \in I_{rT}}$. The output of the local counting algorithm is written

$$\widehat{\mathcal{L}}_{d-1}(M; r, T) := \sum_{\substack{i, j \in I_{rT} \\ i \neq j}} |Z'_i - Z'_j| \mathcal{L}_{d-1}(R'_i \cap R'_j), \quad (1.10)$$

with the labels denoted as

$$Z'_k := \begin{cases} 1, & \text{if } \xi'_k \in A, \\ 0, & \text{if } \xi'_k \in A^c, \end{cases}$$

for $k \in I_{rT}$.

Theorem 1.1 (See also, Theorem 2.2). *If M is isotropic, and (1.6) holds, then, under the conditions of Proposition 1.2,*

$$\mathbb{E} \left[\widehat{\mathcal{L}}_{d-1}(M; r, T) \right] \xrightarrow[r \rightarrow \infty]{} K_d \mathbb{E} [\mathcal{L}_{d-1}(M \cap T)],$$

with

$$K_d := \frac{d \Gamma(\frac{d}{2})}{\sqrt{\pi} \Gamma(\frac{d+1}{2})}.$$

Proof. Let us begin by taking the expectation of $\widehat{\mathcal{L}}_{d-1}(M; r, T)$. We use (1.6) to write,

$$\begin{aligned} \mathbb{E} \left[\widehat{\mathcal{L}}_{d-1}(M; r, T) \right] &= \sum_{\substack{i, j \in I_{rT} \\ i \neq j}} \mathbb{P}(Z'_i \neq Z'_j) \mathcal{L}_{d-1}(R'_i \cap R'_j) \\ &= \sum_{\substack{i, j \in I_{rT} \\ i \neq j}} (g_M^*(v_{ij}) + \epsilon_{i,j}(r)) \|\xi'_j - \xi'_i\| \mathcal{L}_{d-1}(R'_i \cap R'_j), \end{aligned} \quad (1.11)$$

where $v_{ij} := (\xi_j - \xi_i) / \|\xi_j - \xi_i\|$, and the $\epsilon_{i,j}$'s are functions that tend to 0 as $r \rightarrow \infty$. Now, we exploit the assumption that M is isotropic to argue that $g_M^*(v_{ij})$ is constant for all terms in the sum. Moreover, by (1.7), this constant can be written in terms of d and $\mathcal{L}_{d-1}^*(M)$. The functions $\epsilon_{i,j}(r)$ converge to 0 uniformly over i and j , since the points ξ_i and ξ_j are at a distance of at most twice the upper bound on the diameter of the polytopes in $(R_k)_{k \in \mathbb{N}}$. Hence,

$$\begin{aligned} \lim_{r \rightarrow \infty} \mathbb{E} \left[\widehat{\mathcal{L}}_{d-1}(M; r, T) \right] &= \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi} \Gamma(\frac{d+1}{2})} \mathcal{L}_{d-1}^*(M) \times \lim_{r \rightarrow \infty} \sum_{\substack{i, j \in I_{rT} \\ i \neq j}} \|\xi'_j - \xi'_i\| \mathcal{L}_{d-1}(R'_i \cap R'_j) \\ &= \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi} \Gamma(\frac{d+1}{2})} \mathcal{L}_{d-1}^*(M) \times d \mathcal{L}_d(T) \\ &= \frac{d \Gamma(\frac{d}{2})}{\sqrt{\pi} \Gamma(\frac{d+1}{2})} \mathbb{E} [\mathcal{L}_{d-1}(M \cap T)]. \end{aligned}$$

□

Theorem 1.1 illustrates that the act of approximating a stationary, isotropic, $(d-1)$ -dimensional

surface by the facets of a point-referenced d -honeycomb induces a bias. Quite remarkably, this bias is independent of the precise geometry of the d -honeycomb, and an explicit expression for the bias is provided in any dimension d . In dimensions 1, 2, and 3, the bias factor K_d takes the values 1, $4/\pi$, and $3/2$ respectively.

Implications and related analyses

Theorem 1.1 can be seen to be a generalization of the expected biases for square and hexagonal tessellations in [Biermé and Desolneux \(2021, Proposition 5\)](#), and the results on first moments in [Miller \(1999\)](#).

Remark that Voronoi tessellations in dimension d are point-referenced d -honeycombs. Theorem 1.1 treats A as a random set, and the tessellation of points as being fixed. However, the analysis in the proof of Theorem 1.1 can be slightly modified to show that the bias is the same when A is deterministic, and the point-referenced d -honeycomb is a homogeneous Poisson-Voronoi tessellation (see [Corollary 2.1](#)).

In an earlier work, [Yukich \(2015, Section 2.2\)](#) remarks that in the same Poisson-Voronoi setting, the bias factor is independent of the target set A . Later, [Thäle and Yukich \(2016\)](#) points out that in any dimension, “the surface area asymptotics involve a universal correction factor”. The correction factor identified by Theorem 1.1 is not explicitly computed in the aforementioned references.

Over a span of over two decades, another cluster of related papers has taken an interest in computing distances along Delaunay triangulations, that is, the geometric graphs with edges between nodes if their corresponding Voronoi cells are adjacent. It was first shown in [Baccelli et al. \(2000, Theorem 3\)](#) that approximating the length of a straight line by the distance along a Delaunay triangulation on \mathbb{R}^2 results in a bias factor of $4/\pi$ when the points are randomly decided by a Poisson point process. This result has applications in telecommunication systems, in which messages pass between network stations from the sender to the receiver. Recently, [Edelsbrunner and Nikitenko \(2021, Equation \(1\)\)](#) gives an expression of this correction factor in higher dimensions. Interestingly, it is the same bias factor of K_d in Theorem 1.1.

Figure 1.6 illustrates the relationship between the two problems in two dimensions. Notice that the red path along the Delaunay triangulation and the blue path along Voronoi edges deviate to a similar extent from the straight line passing from s to t in the diagram. To gain some intuition as to why these two seemingly unrelated problems yield the same bias factor, consider a local counting algorithm that counts edges in the Delaunay tessellation if the corresponding facet in the Voronoi tessellation is intersected by the straight line. Since the right angle property is preserved when considering the dual tessellation, the analysis follows similarly to the proof of Theorem 1.1, and the same bias factor is obtained in all dimensions d . The importance of this right angle property is noted in the discussion of [Edelsbrunner and Nikitenko \(2021\)](#), in which the authors also note that *power diagrams* (see, e.g., [Aurenhammer \(1987\)](#)) possess this property.

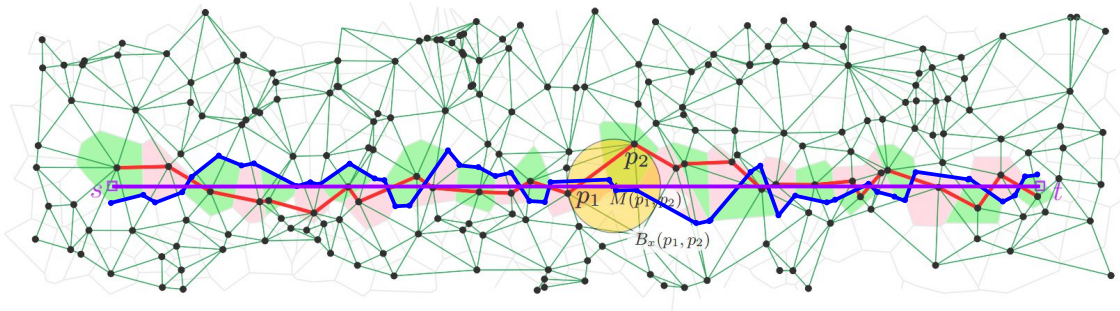


Figure 1.6: The red path along the Delaunay edges that correspond to the Voronoi edges that are intersected by a line segment has the same expected length as the blue path along the Voronoi edges that correspond to the Delaunay edges that are intersected by the line segment. Adapted from [de Castro and Devillers \(2017\)](#) with permission from Springer Nature.

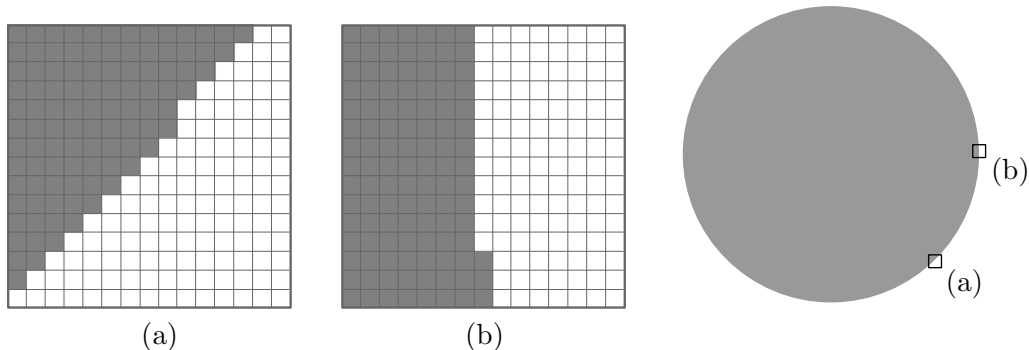


Figure 1.7: The bias factor of the perimeter length estimator in Equation (1.10) is roughly $\sqrt{2}$ (resp. 1) for a smooth set yielding the discretization in (a) (resp. (b)). The orientation of the smooth sets underlying the images in (a) and (b) can be selected by choosing a point on the edge of a circle, as shown.

The bias factor K_d on a hyper-cubic lattice

The bias K_d is the ratio of the surface area of a d -dimensional hypercube to the surface area of the inscribed hypersphere. This can be easily verified by computing the ratio explicitly. Assessing the performance of the local counting algorithm on a hyper-cubic tessellation provides a better intuition as to why this is.

Let us first consider the two-dimensional case. Figure 1.7 shows two possible resulting partitions of pixels when the partition boundary M is a straight line. In panel (a) of Figure 1.7, it is quite clear that the bias in the length estimate of M is a factor of approximately $\sqrt{2}$ too large. However, in panel (b) of the same figure, the resulting estimator is much closer to the true length of the partitioning line.

Selecting the orientation of M uniformly at random is equivalent to selecting a point on the boundary of the circle in Figure 1.7 uniformly at random. Using the local counting algorithm to

estimate the circumference of the circle accumulates local contributions along the perimeter. Thus, the total estimate, for sufficiently fine grid structure, converges to the integral of the bias factor corresponding to each point along the perimeter of the circle. It is not hard to check that the local counting algorithm converges to the perimeter length of the bounding square. Dividing by the length of the interval of integration—the true circumference—yields the average bias factor. Essentially, by considering the performance of the local counting algorithm on the entire circle, we are considering all equally weighted orientations of the perimeter simultaneously. This argument is easily generalized to higher dimensions.

There have been several works that present pixel and voxel based LCAs for estimating the surface area (Ziegel and Kiderlen, 2010; Lindblad, 2005) as well as other intrinsic volumes (Svane, 2014a) by considering weighted averages of various voxel-pair configurations. The analysis that we have provided in Section 1.3 suggests how the bias in the resulting estimator is related to the choice of weights, by considering the overall contribution of each voxel-pair configuration separately. For stationary random sets, this amounts to computing g_M^* in the orientation of the considered voxel pair, the distance between the voxel pair, and the frequency at which the voxel pair is repeated throughout space.

Higher moments

In the proof of Theorem 1.1, we used Fubini’s theorem to distribute the expectation to each term in the sum of Equation (1.11). This allows us to consider the expected contribution of each edge, without reference to the positions and orientations of the other edges in the tessellation. In a way, considering only the first moments allows one to abstract the global geometry of the tessellation, and utilize only the key property of orthogonality between pairs of neighboring reference points and the corresponding facet between them. An analysis of the higher moments would be much more involved. Nonetheless, we prove that the LCA estimator in Equation (1.10) obeys a central limit theorem in the case where the point-referenced d -honeycomb is a hypercubic lattice and where the random field obeys certain mixing conditions.

1.3.5 A pseudo-local counting algorithm

We begin with a lemma that motivates a new, multigrid convergent surface area estimation strategy.

Lemma 1.2 (Donchian and Coxeter (1935)). *Suppose that $M \subset \mathbb{R}^d$ is contained in a $(d - 1)$ -dimensional hyperplane and has finite \mathcal{L}_{d-1} measure almost surely. Then, for any set of orthonormal vectors e_1, \dots, e_d , one has*

$$\left(\sum_{k=1}^d g_M(e_k)^2 \right)^{\frac{1}{2}} = \mathcal{L}_{d-1}(M), \quad (1.12)$$

almost surely.

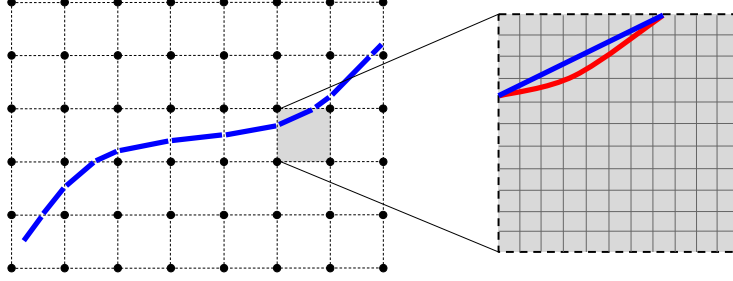


Figure 1.8: The blue, piecewise linear path $\bigcup_{j \in J} N_j$ approximates the red curve M in Figure 1.3. The total length of $\bigcup_{j \in J} N_j$ can be computed via Equation (1.14).

Proof of Lemma 1.2. The unit normal vector to M at the point $s \in M$ is independent of s . And so if we denote this vector by n_M , then by (1.3), one has

$$\sum_{k=1}^d g_M(e_k)^2 = \sum_{k=1}^d |\langle e_k, n_M \rangle|^2 \mathcal{L}_{d-1}(M)^2 = \|n_M\|^2 \mathcal{L}_{d-1}(M)^2 = \mathcal{L}_{d-1}(M)^2,$$

almost surely, and the statement holds by taking square roots. \square

Now, let us return to the framework where M is a smooth $(d-1)$ -dimensional manifold, not necessarily flat, with finite \mathcal{L}_{d-1} measure. In addition, suppose that M is observed over a hypercubic lattice using our local counting algorithm. By only considering only pairs aligned with one lattice direction parallel to the unit vector v , our counting algorithm will approximate $g_M(v)$. Thus, the d -tuple $(g_M(v_1), \dots, g_M(v_d))$ can be approximated by performing the counts separately for each lattice direction v_1, \dots, v_d . Unfortunately, this d -tuple is not sufficient to calculate $\mathcal{L}_{d-1}(M)$ in general. However, in the special case that M is contained in a $(d-1)$ -dimensional hyperplane, Lemma 1.2 applies.

Recall that our goal of this section is to build a multigrid convergent estimator for the surface area of excursion sets. Lemma 1.2 provides such a strategy for hyperplanar surfaces, but this cannot be applied to the excursion set directly. Even worse yet, Lemma 1.2 fails when M is piecewise planar; the surface area of each planar component of M would have to be calculated *individually*, due to the non-linearity of Equation (1.12). This is greatly simplified in dimension 2, where a piecewise linear approximation of a C^1 smooth curve M can be constructed, and Equation (1.12) can be applied to each line segment.

In dimension 2, the construction of a piecewise linear approximation of M is relatively straightforward. Given a square lattice in \mathbb{R}^2 , suppose that the edges of the square lattice cut the curve M into segments. Each segment can then be approximated by the straight line from the entry point in the square cell, to the exit point. In the left panel of Figure 1.8, we show the resulting piecewise linear curve that approximates the smooth red curve from Figure 1.3.

Consider the smooth red curve in Figure 1.3. If each cell of the square lattice is decomposed further into smaller square cells (by refining the grid and adding points to Ξ), then Equation (1.12) can be applied within the larger cell to obtain an approximation of the vector $(g_{M_j}(e_1), g_{M_j}(e_2))$, where M_j is the portion of the smooth red curve in the large cell with index j in the index set J enumerating the larger cells. In the right-hand side of Figure 1.8, the blue line segment that approximates M_j can be labeled N_j , and we see that, in this case,

$$(g_{M_j}(e_1), g_{M_j}(e_2)) = (g_{N_j}(e_1), g_{N_j}(e_2)) \quad (1.13)$$

Equation (1.13) fails in general if M_j is anywhere tangent to e_1 or e_2 , but this is a minor detail since, in these cases, the definition of N_j can be modified accordingly such that Equation (1.13) holds. By Lemma 1.2,

$$\sum_{j \in J} \sqrt{g_{M_j}(e_1)^2 + g_{M_j}(e_2)^2} = \sum_{j \in J} \mathcal{L}_1(N_j) = \mathcal{L}_1 \left(\bigcup_{j \in J} N_j \right). \quad (1.14)$$

There is a trade-off in how the sizes of the larger cells are chosen for a given pixel width. If many large cells are used, then the discrepancy between the length of M and the length of the union of the N_j 's is mitigated. However, if too many are used, and too few pixels are included when computing Equation (1.12) in each large cell, there are larger discretization effects in each computation. Thus, the size of the cells on which Equation (1.12) is computed is a hyperparameter of the plug-in estimator for the length of M . A method for automatically choosing this hyperparameter based on available information is described in Section 3.4.4.

Chapter 3 covers the analysis of the estimator in Equation (1.14) in detail. Theorem 3.1 provides the multigrid convergence of the estimator for non-isotropic, non-stationary excursion sets, along with the rate of convergence. It is shown that the estimator can be used to estimate the perimeter of excursion sets on growing domains in \mathbb{R}^2 , so long as the pixel size decreases sufficiently fast (see Proposition 3.2). Finally, the asymptotic normality of the estimator is proven in Theorem 3.2.

The degree to which it is possible to approximate the manifold M by a union of non-interfering line segments, is well-summarized by the reach of M . The reach of M is used to determine the sufficient rate at which the pixel density increases in the case of a growing domains (see Definition 3.5). In addition, an accurate estimate of $\text{reach}(M)$ would greatly assist in choosing the size of the line segments used to approximate M . In the next section, we explore methods for obtaining such an estimate from the discrete data in Ξ .

Section 1.3: Summary of contributions

- For a $(d - 1)$ -dimensional random surface M that is both stationary and isotropic, the probability that M intersects a line segment of length $q \ll 1$ is, to first order, q times $\mathcal{L}_{d-1}^*(M)$ times a dimensional constant (see Equation (1.6) and Theorem 2.1).
- Using the LCA in Equation (1.10) to estimate the surface area of a stationary, isotropic excursion set on a point-referenced d -honeycomb (see Definitions 1.8 and 2.2) induces a bias factor of K_d , depending only on the dimension d (see Theorems 1.1 and 2.2). It obeys a central limit theorem for a growing domain of observation when observed over a hypercubic lattice (see Theorem 2.3).
- With no assumptions of stationarity and isotropy, the newly introduced pseudo-local counting algorithm in Equation (1.14) is multigrid convergent in dimension 2 with a known rate of convergence (see Theorem 3.1). By assuming stationarity, the estimator is asymptotically normal in the case of a growing domain (see Theorem 3.2).
- An implementation of the perimeter length estimator in \mathbb{R} is made available on [GitHub](#) along with other useful functions for analyzing binary digital images.

1.4 The reach and related notions

1.4.1 The reach as a geometric quantity

The reach, as mentioned before in Section 1.1.2, is a geometric property of sets that characterizes the smoothness of the set's boundary; see Definition 4.2. The reach in itself is an interesting quantity to study, not only for its useful characterization of sets with well-defined intrinsic volumes, but also its relation to what is known as the Steiner formula (Federer, 1959, Theorem 5.6). The Steiner formula states that the d -volume of a set A enlarged by a radius r (by considering all points within a distance r of A) is polynomial in r up to $\text{reach}(A)$. Moreover, the coefficients of this polynomial depend on the set A only through the intrinsic volumes of A . In geometric data analysis, the reach is also known as *the condition number* since it provides a measure of regularity of a set's topological boundary.

Several generalizations of the reach have been developed in existing literature. One such generalization, the β -reach, is introduced in Cotsakis (2023), included as Chapter 4 of the present manuscript. Other examples include the *spherical distortion radius* (Aamari et al., 2023), the μ -reach (Chazal et al., 2009a), and the λ -reach (Chazal and Lieutier, 2005a). Each of these generalized notions is expressed as an infimum or supremum over some set parametrized by some real parame-

ter, and identified with the reach for a particular value of the tuning parameter (see Remark 4.3). Typically, these generalized notions are used as a means of estimating the reach of a set from discrete samples, an endeavour that has been treated extensively in the literature. In fact, we dedicate Chapter 4 to this area of study.

The reach of a set, like the intrinsic volumes, can be seen as its own geometric summary of sets in \mathbb{R}^d . Indeed, a useful interpretation as pointed out in Aamari et al. (2019) is that the reach is determined by either a region of high curvature on the boundary, or a “bottleneck structure” where the set approaches itself closely. Loosely speaking, the reach provides a notion of the scale at which each part of the set is distinguishable from other parts. Although the reach is not additive as defined in Section 1.1.1, it serves as a useful measure when summarizing aspects of excursion set geometry.

1.4.2 Equivalent notions of the reach

There are several equivalent formulations of the reach. Here we outline the most common ones:

- **Distance to medial axis.** The *medial axis* of A^c is the set of points in A^c having more than one closest point in A . The reach of A is the shortest distance between A and the medial axis of A^c .
- **Federer’s tangent space formulation.** The *tangent vectors* of A at a point $a \in A$, namely $\text{Tan}(A, a)$, are the vectors $u \in \mathbb{R}^d$ such that for $\epsilon > 0$, the distance from $a + \epsilon u$ to A is $o(\epsilon)$. From Federer (1959, Theorem 4.18), the inverse of the reach of a closed set A can be expressed as

$$\text{reach}(A) = \inf \left\{ \frac{\|b - a\|^2}{2 \text{dist}(b - a, \text{Tan}(A, a))} : a, b \in A, a \neq b \right\}. \quad (1.15)$$

- **Geodesic distance.** For a closed set $A \subset \mathbb{R}^d$ and $a, b \in A$, let $d_A(a, b)$ denote the infimum of lengths of paths between a and b in A . Then, by Boissonnat et al. (2019, Theorem 1),

$$\text{reach}(A) = \sup \left\{ r \in \mathbb{R}^+ : \forall a, b \in A, \|b - a\| < 2r \Rightarrow d_A(a, b) \leq 2r \arcsin \frac{\|b - a\|}{2r} \right\}. \quad (1.16)$$

- **Global and local decomposition.** The *weak feature size* of a set A is the minimum distance from A to the set of critical points of the generalized gradient of the function $\text{dist}(\cdot, A)$. If A is a C^2 smooth submanifold in \mathbb{R}^d , then $\text{reach}(A)$ is the minimum of the weak feature size of A and the minimal radius of curvature of A (Aamari et al., 2019, Theorem 3.4).

A major contribution of Chapter 4 is a new expression for the reach. It is shown in Boissonnat et al. (2019, Lemma 1) that for a set A of positive reach, a non-increasing function of $\text{reach}(A)$ limits the distance between A and midpoints of certain pairs of points in A . Expanding upon this result, our Theorem 4.2 shows that the reach may equivalently be expressed as follows.

- **Midpoint distances.** The reach of a closed set A is given by

$$\text{reach}(A) = \inf \left\{ g_{\|b-a\|} \left(\text{dist} \left(\frac{a+b}{2}, A \right) \right) : a, b \in A \right\}, \quad (1.17)$$

where the function $g_{(\cdot)}(\cdot)$ is given in Definition 4.5.

A number of authors have leveraged the various expressions of the reach to define several, more general quantities. For instance, the μ -reach and the λ -reach, for real parameters λ and μ , are obtained by computing the shortest distance from A to a filtered version of the medial axis of A^c , parametrized by either μ or λ Chazal et al. (2009a); Chazal and Lieutier (2005a). The spherical distortion radius, introduced in Aamari et al. (2023), takes the geodesic distance formulation of the reach and weakens the condition by only considering points a and b in A if $\|b - a\| \geq \delta$ for some parameter $\delta > 0$.

Similarly, the midpoint distances formulation of the reach introduced in Equation (1.17) leads to a relaxation of the reach if one only considers pairs of points a and b satisfying

$$\text{dist} \left(\frac{a+b}{2}, A \right) \geq \beta$$

for some parameter $\beta \geq 0$. The resulting quantity is called the β -reach (see Definition 4.6). A significant advantage of the β -reach is its computability for point cloud data. Referring to the context of Setting 1.1, the β -reach of Ξ_A is close to the β -reach of A for β much greater than the Hausdorff distance from Ξ_A to A . Moreover, the exact β -reach of a countable set of points Ξ_A can be calculated numerically for all β in an interval; this is shown on real data in Example 4.8.

Intuitively, the β -reach excludes small-scale features when computing the reach, which leads to a quantity that is larger than the reach for $\beta > 0$. The β -reach, seen as a function of β , is continuous at $\beta = 0$, implying that as smaller features of A are counted, the β -reach converges to the reach.

1.4.3 Obtaining bounds on the reach of a set from a noiseless sample

Existing methods

There has been a focus in recent literature on the estimation of the reach based on a noiseless sample of points Ξ_A in Setting 1.1. The statistical framework of Aamari et al. (2019) and Aamari and Levrard (2019) treats Ξ_A as a random set distributed on A , and both works suggest estimating $\text{reach}(A)$ based on the formulation of the reach in Equation (1.15), using pairs of points $a, b \in \Xi_A$. These works obtain bounds on the minimax rate of convergence for this estimator when A is assumed to be a C^3 -smooth manifold without boundary whose tangent spaces are known. Later, Aamari et al. (2023) establishes an optimal convergence rate for minimax estimators of the reach of C^k -smooth submanifolds of \mathbb{R}^d without boundary assuming that the tangent spaces are known. In Berenfeld et al. (2022), the authors employ the formulation of the reach in Equation (1.15), along

with the *convexity defect function* from Attali et al. (2013), to define an estimator for the reach of C^k -smooth manifolds for $k \geq 3$.

In a more general setting, where the set A is not necessarily a manifold, but assumed to have positive reach, Cholaquidis et al. (2023) introduces a tractable method for estimating $\text{reach}(A)$ from the set of sample points Ξ_A . Their approach uses the formulation of the reach in Equation (1.16), and approximates the geodesic distance function d_A by first constructing an ϵ -neighborhood graph over the points in Ξ_A and computing the distance along the graph. The choice of the parameter ϵ is important for guaranteeing convergence rates of the estimator to the true reach.

A converging upper bound for the reach

The advantage of the expression in Equation (1.17) is that there is a very natural extension to estimating $\text{reach}(A)$ from a set of sample points Ξ_A with no assumptions on A other than closedness. If the Hausdorff distance between the sets Ξ_A and A is known to not exceed $\epsilon > 0$, then

$$\text{reach}(A) \leq \inf \left\{ g_{\|b-a\|}(x - \epsilon) : a, b \in \Xi_A, x = \text{dist} \left(\frac{a+b}{2}, \Xi_A \right) \geq \epsilon \right\}.$$

Moreover, it is shown in Theorem 4.4 that this upper bound tends to $\text{reach}(A)$ as $\epsilon \rightarrow 0$, and the rate of convergence is provided.

1.4.4 r -convexity and its relation to the reach

Another geometric quantity closely related to the reach is the radius of r -convexity, denoted $\text{rconv}(A)$ (see Definition 4.4). It is known that the reach of a compact set is bounded above by the radius of r -convexity (Cuevas et al., 2012), *i.e.*,

$$\text{reach}(A) \leq \text{rconv}(A). \tag{1.18}$$

Equation (1.18) provides an alternative approach to bounding the reach from a sample of points, in that an upper bound on the radius of r -convexity also bounds the reach from above. Moreover, in Chapter 4, we provide several sufficient conditions for the reach of a set to be equal to its radius of r -convexity. The weakest of these assumptions is given in Theorem 4.1, where we show the equality of $\text{reach}(A)$ and $\text{rconv}(A)$ for any closed set A whose topological boundary is a C^1 -smooth, $(d-1)$ -dimensional manifold without boundary.

Literature treating the estimation of the radius of r -convexity is limited. In Rodríguez Casal and Saavedra-Nieves (2016), the authors introduce a statistical estimator of the radius of r -convexity by choosing the largest value of r that does not reject a certain hypothesis test at a prescribed level of significance.

In Chapter 4, we study a special case of Setting 1.1 for which it is possible to bound the radius of r -convexity of a compact set $A \subset \mathbb{R}^d$. Our result is not statistical, and so in a non-deterministic

setting, the bound on the radius of r -convexity holds with probability 1.

We suppose that there is some $\epsilon > 0$ such that a union of balls of radius ϵ centered at the points in the points $(\xi_i)_{i \in I}$ covers \mathbb{R}^d . The compactness of A implies that the set Ξ_{A^ϵ} is non-empty. We show that, in this framework, using the discrete analogue of dilation and erosion commonly used in computational geometry, a dilation of the set Ξ_A by a distance $r < \text{rconv}(A)$ followed by an erosion of the resulting set by the same radius r may result in a subset of $(\xi_i)_{i \in I}$ that is strictly greater than Ξ_A (see Figure 4.8 for an example in dimension 2). Since $\text{rconv}(A)$ is unknown, the naive (and false) conclusion is that $r > \text{rconv}(A)$.

We adapt the above strategy by appropriately altering the discrete dilation and erosion distances such that, after applying the manipulations on the set Ξ_A , obtaining anything that is not contained in Ξ_A correctly implies $r > \text{rconv}(A)$. The infimum of all such r is an upper bound for $\text{rconv}(A)$ and is shown to converge to $\text{rconv}(A)$ as $\epsilon \rightarrow 0$ (see Theorem 4.3).

Section 1.4: Summary of contributions


- We introduce the β -reach (see Definition 4.6) for $\beta \in [0, \infty)$, and show that it converges monotonically to the reach as $\beta \rightarrow 0$ for any closed set in \mathbb{R}^d (see Theorem 4.2).
- We define an upper bound for the reach of a closed set $A \subset \mathbb{R}^d$ that is approximated by discrete samples Ξ_A , and show that the bound converges to $\text{reach}(A)$ with an explicit rate of convergence as $\Xi_A \rightarrow A$ in the Hausdorff metric (see Theorem 4.4).
- We show that the traditional discrete closing operation from mathematical morphology does not directly lead to an upper bound on the radius of r -convexity of a closed set $A \subset \mathbb{R}^d$ (see Figure 4.8). We modify the discrete closing operation and show that the result can be used to construct an upper bound for $\text{rconv}(A)$. Moreover, we show that the bound converges to $\text{rconv}(A)$ (see Theorem 4.3).
- These bounds and the β -reach are shown to perform well on real, three-dimensional data (see Example 4.8).

1.5 Organization of the manuscript

Throughout Chapter 1, we have summarized the main findings that are detailed in the rest of this manuscript. Each following chapter is a self-contained article, either published or under revision for publication in an international, peer-reviewed journal. At the start of each chapter, we detail the independent notation used, that may differ slightly from other chapters. We apologize that this may lead to some redundancy between chapters. Furthermore, a short abstract is provided at the


beginning of each chapter. Using the harmonized notation introduced in this general introduction, each chapter is summarized as follows.

- **Chapter 2.** This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Duval C. (2023). Surface area and volume of excursion sets observed on point cloud based polytopic tessellations. *The Annals of Applied Probability* (to appear). [Paper here](#).


Theorem 1.1 is stated, proven, and expanded upon. Some related analyses are provided, including bias calculations for random Poisson-Voronoi tessellations. We also establish a joint central limit theorem for the surface area and volume of excursion sets observed over hypercubic lattices under certain mixing conditions on the random field.

- **Chapter 3.** This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Opitz, T. (2023). On the perimeter estimation of pixelated excursion sets of two-dimensional anisotropic random fields. *Scandinavian Journal of Statistics*, 1–34. [Paper here](#).


The statistical properties of the piecewise-linear approximation in (1.14) is studied as a length estimator for the perimeter of excursion sets of random fields on \mathbb{R}^2 . We prove that this method is multigrid convergent, even without the assumptions of isotropy and stationarity. Under slightly stronger assumptions, we prove a convergence result as the size of the domain grows to cover \mathbb{R}^2 . A Central Limit Theorem is proven for our estimator when multiple excursion levels are considered simultaneously. Several numerical studies are conducted to test our methods on simulated random fields.

- **Chapter 4.** This chapter is based on the work

 Cotsakis, R. (2023). Identifying the reach from high-dimensional point cloud data with connections to r -convexity. *Discrete & Computational Geometry* (to appear). [Paper here](#).

We elaborate on the analysis in Section 1.4. Algorithms are presented for bounding the reach and the radius of r -convexity for a closed subset of \mathbb{R}^d . The β -reach is defined and numerical studies suggest how it can be used in high-dimension to infer the reach of lower dimensional smooth submanifolds.

- **Chapter 5.** This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Opitz, T. (2023). A local statistic for the spatial extent of extreme threshold exceedances. *Submitted for publication in an international peer-reviewed journal*. [Paper here](#).


The extremal range in Definition 1.5 is studied, both in terms of its theoretical properties and its applicability to daily reanalysis data of French temperatures.

- **Chapter 6.** We conclude with several perspectives for improvements to the results detailed in this manuscript.

Chapter 2

Surface area and volume of excursion sets observed on point cloud based polytopic tessellations

This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Duval C. (2023). Surface area and volume of excursion sets observed on point cloud based polytopic tessellations. *Annals of Applied Probability* (to appear). [Paper here](#).

Abstract: The excursion set of a C^2 smooth random field carries relevant information in its various geometric measures. From a computational viewpoint, one never has access to the continuous observation of the excursion set, but rather to observations at discrete points in space. It has been reported that for specific regular lattices of points in dimensions 2 and 3, the usual approximation of the surface area of the excursions does not converge when the lattice becomes dense in the domain of observation to the desired limit. In the present work, under the key assumptions of stationarity and isotropy, we demonstrate that this limiting factor is invariant to the locations of the observation points. Indeed, we identify an explicit formula for the correction factor, showing that it only depends on the spatial dimension d . This enables us to define an approximation for the surface area of excursion sets for general tessellations of polytopes in \mathbb{R}^d , including Poisson-Voronoi tessellations. We also establish a joint central limit theorem for the surface area and volume of excursion sets observed over hypercubic lattices.

2.1 Introduction

2.1.1 Motivations

The study of random fields through the geometry of their excursion sets has received a lot of interest in recent literature. This is mainly stimulated by their wide range of applications in domains such as cosmology, for the study of Cosmic Microwave Background radiation and the distribution of galaxies (see, *e.g.*, [Casaponsa et al. \(2016\)](#), [Schmalzing and Górski \(1998\)](#), [Gott et al. \(2007\)](#), [Gott et al. \(2008\)](#)), brain imaging (see [Adler and Taylor \(2011\)](#), Section 5, and the references therein), the modelling of sea waves (see, *e.g.*, [Longuet-Higgins \(1957\)](#), [Wschebor \(1985\)](#), [Eymard et al. \(2000\)](#)), the routing algorithms for mobile communications networks (see, *e.g.*, [Baccelli et al. \(1997\)](#), [Baccelli et al. \(2000\)](#)) or the shape analysis (see, *e.g.*, [Lachaud et al. \(2023\)](#), [Lachaud et al. \(2020\)](#)).

The geometric features considered are referred to as either Lipschitz-Killing curvatures in stochastic geometry, curvature measures in physics and differential geometry, intrinsic volumes or Minkowski functionals in integral and convex geometry. In stochastic geometry, many studies have been dedicated to computing these objects from the observation of one excursion set of a random field on a compact domain T in \mathbb{R}^d (see, *e.g.*, [Adler and Taylor \(2007\)](#)). The level perimeter and level total curvature integrals for two-dimensional random fields are investigated in [Biermé and Desolneux \(2020\)](#). Limit results when the size of T grows to \mathbb{R}^d have been established under specific conditions on the random field (see, *e.g.*, [Bulinski \(2010\)](#), [Bulinski et al. \(2012\)](#), [Kratz and Vadlamani \(2018\)](#), [Meschenmoser and Shashkin \(2013\)](#), [Berzin \(2021\)](#) or [Spodarev \(2014\)](#)). The main motivation of this work comes from the following consideration: many articles rely on the assumption that “*the excursion set is observed on $T \subset \mathbb{R}^d$* ”, which is meant to be understood as “the field is *continuously* observed over T ”. This seems unrealistic, as in practice, for instance in two dimensions where excursion sets can be viewed as images, they are encoded through a matrix whose entries make a one-to-one connection with the pixels of the image. Even in cases where the resolution of the image is very high, the image of the excursion set remains a discretization of the continuous object that is the excursion set on T . Many methods allow to compute the Lipschitz-Killing curvatures of a set given in a pixelated image, taking into account the discrete nature of the observations. However, most of these methods are not *multigrid convergent*, *i.e.*, the computed values do not converge to the expected values when the resolution increases. The interested reader is referred for instance to [Guderlei et al. \(2007\)](#), [Meschenmoser and Spodarev \(2010\)](#), [Svane \(2014b\)](#), [Edelsbrunner and Pausinger \(2016\)](#). In the present work, we focus on two specific Lipschitz-Killing curvatures for the excursion sets of a C^2 stationary random field in \mathbb{R}^d : the $(d - 1)$ and d -dimensional Hausdorff measures (see their definition in Section 2.1.3). They correspond respectively to what we call in the sequel *surface area* (with a slight abuse of language), and *volume* in dimension d . It has been observed that local counting algorithms for the expected surface area of the discretized excursion set do not converge to that of the continuous excursion set. This multigrid convergence fails due to a dimensional bias factor.

Over the last two decades, significant efforts have been made to determine this factor for the estimated surface area of a *deterministic* Borelian set A of \mathbb{R}^d whose approximation is obtained from a *Poisson-Voronoi tessellation*. To the best of our knowledge, [Yukich \(2015\)](#) is the first to remark that in dimension d , the correction factor is independent of the target set A (see Section 2.2). In [Thäle and Yukich \(2016\)](#) it is also pointed out that in any d -dimension, “the surface area asymptotics involve a universal correction factor”. This correction factor is not explicitly computed in the aforementioned references. In an earlier work, ([Baccelli et al., 2000](#), Theorem 3) identifies that there is a $4/\pi$ bias factor when approximating straight paths by paths along the corresponding Delaunay tessellation of a Poisson-Voronoi diagram in dimension 2. This result is motivated by the telecommunication problem of estimating the path between two network stations modelled by a Poisson point process. Recently, ([Edelsbrunner and Nikitenko, 2021](#), Equation (1)) gives an expression of this correction factor as distortions of $(d - 1)$ -dimensional Voronoi scapes, which are higher dimensional generalizations of these Voronoi paths (see [Baccelli et al. \(2000\)](#)). Some results on approximations of the surface area of a Euclidean subset A by regular lattices (square, hypercubic, triangular, hexagonal, . . .), which cannot be obtained by Poisson-Voronoi tessellation, have been obtained by [Miller \(1999\)](#). In dimension 2 the article finds the same bias factor of $4/\pi$ for different lattices and in dimension 3 a similar behaviour is observed, with a bias factor of $3/2$ for the cubic mosaic. For these regular lattices (square and hexagonal) and in dimension 2, [Biermé and Desolneux \(2021\)](#) studies the case where the set A is *random* and is given by the excursion set of a random field. This same bias factor $4/\pi$ appears when one approximates the perimeter length of the boundary of the excursion set of a random field from its approximation in a regular lattice (see ([Biermé and Desolneux, 2021](#), Proposition 5)).

The main result of this paper unifies some of the aforementioned results. It gives a general picture in any d -dimension of the average surface area of the boundary of the excursion set of a random field approximated by a polytopic tessellations (in the sense of Definition 2.2). This includes both deterministic lattices and Poisson-Voronoi tessellations (see Figure 2.2). As anticipated in the above articles (in particular in [Yukich \(2015\)](#), [Thäle and Yukich \(2016\)](#), [Edelsbrunner and Nikitenko \(2021\)](#), [Baccelli et al. \(2000\)](#)), the bias factor is universal, independent of the chosen tessellation geometry and depends only on the d -dimension (see Equation (2.22) and (2.14)). In all dimensions, it is shown that this bias factor is identical to the expected distortion of Voronoi scapes in the dual problem solved by [Edelsbrunner and Nikitenko \(2021\)](#).

Moreover, thanks to a second order expansion (see Theorem 2.1), it is possible, to derive for a *hypercubic* lattice a *joint* central limit theorem for the approximated surface area and volume (see Theorem 2.3) by imposing additional strong mixing assumptions on the underlying field. This limit result is novel among existing limit results since it gives the joint asymptotics of two *different* Lipschitz-Killing curvatures, the surface area and the volume, whereas most multivariate limit theorems hold for a single curvature measure at multiple levels (see for instance [Bulinski et al. \(2012\)](#), [Di Bernardino et al. \(2017\)](#)).

The outline of the paper is the following. In Section 2.1.2 we define the geometric measures that we consider, as well as the Crofton formula, which is an essential tool to prove the main result (Theorem 2.1). In Section 2.1.3, the surface area and volume as well as their corresponding approximations on general point clouds are introduced. Since the bias of the approximate volume is a well understood deterministic quantity that is asymptotically negligible, Section 2.2 focuses on the study of the approximated surface area; the main results, which hold for general point clouds in any dimension (Theorems 2.1 and 2.2), are stated and proved. Section 2.3 restricts to the hypercubic lattice and proposes under additional strongly-mixing assumptions the joint CLT (Theorem 2.3) for the approximated surface area and volume. Sections 2.4 and 2.5 contain additional results and proofs related to Sections 2.2 and 2.3 respectively.

Finally, an Appendix Section includes several examples (see Section 2.A.1), some considerations on the convergence of the bias factor (see Section 2.A.2) and alternative approaches to recover the dimensional constant appearing in Theorem 2.1 (see Section 2.A.3).

2.1.2 Geometric measures and the Crofton formula

In the following, $\|\cdot\|_p$ denotes the L_p norm; $\|\cdot\|_\infty$, the supremum norm; $|\cdot|$, the absolute value; $\mathbb{1}_A$, the indicator of a set A ; and ∂A , the boundary of a set A . The closed ball of radius r centered at the origin $\mathbf{0}$ in \mathbb{R}^d is denoted B_r^d . Finally, recall that $(\mathbf{e}_i)_{1 \leq i \leq d}$ denotes the canonical basis of \mathbb{R}^d .

Hausdorff measures We first introduce the different measures considered in this article. For $k \in \{0, \dots, d\}$, let $\sigma_k(B)$ be the k -dimensional Hausdorff measure of a measurable set $B \subset \mathbb{R}^d$,

$$\sigma_k(B) := \frac{\pi^{k/2}}{2^k \Gamma(\frac{k}{2} + 1)} \lim_{\delta \rightarrow 0} \sigma_k^\delta(B), \quad (2.1)$$

where Γ denotes the gamma function and

$$\sigma_k^\delta(B) := \inf \left\{ \sum_{i \in \mathbb{N}} \text{diam}(U_i)^k : \text{diam}(U_i) < \delta, \bigcup_{i=1}^{\infty} U_i \supseteq B \right\}, \quad (2.2)$$

where $\text{diam}(U) := \sup\{\|u - v\|_2, u, v \in U\}$ and the infimum is taken over all countable covers of B by arbitrary subsets U_i of \mathbb{R}^d (see, e.g., [Schneider and Weil \(2008\)](#) p.634). The Hausdorff dimension of B is the unique integer value d_B such that $\sigma_k(B) = 0$ if $k < d_B$ and $\sigma_k(B) = +\infty$ if $k > d_B$ (see, e.g., [Rogers \(1998\)](#)). We have chosen to normalize $\sigma_k(B)$ in (2.1) such that for $k = \{0, \dots, d\}$, $\sigma_k(B)$ corresponds to the k -dimensional Lebesgue measure of B .

In the present work we focus on the measures in (2.1) for $k = d - 1$ and $k = d$. They correspond respectively to what we call *surface area* and *volume* in dimension d . In the proofs, two other measures play an important role: σ_0 , the counting measure for sets of isolated points, and σ_1 , the

measure of length.

Main tool: Crofton formula In the present work, we heavily rely on the Crofton formula. For a k -dimensional rectifiable set embedded in \mathbb{R}^d , this classical result in integral geometry relates the σ_k measure of the object with the average number of times it is intersected by randomly oriented $(d - k)$ -flats. We will be particularly interested in the case of $k = d - 1$, in which the 1-flats correspond to randomly oriented lines.

To state the Crofton formula, we first need to introduce the affine Grassmanian $A(d, m)$, for $m \in \{1, \dots, d\}$, which is the set of affine m -dimensional subspaces of \mathbb{R}^d . Since the set $A(d, 1)$ of lines in \mathbb{R}^d plays a crucial role in the Crofton formula we use, we propose a particular parametrization.

It is shown in [Schneider and Weil \(2008\)](#) p.168 that $A(d, 1)$ is equipped with a unique locally finite motion invariant measure μ_1 , that is normalized such that

$$\mu_1(\{l \in A(d, 1) : l \cap B_1^d \neq \emptyset\}) = \sigma_{d-1}(B_1^{d-1}).$$

For $\mathbf{s} \in \partial B_1^d$ and $\mathbf{v} \in \text{vect}(\mathbf{s}^\perp)$, denote by $l_{\mathbf{s}, \mathbf{v}}$ the element of $A(d, 1)$ that is parallel with \mathbf{s} and passes through the point \mathbf{v} ,

$$l_{\mathbf{s}, \mathbf{v}} := \{\mathbf{v} + \lambda \mathbf{s} : \lambda \in \mathbb{R}\}. \quad (2.3)$$

For each $\mathbf{s} \in \partial B_1^d$, there exists a unitary (rotation) operator $\theta_{\mathbf{s}}$ that maps \mathbf{e}_1 to \mathbf{s} . Therefore, we define the parametrization $\varphi : \partial B_1^d \times \mathbb{R}^{d-1} \rightarrow A(d, 1)$ satisfying $\varphi(\mathbf{s}, \mathbf{u}) = l_{\mathbf{s}, \mathbf{v}_{\mathbf{s}}(\mathbf{u})}$, with

$$\mathbf{v}_{\mathbf{s}}(\mathbf{u}) := \theta_{\mathbf{s}} \circ \begin{pmatrix} 0 \\ \mathbf{u} \end{pmatrix} \in \text{vect}(\mathbf{s}^\perp).$$

Notice that for $E \subset A(d, 1)$,

$$\mu_1(E) = \frac{(\sigma_{d-1} \otimes \sigma_{d-1})(\varphi^{-1}(E))}{\sigma_{d-1}(\partial B_1^d)},$$

where \otimes denotes the product measure.

We recall here a particular version of the Crofton formula ([Schneider and Weil, 2008, Theorem 5.4.3](#)), which implies that for a manifold $M \subset \mathbb{R}^d$ satisfying $0 < \sigma_{d-1}(M) < \infty$, it holds that

$$\sigma_{d-1}(M) = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \int_{A(d, 1)} \sigma_0(M \cap l) \mu_1(dl). \quad (2.4)$$

By writing $A(d, 1)$ in terms of the above parametrization φ , Equation (2.4) takes the form

$$\sigma_{d-1}(M) = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \int_{\mathbb{R}^{d-1}} \int_{\partial B_1^d} \frac{\sigma_0(M \cap l_{\mathbf{s}, \mathbf{v}_{\mathbf{s}}(\mathbf{u})})}{\sigma_{d-1}(\partial B_1^d)} \, d\mathbf{s} \, d\mathbf{u}. \quad (2.5)$$

A helpful interpretation of the Crofton formula in Equation (2.5) is as follows: the expected σ_{d-1} measure of the projection of M on a $(d-1)$ -dimensional hyperplane with uniformly random orientation is simply a constant multiple of $\sigma_{d-1}(M)$.

The Crofton formula can also be exploited to propose algorithms to compute the surface area, *e.g.*, it is considered in [Lehmann and Legland \(2012\)](#) for objects in 2 and 3 dimension and recently in [Aaron et al. \(2022\)](#) which provides consistent estimators for the surface area of a compact domain S from the observation of *i.i.d.* random variables supported on S . The interested reader is referred to Appendix 2.A.1 for an illustration of Equation (2.5) on two simple examples.

2.1.3 Approximated volume and surface area of excursion sets observed over a point cloud

Excursion sets, level surfaces, and associated measures We now apply the previous σ_d and σ_{d-1} measures to specific manifolds: the excursion sets and associated level surfaces of d -dimensional smooth random fields.

Definition 2.1 (σ_d and σ_{d-1} measures of excursion sets and associated level surfaces). Let $\{X(t), t \in \mathbb{R}^d\}$, for $d \geq 2$, be a random field satisfying the following assumption

(A0) X is stationary with positive finite variance and is almost surely twice differentiable. Furthermore, the probability density of $(X(\mathbf{0}), \nabla X(\mathbf{0}))$ is bounded uniformly on \mathbb{R}^{d+1} .

Let $u \in \mathbb{R}$ and $T \subset \mathbb{R}^d$ be a bounded closed hypercube with non empty interior. We consider the excursion set within T above level u :

$$E_X^T(u) := \{t \in T : X(t) \geq u\} = T \cap E_X(u), \quad \text{where } E_X(u) := X^{-1}([u, +\infty)).$$

Similarly, the level surfaces within T are defined by

$$L_X^T(u) := \{t \in T : X(t) = u\} = T \cap \partial E_X(u), \quad a.s.$$

Remark that Assumption (A0) guarantees that X admits no critical points at the level u almost surely, which implies that $L_X^T(u)$ is a $(d-1)$ -dimensional manifold possessing a σ_{d-1} measure with finite first and second moments (see, *e.g.*, [Cabaña \(1987\)](#) and [\(Adler and Taylor, 2007, Theorem 11.2.1 and Lemma 11.2.11\)](#)). In addition, since the considered random field X is of class C^2 *a.s.*, the random set $E_X(u)$ is a C^2 submanifold of \mathbb{R}^d and its intersection with the compact, convex hypercube T provides the positive reach property (see [Biermé et al. \(2019\)](#)).

Define the normalized σ_d and σ_{d-1} measures of the excursion set and associated level surfaces, for

$u \in \mathbb{R}$, as

$$C_{d-1}^T(u) := \frac{1}{\sigma_d(T)} \sigma_{d-1}(L_X^T(u)) = \frac{1}{\sigma_d(T)} \int_{L_X^T(u)} \sigma_{d-1}(ds), \quad (2.6)$$

$$C_d^T(u) := \frac{1}{\sigma_d(T)} \sigma_d(E_X^T(u)) = \frac{1}{\sigma_d(T)} \int_T \mathbb{1}_{\{X(t) \geq u\}} dt. \quad (2.7)$$

Assumption (A0) guarantees the existence of the associated densities

$$C_k^*(u) := \mathbb{E}[C_k^T(u)], \quad \text{for } k = d, d-1, \quad (2.8)$$

which are independent of the size of the hypercube T .

The independence of C_d^* from the size of T is trivially verified using that X is stationary and Fubini-Tonelli theorem which give immediately that the density of the normalised volume satisfies

$$C_d^*(u) = \frac{1}{\sigma_d(T)} \mathbb{E} \left[\int_T \mathbb{1}_{\{X(t) \geq u\}} dt \right] = \mathbb{P}(X(\mathbf{0}) \geq u). \quad (2.9)$$

Furthermore, note that we consider in (2.6) the σ_{d-1} measure of $L_X^T(u) = T \cap \partial E_X(u)$ and not of $\partial E_X^T(u)$. Therefore, from Definition 2.1 and Proposition 2.5 in [Biermé et al. \(2019\)](#), we get via kinematic formulas that $\mathbb{E}[C_{d-1}^T(u)]$ is equal to the surface area density. Indeed we do not add the artificial contribution of ∂T to the level surfaces in Definition 2.1. Notice that the density $C_{d-1}^*(u)$ can be explicitly obtained for certain specific random fields. Two classical examples (the isotropic Gaussian and chi-square random fields) are presented in [Appendix 2.A.1](#).

The random quantities in (2.6)-(2.7) can only be used as approximations of $C_d^*(u)$ and $C_{d-1}^*(u)$ if we observe the excursion set $E_X^T(u)$ on the whole domain T . In practice, or at least numerically, images of excursion sets are not objects defined on all T but discretely encrypted objects, *i.e.*, for each point of a discrete grid. Then, quantities in (2.6)-(2.7) are never empirically accessible. In the remainder of this section we propose computable counterparts of $C_d^*(u)$ and $C_{d-1}^*(u)$ based on the observation of the excursion set on a general point cloud (*i.e.*, based on the knowledge of which points fall in the excursion set).

Polytopical tessellations based on point clouds For an arbitrary point cloud, we describe the set of tessellations of \mathbb{R}^d that are permissible for the construction of our estimates for $C_d^*(u)$ and $C_{d-1}^*(u)$.

Definition 2.2. Let \mathcal{H} be a set of convex, closed polytopes that tessellates \mathbb{R}^d in such a way that satisfies the following condition. To each $P \in \mathcal{H}$, one can assign a reference point $P^\bullet \in P$ such that for any two adjacent cells $P_1, P_2 \in \mathcal{H}$, the intersection of their boundaries is normal to the vector spanned between P_1^\bullet and P_2^\bullet . We say that \mathcal{H} is *point-referenceable*, and that the set

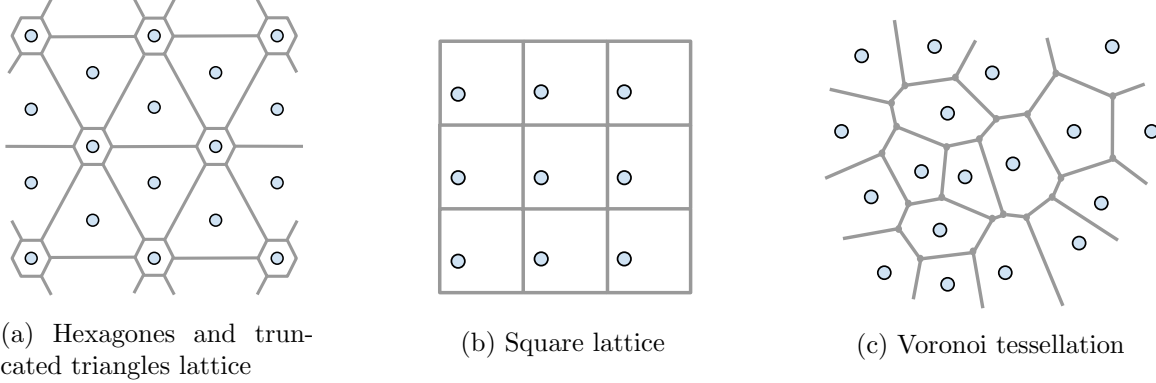


Figure 2.1: Three examples of tilings in \mathfrak{T}^2 (see Definition 2.2) with a particular choice of reference points shown as light-blue circles. The resulting structures are point-referenced 2-honeycombs.

of pairs $\dot{\mathcal{H}} = \{(P, P^\bullet) : P \in \mathcal{H}\}$ is a *point-referenced d -honeycomb*. Let \mathfrak{T}^d denote the space of point-referenceable d -honeycombs, so that $\mathcal{H} \in \mathfrak{T}^d$.

Recall that the interiors of the polytopes in a tessellation do not intersect, *i.e.*, $\sigma_d(P_1 \cap P_2) = 0$, for any different $P_1, P_2 \in \mathcal{H}$. However, if two polytopes P_1 and P_2 are *adjacent* in \mathcal{H} , then $\sigma_{d-1}(P_1 \cap P_2) > 0$.

Remark 2.1. The set of point-referenceable d -honeycombs \mathfrak{T}^d in Definition 2.2 contains a wide variety of tessellations. For example, the Voronoi diagram of any point cloud in \mathbb{R}^d is in \mathfrak{T}^d . It contains also d -honeycombs that cannot be realised as a Voronoi diagram as, for example, the Archimedean tessellations or the tiling of \mathbb{R}^2 in Figure 2.1 (a). Simple examples of tessellations of \mathbb{R}^2 that are not in \mathfrak{T}^2 are Pythagorean tilings and, of course, tilings with curved tiles. Tessellations similar to the point-referenceable d -honeycombs are used in finite element analysis in (Eymard et al., 2000, Definition 9.1).

Figure 2.1 provides some examples of point-referenceable tessellations of \mathbb{R}^2 . Notice that regular triangular and regular hexagonal lattices can be seen as limiting cases of the tessellation in Figure 2.1 (a); furthermore, as depicted, it is not a Voronoi tessellation. Figure 2.1 (b) represents the well-known square lattice, and panel (c) depicts the Voronoi tessellation of an arbitrary point cloud.

Estimates for $C_d^T(u)$ and $C_{d-1}^T(u)$ For a given point cloud in \mathbb{R}^d , that can be random, it is always possible to construct a point-referenced d -honeycomb (see Definition 2.2) with the point cloud as its reference points (take the Voronoi diagram, for example). Thus, for each element of \mathfrak{T}^d there is at least one corresponding point cloud, and to each point cloud in \mathbb{R}^d there is at least one corresponding element of \mathfrak{T}^d . We propose the following estimates for the quantities in (2.6) and (2.7) based on the knowledge of at which points in the point cloud the random field X exceeds the level u .

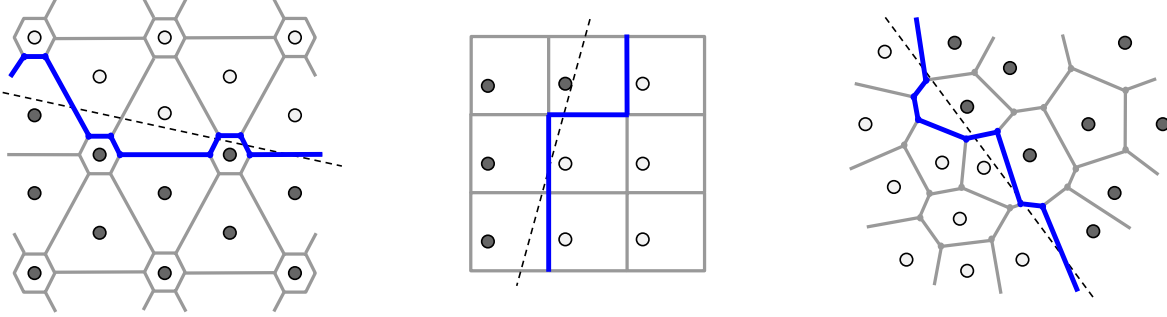


Figure 2.2: For each of the point-referenced 2-honeycombs in Figure 2.1, a portion of the level curve $\{X = u\}$ is illustrated as a dotted line. The estimator $C_{d-1}^{(\mathcal{H}, T)}(u)$ in (2.10) corresponds to the length of the resulting blue curve that separates the reference points in the same way as the curve $\{X = u\}$.

Definition 2.3. Let $T \subset \mathbb{R}^d$ be a compact domain with non empty interior. Let $\mathcal{H} \in \mathfrak{T}^d$, and let $\dot{\mathcal{H}} = \{(P, P^\bullet) : P \in \mathcal{H}\}$ be a corresponding point-referenced d -honeycomb in the sense of Definition 2.2. Let $\mathcal{H}^T \subset \mathcal{H}$ be the set of polytopes P in \mathcal{H} such that $P \subseteq T$. We define an estimator of $C_{d-1}^T(u)$ in (2.6) as

$$C_{d-1}^{(\dot{\mathcal{H}}, T)}(u) := \frac{1}{\sigma_d(T)} \sum_{\substack{P_1, P_2 \in \mathcal{H}^T \\ P_1 \neq P_2}} \sigma_{d-1}(P_1 \cap P_2) \mathbb{1}_{\{X(P_1^\bullet) \leq u < X(P_2^\bullet)\}} \quad (2.10)$$

and of $C_d^T(u)$ in (2.7) as

$$C_d^{(\dot{\mathcal{H}}, T)}(u) := \frac{1}{\sigma_d(T)} \sum_{P \in \mathcal{H}^T} \sigma_d(P) \mathbb{1}_{\{X(P^\bullet) \geq u\}}. \quad (2.11)$$

Computing the quantities in Definition 2.3 only requires the knowledge of the excursion set $E_X^T(u)$ on the reference points of a point-referenced d -honeycomb *i.e.*, a black and white image indicating at which points P^\bullet , for $P \in \mathcal{H}^T$, the field is above level u . In Equation (2.10), since the role of P_1 and P_2 is symmetric, both $\mathbb{1}_{\{X(P_1^\bullet) \leq u < X(P_2^\bullet)\}}$ and $\mathbb{1}_{\{X(P_2^\bullet) \leq u < X(P_1^\bullet)\}}$ are evaluated in the sum. Notice that it is crucial that the cells $P \in \mathcal{H}$ are closed to ensure that for adjacent cells $\sigma_{d-1}(P_1 \cap P_2) > 0$. Figure 2.2 provides an illustration of the behaviour of the estimate in (2.10) for the point-referenced 2-honeycombs in Figure 2.1.

It is easy to see that $C_d^{(\dot{\mathcal{H}}, T)}(u)$ in (2.11) is, in general, biased. Indeed, it follows from the stationarity in Assumption (A0) that

$$\mathbb{E}[C_d^{(\dot{\mathcal{H}}, T)}(u)] = \frac{\sigma_d\left(\bigcup_{P \in \mathcal{H}^T} P\right)}{\sigma_d(T)} C_d^*(u). \quad (2.12)$$

Nevertheless, the bias factor $\sigma_d(\bigcup_{P \in \mathcal{H}^T} P)/\sigma_d(T) \leq 1$ is deterministic and it only relies on the structure of the point-referenced d -honeycomb. It is nearly unity if $\sup\{\text{diam}(P) : P \in \mathcal{H}\} \ll \text{diam}(T)$ and it is exactly unity for tessellations that satisfy $\bigcup_{P \in \mathcal{H}^T} P = T$. This can be easily obtained for simple tessellations such as the hypercubic one, specifically described in Section 2.3.1. This is why we focus our attention to the bias in (2.10) for the $(d-1)$ -volume of $L_X^T(u)$. In this case, the bias factor does not approach unity as the tile sizes become negligibly small with regards to the size of T , as previously observed *e.g.*, in Biermé and Desolneux (2021) and Miller (1999) for some periodic tessellations and in Thäle and Yukich (2016) for Poisson-Voronoi mosaics. Explicit calculations of the bias have been made for square and hexagonal tilings in dimension 2 (Biermé and Desolneux (2021)) and cubic lattices in dimension 3 (Miller (1999)). The approach in Miller (1999) is adapted to the d -dimensional hypercubic setting in Appendix 2.A.3 (see Equation (2.50)). Based on a preliminary result on crossings (see Theorem 2.1), our Theorem 2.2 provides a general formula for this limiting factor that holds in arbitrary dimension d , for a large subset of point-referenced d -honeycombs.

2.2 Approximated surface area

Assumptions We require additional assumptions related to the regularity of the derivatives of X in univariate directions.

- (A1) Fix $\mathbf{w} \in \partial B_1^d$. Assume that $Y := \{X(t\mathbf{w}), t \in \mathbb{R}\}$ is such that $Y(0) \setminus \{Y'(0) = 0\}$ has a bounded density, *i.e.*, that the one dimensional density function of the heights of the local extrema of Y is bounded, and that $\|Y''\|_{\infty, [0,1]} := \sup_{t \in [0,1]} |Y''(t)| \in L^2$.
- (A2) Let $L_X(u) := X^{-1}(\{u\})$, $\mathbf{w} \in \partial B_1^d$ and $\varepsilon \in (0, 1)$. Assume that $\mathbb{E}[\sigma_0(L_X(u) \cap l_{\mathbf{w}, \mathbf{0}} \cap [0, 1]^d)^{1/\varepsilon}]$ is finite, where $l_{\mathbf{w}, \mathbf{0}}$ is defined in (2.3).

Note that if the random field X is isotropic, the vector $\mathbf{w} \in \partial B_1^d$ appearing in (A1) and (A2) can be chosen arbitrarily. Assumption (A1) permits to apply Lemma 2.1 to the unidirectional process Y . The aforementioned lemma controls the probability that a one-dimensional field crosses a level u more than once in a small time interval. Assumption (A2) is technical and allows for a Hölder control in the proof of Theorem 2.1.

Remark that Assumptions (A1) and (A2) are satisfied, for example, by Gaussian random fields, and more generally, some Gaussian mixture models discussed in Di Bernardino et al. (2022). This general model consists of scale mixture models of the form $\Lambda \times W$, and location mixture models of the form $\Lambda + W$, for some random variable Λ independent of the Gaussian random field W . To see that Assumption (A1) is satisfied by both mixture models, remark that their one-dimensional cross sections are mixture models with the same latent variable Λ . The critical points of the mixture model are those of the Gaussian process, and their height distribution can be made to have bounded density with suitable choice of Λ (see (Cheng and Schwartzman, 2018)). As the second derivative

is again a mixture model, the second moment of its supremum on $[0, 1]$ can be shown to be finite if $\mathbb{E}[\Lambda^2] < \infty$ by the Borell-TIS inequality (Adler and Taylor, 2007, Theorem 2.1.1). Assumption (A2) is seen to hold for all $\epsilon \in (0, 1)$, since the moments of the number of level crossings N of a the mixture process on $[0, 1]$ are finite when conditioning on Λ (Kratz, 2006). In addition, the random variable $\mathbb{E}[N^{1/\epsilon} | \Lambda]$ is bounded, and hence integrable.

A first result on crossings The following key result provides a first order approximation of the probability of crossings for the isotropic random field X in a given arbitrary direction. This result plays a central role in deriving a formula for the expected surface area of the limiting approximation of an excursion set by elements of a point-referenceable d -honeycomb in the sense of Definition 2.2 (see Theorem 2.2 below).

Theorem 2.1. *Consider a scalar $q \in \mathbb{R}^+$ and a fixed $\mathbf{w} \in \partial B_1^d$. Let X be an isotropic random field satisfying Assumption (A0). Then for any fixed $u \in \mathbb{R}$,*

$$\lim_{q \rightarrow 0} \frac{1}{q} \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) = \frac{C_{d-1}^*(u)}{\beta_d}, \quad (2.13)$$

where

$$\beta_d = \frac{2\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \quad (2.14)$$

is a dimensional constant and the limit in (2.13) is approached from below. Moreover, if X also satisfies Assumptions (A1) and (A2) for \mathbf{w} and some $\epsilon \in (0, 1)$, then there exists a constant $K \in \mathbb{R}^+$ such that for all $q \in \mathbb{R}^+$,

$$0 \leq \frac{C_{d-1}^*(u)}{\beta_d} q - \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \leq Kq^{2-\epsilon}. \quad (2.15)$$

The proof of Theorem 2.1 is based on the Crofton formula in (2.5).

Proof of Theorem 2.1. Firstly, we prove Equation (2.13). We apply Equation (2.5) to the $(d-1)$ -manifold $L_X(u) \cap B_1^d$,

$$\sigma_{d-1}(L_X(u) \cap B_1^d) = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \int_{\partial B_1^d} \int_{\text{vect}(\mathbf{s}^\perp)} \frac{\sigma_0(L_X(u) \cap B_1^d \cap l_{\mathbf{s}, \mathbf{v}})}{\sigma_{d-1}(\partial B_1^d)} d\mathbf{v} d\mathbf{s}, \quad (2.16)$$

with $l_{\mathbf{s}, \mathbf{v}}$ as in (2.3). The quantity in (2.16) is a positive random variable in L^1 , and thus by taking the expectation, we get

$$\mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)] = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \int_{\partial B_1^d} \int_{\text{vect}(\mathbf{s}^\perp)} \frac{\mathbb{E}[\sigma_0(L_X(u) \cap B_1^d \cap l_{\mathbf{s}, \mathbf{v}})]}{\sigma_{d-1}(\partial B_1^d)} d\mathbf{v} d\mathbf{s}.$$

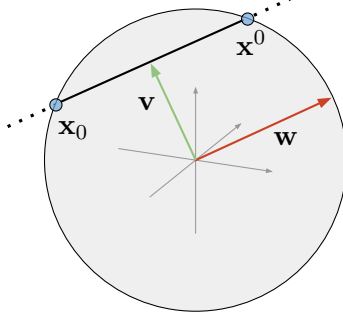


Figure 2.3: In dimension 3, the unit ball is represented; the red vector is a possible $\mathbf{w} \in \partial B_1^3$ and the green vector, a possible $\mathbf{v} \in \text{vect}(\mathbf{w}^\perp) \cap B_1^3$. The two endpoints \mathbf{x}_0 and \mathbf{x}^0 , marked in blue, describe the segment $\overline{l_{\mathbf{w},\mathbf{v}}}$.

By the isotropy of X , we obtain

$$\mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)] = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \int_{\text{vect}(\mathbf{w}^\perp)} \mathbb{E}[\sigma_0(L_X(u) \cap B_1^d \cap l_{\mathbf{w},\mathbf{v}})] d\mathbf{v}, \quad (2.17)$$

for $l_{\mathbf{w},\mathbf{v}}$ as in (2.3).

For $\mathbf{v} \in \text{vect}(\mathbf{w}^\perp) \cap B_1^d$, define $\overline{l_{\mathbf{w},\mathbf{v}}} := B_1^d \cap l_{\mathbf{w},\mathbf{v}}$. Then $\overline{l_{\mathbf{w},\mathbf{v}}}$ either contains a single point, or it is a line segment in \mathbb{R}^d with orientation \mathbf{w} . Define

$$n_q = \left\lfloor \frac{\sigma_1(\overline{l_{\mathbf{w},\mathbf{v}}})}{q} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Define $\mathbf{x}_0 := \mathbf{v} - \frac{\sigma_1(\overline{l_{\mathbf{w},\mathbf{v}}})}{2} \mathbf{w}$ and $\mathbf{x}^0 := \mathbf{v} + \frac{\sigma_1(\overline{l_{\mathbf{w},\mathbf{v}}})}{2} \mathbf{w}$ to be the two endpoints of the line segment $\overline{l_{\mathbf{w},\mathbf{v}}}$. Figure 2.3 provides a graphical visualisation to aid this construction.

Now define $\mathbf{x}_j := \mathbf{x}_0 + jq\mathbf{w}$ for $j \in \{0, \dots, n_q\}$, each of which being an element of $\overline{l_{\mathbf{w},\mathbf{v}}}$. With this construction, it follows that

$$\sum_{j=0}^{n_q-1} (\mathbb{1}_{\{X(\mathbf{x}_j) \leq u < X(\mathbf{x}_{j+1})\}} + \mathbb{1}_{\{X(\mathbf{x}_j) > u \geq X(\mathbf{x}_{j+1})\}}) \xrightarrow{\text{a.s.}} \sigma_0(L_X(u) \cap \overline{l_{\mathbf{w},\mathbf{v}}}), \quad (2.18)$$

as $q \rightarrow 0$. The left-hand side of (2.18) is a count of the number of elements in $L_X(u) \cap \overline{l_{\mathbf{w},\mathbf{v}}}$, which approaches the exact value from below almost surely. This construction is well known in the literature as the *discretization method*. The interested reader is referred for instance to (Kratz, 2006, Section 2.1) and references therein. In particular, by using the stationarity of X in Assumption (A0), the monotonicity of the convergence in (2.18) implies

$$2n_q \mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \xrightarrow{q \rightarrow 0} \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w},\mathbf{v}}})].$$

Moreover, this convergence is uniform for $\mathbf{v} \in \text{vect}(\mathbf{w}^\perp)$, so

$$\mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} 2n_q d\mathbf{v} \xrightarrow{q \rightarrow 0} \int_{\text{vect}(\mathbf{w}^\perp)} \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] d\mathbf{v}.$$

By (2.17), this simplifies to

$$\mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} n_q d\mathbf{v} \xrightarrow{q \rightarrow 0} \frac{1}{\beta_d} \mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)].$$

Since $0 \leq \sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}})/q - n_q < 1$ and $\mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \xrightarrow{q \rightarrow 0} 0$, it holds that

$$\mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} \frac{\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}})}{q} d\mathbf{v} \xrightarrow{q \rightarrow 0} \frac{1}{\beta_d} \mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)].$$

By noticing that $\int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} \sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) d\mathbf{v} = \sigma_d(B_1^d)$, one obtains

$$\frac{1}{q} \mathbb{P}(X(\mathbf{0}) \leq u < X(\mathbf{w}q)) \xrightarrow{q \rightarrow 0} \frac{1}{\beta_d \sigma_d(B_1^d)} \mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)] = \frac{C_{d-1}^*(u)}{\beta_d}.$$

This proves the first order approximation in (2.13).

To show that

$$\frac{C_{d-1}^*(u)}{\beta_d} q \geq \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \tag{2.19}$$

for all $q > 0$, suppose that there exists a $t \in \mathbb{R}^+$ such that

$$\frac{C_{d-1}^*(u)}{\beta_d} t < \mathbb{P}(X(\mathbf{0}) \leq u < X(t\mathbf{w})).$$

Then, for all $n \in \mathbb{N}^+$, $\{X(\mathbf{0}) \leq u < X(t\mathbf{w})\} \subseteq \{X(\mathbf{0}) \leq u < X(\frac{t}{n}\mathbf{w})\} \cup \dots \cup \{X(\frac{n-1}{n}t\mathbf{w}) \leq u < X(t\mathbf{w})\}$, leading under (A0) to

$$\frac{C_{d-1}^*(u)}{\beta_d} < \frac{1}{t} \mathbb{P}(X(\mathbf{0}) \leq u < X(t\mathbf{w})) \leq \frac{n}{t} \mathbb{P}(X(\mathbf{0}) \leq u < X(\frac{t}{n}\mathbf{w})).$$

Keeping t fixed and having $n \rightarrow \infty$ contradicts the convergence in (2.13). Thus, (2.19) holds for all $q \in \mathbb{R}^+$.

We proceed by showing (2.15). For fixed $q > 0$, there can only be a difference between the expression in (2.18) and its limit—both taking values in \mathbb{N}_0 —if at least one of the two following conditions hold,

- (i) There exists a $j \in \{0, \dots, n_q - 1\}$ such that the line segment with endpoints \mathbf{x}_j and \mathbf{x}_{j+1} contains two points $\mathbf{s}_1 \neq \mathbf{s}_2$ such that $X(\mathbf{s}_1) = X(\mathbf{s}_2) = u$.

(ii) The line segment with endpoints \mathbf{x}_{n_q} and \mathbf{x}^0 contains a point \mathbf{s} such that $X(\mathbf{s}) = u$.

The probability of event in (i) is bounded above by $n_q K_2 q^2$ for some $K_2 \in \mathbb{R}^+$, under (A1) by applying Lemma 2.1 with $p = 2$. Furthermore, by using Markov's inequality and (Beutler and Leneman, 1966, Theorem 3.3.1), the probability of event (ii) is bounded above by $K_1 q$ for some $K_1 \in \mathbb{R}^+$. Let A denote the event $\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}) > \sum_{j=0}^{n_q-1} (\mathbb{1}_{\{X(\mathbf{x}_j) \leq u < X(\mathbf{x}_{j+1})\}} + \mathbb{1}_{\{X(\mathbf{x}_j) > u \geq X(\mathbf{x}_{j+1})\}})$, then applying Hölder's inequality together with (A2) implies

$$\begin{aligned}
& \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2n_q \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \\
&= \mathbb{E}\left[\left|\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}) - \sum_{j=0}^{n_q-1} (\mathbb{1}_{\{X(\mathbf{x}_j) \leq u < X(\mathbf{x}_{j+1})\}} + \mathbb{1}_{\{X(\mathbf{x}_j) > u \geq X(\mathbf{x}_{j+1})\}})\right|\right] \\
&\leq \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}) \mathbb{1}_{\{A\}}] \leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon \mathbb{P}(A)^{1-\varepsilon} \\
&\leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon (n_q K_2 q^2 + K_1 q)^{1-\varepsilon} \\
&\leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon (\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) K_2 + K_1)^{1-\varepsilon} q^{1-\varepsilon} \\
&\leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon (2K_2 + K_1)^{1-\varepsilon} q^{1-\varepsilon}.
\end{aligned}$$

Furthermore, note that

$$\begin{aligned}
& \left| \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2 \frac{\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}})}{q} \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right| \\
&= \left| \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2n_q \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) + 2(n_q - \frac{\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}})}{q}) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right| \\
&\leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon (2K_2 + K_1)^{1-\varepsilon} q^{1-\varepsilon} + 2\mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \\
&\leq \left(\mathbb{E}[(\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}}))^{\frac{1}{\varepsilon}}]\right)^\varepsilon (2K_2 + K_1)^{1-\varepsilon} q^{1-\varepsilon} + 2K_1 q.
\end{aligned}$$

Summarizing, we have shown that

$$\left| q \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right| = O(q^{2-\varepsilon}),$$

for all $\mathbf{v} \in \text{vect}(\mathbf{w}^\perp) \cap B_1^d$. By integrating over \mathbf{v} and taking the absolute value, one preserves

$$\left| \int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} \left(q \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right) d\mathbf{v} \right| = O(q^{2-\varepsilon}). \quad (2.20)$$

Evaluating the integral, and again applying $\int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} \sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) d\mathbf{v} = \sigma_d(B_1^d)$, one obtains

$$\begin{aligned}
& \int_{\text{vect}(\mathbf{w}^\perp) \cap B_1^d} \left(q \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] - 2\sigma_1(\overline{l_{\mathbf{w}, \mathbf{v}}}) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right) d\mathbf{v} \\
&= q \int_{\text{vect}(\mathbf{w}^\perp)} \mathbb{E}[\sigma_0(L_X(u) \cap \overline{l_{\mathbf{w}, \mathbf{v}}})] d\mathbf{v} - 2\sigma_d(B_1^d) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \\
&= \frac{2q}{\beta_d} \mathbb{E}[\sigma_{d-1}(L_X(u) \cap B_1^d)] - 2\sigma_d(B_1^d) \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \\
&= 2\sigma_d(B_1^d) \left(\frac{C_{d-1}^*(u)}{\beta_d} q - \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right),
\end{aligned}$$

where the second equality follows from (2.17). Thus, by (2.20), we obtain the desired result

$$\left| \frac{C_{d-1}^*(u)}{\beta_d} q - \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{w})) \right| = O(q^{2-\varepsilon}).$$

□

Remark 2.2. Theorem 2.1 can also be very useful for sample simulations. To have a rapid evaluation of the surface area of excursion sets of a given random field satisfying above assumptions, it is not necessary to generate the whole isotropic random field on \mathbb{R}^d but only *i.i.d.* r.v. with the same bivariate distribution as $(X(\mathbf{0}), X(q\mathbf{e}_1))$, for q small enough. Then, the numerical first order approximation of the surface area would be

$$\frac{\widehat{\mathbb{P}}(X(\mathbf{0}) \leq u < X(q\mathbf{e}_1))}{q}.$$

This induces an error of the order $q^{1-\varepsilon}$ according to Equation (2.15). This approach can be compared for instance with the tedious computations encountered when using tube formulas (see, *e.g.*, Adler and Taylor (2007)) to obtain an exact value of the expected surface area (see also Bierné and Desolneux (2020)).

A related result Notice that Theorem 2.1 fully identifies the limit of the level crossing and relies on classical assumptions. Another result on level crossings can be obtained by adapting a result in Leadbetter et al. (1983) to dimension d (see Proposition 2.1 below). For clarity, the proof of this adaptation in general dimension d is given in Section 2.4.

Proposition 2.1 (A d -dimensional formulation of (Leadbetter et al., 1983, Theorem 7.2.4)). *Let $X : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuous stationary random field such that $X(\mathbf{0})$ and $X_q := \frac{1}{q}(X(q\mathbf{e}_1) - X(\mathbf{0}))$ have a joint density denoted $g_q(u, z)$ that is continuous in u for all z and all sufficiently small $q > 0$, and that there exists a function $p(u, z)$ such that $g_q(u, z) \rightarrow p(u, z)$ uniformly in u for fixed z as $q \rightarrow 0$. Assume furthermore that there is a function $h(z)$ such that $\int_0^\infty zh(z)dz < \infty$ and*

$g_q(u, z) \leq h(z)$ uniformly for all u, q . Then, it holds that

$$\lim_{q \rightarrow 0} \frac{\mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{e}_1))}{q} = \int_0^\infty z p(u, z) dz.$$

Remark that it is difficult to apply Proposition 2.1 in practice, since computing the asymptotic joint density p is not easy outside of specific cases such as the Gaussian one. Moreover, it requires that the convergence of the density of X_q as $q \rightarrow 0$ is established. Even if it is clear that X_q converges almost surely to $\partial_1 X(\mathbf{0})$, where $\partial_1 X$ denotes the partial derivative of X in the direction \mathbf{e}_1 , having the convergence of the corresponding densities is much more delicate to establish (see, e.g., Boos (1985) or Sweeting (1986)). If $d = 1$, Leadbetter et al. (1983) explains that “In many cases, the limit $p(u, z)$ is simply the joint density of $(X(\mathbf{0}), \partial_1 X(\mathbf{0}))$.” This holds for example, for Gaussian processes. Translating this in dimension $d \geq 1$, if we write $p_{X(\mathbf{0})}(u)$ for the density of $X(\mathbf{0})$, we see that

$$\lim_{q \rightarrow 0} \frac{\mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{e}_1))}{q} = p_{X(\mathbf{0})}(u) \mathbb{E} [\partial_1 X(\mathbf{0}) \mathbb{1}_{\{\partial_1 X(\mathbf{0}) > 0\}} | X(\mathbf{0}) = u]. \quad (2.21)$$

Correction factor for the approximated surface area The following theorem provides the explicit formula for the bias of the approximated surface area in any dimension d and for arbitrary point-referenced d -honeycombs in Definition 2.2 whose polytopes do not grow too large near ∞ . In this way it generalizes and unifies existing results (see, e.g., (Biermé and Desolneux, 2021, Proposition 5) for the square and hexagonal tiling in dimension 2 and Miller (1999) for the triangular tiling in dimension 2 and hypercubic lattice in dimension 3). Furthermore, Theorem 2.2 applies to Poisson-Voronoi tessellations, as shown in Corollary 2.1.

Theorem 2.2. *Let $T \subset \mathbb{R}^d$ be a compact domain with non empty interior containing the origin. Let X be an isotropic random field on \mathbb{R}^d that satisfies Assumption (A0).*

- i) *Let $\mathcal{H} \in \mathfrak{T}^d$ (see Definition 2.2) and let $\dot{\mathcal{H}}$ be a corresponding point-referenced d -honeycomb. Define $D^{(\mathcal{H})} := \sup\{\text{diam}(P \cap T) : P \in \mathcal{H}\}$ and $\delta\mathcal{H} := \{\delta P : P \in \mathcal{H}\}$ for $\delta \in \mathbb{R}^+$. Let $C_{d-1}^{(\dot{\mathcal{H}}, T)}(u)$ be as in Definition 2.3. Suppose that $\lim_{\delta \rightarrow 0} D^{(\delta\mathcal{H})} = 0$. Then, it holds*

$$\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d} C_{d-1}^*(u), \quad (2.22)$$

where $\delta\dot{\mathcal{H}} := \{(\delta P, \delta P^\bullet) : (P, P^\bullet) \in \dot{\mathcal{H}}\}$ is $\dot{\mathcal{H}}$ linearly rescaled by δ , β_d is as in (2.14), and $C_{d-1}^*(u)$ is as in (2.8).

- ii) *Furthermore, if ξ is a point process on \mathbb{R}^d that is independent of X , let \mathcal{V}_ξ be the Voronoi diagram in \mathfrak{T}^d generated from the points in ξ . Reference the polytopes in \mathcal{V}_ξ by their corresponding points in ξ and denote the resulting point-referenced d -honeycomb by $\dot{\mathcal{V}}_\xi$. If $D^{(\delta\mathcal{V}_\xi)} \xrightarrow{\mathbb{P}} 0$ as*

$\delta \rightarrow 0$, then

$$\mathbb{E}[C_{d-1}^{(\delta\dot{V}_\xi, T)}(u)] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d} C_{d-1}^*(u).$$

Remark 2.3. The universal correction factor $2d/\beta_d$ in Theorem 2.2 is equal to $4/\pi \approx 1.27$ for $d = 2$; $3/2$ for $d = 3$; and $16/(3\pi)$ for $d = 4$. The correction factor $4/\pi$ can be found in (Baccelli et al., 2000, Theorem 3) for the Poisson-Voronoi paths in dimension 2. One can show that $2d/\beta_d \sim \sqrt{2d/\pi}$, as $d \rightarrow \infty$, where \sim indicates that the ratio of the two expressions tends to 1. Furthermore, (Edelsbrunner and Nikitenko, 2021, Table 1) provides the same correction factor as the expected distortion of higher dimensional generalizations of Voronoi paths introduced in Baccelli et al. (2000), called *Voronoi scapes*. Given a Voronoi diagram in \mathbb{R}^d and a $(d-1)$ -dimensional set S with finite surface area, the authors define the resulting Voronoi scape as the union of $(d-1)$ -cells in the dual Delaunay tessellation that correspond to the edges in the Voronoi diagram that are intersected by S . The authors in Edelsbrunner and Nikitenko (2021) show that when S is randomly oriented with respect to the Voronoi diagram, the expected ratio of the surface area of the Voronoi scape to that of S is $2d/\beta_d$. This result can be seen as the dual of Theorem 2.2 applied to Voronoi tessellations, which says that $C_{d-1}^{(\delta\dot{V}_\xi, T)}(u)$ is the surface area density of the Voronoi faces corresponding to Delaunay edges that are intersected by $L_X^T(u)$.

Proof of Theorem 2.2. First, we establish Equation (2.22) where the point-referenced d -honeycomb $\dot{\mathcal{H}}$ is deterministic. The estimator $C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)$ in (2.10) is written

$$C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u) = \frac{1}{\sigma_d(T)} \sum_{\substack{P_1, P_2 \in \mathcal{H} \\ (P_1 \neq P_2), (\delta P_1, \delta P_2 \subseteq T)}} \sigma_{d-1}(\delta P_1 \cap \delta P_2) \mathbb{1}_{\{X(\delta P_1^\bullet) \leq u < X(\delta P_2^\bullet)\}}.$$

By the linearity of the expectation, for $\delta > 0$,

$$\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)] = \frac{1}{\sigma_d(T)} \sum_{\substack{P_1, P_2 \in \mathcal{H} \\ (P_1 \neq P_2), (\delta P_1, \delta P_2 \subseteq T)}} \sigma_{d-1}(\delta P_1 \cap \delta P_2) \mathbb{P}(X(\delta P_1^\bullet) \leq u < X(\delta P_2^\bullet)). \quad (2.23)$$

Let $P_1, P_2 \in \mathcal{H}$ be such that $P_1 \neq P_2$ and $\sigma_{d-1}(P_1 \cap P_2) > 0$, *i.e.*, such that P_1 and P_2 are adjacent in \mathcal{H} . Then as $\delta \rightarrow 0$, clearly $\|\delta P_2^\bullet - \delta P_1^\bullet\|_2 \rightarrow 0$, and so by Equation (2.13) in Theorem 2.1,

$$\frac{\mathbb{P}(X(\delta P_1^\bullet) \leq u < X(\delta P_2^\bullet))}{\|\delta P_2^\bullet - \delta P_1^\bullet\|_2} \xrightarrow{\delta \rightarrow 0} \frac{C_{d-1}^*(u)}{\beta_d}. \quad (2.24)$$

Moreover, if one restricts to neighbouring polytopes P_1 and P_2 in \mathcal{H} such that $\delta P_1, \delta P_2 \subseteq T$, then $D^{(\delta\mathcal{H})} \rightarrow 0$ implies that the convergence in (2.24) is uniform as T remains fixed, *i.e.*,

$$\sup \left\{ \left| \frac{\mathbb{P}(X(\delta P_1^\bullet) \leq u < X(\delta P_2^\bullet))}{\|\delta P_2^\bullet - \delta P_1^\bullet\|_2} - \frac{C_{d-1}^*(u)}{\beta_d} \right| : P_1 \text{ neighbours } P_2, \delta P_1, \delta P_2 \subseteq T \right\} \xrightarrow{\delta \rightarrow 0} 0.$$

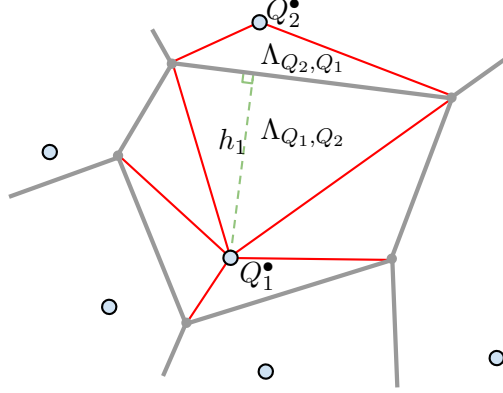


Figure 2.4: A pentagonal cell Q_1 in a point-referenced 2-honeycomb is partitioned into five pyramids (triangles in the case $d = 2$) each with summit Q_1^\bullet . A sixth pyramid Λ_{Q_2, Q_1} with summit Q_2^\bullet is drawn in the cell Q_2 , which shares its base $Q_1 \cap Q_2$ with the pyramid Λ_{Q_1, Q_2} contained in Q_1 . The height of the pyramid Λ_{Q_1, Q_2} is marked h_1 .

Therefore, Equation (2.23) can be written

$$\mathbb{E}[C_{d-1}^{(\delta\mathcal{H}, T)}(u)] - \frac{C_{d-1}^*(u)}{\beta_d \sigma_d(T)} \sum_{\substack{P_1, P_2 \in \mathcal{H} \\ (P_1 \neq P_2), (\delta P_1, \delta P_2 \subseteq T)}} \sigma_{d-1}(\delta P_1 \cap \delta P_2) \|\delta P_2^\bullet - \delta P_1^\bullet\|_2 \xrightarrow{\delta \rightarrow 0} 0.$$

With the change of variables $Q_1 = \delta P_1$ and $Q_2 = \delta P_2$, we have

$$\mathbb{E}[C_{d-1}^{(\delta\mathcal{H}, T)}(u)] - \frac{C_{d-1}^*(u)}{\beta_d \sigma_d(T)} \sum_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \neq Q_2}} \sigma_{d-1}(Q_1 \cap Q_2) \|Q_2^\bullet - Q_1^\bullet\|_2 \xrightarrow{\delta \rightarrow 0} 0. \quad (2.25)$$

For adjacent $Q_1, Q_2 \in (\delta\mathcal{H})^T$, denote by Λ_{Q_1, Q_2} the d -dimensional hyperpyramid with base $Q_1 \cap Q_2$ and vertex Q_1^\bullet (see Figure 2.4).

The d dimensional analogue of the area of a triangle implies that $\sigma_d(\Lambda_{Q_1, Q_2}) = \frac{1}{d} \sigma_{d-1}(Q_1 \cap Q_2) h_1$, where h_1 is the distance from Q_1^\bullet to the hyperplane containing $Q_1 \cap Q_2$. Now, $\sigma_d(\Lambda_{Q_1, Q_2} \cup \Lambda_{Q_2, Q_1}) = \frac{1}{d} \sigma_{d-1}(Q_1 \cap Q_2) \|Q_2^\bullet - Q_1^\bullet\|_2$, and so

$$\begin{aligned} \sum_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \neq Q_2}} \sigma_{d-1}(Q_1 \cap Q_2) \|Q_2^\bullet - Q_1^\bullet\|_2 &= \sum_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \text{ adjacent to } Q_2}} d \sigma_d(\Lambda_{Q_1, Q_2} \cup \Lambda_{Q_2, Q_1}) \\ &= 2d \sum_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \text{ adjacent to } Q_2}} \sigma_d(\Lambda_{Q_1, Q_2}) = 2d \sigma_d \left(\bigcup_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \text{ adjacent to } Q_2}} \Lambda_{Q_1, Q_2} \right). \end{aligned}$$

Since every point in T at distance of at least $2D^{(\delta\mathcal{H})}$ from the boundary of T is contained in a

Λ_{Q_1, Q_2} for some $Q_1, Q_2 \in (\delta\mathcal{H})^T$, and Λ_{Q_1, Q_2} is contained in $Q_1 \subseteq T$,

$$\lim_{\delta \rightarrow 0} \bigcup_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \text{ adjacent to } Q_2}} \Lambda_{Q_1, Q_2} = T.$$

Thus, by the continuity of the measure σ_d ,

$$\sum_{\substack{Q_1, Q_2 \in (\delta\mathcal{H})^T \\ Q_1 \neq Q_2}} \sigma_{d-1}(Q_1 \cap Q_2) \|Q_2^\bullet - Q_1^\bullet\|_2 \xrightarrow{\delta \rightarrow 0} 2d \sigma_d(T), \quad (2.26)$$

and by (2.25),

$$\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d} C_{d-1}^*(u).$$

This proves (2.22). Now if the tessellation is produced from a point process ξ and if $D^{(\delta\mathcal{V}_\xi)} \xrightarrow{\mathbb{P}} 0$ as $\delta \rightarrow 0$, then (2.22) implies that

$$\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u) \mid \xi] \xrightarrow{\mathbb{P}} \frac{2d}{\beta_d} C_{d-1}^*(u), \quad \delta \rightarrow 0. \quad (2.27)$$

What remains to be shown is that the convergence in (2.27) holds in L^1 . First, remark that the sequence in (2.24) converges to its limit from below by Theorem 2.1. Remark also that the sequence in (2.26) converges to its limit from below. These two remarks, together with the convergence in probability in (2.27), imply that

$$\sup_{\delta > 0} \mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u) \mid \xi] = \frac{2d}{\beta_d} C_{d-1}^*(u), \quad a.s.$$

Therefore, the convergence in (2.27) holds in L^1 and $\mathbb{E}[\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u) \mid \xi]] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d} C_{d-1}^*(u)$ as desired. \square

Remark 2.4. In the second part of Theorem 2.2, the assumption that the point process ξ —that generates the point cloud of sample locations—is independent of X is crucial. Otherwise, for any realization X , it would be possible to build a dependent point-referenced d -honeycomb adapted to the realized $L_X^T(u)$ such that $\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u)] \xrightarrow{\delta \rightarrow 0} C_{d-1}^*(u)$ and no corrective constant is required (see [Cotsakis et al. \(2023c\)](#), for $d = 2$). More precisely, for a square lattice, the strategy developed in [Cotsakis et al. \(2023c\)](#) avoids this asymptotic bias (see ([Cotsakis et al., 2023c](#), Remark 3) for further details) at the cost of an introducing an hyperparameter. On the contrary, with a deterministic point-referenced d -honeycomb, the dimensional constant $2d/\beta_d$ is unavoidable.

In the following corollary, we apply Theorem 2.2 to Poisson-Voronoi tessellations. The interested reader is also referred to ([Thäle and Yukich, 2016](#), Theorem 1.1).

Corollary 2.1. *Let $T \subset \mathbb{R}^d$ be a compact domain with non empty interior containing the origin. Let X be an isotropic random field on \mathbb{R}^d satisfying Assumption (A0), and let ξ be a homogeneous Poisson point process on \mathbb{R}^d with unit rate, independent of X . Let \mathcal{V}_ξ be the Poisson-Voronoi tessellation constructed from ξ , thus making $\dot{\mathcal{V}}_\xi$ a random point-referenced d -honeycomb. With $\delta > 0$ and $C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u)$ defined as in Definition 2.3, it holds that*

$$\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{V}}_\xi, T)}(u)] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d} C_{d-1}^*(u).$$

Proof. By Theorem 2.2, one only needs to show that $D^{(\delta\mathcal{V}_\xi)} \xrightarrow{\mathbb{P}} 0$ as $\delta \rightarrow 0$, where $D^{(\delta\mathcal{V}_\xi)} = \sup\{\text{diam}(\delta P \cap T) : \delta P \in \delta\mathcal{V}_\xi\}$. Let $M > 0$, denote by A_M the event $\{D^{(\delta\mathcal{V}_\xi)} > M\}$ and set $T+1 = \{r \in \mathbb{R}^d, d(r, T) \leq 1\}$. Note that if $\delta P \in \delta\mathcal{V}_\xi$ is such that $\text{diam}(\delta P \cap T) > M$, then there exists $r \in \delta P \cap T$ such that $(r + B_{\frac{M}{3}}^d) \cap \delta\xi = \emptyset$. To see this, consider that if $\text{diam}(\delta P \cap T) > M$, then there exists $r \in \delta P$ such that $d(r, \delta P^\bullet) > \frac{M}{3}$, and by the definition of the Voronoi diagram, $d(r, \delta P^\bullet) \leq d(r, \delta\tilde{P}^\bullet)$ for all $\delta\tilde{P}^\bullet \in \delta\xi$. It follows that

$$\begin{aligned} A_M = \{D^{(\delta\mathcal{V}_\xi)} > M\} &\subset \left\{ \exists r \in (T+1) : \left(\frac{r}{\delta} + B_{\frac{M}{3\delta}}^d \right) \cap \xi = \emptyset \right\} \\ &\subset \bigcup_{y \in \frac{M}{12\delta} \mathbb{Z} \cap \frac{T+1}{\delta}} \left\{ (y + B_{\frac{M}{6\delta}}^d) \cap \xi = \emptyset \right\}. \end{aligned}$$

It follows from the stationarity of ξ and the compactness of T that

$$\mathbb{P}(A_M) \leq \frac{\sigma_d(T+1)12^d}{M^d} \mathbb{P}\left(B_{\frac{M}{6\delta}}^d \cap \xi = \emptyset\right) = \frac{\sigma_d(T+1)12^d}{M^d} e^{-\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} \left(\frac{M}{6\delta}\right)^d} \rightarrow 0 \text{ as } \delta \rightarrow 0.$$

□

Remark 2.5 (Convergence of the bias factor). Theorem 2.2 *i*) offers insight about the limiting value of $\mathbb{E}[C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)]$ as $\delta \rightarrow 0$. However, this result does not imply the convergence of the bias factor to $2d/\beta_d$. It can be shown by slightly modifying the proofs of Theorems 2.1 and 2.2 that

$$E \left[\frac{C_{d-1}^{(\delta\dot{\mathcal{H}}, T)}(u)}{C_{d-1}^T(u)} \mid C_{d-1}^T(u) > 0 \right] \xrightarrow{\delta \rightarrow 0} \frac{2d}{\beta_d}, \quad (2.28)$$

with X an isotropic random field on \mathbb{R}^d satisfying Assumption (A0) and $\dot{\mathcal{H}}$ a point-referenced d -honeycomb as in Theorem 2.2 *i*). A sketch of the proof of the convergence in (2.28) is provided in Appendix 2.A.2.

2.3 Joint central limit theorem for the hypercubic lattice

In this section we prove a joint central limit theorem for the approximated volume and surface area of excursion sets of d -dimensional isotropic smooth random fields observed over a hypercubic lattice. Firstly, we rewrite estimates in (2.10)-(2.11) in this specific setting.

2.3.1 Estimators for the hypercubic lattice

For $\mathbf{i} \in \mathbb{Z}^d$ and $\delta \in \mathbb{R}^+$, let $V_{\mathbf{i}}(\delta) := \delta\mathbf{i} + [0, \delta]^d$. Define $\delta\mathcal{G} := \{V_{\mathbf{i}}(\delta) : \mathbf{i} \in \mathbb{Z}^d\} \in \mathfrak{T}^d$ (see Definition 2.2) and the point-referenced d -honeycomb

$$\delta\dot{\mathcal{G}} := \{(V_{\mathbf{i}}(\delta), \delta\mathbf{i}) : \mathbf{i} \in \mathbb{Z}^d\}.$$

Define

$$\mathbb{G}(\delta, T) := \{\delta\mathbf{i} : \mathbf{i} \in \mathbb{Z}^d, V_{\mathbf{i}}(\delta) \subseteq T\} \subset \delta\mathbb{Z}^d. \quad (2.29)$$

Then, referring to Definition 2.3, we write

$$C_d^{(\delta\dot{\mathcal{G}}, T)}(u) = \frac{\delta^d}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta, T)} \mathbb{1}_{\{X(t) \geq u\}}, \quad (2.30)$$

and

$$C_{d-1}^{(\delta\dot{\mathcal{G}}, T)}(u) = \frac{\delta^{d-1}}{\sigma_d(T)} \sum_{j=1}^d \sum_{\substack{t \in \mathbb{G}(\delta, T) \\ t + \delta\mathbf{e}_j \in \mathbb{G}(\delta, T)}} |\mathbb{1}_{\{X(t) \geq u\}} - \mathbb{1}_{\{X(t + \delta\mathbf{e}_j) \geq u\}}|. \quad (2.31)$$

In the following, without loss of generality, we assume that T is centered at the origin. Furthermore, we suppose that $0 < \delta \leq 1$ is chosen such that $T = [-\delta N, \delta N]^d$ for some $N \in \mathbb{N}^+$, which implies that $\bigcup_{P \in (\delta\mathcal{G})^T} P = T$, and that the $\text{Card}(\mathbb{G}(\delta, T)) = (2N)^d$ and $\sigma_d(T) = (2N\delta)^d$. Remark that this constricton implies that the deterministic bias ratio in Equation (2.12) is exactly unity.

In the sequel we simplify the notation of Equations (2.30)-(2.31) by writing $C_d^{(\delta, T)}(u)$ and $C_{d-1}^{(\delta, T)}(u)$, respectively. Furthermore, we write $T = T_N$ when the dependence in N needs to be explicitly specified.

2.3.2 The dominant role of the L_1 -norm for the hypercubic lattice

Note that, $C_{d-1}^T(u)$ in (2.6) can be rewritten, by the Crofton formula in (2.5), as, *a.s.*,

$$C_{d-1}^T(u) = \frac{1}{\sigma_d(T)} \int_{L_X^T(u)} \sigma_{d-1}(ds) = \frac{\sqrt{\pi} \Gamma(\frac{d+1}{2})}{\sigma_d(T) \Gamma(\frac{d}{2})} \int_{\mathbb{R}^{d-1}} \int_{\partial B_1^d} \frac{\sigma_0(L_X^T(u) \cap l_{\mathbf{s}, \mathbf{v}_{\mathbf{s}}(\mathbf{u})})}{\sigma_{d-1}(\partial B_1^d)} ds d\mathbf{u}. \quad (2.32)$$

A first intuition is that when δ gets small, the estimated surface area (2.31) gets close to (2.32). However, this is incorrect and we show in the following result that when δ gets small our estimate is close in the L^1 sense to the following random variable

$$\tilde{C}_{d-1}^T(u) := \frac{1}{\sigma_d(T)} \int_{L_X^T(u)} \frac{\|\nabla X(s)\|_1}{\|\nabla X(s)\|_2} \sigma_{d-1}(ds). \quad (2.33)$$

The difference between (2.32) (left side) and (2.33) is the ratio $\|\nabla X(s)\|_1/\|\nabla X(s)\|_2$ that induces the asymptotic bias. A similar competing behavior between the L^1 and L^2 norms was already visible in Equation (2.21) which leads to

$$\lim_{\delta \rightarrow 0} \mathbb{E} \left[C_{d-1}^{(\delta, T)}(u) \right] = p_{X(\mathbf{0})}(u) \mathbb{E} [\|\nabla X(\mathbf{0})\|_1 | X(\mathbf{0}) = u], \quad (2.34)$$

with $C_{d-1}^{(\delta, T)}(u)$ as in (2.31). Furthermore, by using the well-known Coarea formula (see, *e.g.* (Adler and Taylor, 2007, Equation (7.4.14))), \tilde{C}_{d-1}^T in (2.33) can be rewritten as

$$\tilde{C}_{d-1}^T(u) := \frac{1}{\sigma_d(T)} \lim_{\varrho \rightarrow 0} \frac{1}{2\varrho} \int_T \mathbf{1}_{\{|X(t)-u| \leq \varrho\}} \|\nabla X(t)\|_1 dt. \quad (2.35)$$

Then, from (2.35), one gets

$$\begin{aligned} \mathbb{E}[\tilde{C}_{d-1}^T(u)] &= \frac{1}{\sigma_d(T)} \int_T dt \lim_{\varrho \rightarrow 0} \frac{1}{2\varrho} \int_{u-\varrho}^{u+\varrho} \mathbb{E}[\|\nabla X(t)\|_1 | X(t) = x] p_{X(\mathbf{0})}(x) dx \\ &= p_{X(\mathbf{0})}(u) \mathbb{E} [\|\nabla X(\mathbf{0})\|_1 | X(\mathbf{0}) = u]. \end{aligned}$$

Equation (2.34) should be put in parallel with the desired limit given by Rice's formula (see, *e.g.*, (Azaïs and Wschebor, 2009, Equation (6.27)) or (Berzin et al., 2017, Proposition 2.2.1) (with $j = 1$) written for a stationary process), *i.e.*,

$$\mathbb{E}[C_{d-1}^T(u)] = p_{X(\mathbf{0})}(u) \mathbb{E}[\|\nabla X(\mathbf{0})\|_2 | X(\mathbf{0}) = u], \quad (2.36)$$

with $C_{d-1}^T(u)$ as in (2.6). This difference of norms in the limits in (2.34) and (2.36) motivates the presence of the dimensional constant β_d relating the L_1 and L_2 norms (see main Theorem 2.3 and Equation (2.14)).

Taking inspiration from the relationship between (2.6) and (2.32), the following proposition provides a similar result for \tilde{C}_{d-1}^T in (2.33) (item i) and a L^1 control between the estimator $C_{d-1}^{(\delta, T)}$ and \tilde{C}_{d-1}^T (item ii) as $\delta \rightarrow 0$. To this end, we introduce the following technical assumption.

(A3) Fix $j \in \{1, \dots, d\}$. Let $f(t) := (X(t), \partial_j X(t))$ for $t \in \mathbb{R}^d$, where $\partial_j X$ denotes the partial derivative of X in the direction \mathbf{e}_j . Fix $u \in \mathbb{R}$, and let $U \subset \mathbb{R}^2$ be a neighbourhood containing

the point $(u, 0)$. Suppose that the density of $f(\mathbf{0})$ is bounded uniformly on U . Define

$$W := \sup_{t \in [0,1]^d} \max_{1 \leq k \leq d} \{|\partial_k X(t)|, |\partial_k \partial_j X(t)|\}$$

and suppose that $\mathbb{E}[W^2 \mid f(\mathbf{0}) = s] < \infty$ for all $s \in U$.

Similarly, suppose that for some interval $I \subset \mathbb{R}$ containing u , the marginal density of $X(\mathbf{0})$ is bounded uniformly on I , and that $\mathbb{E}[\sup_{t \in [0,1]^d} \|\nabla X(t)\|_2 \mid X(\mathbf{0}) = s] < \infty$, for all $s \in I$.

This conditional expectation $\mathbb{E}[W^2 \mid f(\mathbf{0}) = s]$ is defined via the limit procedure in (Feller, 1966, Section III.2). The associated condition in (A3) is used in the proof of the key Lemma 2.3 in Section 2.4. It guarantees the integrability of the function $w^2 g(w, u + \epsilon, \epsilon)$, for small ϵ values, where g is the joint probability distribution of W , $X(\mathbf{0})$, and $\partial_j X(\mathbf{0})$. It allows for the use of techniques similar to the ones used in Leadbetter et al. (1983) (see the proof of Proposition 2.1 in Section 2.4). Lemma 2.3 is used in the proof of the following proposition.

Proposition 2.2. *Let X be a random field on \mathbb{R}^d as in Assumption (A0). It holds that*

$$i) \quad \tilde{C}_{d-1}^T(u) = \frac{1}{\sigma_d(T)} \sum_{j=1}^d \int_{\text{vect}(\mathbf{e}_j^\perp)} \sigma_0(L_X^T(u) \cap l_{\mathbf{e}_j, \mathbf{v}}) \, d\mathbf{v}, \quad a.s.$$

ii) *If, furthermore, X satisfies Assumptions (A1) and (A2) for $\mathbf{w} = \mathbf{e}_j$, for each $1 \leq j \leq d$, and for some $\varepsilon \in (0, 1)$, and (A3), then there exists a constant K such that for all $\delta \in (0, 1)$*

$$\mathbb{E} \left[\left| C_{d-1}^{(\delta, T)}(u) - \tilde{C}_{d-1}^T(u) \right| \right] \leq K \delta^{1-\varepsilon}. \quad (2.37)$$

The proof of Proposition 2.2 is postponed to Section 2.5.1. Remark that isotropy is not required in Proposition 2.2.

2.3.3 Strong alpha mixing random fields

We present and discuss sufficient hypotheses to prove the asymptotic normality in Theorem 2.3, below. We impose some spatial asymptotic independence conditions for the random field X which also apply to integrals of continuous functions over the level-curves of X (such as the surface area in (2.6)). To this end, mixing conditions are particularly appropriate (see for instance Cabaña (1987), Iribarren (1989)).

Definition 2.4 (Strongly mixing random field). Let $X := \{X(t) : t \in \mathbb{R}^d\}$ be a random field satisfying Assumption (A0), and let $\sigma_U := \sigma\{X(t) : t \in U\}$ for a subset $U \subset \mathbb{R}^d$, i.e., the σ -field generated by $\{X(t) : t \in U\}$. We define the following mixing coefficient for Borelian subsets $U, V \subset \mathbb{R}^d$,

$$\alpha(U, V) := \sup_{A \in \sigma_U, B \in \sigma_V} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\}. \quad (2.38)$$

Further we define

$$\alpha(s) := \sup\{\alpha(U, V) : d_2(U, V) > s\},$$

where $d_2(U, V) := \inf\{\|u - v\|_2 : u \in U, v \in V\}$. A random field X is said to be strongly mixing if $\alpha(s) \rightarrow 0$ for $s \rightarrow \infty$.

Existing results that establish asymptotic normality of geometric quantities classically rely on the continuous observation of X on T and on a quasi-association notion of dependence (see [Bulinski and Shashkin \(2007\)](#), [Meschenmoser and Shashkin \(2011\)](#), [Bulinski et al. \(2012\)](#), [Spodarev \(2014\)](#)) or are specific to Gaussian random fields (see [Meschenmoser and Shashkin \(2013\)](#), [Müller \(2017\)](#), [Kratz and Vadlamani \(2018\)](#)). There are also results for fields observed on the fixed lattice grid \mathbb{Z}^d in [Reddy et al. \(2018\)](#), where the notion of clustering spin model is introduced and is implied by either mixing assumptions or quasi-association. Finally, it is worth noting that [Bulinski \(2010\)](#) proposes a limit theorem for the empirical mean of X observed on a grid with vanishing size and large T , *i.e.*, the same grid of observation $\mathbb{G}(\delta, T)$ as in the present work.

Corollary 3 in [Dedecker \(1998\)](#) (see also [Bolthausen \(1982\)](#)) proposes minimal α -mixing conditions to get central limit results for stationary random fields observed on the fixed lattice \mathbb{Z}^d . However, we can not rely on such results as we are interested in non-trivial functionals of the field and we aim at imposing the conditions on the underlying field X and not on the observed sequence on the varying lattice $\delta\mathbb{Z}^d$. Instead, to establish [Theorem 2.3](#) below, we heavily rely on the inheritance properties of mixing sequences: if a random field X is strongly alpha-mixing (see [Definition 2.4](#)), so is any measurable transformation of it. The latter property does not hold for quasi-association properties. Examples of mixing random fields include Gaussian random fields (see [Doukhan \(1994\)](#) [Section 2.1](#) and [Corollary 2](#) for an explicit control of [\(2.38\)](#)), and therefore any transformation of Gaussian random fields such as Student or chi-square fields, or Max-infinitely divisible random fields (see [Dombry and Eyi-Minko \(2012\)](#)). We refer the reader to [Bradley \(2005\)](#) for other examples of processes satisfying various mixing conditions.

2.3.4 Joint Central Limit Theorem

In the following, $\xrightarrow{\mathcal{L}}$ denotes the convergence in law. Taking advantage of the results of [Section 2.2](#) and adapting the results of [Iribarren \(1989\)](#), we state the following joint limit result for the approximated volume and surface area.

Theorem 2.3 (Joint central limit theorem for $C_d^{(\delta, T_N)}$ and $C_{d-1}^{(\delta, T_N)}$). *Let X be an isotropic random field satisfying Assumptions [\(A0\)](#), [\(A1\)](#), [\(A2\)](#) for some $\varepsilon \in (0, 1)$, and [\(A3\)](#). Assume that X is strongly mixing as in [Definition 2.4](#) and that for some $\eta > 0$, the mixing coefficients satisfy the rate condition*

$$\sum_{r=1}^{+\infty} r^{3d-1} \alpha(r)^{\frac{\eta}{2+\eta}} < +\infty, \tag{2.39}$$

and $\mathbb{E}[\sigma_{d-1}(L_X(u) \cap [0, 1]^d)^{2+\eta}] < +\infty$. Let $(T_N)_{N \geq 1}$ be a sequence of hypercubes in \mathbb{R}^d such that $\sigma_d(T_N) = (2N\delta)^d$. Define

$$C^{(\delta, T_N)}(u) := \left(C_d^{(\delta, T_N)}(u), C_{d-1}^{(\delta, T_N)}(u) \right)^t \quad \text{and} \quad C^*(u) := \left(C_d^*(u), \frac{2d}{\beta_d} C_{d-1}^*(u) \right)^t,$$

with C_d (resp. C_{d-1}) as in (2.30) (resp. in (2.31)) on the hypercubic lattice $\mathbb{G}(\delta, T)$ in (2.29), $C_k^*(u)$ as in (2.8), where $(\cdot)^t$ denotes the matrix transposition and β_d is as in (2.14). Then, there exists a finite and positive definite covariance matrix $\Sigma(u)$ such that, it holds

$$\sqrt{\sigma_d(T_N)} \left(C^{(\delta, T_N)}(u) - C^*(u) \right) \xrightarrow{\mathcal{L}} \mathcal{N}_2(0, \Sigma(u)),$$

as $\delta \rightarrow 0$, $N\delta \rightarrow \infty$, and $(N\delta)^d \delta^{1-\varepsilon} \rightarrow 0$, with ε as in Assumption (A2), with variances

$$\begin{aligned} \Sigma_{(1,1)}(u) &= \int_{\mathbb{R}^d} \text{Cov}(\mathbf{1}_{\{X(\mathbf{0}) \geq u\}}, \mathbf{1}_{\{X(t) \geq u\}}) dt, \\ \Sigma_{(2,2)}(u) &= \int_{\mathbb{R}^d} \left[p_{X(\mathbf{0}), X(t)}(u, u) \mathbb{E}[\|\nabla X(\mathbf{0})\|_1 \|\nabla X(t)\|_1 \mid X(\mathbf{0}) = X(t) = u] \right. \\ &\quad \left. - \left(p_{X(\mathbf{0})}(u) \mathbb{E}[\|\nabla X(\mathbf{0})\|_1 \mid X(\mathbf{0}) = u] \right)^2 \right] dt. \end{aligned}$$

and covariances $\Sigma_{(1,2)}(u) = \lim_{N \rightarrow \infty} \sigma_d(T_N) \text{Cov}(C_d^{T_N}(u), C_{d-1}^{T_N}(u))$.

The mixing condition (2.39) come from (Iribarren, 1989, Proposition 1 (ii)), by choosing $q = 2$. This mixing rate condition relies on the dimension d of the field X , the moment condition on $\sigma_{d-1}(L_X(u) \cap [0, 1]^d)$ and the decay behavior of the mixing coefficient $\alpha(s)$, as $s \rightarrow \infty$.

Proof of Theorem 2.3. Let $C^{T_N}(u) := \left(C_d^{T_N}(u), \tilde{C}_{d-1}^{T_N}(u) \right)^t$, with $C_d^{T_N}(u)$ as in (2.7) and $\tilde{C}_{d-1}^{T_N}(u)$ as in (2.33). We decompose

$$\begin{aligned} \sqrt{\sigma_d(T_N)} \left(C^{(\delta, T_N)}(u) - C^*(u) \right) &= \sqrt{\sigma_d(T_N)} \left(C^{(\delta, T_N)}(u) - C^{T_N}(u) \right) \\ &+ \sqrt{\sigma_d(T_N)} \left(C^{T_N}(u) - \mathbb{E}[C^{T_N}(u)] \right) + \sqrt{\sigma_d(T_N)} \left(\mathbb{E}[C^{T_N}(u)] - \mathbb{E}[C^{(\delta, T_N)}(u)] \right) \\ &+ \sqrt{\sigma_d(T_N)} \left(\mathbb{E}[C^{(\delta, T_N)}(u)] - C^*(u) \right) := I_1 + I_2 + I_3 + I_4. \end{aligned}$$

From Proposition 2.3, we get that first coordinates of I_1 and I_3 go in probability to zero. The second coordinates are handled with Proposition 2.2. It follows that $I_1 \xrightarrow{\mathbb{P}} 0$ and $I_3 \xrightarrow{\mathbb{P}} 0$, for $N\delta \rightarrow \infty$, $\delta \rightarrow 0$ and $(N\delta)^d \delta^{1-\varepsilon} \rightarrow 0$ as $N \rightarrow \infty$, with ε as in Assumption (A2).

The joint central limit theorem under mixing conditions (see Theorem 2.4, for $n = N\delta$) for the random vector $C^{T_N}(u)$ gives that $I_2 \xrightarrow{\mathcal{L}} \mathcal{N}_2(0, \Sigma(u))$. The given asymptotic variances come from Cotsakis et al. (2023c) (see Equation (11)), Equation (2.33) and covariances in Bulinski et al. (2012). Notice that $C_d^{(\delta, T)}(u)$ in (2.30) does not generate any bias for the volume measure σ_d (see (2.9)),

so the first coordinate of I_4 is trivially equal to zero. The second coordinate of I_4 is obtained via Corollary 2.2. Then, by Slutsky's Theorem, we obtain the result. \square

All auxiliary results necessary for the proof of the joint central limit theorem for $(C_d^{(\delta, T_N)}, C_{d-1}^{(\delta, T_N)})$ are provided in Section 2.5.2.

Remark 2.6. The regime restriction $(N\delta)^d \delta^{1-\varepsilon} \rightarrow 0$ (where ε can be small) is imposed by Proposition 2.2 and Proposition 2.3. It can be improved by requiring more stringent assumptions: *e.g.*, if X is quasi-associated (see Bulinski and Shashkin (2007) and Bulinski et al. (2012)) and under a decay assumption for its correlation function (see Appendix 2.A). However, improving the rate in Proposition 2.2 seems much more delicate and is out of the scope of the present work.

Generalizing Theorem 2.3 to point clouds other than hypercubic lattices requires that one first identifies the associated random variable when δ gets small, *i.e.*, an analog to (2.33), and that one proves an analog of (2.37). In this sense, this asymptotic result is lattice-dependent and generalizing it to general tessellations is also an interesting open point.

2.4 Additional results and proofs associated to Section 2.2

2.4.1 Auxiliary lemmas on crossings

The following technical lemma controls the probability that a one-dimensional random process crosses the level u more than once on a small interval of length t .

Lemma 2.1 (Level crossings of random processes). *Let $Y = \{Y(s) : s \in \mathbb{R}\}$ be a one-dimensional stationary random process with twice continuously differentiable sample paths. Suppose that the probability density function of $Y(0) | \{Y'(0) = 0\}$ is uniformly bounded by $M < \infty$ and that $\|Y''\|_{\infty, [0,1]} \in L^p$ for some $p > \frac{1}{2}$. For $u \in \mathbb{R}$, let $Y^{-1}(u) := \{s \in \mathbb{R} : Y(s) = u\}$. Then, there exists a constant $K \in \mathbb{R}^+$ such that*

$$\mathbb{P}(\text{Card}(Y^{-1}(u) \cap [0, t]) \geq 2) \leq K t^{\frac{3p}{p+1}},$$

for all $t \in [0, 1]$.

Proof of Lemma 2.1. If there are to exist two points $s_1, s_2 \in [0, t]$, $s_1 < s_2$ such that $Y(s_1) = Y(s_2) = u$, then by Rolle's theorem, there exists a $c \in [s_1, s_2]$ such that $Y'(c) = 0$. Moreover, by Taylor's theorem,

$$Y(s_2) = u = Y(c) + \frac{(s_2 - c)^2}{2} Y''(\eta),$$

for some $\eta \in [c, s_2]$, which implies $2|Y(c) - u| \leq t^2 \|Y''\|_{\infty, [0,1]}$, where as Y is twice continuously differentiable, $\|Y''\|_{\infty, [0,1]} < \infty$ almost surely. It implies that for any $x \in (0, 1)$, $\{\text{Card}(Y^{-1}(u) \cap [0, t]) \geq 2\} \subseteq \{\|Y''\|_{\infty, [0,1]} \geq (2/t^2)^x\} \cup \{\exists c \in [0, t] : Y'(c) = 0, |Y(c) - u| \leq (t^2/2)^{1-x}\}$.

To control the probability of the first event, Markov's inequality gives

$$\mathbb{P}(\|Y''\|_{\infty,[0,1]} \geq (2/t^2)^x) = \mathbb{P}(\|Y''\|_{\infty,[0,1]}^p \geq (2/t^2)^{px}) \leq \left(\frac{t^2}{2}\right)^{px} \mathbb{E}[\|Y''\|_{\infty,[0,1]}^p].$$

As for the second event, we write

$$\begin{aligned} & \mathbb{P}(\exists c \in [0, t] : Y'(c) = 0, |Y(c) - u| \leq (t^2/2)^{1-x}) \\ &= \mathbb{P}(\text{Card}(\{c \in [0, t] : Y'(c) = 0\}) \geq 1) \mathbb{P}(|Y(s) - u| \leq (t^2/2)^{1-x} \mid Y'(s) = 0, s \in [0, t]) \\ &\leq \mathbb{E}[\text{Card}(\{c \in [0, t] : Y'(c) = 0\})] \mathbb{P}(|Y(s) - u| \leq (t^2/2)^{1-x} \mid Y'(s) = 0). \end{aligned}$$

By Theorem 3.3.1 in [Beutler and Leneman \(1966\)](#), $\mathbb{E}[\text{Card}(\{c \in [0, t] : Y'(c) = 0\})] = t\mathbb{E}[\text{Card}(\{c \in [0, 1] : Y'(c) = 0\})]$. In addition, $\mathbb{P}(|Y(s) - u| \leq (t^2/2)^{1-x} \mid Y'(s) = 0) \leq 2M(t^2/2)^{1-x}$ as Y is stationary and $Y(0) \mid \{Y'(0) = 0\}$ has bounded density. In total, we have shown that

$$\begin{aligned} \mathbb{P}(\text{Card}(Y^{-1}(u) \cap [0, t]) \geq 2) &\leq \left(\frac{t^2}{2}\right)^{px} \mathbb{E}[\|Y''\|_{\infty,[0,1]}^p] + \\ &2M\mathbb{E}[\text{Card}(\{c \in [0, 1] : Y'(c) = 0\})] t \left(\frac{t^2}{2}\right)^{1-x}. \end{aligned}$$

Thus, optimizing in x leads to $x := 3/(2p + 2) \in (0, 1)$, if $p > \frac{1}{2}$, and to the result. \square

The following lemma provides a bound on the crossing probability for general, possibly anisotropic, random fields.

Lemma 2.2. *Let X be a random field on \mathbb{R}^d satisfying [\(A0\)](#). Fix $t \in \mathbb{R}^d$ and $u \in \mathbb{R}$. Then,*

$$\mathbb{P}(X(\mathbf{0}) \leq u < X(t)) \leq C_{d-1}^*(u) \|t\|_2.$$

Proof of Lemma 2.2. Let \bar{t} denote the line segment in \mathbb{R}^d whose endpoints are $\mathbf{0}$ and t . Define $K_X(u) := \{L_X(u) - \lambda t : \lambda \in (0, 1)\}$ to be the set of points that are at a distance of at most $\|t\|_2$ from $L_X(u)$ in the direction of $-t$ (see [Figure 2.5](#) for an illustration of the set $K_X(u)$ in \mathbb{R}^2). By the assumption of stationarity it holds that

$$\mathbb{P}(X(\mathbf{0}) \leq u < X(t)) \leq \mathbb{P}(\bar{t} \cap L_X(u) \neq \emptyset) = \mathbb{P}(\mathbf{0} \in K_X(u)) = \mathbb{E} \left[\frac{\sigma_d(K_X(u) \cap B_r^d)}{\sigma_d(B_r^d)} \right],$$

for all $r \in \mathbb{R}^+$. If we denote $r^+ := r + \|t\|_2$ for $r \in \mathbb{R}^+$, then

$$\begin{aligned} \sigma_d(K_X(u) \cap B_r^d) &\leq \int_{L_X(u) \cap B_{r^+}^d} \left\langle t, \frac{\nabla X(s)}{\|\nabla X(s)\|_2} \right\rangle \sigma_{d-1}(ds) \leq \int_{L_X(u) \cap B_{r^+}^d} \|t\|_2 \sigma_{d-1}(ds) \\ &= \|t\|_2 \sigma_{d-1}(L_X(u) \cap B_{r^+}^d), \quad a.s. \end{aligned}$$

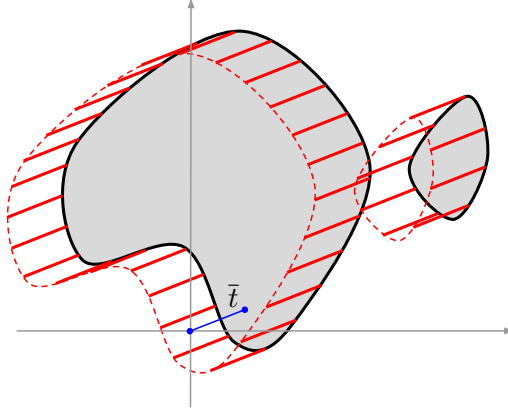


Figure 2.5: In dimension 2, for the level curve $L_X(u)$ in black, $K_X(u)$ is the area covered by solid red line segments. Note that the line segment \bar{t} (in blue) crosses $L_X(u)$ (in black) if and only if $\mathbf{0} \in K_X(u)$.

Taking the expectation and sending $r \rightarrow \infty$ yields the result, since $\sigma_d(B_r^d)/\sigma_d(B_{r+}^d) \rightarrow 1$. \square

Definition 2.5. Define the map $\pi_j : \mathbb{R}^d \rightarrow \mathbb{R}^d$ to be the orthogonal projection onto the $(d-1)$ -dimensional subspace $\text{vect}(\mathbf{e}_j^\perp)$. That is, for $S \subseteq \mathbb{R}^d$,

$$\pi_j(S) = \left\{ \sum_{i \neq j} s_i \mathbf{e}_i : \mathbf{s} \in S \right\} = \{ \mathbf{v} \in \text{vect}(\mathbf{e}_j^\perp) : l_{\mathbf{e}_j, \mathbf{v}} \cap S \neq \emptyset \},$$

with $l_{\mathbf{e}_j, \mathbf{v}}$ as in (2.3).

The following lemma allows to obtain the rate of convergence of a Riemann sum used in Proposition 2.2. The techniques used in the proof can be used to bound the probability that a random manifold intersects a small region of space.

Lemma 2.3. *Let X be a stationary random field on \mathbb{R}^d satisfying Assumption (A0) and (A3). Fix $j \in \{1, \dots, d\}$. For $\mathbf{v} \in \text{vect}(\mathbf{e}_j^\perp)$, denote $n_{\mathbf{v}} := \sigma_0(L_X(u) \cap l_{\mathbf{e}_j, \mathbf{v}} \cap [0, 1]^d)$, with $l_{\mathbf{e}_j, \mathbf{v}}$ as in (2.3). Then, there exists a constant $K \in \mathbb{R}^+$ such that for all $\delta \in [0, 1)$,*

$$\mathbb{P} \left(\sup_{\mathbf{v}_1, \mathbf{v}_2 \in \pi_j([0, \delta]^d)} |n_{\mathbf{v}_1} - n_{\mathbf{v}_2}| > 0 \right) \leq K\delta.$$

Proof of Lemma 2.3. Note that $n_{\mathbf{v}}$ seen as a function that maps $\mathbf{v} \in \text{vect}(\mathbf{e}_j^\perp)$ into \mathbb{N}_0 is almost surely piecewise constant on $\pi_j([0, 1]^d)$. The discontinuities occur at points in $\pi_j(H_1 \cup H_2 \cup H_3)$, where $H_1 := f^{-1}((u, 0)) \cap [0, 1]^d$ —recall from Assumption (A3) that $f = (X, \partial_j X)$, $H_2 := \{\mathbf{s} \in L_X(u) : s_j = 0\}$, and $H_3 := \{\mathbf{s} \in L_X(u) : s_j = 1\}$ (see Figure 2.6 for an example with $d = 2$).

For $i \in \{1, 2, 3\}$, we aim to show that $\mathbb{P}(\pi_j([0, \delta]^d) \cap \pi_j(H_i) \neq \emptyset)$ is of the order of δ , which implies the desired result.

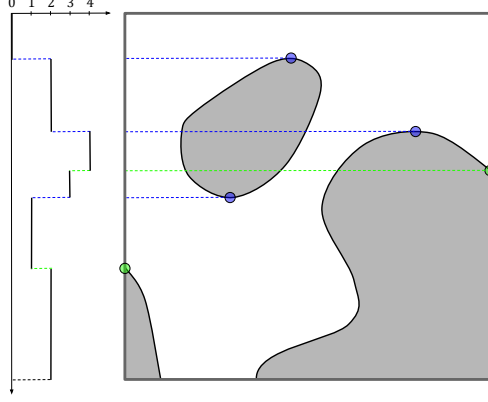


Figure 2.6: On the left a representation of the piecewise constant function $n_{\mathbf{v}}$ of $\mathbf{v} \in \text{vect}(\mathbf{e}_1^\perp)$ computed for the excursion set represented in grey on the right. The points in blue are in H_1 and the points in green are in H_2 and H_3 .

We begin with the case of H_1 . By stationarity, we apply the union bound,

$$\mathbb{P}(\pi_j([0, \delta]^d) \cap \pi_j(H_1) \neq \emptyset) \leq \lceil 1/\delta \rceil \mathbb{P}([0, \delta]^d \cap H_1 \neq \emptyset), \quad (2.40)$$

where $\lceil \cdot \rceil$ denotes the ceiling function.

By Taylor's theorem, for each $t \in [0, \delta]^d$, there exists an $s \in [0, \delta]^d$ such that $f(t) = f(\mathbf{0}) + \langle t, \nabla f(s) \rangle$. If $[0, \delta]^d \cap H_1 \neq \emptyset$, then $(u, 0) - f(\mathbf{0}) = \langle t, \nabla f(s) \rangle$, for some $s, t \in [0, \delta]^d$.

Let us denote by $M_1 \in \mathbb{R}^+$ the constant which uniformly bounds the density of $f(\mathbf{0})$ on U , and M_2 the constant which bounds the marginal density of $X(\mathbf{0})$ on I (see Assumption [\(A3\)](#)).

For $\delta \leq 1$, we bound the variation of $f(\mathbf{0})$ by writing

$$\|(u, 0) - f(\mathbf{0})\|_\infty \leq \|t\|_1 W \leq d\delta W.$$

Therefore, we obtain

$$\mathbb{P}([0, \delta]^d \cap H_1 \neq \emptyset) \leq \mathbb{P}(\|(u, 0) - f(\mathbf{0})\|_\infty \leq d\delta W).$$

let $g(w, x, \dot{x})$ be the joint probability distribution of W , $X(\mathbf{0})$, and $\partial_j X(\mathbf{0})$. Note that

$$\begin{aligned} \mathbb{P}(\|(u, 0) - f(\mathbf{0})\|_\infty \leq d\delta W) &= \int_0^\infty \int_{u-\delta wd}^{u+\delta wd} \int_{-\delta wd}^{\delta wd} g(w, x, \dot{x}) \, d\dot{x} dx dw \\ &= (2d\delta)^2 \int_0^\infty \int_{-w/2}^{w/2} \int_{-w/2}^{w/2} g(w, u + 2d\delta z, 2d\delta \dot{z}) \, d\dot{z} dz dw, \end{aligned}$$

where a change of the variables $x = u + 2d\delta z$ and $\dot{x} = 2d\delta\dot{z}$ was used. Now, by the dominated convergence theorem,

$$\begin{aligned} \lim_{\delta \rightarrow 0} \frac{1}{\delta^2} \mathbb{P}(\|(u, 0) - f(\mathbf{0})\|_\infty \leq d\delta W) &= \lim_{\delta \rightarrow 0} 4d^2 \int_0^\infty \int_{-w/2}^{w/2} \int_{-w/2}^{w/2} g(w, u + 2d\delta z, 2d\delta \dot{z}) \, d\dot{z} dz dw \\ &= 4d^2 \int_0^\infty w^2 g(w, u, 0) \, dw \leq 4d^2 M_1 \mathbb{E}[W^2 \mid f(\mathbf{0}) = (u, 0)], \end{aligned}$$

which is finite by (A3). Thus, $\mathbb{P}([0, \delta]^d \cap H_1 \neq \emptyset) = O(\delta^2)$, and by (2.40), $\mathbb{P}(\pi_j([0, \delta]^d) \cap \pi_j(H_1) \neq \emptyset) = O(\delta)$, as desired.

Now, we consider the case of H_2 (the case of H_3 being identical). Using similar arguments as in the case of H_1 , we readily obtain

$$\begin{aligned} \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P}(\pi_j([0, \delta]^d) \cap \pi_j(H_2) \neq \emptyset) &= \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P}([0, \delta]^d \cap H_2 \neq \emptyset) \\ &\leq \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P}\left(|u - X(\mathbf{0})| \leq \sqrt{d}\delta \sup_{t \in [0, 1]^d} \|\nabla X(t)\|_2\right) \\ &\leq 2\sqrt{d}M_2 \mathbb{E}\left[\sup_{t \in [0, 1]^d} \|\nabla X(t)\|_2 \mid X(\mathbf{0}) = u\right] < \infty. \end{aligned}$$

Thus, $\mathbb{P}(\pi_j([0, \delta]^d) \cap \pi_j(H_2) \neq \emptyset) = O(\delta)$ as desired. \square

2.4.2 Proof of Proposition 2.1

First note that

$$\{X(\mathbf{0}) \leq u < X(q\mathbf{e}_1)\} = \{X(\mathbf{0}) \leq u < X(\mathbf{0}) + qX_q\} = \{X(\mathbf{0}) \leq u\} \cap \left\{\frac{1}{q}(u - X(\mathbf{0})) < X_q\right\}.$$

It follows that

$$\begin{aligned} \frac{1}{q} \mathbb{P}(X(\mathbf{0}) \leq u < X(q)) &= \frac{1}{q} \mathbb{P}(X(\mathbf{0}) \leq u, X_q > q^{-1}(u - X(\mathbf{0}))) \\ &= \frac{1}{q} \int_{-\infty}^u \int_{\frac{1}{q}(u-x)}^\infty g_q(x, y) \, dy dx = \frac{1}{q} \int_0^\infty \int_{u-xy}^u g_q(x, y) \, dx dy = \int_0^\infty z \int_0^1 g_q(u - qzv, z) \, dv dz, \end{aligned}$$

using the change of variable $v = \frac{u-x}{qy}$, $z = y$ ($x = u - qzv$, $y = z$). From the assumptions, we derive that $g_q(u - qzv, z) \rightarrow p(u, z)$ pointwise as $q \rightarrow 0$, and the dominated convergence theorem applies and leads to the result.

2.5 Additional results and proofs associated to Section 2.3

2.5.1 Proof of Proposition 2.2

Firstly, we prove the first item in Proposition 2.2. Let (A_1, \dots, A_k) be a sequence of polytopes residing in $(d-1)$ -dimensional hyperplanes in \mathbb{R}^d . For $i \in \{1, \dots, k\}$, let $\mathbf{n}_i \in \mathbb{R}^d$ denote the unit normal vector to A_i . We aim to show

$$\sum_{i=1}^k \sum_{j=1}^d |\langle \mathbf{n}_i, \mathbf{e}_j \rangle| \sigma_{d-1}(A_i) = \sum_{j=1}^d \int_{\text{vect}(\mathbf{e}_j^\perp)} \sigma_0 \left(\left(\bigcup_{i=1}^k A_i \right) \cap l_{\mathbf{e}_j, \mathbf{v}} \right) d\mathbf{v}, \quad (2.41)$$

with $l_{\mathbf{e}_j, \mathbf{v}}$ as in (2.3). By noticing that for all $\mathbf{v} \in \text{vect}(\mathbf{e}_j^\perp)$,

$$\sigma_0 \left(\left(\bigcup_{i=1}^k A_i \right) \cap l_{\mathbf{e}_j, \mathbf{v}} \right) = \sum_{i=1}^k \mathbb{1}_{\{A_i \cap l_{\mathbf{e}_j, \mathbf{v}} \neq \emptyset\}},$$

we see that it suffices to show that for each pair of indices $(i, j) \in \{1, \dots, k\} \times \{1, \dots, d\}$,

$$|\langle \mathbf{n}_i, \mathbf{e}_j \rangle| \sigma_{d-1}(A_i) = \int_{\text{vect}(\mathbf{e}_j^\perp)} \mathbb{1}_{\{A_i \cap l_{\mathbf{e}_j, \mathbf{v}} \neq \emptyset\}} d\mathbf{v}. \quad (2.42)$$

Equation (2.42) is evident, since both sides of the equality describe the σ_{d-1} measure of the projection of A_i onto the hyperplane $\text{vect}(\mathbf{e}_j^\perp)$. Note that $\sum_{j=1}^d |\langle \mathbf{n}_i, \mathbf{e}_j \rangle| = \|\mathbf{n}_i\|_1$.

Let us consider $t \in L_X(u)$ and $\nabla X(t)/\|\nabla X(t)\|_2$ be the unit normal vector to $L_X(u)$ at t . Notice that $L_X^T(u)$ is a C^1 smooth manifold. Then, one can consider a set of points $t_i \in L_X(u)$, for $i \in \{1, \dots, k\}$ and build a sequence of polygons tangent to the surface $L_X^T(u)$ in these points. For this choice of A_i , the left-hand side of Equation (2.41) can be seen as a Riemann sum that approximates the integral expression for $\tilde{C}_{d-1}^T(u)$ in (2.33), when k goes to infinity. This proves item i).

Now, we prove the second item in Proposition 2.2. For $\mathbf{g} \in \mathbb{G}(\delta, T)$ in (2.29), let $Z_{\mathbf{g}} := \mathbb{1}_{\{X(\mathbf{g}) \geq u\}}$. Equations (2.30)-(2.31) can be rewritten as

$$\begin{aligned} C_d^{(\delta, T)}(u) &= \frac{\delta^d}{\sigma_d(T)} \sum_{\mathbf{g} \in \mathbb{G}(\delta, T)} Z_{\mathbf{g}} = \frac{1}{(2N)^d} \sum_{\mathbf{g} \in \mathbb{G}(\delta, T)} Z_{\mathbf{g}}, \\ C_{d-1}^{(\delta, T)}(u) &= \frac{\delta^{d-1}}{\sigma_d(T)} \sum_{j=1}^d \sum_{\substack{\mathbf{g} \in \mathbb{G}(\delta, T) \\ \mathbf{g} + \delta \mathbf{e}_j \in \mathbb{G}(\delta, T)}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}}. \end{aligned}$$

Set $G := \mathbb{G}(\delta, T) \cap [0, 1]^d$ and $n_{\mathbf{v}} := \sigma_0(L_X(u) \cap l_{\mathbf{e}_j, \mathbf{v}} \cap [0, 1]^d)$, suppose that we have established

$$\mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{g} \in G} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - \int_{\text{vect}(\mathbf{e}_j^\perp) \cap [0, 1]^d} n_{\mathbf{v}} d\mathbf{v} \right| \right] = O(\delta^{1-\varepsilon}), \quad (2.43)$$

for all $j \in \{1, \dots, d\}$. Then, by stationarity and the triangle inequality,

$$\mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{g} \in \mathbb{G}(\delta, T)} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - \int_{\text{vect}(\mathbf{e}_j^\perp)} \sigma_0(L_X^T(u) \cap l_{\mathbf{e}_j, \mathbf{v}}) d\mathbf{v} \right| \right] = O((N\delta)^d \delta^{1-\varepsilon})$$

so the desired result holds by summing over the d dimensions and again applying the triangle inequality together with the first item. Therefore, it suffices to show (2.43). This is done in two steps. First, we show that (recall Definition 2.5)

$$\mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{g} \in G} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} n_{\mathbf{v}} \right| \right] = O(\delta^{1-\varepsilon}). \quad (2.44)$$

Second, we show that

$$\mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} n_{\mathbf{v}} - \int_{\text{vect}(\mathbf{e}_j^\perp) \cap [0, 1]^d} n_{\mathbf{s}} d\mathbf{s} \right| \right] = O(\delta^{1-\varepsilon}). \quad (2.45)$$

Let $M^{(\delta)} := (\sigma_0(G))^{1/d} = \lfloor \delta^{-1} \rfloor + 1$ be the number of rows in the grid G that is contained in $[0, 1]^d$. By construction, $M^{(\delta)}\delta \leq 1 + \delta$ for all $\delta \in \mathbb{R}^+$.

To see that (2.44) holds, we use the triangle inequality to write

$$\mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{g} \in G} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} n_{\mathbf{v}} \right| \right] \leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \mathbb{E} \left[\left| \sum_{\mathbf{g} \in G \cap l_{\mathbf{e}_j, \mathbf{v}}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - n_{\mathbf{v}} \right| \right]. \quad (2.46)$$

Moreover, for fixed $\mathbf{v} \in \pi_j(G)$, $\sum_{\mathbf{g} \in G \cap l_{\mathbf{e}_j, \mathbf{v}}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}}$ approaches $n_{\mathbf{v}}$ from below, and both quantities take values in \mathbb{N}_0 . If the two quantities are distinct, then there must be two elements of $L_X(u) \cap [0, 1]^d \cap l_{\mathbf{e}_j, \mathbf{v}}$ with a spacing of less than δ . Let $A_1^{(\mathbf{v})}$ denote the event $\left\{ \sum_{\mathbf{g} \in G \cap l_{\mathbf{e}_j, \mathbf{v}}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} \neq n_{\mathbf{v}} \right\}$. With $(\cdot)_j$ denoting the projection onto the j^{th} component, note that

$$A_1^{(\mathbf{v})} \subset \bigcup_{i=0}^{M^{(\delta)}-1} \left\{ \text{Card}([i\delta, (i+2)\delta] \cap (L_X(u) \cap l_{\mathbf{e}_j, \mathbf{v}})_j) \geq 2 \right\},$$

since any two points in \mathbb{R} with a spacing of less than δ will be contained in an interval $[i\delta, (i+2)\delta]$ for some $i \in \mathbb{Z}$. It follows from Assumption (A1), Lemma 2.1 with $p = 2$, and the definition of $M^{(\delta)}$

that $\mathbb{P}(A_1^{(\mathbf{v})}) = O(\delta)$. This fact, with Hölder's inequality yields for the ε defined in (A2),

$$\begin{aligned} \mathbb{E} \left[\left| \sum_{\mathbf{g} \in G \cap l_{\mathbf{e}_j, \mathbf{v}}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - n_{\mathbf{v}} \right| \right] &= \mathbb{E} \left[\left| \sum_{\mathbf{g} \in G \cap l_{\mathbf{e}_j, \mathbf{v}}} \mathbb{1}_{\{Z_{\mathbf{g}} \neq Z_{\mathbf{g} + \delta \mathbf{e}_j}\}} - n_{\mathbf{v}} \right| \mathbb{1}_{A_1^{(\mathbf{v})}} \right] \\ &\leq \mathbb{E} [n_{\mathbf{v}} \mathbb{1}_{A_1^{(\mathbf{v})}}] \leq (\mathbb{E}[n_{\mathbf{v}}^{1/\varepsilon}])^\varepsilon \mathbb{P}(A_1^{(\mathbf{v})})^{1-\varepsilon} = O(\delta^{1-\varepsilon}). \end{aligned}$$

By the stationarity of X , each of the $(M^{(\delta)})^{d-1}$ terms in (2.46) is identical, and as $M^{(\delta)}\delta \leq 1 + \delta$, (2.44) follows immediately. We continue by showing (2.45). For $\mathbf{v} \in \text{vect}(\mathbf{e}_j^\perp)$, define the event $A_2^{(\mathbf{v})} := \left\{ \sup_{\mathbf{s} \in \mathbf{v} + \pi_j([0, \delta]^d)} |n_{\mathbf{v}} - n_{\mathbf{s}}| > 0 \right\}$. Then, the triangle inequality gives

$$\begin{aligned} \mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} n_{\mathbf{v}} - \int_{\text{vect}(\mathbf{e}_j^\perp) \cap [0, 1]^d} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right| \right] &\leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \mathbb{E} \left[\left| n_{\mathbf{v}} - \delta^{1-d} \int_{\mathbf{v} + \pi_j([0, \delta]^d)} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right| \right] \\ &= \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \mathbb{E} \left[\left| n_{\mathbf{v}} - \delta^{1-d} \int_{\mathbf{v} + \pi_j([0, \delta]^d)} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right| \mathbb{1}_{A_2^{(\mathbf{v})}} \right] \\ &\leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \left(\mathbb{E} \left[\left| n_{\mathbf{v}} - \delta^{1-d} \int_{\mathbf{v} + \pi_j([0, \delta]^d)} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right|^{\frac{1}{\varepsilon}} \right] \right)^\varepsilon \mathbb{P}(A_2^{(\mathbf{v})})^{1-\varepsilon} \\ &\leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \left((\mathbb{E}[n_{\mathbf{0}}^{\frac{1}{\varepsilon}}])^\varepsilon + \left(\mathbb{E} \left[\left| \delta^{1-d} \int_{\mathbf{v} + \pi_j([0, \delta]^d)} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right|^{\frac{1}{\varepsilon}} \right] \right)^\varepsilon \right) \mathbb{P}(A_2^{(\mathbf{v})})^{1-\varepsilon} \\ &\leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} \left((\mathbb{E}[n_{\mathbf{0}}^{\frac{1}{\varepsilon}}])^\varepsilon + \left(\delta^{1-d} \int_{\mathbf{v} + \pi_j([0, \delta]^d)} \mathbb{E}[n_{\mathbf{s}}^{\frac{1}{\varepsilon}}] \, \text{d}\mathbf{s} \right)^\varepsilon \right) \mathbb{P}(A_2^{(\mathbf{v})})^{1-\varepsilon} \\ &\leq \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} 2(\mathbb{E}[n_{\mathbf{0}}^{\frac{1}{\varepsilon}}])^\varepsilon \mathbb{P}(A_2^{(\mathbf{v})})^{1-\varepsilon}, \end{aligned}$$

where we used Jensen inequality for the penultimate inequality and the fact that $\sigma_{d-1}(\pi_j([0, \delta]^d)) = \delta^{d-1}$ for the last inequality. Of the $(M^{(\delta)})^{d-1}$ terms in the sum over $\pi_j(G)$, there are $(M^{(\delta)} - 1)^{d-1}$ terms such that $\mathbf{v} + \pi_j([0, \delta]^d) \subset \pi_j([0, 1]^d)$. For these terms, by stationarity, $\mathbb{P}(A_2^{(\mathbf{v})}) = \mathbb{P}(A_2^{(\mathbf{0})}) \leq K_1 \delta$ for some $K_1 \in \mathbb{R}^+$ by Lemma 2.3. For the remaining $M^{(\delta)d-1} - (M^{(\delta)} - 1)^{d-1}$ terms in the sum, the bound $\mathbb{P}(A_2^{(\mathbf{v})}) \leq 1$ suffices. Thus,

$$\begin{aligned} \mathbb{E} \left[\left| \delta^{d-1} \sum_{\mathbf{v} \in \pi_j(G)} n_{\mathbf{v}} - \int_{\text{vect}(\mathbf{e}_j^\perp) \cap [0, 1]^d} n_{\mathbf{s}} \, \text{d}\mathbf{s} \right| \right] \\ \leq 2\delta^{d-1} \left((M^{(\delta)} - 1)^{d-1} (K_1 \delta)^{1-\varepsilon} + (M^{(\delta)d-1} - (M^{(\delta)} - 1)^{d-1}) \right) (\mathbb{E}[|n_{\mathbf{0}}|^{\frac{1}{\varepsilon}}])^\varepsilon = O(\delta^{1-\varepsilon}) \end{aligned}$$

and (2.45) holds which completes the proof. \square

2.5.2 Auxiliary results for the proof of the joint central limit theorem

Analogously to (2.37), the following result provides an L^1 control of the approximation error of the first coordinate of I_1 (i.e., the estimated volume) in the proof of Theorem 2.3. To this end, notice that we cannot directly apply an approximation inequality such as Proposition 4 in [Biermé and Desolneux \(2021\)](#) as the function $y \mapsto \mathbb{1}_{\{y \geq u\}}$ appearing in the definition of the volume in (2.7) is not Lipschitz.

Proposition 2.3. *Let X be an isotropic random field satisfying Assumption (A0). Let $C_d^T(u)$ as in (2.7) and $C_d^{(\delta,T)}(u)$ as in (2.30). It holds that*

$$\mathbb{E} \left[\left| C_d^{(\delta,T)}(u) - C_d^T(u) \right| \right] \leq K\delta, \quad (2.47)$$

where $K \in \mathbb{R}^+$ is a constant independent of δ and $\sigma_d(T)$.

Remark 2.7. Proposition 2.3 provides a control of the discretization error in the computation of the volume. This error is of the same order as in Proposition 2.2 (item *ii*). This control can be greatly improved by asking more stringent assumptions: e.g., if X is quasi-associated (see [Bulinski and Shashkin \(2007\)](#) and [Bulinski et al. \(2012\)](#)) and under a decay assumption for its correlation function. In this setting, the bound in (2.47) can be improved to get an upper bound in L^2 with a faster rate in $\delta(\sigma_d(T))^{-1}$ instead of δ . The interested reader is referred to the Appendix 2.A.

Proof of Proposition 2.3. Remark that

$$\begin{aligned} \left| C_d^{(\delta,T)}(u) - C_d^T(u) \right| &= \left| \frac{\delta^d}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta,T)} \left(\mathbb{1}_{\{X(t) \geq u\}} - \frac{1}{\delta^d} \int_{t+[0,\delta]^d} \mathbb{1}_{\{X(s) \geq u\}} \, ds \right) \right| \\ &= \left| \frac{1}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta,T)} \int_{t+[0,\delta]^d} (\mathbb{1}_{\{X(t) \geq u\}} - \mathbb{1}_{\{X(s) \geq u\}}) \, ds \right| \\ &\leq \frac{1}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta,T)} \int_{t+[0,\delta]^d} |\mathbb{1}_{\{X(t) \geq u\}} - \mathbb{1}_{\{X(s) \geq u\}}| \, ds. \end{aligned}$$

Therefore, by Lemma 2.2, it holds that

$$\begin{aligned} \mathbb{E} \left[\left| C_d^{(\delta,T)}(u) - C_d^T(u) \right| \right] &\leq \frac{1}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta,T)} \int_{t+[0,\delta]^d} \mathbb{P}(X(t) < u \leq X(s)) + \mathbb{P}(X(t) \geq u > X(s)) \, ds \\ &\leq \frac{1}{\sigma_d(T)} \sum_{t \in \mathbb{G}(\delta,T)} \int_{t+[0,\delta]^d} 2C_{d-1}^*(u) \|t - s\|_2 \, ds \leq 2\sqrt{d}C_{d-1}^*(u)\delta. \end{aligned}$$

□

Corollary 2.2 below describes the behaviour of the second coordinate of the term I_4 in our main

Theorem 2.3, by using the second order approximation in Theorem 2.1 for the specific hypercubic point-referenced d -honeycomb $\delta\dot{\mathcal{G}}$ in Section 2.3.1.

Corollary 2.2. *Consider a discrete regular hypercubic grid $\mathbb{G}(\delta, T)$ as in (2.29). Let X be an isotropic random field satisfying Assumptions (A0), (A1) and (A2) for some $\varepsilon \in (0, 1)$. Let $C_{d-1}^*(u)$ be as in (2.8) and $C_{d-1}^{(\delta, T_N)}(u)$ as in (2.31). Then, it holds that*

$$\sqrt{\sigma_d(T_N)} \left(\mathbb{E}[C_{d-1}^{(\delta, T_N)}(u)] - \frac{2d}{\beta_d} C_{d-1}^*(u) \right) \rightarrow 0$$

for $N\delta \rightarrow \infty$, $(N\delta)^{d/2} \delta^{1-\varepsilon} \rightarrow 0$ with β_d as in (2.14) and ε as in Assumption (A2).

Proof. It is a direct consequence of (2.31) the stationarity and isotropy of X , Equation (2.15) in Theorem 2.1, and the fact that by construction $\text{Card}(\mathbb{G}(\delta, T)) = (2N)^d$ and $\sigma_d(T_N) = (2N\delta)^d$. \square

Finally, we focus on the term I_2 in the proof of Theorem 2.3. The following theorem establishes a joint central limit theorem for the vector $(C_d^{T_n}(u), \tilde{C}_{d-1}^{T_n}(u))$ in the case where T_n is a sequence of hypercubes in \mathbb{R}^d such that $T_n \nearrow \mathbb{R}^d$ as $n \rightarrow \infty$. Our result is based on techniques used in Iribarren (1989).

Theorem 2.4. *Let u be a fixed level in \mathbb{R} . Assume that X is strongly mixing as in Definition 2.4 such that for some $\eta > 0$, the mixing satisfies the rate condition*

$$\sum_{r=1}^{+\infty} r^{3d-1} \alpha(r)^{\frac{\eta}{2+\eta}} < +\infty,$$

and $\mathbb{E}[\sigma_{d-1}(L_X(u) \cap [0, 1]^d)^{2+\eta}] < +\infty$. Let $C^{T_n}(u) := (C_d^{T_n}(u), \tilde{C}_{d-1}^{T_n}(u))^t$. Let $(T_n)_{n \geq 1}$ be a sequence of hypercubes in \mathbb{R}^d such that $\sigma_d(T_n) = (2n)^d$. Then there exists a finite covariance matrix $\Sigma(u)$ such that, if $\Sigma(u) > 0$,

$$\sqrt{\sigma_d(T_n)} (C^{T_n}(u) - \mathbb{E}[C^{T_n}(u)]) \xrightarrow{\mathcal{L}} \mathcal{N}_2(0, \Sigma(u)),$$

as $n \rightarrow \infty$.

Proof. Let $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$ and $V_{\mathbf{i}}(1)$ be as in Section 2.3.1. Write

$$\xi_{\mathbf{i}} := t_1 \int_{L_X(u) \cap V_{\mathbf{i}}(1)} \frac{\|\nabla X(s)\|_1}{\|\nabla X(s)\|_2} \sigma_{d-1}(ds) + t_2 \sigma_d(E_X(u) \cap V_{\mathbf{i}}(1)),$$

for $\mathbf{i} \in \mathbb{Z}^d$, which makes $\xi := \{\xi_{\mathbf{i}} : \mathbf{i} \in \mathbb{Z}^d\}$ a stationary random field on \mathbb{Z}^d . It is straightforward to see that ξ inherits the mixing property of the random field X and that the mixing coefficients of ξ satisfy $\alpha_{\xi}(n + \lceil \sqrt{d} \rceil) \leq \alpha_X(n)$ for all $n \in \mathbb{N}^+$, since the diameter of $V_{\mathbf{i}}(1)$ is \sqrt{d} . Notice also

that $|\xi_{\mathbf{i}}| \leq \|\mathbf{t}\|_2(\sqrt{d}\sigma_{d-1}(L_X(u) \cap V_{\mathbf{i}}(1)) + \sigma_d(E_X(u) \cap V_{\mathbf{i}}(1))) \leq \|\mathbf{t}\|_2(\sqrt{d}\sigma_{d-1}(L_X(u) \cap V_{\mathbf{i}}(1)) + 1)$ almost surely, so the $2 + \eta$ moment of $\xi_{\mathbf{i}}$ is finite. Finally, since

$$\langle \mathbf{t}, C^{T_n}(u) \rangle = \sum_{\mathbf{i} \in \mathbb{Z}^d} \xi_{\mathbf{i}} \mathbf{1}_{\{V_{\mathbf{i}}(1) \subset T_n\}},$$

the proof is completed with an application of Proposition 1 and Lemma 1 in [Iribarren \(1989\)](#) and the Cramér-Wold device. \square

2.A Supplementary results

2.A.1 Examples

We begin with two simple examples that illustrate the use of the Crofton formula in Equation (2.5).

Example 2.1. Let $d = 2$, Equation (2.5) takes the form:

$$\sigma_1(M) = \frac{1}{4} \int_{-\infty}^{\infty} \int_0^{2\pi} \sigma_0(M \cap l_{\mathbf{s}(\theta), v}) \, d\theta \, dv, \quad (2.48)$$

where $\mathbf{s}(\theta) := (\cos(\theta), \sin(\theta))$. Let M be a circle of radius R in \mathbb{R}^2 . For all $\theta \in [0, 2\pi)$, the function $f_{\theta}(v) := \sigma_0(M \cap l_{\mathbf{s}(\theta), v})$ is equal to 2 on an interval of length $2R$, and 0 elsewhere. Therefore, we easily recover

$$\sigma_1(M) = \frac{1}{4} \int_0^{2\pi} \int_{-\infty}^{\infty} f_{\theta}(v) \, dv \, d\theta = \frac{1}{4} \int_0^{2\pi} 4R \, d\theta = 2\pi R.$$

Example 2.2. Let M be the boundary of a square K with side-length a . As $\partial K = \bigcup_{i=1}^4 \text{side}_i$, using the additivity of (2.48), it suffices to consider only a single side of the square. Without loss of generality, let us suppose that side_1 is horizontal. Then, the function $f_{\theta}(v) := \sigma_0(\text{side}_1 \cap l_{\mathbf{s}(\theta), v})$ is equal to 1 on an interval of length $a|\sin \theta|$, and 0 elsewhere. Thus, we easily recover that

$$\sigma_1(\text{side}_1) = \frac{1}{4} \int_0^{2\pi} \int_{-\infty}^{\infty} f_{\theta}(v) \, dv \, d\theta = \frac{1}{4} \int_0^{2\pi} a|\sin \theta| \, d\theta = \frac{a}{2} \int_0^{\pi} \sin \theta \, d\theta = a.$$

We now present two classical examples where the density $C_{d-1}^*(u)$ in (2.8) can be explicitly obtained. The interested reader is referred to Exercises 6.2.c and 6.3 in [Azaïs and Wschebor \(2009\)](#) and to [Biermé et al. \(2019\)](#).

Example 2.3. Let $X = \{X(t), t \in \mathbb{R}^d\}$, be an isotropic Gaussian field, with zero mean, unit variance and second spectral moment λ finite satisfying Assumption (A0). Then we get

$$C_{d-1}^*(u) = \sqrt{\frac{\lambda}{\pi}} e^{-\frac{u^2}{2}} \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}.$$

Let $X^{(K)} = \{X^{(K)}(t), t \in \mathbb{R}^d\}$ be an isotropic chi-square field satisfying Assumption (A0) with $K \in \mathbb{N}^+$ degree of freedom such that $X^{(K)}(t) = X_1(t)^2 + \dots + X_K(t)^2$ where $X_1(t), \dots, X_K(t)$ are K *i.i.d.* stationary isotropic Gaussian random fields defined as above. Then one can get

$$C_{d-1}^*(u) = \sqrt{\lambda} \left(\frac{u}{2}\right)^{\frac{K-1}{2}} e^{-\frac{u}{2}} \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{K}{2}\right) \Gamma\left(\frac{d}{2}\right)}.$$

2.A.2 Convergence of the bias factor

In this appendix, we provide a short justification for Equation (2.28). The Crofton formula in (2.5) applied to a random manifold M can be written in terms of the conditional expectation,

$$\sigma_{d-1}(M) = \frac{\sqrt{\pi} \Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \int_{\mathbb{R}^{d-1}} \int_{\partial B_1^d} \frac{\mathbb{E}[\sigma_0(M \cap l_{\mathbf{s}, \mathbf{v}_s(\mathbf{u})}) \mid \sigma_{d-1}(M)]}{\sigma_{d-1}(\partial B_1^d)} \, d\mathbf{s} \, d\mathbf{u}, \quad a.s. \quad (2.49)$$

By slightly modifying the proof of Theorem 2.1 so as to leverage (2.49), one obtains

$$\frac{1}{q} \mathbb{P}(X(\mathbf{0}) \leq u < X(q\mathbf{e}_1) \mid C_{d-1}^T(u)) \xrightarrow[q \rightarrow 0]{L^1} \frac{C_{d-1}^T(u)}{\beta_d}.$$

Likewise, by slightly adjusting the proof of Theorem 2.2 accordingly, one writes

$$\mathbb{E}[C_{d-1}^{(\delta \dot{\mathcal{H}}, T)}(u) \mid C_{d-1}^T(u)] \xrightarrow[\delta \rightarrow 0]{L^1} \frac{2d}{\beta_d} C_{d-1}^T(u).$$

Equation (2.28) follows.

2.A.3 Further elements on the dimensional constant β_d

Here, we provide an alternative justification for the value of the bias factor $\frac{2d}{\beta_d}$ in (2.22) for the simple case of the hypercubic lattice by adapting in any dimension d the methodology proposed for instance by Miller (1999).


Remark that the $(d-1)$ -dimensional surface $L_X(u)$ is C^1 , and so it can be approximated arbitrarily well by a union of $(d-1)$ -dimensional hyperplanar surfaces. Then, the expectation of $C_{d-1}^{(\delta \dot{\mathcal{G}}, T)}$ in (2.31) is linear over these hyperplanar surfaces. Thus, it suffices to consider the bias in the estimate of the σ_{d-1} measure of a single hyperplanar surface with orientation vector distributed uniformly on ∂B_1^d . It is not difficult to see that the contribution to $C_{d-1}^{(\delta \dot{\mathcal{G}}, T)}$ of a hyperplanar surface with orientation vector $\mathbf{r} \in \partial B_1^d$ is $\|\mathbf{r}\|_1$ times its σ_{d-1} measure. Thus, the expected bias factor should be the average value of $\|\mathbf{r}\|_1$, when \mathbf{r} is distributed uniformly on ∂B_1^d . Indeed,

$$\frac{1}{\sigma_{d-1}(\partial B_1^d)} \int_{\partial B_1^d} \|\mathbf{r}\|_1 \, \sigma_{d-1}(d\mathbf{r}) = \frac{2d}{\beta_d}. \quad (2.50)$$

Chapter 3

On the perimeter estimation of pixelated excursion sets of two-dimensional anisotropic random fields

This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Opitz, T. (2023). On the perimeter estimation of pixelated excursion sets of two-dimensional anisotropic random fields. *Scandinavian Journal of Statistics*, 1–34. [Paper here](#).

Abstract: We are interested in creating statistical methods to provide informative summaries of random fields through the geometry of their excursion sets. To this end, we introduce an estimator for the length of the perimeter of excursion sets of random fields on \mathbb{R}^2 observed over regular square tilings. The proposed estimator acts on the empirically accessible binary digital images of the excursion regions and computes the length of a piecewise linear approximation of the excursion boundary. The estimator is shown to be consistent as the pixel size decreases, without the need of any normalization constant, and with neither assumption of Gaussianity nor isotropy imposed on the underlying random field. In this general framework, even when the domain grows to cover \mathbb{R}^2 , the estimation error is shown to be of smaller order than the side length of the domain. For affine, strongly mixing random fields, this translates to a multivariate Central Limit Theorem for our estimator when multiple levels are considered simultaneously. Finally, we conduct several numerical studies to investigate statistical properties of the proposed estimator in the finite-sample data setting.

3.1 Introduction

Random fields play a central role in the study of several real-world phenomena. In many applications, the excursion set of a random field (*i.e.*, the subset of the observation domain on which the random field exceeds a certain threshold) is observed—or partially observed—and its geometry can be used to make meaningful inferences about the underlying field. Such techniques have been used in disciplines such as astrophysics (Gott et al., 1990; Ade et al., 2016), brain imaging (Worsley et al., 1992), and environmental sciences (Angulo and Madrid, 2010; Lhotka and Kyselý, 2015; Frölicher et al., 2018). In certain cases, for example in landscape ecology, land-use analysis, and statistical modeling, understanding the geometry of excursions is of primary importance (McGarigal, 1995; Nagendra et al., 2004; Bolin and Lindgren, 2015).

Lipschitz-Killing curvatures (abbreviated LKCs; also known as *intrinsic volumes*) form a rich, well-known class of geometric summaries of stratified manifolds. *Hadwiger’s characterization theorem* states that LKCs form a basis for all rigid motion invariant valuations of convex bodies, which makes them central in the study of the geometry of random sets (Schneider and Weil, 2008). From a theoretical point of view, probabilistic and statistical properties of the LKCs of excursion sets have been widely studied in the last decades (Adler and Taylor, 2007). For Gaussian random fields, the *Euler-Poincaré characteristic* (a well-studied, topological LKC) is studied in Estrade and León (2016) and Di Bernardino et al. (2017); the excursion volume (another LKC, better known as the *sojourn time* for one-dimensional processes) is studied in Bulinski et al. (2012) and Pham (2013). The reader is also referred to Müller (2017) and Kratz and Vadlamani (2018) for a joint analysis of LKCs and to Meschenmoser and Shashkin (2013) and Shashkin (2013) for functional central limit theorems.

LKCs have recently been used to create several statistical procedures including parametric inference (Biermé et al., 2019; Di Bernardino and Duval, 2022) and tests of Gaussianity (Di Bernardino et al., 2017), isotropy (Cabaña, 1987; Fournier, 2018; Berzin, 2021), and symmetry of marginal distributions the underlying fields (Abaach et al., 2021). Di Bernardino et al. (2020) quantifies perturbation via the LKCs and provides a quantitative non-Gaussian limit theorem of the perturbed excursion area behaviour. To further emphasize their importance, LKCs of excursions have deep links to extreme value theory; these insights are summarized in Adler and Taylor (2007) and Azaïs and Wschebor (2009). LKCs can thus provide meaningful and parsimonious summaries of the spatial properties of the studied random fields.

In this manuscript, we focus on the two-dimensional setting—specifically, random fields defined on \mathbb{R}^2 endowed with the standard Euclidean metric. In this case, there are exactly three LKCs that can be leveraged to describe excursion sets of random fields in \mathbb{R}^2 : the excursion volume (*i.e.*, the area), half the value of the perimeter of the excursion set, and the Euler-Poincaré characteristic (which is equal to the number of connected components minus the number of holes of the excursion set).

Analyzed jointly with information on the area and Euler characteristic of an excursion set, the perimeter provides valuable information about the fragmentation of the excursion set. Examples can be found in medical imaging where certain diseases can change fragmentation patterns in biological tissues (Yao et al., 2016; Jurdi et al., 2021), or in ecology where suitable habitats of species are often characterized by exceedances of variables describing favorable conditions, and where edge effects near the boundary the excursion sets play an important role (Debinski and Holt, 2000; Taubert et al., 2018). In spatial risk analysis, the perimeter can give information about the length of the interface between a high-risk zone (associated with exceedances of the threshold level) and moderate-to-low risk zones.

Most of the results presented in the previous literature are based on the empirically inaccessible knowledge of the continuous random field X on a compact domain $T \subset \mathbb{R}^2$. In practice, spatial data are often observed only at sampling locations on a discrete grid $\{s_{i,j} : i, j \in \mathbb{N}_0\} \cap T$, and in such cases, the values of the random field at intermediate points between the sampling locations are not empirically accessible. This regular lattice setting is popular, for example, in the areas of remote sensing, computer vision, biomedical imaging, surface meteorology. The datum at the sampling location $s_{i,j}$ could conceivably be a floating point number representing the value of the random field at $s_{i,j}$, however, it may be the case that this level of precision is not available. One can also consider the more general case where the accessible information at the sampling location $s_{i,j}$ is a boolean value corresponding to whether the random field evaluated at $s_{i,j}$ falls within a predetermined interval—normally $[u, \infty)$ for fixed $u \in \mathbb{R}$. In this general case, one obtains a pixelated representation of the excursion set of X at the fixed level u .

From these sparse-information, binary digital images of excursion sets, we aim in the present work to infer the second Lipschitz-Killing curvature, *i.e.*, the perimeter of the excursion set, for a fixed level u . The perimeter is a particularly difficult quantity to estimate, since, in a digital image, the boundary of an object is comprised of vertical and horizontal pixel edges, which obviously does not correspond to the object’s true boundary. There exists a number of algorithms for computing the perimeter of objects in hard segmented (*i.e.* binary) digital images, many of which are summarized in Coeurjolly and Klette (2002) with further developments made in de Vieilleville et al. (2007). It seems, however, intractable to evaluate the performance of these algorithms on excursion sets of two-dimensional random fields. Bierné and Desolneux (2021) studies how the integrated perimeter of excursion sets over a set of levels changes when considering discretized versions of the underlying *stationary, isotropic* random fields (*i.e.*, those with translation- and rotation-invariant distributions). This gives rise to a perimeter estimator for a single level, complete with its own probabilistic analysis for isotropic random fields (Bierné and Desolneux, 2021). The estimator is further analyzed and given explicit covariance formulas in Abaach et al. (2021) for the case of complete spatial independence. Although this particular perimeter estimator is quite natural to study, it suffers from certain defects; namely, an intrinsic inadequacy for anisotropic random fields.

We introduce a class of estimators for the perimeter of objects in binary digital images, one

of which being particularly suitable for estimating the perimeter of excursion sets of anisotropic random fields on \mathbb{R}^2 . The elements of the class are uniquely associated to the choice of norm that is used to measure a piecewise linear approximation of the excursion’s boundary. The estimator derived from the work of [Biermé and Desolneux \(2021\)](#) arises as the element of the proposed class associated to the 1-norm. The novel estimator associated to the 2-norm (the primary focus of this paper) possesses the desirable property of *multigrid convergence* (*i.e.*, strong consistency as the pixel size tends to zero; see [Theorem 3.1](#)), which we extend to convergence in mean (see [Proposition 3.1](#)). These general results hold under weak assumptions about the smoothness of the random field that do not include Gaussianity, nor isotropy. As the domain grows to cover \mathbb{R}^2 , sufficient conditions are given such that the error in the estimation is of smaller order than the fluctuations of the perimeter—making the limiting distributions of the perimeter and the estimator identical. In particular, by further supposing that the underlying random field is *affine* and *strongly mixing* (notions described in [Section 3.3.2](#)), the estimator associated to the 2-norm is asymptotically normal with the same asymptotic variance as perimeter itself (see [Theorem 3.2](#)).

The organization of the paper is as follows. [Section 3.2](#) specifies key notions including: excursion sets, the hypotheses on the underlying random fields, the regular grid on which the excursion sets are observed, and the novel class of considered perimeter estimators. In [Section 3.3](#), the statistical properties of the perimeter estimate based on the 2-norm are discussed for a fixed domain ([Section 3.3.1](#)) and for a sequence of growing domains ([Section 3.3.2](#)). [Section 3.4](#) provides extensive numerical results to support and illustrate the theory developed in [Section 3.3](#). Proofs and auxiliary notions are postponed to [Section 3.5](#). We conclude with a discussion section. Some supplementary elements are provided in the Appendix Section.

3.2 Definitions and Notation

Let us begin by introducing some notation. Calligraphic font is used to denote sets of isolated points in \mathbb{R}^2 . For a set $S \subset \mathbb{R}^2$, its boundary is denoted $\partial(S)$; its cardinality $\#(S)$; and its Lebesgue measure $\nu(S)$. We use \mathcal{H}^1 to denote the one-dimensional Hausdorff measure, and C^k to denote the space of real-valued functions on \mathbb{R}^2 with k continuous derivatives. Between the nomenclatures *sample paths* and *trajectories*, we choose to use the former when describing the realizations of a random field.

The following assumption ensures that the random objects that we consider are well defined.

Assumption 3.1. *The real-valued random field $X = \{X(s) : s \in \mathbb{R}^2\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has C^2 sample paths.*

Definition 3.1. Denote the *excursion set* of X at the level $u \in \mathbb{R}$ by $E_X(u) := \{s \in \mathbb{R}^2 : X(s) \geq u\}$. For compact $T \subset \mathbb{R}^2$, we denote the restriction of $E_X(u)$ and $\partial(E_X(u))$ to T by

$$E_X(T, u) := T \cap E_X(u) \quad \text{and} \quad E_X^\partial(T, u) := T \cap \partial(E_X(u))$$

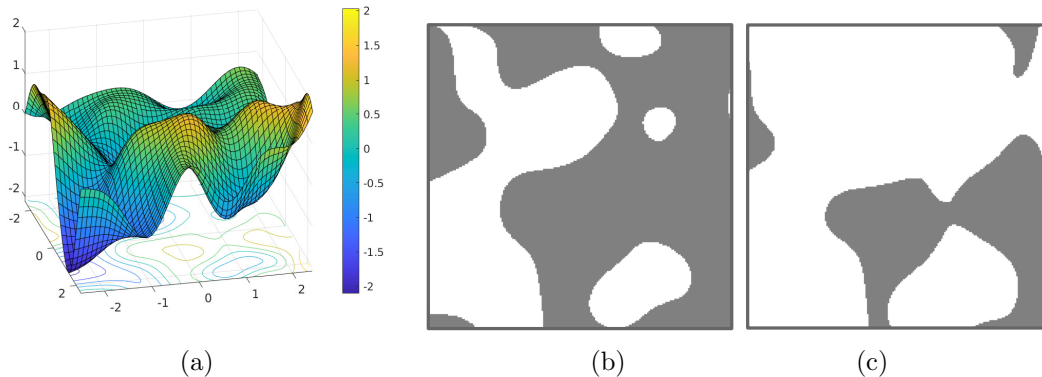


Figure 3.1: Panel (a): a C^2 realization of a stationary, centered, Gaussian random field X with covariance function $r_X(h) = \exp(-\|h\|_2^2)$ is depicted in the square-shaped observation window $T = [-2.5, 2.5]^2$ (generated using the R package `RandomFields` (Schlather et al., 2017)). Underneath the sample path, the curves $E_X^\partial(T, u)$ are drawn for different values of u . Panel (b) (*resp.* panel (c)): the dark region $E_X(T, u)$ is shown for $u = 0$ (*resp.* $u = 0.5$).

respectively. Finally, the quantity of interest in this paper:

$$P_X^T(u) := \mathcal{H}^1(E_X^\partial(T, u)).$$

In Figure 3.1 (a), a C^2 sample path of a Gaussian random field X is depicted in a square domain T with the contours $E_X^\partial(T, u)$ drawn on the domain for various levels u . In Figure 3.1 (b) and (c), $E_X(u)$ is represented by the dark regions, for two different levels u .

In what follows, let

$$T := [-t, t]^2 \subset \mathbb{R}^2, \tag{3.1}$$

for fixed $t > 0$. Before proceeding, it is helpful to specify additional assumptions on the considered random fields.

Assumption 3.2. *Let X_1 and X_2 denote the partial derivatives of X in the two principle Cartesian directions in \mathbb{R}^2 , and let X_{11} and X_{22} denote the corresponding second order partials. For any $u \in \mathbb{R}$, the following three conditions hold almost surely:*

1. X has no critical points in T at the level u .
2. The restriction of X to each face of the square boundary $\partial(T)$ has no local extrema at the level u .
3. For $k \in \{1, 2\}$, there are no $s \in T$ such that $X(s) - u = X_k(s) = X_{kk}(s) = 0$.

Together, Assumptions 3.1 and 3.2 ensure that the random field X is almost surely *suitably regular* at the level u in T as defined in Adler and Taylor (2007, Definition 6.2.1). The third condition of Assumption 3.2 is made to be slightly stronger than item (C) in Definition 6.2.1 of Adler and Taylor

(2007) so that the suitably regular condition holds even after a permutation of the two principal Cartesian directions. This is useful when considering the set

$$\mathcal{Y}_X^T(u) := \bigcup_{k=1,2} \{s \in E_X^\partial(T, u) : X_k(s) = 0\}. \quad (3.2)$$

Indeed, under Assumptions 3.1 and 3.2, it follows directly from Adler and Taylor (2007, Lemma 6.2.3) that

$$\#(\mathcal{Y}_X^T(u)) < \infty, \quad a.s. \quad (3.3)$$

Recall that the *reach* of a set $S \subset \mathbb{R}^d$ is given by

$$\text{reach}(S) := \sup\{\delta \geq 0 : \forall y \in S_\delta \exists! x \in S \text{ nearest to } y\}, \quad (3.4)$$

where $S_\delta = \{y \in \mathbb{R}^d : \exists x \in S \text{ s.t. } \|x - y\|_2 \leq \delta\}$ is the dilation of the set S by a radius $\delta \geq 0$ (see, e.g., Definition 11 in Thäle (2008)). Equations (3.3) and (3.4) will be useful later (see, for example, Remark 3.4).

Recall that a curve $\gamma \subset \mathbb{R}^2$ is connected if it cannot be expressed as the union of two disjoint nonempty closed sets in \mathbb{R}^2 . For sets $B \subseteq A \subseteq \mathbb{R}^2$, B is maximally connected in A if B is connected and there does not exist a connected $C \subseteq A$ such that $B \subset C$.

Definition 3.2. Let $\Gamma_X^T(u)$ be the set of maximally connected subsets of $E_X^\partial(T, u)$.

Assumption 3.3. The random variables $P_X^T(u)$ and $\#(\Gamma_X^T(u))$ are in $L^1(\Omega)$, the space of integrable random variables, for all $u \in \mathbb{R}$.

We emphasize that none of the assumptions stated thus far restrict to stationary or isotropic random fields. Although stationarity is assumed in Theorem 3.2 and Corollary 3.2, these results and all other results are applicable to anisotropic random fields—a crucial point that we investigate numerically in Section 3.4.2.

In what follows, we study a novel estimator of the random quantity $P_X^T(u)$ for arbitrary but fixed $u \in \mathbb{R}$, based only on the random field $Z_X(\cdot; u) = \{Z_X(s; u) : s \in \mathbb{R}^2\}$ defined by

$$Z_X(s; u) := \mathbf{1}_{\{s \in E_X(u)\}} = \mathbf{1}_{\{X(s) \geq u\}}, \quad s \in \mathbb{R}^2.$$

Note that $Z_X(s; u)$ has dependent Bernoulli margins with parameter $\mathbb{P}(X(s) \geq u)$. We will assume that $Z_X(\cdot; u)$ is empirically accessible only at sampling locations on a regular grid, one that is defined in Section 3.2.1 below.

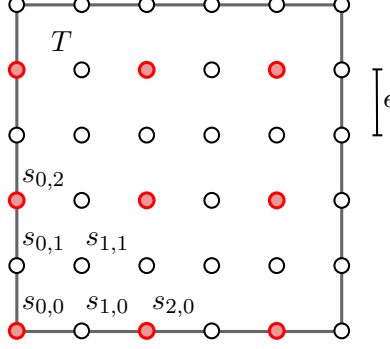


Figure 3.2: An illustration of the quantities defined in Definition 3.3. The positions of the elements of $\mathcal{G}^{(T,\epsilon)}$ in \mathbb{R}^2 are shown as circles, and the subset $\{s_{i,j} : i, j \in I^{(T,\epsilon,m)}\}$ with $m = 2$ is highlighted in red. Here, $M = 6$, and the side length of T is $\sqrt{\nu(T)} = (M - 1)\epsilon = 5\epsilon$.

3.2.1 Sampling locations on a regular grid

Definition 3.3. Fix $\epsilon > 0$, and define a square grid of points in \mathbb{R}^2 as

$$\mathcal{G}^{(T,\epsilon)} := \{s_{i,j} : i, j \in \mathbb{N}_0\} \cap T, \quad \text{with } s_{i,j} := (-t + i\epsilon, -t + j\epsilon) \in \mathbb{R}^2, \quad (3.5)$$

and with T and t as in Equation (3.1). Let M be the number of rows (which is consequentially identical to the number of columns) of $\mathcal{G}^{(T,\epsilon)}$. Define the index set

$$I^{(T,\epsilon)} := \{0, \dots, M - 1\} \subset \mathbb{N}_0$$

and the random matrix $\zeta_X^{(T,\epsilon)}(u)$ with binary elements

$$\zeta_{X,i,j}^{(T,\epsilon)}(u) := Z_X(s_{i,j}; u) = \mathbf{1}_{\{X(s_{i,j}) \geq u\}}, \quad (3.6)$$

for $i, j \in I^{(T,\epsilon)}$. For $m \in \mathbb{N}^+$, let us define

$$I^{(T,\epsilon,m)} := \{i \in I^{(T,\epsilon)} : i \equiv 0 \pmod{m}\}.$$

Notice that $\mathcal{G}^{(T,\epsilon)} = \{s_{i,j} : i, j \in I^{(T,\epsilon)}\}$. We provide an illustration of $\mathcal{G}^{(T,\epsilon)}$ in Figure 3.2, where the elements with indices in $I^{(T,\epsilon,m)}$, with $m = 2$, are highlighted in red. We highlight that our proposed estimator for $P_X^T(u)$ will be based only on the sparse observations $\zeta_{X,i,j}^{(T,\epsilon)}(u)$ for $i, j \in I^{(T,\epsilon)}$ (see Section 3.2.2).

Remark 3.1. The data matrix $\zeta_X^{(T,\epsilon)}(u)$ in (3.6) can be represented as a binary digital image as depicted in Figure 3.3 (b). In this framework, M corresponds to the pixel density or grid size of the image (an integer number of pixels per distance of $2t$, the side length of T), and ϵ corresponds to the pixel width. The quantities are related by $|M\epsilon - 2t| \leq \epsilon$.

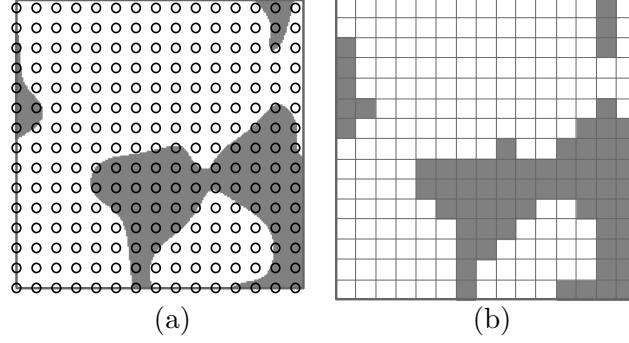


Figure 3.3: Panel (a): $E_X(T, 0.5)$, as shown in Figure 3.1 panel (c), superposed with the elements of the grid $\mathcal{G}^{(T, \epsilon)}$ shown as black circles. Here, $\epsilon \approx 0.32$. Panel (b): the binary matrix $\zeta_X^{(T, \epsilon)}(0.5)$, defined in (3.6), represented as a binary digital image (dark pixels corresponding to 1, and white to 0).

3.2.2 Definition of the estimators

Here, we introduce a class of estimators of $P_X^T(u)$ that use only the information contained in $\zeta_X^{(T, \epsilon)}(u)$, defined in (3.6). Loosely speaking, $\zeta_X^{(T, \epsilon)}(u)$ is separated into submatrices, and in each submatrix the length of the line segment that approximately separates the 1's from the 0's is computed. In this way, the estimator obtained depends on the choice of norm used.

Definition 3.4. With $\|\cdot\|_p$ denoting the p -norm, for $p \in \mathbb{N}^+$, define

$$\widehat{P}_X^{(p)}(\epsilon, m; T, u) := \epsilon \sum_{a \in I^{(T, \epsilon, m)}} \sum_{b \in I^{(T, \epsilon, m)}} \left\| (N_{X, h}(a, b; u), N_{X, v}(a, b; u)) \right\|_p, \quad (3.7)$$

where

$$N_{X, h}(a, b; u) := \sum_{i=a}^{(a+m-1) \wedge (M-1)} \sum_{j=b}^{(b+m-1) \wedge (M-2)} |\zeta_{X, i, j}^{(T, \epsilon)}(u) - \zeta_{X, i, j+1}^{(T, \epsilon)}(u)|, \quad a, b \in I^{(T, \epsilon, m)},$$

and

$$N_{X, v}(a, b; u) := \sum_{i=a}^{(a+m-1) \wedge (M-2)} \sum_{j=b}^{(b+m-1) \wedge (M-1)} |\zeta_{X, i, j}^{(T, \epsilon)}(u) - \zeta_{X, i+1, j}^{(T, \epsilon)}(u)|, \quad a, b \in I^{(T, \epsilon, m)}.$$

Continuing from the framework discussed in Remark 3.1, $N_{X, v}$ (*resp.* $N_{X, h}$) counts the number of pixels in a subrectangle—of size at most $m \times m$ pixels—of T that differ in shade from the neighbouring pixel to the right (*resp.* above). In other words, $N_{X, v}$ (*resp.* $N_{X, h}$) provides a count of significant vertical (*resp.* horizontal) pixel edges in the subrectangle.

By considering the estimator in (3.7) with norm $p = 1$, one recovers the estimator that is extensively studied in Biermé and Desolneux (2021) and Abaach et al. (2021). It counts the number of pixel edges that separate pixels of different color, and rescales the count by ϵ . Thus, $\widehat{P}_X^{(1)}(\epsilon, m; T, u)$

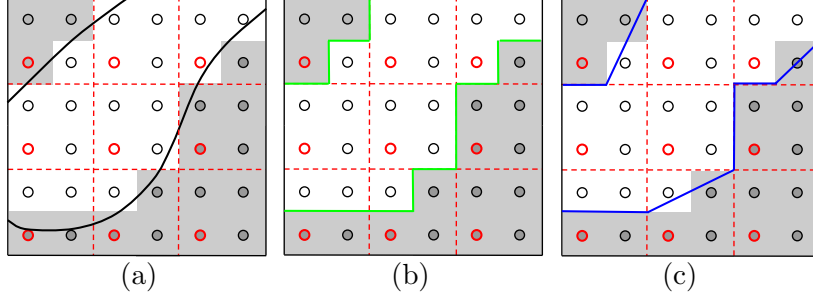


Figure 3.4: Panel (a): the curve $E_X^\partial(T, u)$ is shown in relation to the points in $\mathcal{G}^{(T, \epsilon)}$ in (3.5). Points in the dark regions are assigned a value of 1 in the matrix $\zeta_X^{(T, \epsilon)}(u)$, and points in white are assigned a value of 0. The points outlined in red have indices in $I^{(T, \epsilon, m)}$ with $m = 2$. In effect, $\widehat{P}_X^{(1)}(\epsilon; T, u)$ is calculated by counting the pixel edges shown in green (see panel (b)), whereas $\widehat{P}_X^{(2)}(\epsilon, 2; T, u)$ is calculated by summing the lengths of the blue piecewise linear curves (see panel (c)).

will not depend on m , so we write $\widehat{P}_X^{(1)}(\epsilon; T, u)$ in place of $\widehat{P}_X^{(1)}(\epsilon, m; T, u)$.

Figure 3.4 illustrates the behavior of the estimator in Equation (3.7) constructed with two different norms; the norms associated to $p = 1$ and $p = 2$. In addition, Table 3.1 provides the corresponding terms in Equation (3.7) for this example, for each $a, b \in I^{(T, \epsilon, 2)} = \{0, 2, 4\}$, for both $p = 1$ (second-last column) and $p = 2$ (last column).

The estimator in (3.7) with norm $p = 2$ approximates the length of $E_X^\partial(T, u)$ by the total length of a set of line segments that approximate the curve (see Figure 3.4 (c)). The number of possible orientations of each line segment grows with m ; so does the length of each line segment, which, loosely speaking, is on the order of $m\epsilon$. Therefore, it is not surprising that $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$ depends on m , and our statistical analysis in Section 3.3 therefore takes place in the regime where m is large and $m\epsilon$ is small. In Section 3.4.4, we provide an adaptive method to select the hyperparameter m when ϵ is given as a feature of the data.

3.3 Main Results

The focus of this section is to prove convergence results for the estimator $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$. The statistical analysis is separated into two regimes. In Section 3.3.1, we consider the domain T to be fixed and decrease the pixel width while sending m to infinity. Section 3.3.2 studies the behaviour of the estimator on a sequence of growing domains. In particular, in Section 3.3.2, we study the asymptotic relationships between ϵ , m , and the Lebesgue measure of the sequence of domains, and provide sufficient conditions for good convergence properties. We conclude with a multivariate Central Limit Theorem in the case where multiple levels (u_1, \dots, u_k) are considered simultaneously under the assumption that the underlying random field X is affine and strongly mixing (see Section 3.3.2 for the theorem and the notions of affinity and strongly mixing).

Table 3.1: $\widehat{P}_X^{(p)}(\epsilon, m; T, u)$ in (3.7) computed for the discretized excursion set in Figure 3.4. The last two columns correspond to the terms $(N_{X,v}(a, b; u)^p + N_{X,h}(a, b; u)^p)^{1/p}$ for $p = 1$ and $p = 2$. Summing each term yields $\widehat{P}_X^{(p)}(\epsilon, m; T, u)$, as shown in bold in the final row.

$a \in I^{(T, \epsilon, m)}$ (column)	$b \in I^{(T, \epsilon, m)}$ (row)	$N_{X,h}(a, b; u)$	$N_{X,v}(a, b; u)$	$p = 1$	$p = 2$
0	0	2	0	2	2
0	2	1	0	1	1
0	4	1	2	3	$\sqrt{5}$
2	0	2	1	3	$\sqrt{5}$
2	2	0	2	2	2
2	4	0	0	0	0
4	0	0	0	0	0
4	2	1	0	1	1
4	4	1	1	2	$\sqrt{2}$
$\widehat{P}_X^{(p)}(\epsilon, m; T, u) :$				14ϵ	11.89ϵ

3.3.1 On a fixed domain with decreasing pixel width

Here, we are interested in the behaviour of the estimator $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$ in the case where the domain $T = [-t, t]^2$ is fixed, and the spacing between the locations of the observations in the matrix $\zeta_X^{(T, \epsilon)}(u)$ tends to 0. We proceed to show that the resulting perimeter estimate converges almost surely to $P_X^T(u)$ and give the rate of convergence.

Theorem 3.1. *Let $(m_n)_{n \geq 1}$ be a non-decreasing sequence in \mathbb{N}^+ tending to ∞ as $n \rightarrow \infty$. Let $(\epsilon_n)_{n \geq 1}$ be a sequence in \mathbb{R}^+ such that $m_n \epsilon_n^{2/3}$ converges to a constant $C \in \mathbb{R}^+$ and that the vertices of T are contained in $\mathcal{G}^{(T, \epsilon_n)}$ for all $n \in \mathbb{N}^+$. Then, under Assumptions 3.1 and 3.2, for fixed $u \in \mathbb{R}$, it holds that*

$$g_n |\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u) - P_X^T(u)| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty,$$

where $(g_n)_{n \geq 1}$ is any non-decreasing sequence such that $g_n = o(m_n)$.

The proof of Theorem 3.1 is postponed to Section 3.5.

Remark 3.2. Theorem 3.1 is a statement about the *multigrid convergence* (see, for instance, Definition 2 of Coeurjolly and Klette (2002)) of $\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u)$ to $P_X^T(u)$ as $n \rightarrow \infty$ for almost all sample paths of the random field X . The *speed* of this convergence is $O(1/m_n)$.

Theorem 3.1 requires that the vertices of T are in $\mathcal{G}^{(T, \epsilon_n)}$ for all $n \in \mathbb{N}^+$, for example as depicted in Figure 3.2. This prevents the possibility of there being long segments of $E_X^\partial(T, u)$ that remain close to the border of T so as to not pass between elements of $\mathcal{G}^{(T, \epsilon_n)}$. In addition, it is supposed that the

sequence $(m_n)_{n \geq 1}$ is asymptotically equivalent to $(\epsilon_n^{-2/3})_{n \geq 1}$, which gives the fastest possible rate of convergence of $\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u)$ to $P_X^T(u)$. By relaxing this condition, we obtain the following corollary.

Corollary 3.1. *Under the conditions of Theorem 3.1, if the requirement that $m_n \epsilon_n^{2/3} \rightarrow C$ is relaxed to $m_n \epsilon_n \rightarrow 0$, it holds that*

$$\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u) \xrightarrow{\text{a.s.}} P_X^T(u), \quad n \rightarrow \infty.$$

The proof is postponed to Section 3.5. The following proposition shows that convergence in $L^1(\Omega)$ holds under slightly stronger assumptions. The proof can also be found in Section 3.5.

Proposition 3.1. *Let $(m_n)_{n \geq 1}$ be a non-decreasing sequence in \mathbb{N}^+ tending to ∞ as $n \rightarrow \infty$. Let $(\epsilon_n)_{n \geq 1}$ be a sequence in \mathbb{R}^+ such that $m_n \epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, and that the vertices of T are contained in $\mathcal{G}^{(T, \epsilon_n)}$ for all $n \in \mathbb{N}^+$. Then under Assumptions 3.1, 3.2, and 3.3,*

$$|\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u) - P_X^T(u)| \xrightarrow{L^1} 0, \quad n \rightarrow \infty,$$

for any fixed $u \in \mathbb{R}$.

Remark 3.3. It is shown in Proposition 5 of [Biermé and Desolneux \(2021\)](#) that for a random field X satisfying Assumption 3.1, if, in addition, X is stationary, Gaussian, isotropic, and the supremum of the first and second order partial derivatives of X in the domain T are in $L^1(\Omega)$, then

$$\mathbb{E}[\widehat{P}_X^{(1)}(\epsilon; T, u)] \rightarrow \frac{4}{\pi} \mathbb{E}[P_X^T(u)], \quad (3.8)$$

as $\epsilon \rightarrow 0$. Proposition 3.1 is a stronger result under weaker assumptions on X . With neither Gaussianity, stationarity, nor isotropy imposed on X , it holds that

$$\mathbb{E}[\widehat{P}_X^{(2)}(\epsilon, m; T, u)] \rightarrow \mathbb{E}[P_X^T(u)],$$

as $\epsilon \rightarrow 0$ and $m \rightarrow \infty$ under the constraint $m\epsilon \rightarrow 0$. Thus, the estimator $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$ does not suffer from the asymptotic bias factor of $4/\pi$.

3.3.2 On a growing domain with decreasing pixel width

In this section, the performance of $\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, u)$ is investigated for sequences $(\epsilon_n)_{n \geq 1}$, $(m_n)_{n \geq 1}$, and $(T_n)_{n \geq 1}$ satisfying $\epsilon_n \rightarrow 0$, $m_n \rightarrow \infty$, and $T_n \nearrow \mathbb{R}^2$ as $n \rightarrow \infty$. To manage the added complexity of the sequence of growing domains, first define

$$T_n := \{ns : s \in T\},$$

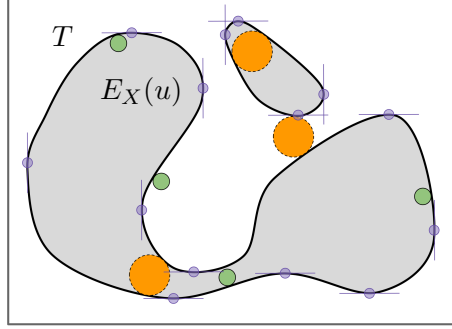


Figure 3.5: Illustration of the notions of reach and resolution in Definition 3.5. The reach of $E_X(T, u)$ is greater than the radius, r_{green} , of the small green circles with solid border. The reach of $T \setminus E_X(u)$ is also greater than r_{green} . Moreover, the minimum distance between points in $\mathcal{Y}_X^T(u)$, highlighted in purple, exceeds $2r_{\text{green}}$. Therefore, $E_X(u)$ is *resolved* by r_{green} in T (see Definition 3.5). Conversely, it is clear that $E_X(u)$ is not resolved in T by the radius of the larger orange circles with dashed border.

such that T_n is a dilation of the fixed domain $T = [-t, t]^2$. The side length of the square domain T_n is then $2tn$. The challenge then becomes determining sufficient asymptotic relations for the sequences $(\epsilon_n)_{n \geq 1}$ and $(m_n)_{n \geq 1}$ to ensure desirable statistical properties of our estimator.

Asymptotics for the pixel width

We relate the domain size with an appropriate pixel width by defining *resolution* in the context of excursion sets of random fields, inspired by the notion of optical resolution.

Definition 3.5. Define the random variable

$$\Lambda_X^T(u) := \min \left\{ \text{reach}(E_X(T, u)), \text{reach}(T \setminus E_X(u)), \text{reach}(\mathcal{Y}_X^T(u)) \right\}.$$

For $\lambda \in \mathbb{R}^+$, we say that “ $E_X(u)$ is resolved by λ in T ” whenever the random event $\{\lambda < \Lambda_X^T(u)\}$ occurs.

This makes $\Lambda_X^T(u)$ a random geometrical description of $E_X(u)$ in the domain T : $\Lambda_X^T(u)$ is the supremum of the set of $\lambda \in \mathbb{R}^+$ such that one can roll a ball of radius λ along both sides of the curve $E_X^\partial(T, u)$, and that the distances between points in $\mathcal{Y}_X^T(u)$ are all at least 2λ . Figure 3.5 clarifies some of the notions introduced in Definition 3.5. This definition allows us to relate the domain size with the pixel width, since the estimation error can be bounded in the case where $E_X(u)$ is resolved by $m_n \epsilon_n$ in T_n (see the proof of Theorem 3.1).

Remark 3.4. Under Assumptions 3.1 and 3.2, the random sets $E_X(T, u)$ and $T \setminus E_X(u)$ have positive reach almost surely, since $E_X(u)$ and $E_{-X}(u)$ have a twice differentiable boundary everywhere in T , almost surely, for all $u \in \mathbb{R}$. The intersection of these sets with the compact rectangle T guarantees that the reach of each intersection is positive (Biermé et al., 2019, p. 541). The

minimum distance between points in $\mathcal{Y}_X^T(u)$ is positive by Equation (3.3) and the compactness of T . Therefore, $\Lambda_X^T(u)$ in Definition 3.5 is almost surely positive for all $u \in \mathbb{R}$. Equivalently, for any $u \in \mathbb{R}$,

$$\mathbb{P}\left(\liminf_{\lambda \rightarrow 0} \{\lambda < \Lambda_X^T(u)\}\right) = 1,$$

i.e., with probability 1, there exists a sufficiently small positive λ that resolves $E_X(u)$ in T .

With the notion of resolution established, we state an important convergence result for the sequence of growing domains $(T_n)_{n \geq 1}$ under general regularity assumptions.

Proposition 3.2. *Let X be a random field satisfying Assumptions 3.1, 3.2, and 3.3. Let $(m_n)_{n \geq 1}$ be a non-decreasing sequence in \mathbb{N}^+ such that $m_n/n \rightarrow \infty$. Let $(\epsilon_n)_{n \geq 1}$ be a non-increasing sequence in \mathbb{R}^+ satisfying $\epsilon_n = O(m_n^{-3/2})$. Moreover, suppose that $2t$ is an integer multiple of ϵ_n for all $n \in \mathbb{N}^+$, and $\mathbb{P}(m_n \epsilon_n < \Lambda_X^{T_n}(u)) \rightarrow 1$ as $n \rightarrow \infty$. Then for any $u \in \mathbb{R}$,*

$$\frac{\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, u) - P_X^{T_n}(u)}{\sqrt{\nu(T_n)}} \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$.

The proof of Proposition 3.2 is postponed to Section 3.5.

Remark 3.5. One example of a sequence $(\epsilon_n)_{n \geq 1}$ satisfying the constraints in Proposition 3.2 is constructed by letting ϵ_n be the largest element in the sequence $(2t/k)_{k \geq 1}$ such that $\epsilon_n \leq m_n^{-3/2}$ and $\mathbb{P}(\Lambda_X^{T_n}(u) \leq m_n \epsilon_n) \leq 1/n$, where $\Lambda_X^{T_n}(u)$ is defined in Definition 3.5. Such a sequence $(\epsilon_n)_{n \geq 1}$ exists since $\mathbb{P}(\Lambda_X^{T_n}(u) \leq 0) = 0$ for all $n \in \mathbb{N}^+$ as discussed in Remark 3.4. The idea is to have the sequence $\lambda_n := m_n \epsilon_n$ tend to 0 faster than the quantiles of $\Lambda_X^{T_n}(u)$, which is difficult to verify analytically. However, in practice, for a given realization of $E_X(u)$, one can estimate $\Lambda_X^T(u)$ by first estimating the reach of the sets $E_X(T, u)$ and $T \setminus E_X(u)$ (Aamari et al., 2019; Cotsakis, 2023) and the vector coordinates of the points in $\mathcal{Y}_X^T(u)$, defined in (3.2).

Proposition 3.2 establishes that for a large class of random fields, as the domain grows and the grid spacing decreases, the error in the perimeter estimation is negligible compared to the side length of the domain. Such a comparison is made possible by the conditions on the sequences $(m_n)_{n \geq 1}$ and $(\epsilon_n)_{n \geq 1}$, since the indexing variable n is proportional to the side length of T_n .

Asymptotic normality of the perimeter estimator

In this section, we prove a multivariate Central Limit Theorem for our estimator as stated in Theorem 3.2 below, based on the results from Iribarren (1989). The interested reader is also referred to Cabaña (1987).

First, we recall two important notions regarding the random fields for which the theorem applies. Recall that a random field $X = \{X(s) : s \in \mathbb{R}^2\}$ is said to be *affine* if it is equal in distribution

to $\{Y(As) : s \in \mathbb{R}^2\}$, where Y is stationary, isotropic, and A is a positive-definite 2×2 matrix. Consequentially, the resulting X is stationary but may be anisotropic. Note that it is common in geostatistics literature to use the nomenclature *geometric anisotropy* when referring to affine random fields (Chiles and Delfiner, 2009).

In the case of X affine, a useful expression for $\mathbb{E}[P_X^T(u)]$, when it exists, is provided in Cabaña (1987, Section 1.1); that is,

$$\mathbb{E}[P_X^T(u)] = \frac{\text{ellipse}(\lambda_1, \lambda_2)}{2\pi} \mathbb{E}[P_Y^T(u)], \quad (3.9)$$

with λ_1 and λ_2 denoting the eigenvalues of A , and $\text{ellipse}(a, b)$ denoting the perimeter of an ellipse with semi-minor and semi-major axes a and b .

Recall that X is said to be *strongly mixing*, or *uniformly mixing*, if there exists a function $\psi(\rho) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ tending to 0 as $\rho \rightarrow \infty$, such that for any two measurable sets $S_1, S_2 \subset \mathbb{R}^2$ that satisfy $\inf\{\|s_1 - s_2\|_2 : s_1 \in S_1, s_2 \in S_2\} =: \rho > 0$, and for any events A_1 and A_2 in the the sigma fields generated by $\{X(s) : s \in S_1\}$ and $\{X(s) : s \in S_2\}$ respectively, it holds that $|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| < \psi(\rho)$.

Under the assumption that the underlying random field is affine and strongly mixing, we prove the multivariate central limit theorem for our estimator. The proof of Theorem 3.2 is postponed to Section 3.5.

Theorem 3.2. *Let X be a stationary, affine, strongly mixing random field satisfying Assumptions 3.1–3.3. With ∇X denoting the gradient of X , suppose that the joint density function of $(X, \nabla X)$ is bounded. Let $k \in \mathbb{N}^+$ and fix the vector $\mathbf{u} := (u_1, \dots, u_k) \in \mathbb{R}^k$ such that $u_i \neq u_j$ for $1 \leq i < j \leq k$. Let the sequences $(m_n)_{n \geq 1}$ and $(\epsilon_n)_{n \geq 1}$ satisfy the constraints in Proposition 3.2 for all u_j , with $j = 1, \dots, k$. Let*

$$\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, \mathbf{u}) := (\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, u_1), \dots, \widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, u_k))$$

and

$$P_X^{T_n}(\mathbf{u}) := (P_X^{T_n}(u_1), \dots, P_X^{T_n}(u_k)).$$

Then there exists a finite, non-degenerate (i.e., full-rank) covariance matrix $\Sigma(\mathbf{u})$ such that

$$\frac{\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, \mathbf{u}) - \mathbb{E}[P_X^{T_n}(\mathbf{u})]}{\sqrt{\nu(T_n)}} \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \Sigma(\mathbf{u})), \quad n \rightarrow \infty, \quad (3.10)$$

with $\mathbb{E}[P_X^{T_n}(u_j)]$ as in (3.9) for all u_j , $j = 1, \dots, k$. The elements of $\Sigma(\mathbf{u})$ are of the form

$$\Sigma_{ij}(\mathbf{u}) = \int_{\mathbb{R}^2} H_s(u_i, u_j) \, ds, \quad (3.11)$$

where

$$H_s(u_i, u_j) = g_s(u_i, u_j) \mathbb{E} \left[\|\nabla X(0)\|_2 \|\nabla X(s)\|_2 \mid X(0) = u_i, X(s) = u_j \right] \\ - f(u_i) f(u_j) \mathbb{E} \left[\|\nabla X(0)\|_2 \mid X(0) = u_i \right] \mathbb{E} \left[\|\nabla X(s)\|_2 \mid X(s) = u_j \right],$$

with f denoting the marginal density function of X , and g_s , the joint density function of $(X(0), X(s))$.

As seen in the proof of Theorem 3.2, the rescaled limiting Gaussian distribution of our perimeter estimator—in our pixelated framework—coincides with that of $P_X^{T_n}(u)$, the true perimeter in the continuous framework.

Corollary 3.2, stated below, provides a succinct set of conditions on X that imply the result of Theorem 3.2. In particular, the additional assumption of Gaussianity of the underlying random fields is introduced.

Corollary 3.2. *Suppose that there exists a positive-definite matrix A such that the random field X is equal in distribution to $\{Y(As) : s \in \mathbb{R}^2\}$, for some C^2 , stationary, isotropic, centered, Gaussian random field Y with covariance function $r(h)$, $h \in \mathbb{R}^2$. Define*

$$\Psi(s) = \max \left\{ |r(s)|, |r_1(s)|, |r_2(s)|, |r_{11}(s)|, |r_{22}(s)|, |r_{12}(s)| \right\},$$

for $s \in \mathbb{R}^2$, where $r_i := \partial r / \partial s_i$ and $r_{ij} := \partial^2 r / (\partial s_i \partial s_j)$ for $i, j \in \{1, 2\}$. Suppose further that $\Psi(s) \rightarrow 0$ as $\|s\|_2 \rightarrow \infty$, $\int_{\mathbb{R}^2} |\Psi(s)| \, ds < \infty$, and $\int_{\mathbb{R}^2} r(s) \, ds > 0$. Then the result of Theorem 3.2 holds.

The proof can be found in Section 3.5. We remark that a vast literature exists on the asymptotic distribution of level functionals of Gaussian random fields (Wschebor, 1985; Meschenmoser and Shashkin, 2013; Shashkin, 2013; Di Bernardino et al., 2017; Beliaev et al., 2020; Di Bernardino and Duval, 2022), in which case, the asymptotic variance-covariance matrix in (3.11) can be written by projecting the Gaussian functionals of interest onto the Itô-Wiener chaos (the interested reader is referred, for instance, to Kratz and León, 2001; Estrade and León, 2016; Müller, 2017; Kratz and Vadlamani, 2018; Berzin, 2021).

3.4 Simulation studies

In this section, we illustrate finite sample performances of our estimator $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ on simulated data. More precisely, we wish to showcase the results of Proposition 3.1 and Theorem 3.2. Furthermore, we aim to compare the estimators constructed from the norms $p = 1$ and $p = 2$ in (3.7). Our simulation studies are implemented both for anisotropic (see Section 3.4.2) and isotropic (see Section 3.4.3) random fields. In addition, we provide an adaptive method for choosing the hyperparameter m for the estimator $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$ (see Section 3.4.4). The random fields used in each simulation are elements of the class in Example 3.1 below.

Example 3.1. Let Y be a stationary, isotropic, centered, Gaussian random field with a Matérn covariance function

$$r(h) := \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu} \|h\|_2)^\nu K_\nu(\sqrt{2\nu} \|h\|_2), \quad h \in \mathbb{R}^2,$$

where K_ν is the modified Bessel function of the second kind and $\nu = 2.5$. To clarify, the range parameter in the covariance function is fixed as 1.

Let $\{X(s; \sigma_1, \sigma_2, \theta) : s \in \mathbb{R}^2\}$ be a random field equal in distribution to $\{Y(As) : s \in \mathbb{R}^2\}$, where

$$A := \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \quad (3.12)$$

$\sigma_1, \sigma_2 \in \mathbb{R}^+$, $\sigma_1 \geq \sigma_2$, and $\theta \in [0, \pi)$. In this way, $X(\cdot; \sigma_1, \sigma_2, \theta)$ is affine with affinity parameters $k = (1 - \sigma_2^2/\sigma_1^2)^{1/2}$ and θ (Cabaña, 1987). Notice that $X(\cdot; \sigma_1, \sigma_2, \theta)$ is also Gaussian with covariance function given by $r_X(h) = r(Ah)$. Although A is not necessarily positive-definite, there exists a unique positive-definite matrix B with eigenvalues σ_1 and σ_2 such that $\|Ah\|_2 = \|Bh\|_2$ for all $h \in \mathbb{R}^2$. Note also that $\sigma_1 = \sigma_2$ if and only if X is isotropic, in which case, X does not depend on θ .

Throughout Section 3.4, $X(\cdot; \sigma_1, \sigma_2, \theta)$ and Y denote the random fields in Example 3.1. The former is sometimes abbreviated as X , and the dependence on σ_1 , σ_2 , and θ should be understood implicitly. The results in this section can be reproduced using the code made available at <https://github.com/RyanCotsakis/excursion-sets>.

3.4.1 A proxy for the true perimeter

In what follows, the R package `RandomFields` is used to generate realizations of random fields on regular grids. However, when simulating the random field $X(\cdot; \sigma_1, \sigma_2, \theta)$ in this way, it is impossible to infer the exact value of $P_X^T(u)$ for any level $u \in \mathbb{R}$ due to the discretization of the domain T . To overcome this issue, a proxy is used for the true perimeter. In Appendix B of Biermé and Desolneux (2021), the authors introduce an estimator that they show to be multigrid convergent for $P_X^T(u)$, for any $u \in \mathbb{R}$. Moreover, the estimator takes as its arguments the values of X , a random field with C^2 sample paths, evaluated on a regular grid, *i.e.*, $X(s_{i,j})$ for $i, j \in I^{(T, \epsilon)}$ —precisely the output of the simulation from the `RandomFields` package. For a pixel width of ϵ , denote this estimator by $\tilde{P}_X(\epsilon; T, u)$. Notice that $\tilde{P}_X(\epsilon; T, u)$ requires more information than $\hat{P}_X^{(2)}(\epsilon, m; T, u)$. While \tilde{P}_X has access to the value of X evaluated on the regular square tiling $\mathcal{G}^{(T, \epsilon)}$, defined in (3.5), $\hat{P}_X^{(2)}$ only has access to the binary black-and-white matrix $\zeta_X^{(T, \epsilon)}(u)$, defined in (3.6).

Convergence of $\tilde{P}_X(\epsilon_n; T, u)$ to $P_X^T(u)$ in $L^1(\Omega)$ follows from the same arguments that we use in the proof of our Proposition 3.1. Therefore, for any sequence $(h_n)_{n \geq 1}$,

$$|h_n - \tilde{P}_X(\epsilon_n; T, u)| \xrightarrow{L^1} 0 \iff h_n \xrightarrow{L^1} P_X^T(u) \quad (3.13)$$

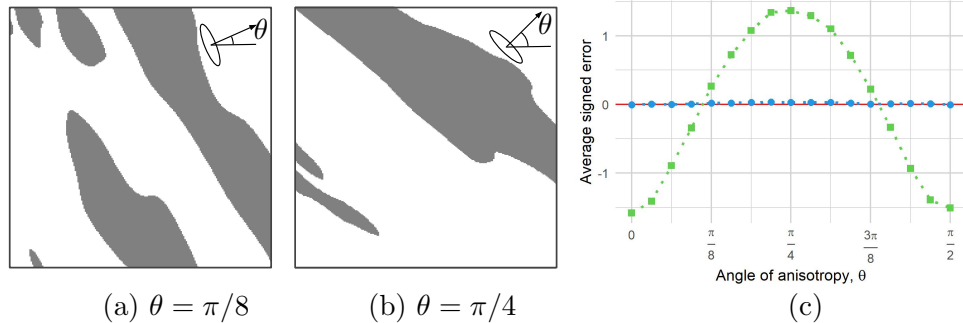


Figure 3.6: Illustration of the effect of anisotropy on the perimeter length estimation. The anisotropic random field $X(\cdot; 2, 0.5, \theta)$ is described in Example 3.1. Here, $T = [-2.5, 2.5]^2$ and $\epsilon = 5/255$. Panels (a, b): a realization of $E_X(T, 0.5)$ shown as the dark region for the corresponding value of θ . The matrix A , defined in (3.12), maps the drawn ellipse to a circle. Panel (c): for several $\theta \in [0, \pi/2]$, 200 independent realizations of X are simulated, and the mean values of $(\pi/4)\widehat{P}_X^{(1)}(\epsilon; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ (green squares) and $\widehat{P}_X^{(2)}(\epsilon, 11; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ (blue circles) are plotted.

as $n \rightarrow \infty$.

3.4.2 The anisotropic case

None of the assumptions established thus far prohibit anisotropy. In fact, all of the results developed in Section 3.3 are applicable to all of the random fields parameterized as in Example 3.1. In Sections 3.4.2, 3.4.2, and 3.4.2, we consider such random fields that are anisotropic (*i.e.*, parameterized by $\sigma_1 \neq \sigma_2$). To avoid confusion, we consistently choose $(\sigma_1, \sigma_2) = (2, 0.5)$.

Mean perimeter estimate as a function of the angle θ

The random fields X in Example 3.1 parameterized by $(\sigma_1, \sigma_2) = (2, 0.5)$ and several $\theta \in [0, \pi/2]$ are simulated in the domain $T = [-2.5, 2.5]^2$, discretized into 256×256 pixels. With ϵ denoting the resulting pixel width, the performances of the estimators $(\pi/4)\widehat{P}_X^{(1)}(\epsilon; T, u)$ and $\widehat{P}_X^{(2)}(\epsilon, m; T, u)$ with $m = 11$ are compared at the level $u = 0.5$. For each of the several values of θ chosen in $[0, \pi/2]$, 200 independent replications of $X(\cdot; 2, 0.5, \theta)$ are simulated in the domain T and the mean error in the estimates of $P_X^T(0.5)$ is plotted for each of the two estimators: the sample means of $(\pi/4)\widehat{P}_X^{(1)}(\epsilon; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ (shown in green) and $\widehat{P}_X^{(2)}(\epsilon, 11; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ (shown in blue), and $\widetilde{P}_X(\epsilon; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ (shown in black) in Figure 3.6 (c). Notice that $\mathbb{E}[(\pi/4)\widehat{P}_X^{(1)}(\epsilon; T, 0.5)]$ depends on θ , since $\mathbb{E}[P_X^T(0.5)] = 19.4$ for all θ . The latter expectation is computed via Equation (3.9) and the Gaussian Kinematic Formula in Adler and Taylor (2007, Theorem 15.9.5). The sample average of $\widehat{P}_X^{(2)}(\epsilon, 11; T, 0.5) - \widetilde{P}_X(\epsilon; T, 0.5)$ shown in Figure 3.6 is nearly 0 for all θ , thus supporting our claim that that our estimator adapts to anisotropic random fields.

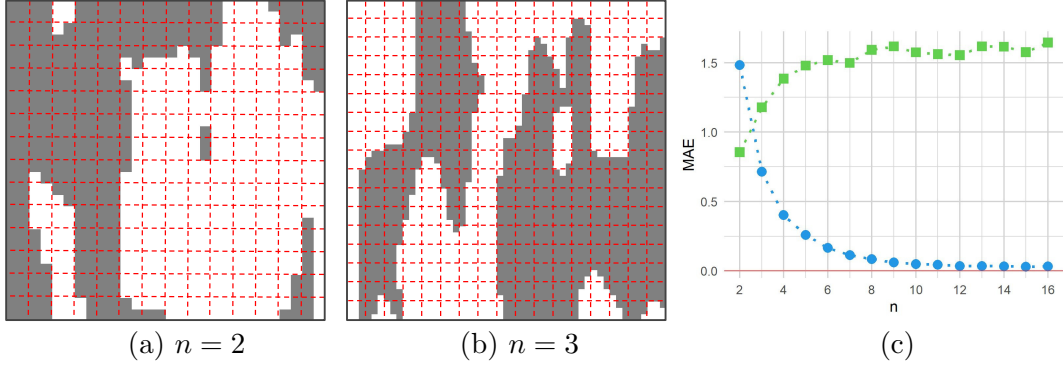


Figure 3.7: The case of decreasing pixel width with the domain $T = [-2.5, 2.5]^2$ fixed. Here, $u = 0.5$; M_n , m_n , and ϵ_n are given in (3.14); and X in Example 3.1, parametrized by $(\sigma_1, \sigma_2, \theta) = (2, 0.5, 0)$, is anisotropic. Panel (a): the excursion set (shown as the dark region) is generated using $M_2 \times M_2$ pixels, and the dashed red lines have a spacing of $2\epsilon_2$, where ϵ_2 is the pixel width. Panel (b): the size of the image (measured in pixels) is $M_3 \times M_3$, and the dashed red lines have a spacing of $3\epsilon_3$, where ϵ_3 is the pixel width. Panel (c): the approximation of $\tilde{P}_X(\epsilon_n; T, 0.5)$ by $(\pi/4)\hat{P}_X^{(1)}(\epsilon_n; T, 0.5)$ (green squares) and by $\hat{P}_X^{(2)}(\epsilon_n, m_n; T, 0.5)$ (blue circles) is shown for different values of n . For each n , the MAE of the approximations are calculated from 500 independent replications of the process X .

Convergence in mean in the anisotropic case

Let $\lfloor \cdot \rfloor$ denote the floor function. For $n \in \mathbb{N}^+$, fix the domain $T = [-2.5, 2.5]^2$ and let

$$M_n = \lfloor 10n^{3/2} \rfloor, \quad m_n = n, \quad \text{and} \quad \epsilon_n = 5/(M_n - 1), \quad (3.14)$$

so that the constraints in Theorem 3.1 and Proposition 3.1 are satisfied. Let $X(\cdot; 2, 0.5, 0)$ be the random field in Example 3.1 associated to $(\sigma_1, \sigma_2, \theta) = (2, 0.5, 0)$. As noted in Remark 3.1, the quantity M_n should be interpreted as the pixel density of the discretized domain T , and ϵ_n should be understood as the corresponding pixel width. Figure 3.9 provides two illustrations of $E_X(u)$, with $u = 0.5$, in the domain T ; one containing $M_2 \times M_2$ pixels, and another containing of $M_3 \times M_3$ pixels. In this study, $\mathbb{E}[P_X^T(0.5)] = 21.3$ (computed via Equation (3.9) and the Gaussian Kinematic Formula in Adler and Taylor (2007, Theorem 15.9.5)).

To illustrate the convergence of $\hat{P}_X^{(2)}(\epsilon_n, m_n; T, 0.5)$ to $P_X^T(0.5)$ in $L^1(\Omega)$, the left-hand side of Equation (3.13) is shown numerically with $h_n = \hat{P}_X^{(2)}(\epsilon_n, m_n; T, 0.5)$. Figure 3.7 shows how the mean absolute error (MAE) of the approximation of $\tilde{P}_X(\epsilon_n; T, 0.5)$ (the proxy for $P_X^T(0.5)$; see Section 3.4.1) by the estimator $\hat{P}_X^{(2)}(\epsilon_n, m_n; T, 0.5)$ (shown in blue) approaches 0 as $n \rightarrow \infty$. There is no convergence result for the estimator $(\pi/4)\hat{P}_X^{(1)}(\epsilon_n; T, 0.5)$ (shown in green) since it is not well-suited for anisotropic random fields.

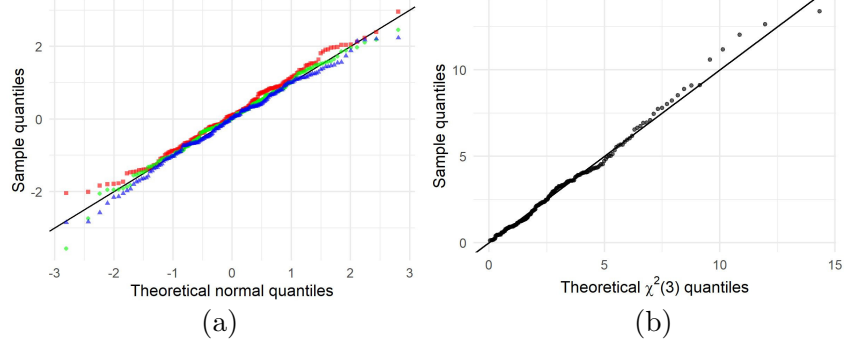


Figure 3.8: An illustration of the asymptotic normality of our estimator for the anisotropic random field $X(\cdot; 2, 0.5, \pi/4)$ in Example 3.1. We simulated 200 independent replications of the vector $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ with $\mathbf{u} = (0, 0.5, 1)$, $T = [-15, 15]^2$, $m = 7$, $\epsilon = 30/1023$. Panel (a): the margins of $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u}) - \mathbb{E}[P_X^T(\mathbf{u})]$, rescaled using the sample variances, plotted on a normal qq-plot. Panel (b): the squared Mahalanobis distance of $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ to $\mathbb{E}[P_X^T(\mathbf{u})]$, calculated via the sample covariance matrix of $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$, plotted against the quantiles of a $\chi^2(3)$ random variable with 3 degrees of freedom.

Asymptotic normality in the anisotropic case

To illustrate the Central Limit Theorem for multiple levels (see Theorem 3.2), we compute $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ in a large domain $T = [-15, 15]^2$ divided into 1024×1024 pixels, with $m = 7$, $\mathbf{u} = (0, 0.5, 1)$, and X as in Example 3.1 with $(\sigma_1, \sigma_2) = (2, 0.5)$ and $\theta = \pi/4$. Figure 3.8 shows how the distribution of the random vector $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ is close to a 3-variate normal distribution with mean $\mathbb{E}[P_X^T(\mathbf{u})] = (793, 700, 481)$ (computed via Equation (3.9)).

For each component u_i of \mathbf{u} , we test the null hypothesis that $\widehat{P}_X^{(2)}(\epsilon, m; T, u_i)$ follows a Gaussian distribution using the Shapiro-Wilk test. The resulting p -values from the tests are 0.39, 0.49, and 0.31, respectively. Thus, the hypothesis of Gaussianity cannot be rejected at a significant level for any margin of $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$. Using the R package `mvmnormtest` (Jarek, 2012), we test the null hypothesis that $\widehat{P}_X^{(2)}(\epsilon, m; T, \mathbf{u})$ follows a multivariate normal distribution with a multivariate Shapiro-Wilk test. The test statistic corresponds to a p -value of 0.14, hence, multivariate normality cannot be rejected at a significant level.

3.4.3 The isotropic case

In what follows, Y denotes the isotropic random field in Example 3.1. This isotropic case allows for a fair comparison between the estimators $(\pi/4)\widehat{P}_Y^{(1)}(\epsilon; T, u)$ and $\widehat{P}_Y^{(2)}(\epsilon, m; T, u)$.

Convergence in mean in the isotropic case

The experiment in Section 3.4.2 is repeated for the isotropic random field Y . Figure 3.9 summarizes the new results. The MAE of the approximation of $\widetilde{P}_Y(\epsilon_n; T, 0.5)$ by $\widehat{P}_Y^{(1)}(\epsilon_n; T, 0.5)$ (shown in green)

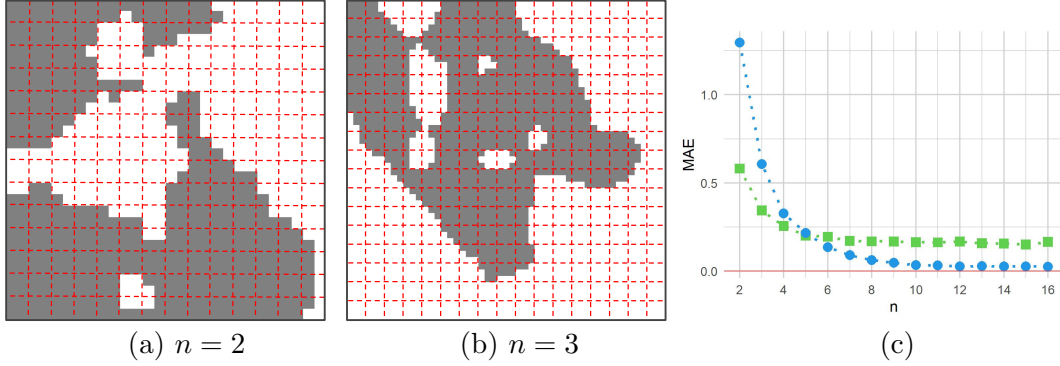


Figure 3.9: The case of decreasing pixel width and fixed domain $T = [-2.5, 2.5]^2$, where $u = 0.5$; M_n , m_n , and ϵ_n are given in (3.14); and Y is the isotropic random field in Example 3.1. See the caption of Figure 3.7 for a more detailed description of each panel.

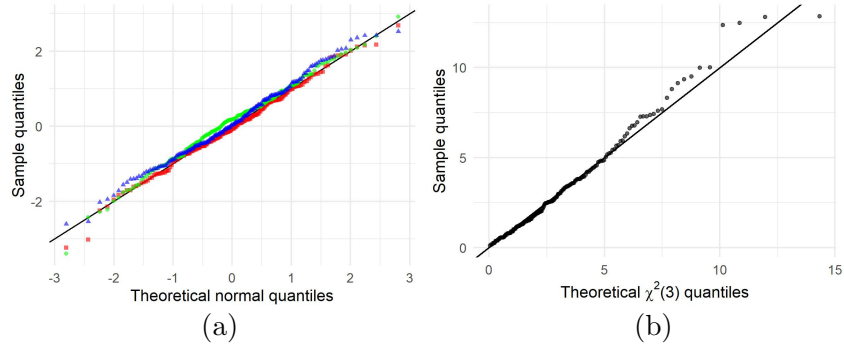


Figure 3.10: An illustration of the asymptotic normality of our estimator when considering the isotropic random field Y in Example 3.1. We simulated 200 independent replications of the vector $\widehat{P}_Y^{(2)}(\epsilon, m; T, \mathbf{u})$ with $\mathbf{u} = (0, 0.5, 1)$, $T = [-15, 15]^2$, $m = 7$, $\epsilon = 30/1023$. See the caption of Figure 3.8 for a description of each panel.

tends to a positive value, so by (3.13), $(\pi/4)\widehat{P}_Y^{(1)}(\epsilon_n; T, 0.5)$ does not converge to $P_Y^T(0.5)$ in $L^1(\Omega)$, even though $\mathbb{E}[(\pi/4)\widehat{P}_Y^{(1)}(\epsilon_n; T, 0.5)] \rightarrow \mathbb{E}[P_Y^T(0.5)]$ as $n \rightarrow \infty$ (see Equation (3.8)). The interested reader is referred to Theorem 3 in [Biermé and Desolneux \(2021\)](#). For reference, $\mathbb{E}[P_Y^T(0.5)] = 15.6$ (computed via the Gaussian Kinematic Formula in [Adler and Taylor \(2007, Theorem 15.9.5\)](#)).

Asymptotic normality in the isotropic case

We repeat the experiment in Section 3.4.2, which tests the asymptotic normality of our estimator, but now with Y as the underlying random field. The p -values corresponding to the Gaussianity tests for the levels $u = 0, 0.5$, and 1 are $0.80, 0.68$, and 0.43 , respectively. For the multivariate normality test, the resulting p -value is 0.37 . The same diagnostic plots in Section 3.4.2 are provided in Figure 3.10 for this isotropic case.

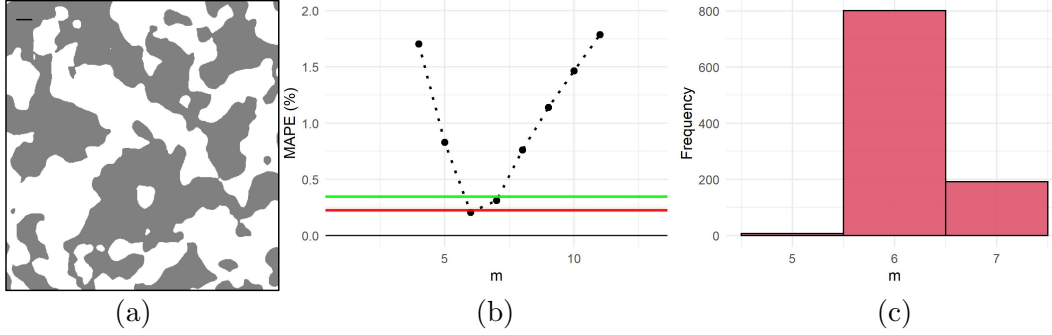


Figure 3.11: Illustration of the influence of the hyperparameter m . The mean absolute percentage error (MAPE) of several perimeter estimators is calculated for 1000 independent replications of the stationary, isotropic, Gaussian random field Y in Example 3.1, with $T = [-10, 10]^2$, $u = 0$, and $\epsilon = 20/511$. The proxy $\hat{P}_Y(\epsilon; T, 0)$ is used to represent the true perimeter $P_Y^T(0)$ for each sample path (see Section 3.4.1). Panel (a): one particular realization of $E_Y(0)$ is depicted in T . Shown for scale in the top-left of the image is a line segment with length 30ϵ . Panel (b): the points plotted in black correspond to the MAPE of $\hat{P}_Y^{(2)}(\epsilon, m; T, 0)$ for various values of m . The green horizontal line (0.35%) corresponds to the MAPE of $(\pi/4)\hat{P}_Y^{(1)}(\epsilon; T, 0)$, which obviously does not depend on m . The red horizontal line (0.22%) corresponds to the MAPE of $\hat{P}_Y^{(2)}(\epsilon, m_Y^T; T, 0)$, with m_Y^T as in (3.15). Panel (c): the values of m_Y^T computed from the 1000 independent replications of Y .

3.4.4 Hyperparameter selection

In practice, sampling locations often have a fixed spacing, and it is not possible to further decrease the grid spacing in the discretization. In these cases, the pixel width ϵ is a feature of the data. So, to use $\hat{P}_X^{(2)}(\epsilon, m; T, u)$ (for an arbitrary model X), the hyperparameter m must be chosen appropriately. As a rule-of-thumb, empirical studies suggest that it is reasonable to choose

$$m = m_X^T := \lfloor C\epsilon^{-2/3} \rfloor, \quad (3.15)$$

with

$$C := \frac{1}{3} \left(\frac{\nu(T)}{N_{cc} + N_{holes}} \right)^{1/3},$$

where N_{cc} (*resp.* N_{holes}) corresponds to the number of connected components (*resp.* holes) of $E_X(T, u)$. For a sequence $(\epsilon_n)_{n \geq 1}$ tending to 0, the corresponding sequence $(m_n)_{n \geq 1}$ determined by (3.15) satisfies the asymptotic relationship required by Theorem 3.1.

In practice, the quantities N_{cc} and N_{holes} can be estimated by considering the sites in $\mathcal{G}^{(T, \epsilon)}$ to be either 4-connected or 8-connected, and colouring each site based on its corresponding value in $\zeta_X^{(T, \epsilon)}(u)$.

Figures 3.11 and 3.12 showcase the performance of $\hat{P}_Y^{(2)}(\epsilon, m_Y^T; T, 0)$, with m_Y^T as in (3.15), for two different levels of discretization of the isotropic random field Y in Example 3.1.

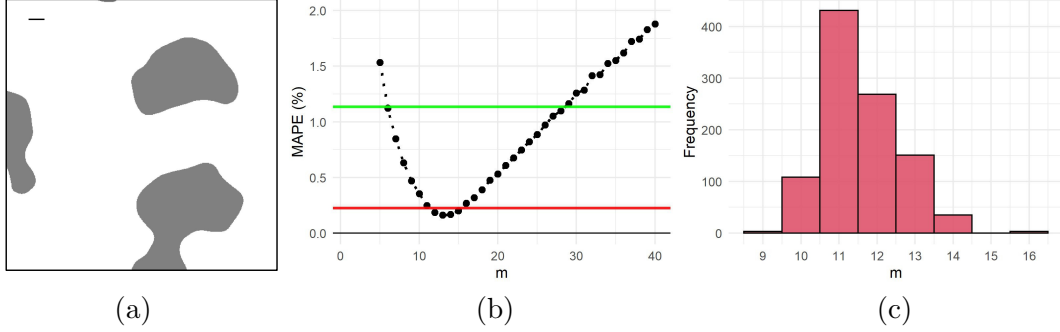


Figure 3.12: See the caption of Figure 3.11 for a description of each panel. In this case, $T = [-2.5, 2.5]^2$ and $\epsilon = 5/511$. The MAPE of $(\pi/4)\widehat{P}_Y^{(1)}(\epsilon; T, 0)$ is 1.13%, and that of $\widehat{P}_Y^{(2)}(\epsilon, m_Y^T; T, 0)$ is 0.22%.

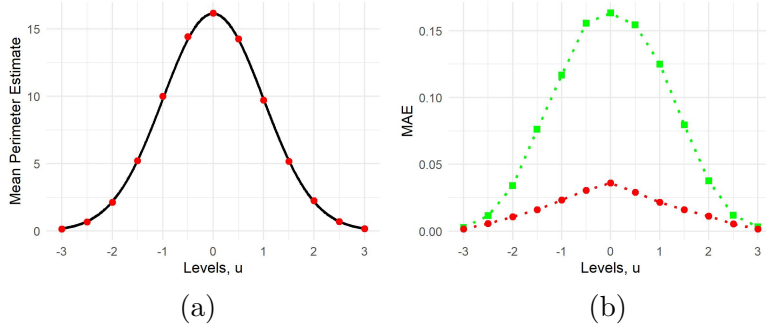


Figure 3.13: Illustration of perimeter estimation for several levels u . The stationary, isotropic, Gaussian random field Y in Example 3.1 is considered on $T = [-2.5, 2.5]^2$ with a discretization of $\epsilon = 5/511$. Panel (a): the sample mean of 500 independent replications of $\widehat{P}_Y^{(2)}(\epsilon, m_Y^T; T, u)$ plotted in red for several values of u , shown against $\mathbb{E}[P_Y^T(u)]$ in black (computed via the Gaussian Kinematic Formula in Adler and Taylor (2007, Theorem 15.9.5)). Panel (b): the MAE of the approximation of $\widehat{P}_Y(\epsilon; T, u)$ by $\widehat{P}_Y^{(2)}(\epsilon, m_Y^T; T, u)$ (red circles) and $(\pi/4)\widehat{P}_Y^{(1)}(\epsilon; T, u)$ (green squares).

3.4.5 Behaviour of the perimeter estimator as a function of the level u

Differently from our previous numerical studies, we illustrate the behaviour of $\widehat{P}_Y^{(2)}(\epsilon, m_Y^T; T, u)$ as a function of the level u in Figure 3.13, where Y is the isotropic random field in Example 3.1. The same is done for an anisotropic field X in Figure 3.14.

3.5 Proofs

This section provides detailed justifications for the theoretical results stated thus far. The following definition is used throughout this section.

Definition 3.6. For $s \in \mathbb{R}^2$, define the set $B_s^{(l)} := [0, l]^2 + s$, where “+” in this context denotes

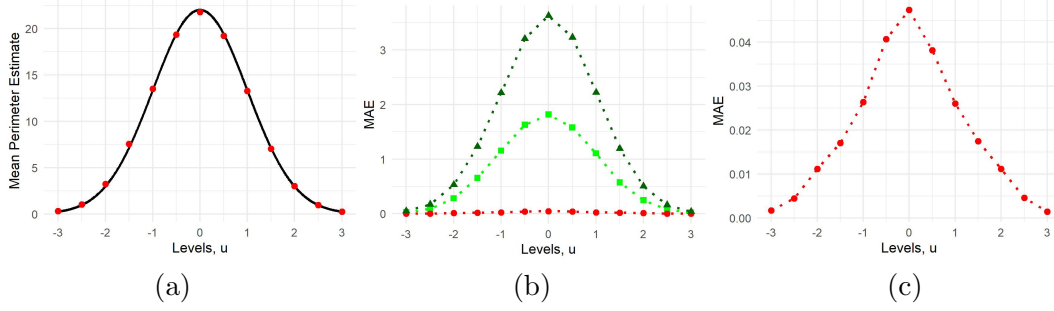


Figure 3.14: The same experiment as depicted in Figure 3.13, but using the stationary, anisotropic, Gaussian random field $X(\cdot; 2, 0.5, 0)$ in Example 3.1. Panel (a): $\mathbb{E}[P_X^T(u)]$, shown in black, is calculated via Equation (3.9) and the Gaussian Kinematic Formula in Adler and Taylor (2007, Theorem 15.9.5). Panel (b): the MAE of the approximation of $\tilde{P}_X(\epsilon; T, u)$ by $\hat{P}_X^{(2)}(\epsilon, m_X^T; T, u)$ (red circles), $(\pi/4)\hat{P}_X^{(1)}(\epsilon; T, u)$ (green squares), and $\hat{P}_X^{(1)}(\epsilon; T, u)$ (dark green triangles). Panel (c): the MAE associated to $\hat{P}_X^{(2)}(\epsilon, m_X^T; T, u)$ shown again on a more appropriate y-axis scale.

the Minkowski sum. Let $\epsilon > 0$ and $m \in \mathbb{N}^+$. Define

$$\mathcal{V}_X^T(\epsilon, m; u) := \{s_{i,j} \in \mathcal{G}^{(T,\epsilon)} : i, j \in I^{(T,\epsilon,m)}, B_{s_{i,j}}^{(m\epsilon)} \cap E_X^\partial(T, u) \neq \emptyset\}.$$

The following lemma allows us to bound $\#(\mathcal{V}_X^T(\epsilon, m; u))$, which amounts to an upper bound on the number of nonzero terms in the sum given by Equation (3.7). See Figure 3.16 in the appendix for an illustration that complements Lemma 3.1.

Lemma 3.1. *Let X be a random field satisfying Assumption 3.1. For any $\epsilon > 0$ and $m \in \mathbb{N}^+$,*

$$\#(\mathcal{V}_X^T(\epsilon, m; u)) \leq 4 \left(\frac{P_X^T(u)}{m\epsilon} + \#(\Gamma_X^T(u)) \right), \quad a.s.$$

Proof. The squares of side length $m\epsilon$ in the set $\mathcal{B} := \{B_{s_{i,j}}^{(m\epsilon)} : i, j \in I^{(T,\epsilon,m)}\}$ are disjoint and cover T . For each $\gamma \in \Gamma_X^T(u)$, it is possible to find connected subsets of γ , namely $\beta_{\gamma,1}, \dots, \beta_{\gamma,M_\gamma}$, that satisfy

$$\gamma = \bigcup_{i=1}^{M_\gamma} \beta_{\gamma,i},$$

where

$$M_\gamma := \left\lfloor \frac{\mathcal{H}^1(\gamma)}{m\epsilon} \right\rfloor + 1,$$

and for all $i \in \{1, \dots, M_\gamma\}$,

$$\mathcal{H}^1(\beta_{\gamma,i}) \leq m\epsilon.$$

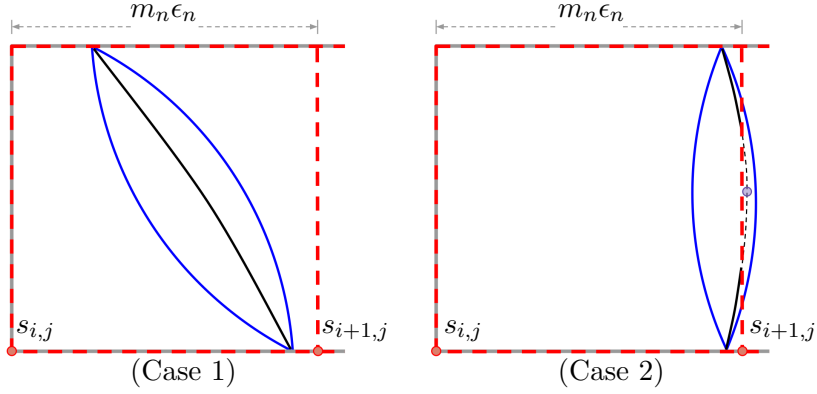


Figure 3.15: (Case 1) The curve γ shown in black is bounded by the planar arcs of radius $m_{n_0}\epsilon_{n_0}$ shown in blue. (Case 2) Here, γ shown in black is not connected, and the only point in $\mathcal{Y}_{X(\omega)}^T(u) \cap B_{s_{i+1,j}}^{(m_n\epsilon_n)}$ is highlighted in purple.

Each $\beta_{\gamma,i}$ can intersect at most 4 elements of \mathcal{B} . Since

$$E_X^\partial(T, u) = \bigcup_{\gamma \in \Gamma_X^T(u)} \bigcup_{i=1}^{M_\gamma} \beta_{\gamma,i}$$

it follows that

$$\begin{aligned} \#(\mathcal{V}_X^T(\epsilon, m; u)) &= \#\{b \in \mathcal{B} : b \cap E_X^\partial(T, u) \neq \emptyset\} \\ &\leq 4 \sum_{\gamma \in \Gamma_X^T(u)} M_\gamma \leq 4 \left(\frac{P_X^T(u)}{m\epsilon} + \#\Gamma_X^T(u) \right), \quad a.s. \end{aligned}$$

□

Proof of Theorem 3.1. Let $\omega \in \Omega$ be such that $\Lambda_{X(\omega)}^T(u)$, defined in Definition 3.5, is positive (note that almost any $\omega \in \Omega$ will suffice, as discussed in Remark 3.4). There exists $n_0 \in \mathbb{N}^+$ such that $E_{X(\omega)}(u)$ is resolved by $m_n\epsilon_n$ in T for all $n \geq n_0$ (see Definition 3.5). Fix $s_{i,j} \in \mathcal{V}_{X(\omega)}^T(\epsilon_n, m_n; u)$ and $n \geq n_0$. Let $\gamma := B_{s_{i,j}}^{(m_n\epsilon_n)} \cap E_{X(\omega)}^\partial(T, u)$. It follows from our construction that $\mathcal{Y}_{X(\omega)}^T(u) \cap \gamma$ contains at most one element, since the spacing between points in $\mathcal{Y}_{X(\omega)}^T(u)$ is larger than the diameter of $B_{s_{i,j}}^{(m_n\epsilon_n)}$. It also follows from our construction that γ is either connected, or the union of two maximally connected subsets. To see this, note that the planar curvature of γ does not exceed $1/(m_{n_0}\epsilon_{n_0})$ since $m_{n_0}\epsilon_{n_0}$ is smaller than the reach of both $E_{X(\omega)}(T, u)$ and $T \setminus E_{X(\omega)}(u)$. Therefore, the curve is bounded by the planar arcs of radius $m_{n_0}\epsilon_{n_0}$ as shown in Figure 3.15 (Dubins, 1961). We aim to bound the absolute difference between the length of γ and its contribution to $\widehat{P}_{X(\omega)}^{(2)}(\epsilon_n, m_n; T, u)$. To this end, the two cases shown in Figure 3.15 are considered separately.

Case 1: The curve γ is connected (see the left panel of Figure 3.15). The closure of γ can be

parametrized by a continuous injective vector function $\mathbf{x} : [0, 1] \rightarrow \mathbb{R}^2$. For $\alpha \in [0, 1]$, define

$$\mathrm{TV}_k(\alpha; s_{i,j}) := \int_0^\alpha |x'_k(s)| \, ds, \quad k \in \{1, 2\}, \quad (3.16)$$

so that $\mathrm{TV}_k(1; s_{i,j})$ corresponds to the total variation of γ in the k^{th} principle Cartesian direction of \mathbb{R}^2 . As a consequence of the coarea formula (Adler and Taylor, 2007, Equation (7.4.15)), the quantity $\epsilon_n N_{X(\omega),h}(i, j; u)$ (see Definition 3.4) is a Riemann sum that approximates the definite integral $\mathrm{TV}_1(1; s_{i,j})$. The total error can therefore be bounded above by

$$|\epsilon_n N_{X(\omega),h}(i, j; u) - \mathrm{TV}_1(1; s_{i,j})| \leq 4\epsilon_n, \quad (3.17)$$

as suggested by Figure 3.17, found in the appendix. Analogously,

$$|\epsilon_n N_{X(\omega),v}(i, j; u) - \mathrm{TV}_2(1; s_{i,j})| \leq 4\epsilon_n.$$

Let

$$\widehat{l}_n(s_{i,j}) := \epsilon_n \left\| (N_{X(\omega),v}(i, j; u), N_{X(\omega),h}(i, j; u)) \right\|_2, \quad (3.18)$$

and we achieve the following bound by the triangle inequality

$$\left| \widehat{l}_n(s_{i,j}) - \left\| (\mathrm{TV}_1(1; s_{i,j}), \mathrm{TV}_2(1; s_{i,j})) \right\|_2 \right| \leq 4\sqrt{2}\epsilon_n. \quad (3.19)$$

It is clear that

$$\|\mathbf{x}(1) - \mathbf{x}(0)\|_2 \leq \left\| (\mathrm{TV}_1(1; s_{i,j}), \mathrm{TV}_2(1; s_{i,j})) \right\|_2, \quad (3.20)$$

since the computation of the left-hand side of Equation (3.20) involves the same integral as in (3.16) but without the absolute values. In addition, let

$$l(\alpha; s_{i,j}) := \int_0^\alpha \|\mathbf{x}'(s)\|_2 \, ds$$

denote the length of $\mathbf{x}(s)$ for $s \in [0, \alpha]$. It follows from the definition of the derivative and the reverse triangle inequality that for all $\alpha \in (0, 1)$,

$$\left| \frac{\partial}{\partial \alpha} \left\| (\mathrm{TV}_1(\alpha; s_{i,j}), \mathrm{TV}_2(\alpha; s_{i,j})) \right\|_2 \right| \leq \left\| (x'_1(\alpha), x'_2(\alpha)) \right\|_2 = \frac{\partial}{\partial \alpha} l(\alpha; s_{i,j}).$$

Therefore,

$$\left\| (\mathrm{TV}_1(1; s_{i,j}), \mathrm{TV}_2(1; s_{i,j})) \right\|_2 \leq l(1; s_{i,j}). \quad (3.21)$$

Since the curvature of γ is bounded above by the inverse of $\Lambda_{X(\omega)}^T(u)$, we apply a well known

result from Schwartz (Dubins, 1961) that guarantees that

$$l(1; s_{i,j}) \leq a(s_{i,j}), \quad (3.22)$$

where $a(s_{i,j})$ is the length of the smallest planar arc with radius $m_{n_0}\epsilon_{n_0}$ that has endpoints $\mathbf{x}(0)$ and $\mathbf{x}(1)$. The Taylor expansion of the sine function shows the existence of $K \in \mathbb{R}^+$ independent of $s_{i,j}$ and n such that

$$\left| a(s_{i,j}) - \|\mathbf{x}(1) - \mathbf{x}(0)\|_2 \right| \leq K \|\mathbf{x}(1) - \mathbf{x}(0)\|_2^3 \leq K(\sqrt{2}m_n\epsilon_n)^3. \quad (3.23)$$

Assembling the bounds demonstrated in Equations (3.20), (3.21), and (3.22), we get

$$\|\mathbf{x}(1) - \mathbf{x}(0)\|_2 \leq \left\| (\text{TV}_1(1; s_{i,j}), \text{TV}_2(1; s_{i,j})) \right\|_2 \leq l(1; s_{i,j}) \leq a(s_{i,j}),$$

which in combination with (3.23) implies

$$\left| l(1; s_{i,j}) - \left\| (\text{TV}_1(1; s_{i,j}), \text{TV}_2(1; s_{i,j})) \right\|_2 \right| \leq K(\sqrt{2}m_n\epsilon_n)^3. \quad (3.24)$$

Now, combining Equations (3.24) and (3.19) by the triangle inequality yields

$$\left| \widehat{l}_n(s_{i,j}) - l(1; s_{i,j}) \right| \leq K(\sqrt{2}m_n\epsilon_n)^3 + 4\sqrt{2}\epsilon_n. \quad (3.25)$$

Case 2: *The curve γ has two connected components (see the right panel of Figure 3.15).* Similarly to Case 1, we parametrize the closure of each maximally connected subset of γ with continuous injective vector functions $\mathbf{x} : [0, 1] \rightarrow \mathbb{R}^2$ and $\mathbf{y} : [0, 1] \rightarrow \mathbb{R}^2$. For $\alpha \in [0, 1]$, define

$$\text{TV}_k(\alpha; s_{i,j}) := \int_0^\alpha (|x'_k(s)| + |y'_k(s)|) \, ds, \quad k \in \{1, 2\}.$$

With $\widehat{l}_n(s_{i,j})$ defined as in (3.18), Equation (3.19) holds. Now, consider the curve $\widetilde{\gamma} := (B_{s_{i,j}}^{(m_n\epsilon_n)} \cup B_{s_{i+1,j}}^{(m_n\epsilon_n)}) \cap E_{X(\omega)}^\partial(T, u)$, which is γ in union with the middle section in the adjacent box $B_{s_{i+1,j}}^{(m_n\epsilon_n)}$ (where we have assumed, without loss of generality, that the “middle section” is in the box to the right). It is clear that $\widetilde{\gamma}$ is connected, so its closure can be parametrized by the continuous injective vector function $\mathbf{z} : [0, 1] \rightarrow \mathbb{R}^2$. Define

$$\widetilde{\text{TV}}_k(\alpha; s_{i,j}) := \int_0^\alpha |z'_k(s)| \, ds, \quad k \in \{1, 2\}$$

and

$$\widetilde{l}(\alpha; s_{i,j}) := \int_0^\alpha \|\mathbf{z}'(s)\|_2 \, ds.$$

By the same arguments that led to Equation (3.24), it holds that

$$0 \leq \tilde{l}(1; s_{i,j}) - \left\| (\widetilde{\text{TV}}_1(1; s_{i,j}), \widetilde{\text{TV}}_2(1; s_{i,j})) \right\|_2 \leq K(\sqrt{2}m_n\epsilon_n)^3, \quad (3.26)$$

where $K \in \mathbb{R}^+$ is independent of $s_{i,j}$ and n . Let

$$l(1; s_{i,j}) := \int_0^1 (\|\mathbf{x}'(s)\|_2 + \|\mathbf{y}'(s)\|_2) ds$$

be the total length of γ . Then $\tilde{l}(1; s_{i,j}) = l(1; s_{i,j}) + l(1; s_{i+1,j})$, and

$$\begin{aligned} \left\| (\widetilde{\text{TV}}_1(1; s_{i,j}), \widetilde{\text{TV}}_2(1; s_{i,j})) \right\|_2 &\leq \left\| (\text{TV}_1(1; s_{i,j}), \text{TV}_2(1; s_{i,j})) \right\|_2 \\ &\quad + \left\| (\text{TV}_1(1; s_{i+1,j}), \text{TV}_2(1; s_{i+1,j})) \right\|_2 \end{aligned}$$

by the triangle inequality. Therefore, (3.26) can be written as

$$\begin{aligned} &\left(l(1; s_{i,j}) - \left\| (\text{TV}_1(1; s_{i,j}), \text{TV}_2(1; s_{i,j})) \right\|_2 \right) + \\ &\left(l(1; s_{i+1,j}) - \left\| (\text{TV}_1(1; s_{i+1,j}), \text{TV}_2(1; s_{i+1,j})) \right\|_2 \right) \leq K(\sqrt{2}m_n\epsilon_n)^3. \end{aligned} \quad (3.27)$$

By the arguments in Case 1 that led to Equation (3.21), it follows that

$$l(1; s_{i+1,j}) \geq \left\| (\text{TV}_1(1; s_{i+1,j}), \text{TV}_2(1; s_{i+1,j})) \right\|_2,$$

and by the same arguments,

$$l(1; s_{i,j}) \geq \left\| (\text{TV}_1(1; s_{i,j}), \text{TV}_2(1; s_{i,j})) \right\|_2.$$

Therefore, both (3.24) and (3.25) follow from Equation (3.27).

Following from Equation (3.25), we have

$$\begin{aligned} \left| \widehat{P}_{X(\omega)}^{(2)}(\epsilon_n, m_n; T, u) - P_{X(\omega)}^T(u) \right| &= \left| \sum_{s_{i,j} \in \mathcal{V}_{X(\omega)}^T(\epsilon_n, m_n; u)} \left(\widehat{l}_n(s_{i,j}) - l(1; s_{i,j}) \right) \right| \\ &\leq \sum_{s_{i,j} \in \mathcal{V}_{X(\omega)}^T(\epsilon_n, m_n; u)} \left| \widehat{l}_n(s_{i,j}) - l(1; s_{i,j}) \right| \\ &\leq \#(\mathcal{V}_{X(\omega)}^T(\epsilon_n, m_n; u)) 2\sqrt{2}(Km_n^3\epsilon_n^3 + 2\epsilon_n). \end{aligned}$$

By Lemma 3.1,

$$\begin{aligned}
& g_n |\widehat{P}_{X(\omega)}^{(2)}(\epsilon_n, m_n; T, u) - P_{X(\omega)}^T(u)| \\
& \leq 8\sqrt{2}g_n \left(\frac{P_{X(\omega)}^T(u)}{m_n\epsilon_n} + \#(\Gamma_{X(\omega)}^T(u)) \right) (Km_n^3\epsilon_n^3 + 2\epsilon_n) \\
& = 8\sqrt{2} \frac{g_n}{m_n} \left(P_{X(\omega)}^T(u) + m_n\epsilon_n\#(\Gamma_{X(\omega)}^T(u)) \right) (Km_n^3\epsilon_n^2 + 2), \tag{3.28}
\end{aligned}$$

which tends to 0 as $n \rightarrow \infty$. This convergence holds for almost every $\omega \in \Omega$, since $\Lambda_X^T(u)$ is almost surely positive. \square

Proof of Corollary 3.1. The last expression in Equation (3.28) tends to 0 under the relaxed constraint on $(\epsilon_n)_{n \geq 1}$ if $g_n \equiv 1$ for all $n \in \mathbb{N}^+$. \square

Proof of Proposition 3.1. If a sequence is uniformly integrable, convergence in $L^1(\Omega)$ is equivalent to convergence in probability. Therefore, by Corollary 3.1, it suffices to show that $(\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u))_{n \geq 1}$ is bounded above by an element of $L^1(\Omega)$ uniformly in n . Note that for each $n \geq 1$,

$$\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u) \leq \widehat{P}_X^{(1)}(\epsilon_n; T, u), \quad a.s.$$

since the 2-norm is inferior to the 1-norm. Now, consider the quantity

$$G_n := \#(\{s \in \mathcal{G}^{(T, \epsilon_n)} : B_s^{(\epsilon_n)} \cap E_X^\partial(T, u) \neq \emptyset\}),$$

which represents the number of pixels of side length ϵ_n that the curve $E_X^\partial(T, u)$ intersects. Almost surely, $\widehat{P}_X^{(1)}(\epsilon_n; T, u)$ is at most $4\epsilon_n$ (the perimeter of one pixel) times G_n . By the same arguments used to prove Lemma 3.1, we have for all $n \geq 1$,

$$G_n \leq 4 \left(\frac{P_X^T(u)}{\epsilon_n} + \#(\Gamma_X^T(u)) \right), \quad a.s.$$

and

$$\widehat{P}_X^{(2)}(\epsilon_n, m_n; T, u) \leq \widehat{P}_X^{(1)}(\epsilon_n; T, u) \leq 4\epsilon_n G_n \leq 16 \left(P_X^T(u) + \sup_n \epsilon_n \#(\Gamma_X^T(u)) \right), \quad a.s.$$

which is in $L^1(\Omega)$ by Assumption 3.3. \square

Proof of Proposition 3.2. Let

$$W_n := \frac{\widehat{P}_X^{(2)}(\epsilon_n, m_n; T_n, u) - P_X^{T_n}(u)}{\sqrt{\nu(T_n)}}.$$

Given that $E_X(u)$ is resolved by $m_n\epsilon_n$ in T_n for fixed $n \in \mathbb{N}^+$, Equation (3.28) holds with $g_n =$

$1/\sqrt{\nu(T_n)}$, implying

$$\begin{aligned} |W_n| &\leq \frac{8}{m_n} \sqrt{\frac{2}{\nu(T_n)}} \left(P_X^{T_n}(u) + m_n \epsilon_n \#(\Gamma_X^{T_n}(u)) \right) \left(K m_n^3 \epsilon_n^2 + 2 \right) \\ &= \frac{8\sqrt{2\nu(T_n)}}{m_n} \left(\frac{P_X^{T_n}(u)}{\nu(T_n)} + m_n \epsilon_n \frac{\#(\Gamma_X^{T_n}(u))}{\nu(T_n)} \right) \left(K m_n^3 \epsilon_n^2 + 2 \right), \end{aligned} \quad (3.29)$$

where $K \in \mathbb{R}^+$ is independent of n . Note that for any $n \in \mathbb{N}^+$,

$$T_n = \bigcup_{i=1}^{n^2} T_n^{(i)},$$

for a family of sets $(T_n^{(i)})_{i=1, \dots, n^2}$, each of which being congruent to T_1 . Then

$$\mathbb{E} \left[\frac{P_X^{T_n}(u)}{\nu(T_n)} \right] = \mathbb{E} \left[\frac{\sum_{i=1}^{n^2} P_X^{T_n^{(i)}}(u)}{n^2 \nu(T_1)} \right] = \mathbb{E} \left[\frac{P_X^{T_1}(u)}{\nu(T_1)} \right] < \infty$$

and

$$\mathbb{E} \left[\frac{\#(\Gamma_X^{T_n}(u))}{\nu(T_n)} \right] \leq \mathbb{E} \left[\frac{\sum_{i=1}^{n^2} \#(\Gamma_X^{T_n^{(i)}}(u))}{n^2 \nu(T_1)} \right] = \mathbb{E} \left[\frac{\#(\Gamma_X^{T_1}(u))}{\nu(T_1)} \right] < \infty,$$

by Assumption 3.3. This implies that both

$$\limsup_{n \rightarrow \infty} \frac{P_X^{T_n}(u)}{\nu(T_n)} \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{\#(\Gamma_X^{T_n}(u))}{\nu(T_n)}$$

are finite almost surely. Therefore, the final expression in (3.29) tends to 0 almost surely, since $\sqrt{\nu(T_n)}/m_n \rightarrow 0$ by assumption. Now, denote the random event $A_n := \{m_n \epsilon_n < \Lambda_X^{T_n}(u)\}$, and let A_n^C denote its complement. Since $\mathbb{P}(A_n) \rightarrow 1$ as $n \rightarrow \infty$ by assumption, it holds that for any $\eta > 0$,

$$\mathbb{P}(|W_n| > \eta) \leq \mathbb{P}(|W_n| > \eta \mid A_n) \mathbb{P}(A_n) + \mathbb{P}(A_n^C) \rightarrow 0$$

as $n \rightarrow \infty$. □

Proof of Theorem 3.2. The Central Limit Theorem in Iribarren (1989) for $P_X^{T_n}(u)$ at the fixed level $u \in \mathbb{R}$ is implied by the constraints on X . The result is proven for a single level u , but as noted in the Discussion of Kratz and Vadlamani (2018) and in Shashkin (2013), the Cramér-Wald device can be used to extend the arguments to the multivariate setting.

The Central Limit Theorem for the perimeter is then written as follows. For any $\mathbf{u} \in \mathbb{R}^k$ satisfying the given constraints, it holds that

$$\frac{P_X^{T_n}(\mathbf{u}) - \mathbb{E}[P_X^{T_n}(\mathbf{u})]}{\sqrt{\nu(T_n)}} \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \Sigma(\mathbf{u})), \quad n \rightarrow \infty. \quad (3.30)$$

Equation (3.10) is obtained by combining Equation (3.30), Proposition 3.2, and Slutsky's theorem.

By writing $P_X^T(u) = \lim_{\epsilon \rightarrow 0} 1/(2\epsilon) \int_T \mathbf{1}_{\{|X(s)-u|<\epsilon\}} \|\nabla X(s)\|_2 ds$, (see, for instance, Proposition 6.13 in Azaïs and Wschebor (2009)), it is easily checked that for $u_1, u_2 \in \mathbb{R}$,

$$\mathbb{E}[P_X^T(u_1)] = \nu(T) f(u_1) \mathbb{E}[\|\nabla X(s)\|_2 \mid X(s) = u_1]$$

and

$$\begin{aligned} \mathbb{E}[P_X^T(u_1)P_X^T(u_2)] &= \int_T \int_T g_{s_2-s_1}(u_1, u_2) \\ &\quad \times \mathbb{E}[\|\nabla X(s_1)\|_2 \|\nabla X(s_2)\|_2 \mid X(s_1) = u_1, X(s_2) = u_2] ds_1 ds_2, \end{aligned}$$

where f denotes the marginal density function of X , and g_s , the joint density function of $(X(0), X(s))$. Hence the result in (3.11). \square

Proof of Corollary 3.2. Under the given constraints, it is clear that Assumption 3.1 is satisfied. Also following from the hypotheses, the gradient of X and the Hessian matrix of X are independent with Gaussian entries, and thus the conditions of Theorem 11.3.3 of Adler and Taylor (2007) are satisfied. Therefore, X is almost surely *suitably regular* (Adler and Taylor, 2007, Definition 6.2.1) over bounded rectangles, which implies the conditions of Assumption 3.2. The expectations $\mathbb{E}[P_X^{T_n}(u)]$ and $\mathbb{E}[\#\{\Gamma_X^{T_n}(u)\}]$ are shown to be finite in Adler and Taylor (2007, Theorem 13.2.1) and Beliaev et al. (2020) respectively, implying the conditions of Assumption 3.3. Therefore, Proposition 3.2 holds, which in combination with the Central Limit Theorem in Berzin (2021, Theorem 4.7) yields the result. \square

Discussion

We have shown for a large class of random fields that $\widehat{P}_X^{(p)}(\epsilon, m; T, u)$ with $p = 2$ is a consistent and asymptotically normal estimator for $P_X^T(u)$. Our numerous simulation studies showcase the various cases where it is advantageous to use the norm $p = 2$ as opposed to $p = 1$. An obvious example is when X is not known to be isotropic. For $p > 2$, we do not expect $\widehat{P}_X^{(p)}(\epsilon, m; T, u)$ to have desirable properties, since there is a bias introduced for certain orientations of the curve $E_X^\partial(T, u)$. There is a natural extension of $\widehat{P}_X^{(p)}(\epsilon, m; T, u)$ to random fields defined on \mathbb{R}^d , with $d > 2$, and it is plausible that analogous results hold in this multivariate setting. Results such as the central limit theorems in Shashkin (2013), Müller (2017), and Kratz and Vadlamani (2018), which hold in arbitrary dimension, will be useful to study the Gaussian fluctuations of our estimate.

Future work might also investigate the rate at which $\Lambda_X^T(u)$ tends weakly to 0 as $T \nearrow \mathbb{R}^2$, which would provide a more explicit constraint on the rate at which $\epsilon_n \rightarrow 0$ in Proposition 3.2.

Furthermore, we plan to study how the proposed perimeter estimate can be used to build a test statistics for isotropy testing based on the length of level curves of smooth random fields. This future analyse could enrich the existing literature of isotropy testing based on functionals of level curves (Wschebor, 1985; Cabaña, 1987; Fournier, 2018; Berzin, 2021).

The proposed estimator works with observations available at a set of locations forming a regular grid. A large variety of datasets possess this format, such as outputs of various types of models (*e.g.*, climate, hydrology), remote sensing data, or imaging data (*e.g.*, in medicine). However, geostatistical spatial data are sometimes not observed on regular grids, such as meteorological data observed over a network of weather stations not organised in any grid structures. In such cases, one could first apply a deterministic or stochastic interpolation method (*e.g.*, bilinear interpolation, geostatistical kriging) to pre-process data to make them available on a regular grid, and then use the grid-based estimator.

In this paper, we have focused on perimeter estimator properties in the case of a single replicate of the random field with one or several fixed levels u . Properties of estimators of Lipschitz–Killing curvatures, including the perimeter, could further be studied when the level u tends towards the upper endpoint of the marginal distribution of X . This setting is relevant for extreme-value theory of stochastic processes (de Haan and Ferreira, 2006, Chapters 9–10). Jointly with decreasing pixel size and increasing domain T , we would further have to control the rate at which the perimeter tends towards zero as u increases, where ultimately the excursion set is almost surely empty. The combination of the results obtained for our perimeter estimator with asymptotics of the exact perimeter for increasing level u (Adler and Taylor, 2007) could be useful to establish asymptotic results and appropriate estimators for the perimeter and for other excursion-set geometrical features at extreme thresholds.

Appendix

Here, we provide two figures; one to complement Lemma 3.1, and the other, Equation (3.17).

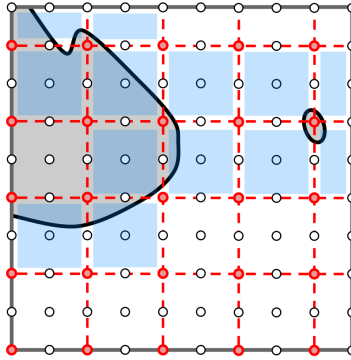


Figure 3.16: An illustration to aid Lemma 3.1. With $m = 2$, the curve $E_X^\partial(T, u)$ shown in black intersects 13 elements of $\{B_{s_{i,j}}^{(m\epsilon)} : i, j \in I^{(T, \epsilon, m)}\}$, which are highlighted in blue. Thus, $\#\mathcal{V}_X^T(\epsilon, m; u) = 13$.

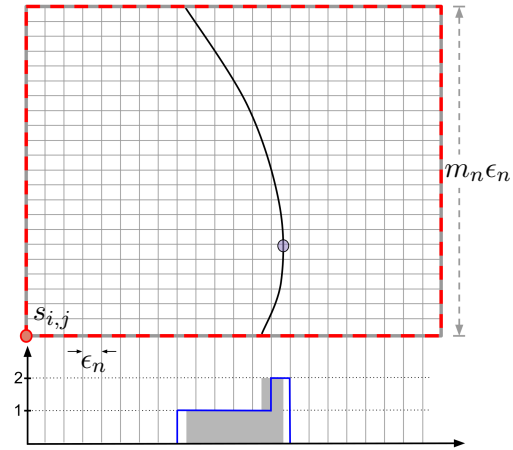



Figure 3.17: The approximation of $TV_1(1, s_{i,j})$ in (3.16) by $\epsilon_n N_{X(\omega), h}(i, j; u)$ (see Definition 3.4). The black curve γ is shown in $B_{s_{i,j}}^{(m_n \epsilon_n)}$, which we outline in dashed red. The definite integral $TV_1(1, s_{i,j})$ is represented by the grey area, and is approximated by $\epsilon_n N_{X(\omega), h}(i, j; u) = 7\epsilon_n$, the area under the blue curve. The absolute error of this approximation is clearly bounded above by $4\epsilon_n$ as stated in Equation (3.17). Highlighted in purple is a point in $\mathcal{Y}_{X(\omega)}^T(u)$ (see Equation (3.2)).

Chapter 4

Computable bounds for the reach and r -convexity of subsets of \mathbb{R}^d

This chapter is based on the work

 Cotsakis, R. (2023). Identifying the reach from high-dimensional point cloud data with connections to r -convexity. *Discrete & Computational Geometry* (to appear). [Paper here](#).

Abstract: The convexity of a set can be generalized to the two weaker notions of positive reach and r -convexity; both describe the regularity of a set’s boundary. For any compact subset of \mathbb{R}^d , we provide methods for computing upper bounds on these quantities from point cloud data. The bounds converge to the respective quantities as the sampling scale of the point cloud decreases, and the rate of convergence for the bound on the reach is given under a weak regularity condition. We also introduce the β -reach, a generalization of the reach that excludes small-scale features of size less than a parameter $\beta \in [0, \infty)$. Numerical studies suggest how the β -reach can be used in high-dimension to infer the reach and other geometric properties of smooth submanifolds.

4.1 Introduction

A number of concepts from convex geometry generalize from convex sets to much larger classes of sets. A classic example from [Federer \(1959\)](#) is the extension of kinematic formulas (in particular, Steiner’s formula) for convex sets to sets with *positive reach* (see [Definition 4.2](#)). Another example is the notion of the convex hull of a set, which can be weakened to the *r -convex hull*, for $r > 0$. This weak notion of a convex hull, instead of being expressed as the intersection of half-spaces, is expressed in terms of intersections of the complements of open balls of radius r . The resulting intersection is said to be *r -convex* (see [Definition 4.4](#)), which differs subtly from the notion of reach, and these differences have been studied in [Colesanti and Manselli \(2007\)](#) and [Cuevas et al. \(2012\)](#) for example. Both the reach and the radius of r -convexity—ranging from 0 to ∞ inclusive—can

be seen as measures of the degree to which a set is convex. This paper introduces methods of computing upper bounds for these quantities from point cloud data in several, general settings.

There is a vast literature on properties of r -convex sets, dating from [Perkal \(1956\)](#). Their relations to other classes of sets that generalize convexity have been studied in [Serra \(1984\)](#); [Walther \(1999\)](#); [Colesanti and Manselli \(2007\)](#); [Cuevas et al. \(2012\)](#). The use of r -convexity in image smoothing has been suggested in [Serra \(1984\)](#); [Walther \(1999\)](#); [Cuevas et al. \(2007\)](#) as well as for set estimation in [Mani-Levitska \(1993\)](#); [Cuevas et al. \(2007\)](#); [Pateiro López \(2008\)](#); [Cuevas \(2009\)](#); [Cuevas et al. \(2012\)](#); [Aaron et al. \(2022\)](#). In other literature, the r -convex hull of a set is referred to as its *double offset* ([Chazal et al., 2007, 2009b](#); [Kim et al., 2019](#)). In [Cuevas and Rodríguez-Casal \(2004\)](#), the authors use a rolling-type condition (weaker than r -convexity) to improve the rate of convergence of their set boundary estimator. Efforts towards estimating the largest r such that a set is r -convex have been made in [Rodríguez Casal and Saavedra-Nieves \(2016\)](#); the authors test the uniformity of a point cloud on its r -convex hull using the statistical test of uniformity proposed in [Berrendero et al. \(2012\)](#).

The reach is a popular measure of convexity largely due to its link with the Steiner formula; for the larger the reach of a set, the larger the interval on which the volume of a dilation of the set is polynomial in the dilation radius ([Federer, 1959](#), Theorem 5.6). Positive reach is a common hypothesis to guarantee convergence rates of statistical estimators of geometric quantities ([Chazal and Lieutier, 2008](#); [Thäle, 2008](#); [Cuevas, 2009](#); [Rataj and Zähle, 2019](#); [Biermé et al., 2019](#); [Cotsakis et al., 2023c](#)). In the field of geometric data analysis, the reach constitutes a commonly used measure of regularity of a set’s boundary, and is hence also referred to as the *condition number* [Niyogi et al. \(2008\)](#). In topological data analysis, a set’s reach is shown to quantify the ability to infer its homology from point cloud data ([Kim et al., 2019](#); [Niyogi et al., 2008](#); [Lieutier, 2004](#); [Chazal and Lieutier, 2005b](#)).

The statistical estimation of the reach has been of particular interest in recent literature. Both [Aamari et al. \(2019\)](#) and [Aamari and Levrard \(2019\)](#) suggest estimating the reach of a smooth submanifold of \mathbb{R}^d by measuring distances between it and its tangent spaces—a strategy inspired by the formulation of the reach in ([Federer, 1959](#), Theorem 4.18). These works obtain bounds on the minimax rate of convergence for this estimator when the manifold is C^3 -smooth, and when its tangent spaces are known. Following up to these works, [Aamari et al. \(2023\)](#) establishes an optimal convergence rate for minimax estimators of the reach of C^k -smooth submanifolds of \mathbb{R}^d without boundary. The rates that they establish adapt to the smoothness of the manifold, and to the type of phenomenon that limits the reach (see Remark 4.8). In [Berenfeld et al. \(2022\)](#), the *convexity defect function* from [Attali et al. \(2013\)](#) is used to define an estimator for the reach of C^k -smooth manifolds. Convergence rates of their estimator are given for $k \geq 3$. In [Cholaquidis et al. \(2023\)](#), the authors introduce a complete and tractable method for estimating the reach from point cloud data involving the computation of graph distances in a spatial network defined over the point cloud. Their approach is based on the formulation of the reach introduced in ([Boissonnat et al., 2019](#),

Theorem 1).

Other strategies for estimating the reach involve a prior estimation of the *medial axis* (see Definition 4.3) (Dey and Zhao, 2003; Dey and Sun, 2006). The λ -medial axis (Chazal and Lieutier, 2005a), μ -medial axis (Chazal et al., 2009a), and (λ, α) -medial axis (Lieutier and Wintraecken, 2023) are all generalizations of the medial axis, and the reach can conceivably be approximated by the minimal distance from a set to the estimate of the medial axis of its complement. Such a strategy is suggested in Cuevas et al. (2014) for the λ -medial axis, where the authors heavily rely on the notion of r -convexity in their construction.

A number of other authors have taken an interest in the mathematical properties of the reach. In Poliquin et al. (2000), sets of reach r are identified with the r -proximally smooth sets introduced in Clarke and Wolenski (1995), implying that the reach can be characterized by the gradients of the distance-to-set function. A number of insightful connections between the reach and other geometrical properties of sets are made in Colesanti and Manselli (2007). Vietoris–Rips complexes are studied in relation to the reach in Attali and Lieutier (2015), and in the same work, the authors prove that the reach of a set can only increase if intersected by sufficiently small balls. An alternative characterization of the reach, involving pairwise geodesic distances, is provided in Boissonnat et al. (2019). The authors also study the relationship between midpoints of pairs of points in a set, and the reach (Boissonnat et al., 2019, Lemma 1). We base the construction of our bound for the reach on this result (see Theorem 4.2).

This paper makes steps towards providing computationally tractable methods for bounding the reach and r -convexity of subsets of \mathbb{R}^d given point cloud data that represent the underlying sets.

Firstly, we establish some facts about the reach of closed subsets of \mathbb{R}^d . We prove that the radius of r -convexity and the reach are equivalent for compact subsets of \mathbb{R}^d whose topological boundary is a C^1 -smooth, $(d - 1)$ -dimensional manifold without boundary (see Theorem 4.1). In addition, for closed subsets of \mathbb{R}^d , we introduce the β -reach (see Definition 4.6), a quantity that loosely represents the reach of a set when features of size less than $\beta \in [0, \infty)$ are ignored. Indeed, the β -reach converges to the reach from above as $\beta \searrow 0$ (see Theorem 4.2).

These ideas are used to create methods of inferring bounds on the reach and r -convexity of sets from point cloud data. For general, closed subsets of \mathbb{R}^d (possibly having finite d -volume, and having *no smoothness conditions* on their topological boundaries), we provide an upper bound of the r -convexity of the set based on samples of the set and its complement at sampling locations that extend over \mathbb{R}^d . We show that, as the spacing between sampling locations diminishes, this bound converges to the largest r such that the set is r -convex (see Theorem 4.3). An example on real data in 3 dimensions shows that this method identifies regions where the underlying set is not locally r -convex, for $r > 0$, with a test specificity of 100%. Similarly, for any closed set $A \subseteq \mathbb{R}^d$, we define an upper bound on the reach based on a set of points known to reside in A . As the set of points converges in the Hausdorff metric to A , we show that the bound converges to the reach of A , and provide the rate of convergence in terms of the Hausdorff distance between A and the sample

points (see Theorem 4.4). A regularity condition on the β -reach of A , for β near 0, is used to show the convergence of our upper bound with a rate. In practice, both the β -reach of a point cloud and the upper bound on the reach can be computed efficiently in high-dimension. If the number of data points is held fixed, the computational complexity of the method increases only linearly with the dimension of the ambient space d .

The organization of the paper is as follows. Section 4.2 introduces the notation that we use throughout the document, explores the relationships between the reach and r -convexity, and introduces the β -reach. Section 4.3 describes the three methods that we propose for inferring bounds and approximations for the reach and r -convexity of general compact subsets of \mathbb{R}^d from point cloud data. The bounds for the r -convexity and reach are studied in Sections 4.3.1 and 4.3.2 respectively. Section 4.3.3 elaborates on how the β -reach of point clouds can be used to approximate the β -reach of the sets that they represent. In Section 4.4, we provide numerical studies that underline the computability of our methods. An application of the methods on real data is given in Section 4.4.1. In Section 4.4.2, the methods are tested against simulated data for which the reach and r -convexity are known, and empirical rates of convergence are provided. Some technical proofs and auxiliary results are postponed to Section 4.5.

4.2 Definitions and important notions

The sets that we study in this paper are subsets of \mathbb{R}^d , endowed with the Euclidean metric $\|\cdot\|$. For a set $S \subseteq \mathbb{R}^d$, let ∂S denote its topological boundary, let $\text{cl}(S) := S \cup \partial S$ denote the closure of S , and let S^c denote the complement of S in \mathbb{R}^d . Denote the closed ball with radius $r \in \mathbb{R}^+$ centered at $s \in \mathbb{R}^d$ by $B(s, r) := \{t \in \mathbb{R}^d : \|t - s\| \leq r\}$. For $t \in \mathbb{R}^d$, denote the distance between t and a non-empty set S by $\delta_S(t) := \inf\{\|t - s\| : s \in S\}$.

4.2.1 Set dilation, set erosion, and combinations of the two

Definition 4.1 (Operations on subsets of \mathbb{R}^d). We recall the Minkowski addition of two sets $A, B \subseteq \mathbb{R}^d$,

$$A \oplus B := \{x + y : x \in A, y \in B\}.$$

The Minkowski difference is given by

$$A \ominus B := \{x \in \mathbb{R}^d : \{x\} \oplus B \subseteq A\} = (A^c \oplus (-B))^c,$$

where A^c denotes the complement of A in \mathbb{R}^d , and $-B := \{-x : x \in B\}$. For $r \in \mathbb{R}$, let

$$A_r := \begin{cases} A \oplus B(0, r), & \text{for } r \geq 0, \\ A \ominus B(0, -r), & \text{for } r < 0, \end{cases}$$

denote either the dilation or erosion of a set A , depending on the sign of r . Finally, define the set $A_{\bullet r} := (A_r)_{-r}$. In mathematical morphology literature, $A_{\bullet r}$ is referred to as the *closing* (resp. *opening*) of A by $B(0, |r|)$ if $r > 0$ (resp. $r < 0$).

For $r \geq 0$ and $A \subseteq \mathbb{R}^d$ closed, it follows from Definition 4.1 that A_r (also known as the *r-offset* of A in, e.g., Chazal et al. (2007)) denotes all the points in \mathbb{R}^d within a distance r of the set A . The set A_{-r} denotes all the points in A a distance of at least r from ∂A .

With these notions established, the Hausdorff distance between two closed sets $A, B \subseteq \mathbb{R}^d$ is defined as $d_H(A, B) := \inf\{r \in \mathbb{R}^+ : A \subseteq B_r, B \subseteq A_r\}$.

Lemma 4.1. *Let $r, s > 0$, and let $A \subseteq \mathbb{R}^d$. The following identities hold:*

- (a) $(A_r)^c = (A^c)_{-r}$,
- (b) $A \subseteq A_{\bullet r}$,
- (c) $(A_r)_s = A_{r+s}$,
- (d) $(A_r)_{-s} \supseteq A_{r-s}$,
- (e) $(A_{-r})_s \subseteq A_{s-r}$.

Proof of Lemma 4.1. Fix $r, s > 0$. The identity in (a) follows directly from Definition 4.1. Item (b) is proved by contradiction. Let $a \in A$ and suppose $a \in (A_{\bullet r})^c = (A_r)^c \oplus B(0, r)$. Then there is a $p \in (A_r)^c$ such that $a \in B(p, r) \Leftrightarrow \|p - a\| \leq r \Leftrightarrow p \in B(a, r) \Leftrightarrow p \in A_r$; thus, a contradiction. To prove (c), remark that $B(0, r) \oplus B(0, s) = B(0, r + s)$ and that Minkowski addition is associative. To prove (d), consider the case where $r \geq s$, then $(A_r)_{-s} = (A_{r-s+s})_{-s} = ((A_{r-s})_s)_{-s} = (A_{r-s})_{\bullet s} \supseteq A_{r-s}$. For the case $r < s$, write $(A_r)_{-s} = (A_r)_{r-s-r} = ((A_r)_{-r})_{r-s} = (A_{\bullet r})_{r-s} \supseteq A_{r-s}$. To show (e), consider the complements of the sets in (d) and apply (a) repeatedly. \square

4.2.2 The reach and related concepts

Definition 4.2 (The reach). Recall from Federer (1959) that the *reach* of a set $A \subseteq \mathbb{R}^d$ is given by

$$\text{reach}(A) := \sup\{r \in \mathbb{R}^+ : \forall y \in A_r \exists! x \in A \text{ nearest to } y\}.$$

If $\text{reach}(A) > 0$, then A is said to have *positive reach*.

It follows from Definition 4.2 that if A has positive reach, then A is closed (Federer, 1959). In the remainder of this document, A denotes a closed set. A useful notion related to the reach of A is the *medial axis* of A^c , originally proposed in Blum (1967).

Definition 4.3 (The medial axis). Let $O \subseteq \mathbb{R}^d$ be open. Its *medial axis* $\mathcal{M}(O)$ is the set of points in O with at least two closest points in ∂O .

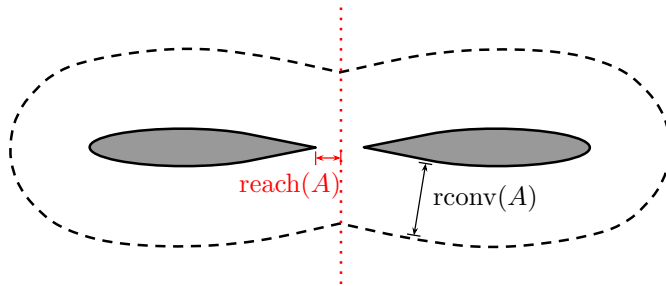


Figure 4.1: The closed set A (in grey) has $\text{rconv}(A) > \text{reach}(A)$. The set A can be expressed as the complement of a union of open balls of radius $\text{rconv}(A)$ whose centers lie outside the dotted black curve. The dotted red line is the medial axis $\mathcal{M}(A^c)$.

The reach of a closed set A can be alternatively expressed as

$$\text{reach}(A) = \inf\{\|a - x\| : a \in A, x \in \mathcal{M}(A^c)\}. \quad (4.1)$$

Connections to r -convexity

Definition 4.4 (r -convexity). A set $A \subseteq \mathbb{R}^d$ is said to be r -convex for $r \in \mathbb{R}^+$ if it is closed and $A_{\bullet s} = A$ for all $s \in (0, r)$ (see, e.g., [Perkal \(1956\)](#)). Define the quantity $\text{rconv}(A) := \sup\{r \in \mathbb{R}^+ : A_{\bullet r} = A\}$.

An equivalent definition of r -convexity is as follows: a set $A \subseteq \mathbb{R}^d$ is r -convex if and only if it can be expressed as the complement of a union of open balls of radius r .

Theorem 4.1. *Let A be closed in \mathbb{R}^d . It holds that*

$$\text{reach}(A) \leq \text{rconv}(A). \quad (4.2)$$

Moreover, if ∂A is a C^1 -smooth $(d - 1)$ -dimensional manifold without boundary, then

$$\text{reach}(A) = \text{rconv}(A). \quad (4.3)$$

Equation (4.2) is proven in [Cuevas et al. \(2012, Proposition 1\)](#) for compact A . Nonetheless, we reprove the statement for closed A in the proof of [Theorem 4.1](#), which we postpone to [Section 4.5](#). The novelty in [Theorem 4.1](#) is that it provides a class of subsets of \mathbb{R}^d for which the reach and r -convexity are equal. This class contains Serra's regular model ([Serra, 1984](#), p. 144) as a subclass, as shown by ([Walther, 1999, Theorem 1](#)).

To see that the r -convexity and the reach of a set are indeed distinct notions for general subsets of \mathbb{R}^d , [Figure 4.1](#) provides an example of a closed set A for which $\text{reach}(A) < \text{rconv}(A)$. This shows that the C^1 -smooth condition cannot be weakened to *piecewise* C^1 -smooth. Consider also the following remark.

Remark 4.1. Any closed subset of \mathbb{R}^d contained in a $(d - 1)$ -dimensional affine linear subspace is r -convex for all $r > 0$. This is easy to see since the complement of the set in the $(d - 1)$ -dimensional affine linear subspace is open, and is the union of open $(d - 1)$ -balls of radius less than r . Each $(d - 1)$ -ball can be expressed as the intersection of the affine linear subspace with an open ball of radius r in \mathbb{R}^d ; therefore, the closed subset in question can be expressed as the complement of a union of open balls of radius r and is hence r -convex.

Recently, a counterexample to Borsuk’s conjecture that r -convex sets are locally contractible, was published in [Cholaquidis \(2023\)](#). The conjecture is easily seen to be false by mapping a closed set in \mathbb{R}^{d-1} that is not locally contractible to a $(d - 1)$ -dimensional hyperplane in \mathbb{R}^d under an isometry.

Here, we provide some corollaries to [Theorem 4.1](#) that provide alternative sets of sufficient conditions for the equality of the reach and the r -convexity. The first applies to sets in Serra’s regular model ([Serra, 1984](#)).

Corollary 4.1. *Let $A \subset \mathbb{R}^d$ be non-empty, compact, and path-connected. If $\text{rconv}(\text{cl}(A^c)) > 0$, then $\text{reach}(A) = \text{rconv}(A)$.*

The proof of [Corollary 4.1](#), which relies heavily on [Theorem 1](#) in [Walther \(1999\)](#), is postponed to [Section 4.5](#). Likewise, we prove the following result in [Section 4.5](#).

Proposition 4.1. *Let A be a closed set in \mathbb{R}^d and let $\epsilon \in \mathbb{R}^+$. Then $\text{reach}(A_\epsilon) = \text{rconv}(A_\epsilon)$.*

The β -reach

We show in [Theorem 4.2](#) below, that the reach of a set can be formulated in terms of pairs of points in the set, and the distance to the set from their midpoints. We define a parametrized version of the reach by restricting to pairs of points whose midpoints are sufficiently far from the set. The so-called β -reach is constructed from the following family of functions.

Definition 4.5 (Spherical cap geometry). Define for $\alpha \in [0, \infty)$ and $x \in [0, \alpha/2]$,

$$g_\alpha(x) := \begin{cases} \frac{\alpha^2}{8x} + \frac{x}{2}, & x > 0, \\ \infty, & x = 0, \end{cases} \quad (4.4)$$

and its inverse for $r \geq \alpha/2$,

$$g_\alpha^{-1}(r) := r - \sqrt{r^2 - \frac{\alpha^2}{4}}. \quad (4.5)$$

See [Figure 4.3](#) for a geometric interpretation of the function in [\(4.4\)](#) and its inverse in [\(4.5\)](#). Evidently from the figure, this function is derived from the height of a spherical cap; it is written about in [Attali et al. \(2013\)](#); [Attali and Lieutier \(2015\)](#); [Berenfeld et al. \(2022\)](#); [Boissonnat et al. \(2019\)](#); [Divol \(2021\)](#) in reference to the reach.

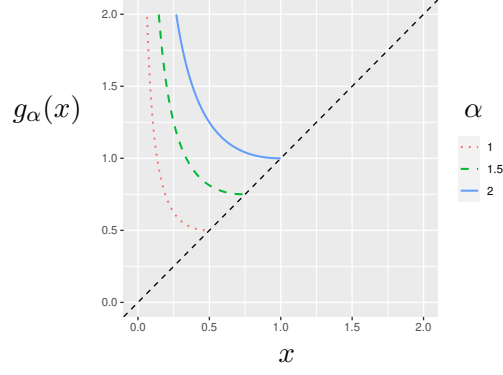


Figure 4.2: $g_\alpha(x)$ in Definition 4.5 plotted for several values of α .

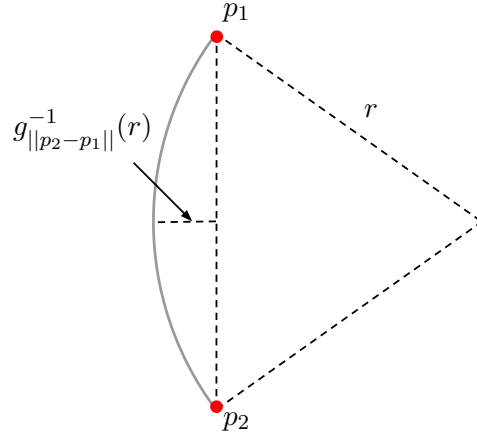


Figure 4.3: A geometric interpretation of $g_\alpha(x)$ in Definition 4.5.

Definition 4.6 (The β -reach). For a closed set $A \subseteq \mathbb{R}^d$ and $\beta \in [0, \infty)$, let the β -reach of A be defined as

$$\text{reach}_\beta(A) := \inf \left\{ g_{\|a_2 - a_1\|}(x) : a_1, a_2 \in A, x = \delta_A \left(\frac{a_1 + a_2}{2} \right) \geq \beta \right\},$$

where $g_{\|a_2 - a_1\|}(x)$ is defined in (4.4). Recall that $\delta_A : \mathbb{R}^d \rightarrow \mathbb{R}$ maps each point in \mathbb{R}^d to its distance from A . See Figure 4.4 for a visual aid.

By restricting to pairs of points $a_1, a_2 \in A$ in Figure 4.4 that yield $x \geq \beta$, the β -reach of A is the largest lower bound of the resulting values of $g_{\|a_2 - a_1\|}(x)$. If one does not restrict the size of x (i.e., for $\beta = 0$), then this largest lower bound is precisely $\text{reach}(A)$. This is formalized by the following theorem.

Theorem 4.2. *Let A be closed in \mathbb{R}^d . The map $\beta \mapsto \text{reach}_\beta(A)$ for $\beta \in \mathbb{R}^+$ is non-decreasing; moreover,*

$$\lim_{\beta \searrow 0} \text{reach}_\beta(A) = \text{reach}_0(A) = \text{reach}(A). \quad (4.6)$$

Proof of Theorem 4.2. In the following, we assume that A is not convex, for if it is convex, then all

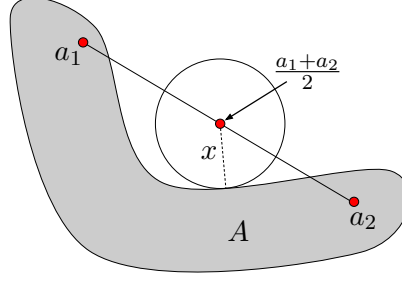


Figure 4.4: For $\beta \in [0, \infty)$, the β -reach of A (see Definition 4.6) is the largest lower bound of $g_{\|a_2-a_1\|}(x)$ (see Definition 4.5), for the pairs $a_1, a_2 \in A$ satisfying $x \geq \beta$.

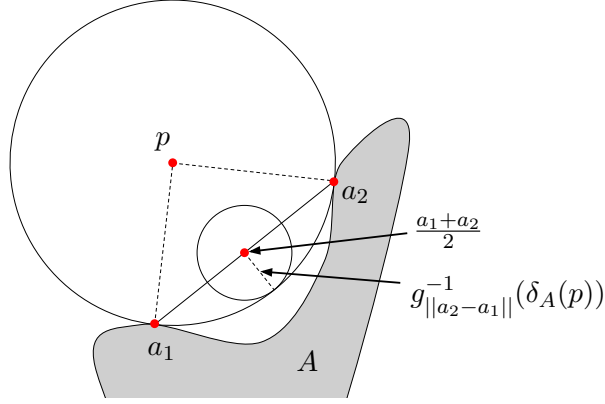


Figure 4.5: The interior of the large ball centered at p does not intersect A , so the distance from the midpoint $(a_1 + a_2)/2$ to the set A is at least $g_{\|a_2-a_1\|}^{-1}(\delta_A(p))$. This construction is used in the proof of Theorem 4.2.

of the quantities in (4.6) are infinite and the theorem holds (Federer, 1959). The non-decreasing property is seen immediately via the inclusion

$$\left\{ (a_1, a_2) \in A^2 : \delta_A \left(\frac{a_1 + a_2}{2} \right) \geq \beta_2 \right\} \subseteq \left\{ (a_1, a_2) \in A^2 : \delta_A \left(\frac{a_1 + a_2}{2} \right) \geq \beta_1 \right\},$$

for all $\beta_1, \beta_2 \in \mathbb{R}$ satisfying $\beta_1 < \beta_2$.

Now, we start by proving the second equality in (4.6). The proof is largely supplied by Lemma 1 of Boissonnat et al. (2019) which provides

$$\text{reach}(A) \leq g_{\|a_2-a_1\|} \circ \delta_A \left(\frac{a_1 + a_2}{2} \right), \quad (4.7)$$

for all $a_1, a_2 \in A$, since $g_\alpha(x)$ is non-increasing in x . What remains to show is that $\text{reach}(A)$ is the largest lower bound in (4.7); *i.e.*, if $r > \text{reach}(A)$ then there exist $a_1, a_2 \in A$ such that r exceeds the right-hand side of (4.7). Let $r > \text{reach}(A)$ and let $\tilde{r} \in (\text{reach}(A), r)$. By the definition of $\text{reach}(A)$ (Definition 4.2), $\exists p \in A_{\tilde{r}}$ and $a_1, a_2 \in A$ such that $\|a_1 - p\| = \|a_2 - p\| = \delta_A(p) \leq \tilde{r}$. Since the

interior of $B(p, \delta_A(p))$ does not intersect A , we have

$$\delta_A\left(\frac{a_1 + a_2}{2}\right) \geq g_{\|a_2 - a_1\|}^{-1}(\delta_A(p)),$$

(see Figure 4.5) and

$$g_{\|a_2 - a_1\|} \circ \delta_A\left(\frac{a_1 + a_2}{2}\right) \leq \delta_A(p) \leq \tilde{r} < r.$$

This proves that $\text{reach}_0(A) = \text{reach}(A)$.

Finally, we show the first equality in (4.6). The non-decreasing property gives $\lim_{\beta \searrow 0} \text{reach}_\beta(A) \geq \text{reach}_0(A)$. Now, it suffices to show the reverse inequality. For any $\epsilon > 0$, there exists $a_1, a_2 \in A$ satisfying

$$g_{\|a_2 - a_1\|} \circ \delta_A\left(\frac{a_1 + a_2}{2}\right) < \text{reach}_0(A) + \epsilon.$$

Any such pair (a_1, a_2) must satisfy $\delta_A\left(\frac{a_1 + a_2}{2}\right) > 0$, and so for $\beta \in (0, \delta_A\left(\frac{a_1 + a_2}{2}\right))$, it holds that $\text{reach}_\beta(A) < \text{reach}_0(A) + \epsilon$. Thus, $\lim_{\beta \searrow 0} \text{reach}_\beta(A) \leq \text{reach}_0(A)$. \square

Remark 4.2. For any $\alpha \in [0, \infty)$ and $x \in [0, \alpha/2]$, one has $g_\alpha(x) \geq x$ (see Figure 4.2). Thus, for any closed set A and $\beta \in \mathbb{R}^+$,

$$\text{reach}_\beta(A) \geq \beta.$$

Intuitively, this means that the β -reach excludes small, reach-limiting features of the set that have scale less than β .

The values of β for which $\text{reach}_\beta(A) = \beta$ are the distances from the critical points of the generalized gradient function (Chazal et al., 2009a) to the set A .

Remark 4.3. Other generalizations of the reach are constructed similarly to the β -reach, in that they formulate the reach as an infimum or supremum over some set, and add or remove elements in the set using some real parameter (in our case, β). For example, (Boissonnat et al., 2019, Theorem 1) expresses the reach of a set $A \subset \mathbb{R}^d$ as the supremum over a subset of \mathbb{R}^+ that satisfies a certain condition relating to A . By “continuously” weakening the condition with a parameter $\delta \in [0, \infty)$, the authors in Aamari et al. (2023) introduce the *spherical distortion radius* as the supremum of the larger subset of \mathbb{R}^+ satisfying the weaker condition parametrized by δ . The spherical distortion radius is identified with the reach for $\delta = 0$. In an earlier work, Chazal et al. (2009a) introduces the μ -reach by considering the shortest distance from an element in the set to the μ -medial axis, a filtered version of the medial axis by considering regions where the generalized gradient function (Lieutier, 2004) is less than some $\mu \in (0, 1]$. The β -reach, the μ -reach, and the spherical distortion radius, are all generalizations of the reach that exclude “small-scale” features as decided by the corresponding parameter β , μ , or δ .

One significant advantage of the β -reach is its computability for high-dimensional point cloud data (see Section 4.3.3). In this setting, the formulation of the reach in terms of the β -reach for

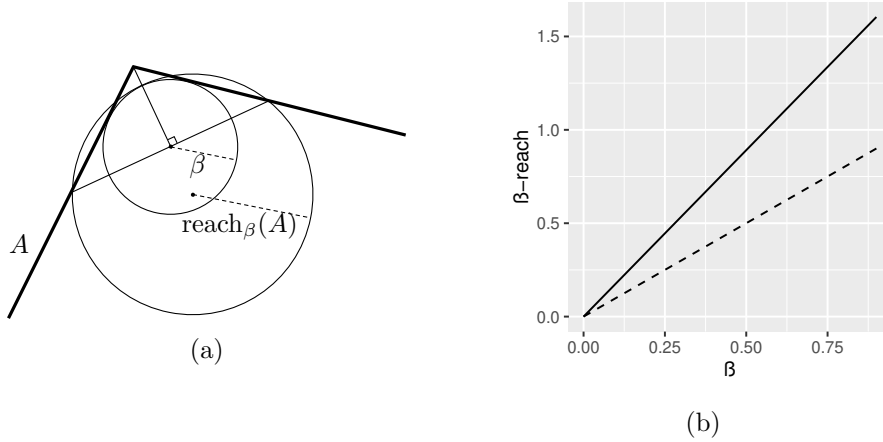


Figure 4.6: **(a)**: The construction of $\text{reach}_\beta(A)$ for $\beta > 0$ with A , the union of two rays in \mathbb{R}^d that originate from the same point. **(b)**: The β -reach profile of A shown as a solid line. The lower bound with slope 1 is shown as a dashed line.

$\beta = 0$ can also be used to construct an upper bound for the reach of any compact $A \subset \mathbb{R}^d$ (see Section 4.3.2).

For a closed set $A \subseteq \mathbb{R}^d$, one can study the map $\beta \mapsto \text{reach}_\beta(A)$ for $\beta \in [0, \infty)$, which we refer to as the β -reach profile of the set A . We will see in Section 4.3 that a set's β -reach profile provides pertinent information regarding the estimation of the reach from point cloud data, especially through its first-order approximation at $\beta = 0$.

Some exemplary sets and their β -reach profiles are considered in the several examples that follow.

Example 4.1 (The β -reach of a corner). For two line segments in \mathbb{R}^d with ends joined by an angle $\theta \in (0, \pi)$, the β -reach of their union is $\frac{\beta}{2} (1 + \sec^2(\frac{\theta}{2}))$, for sufficiently small $\beta > 0$. See Figure 4.6 for an illustration.

Example 4.2 (The β -reach of an arc). The β -reach of a circular arc of angle at most π is equal to the radius of the arc, for sufficiently small $\beta > 0$.

Example 4.3 (The β -reach of a bottleneck structure). If the reach of a set $A \subset \mathbb{R}^d$ is determined by a bottleneck structure like those described in Aamari et al. (2019), then $\text{reach}_\beta(A) = \text{reach}(A)$ for $\beta \in [0, \text{reach}(A)]$. This is the case, for example, when A is a finite set of points in \mathbb{R}^d .

Example 4.4 (The β -reach of a C^2 -smooth curve). Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a C^2 -smooth function with $h'(0) > 0$. Suppose that the graph of the function $f : [-1, 1] \rightarrow \mathbb{R}$ defined by $f(x) := h(x^2)$ obtains its maximal curvature at $x = 0$. Then, the graph $A := \{(x, f(x)) : -1 \leq x \leq 1\} \subset \mathbb{R}^2$ satisfies

$$\text{reach}(A) = \frac{1}{2h'(0)} \tag{4.8}$$

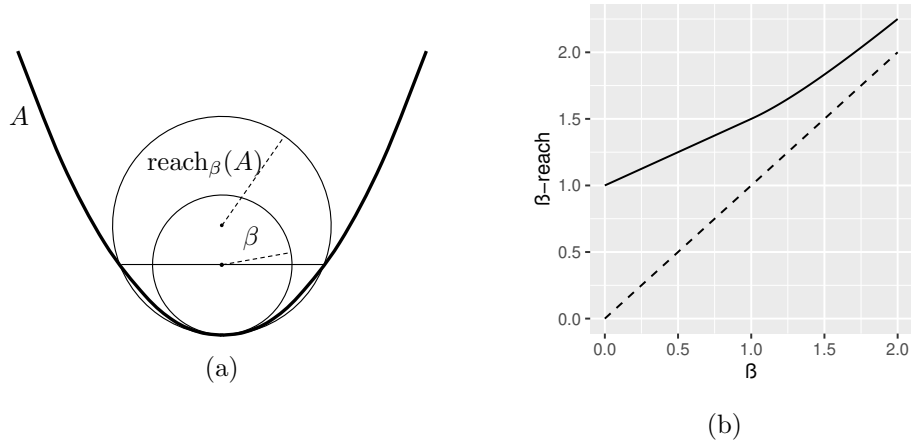


Figure 4.7: **(a)**: The construction of $\text{reach}_\beta(A)$ for $\beta > 0$ with A , a paraboloid embedded in \mathbb{R}^d . **(b)**: The β -reach profile of A shown as a solid line. The lower bound with slope 1 is shown as a dashed line.

and

$$\text{reach}_\beta(A) - \text{reach}(A) \leq \left(\frac{1}{2} - \frac{h''(0)}{4h'(0)^3} \right) \beta + o(\beta), \quad (4.9)$$

for sufficiently small $\beta > 0$. Proofs for Equations (4.8) and (4.9) are provided at the end of Section 4.5. Figure 4.7 depicts a special case of this example with $h(x) = x/2$.

4.3 Methods for point cloud data

In practice, one might be interested in identifying bounds on the reach and r -convexity of a set $A \subseteq \mathbb{R}^d$ from a discrete sample of A without noise. The goal of this section is to provide computational methods for bounding the reach and the r -convexity from above, and to produce methods for approximating these quantities. The three main settings that we consider are:

- (a) One has access to a point cloud that extends over \mathbb{R}^d in the sense that \mathbb{R}^d can be covered by balls of fixed radius ϵ centered at each point. Moreover, one knows the partition of the points that lie in $A \subseteq \mathbb{R}^d$, and those that lie in A^c . See Section 4.3.1 for a treatment of this setting.
- (b) One has access to a set of points for which it is known that $A \subseteq \mathbb{R}^d$ can be covered by balls of fixed radius ϵ centered at each point. Here, ϵ is known. See Section 4.3.2.
- (c) The set A is a submanifold of \mathbb{R}^d of dimension $m < d$. One has access to a set of points that is known to be contained in A . See Section 4.3.3.

The mathematical results that we present in this section hold in arbitrary dimension d . Nevertheless, the computational complexity of the method described in Section 4.3.1 for setting (a) increases quite drastically as higher dimensions are considered (see Remark 4.6).

The method that we present for setting (b), described in Section 4.3.2, involves computing a nearest neighbor search for each midpoint of pairs of points in the dataset. This can be achieved in $\mathcal{O}(dN^3)$ time by naively searching each of the N points for the nearest neighbor to each of the $\mathcal{O}(N^2)$ midpoints. Improvements to the nearest neighbor search using binary space partitioning data structures such as k - d trees (Bentley, 1975) or ball trees (Omohundro, 1989; Liu et al., 2006) improve the time complexity of the overall strategy to $\mathcal{O}(N^2 \log N)$ if d is kept fixed. This can be related to the desired precision ϵ , since the number of points N necessary to cover A with balls of size ϵ is $\Omega(\epsilon^{-m})$, where m is the Hausdorff dimension of A .

Calculating the β -reach profile of the manifold in setting (c) has the same time complexity as the method for setting (b). Thus, an isometric embedding of a manifold of fixed dimension m into \mathbb{R}^d affects the computational complexity of the method linearly in d . Contrast this with existing methods that aim to approximate the reach by first approximating the medial axis Cuevas et al. (2014); Chazal and Lieutier (2005a); Chazal et al. (2009a); Dey and Zhao (2003); Dey and Sun (2006); Lieutier and Wintraecken (2023), where in high dimension, accessing the medial axis becomes computationally challenging. The relationship between d , m , n , and our method’s performance is elaborated on in Remark 4.10 in Section 4.3.3.

4.3.1 An upper bound for the r -convexity of a set

In this section, we introduce a method for identifying when sets are not r -convex, for $r > 0$, with a true negative rate of 100%. Given a test radius $r > 0$, and points labeled as either residing in A or A^c , our method identifies if $r > \text{rconv}(A)$ with perfect specificity. This establishes an upper bound for $\text{rconv}(A)$, and by Equation (4.2), an upper bound for $\text{reach}(A)$. Moreover, as more samples are included to decrease the Hausdorff distances to the sets A and A^c , we show that this upper bound converges to $\text{rconv}(A)$.

Here, we define operations analogous to dilation and erosion, for discrete sets of points.

Definition 4.7. Let \mathcal{P} be a point cloud in \mathbb{R}^d , *i.e.*, a countable subset of \mathbb{R}^d . For a set $\hat{A} \subseteq \mathcal{P}$, we make a slight abuse of notation and denote for $r \in \mathbb{R}$,

$$\hat{A}_r := \begin{cases} \{p \in \mathcal{P} : \delta_{\hat{A}}(p) \leq r\}, & \text{for } r \geq 0, \\ \{a \in \hat{A} : \delta_{\mathcal{P} \setminus \hat{A}}(a) > -r\}, & \text{for } r < 0. \end{cases}$$

Remark 4.4. We emphasize that for a point cloud \mathcal{P} , a subset $\hat{A} \subseteq \mathcal{P}$, and a real number $r \in \mathbb{R}$, Definition 4.7 implies that $\hat{A}_r \subseteq \mathcal{P}$. Contrast this with the set $\hat{A} \oplus B(0, |r|)$, which is not contained in \mathcal{P} for $r > 0$.

Let us illustrate the operations in Definition 4.7 using the example in Figure 4.8. The set $A \subset \mathbb{R}^2$ occupies the bottom left corner of the domain, just until the dashed line. When sampled on a square lattice \mathcal{P} , one obtains the image in panel (a); each pixel is centered on a point in \mathcal{P} , and

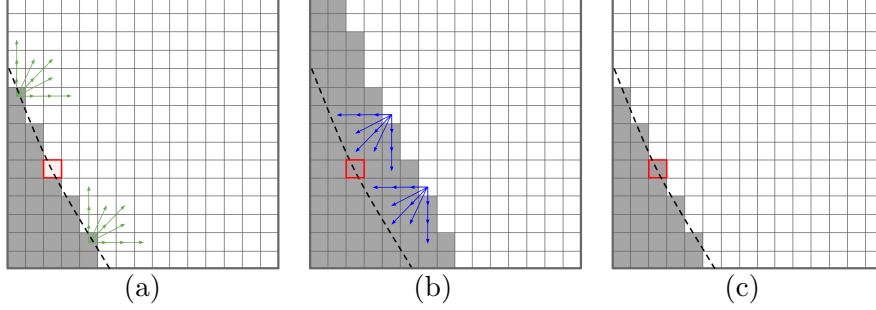


Figure 4.8: An example where $(\widehat{A}_r)_{-r}$ in (c) is strictly larger than \widehat{A} in (a), while the underlying set A is r -convex. The boundary ∂A is shown as a dashed line and r is three times the pixel width. The point cloud Φ is a square lattice of points, and pixels are centered on the elements of Φ .

grey if the point lies in A . Dilating the grey pixels \widehat{A} by $r = 3 \times (\text{pixel width})$, one obtains the grey pixels in panel (b), \widehat{A}_r . If one then erodes \widehat{A}_r by r (equivalent to dilating $\mathcal{P} \setminus \widehat{A}_r$ by r and taking complements), one obtains the grey pixels in panel (c), $(\widehat{A}_r)_{-r}$. Notice, however, that there is a point in $(\widehat{A}_r)_{-r}$ that is not in \widehat{A} . This might be surprising since we chose a set A such that $A_{\bullet r} = A$. This illustration shows that naively testing for r -convexity using the discrete dilation and erosion operations *mistakenly classifies sets as not r -convex*, when indeed they are.

The following theorem shows how to correctly identify sets as not being r -convex, and gives an upper bound for the r -convexity of a set that is tight in some sense.

Theorem 4.3. For $n \in \mathbb{N}^+$, let $\mathcal{P}^{(n)}$ be a point cloud in \mathbb{R}^d , and suppose that

$$\epsilon_n := \sup \left\{ \delta_{\mathcal{P}^{(n)}}(q) : q \in \mathbb{R}^d \right\} \quad (4.10)$$

is finite. Let A be a compact subset of \mathbb{R}^d , and for $n \in \mathbb{N}^+$, define $\widehat{A}^{(n)} := A \cap \mathcal{P}^{(n)}$ and the corresponding bound,

$$\widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) := \inf \left\{ r > \epsilon_n : (\widehat{A}_{r-\epsilon_n}^{(n)})_{-(r+\epsilon_n)} \not\subseteq \widehat{A}^{(n)} \right\}. \quad (4.11)$$

It holds that

$$\inf_{n \in \mathbb{N}^+} \widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) \geq \text{rconv}(A). \quad (4.12)$$

Moreover, if $\epsilon_n \rightarrow 0$ and $d_H(\widehat{A}^{(n)}, A) \rightarrow 0$ as $n \rightarrow \infty$, then

$$\lim_{n \rightarrow \infty} \widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) = \text{rconv}(A). \quad (4.13)$$

Theorem 4.3 provides a method for correctly identifying which subsets of \mathbb{R}^d are not r -convex. If a dilation of the the discretization $\widehat{A}^{(n)}$ by $r - \epsilon_n$ followed by an erosion of $r + \epsilon_n$ produces a set that is not contained in $\widehat{A}^{(n)}$, then $\text{rconv}(A) \leq r$. See Figure 4.9 for an example of a set A for which $(\widehat{A}_{r-\epsilon_n}^{(n)})_{-(r+\epsilon_n)} \not\subseteq \widehat{A}^{(n)}$. The precise regions where r -convexity does not hold are highlighted by the

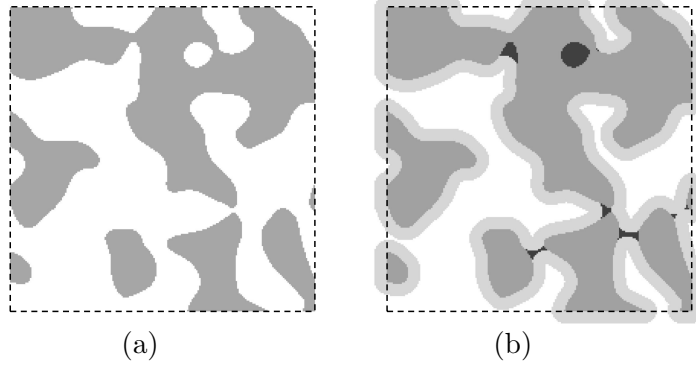


Figure 4.9: **(a):** A set A is shown as a pixelated image in grey. The set of grey pixel centers is $\widehat{A}^{(n)}$. **(b):** The set $(\widehat{A}_{r-\epsilon_n}^{(n)})_{-(r+\epsilon_n)}$ contains the dark grey pixels, which are not contained in $\widehat{A}^{(n)}$. Thus, $r \geq \text{rconv}(A)$. Here, r is nine times the pixel width.

method.

Remark 4.5. Note that $\widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)})$ in (4.11) can be computed using entirely available information, since ϵ_n in (4.10) is a feature of the point cloud $\mathcal{P}^{(n)}$ and not of the unknown set A . In particular, there is no need to estimate $d_H(\widehat{A}^{(n)}, A)$ for the construction of $\widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)})$. A binary search algorithm, along with the discrete dilation operations in Definition 4.7, are sufficient for the numerical calculation of $\widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)})$.

Remark 4.6. The requirement that $\mathcal{P}^{(n)}$ extends over all of \mathbb{R}^d (see Equation (4.10)) allows for a mathematical simplification and is not needed in practice. In real applications, one only requires that $d_H(\mathcal{P}^{(n)}, T) = \epsilon_n$ for some compact $T \subset \mathbb{R}^d$ that contains $A_{\text{rconv}(A)+2\epsilon_n}$. In the case where A is lower dimensional than the embedding space \mathbb{R}^d , one can add any finite number of points uniformly distributed on T to $\mathcal{P}^{(n)}$, and they will almost surely not belong to A . However, the number of points needed for a given precision ϵ_n scales like ϵ_n^{-d} times the Lebesgue measure of T , which is computationally expensive in high dimension.

The proof of Theorem 4.3 is quite technical, and so we postpone it to Section 4.5. Nevertheless, Figure 4.10 provides an intuitive illustration that helps to understand Equation (4.12). This is an example of a worst-case scenario, in that a dilation of $\widehat{A}^{(n)}$ by any more than $\text{rconv}(A) - \epsilon_n$ results in all of $\mathcal{P}^{(n)}$ being consumed. The dilation radius $\text{rconv}(A) - \epsilon_n$ is maximal in the sense that, by dilating any less, there is guaranteed to be at least one element of $\mathcal{P}^{(n)}$ that is not in the dilation. The distance between this one remaining element of $\mathcal{P}^{(n)}$ and the other points in $(\mathcal{P}^{(n)} \setminus \widehat{A}^{(n)}) \cap \widehat{A}_{r-\epsilon_n}^{(n)}$ might be anywhere up to $r + \epsilon_n$, where $r - \epsilon_n \in (0, \text{rconv}(A) - \epsilon_n)$ is the dilation radius. Therefore, an erosion by at least $r + \epsilon_n$ is necessary to guarantee that the resulting set is a subset of $\widehat{A}^{(n)}$. In this sense, the bound in (4.12) is tight.

The following result from Rodríguez Casal and Saavedra-Nieves (2016), while interesting on its own, is instrumental in the proof of Equation (4.13).

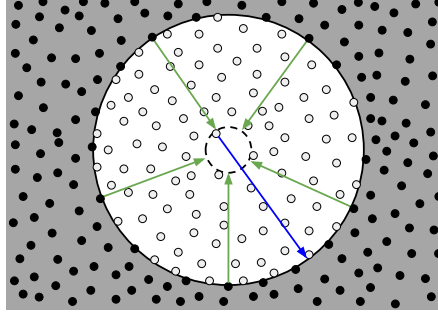


Figure 4.10: A discrete dilation of the set $\widehat{A}^{(n)}$ (black points) by the length of the green arrows ($\text{rconv}(A) - \epsilon_n$) followed by a discrete erosion by the length of the longer blue arrow ($\text{rconv}(A) + \epsilon_n$) leaves all of the white points out of the resulting set. The boundary of $A_{\text{rconv}(A) - \epsilon_n}$ is shown as a dashed line.

Proposition 4.2 (Lemma 8.3 in [Rodríguez Casal and Saavedra-Nieves \(2016\)](#)). *Let $r \in \mathbb{R}^+$ and let $A \subset \mathbb{R}^d$ be a closed set satisfying $\text{rconv}(A) < r$. The set $A_{\bullet, r} \setminus A$ contains an open subset of \mathbb{R}^d .*

The relationship between Proposition 4.2 and Equation (4.13) is that, for $r > \text{rconv}(A)$, as the granularity of the point cloud $\mathcal{P}^{(n)}$ increases, the open subset in $A_{\bullet, r} \setminus A$ fills with points that remain in $(\widehat{A}_{r - \epsilon_n}^{(n)})_{-(r + \epsilon_n)}$ for ϵ_n sufficiently small. Thus,

$$\lim_{n \rightarrow \infty} \widehat{\text{r}}^{(\epsilon_n)}(\widehat{A}^{(n)}) < r,$$

and since $r \in (\text{rconv}(A), \infty)$ is arbitrary, the limit is at most $\text{rconv}(A)$. Equality then holds by Equation (4.12). The full proof of Theorem 4.3 and an alternative proof of Proposition 4.2 are provided in Section 4.5.

Remark 4.7. Equation (4.13) is provided without a rate of convergence. This is due to the fact that Proposition 4.2 provides no guarantees on the size of the open subset that can be found in $A_{\bullet, r} \setminus A$ for $r > \text{rconv}(A)$. To provide a rate of convergence, a deeper analysis is needed to understand the rate at which the size of the largest open ball in $A_{\bullet, r} \setminus A$ decreases as $r \searrow \text{rconv}(A)$. Conversely, the bound that we construct for the reach in Section 4.3.2 converges to the reach at an explicit rate (see Theorem 4.4 below).

4.3.2 An upper bound for the reach of a set

We have already seen that $\widehat{\text{r}}^{(\epsilon_n)}(\widehat{A}^{(n)})$ in (4.11) is an upper bound for the reach by Equations (4.2) and (4.12). In this section, we introduce another computable bound for the reach based on the expression for the β -reach in Definition 4.6. The context in which one can apply this bound was introduced as setting (b) at the start of Section 4.3: a countable set of points $\widehat{A}^{(n)}$ is known to be included in a compact set $A \subset \mathbb{R}^d$, and its Hausdorff distance to A is known to be at most ϵ_n .

A regularity condition is imposed on the set A in terms of its β -reach profile near 0; it is used

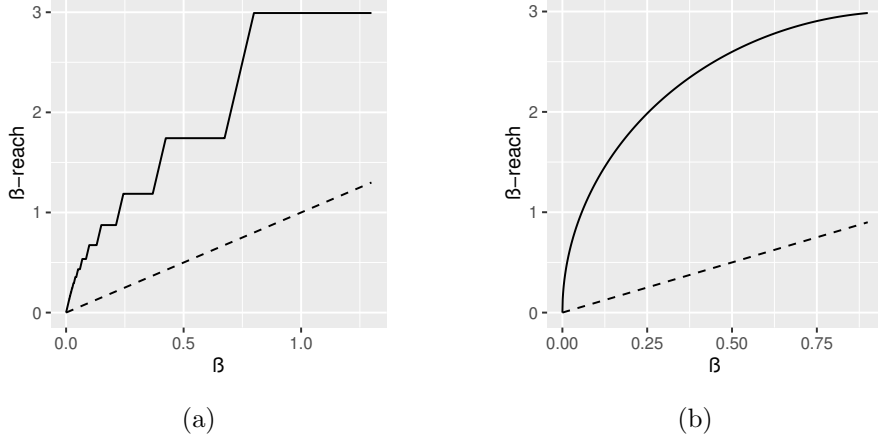


Figure 4.11: Hypothetical β -reach profiles of sets that would *not* satisfy Assumption 4.1. **(a):** The β -reach profile is neither constant nor strictly increasing on any neighborhood of 0. **(b):** The limit in (4.14) is infinite.

to prove the rate at which our upper bound converges to $\text{reach}(A)$ as $\widehat{A}^{(n)} \rightarrow A$ in the Hausdorff metric.

Assumption 4.1. *Suppose that the set $A \subset \mathbb{R}^d$ is compact, not convex, and that there exists $\delta > 0$ such that the map $\beta \mapsto \text{reach}_\beta(A)$ for $\beta \in \mathbb{R}$ is either constant or strictly increasing on $[0, \delta]$. In addition, suppose that*

$$K_A := \lim_{\beta \searrow 0} \frac{\text{reach}_\beta(A) - \text{reach}(A)}{\beta} \quad (4.14)$$

exists and is finite.

Assumption 4.1 allows for most compact, non-convex sets in \mathbb{R}^d , excluding some pathological counterexamples such as hypothetical sets admitting the β -reach profiles in Figure 4.11, or the counterexample constructed in Cholaquidis (2023). The behaviour of the β -reach profile near 0 is entirely determined by features of negligible size, so Assumption 4.1 will hold for any compact, non-convex set whose boundary is *locally* well-behaved. Assumption 1 is seen to hold in all of Examples 4.1, 4.2, and 4.3 in Section 4.2.2, which shows that smoothness is not a necessary condition.

Theorem 4.4. *Let $A \subseteq \mathbb{R}^d$ be closed. For each $n \in \mathbb{N}^+$, let $\widehat{A}^{(n)}$ be a countable subset of A . Suppose that $\widehat{A}^{(n)} \rightarrow A$ in the Hausdorff metric. i.e., there is a sequence $(\epsilon_n)_{n \geq 1}$ in \mathbb{R}^+ tending to 0 as $n \rightarrow \infty$ such that $\epsilon_n \geq d_H(\widehat{A}^{(n)}, A)$ for all $n \in \mathbb{N}^+$. For each $n \in \mathbb{N}^+$, define the corresponding bound on $\text{reach}(A)$ by*

$$\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) := \inf \left\{ g_{\|a_2 - a_1\|}(x - \epsilon_n) : a_1, a_2 \in \widehat{A}^{(n)}, \right. \quad (4.15)$$

$$\left. x = \delta_{\widehat{A}^{(n)}} \left(\frac{a_1 + a_2}{2} \right) \geq \epsilon_n \right\},$$

where $g_{\|a_2-a_1\|}$ is defined in (4.4). Then,

$$\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) \geq \text{reach}(A), \quad (4.16)$$

for $n \in \mathbb{N}^+$. Furthermore, if A satisfies Assumption 4.1, then

$$\limsup_{n \rightarrow \infty} \frac{\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) - \text{reach}(A)}{\sqrt{\epsilon_n}} \leq (2 + K_A)\sqrt{\text{reach}(A)}, \quad (4.17)$$

where K_A is defined as in (4.14).

Proof of Theorem 4.4. If A is convex, then for all $a_1, a_2 \in A$,

$$\delta_{\widehat{A}^{(n)}}\left(\frac{a_1 + a_2}{2}\right) \leq d_H(\widehat{A}^{(n)}, A) \leq \epsilon_n,$$

and so both sides of the inequality in (4.16) are infinite (Federer, 1959). Now, suppose that A is not convex. Let $n \in \mathbb{N}^+$, and write,

$$\begin{aligned} \text{reach}_0(A) &= \inf \left\{ g_{\|a_2-a_1\|} \circ \delta_A\left(\frac{a_1 + a_2}{2}\right) : a_1, a_2 \in A \right\} \\ &\leq \inf \left\{ g_{\|a_2-a_1\|} \circ \delta_A\left(\frac{a_1 + a_2}{2}\right) : a_1, a_2 \in \widehat{A}^{(n)} \right. \\ &\quad \left. \delta_{\widehat{A}^{(n)}}\left(\frac{a_1 + a_2}{2}\right) \geq \epsilon_n \right\}. \end{aligned} \quad (4.18)$$

The equality in (4.18) is an application of Definition 4.6, and the inequality holds since the infimum is taken over a smaller subset. For any $a_1, a_2 \in A$, there exists a projection of $\frac{a_1+a_2}{2}$ onto A , namely $a^\pi \in A$, satisfying

$$\left\| a^\pi - \frac{a_1 + a_2}{2} \right\| = \delta_A\left(\frac{a_1 + a_2}{2}\right).$$

By the triangle inequality and the fact that $\epsilon_n \geq d_H(\widehat{A}^{(n)}, A)$,

$$\delta_{\widehat{A}^{(n)}}\left(\frac{a_1 + a_2}{2}\right) \leq \delta_{\widehat{A}^{(n)}}(a^\pi) + \left\| a^\pi - \frac{a_1 + a_2}{2} \right\| \leq \epsilon_n + \delta_A\left(\frac{a_1 + a_2}{2}\right).$$

Rearranging gives,

$$\delta_{\widehat{A}^{(n)}}\left(\frac{a_1 + a_2}{2}\right) - \epsilon_n \leq \delta_A\left(\frac{a_1 + a_2}{2}\right). \quad (4.19)$$

Since $g_{\|a_2-a_1\|}$ is non-increasing, the rightmost expression in (4.18) cannot decrease when $\delta_A\left(\frac{a_1+a_2}{2}\right)$ is replaced by $\delta_{\widehat{A}^{(n)}}\left(\frac{a_1+a_2}{2}\right) - \epsilon_n$. Thus,

$$\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) \geq \text{reach}_0(A) = \text{reach}(A),$$

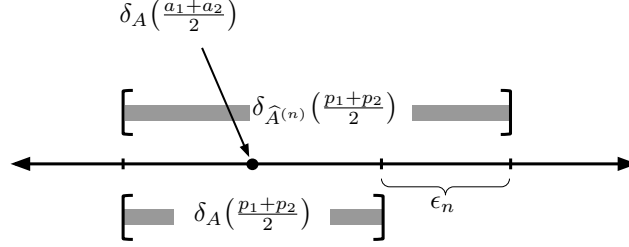


Figure 4.12: Visual aid for Equation (4.21). The real numbers $\delta_{\widehat{A}^{(n)}}\left(\frac{p_1+p_2}{2}\right)$ and $\delta_A\left(\frac{p_1+p_2}{2}\right)$ are contained in their respective intervals, positioned relative to $\delta_A\left(\frac{a_1+a_2}{2}\right)$. Ticks have a spacing of ϵ_n .

where the latter equality is an application of Theorem 4.2.

Now, we proceed to show (4.17). Recall the equality in (4.18), and split the analysis into two cases.

Case 1: *There exists $a_1, a_2 \in A$ such that $a_1 \neq a_2$ and $\text{reach}(A) = g_{\|a_2-a_1\|} \circ \delta_A\left(\frac{a_1+a_2}{2}\right)$.*

For this pair $a_1, a_2 \in A$ there exists $p_1, p_2 \in \widehat{A}^{(n)}$ satisfying $\|p_2 - a_2\|, \|p_1 - a_1\| \leq \epsilon_n$. This implies

$$\left\| \frac{p_1 + p_2}{2} - \frac{a_1 + a_2}{2} \right\| \leq \epsilon_n. \quad (4.20)$$

Equation (4.20) tells us that the two midpoints are close, and so their distances to A cannot differ by more than the distance between them. *i.e.*,

$$\left| \delta_A\left(\frac{a_1 + a_2}{2}\right) - \delta_A\left(\frac{p_1 + p_2}{2}\right) \right| \leq \epsilon_n.$$

Since Equation (4.19) holds for $p_1, p_2 \in A$, and $\delta_A\left(\frac{p_1+p_2}{2}\right) \leq \delta_{\widehat{A}^{(n)}}\left(\frac{p_1+p_2}{2}\right)$, we also have

$$\delta_A\left(\frac{a_1 + a_2}{2}\right) - \epsilon_n \leq \delta_{\widehat{A}^{(n)}}\left(\frac{p_1 + p_2}{2}\right) \leq \delta_A\left(\frac{a_1 + a_2}{2}\right) + 2\epsilon_n, \quad (4.21)$$

by the triangle inequality (see Figure 4.12).

Now, suppose that n is sufficiently large such that $2\epsilon_n < y := \delta_A\left(\frac{a_1+a_2}{2}\right)$. Then, by (4.21), $x := \delta_{\widehat{A}^{(n)}}\left(\frac{p_1+p_2}{2}\right) - \epsilon_n > 0$ and so

$$\begin{aligned} \widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) &\leq g_{\|p_2-p_1\|}(x) = \frac{\|p_2 - p_1\|^2}{8x} + \frac{x}{2} \\ &\leq \frac{(\|a_2 - a_1\| + 2\epsilon_n)^2}{8x} + \frac{x}{2} \\ &\leq \frac{(\|a_2 - a_1\| + 2\epsilon_n)^2}{8(y - 2\epsilon_n)} + \frac{y + \epsilon_n}{2}. \end{aligned} \quad (4.22)$$

The second inequality in (4.22) holds since $\|p_2 - p_1\| \leq \|a_2 - a_1\| + 2\epsilon_n$, and the final inequality holds by Equation (4.21).

Recall that $\text{reach}(A) = \frac{\|a_2 - a_1\|^2}{8y} + \frac{y}{2}$. The difference between this and the final expression in (4.22) is of the order $\mathcal{O}(\epsilon_n)$, which is stronger than required.

Case 2: $\text{reach}(A) < g_{\|a_2 - a_1\|} \circ \delta_A \left(\frac{a_1 + a_2}{2} \right)$ for all distinct $a_1, a_2 \in A$.

By Assumption 4.1, for all sufficiently large n , the function $\text{reach}_\beta(A)$ is strictly increasing for β in a neighborhood of $\beta_n := \sqrt{\text{reach}(A)\epsilon_n} + 2\epsilon_n$. For each of these values of n , define

$$\Psi_n := \{(a_1, a_2) \in A^2 : \delta_A \left(\frac{a_1 + a_2}{2} \right) \geq \beta_n\},$$

which is compact by Assumption 4.1, and the continuous map $f_n : \Psi_n \rightarrow \mathbb{R}^+$ by

$$f_n((\psi_1, \psi_2)) = g_{\|\psi_2 - \psi_1\|} \circ \delta_A \left(\frac{\psi_1 + \psi_2}{2} \right).$$

Remark that $\text{reach}_{\beta_n}(A) = \inf_{\psi \in \Psi_n} f_n(\psi)$, and so there is an element $\psi_n^* := (\psi_{1n}^*, \psi_{2n}^*) \in \Psi_n$ for which

$$\text{reach}_{\beta_n}(A) = f_n(\psi_n^*).$$

By Definition 4.6, $\text{reach}_\beta(A) \leq f_n(\psi_n^*)$ for $\beta = \delta_A \left(\frac{\psi_{1n}^* + \psi_{2n}^*}{2} \right)$. The β -reach profile of A is thus constant on the interval $[\beta_n, \delta_A \left(\frac{\psi_{1n}^* + \psi_{2n}^*}{2} \right)]$ by Theorem 4.2, but by Assumption 4.1, any such interval must have length 0. Therefore,

$$\beta_n = \delta_A \left(\frac{\psi_{1n}^* + \psi_{2n}^*}{2} \right).$$

Again by Assumption 4.1,

$$\text{reach}_{\beta_n}(A) = \frac{\|\psi_{2n}^* - \psi_{1n}^*\|^2}{8\beta_n} + \frac{\beta_n}{2} = \text{reach}(A) + K_A\beta_n + o(\beta_n).$$

Therefore,

$$\begin{aligned} \|\psi_{2n}^* - \psi_{1n}^*\|^2 &= 8\beta_n \text{reach}(A) + (8K_A - 4)\beta_n^2 + o(\beta_n^2) \\ &= 8\sqrt{\epsilon_n} (\text{reach}(A))^{3/2} + (16 + 8K_A - 4) \text{reach}(A)\epsilon_n + o(\epsilon_n). \end{aligned}$$

As seen in Case 1, Equation (4.22),

$$\begin{aligned} \widehat{\text{rch}}^{(\epsilon_n)}(\widehat{A}^{(n)}) &\leq \frac{(\|\psi_{2n}^* - \psi_{1n}^*\| + 2\epsilon_n)^2}{8(\beta_n - 2\epsilon_n)} + \frac{\beta_n + \epsilon_n}{2} \\ &= \text{reach}(A) + (2 + K_A) \sqrt{\text{reach}(A)}\sqrt{\epsilon_n} + o(\sqrt{\epsilon_n}). \end{aligned}$$

□

Remark 4.8. The split of the proof of Theorem 4.4 into two cases is natural in the context of reach estimation. Recall Theorem 3.4 in Aamari et al. (2019): For a compact set A with $0 < \text{reach}(A) < \infty$, there is either a bottleneck structure (two distinct points $a_1, a_2 \in A$ such that $\|a_2 - a_1\| = 2\delta_A\left(\frac{a_1+a_2}{2}\right) = 2\text{reach}(A)$) or an arc-length parametrized geodesic of A with curvature $1/\text{reach}(A)$. The convergence rate developed in Case 1 applies to sets whose reach is decided by a bottleneck structure, and that of Case 2 applies to sets whose reach is decided by a region of high curvature.

Remark 4.9. In Aamari et al. (2023), the authors provide a minimax optimal rates of convergence for statistical estimators of the reach. The authors proceed by assuming that $A \subset \mathbb{R}^d$ is a C^k -smooth manifold without boundary, with $k \geq 3$. The minimax rates in (Aamari et al., 2023, Theorem 6.6) are expressed in terms of k , the number of continuous derivatives of the underlying manifold.

It is not surprising that our bound on the reach—for sets satisfying the very general Assumption 4.1—does not converge to the reach at the optimal rates given in Aamari et al. (2023). Nevertheless, we do observe a similar phenomenon in that the rate of convergence derived for a set whose reach is determined by a bottleneck structure ($\mathcal{O}(\epsilon_n)$ in Case 1 in the proof of Theorem 4.4) is faster than the rate when the reach is determined by a region of high curvature ($\mathcal{O}(\sqrt{\epsilon_n})$ in Case 2).

4.3.3 The β -reach profiles of high-dimensional point clouds

Here, we tackle the problem of approximating the reach of smooth manifolds of low dimension embedded in high-dimensional Euclidean space. As discussed in the introduction, this setting is the focus of many works. Given a C^1 -smooth manifold $M \subset \mathbb{R}^d$ of dimension $m < d$, we suppose that one has access to a set of points $\widehat{M} \subset M$ from which one would like to infer $\text{reach}(M)$.

The β -reach profile of a point cloud can be obtained by considering pairs of points in the point cloud and computing the distance from their midpoint to the closest neighbor in the point cloud. Of course, the β -reach profile of \widehat{M} is not identical to that of M , however, we argue here that it provides a good approximation for the values of β away from 0.

Recall that the bounds developed in Sections 4.3.1 and 4.3.2 are for subsets of \mathbb{R}^d that possibly have positive d -volume. This means that a point cloud \widehat{A} contained in a set $A \subset \mathbb{R}^d$ can recede a distance $\epsilon := d_H(\widehat{A}, A)$ from the topological boundary of A . Thus, the distance from a point in A^c to A can be up to ϵ shorter than the distance to the nearest point in \widehat{A} . Conversely, for a smooth manifold M embedded in \mathbb{R}^d , projection of an element $p \in \mathbb{R}^d$ onto M is along a vector normal to M at the projection in M . Thus, the distance from p to its nearest point in \widehat{M} is more similar to the length of the projection vector, so in this setting, the bounds constructed previously are overly robust. Moreover, existing algorithms for computing a mesh from point cloud data can be leveraged to improve quality of the estimate of the distance to the nearest point in M , thus improving the estimate of the β -reach.

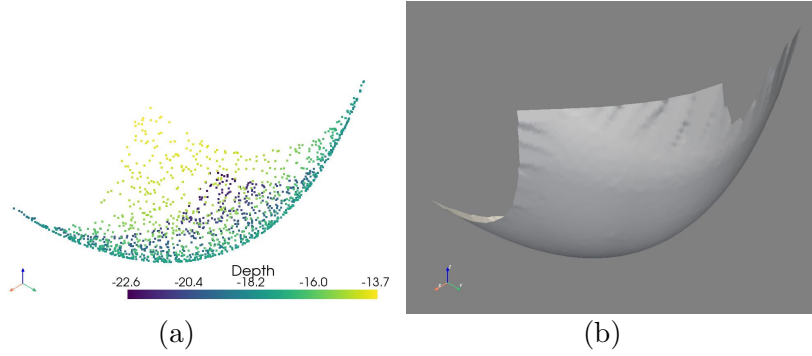


Figure 4.13: (a): The point cloud \widehat{M} in Example 4.5. (b): The mesh $\mathfrak{M}(\widehat{M})$ computed with the Python library `pyvista`.

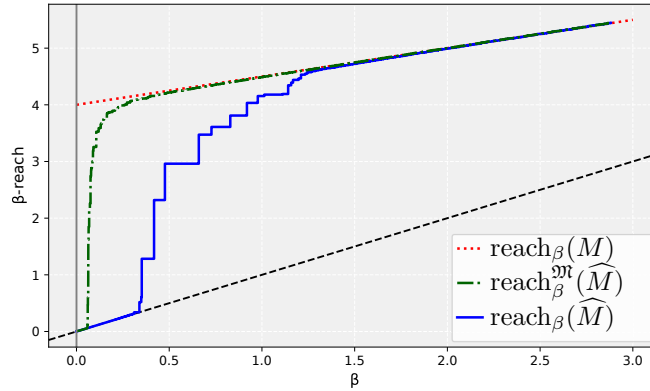


Figure 4.14: The exact β -reach profile (see Definition 4.6) of \widehat{M} in Example 4.5 and Figure 4.13 (a) is shown as the blue solid curve. Using the formulation in (4.23), and the mesh in Figure 4.13 (b), one may compute an improved approximation (green dashed curve) to the β -reach of the underlying manifold M (red dotted curve).

Example 4.5 (3-dimensional point cloud on a 2-dimensional manifold). Let $M \subset \mathbb{R}^3$ be a compact subset of a two-dimensional paraboloid, where its reach is decided by a point of maximal curvature at the vertex ($\text{reach}(M) = 4$). Let $\widehat{M} \subset M$ be a realization of 1500 points uniformly distributed on M (see Remark 4.13 in Section 4.5) shown in Figure 4.13 (a). Using the Python library `pyvista` Sullivan and Kaszynski (2019), we construct a mesh $\mathfrak{M}(\widehat{M})$ over the point cloud \widehat{M} (see Figure 4.13 (b)). By computing the distances of midpoints of pairs of points in \widehat{M} to the mesh $\mathfrak{M}(\widehat{M})$, rather than the nearest neighbor in \widehat{M} , we obtain the modified β -reach,

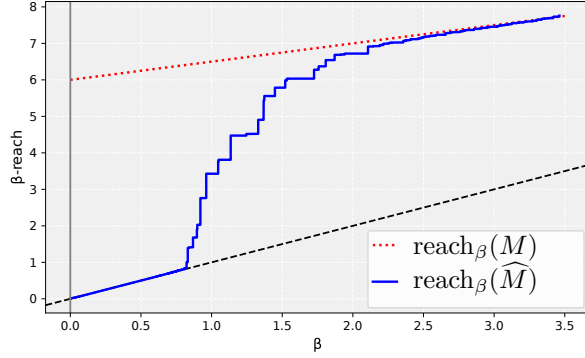


Figure 4.15: The exact β -reach profile of \widehat{M} in Example 4.6 is shown as the blue solid curve. The β -reach of the underlying manifold M (a 3-dimensional paraboloid embedded in \mathbb{R}^d , with $d \geq 4$) is shown as a red dotted curve.

$$\text{reach}_\beta^{\mathfrak{M}}(\widehat{M}) := \inf \left\{ g_{\|a_2 - a_1\|}(x) : a_1, a_2 \in \widehat{M}, \right. \quad (4.23)$$

$$\left. x = \delta_{\mathfrak{M}(\widehat{M})} \left(\frac{a_1 + a_2}{2} \right) \geq \beta \right\}.$$

The β -reach of M can be computed exactly as $\text{reach}_\beta(M) = 4 + \frac{\beta}{2}$, for β between 0 and some positive constant. As seen in Figure 4.14, this is well approximated by $\text{reach}_\beta(\widehat{M})$ for $\beta > 1.3$, and by $\text{reach}_\beta^{\mathfrak{M}}(\widehat{M})$ for $\beta > 0.4$; both approximations are efficiently computable from the point cloud \widehat{M} .

To estimate $\text{reach}(M)$ from the β -reach profiles in Example 4.5, one can perform a linear regression on $\text{reach}_\beta^{\mathfrak{M}}(\widehat{M})$, for β in the range where the curve appears linear, and estimate the reach as the model intercept.

Although the manifold M in Example 4.5 is derived from a simple paraboloid, this example is representative of other two-dimensional manifolds whose reach is determined by a region of high curvature.

The following two examples emphasize the applicability of this method in higher dimensions.

Example 4.6 (d -dimensional point cloud on a 3-dimensional manifold). Let $d \geq 4$, and let $M \subset \mathbb{R}^d$ be a compact subset of the three-dimensional paraboloid defined by $x_1^2 + x_2^2 + x_3^2 - 12x_4 = 0$ and $x_i = 0$, for $i = 5, \dots, d$, such that the reach of M is decided by a point of maximal curvature at the origin. It can be shown that $\text{reach}_\beta(M) = 6 + \frac{\beta}{2}$, for β between 0 and some positive value.

Let $\widehat{M} \subset M$ be a realization of 3000 points, uniformly distributed on M (see Remark 4.13 in Section 4.5). The Python package `pyvista` does not currently have methods for computing meshes in dimension higher than 3, so in this example, we do not compute $\text{reach}_\beta^{\mathfrak{M}}(\widehat{M})$ in (4.23). Mesh

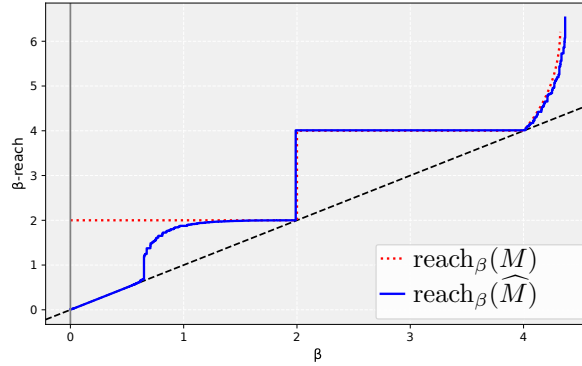


Figure 4.16: The exact β -reach profile of \widehat{M} in Example 4.7. The β -reach of the underlying manifold M is shown as a red dotted curve. Here, M is the union of two 3-dimensional hyperspheres of radius 2 embedded in \mathbb{R}^d .

generation in higher dimension is itself an active field of research (Boissonnat et al., 2008, 2009; Edelsbrunner, 2001). Nonetheless, we plot the exact β -reach profile of \widehat{M} in Figure 4.15.

Remark 4.10. The dimension of the ambient space $d \geq 4$ in Example 4.6 does not affect the shape of the β -reach profile of the point cloud \widehat{M} , computed purely in terms of distances between the points, and distances to midpoints. These distances are preserved under isometries to higher dimensional spaces. However, the number of points in \widehat{M} plays a role in the shape of the β -reach profile through the Hausdorff distance between \widehat{M} and the underlying manifold M . The exact relationship between the number of points in \widehat{M} and the Hausdorff distance $d_H(\widehat{M}, M)$ cannot be made precise with no prior knowledge of how the points are distributed on M —which also plays a role in the shape of the β -reach profile of \widehat{M} . Nonetheless, the dimension of M determines to a large extent the number of points needed to ensure that $d_H(\widehat{M}, M)$ is less than a given tolerance.

Example 4.7 (Two 3-dimensional hyperspheres). Let $d \geq 4$, and let $M \subset \mathbb{R}^d$ be the union of two three-dimensional hyperspheres of radius 2 whose centers are 12 units apart. Let $\widehat{M} \subset M$ be a realization of 3000 points uniformly distributed on M . The β -reach profiles of M and a realization of \widehat{M} are shown in Figure 4.16.

Example 4.7 highlights a few nice features of the β -reach profile. That is, it adapts very well to situations where the reach is determined by a bottleneck structure, as is the case for a hypersphere. For $\beta \in (2, 4]$, the β -reach is no longer determined by the radius of the hypersphere, but by half the distance between the surfaces of the spheres (in this case, $(12 - 2 - 2)/2 = 4$ units). Thus, the β -reach profile also provides information about the large scale features of the data. Notably, it gives the scales at which it becomes possible to distinguish these features from one another.

Finally, recall from Remark 4.2 that one can read from the plot information about the critical points of the generalized gradient function. In the case of Example 4.7, the distances from the three

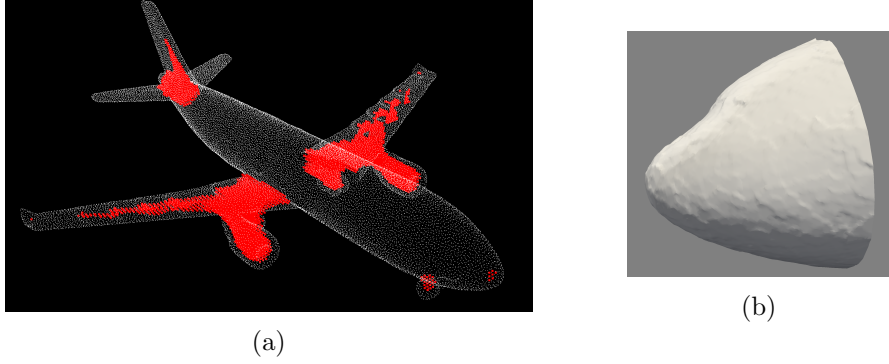


Figure 4.17: **(a)**: The point cloud data \hat{P} in Example 4.8 is used to guarantee that the r -convexity of the underlying surface of the plane P does not exceed the chosen test radius $r = 0.03$. The areas of the plane responsible for this restriction on the r -convexity are highlighted in red. **(b)**: The *pyvista* mesh $\mathfrak{M}(\hat{N})$ constructed from $\hat{N} \subset \hat{P}$, the points at the nose of the plane.

critical points to M are 2, 2, and 4.

4.4 Numerical studies

In this section, we test the methods for bounding the reach and r -convexity introduced in Section 4.3 against numerical data. First, in Section 4.4.1, we study the performance of our methods on real data. Then, in Section 4.4.2, we test the convergence results in Theorems 4.3 and 4.4.

4.4.1 Aircraft data

The real data in Example 4.8, below, is studied using the tools developed in this document. After a short description of the data, we perform several analyses using the tools developed in Section 4.3.

Example 4.8 (Aircraft data). Provided in Baorui et al. (2022) is a point cloud dataset that is sampled over the surface of a commercial aircraft (the white points in Figure 4.17 (a)). The diameter of the raw point cloud data (which corresponds to the length of the plane) is 0.749 units. Denote the surface of the aircraft by P and the approximating point cloud by \hat{P} .

We are also interested in the nose of the aircraft, $N \subset P$, the surface of the first 0.056 length units of the aircraft. A two-dimensional mesh approximating $\hat{N} := N \cap \hat{P}$ is shown in Figure 4.17 (b).

It is possible that in an engineering practice, one may want to identify which regions of a point cloud are not r -convex for a predefined value of r . Panel (a) of Figure 4.17 highlights in red the regions of the plane that are *certainly* not r -convex for the choice of $r = 0.03$.

The surface of the plane P is known to be two-dimensional, and so a cubic lattice of points φ with lattice spacing $a = 0.004$ superimposed over \hat{P} will almost surely not intersect P . The discrete dilation and erosion operations in Definition 4.7 are well defined using the larger point cloud $\mathcal{P} := \varphi \cup \hat{P}$. The red points in Figure 4.17 (a) are the elements of the discrete set $(\hat{P}_{r-\epsilon})_{-(r+\epsilon)} \cap \varphi$,

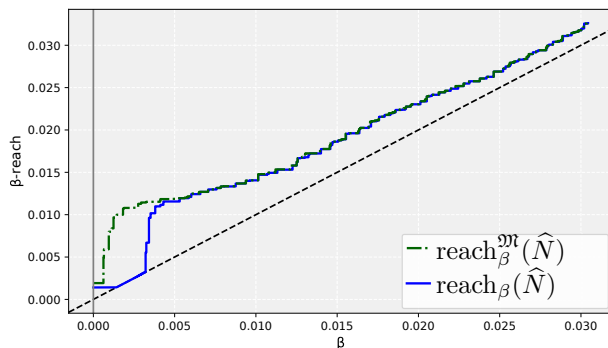


Figure 4.18: The exact β -reach profile of \widehat{N} in Example 4.8 is shown as the blue solid curve. Using the formulation in (4.23), and the mesh in Figure 4.17 (b), one may compute an improved approximation (green dashed curve) to the β -reach of the underlying (unknown) manifold N .

with $\epsilon := \sqrt{3}a/2 = \sup\{\delta_{\mathcal{P}}(q) : q \in \mathbb{R}^3\}$. Since this set is not empty, one has that P is a proper subset of $P_{\bullet r}$. In addition, by (4.2), one has conclusive evidence that $\text{reach}(P) \leq r$.

This example illustrates that, with this method, one can identify the regions responsible for limiting the r -convexity (and thus the reach), with a test specificity of 100%. One can improve the sensitivity of the test by decreasing the lattice spacing a . Then, other regions that are not r -convex (such as the interiors of the horizontal and vertical stabilizers) would be identified as such.

Remark that the nose of the plane is marked in red, due of course to a region of high curvature. By considering only the smooth manifold N in Example 4.8, corresponding to the nose of the plane, we can approximate the β -reach profile of N by $\text{reach}_{\beta}(\widehat{N})$ in Definition 4.6, or by $\text{reach}_{\beta}^m(\widehat{N})$ in Equation (4.23) using a linear nearest neighbor search.

The resulting β -reach profiles are shown in Figure 4.18. From the figure, the reach is clearly seen to be much smaller than 0.03 as indicated by the r -convexity experiment. Assuming that the β -reach profile of N maintains a constant slope near 0 as seen before in Figure 4.14, the approximation $\text{reach}(N) \approx 0.012$ can be read from Figure 4.18.

There are two main issues with using the reach bound in (4.15) on the aircraft data in Example 4.8. First, it is expected to produce a large overestimate of $\text{reach}(N)$ since N is a smooth manifold (see the discussion in Section 4.3.3). In addition, it is impossible to know the Hausdorff distance between N and its approximating point cloud \widehat{N} . Nevertheless, one can obtain a good approximation by considering the persistence diagram of growing balls centered at the points in \widehat{N} . The largest of the death-times of the topological features corresponds to the smallest radius for which the union of balls is homotopic to a point. This is likely to correspond closely to the Hausdorff distance $d_H(N, \widehat{N})$, and is thus a good choice for ϵ in (4.15). The persistence diagram, generated from the Python module `riper` Bauer (2021), gives that the largest death-time is $\epsilon = 0.0065$. From this, one calculates

$$\widehat{\text{rch}}^{(\epsilon)}(\widehat{N}) = 0.022.$$

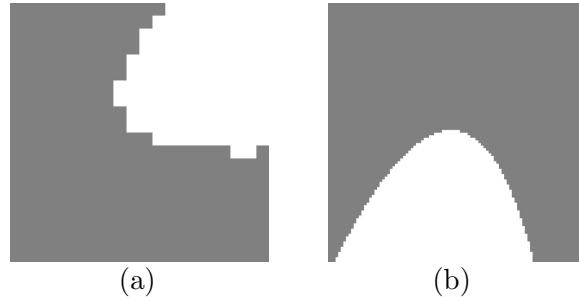


Figure 4.19: Realizations of the random set $\widehat{U}^{(n)}$ in Example 4.9. (a): $n = 2$. (b): $n = 10$.

Although the reach bound in (4.15) is less appropriate to use in this setting, we will see in the following section that it (and the r -convexity bound in (4.11)) is highly applicable when studying binary images.

4.4.2 Numerical convergence of reach and r -convexity bounds

In stochastic geometry literature, the *excursion set* of a random field is the subset of its domain on which the random field surpasses a predefined threshold (see Adler and Taylor (2007) for a comprehensive reference). The set in Figure 4.9, for example, is one realization of the excursion set of a stationary, isotropic Gaussian random field sampled on a square lattice.

The following two examples are meant to imitate the discretized excursion set of C^2 continuous random fields on square lattices (see, e.g., Biermé and Desolneux (2021); Cotsakis et al. (2023a,c)). A key feature of both examples, is that the reach and r -convexity of the sets are known, and so we can study the convergence of the bounds in Theorems 4.3 and 4.4 as the grid of sampling points becomes more fine. Each example illustrates one of the two cases mentioned in Remark 4.8 concerning the relationship between the reach, regions of high curvature, and bottleneck structures.

Example 4.9 (Reach determined by curvature). Let $U := \{(x, y) \in \mathbb{R}^2 : y \leq x^2/2\} \cap B(0, 10)$. It is easy to check that $\text{reach}(U) = \text{rconv}(U) = 1$. Moreover, the reach is determined by a point of maximal curvature at $(0, 0) \in U$.

Let $(\mathcal{P}^{(n)})_{n \geq 1}$ be a sequence of square lattices over \mathbb{R}^2 with lattice spacing $a_n = 0.7/n$, and independent, uniformly random position and orientation. For $n \in \mathbb{N}^+$, let $\widehat{U}^{(n)} := U \cap \mathcal{P}^{(n)}$ be the set U sampled on $\mathcal{P}^{(n)}$. Figure 4.19 depicts square subsets of realizations of $\widehat{U}^{(2)}$ and $\widehat{U}^{(10)}$ shown as binary images.

Example 4.10 (Reach determined by a bottleneck structure). Let $W := \{(x, y) \in \mathbb{R}^2 : |y| \geq x^2/2 + 1\} \cap B(0, 10)$. One has $\text{reach}(W) = \text{rconv}(W) = 1$, which is half the distance between its two connected components.

For the sequence of square lattices $(\mathcal{P}^{(n)})_{n \geq 1}$ in Example 4.9, let $\widehat{W}^{(n)} := W \cap \mathcal{P}^{(n)}$. Figure 4.20 depicts square subsets of realizations of $\widehat{W}^{(2)}$ and $\widehat{W}^{(10)}$ shown as binary images.

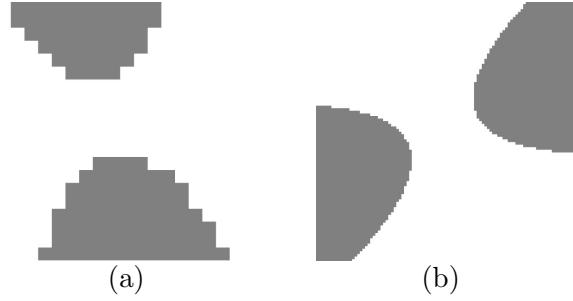


Figure 4.20: Realizations of the random set $\widehat{W}^{(n)}$ in Example 4.10. **(a)**: $n = 2$. **(b)**: $n = 10$.

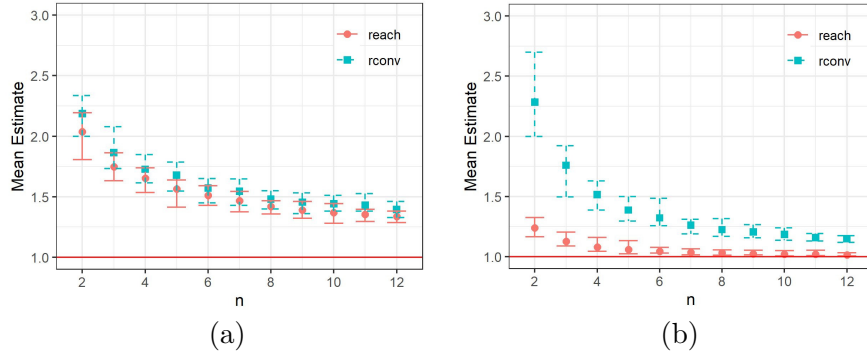


Figure 4.21: **(a)**: The mean and 95% confidence interval for the values of $\widehat{r}^{(a_n/\sqrt{2})}(\widehat{U}^{(n)})$ (squares) and $\widehat{rch}^{(\sqrt{1.25}a_n)}(\widehat{U}^{(n)})$ (circles) for 50 independent realizations of $\widehat{U}^{(n)}$ in Example 4.9 **(b)**: The mean and 95% confidence interval for the values of $\widehat{r}^{(a_n/\sqrt{2})}(\widehat{W}^{(n)})$ (squares) and $\widehat{rch}^{(\sqrt{1.25}a_n)}(\widehat{W}^{(n)})$ (circles) for 50 independent realizations of $\widehat{W}^{(n)}$ in Example 4.10.

Remark 4.11. The choice to use sets with a reach of 1 in Examples 4.9 and 4.10 does not limit their generality. The importance lies in the scale of the sampling lattice $\mathcal{P}^{(n)}$ relative to the reach, and the ratio of the two tends to 0 as $n \rightarrow \infty$ in both examples.

The following observation makes the generality of these examples even clearer: to consider the lattice of sampling points at various scales and orientations is equivalent to considering a fixed lattice and various scales and orientations of the underlying sets U and W .

For each $n \in \{2, \dots, 12\}$, we compute 50 independent realizations of $\widehat{U}^{(n)}$ and 50 independent realizations of $\widehat{W}^{(n)}$. For each of the resulting discrete sets, we compute the bounds of r -convexity and reach in (4.11) and (4.15) respectively. The means of the bounds are plotted in Figure 4.21 with empirical 95% confidence intervals, computed from the Harrell-Davis quantiles (Harrell and Davis, 1982). Panels (a) and (b) correspond to the results for the replications of $\widehat{U}^{(n)}$ and $\widehat{W}^{(n)}$ respectively.

For the r -convexity bound, $\epsilon_n = a_n/\sqrt{2}$ is chosen since $d_H(\mathcal{P}^{(n)}, \mathbb{R}^2) = a_n/\sqrt{2}$, for $n \in \mathbb{N}^+$. For the reach bound, we set $\epsilon_n = \sqrt{1.25}a_n$. The justification of this choice is quite technical, so we postpone the discussion to Remark 4.14 in Section 4.5.

$x \in (A_r)^c \oplus B(0, r) = (A_{\bullet r})^c$. If $x \in A_r \setminus A$, then by [Federer \(1959, Corollary 4.9\)](#),

$$\delta_{(A_r)^c}(x) = r - \delta_A(x) < r,$$

and so $x \in (A_r)^c \oplus B(0, r) = (A_{\bullet r})^c$, which proves (4.2). What remains to be shown is that if ∂A is C^1 -smooth and $(d-1)$ -dimensional, then $\text{rconv}(A) \leq \text{reach}(A)$. This inequality is shown via proof by contradiction. Suppose that

(i) $r > \text{reach}(A)$ and

(ii) $A_{\bullet r} = A$.

By (i), there exists $p \in A_{\frac{r+\text{reach}(A)}{2}}$ with no unique point in A closest to p . In particular, since A is closed, there exists two non-identical points $a_1, a_2 \in \partial A \subseteq A$ such that $\|p - a_1\| = \|p - a_2\| = \delta_A(p) < r$. Let n_1 be the unit normal vector to ∂A at a_1 , pointing towards p . Since a_1 is a limit point of A , one has $A^c \cap B(a_1, \epsilon) \neq \emptyset$ for all $\epsilon > 0$. By (ii), $A^c = \bigcup_{x \in (A_r)^c} B(x, r)$, and so for all $\epsilon > 0$, there exists $x \in (A_r)^c$ such that

(iii) $B(x, r) \cap B(a_1, \epsilon) \neq \emptyset$ and

(iv) $B(x, r) \cap A = \emptyset$.

The boundary ∂A is C^1 -smooth at a_1 , so it is easily checked that for ϵ close to 0, the locations x that satisfy both (iii) and (iv) are necessarily contained in a small neighborhood around $a_1 + rn_1$ (see [Figure 4.22](#)). That is, there exists a mapping $\theta : \epsilon \mapsto \theta(\epsilon) \in \mathbb{R}^+$ that tends to 0 as $\epsilon \rightarrow 0$, such that if $x \in (A_r)^c$ and $\epsilon \in \mathbb{R}^+$ satisfy (iii) and (iv), then $x \in B(a_1 + rn_1, \theta(\epsilon))$. Note that $\|(a_1 + rn_1) - a_2\| < r$ since $a_2 \in B(p, \delta_A(p)) \subset B(a_1 + rn_1, r)$. Choose ϵ such that $\theta(\epsilon) < r - \|(a_1 + rn_1) - a_2\|$. Then for any $x \in (A_r)^c \cap B(a_1 + rn_1, \theta(\epsilon))$,

$$\begin{aligned} \|x - a_2\| &\leq \|x - (a_1 + rn_1)\| + \|(a_1 + rn_1) - a_2\| \\ &\leq \theta(\epsilon) + \|(a_1 + rn_1) - a_2\| < r. \end{aligned}$$

This contradicts (iv). □

Lemma 4.2. *Let U be an open set in \mathbb{R}^d , and let $r \geq 0$. Then,*

$$\text{cl}(U)_{\bullet r} = \text{cl}(U) \quad \implies \quad U_{\bullet r} = U.$$

Proof of Lemma 4.2. By [Lemma 4.1](#), item (b), $U \subseteq U_{\bullet r} \subseteq \text{cl}(U)_{\bullet r}$. Note also that $U_{\bullet r}$ is open. Suppose that $\text{cl}(U)_{\bullet r} = \text{cl}(U)$. Then $U_{\bullet r}$ is an open subset of $\text{cl}(U)$ that contains all of the interior points of $\text{cl}(U)$. Therefore, $U_{\bullet r} = U$. □

Lemma 4.3. *Let $A \subseteq \mathbb{R}^d$ be closed. It holds that*

$$\text{rconv}(\text{cl}(A^c)) = \sup\{r \in \mathbb{R} : A_{\bullet-r} = A\}.$$

Proof of Lemma 4.3. By Lemma 4.2 and item (a) in Lemma 4.1,

$$\{r \in \mathbb{R} : \text{cl}(A^c)_{\bullet r} = \text{cl}(A^c)\} \subseteq \{r \in \mathbb{R} : (A^c)_{\bullet r} = A^c\} = \{r \in \mathbb{R} : A_{\bullet-r} = A\}.$$

Therefore, $\text{rconv}(\text{cl}(A^c)) \leq \sup\{r \in \mathbb{R} : A_{\bullet-r} = A\}$. Now we show the reverse inequality. Let $r, \tilde{r} \in \mathbb{R}^+$ satisfy $r > \tilde{r} > \text{rconv}(\text{cl}(A^c))$. By Proposition 4.2, $\text{cl}(A^c)_{\bullet \tilde{r}} \setminus \text{cl}(A^c)$ contains a ball of radius ϵ for some $\epsilon \in (0, r)$. Now, let $\delta = \min(\epsilon, 2(r - \tilde{r}))$. Note that $(A^c)_{\delta/2} \supseteq \text{cl}(A^c)$, and so by Lemma 4.1, $((A^c)_{\delta/2})_{\bullet r - \delta/2} = ((A^c)_r)_{\bullet r - \delta/2} \supseteq \text{cl}(A^c)_{\bullet r - \delta/2} \supseteq \text{cl}(A^c)_{\bullet \tilde{r}}$, which contains the ball of radius ϵ . Now, eroding by $\delta/2$ preserves a ball of radius $\epsilon - \delta/2 \geq \epsilon/2$. That is, $((A^c)_{\delta/2})_{\bullet r - \delta/2} \supseteq \text{cl}(A^c)_{\bullet r - \delta/2}$ contains a ball of radius $\epsilon/2$ that does not intersect $\text{cl}(A^c)$, and so $(A^c)_{\bullet r} \neq A^c$ which implies the desired $A_{\bullet-r} \neq A$ by Lemma 4.1, item (a). \square

Proof of Corollary 4.1. If $\text{rconv}(A) = 0$, then by Theorem 4.1, $\text{reach}(A) = 0$ and we are done. Now assume $\text{rconv}(A) > 0$ and $\text{rconv}(\text{cl}(A^c)) > 0$, then by Lemma 4.3, there exists $\delta > 0$ such that $A_{\bullet r} = A$ for all $r \in (-\delta, \delta)$. Theorem 1 in Walther (1999) states that ∂A is $(d-1)$ -dimensional and C^1 -smooth, which implies Equation 4.3 in our Theorem 4.1. \square

Proof of Proposition 4.1. By Theorem 4.1, it suffices to show that if $\partial(A_\epsilon)$ is not C^1 -smooth, then $\text{reach}(A_\epsilon) = \text{rconv}(A_\epsilon) = 0$. Indeed, if there is a point c on the boundary of $A_\epsilon = \bigcup_{a \in A} B(a, \epsilon)$ that does not have a continuous derivative, then c is at the boundary of two distinct balls of radius ϵ , contained in A_ϵ . In other words, c is at cusp that points inwards towards the interior of A_ϵ . Suppose that A_ϵ is r -convex for some $r > 0$. Then there is an open ball of radius r , with c as one of its limit points, that does not intersect either of the two balls with c on their boundaries. This is not possible, since there is exactly one ball of radius ϵ tangent to the ball of radius r at c , and so our assumption of r -convexity is false. Thus, we conclude $\text{rconv}(A_\epsilon) = 0$, and by (4.2), $\text{reach}(A_\epsilon) = 0$. \square

Remark 4.12. Remark that $(A_\epsilon)_{\bullet-\epsilon} = A_\epsilon$. Therefore, if the hypotheses of Proposition 4.1 are strengthened to those of Corollary 4.1, then the proof of Proposition 4.1 holds by applying Theorem 1 in Walther (1999) followed by our Theorem 4.1.

Here, we present an auxiliary lemma, and use it in our proof of Proposition 4.2 in Section 4.3.1.

Lemma 4.4. *For closed $A \subseteq \mathbb{R}^d$ and $r \in \mathbb{R}^+$, it holds that*

$$A_{\bullet r} = \left\{ p \in \mathbb{R}^d : \forall x \in B(0, r), B(p+x, r) \cap A \neq \emptyset \right\}. \quad (4.24)$$

Proof of Lemma 4.4. By manipulating the expressions in Definition 4.1, one obtains

$$A_{\bullet r} = \left(\bigcup_{y \in (A \oplus B(0, r))^c} B(y, r) \right)^c, \quad (4.25)$$

which reads: *the elements of $A_{\bullet r}$ are the $p \in \mathbb{R}^d$ such that p is not contained in any closed ball of radius r that does not intersect A .* This statement is equivalent to: *$p \notin A_{\bullet r}$ if and only if there exists a closed ball of radius r that contains p but does not intersect A .* Thus, we have shown that $(A_{\bullet r})^c$ is equal to the complement of the RHS of (4.24), which proves the desired result. \square

Proof of Proposition 4.2. Let $\tilde{r} \in (\text{rconv}(A), r)$ and fix $p \in A_{\bullet \tilde{r}} \setminus A$. Since A is closed, there is an open neighborhood containing p that does not intersect A . Choose $\epsilon \in \mathbb{R}^+$ such that $B(p, \epsilon) \cap A = \emptyset$. There exists a sufficiently small $\delta \in (0, \epsilon)$ such that for every $y \in B(p, \delta + r)$, there exists an $x \in B(0, \tilde{r})$ that satisfies $B(p+x, \tilde{r}) \setminus B(p, \epsilon) \subset B(y, r)$. Since $p \in A_{\bullet \tilde{r}}$, one has that $B(p+x, \tilde{r}) \setminus B(p, \epsilon)$ contains an element of A by Lemma 4.4. By inclusion, $B(y, r)$ contains a point in A as well.

Let $z \in B(p, \delta)$. We have shown that z is in the right-hand side of (4.24), since, for all $x \in B(0, r)$, one has $y := z+x \in B(z, \delta+r)$ by the triangle inequality, and so by previous arguments, $B(z+x, r)$ contains an element of A . Therefore, by Lemma 4.4, $z \in A_{\bullet r}$. But z is not in A since $B(p, \delta) \subseteq B(p, \epsilon) \subseteq A^c$. Therefore, $z \in A_{\bullet r} \setminus A$ and so $B(p, \delta) \subseteq A_{\bullet r} \setminus A$. \square

Proof of Theorem 4.3. We begin by showing (4.12). Let $n \in \mathbb{N}^+$ and fix $p \in \mathcal{P}^{(n)} \setminus \widehat{A}^{(n)}$. If $\epsilon_n \geq \text{rconv}(A)$, then (4.12) holds trivially. Now, let $r \in \mathbb{R}^+$ be such that $\epsilon_n < r < \text{rconv}(A)$ so that $A_{\bullet r} = A$. We aim to show that $p \notin (\widehat{A}_{r-\epsilon_n}^{(n)})_{-(r+\epsilon_n)}$.

Indeed, by the r -convexity of A , there exists $x \in (A_r)^c$ such that $\|x - p\| < r$. In addition, $B(x, \epsilon_n) \cap A_{r-\epsilon_n} = \emptyset$, so there exists $q \in \mathcal{P}^{(n)} \setminus \widehat{A}_{r-\epsilon_n}^{(n)}$ such that $\|q - x\| \leq \epsilon_n$. Therefore, by the triangle inequality, $\|q - p\| \leq r + \epsilon_n$, which implies $\delta_{\mathcal{P}^{(n)} \setminus \widehat{A}_{r-\epsilon_n}^{(n)}}(p) \leq r + \epsilon_n$. Thus, $p \notin (\widehat{A}_{r-\epsilon_n}^{(n)})_{-(r+\epsilon_n)}$ as desired.

Now, to prove (4.13), first fix $r > \text{rconv}(A)$. To simplify notation, let $\delta_n := d_H(\widehat{A}^{(n)}, A)$. By Proposition 4.2, there exists an open subset $O \subseteq A_{\bullet r} \setminus A$ that contains a closed ball of radius $\delta_{n_0} + 3\epsilon_{n_0}$ for sufficiently large $n_0 \in \mathbb{N}^+$. Let $n \geq n_0$. The spacing between points in the point cloud $\mathcal{P}^{(n)}$ is sufficiently small such that there exists $q \in O_{-(\delta_n + 2\epsilon_n)} \cap \mathcal{P}^{(n)}$. Importantly, this implies that $q \in (A_r)_{-(r+\delta_n+2\epsilon_n)} \cap \mathcal{P}^{(n)}$. By Lemma 4.1, for all $s > r$, we have

$$(A_r)_{-(r+\delta_n+2\epsilon_n)} \subseteq ((A_r)_{\bullet(s-r)})_{-(r+\delta_n+2\epsilon_n)} = (A_s)_{-(s+\delta_n+2\epsilon_n)},$$

and therefore $q \in (A_s)_{-(s+\delta_n+2\epsilon_n)} \cap \mathcal{P}^{(n)}$. Notice that $A_s \cap \mathcal{P}^{(n)} \subseteq \widehat{A}_{\delta_n+s}^{(n)}$, which implies $(A_s)^c \supseteq \mathcal{P}^{(n)} \setminus \widehat{A}_{s+\delta_n}^{(n)}$ and thus

$$(A_s)_{-(s+\delta_n+2\epsilon_n)} \cap \mathcal{P}^{(n)} \subseteq (\widehat{A}_{\delta_n+s}^{(n)})_{-(s+\delta_n+2\epsilon_n)}.$$

Summarizing, we have shown that there is a point q in $(\widehat{A}_{\delta_n+s}^{(n)})_{-(s+\delta_n+2\epsilon_n)}$ that is not in $\widehat{A}^{(n)}$. By

the change of variables $\tilde{s} := s + \delta_n + \epsilon_n$, it follows that $(\widehat{A}_{\tilde{s}-\epsilon_n}^{(n)})_{-(\tilde{s}+\epsilon_n)} \setminus \widehat{A}^{(n)} \neq \emptyset$ for all $\tilde{s} > r + \delta_n + \epsilon_n$, which implies $\widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) \leq r + \delta_n + \epsilon_n$.

Sending $n \rightarrow \infty$ yields $\lim_{n \rightarrow \infty} \widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) \leq r$, and since $r \in (\text{rconv}(A), \infty)$ was chosen freely, $\lim_{n \rightarrow \infty} \widehat{r}^{(\epsilon_n)}(\widehat{A}^{(n)}) \leq \text{rconv}(A)$. This result, along with (4.12), gives the convergence in (4.13). \square

Proof of Equations (4.8) and (4.9). The reach of A is equal to the inverse of the curvature of f at $x = 0$ (Aamari et al., 2019, Theorem 3.4). Equation (4.8) holds since the curvature of f at $x = 0$ is $f''(0) = 2h'(0)$.

Now we show (4.9). Without loss of generality, suppose $h(0) = f(0) = 0$. For x in a small neighborhood of 0, consider the symmetric points $a_1 = (x, f(x))$ and $a_2 = (-x, f(x))$ in A , and remark that

$$g_{\|a_2 - a_1\|} \circ \delta_A \left(\frac{a_1 + a_2}{2} \right) = \frac{x^2}{2f(x)} + \frac{f(x)}{2}, \quad (4.26)$$

since $\delta_A \left(\frac{a_1 + a_2}{2} \right) = f(x)$ for sufficiently small x . Thus, the β -reach, for $\beta = f(x)$, satisfies

$$\text{reach}_{f(x)}(A) \leq g_{\|a_2 - a_1\|} \circ \delta_A \left(\frac{a_1 + a_2}{2} \right). \quad (4.27)$$

There is a $\delta > 0$ such that h has an inverse on $[0, \delta]$, and so for $\beta \in [0, \delta]$, choose x such that $f(x) = \beta$. Remark that

$$x^2 = h^{-1}(\beta) = \frac{\beta}{h'(0)} - \frac{\beta^2 h''(0)}{2h'(0)^3} + o(\beta^2),$$

by Taylor's theorem. Plugging into (4.26) and applying (4.27), one obtains (4.9). \square

We believe that the inequality in (4.27) can be strengthened to an equality. This would provide the exact value of K_A in (4.14) for A in Example 4.4.

Remark 4.13 (Generating points uniformly on the graph of a function). In Examples 4.5 and 4.6, we generate points on the graph of a function *uniformly* with respect to the Riemannian metric. This can be achieved as follows. Let $G \subset \mathbb{R}^{m+1}$ be the graph of a C^1 -smooth function $f : T \rightarrow \mathbb{R}$, where $T \subset \mathbb{R}^m$ is compact. To simulate points uniformly on G , it suffices to simulate the points on T with the appropriate probability density, and map the points up to G by the function f . The appropriate density on T is given by

$$\lambda(t) := c \sqrt{1 + \|\nabla f(t)\|^2},$$

for $t \in T$, where ∇f denotes the gradient of f , and c is a normalizing constant. This is true since $\lambda(t)/c$ is the determinant of the Jacobian of the map induced by f from T to G . If X is distributed randomly on T with density λ , then the image of X in G is uniform in G .

Remark 4.14 (The choice of ϵ_n for the reach bound in Section 4.4.2). In Section 4.4.2, we choose $\epsilon_n = \sqrt{1.25}a_n$ when computing $\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{U}^{(n)})$ in Example 4.9 and $\widehat{\text{rch}}^{(\epsilon_n)}(\widehat{W}^{(n)})$ in Example 4.10,

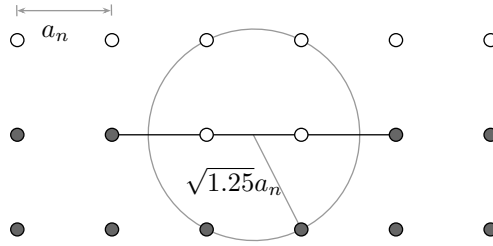


Figure 4.23: The grey points and the white points can be separated by a curve with arbitrarily small curvature (arbitrarily large reach). The midpoint of two grey points may be up to a distance of $\sqrt{1.25}a_n$ from another grey point, while remaining arbitrarily close to the separating curve.


where a_n denotes the lattice spacing of the square grid on \mathbb{R}^2 . Although this choice of ϵ_n leads to $d_H(\hat{U}^{(n)}, U) > \epsilon_n$, or $d_H(\hat{W}^{(n)}, W) > \epsilon_n$ with positive probability, it is sufficient for the result of Theorem 4.4 in this two-dimensional case.

Let us take the case of U for example. The condition $d_H(\hat{U}^{(n)}, U) > \epsilon_n$ can be weakened to “midpoints of pairs of points in $\hat{U}^{(n)}$ are no further from $\hat{U}^{(n)}$ than ϵ_n plus the distance to U ”, and the result of Theorem 4.4 still holds. Figure 4.23 depicts a situation where, for a very fine grid ($a_n \ll 1$), the midpoint of two points in $\hat{U}^{(n)}$ is a distance of $\sqrt{1.25}a_n$ from another point in $\hat{U}^{(n)}$, while its distance to U is negligible. This worst-case scenario justifies the choice of $\epsilon_n = \sqrt{1.25}a_n$.

Chapter 5

A local statistic for the spatial extent of extreme threshold exceedances

This chapter is based on the joint work

 Cotsakis, R., Di Bernardino, E., & Opitz, T. (2023). A local statistic for the spatial extent of extreme threshold exceedances. *Submitted for publication in an international peer-reviewed journal*. [Paper here](#).

Abstract: We introduce the *extremal range*, a local statistic for studying the spatial extent of extreme events in random fields on \mathbb{R}^2 . Conditioned on exceedance of a high threshold at a location s , the extremal range at s is the random variable defined as the smallest distance from s to a location where there is a non-exceedance. We leverage tools from excursion-set theory to study distributional properties of the extremal range, propose parametric models and predict the median extremal range at extreme threshold levels. The extremal range captures the rate at which the spatial extent of conditional extreme events scales for increasingly high thresholds, and we relate its distributional properties with the bivariate tail dependence coefficient and the extremal index of time series in classical Extreme-Value Theory. Consistent estimation of the distribution function of the extremal range for stationary random fields is proven. For non-stationary random fields, we implement generalized additive median regression to predict extremal-range maps at very high threshold levels. An application to two large daily temperature datasets, namely reanalyses and climate-model simulations for France, highlights decreasing extremal dependence for increasing threshold levels and reveals strong differences in joint tail decay rates between reanalyses and simulations.

5.1 Introduction

Assessing the spatial and temporal correlation of extreme events is an important modeling step in applications of environmental statistics where extreme risks arise from concurrence and compounding of extremes (Dombry et al., 2018; AghaKouchak et al., 2020). We here focus on assessing the spatial contiguity of extreme events based on excursion sets with the aim to avoid strong parametric assumptions and high numerical cost when inferring spatial extremal dependence properties from large datasets on regular grids. Analysis of excursion sets has become valuable in spatial statistics (e.g., Bolin and Lindgren, 2015; Sommerfeld et al., 2018) and computer vision (e.g., Bleau and Leon, 2000; Sezgin and Sankur, 2004), especially for data on regular grids, such as climate model output, remote sensing data or medical images. We use the framework of Extreme-Value Theory (EVT, de Haan and Ferreira, 2006), useful to formulate general tail-regularity assumptions and enable statistical extrapolation towards very high and even yet unobserved quantiles. The standard asymptotic models in spatial EVT exhibit asymptotic dependence where the limiting dependence structure of threshold exceedances is characterized by peaks-over-threshold stability (Ferreira and de Haan, 2014; Dombry and Ribatet, 2015; Thibaud and Opitz, 2015). However, strong empirical evidence from many environmental processes advises against this property (Tawn et al., 2018; Huser and Wadsworth, 2022). Often, spatial dependence between threshold exceedances is lost as thresholds are increased, and it may ultimately vanish in the case of asymptotic independence. More flexible subasymptotic models have been proposed to accommodate asymptotic independence or even both situations of asymptotic (in)dependence (Huser et al., 2017; Huser and Wadsworth, 2022; Zhang et al., 2022).

Here, we use a setting borrowing from the idea of spatial conditional extremes (Heffernan and Tawn, 2004; Wadsworth and Tawn, 2022) to better understand spatial joint tail decay behavior near a location of interest. The tail dependence coefficient $\lim_{u \rightarrow 1} \mathbb{P}(F_2(X_2) > u \mid F_1(X_1) > u)$ of two random variables $X_i \sim F_i$, $i = 1, 2$, is a conditional probability that is a routinely used exploratory and diagnostic tool to assess the strength of bivariate extremal dependence (Coles et al., 1999). In spatial statistics, however, one usually has access to observations for a relatively large number of locations, and so methods of assessing extremal dependence based only on pairwise observations exclude pertinent information about multivariate, or spatial dependence structure, and numerical computation may become very costly for data available on regular grids with a large number of locations. There have been a number of methods to overcome this issue. Wadsworth and Tawn (2022); Simpson et al. (2023), for example, propose a parametric inference method based on a spatial extension of the multivariate conditional extremes model of Heffernan and Tawn (2004), which also relies on conditioning on an exceedance in one of the variables. Other works based on the conditional extremes model are limited in their applicability to high-dimensional datasets. As noted by Wadsworth and Tawn (2022), it is common in environmental data for the spatial dependence to weaken as the considered threshold increases. One interpretation of this phenomenon, the inspiration for the statistics introduced in this paper, is that the *spatial extent* of extreme events

tends to decrease with an increase in the threshold level. Thus, in this paper, we focus on the size and other geometric properties of excursion sets of continuous planar random fields—the regions where the random fields exhibit threshold exceedances.

There is a vast literature concerning the geometric features of excursion sets of random fields; see [Adler and Taylor \(2007\)](#) for a comprehensive introduction. For smooth random fields, geometric summaries of excursion sets, namely their intrinsic volumes, carry pertinent information about the asymptotic dependence structure at extreme thresholds ([Di Bernardino et al., 2022](#)). In this paper, we introduce a new local statistic, the *extremal range*. The extremal range at a site $s \in \mathbb{R}^2$ is defined as the largest radius r around s such that all locations within r are extreme, conditioned on a threshold exceedance at s . We will explore how the extremal range relates to the intrinsic volumes of the excursion set and to the notion of asymptotic dependence defined by a positive value of the tail dependence coefficient.

The extremal range can be seen as a spatial analogue to the extremal index ([Moloney et al., 2019](#)), a popular asymptotic statistic for time series extremes that allows for interpretation as the reciprocal of the average number of consecutive time steps over which an extreme cluster spans. In this sense, both quantities provide a notion of the size of clusters of extremes. However, several notable distinctions can be made. Firstly, we consider two-dimensional Euclidean space and not one single time dimension with regular discrete time steps. In one dimension, the distributional properties of the extremal range and its asymptotics at high thresholds can be obtained by studying sojourn times of one dimensional stochastic processes ([Berman, 1971, 1982](#); [Kratz, 2006](#); [Pham, 2013](#); [Dalmao et al., 2019](#)). Where the classical extremal index is equal to unity in the case of asymptotic independence and therefore not informative, the extremal range can be used to quantify the degree of asymptotic dependence for asymptotically independent random fields. An important practical difference further stems from the fact that edge effects at the boundary of the observation window play a more important role in the spatial setting than in the temporal one. Thus, for spatial environmental datasets, care needs to be taken when computing the extremal range when the surrounding data is censored or unavailable.

Our results are organized as follows. Section [5.2](#) introduces the extremal range and notations. In Section [5.3](#), we express the cumulative distribution function of the extremal range through the intrinsic volumes of the excursion regions. In Section [5.4](#), we study the asymptotic behavior of the extremal range for common random field models as the threshold at the conditioning location s is increased, and propose a parametric model for the quantiles of the extremal range. Inference methods for estimating the extremal range and its quantiles are described in Section [5.5](#), and are applied to French temperature data by using a generalized additive quantile regression framework in Section [5.6](#). Some technical definitions and examples are postponed to Appendix [5.A](#). Finally, proofs and supporting lemmas for the theory established in Sections [5.3](#), [5.4](#), and [5.5](#) are provided in Appendix [5.B](#).

5.2 The extremal range and relevant notations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \times \mathbb{R}^2 \rightarrow \mathbb{R}$ be a random field defined on \mathbb{R}^2 , the Euclidean plane, endowed with the Euclidean metric $\|\cdot\|$ and Lebesgue measure $\mathcal{L}(\cdot)$. For a study domain $S \subseteq \mathbb{R}^2$, let ∂S denote its topological boundary. For $x \in \mathbb{R}^2$, denote the distance between x and a non-empty set S by $\text{dist}(x, S) := \inf\{\|x - s\| : s \in S\}$. Throughout this paper, $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ denotes a deterministic threshold that is allowed to vary in space, and we focus on the binary random field of excursion indicators $\{X(t) > u(t)\}_{t \in \mathbb{R}^2}$. This is expressed in terms of the following definition.

Definition 5.1 (Excursion set). Let X be a random field on \mathbb{R}^2 and $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a threshold function that may vary in space. Define the excursion set of X to be

$$E_X(u) := \{t \in \mathbb{R}^2 : X(t) > u(t)\}.$$

Definition 5.2 (Extremal range). For $r > 0$ and $s \in \mathbb{R}^2$, let $B(s, r) := \{t \in \mathbb{R}^2 : \|t - s\| \leq r\}$ denote the closed ball of radius r centered at s . Let $\tilde{R}^{(u)} : \Omega \times \mathbb{R}^2 \rightarrow \mathbb{R}^+ \cup \{0, \infty\}$ be a random field defined by

$$\tilde{R}^{(u)}(t) := \sup \{r \in \mathbb{R}^+ : B(t, r) \subset E_X(u)\} = \text{dist}(t, (E_X(u))^c), \quad t \in \mathbb{R}^2.$$

Let $s \in \mathbb{R}^2$ satisfy $u(s) < x^*(s)$, with $x^*(s) := \inf\{x \in \mathbb{R} : \mathbb{P}(X(s) > x) = 0\}$ denoting the upper end-point of the marginal density of X at s . Define the *extremal range* at s , denoted $R_s^{(u)}$, to be the random variable whose pushforward measure is given by

$$\mathbb{P}(R_s^{(u)} \in A) = \mathbb{P}(\tilde{R}^{(u)}(s) \in A \mid X(s) > u(s)), \quad A \in \mathcal{B}(\mathbb{R}).$$

Remark 5.1. As discussed in [Moloney et al. \(2019\)](#), the inverse of the so-called extremal index quantifies the average size of clusters of threshold exceedances for time series, *i.e.*, for one-dimensional discrete random processes. Analogously, the extremal range provides a notion of the size of the clusters of sites that exhibit threshold exceedances for continuous, two-dimensional random fields.

The definitions below are relevant to establish the main results for the extremal range.

Definition 5.3 (Erosion and dilation). For two nonempty sets $A, B \subseteq \mathbb{R}^2$, let $A \oplus B := \{x + y : x \in A, y \in B\}$ be the Minkowski sum of A and B . For $r \in \mathbb{R}$, and $S \subseteq \mathbb{R}^2$ let

$$S_r := \begin{cases} S \oplus B(0, r), & \text{for } r \geq 0, \\ (S^c \oplus B(0, -r))^c, & \text{for } r < 0, \end{cases}$$

denote respectively the set dilation and the set erosion, depending on the sign of r .

Definition 5.4 (Connected components of the level set). Let $T \subset \mathbb{R}^2$ be compact and define $N_T^{(u)}$ to be the number of connected components of the level curve $\partial E_X(u) \cap T$.

Definition 5.5 (Lipschitz-Killing curvature densities). For a set $S \subset \mathbb{R}^2$, let $\chi(S)$ denote the Euler-Poincaré characteristic of S (equal to the number of connected components of S minus the number of holes in S) and let $\ell(\partial S)$ denote the perimeter length of S (the one-dimensional Hausdorff measure of its boundary). Recall that \mathcal{L} denotes the Lebesgue measure. For a compact set $T \subset \mathbb{R}^2$ with positive Lebesgue measure, assuming the limits exist, define the *curvature densities*,

$$\begin{aligned} C_0^*(E_X(u)) &:= \lim_{n \rightarrow \infty} \frac{\mathbb{E}[\chi(E_X(u) \cap nT)]}{\mathcal{L}(nT)}, \\ C_1^*(E_X(u)) &:= \lim_{n \rightarrow \infty} \frac{\mathbb{E}[\ell(\partial(E_X(u) \cap nT))]}{2\mathcal{L}(nT)}, \\ C_2^*(E_X(u)) &:= \lim_{n \rightarrow \infty} \frac{\mathbb{E}[\mathcal{L}(E_X(u) \cap nT)]}{\mathcal{L}(nT)}, \end{aligned}$$

where nT is the result after linearly rescaling T by n .

Note that $C_i^*(E_X(u))$, for $i = 0, 1, 2$, are the limiting normalized intrinsic volumes of the excursion set $E_X(u)$ seen on large domains (Schneider and Weil, 2008, Theorem 9.3.3). They play an important role in determining the shape of the distribution function of the extremal range; a topic that we investigate in the following section.

5.3 Linking the extremal range and intrinsic volumes

Proposition 5.1. *Suppose that the random field X is continuous and stationary, and that u is constant and less than $x^*(0)$. For any compact set $T \subset \mathbb{R}^2$ with $\mathcal{L}(T) > 0$, the distribution function of $R_0^{(u)}$ is given by*

$$\mathbb{P}(R_0^{(u)} \leq r) = 1 - \frac{\mathbb{E}[\mathcal{L}(E_X(u)_{-r} \cap T)]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]}, \quad (5.1)$$

for $r \geq 0$, and $\mathbb{P}(R_0^{(u)} < 0) = 0$, where the subscript $-r$ denotes set erosion by a radius of r (see Definition 5.3, Equation (5.1)).

The proof of Proposition 5.1 is provided in Appendix 5.B.1. The extremal range has close links with the *spherical erosion function* (Serra, 1984; Ripley, 1988), which describes the distribution function of the distance of a uniform random point in a set to the set's boundary. Proposition 5.1 states that the eroded excursion set $E_X(u)_{-r}$ carries information about the distribution of $R_s^{(u)}$ through its area when intersected with a compact set T . Areas of excursion sets and their erosion can be efficiently estimated with routine algorithms, such that Equation (5.1) can be used for estimating the distribution function of $R_0^{(u)}$ in the stationary setting by replacing expectations with empirical estimates.

Next, we show that under certain regularity conditions, a polynomial expression of the Lebesgue measure of an eroded set in terms of its Lipschitz-Killing curvatures (see Definition 5.5) can be obtained as corollary to the well-known Steiner formula (Federer, 1959, Theorem 5.6). The most general case for which this is known to hold is when the complement of the considered set has *positive reach* (see Definition 5.9 in Appendix 5.A.1).

Assumption 5.1. *Suppose that for the random field X paired with the threshold function u , the excursion set $E_X(u)$ is a stationary random set. In addition, for any compact, convex $T \subset \mathbb{R}^2$ with positive Lebesgue measure, suppose that*

- *the densities in Definition 5.5 exist, are finite, and are independent of T ;*
- *$E_X(u)^c \cap T$ is almost surely a positive reach set;*
- *$\mathbb{E}[N_T^{(u)}] < \infty$ (see Definition 5.4).*

Under Assumption 5.1, the random field X is not necessarily stationary, as u is not necessarily a constant function in space. What is necessary instead is that the *excursion set* at the level u be stationary. An important, easily verifiable consequence of this is that $C_2^*(E_X(u)) = \mathbb{P}(X(0) > u(0))$. The condition that $E_X(u)^c \cap T$ is positive reach implies a certain regularity or smoothness of $\partial E_X(u)$. Conversely, compact subsets of \mathbb{R}^2 with a C^2 smooth boundary have positive reach (Thäle, 2008, Proposition 14). Assumption 5.1 also implies that $E_X(u)$ is almost surely open, as its complement must be closed to satisfy the positive reach property. Examples of random fields that satisfy the last item in Assumption 5.1 are the Gaussian fields discussed in Beliaev et al. (2020). However, Gaussianity is not a necessary condition for our results, except for Proposition 5.2 focusing on results for such fields. A final remark on Assumption 5.1 is that it allows for the random fields X and $X - u$ to be discontinuous with positive probability.

An important property of the extremal range under Assumption 5.1 is asserted by the following Lemma, which we prove in Appendix 5.B.1.

Lemma 5.1. *Under Assumption 5.1, $\mathbb{P}(R_0^{(u)} \leq r)$ is continuous in r , for $r > 0$.*

The main result of this section is the following first-order approximation of the distribution function of the extremal range. The proof of Theorem 5.1 can be found in Appendix 5.B.1.

Theorem 5.1. *Under Assumption 5.1, for $r > 0$, it holds that*

$$\frac{\mathbb{P}(R_0^{(u)} \leq r)}{r} \xrightarrow{r \rightarrow 0} \frac{2C_1^*(E_X(u))}{C_2^*(E_X(u))}. \quad (5.2)$$

Theorem 5.1 shows that the distribution of the extremal range follows a first-order Taylor expansion for positive values of the radius r near 0. Moreover, the linear coefficient is provided by the limit on the right-hand side of Equation (5.2). By studying how this coefficient behaves for large thresholds, we gain insight about the spatial extent of high threshold exceedances.

5.4 The asymptotic behavior of the extremal range

We study the asymptotic behavior of the extremal range as the threshold function u tends to the location-wise upper endpoint of the distribution of X everywhere in space. By studying the extremal range, we aim to capture information about the dependence structure of the random field X . Therefore, we use the threshold function $u_p : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined below as a location-wise quantile, such that it naturally adapts to the margins of the random field X .

Definition 5.6. For $p \in (0, 1)$ and a random variable $Y : \Omega \rightarrow \mathbb{R}$, let $q_p(Y) \in \mathbb{R}$ denote the p -quantile of Y , *i.e.*, $q_p(Y) := \inf\{r \in \mathbb{R} : \mathbb{P}(Y \leq r) \geq p\}$. Now, define the adaptive threshold u_p by the mapping $u_p(s) := q_p(X(s))$, for $s \in \mathbb{R}^2$.

Theorem 5.1 allows us to study how the extremal range decreases as the considered threshold increases, *i.e.*, as $p \rightarrow 1$. This important result is summarized in the following corollary.

Corollary 5.1. *Suppose that Assumption 5.1 holds for the threshold function u_p , for all sufficiently large $p \in (0, 1)$. Then a function $g : (0, 1) \rightarrow \mathbb{R}$ satisfies*

$$\lim_{p \rightarrow 1} g(p) \frac{C_2^*(E_X(u_p))}{2C_1^*(E_X(u_p))} = \frac{1}{K}, \quad (5.3)$$

for some $K \in \mathbb{R}^+$, if and only if

$$\lim_{p \rightarrow 1} \lim_{r \rightarrow 0} \frac{\mathbb{P}(g(p)R_0^{(u_p)} \leq r)}{r} = K. \quad (5.4)$$

Proof. Theorem 5.1 tells us that for any p ,

$$\lim_{r \rightarrow 0} \frac{\mathbb{P}(g(p)R_0^{(u_p)} \leq r)}{r} = \frac{2C_1^*(E_X(u_p))}{g(p)C_2^*(E_X(u_p))}.$$

Sending $p \rightarrow 1$ yields the desired result. □

An interpretation of Corollary 5.1 is that the probability density function of $g(p)R_0^{(u_p)}$ just to the right of 0 approaches 1 if and only if $g(p)$ is asymptotically equivalent to $2C_1^*(E_X(u_p))/C_2^*(E_X(u_p))$ as $p \rightarrow 1$. In this sense, Corollary 5.1 shows how $R_0^{(u_p)}$ scales as $p \rightarrow 1$. We are not able to use Corollary 5.1 to establish a non-degenerate limit distribution of $[2C_1^*(E_X(u_p))/C_2^*(E_X(u_p))]R_0^{(u_p)}$ as $p \rightarrow 1$; it is not always possible to exchange the order of the limits in Equation (5.4). A counterexample is provided in Appendix 5.A.2.

5.4.1 Non-degenerate limit distributions of the extremal range

Here, we study certain cases of widely used spatial random field models where the extremal range is known to have a non-degenerate limit distribution at high thresholds after appropriate rescaling.

This will serve as basis for defining parametric statistical models for the extrapolation behavior of the extremal range at extreme conditioning thresholds. The random fields that we will consider in this section are stationary, so we choose a threshold function u that is constant throughout space. To ease notation, we write u to denote both the constant mapping $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ and its image in \mathbb{R} .

Gaussian random fields

For a smooth, stationary Gaussian process Y on \mathbb{R} , if one is to condition on the event $\{X(0) > u\}$ for some large threshold $u \in \mathbb{R}$, one can show using tools developed in [Kac and Slepian \(1959\)](#) that the connected component of the excursion set containing 0 is a random interval with expected length asymptotically equivalent to $1/u$. By analogy, after appropriately rescaling in the spatial dimension, one finds that the limit process is a random parabola with deterministic shape. These insights are formally generalized for the two-dimensional case in the following proposition formulated for smooth standard Gaussian fields, for which a proof is given in [Appendix 5.B.2](#).

Proposition 5.2. *Suppose that X is a stationary, isotropic, centered Gaussian random field on \mathbb{R}^2 with covariance function*

$$\rho(h) = 1 - \frac{\alpha}{2} \|h\|^2 + o(\|h\|^2), \quad \alpha > 0, \quad (5.5)$$

for h in a neighbourhood of 0. Then $\mathbb{P}(uR_0^{(u)} \in \cdot)$ converges to a non-degenerate probability distribution, as $u \rightarrow \infty$.

A stationary, isotropic Gaussian random field with unit variance and C^1 -smooth sample paths has the covariance function in (5.5) with α equal to its second spectral moment; see [Leadbetter et al. \(1983, page 151\)](#) and [Cambanis \(1973\)](#). The isotropic, Matérn covariance function

$$\rho(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|h\|}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} \|h\|}{l} \right), \quad \nu, l > 0,$$

with K_ν denoting the modified Bessel function of the second kind, satisfies (5.5) for $\nu > 1$ and

$$\alpha = \frac{\nu}{l^2(\nu - 1)}.$$

For a random field X as described in [Proposition 5.2](#), the expressions for $C_1^*(E_X(u))$ and $C_2^*(E_X(u))$ are computed in [Biermé et al. \(2019\)](#) using the Gaussian Kinematic Formula ([Adler and Taylor, 2007, Theorem 15.9.5](#)). By the results of [Gordon \(1941\)](#) concerning the Mill's ratio of the Gaussian distribution,

$$\frac{2C_1^*(E_X(u))}{C_2^*(E_X(u))} = \frac{\sqrt{\alpha} e^{-u^2/2}}{2(1 - \Phi(u))} \sim \sqrt{\frac{\pi\alpha}{2}} u,$$

where Φ denotes the standard Gaussian cumulative distribution function, and α is as in (5.5).

Therefore, by using Corollary 5.1 with $g(p) = u_p(0)$, the probability density of $uR_0^{(u)}$ just to the right of 0 approaches $\sqrt{\pi\alpha/2}$, as $u \rightarrow \infty$. If one were to show in addition the uniform convergence of the density of the extremal range, one may conclude that $\sqrt{\pi\alpha/2}$ is the limiting value as $r \rightarrow 0$ of the limiting density, as $u \rightarrow \infty$.

In practice, a useful approximation of the distribution function $R_0^{(u)}$ for large u and small r for Gaussian random fields is therefore

$$\mathbb{P}(R_0^{(u)} \leq r) \approx \sqrt{\frac{\pi\alpha}{2}}ur,$$

where an estimate $\hat{\alpha}$ of the second spectral moment α based on a parametric covariance function $\rho(h)$ could be plugged in to obtain an estimate for spatial data corresponding to relatively smooth spatial surfaces.

Regularly varying fields

Here, we recall the core elements of the theory of regularly varying random fields, and the related ℓ -Pareto processes from Ferreira and de Haan (2014); Dombry and Ribatet (2015), commonly used as models for spatial processes conditioned on high threshold exceedances of a certain cost functional.

Let T be a compact domain satisfying $r_T := \sup\{r \in \mathbb{R}^+ : B(0, r) \subseteq T\} > 0$. Let X be a continuous, stationary random field defined on \mathbb{R}^2 , and let $X|_T$ be the random field X restricted to the domain T . Let \mathcal{C}_0 be the set of continuous functions from T to $[0, \infty)$, excluding the constant function 0. Let $\mathcal{S} = \{x \in \mathcal{C}_0 : \|x\|_T = 1\}$, where $\|x\|_T := \sup_{t \in T} x(t)$.

In Appendix 5.A.3, we recall from Dombry and Ribatet (2015) what it means for a random field to be *regularly varying with exponent α and spectral measure σ on \mathcal{S}* . The limiting behaviour of these random fields at high thresholds can be well described by *ℓ -Pareto random fields* (see Lemma 5.5 in Appendix 5.B.3); more recently also called *r -Pareto random fields* with r standing for *risk* (de Fondeville and Davison, 2022). These random fields are characterized by a *cost functional* $\ell : \mathcal{C}_0 \rightarrow [0, \infty)$ that is homogeneous of order 1, *i.e.*, $\ell(ux) = u\ell(x)$ for all $x \in \mathcal{C}_0$ and $u > 0$. The precise definition of an ℓ -Pareto random field is provided in Definition 5.10 in Appendix 5.B.3. For now, we borrow the notation of Dombry and Ribatet (2015) and write $\mathbb{P}_{\alpha, \sigma_\ell}^\ell$ for the set of ℓ -Pareto random fields with exponent α and spectral measure σ_ℓ .

For regularly varying X , it is possible to express the limit distribution of the extremal range in terms of two different constructions of ℓ -Pareto processes.

Proposition 5.3. *Suppose that $X|_T$ is regularly varying with exponent $\alpha > 0$ and spectral measure σ on \mathcal{S} . Define the cost functionals f and g mapping from \mathcal{C}_0 to $[0, \infty)$ by $f : x \mapsto \|x\|_T$ and $g : x \mapsto x(0)$, and let $Y_T^f \in \mathbb{P}_{\alpha, \sigma_f}^f$, and $Y_0^g \in \mathbb{P}_{\alpha, \sigma_g}^g$. Here, $\sigma_f(A) := \sigma(\mathcal{S} \cap A)/\sigma(\mathcal{S})$ and $\sigma_g(A) :=$*

$\frac{1}{c} \int_{\mathcal{S}} x(0)^\alpha \mathbf{1}_{\{x/x(0) \in A\}} \sigma(dx)$ for $A \in \mathcal{B}(\mathcal{C}_0)$, with $c := \int_{\mathcal{S}} x(0)^\alpha \sigma(dx)$. Then, for $r \in (0, r_T)$,

$$\lim_{u \rightarrow \infty} \mathbb{P}(R_0^{(u)} \leq r) = 1 - \frac{\mathbb{E}[\mathcal{L}((E_{Y_T}(1) \cap T)_{-r})]}{\mathcal{L}(T_{-r})\mathbb{P}(Y_T(0) > 1)} = \mathbb{P}(\exists t \in B(0, r) \text{ s.t. } Y_0(t) \leq 1). \quad (5.6)$$

The proof of Proposition 5.3 is postponed to Appendix 5.B.3.

5.4.2 Connections with the tail dependence coefficient

Taking a more non-parametric perspective, we continue using the threshold function u_p as defined in Definition 5.6 that adapts to non-stationary random fields.

Recall that for two sites $s_1, s_2 \in \mathbb{R}^2$, the tail dependence coefficient function of a spatial random field X is defined as $\chi(s_1, s_2) := \lim_{p \rightarrow 1} \chi_p(s_1, s_2)$, where,

$$\chi_p(s_1, s_2) := \frac{\mathbb{P}(X(s_1) > u_p(s_1), X(s_2) > u_p(s_2))}{1 - p} = \mathbb{P}(X(s_1) > u_p(s_1) \mid X(s_2) > u_p(s_2)).$$

Here, we use the following definition of asymptotic (in)dependence. The random field X is said to be *asymptotically independent* if $\chi(s_1, s_2) = 0$ for all $s_1 \neq s_2$, and *asymptotically dependent* if $\chi(s_1, s_2) > 0$, for all $s_1, s_2 \in \mathbb{R}^2$. The asymptotic dependence of X forces the asymptotic behavior of $R_s^{(u)}$ as $u \rightarrow \infty$. Indeed, if X exhibits asymptotic independence, then we have immediately that $R_s^{(u)} \xrightarrow{\mathbb{P}} 0$, as $u \rightarrow \infty$. This simple observation is a corollary of the following proposition.

Proposition 5.4. *Let X be any random field on \mathbb{R}^2 . For all $s \in \mathbb{R}^2$ and all $p \in (0, 1)$,*

$$\mathbb{P}(R_0^{(u_p)} \leq \|s\|) \geq 1 - \chi_p(s, 0).$$

Proof. The event $\{\tilde{R}^{(u_p)}(0) > \|s\|, X(0) > u_p\}$ is contained in the event $\{X(s) > u_p, X(0) > u_p\}$. Therefore, $\mathbb{P}(\tilde{R}^{(u_p)}(0) > \|s\|, X(0) > u_p) \leq \mathbb{P}(X(s) > u_p, X(0) > u_p)$. A division by $\mathbb{P}(X(0) > u_p)$ (equal to $1 - p$ if $X(0)$ has a continuous distribution function) implies $\mathbb{P}(\tilde{R}^{(u_p)}(0) > \|s\| \mid X(0) > u_p) \leq \mathbb{P}(X(s) > u_p \mid X(0) > u_p)$, and the result holds by taking compliments. \square

Therefore, asymptotic dependence is a necessary condition for $R_0^{(u_p)}$ to have a non-degenerate limit distribution as $p \rightarrow 1$. However, it is not sufficient. In Appendix 5.A.2, we study a random field for which $R_0^{(u_p)} \xrightarrow[p \rightarrow 1]{\mathbb{P}} 0$; the following theorem makes an important link between the extremal range and the tail dependence coefficient, and establishes that in this specific case, $\chi(0, s) = 1$ for all $s \in \mathbb{R}^2$.

Theorem 5.2. *Let X be an isotropic random field, and suppose that for $p \in (0, 1)$, Assumption 5.1 is satisfied for the threshold function u_p . Let h be a real function of $p \in (0, 1)$ such that*

$$h(p) \frac{C_2^*(E_X(u_p))}{C_1^*(E_X(u_p))} \xrightarrow[p \rightarrow 1]{} \infty.$$

Then, for any fixed $s \in \mathbb{R}^2$, $\chi_p(s/h(p), 0) \xrightarrow{p \rightarrow 1} 1$.

Proof. Assumption 5.1 provides an alternative set of hypotheses to support the result of Cotsakis et al. (2023a, Theorem 2.1) which states that for $p \in (0, 1)$ and $s \in \mathbb{R}^2$,

$$\frac{1}{q} \mathbb{P}(X(qs) \leq u_p(qs), X(0) > u_p(0)) \xrightarrow{q \rightarrow 0} \frac{C_1^*(E_X(u_p))}{\pi} \|s\|,$$

and the limit is approached from below. Thus, for any $q \in \mathbb{R}^+$, a division by $1 - p$ yields

$$\frac{1 - \chi_p(qs, 0)}{q} \leq \frac{C_1^*(E_X(u_p))}{C_2^*(E_X(u_p))\pi} \|s\|.$$

By setting $q = 1/h(p)$, we find that

$$1 - \chi_p(s/h(p), 0) \leq \frac{C_1^*(E_X(u_p))}{h(p)C_2^*(E_X(u_p))\pi} \|s\| \xrightarrow{p \rightarrow 1} 0.$$

The desired result holds since $\chi_p \in [0, 1]$ for all $p \in (0, 1)$. □

Remark 5.2. Recall that $2C_1^*(E_X(u_p))/C_2^*(E_X(u_p))$ is the limit value in Theorem 5.1, giving the first-order approximation of the cumulative distribution function of the extremal range. For many random fields, this is seemingly the rate at which space should be rescaled as $p \rightarrow 1$ if one is to expect the tail dependence coefficient and the distribution function of the extremal range to stabilize to values strictly between 0 and 1.

5.4.3 A parametric model for the extremal range

We have seen that tail dependence coefficient determines whether a random field is asymptotically dependent or independent according to our definition. While it is natural to study pairwise exceedances in discrete data, an attractive alternative is to describe the spatial dependence of continuous data using the extremal range. The extremal range tends to 0 in probability as higher thresholds are considered for asymptotically independent random fields; see Proposition 5.4. The rate of this convergence provides an alternative, more precise notion of extremal independence. To formalize this idea, we propose the following assumption on the random field X .

Assumption 5.2. For each $s \in \mathbb{R}^2$, the distribution of $R_s^{(u_p)}$ is non-degenerate for $p \in (0, 1)$, and there exists $\theta_s \in [0, \infty)$ such that for all $a > 0$ and $\alpha \in (0, 1)$,

$$\frac{q_\alpha(R_s^{(u_{p'})})}{q_\alpha(R_s^{(u_p)})} \xrightarrow{p \rightarrow 1} a^{-\theta_s},$$

where $p' = 1 - (1 - p)^\alpha$.

Assumption 5.2 is equivalent to demanding that the quantile functions $q_\alpha(R_0^{(u)})$, for $\alpha \in (0, 1)$, $u \in \mathbb{R}^+$, are regularly varying in u after X is transformed to have standard exponential margins. The indices of regular variation are allowed to vary throughout space. For asymptotically independent models, the quantile function $q_\alpha(R_s^{(u_p)})$ tends to decrease as $p \rightarrow 1$, and so one expects θ_s to be large. However, for asymptotically dependent models, the same quantile function approaches a positive constant as $p \rightarrow 1$, in which case, one can expect $\theta_s = 0$. Contrary to other popular measures of asymptotic dependence, θ_s distinguishes between varying degrees of asymptotic *independence*. A consistent, local estimator for θ_s is defined in Section 5.5.2, thus providing a measure of the spatial asymptotic independence at high thresholds.

In addition, Section 5.5.3 outlines how linear quantile regression can be used to extrapolate or interpolate the quantiles of the extremal range when the excursion sets are observed at multiple levels.

Remark 5.3. The Gaussian (resp. regularly varying) random fields studied in Section 5.4.1 satisfy Assumption 5.2 with $\theta_s = 1/2$ (resp. $\theta_s = 0$) for all $s \in \mathbb{R}^2$. Therefore, Assumption 5.2 can be applied to both asymptotically dependent and asymptotically independent models.

5.5 Inference

5.5.1 Empirical CDF of the extremal range for stationary fields

The results developed in Sections 5.3 and 5.4 lead to natural statistical procedures for estimating the extremal range. We begin by describing a simple procedure for strongly mixing sequences of random fields that are assumed to be stationary and identically distributed. Independent replications of the random field are trivially strongly mixing.

Proposition 5.5. *Let $T \subset \mathbb{R}^2$ be compact, and suppose that $\bar{r}_T := \sup\{r \in \mathbb{R}^+ : \mathcal{L}(T_{-r}) > 0\} > 0$. Let X_1, \dots, X_n be a strongly mixing sequence of random fields defined on T , each equal in distribution to the stationary random field X (Rosenblatt, 1956). Define*

$$\tilde{R}_i^{(u)}(t) := \text{dist}(t, (E_{X_i}(u))^c), \quad (5.7)$$

for some threshold $u \in \mathbb{R}$ satisfying $\mathbb{P}(X(0) > u) > 0$, and

$$\hat{F}_n(r) := \begin{cases} \frac{\sum_{i=1}^n \mathcal{L}(\{t \in T_{-r} : 0 < \tilde{R}_i^{(u)}(t) \leq r\})}{\sum_{i=1}^n \mathcal{L}(E_{X_i}(u) \cap T_{-r})} & , \text{ if } \sum_{i=1}^n \mathcal{L}(E_{X_i}(u) \cap T_{-r}) > 0 \\ 0 & , \text{ if } \sum_{i=1}^n \mathcal{L}(E_{X_i}(u) \cap T_{-r}) = 0, \end{cases}$$

for $r \in (0, \bar{r}_T)$. It follows that $\hat{F}_n(r) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{P}(R_0^{(u)} \leq r)$, uniformly for $r \in (0, \bar{r}_T)$.

Proof. Fix $r \in (0, \bar{r}_T)$ and $u \in \mathbb{R}$. By the strongly mixing assumption, the sequences $(\mathcal{L}(\{t \in T_{-r} : 0 < \tilde{R}_i^{(u)}(t) \leq r\}))_{i \geq 1}$ and $(\mathcal{L}(E_{X_i}(u) \cap T_{-r}))_{i \geq 1}$ are mean-ergodic, and so we write

$$\begin{aligned} \hat{F}_n(r) &= \frac{\frac{1}{n} \sum_{i=1}^n \mathcal{L}(\{t \in T_{-r} : 0 < \tilde{R}_i^{(u)}(t) \leq r\})}{\frac{1}{n} \sum_{i=1}^n \mathcal{L}(E_{X_i}(u) \cap T_{-r})} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{\mathbb{E}[\mathcal{L}(\{t \in T_{-r} : 0 < \tilde{R}^{(u)}(t) \leq r\})]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})]} \\ &= \frac{\int_{T_{-r}} \mathbb{P}(0 < \tilde{R}^{(u)}(t) \leq r) dt}{\int_{T_{-r}} \mathbb{P}(X(t) > u) dt} = \frac{\mathbb{P}(\tilde{R}^{(u)}(0) \leq r, X(0) > u)}{\mathbb{P}(X(0) > u)} = \mathbb{P}(R_0^{(u)} \leq r). \end{aligned}$$

Uniform convergence follows from the continuity of the limiting distribution function (see Lemma 5.1) and Dini's theorem (Friedman, 2007, p.199). \square

Proposition 5.5 therefore provides a convenient way to estimate the empirical distribution function of the extremal range under the assumption of stationarity for a series of time-replicated spatial fields.

5.5.2 Consistent estimation of the tail decay rate of the extremal range

Recall that under Assumption 5.2, the quantiles of the extremal range are parameterized by an index of regular variation θ_s for each site s in the study domain, which we call tail decay rate. We propose a local estimator for this quantity and show that it is consistent, providing the corresponding rate of convergence.

Definition 5.7. Let X_1, \dots, X_n be n realizations of the random field X . For $p \in (0, 1)$, define the empirical median of $R_s^{(u_p)}$ to be

$$\widehat{q}_{50\%}^{(n)}(R_s^{(u_p)}) := \operatorname{argmin}_{x \in \mathbb{R}^+} \left\{ \sum_{i=1}^n |x - \tilde{R}_i^{(u_p)}(s)| \mathbb{1}_{\{\tilde{R}_i^{(u_p)}(s) > 0\}} \right\},$$

with $\tilde{R}_i^{(u_p)}(s)$ as in (5.7). For $p_1, p_2 \in (0, 1)$, $p_1 \neq p_2$, define

$$\hat{\theta}_s^{(n)}(p_1, p_2) = \frac{\log(\widehat{q}_{50\%}^{(n)}(R_s^{(u_{p_2})})) - \log(\widehat{q}_{50\%}^{(n)}(R_s^{(u_{p_1})}))}{\log(-\log(1 - p_1)) - \log(-\log(1 - p_2))}. \quad (5.8)$$

If $\tilde{R}_i^{(u_p)}(s) = 0$, $\forall i = 1, \dots, n$, for either $p = p_1$ or $p = p_2$, then both $\hat{\theta}_s^{(n)}(p_1, p_2)$ and the empirical median $\widehat{q}_{50\%}^{(n)}(R_s^{(u_p)})$ are understood to be 0.

Proposition 5.6. Let X_1, \dots, X_n be n independent realizations of the random field X , satisfying Assumption 5.2. Fix $p_0 \in (0, 1)$, and let $(p_n)_{n \geq 1}$ be a sequence in $(0, 1)$ tending to 1 such that $n(1 - p_n) \rightarrow \infty$ as $n \rightarrow \infty$. Then, for each $s \in \mathbb{R}^2$, the estimator defined in (5.8) satisfies

$$\hat{\theta}_s^{(n)}(p_0, p_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta_s,$$

where $\widehat{\theta}_s^{(n)}(p_0, p_n)$ is defined in Definition 5.7, and θ_s is defined in Assumption 5.2.

The proof of Proposition 5.6 can be found in Appendix 5.B.4. The estimator in (5.8) can be viewed as the slope of a median regression with response $\log(\widehat{q}_{50\%}^{(n)}(R_s^{(u_p)}))$ and covariate $\log(-\log(1-p))$. This motivates the following median regression approach for predicting the median extremal range at very extreme probability levels p close to 1.

5.5.3 Extrapolation for non-stationary data using median regression

As we have seen previously, under Assumption 5.1, the random excursion set $E_X(u)$ is stationary. This implies that the distribution of $R_s^{(u)}$ is invariant in the spatial location s , allowing one to extract information from the entire spatial domain to gain insights on the behavior of the extremal range at s . When $E_X(u)$ is non-stationary, it is likely that there is insufficient information to infer the entire distribution of $R_s^{(u)}$ (especially when $\mathbb{P}(X(s) > u(s))$ is small). Thus, it is reasonable to instead estimate a summary statistic of $R_s^{(u)}$. The median, $q_{50\%}(R_s^{(u)})$, is an appropriate statistic to estimate for several reasons: it is robust to censorship of large observations and to strongly discretized small values arising with data available on pixel grids; the median commutes with monotonic rescalings; one does not require the existence of moments of the distribution of the extremal range; the behavior of the median is controlled under Assumption 5.2 with $\alpha = 1/2$.

Definition 5.8. The Median Extremal Range (MER) at a site $s \in \mathbb{R}^2$ at the threshold $u_p(s)$, for $p \in (0, 1)$, is defined as

$$\text{MER}(s; p) := q_{50\%}(R_s^{(u_p)}).$$

Suppose that for several realizations X_1, \dots, X_n of X , the excursion set is observed at several high thresholds $u^{(1)}, \dots, u^{(k)}$. That is, we observe the sets $E_{X_i}(u^{(j)})$ for $(i, j) \in \{1, \dots, n\} \times \{1, \dots, k\}$. Under Assumption 5.2,

$$\frac{\log(\text{MER}(s; p')) - \log(\text{MER}(s; p))}{\log(-\log(1-p')) - \log(-\log(1-p))} \xrightarrow{p \rightarrow 1} -\theta_s, \quad (5.9)$$

for any $a \in (0, 1)$ (see Assumption 5.2). Thus, plotted on a log-log plot, the slope of the graph of the $\text{MER}(s; p)$ against $-\log(1-p)$ tends to a constant $-\theta_s$, as $p \rightarrow 1$. This justifies the use of quantile regression to estimate the median of our empirical observations of $\log \widetilde{R}^{(u_p)}(s)$ against $\log(-\log(1-p))$ for several $p \in (0, 1)$ moderately close to 1. In this way, the height of the resulting regression line provides a model for the median of $\log R^{(u_p)}(s)$. Exponentiating the regression line provides a model for the $\text{MER}(s; p)$ for arbitrarily large $p \in (0, 1)$ and therefore allows us to extrapolate extremal-range properties at very high quantile levels for which only few or no exceedances at all are available in the data. Another important benefit of median regression that we can include covariates, such as the spatial coordinates, into the intercept and slope, and we can estimate nonlinear covariate effects thanks to generalized additive quantile regression, as highlighted in the following data application.

5.6 Application to gridded temperature data

We use the estimation methods introduced in Section 5.5 to identify the local spatial extent of extremes of daily average temperatures in climate model outputs and provide insights into their spatial patterns. Temperature is a weather variable that is known to vary relatively smoothly in space (Perkins et al., 2012). We study two datasets available for mainland France for a regular pixel grid at 8km resolution in the metric Lambert-II projection. The primary dataset is based on the SAFRAN reanalysis (Vidal et al., 2010) and spans the 1991–2020 period excluding 1997 and 1998. Reanalysis is climate model simulation conditioned on observational data, and is routinely used in climate-change impact studies as a proxy for real weather and climate. We compare our results for the reanalysis model to those obtained for temperatures simulated for the historical period 1951–2005 using a couple of Global-Circulation-Model and Regional-Climate-Model (*IPSL-WRF*), one of the reference models provided by the French weather service for studying climate change impacts (<http://www.drias-climat.fr/>), with data available on the same spatial grid as the SAFRAN reanalysis. We perform the same analyses for each of the two datasets, so that any differences in the results are due to statistical uncertainties (that we assess), and to fundamental differences in the distributional properties of the datasets.

We consider only data for summer months (June 1 to August 31) and assume temporal stationarity. We first estimate $u_p(s)$, the p -quantile of $X(s)$, for $p \in \{0.85, 0.86, \dots, 0.98\}$, using standard methods, for each location s separately. As we have preselected the summer months, during which the majority of temperature extremes typically manifest, it is justifiable to use a set of quantile levels including values that are relatively low for extreme-value analysis. This provides an estimate of the excursion set $E_X(u_p)$ for each day; see Figure 5.2. For each p and each pixel inside an excursion set, we estimate $\tilde{R}^{(u_p)}$ using the fast marching method (Sethian, 1996) implemented in the R package `fastmarching` (Silva and Steele, 2014). To manage the absence of temperature data outside of France, we define our estimate of $\tilde{R}^{(u_p)}(s)$ to be the smallest distance from s to a site s' in mainland France with $X(s') \leq u_p(s')$. In terms of boundary effects, this is equivalent to considering the modified excursion set $E_X(u_p) \cup T^c$ with T corresponding to mainland France. The histograms of the empirical distribution of extremal ranges $\tilde{R}^{(u_p)}(s)$ —provided that $\tilde{R}^{(u_p)}(s) > 0$ —is shown in Figure 5.1 for the quantile levels $p = 0.85$ and $p = 0.98$, for all locations pooled together. We see a clear shift towards generally smaller extremal ranges at the higher threshold, which means that dependence strength seems to decrease with the quantile level. Moreover, reanalysis data tend to have clearly larger extremal ranges than simulation data at both quantile levels, which hints at structural differences in the geometries of extremes among these two types of data, with generally smaller spatial clusters of extremes in the simulation data.

Next, we more precisely estimate the tail decay behavior of the extremal range and its spatial variability. Operating under Assumption 5.2, we estimate the parameters of a model for the $\text{MER}(s; p)$;

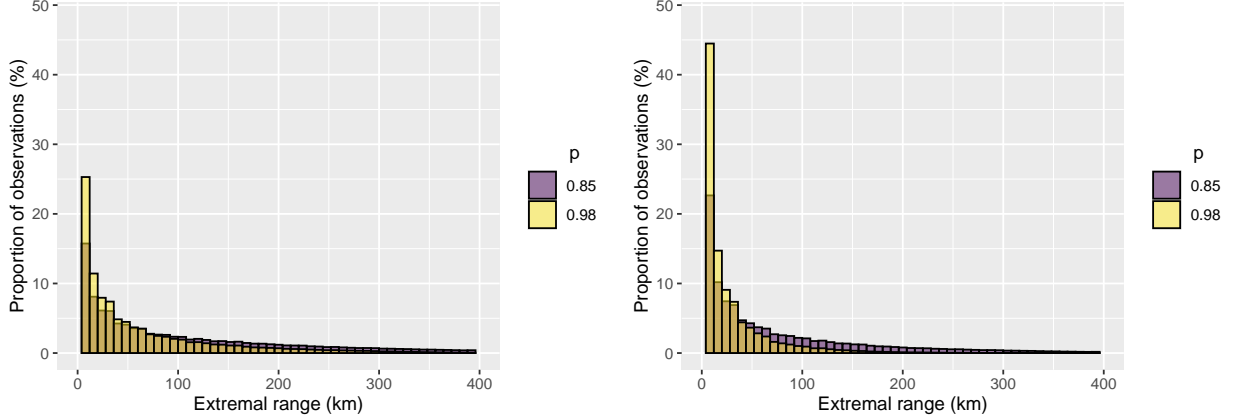


Figure 5.1: Histograms of estimated extremal ranges at two quantile levels for SAFRAN reanalysis (left) and IPSL-WRF simulations (right).

see Definition 5.8. The model, as suggested by (5.9), is given by

$$\log(\text{MER}(s; p)) = \beta_s - \theta_s \log(-\log(1 - p)), \quad (5.10)$$

where the model parameters β_s and θ_s are allowed to vary over space.

Using the set of estimates of $\tilde{R}^{(u_p)}$ for all p and all pixel-days in an excursion set, we use $\log \tilde{R}^{(u_p)}(s)$ —provided that $\tilde{R}^{(u_p)}(s) > 0$ —as the response variable in a generalized additive median regression with covariates s and $x := \log(-\log(1 - p))$. We implement a generalized additive model with spline tensor products defined over spatial coordinates for the parameters β_s and θ_s , such that their estimates vary smoothly in space (Wood, 2017), *i.e.*, we perform a quantile regression with space-varying intercept and slope parameters. This modeling strategy would further allow for incorporating other covariates, such as altitude, or time to account for temporal trends.

We provide several diagnostic plots to illustrate the quantile regression fit. Figure 5.3 shows a satisfactory fit of the regression line for the SAFRAN data at two pixels s_1, s_2 with quite different intercept $\hat{\beta}_s$ and slope $\hat{\theta}_s$. Figure 5.4 provides maps of the estimates of the $\text{MER}(s; p)$ based on model (5.10) for two values of p . Standard deviations of estimates are computed using a block jackknife based on blocks given by years, where the whole chain of estimation is repeated for each jackknife sample. The resulting estimates of the $\text{MER}(s; p)$ correspond well with certain topographical features, even though this information was not provided to the model. The extremal range tends to be larger in planar regions, while it tends to be smaller in mountainous regions, and near the Atlantic coast to the East.

Moreover, all jackknife-based estimates of the negative slopes θ_s , *i.e.*, of the tail decay rate, were positive, such that the signal against asymptotic dependence is significant around all pixels s . Figure 5.5 depicts the estimates of θ_s for the model (5.10), and their standard deviations. This parameter quantifies the joint tail decay rate near each location; see Section 5.4.3. As discussed

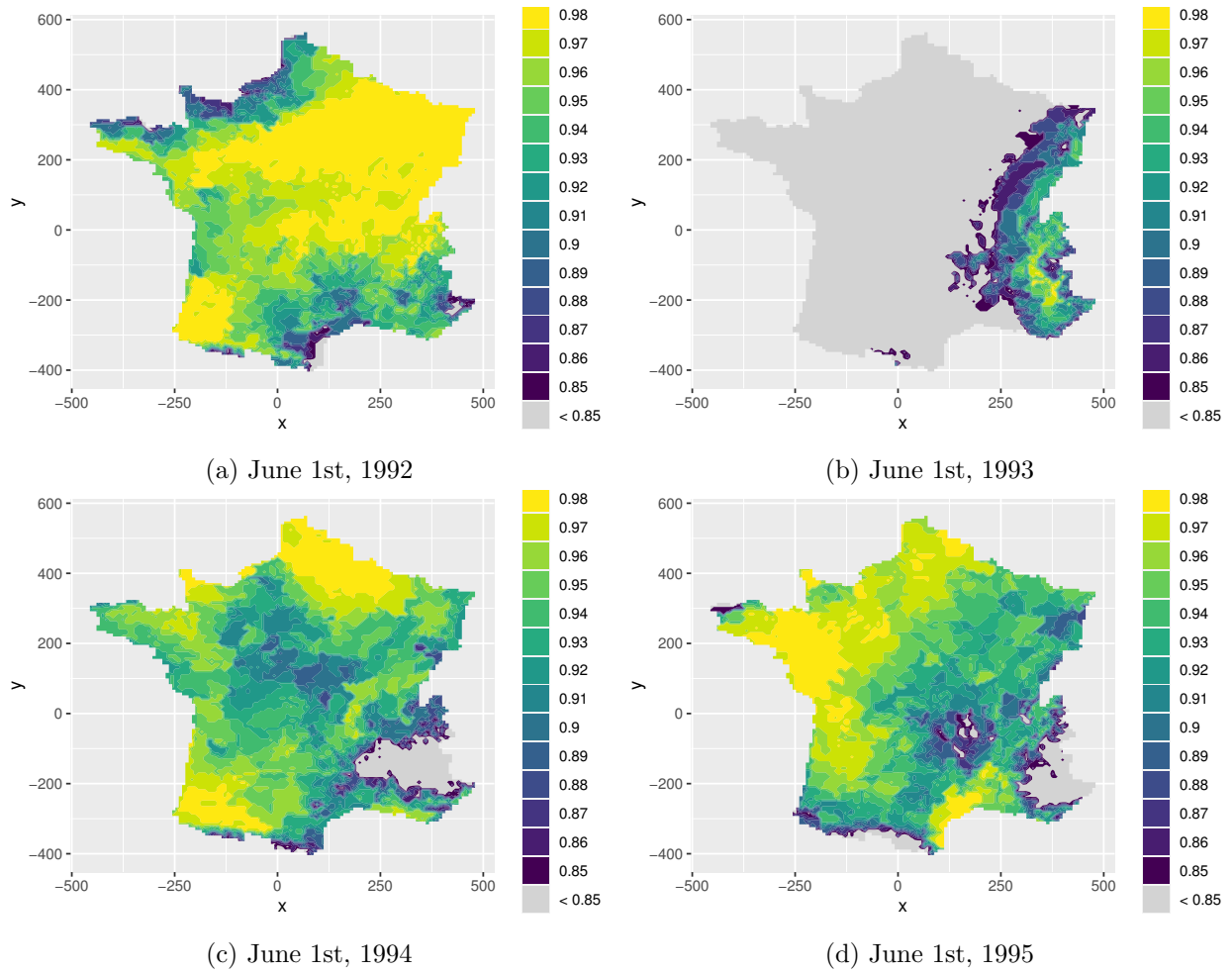


Figure 5.2: Realizations of nested excursion sets $E_X(u_p)$ for $p \in \{0.85, 0.86, \dots, 0.98\}$, for each of the four indicated dates in the SAFRAN reanalysis.

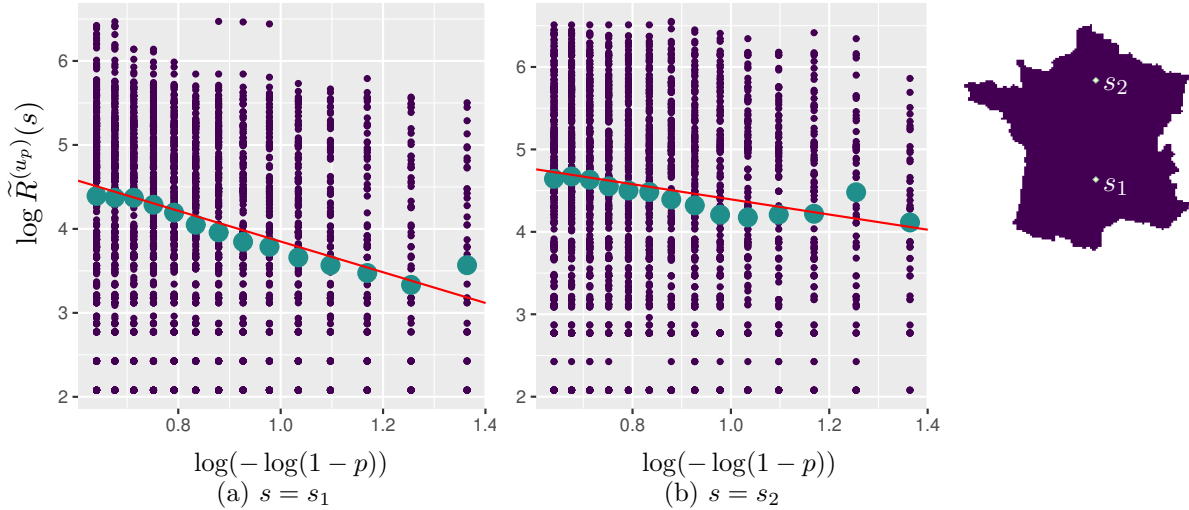


Figure 5.3: Illustration of median regression of the extremal range for SAFRAN reanalysis data. Fitted regression lines (red) of the model (5.10) for the log-median of the extremal range at the two sites s shown in the map on the right. For $p \in \{0.85, 0.86, \dots, 0.98\}$, we record a small blue point for all days where $\tilde{R}^{(u_p)}(s) > 0$, *i.e.*, where $X(s) > u_p(s)$. For each p , the empirical median of $\log \tilde{R}^{(u_p)}(s)$ is plotted as a larger green point to aid with visual diagnostic of the model fit. Multiplicity of points at discrete values is not shown.

previously, $\theta_s \leq 0$ implies asymptotic dependence, whereas $\theta_s > 0$ is a strong indicator of asymptotic independence, as the implication is in the reverse direction (see Proposition 5.4).

The results obtained for IPSL-WRF simulated temperatures differ quite drastically from those obtained for the SAFRAN reanalysis data. Estimated extremal ranges are substantially smaller for simulated temperatures and show less spatial variability and also differences in spatial patterns. Unless these differences arise from non-stationarity in time, our study demonstrates that the IPSL-WRF climate model has strong biases in the extremal range of temperatures when compared to the observation-based SAFRAN reanalysis.

5.7 Conclusion

The extremal range, particularly the evolution of the $\text{MER}(s; p)$ as $p \in (0, 1)$ tends to 1, quantifies the degree of asymptotic independence locally at the site s . It aids flexible exploratory analysis of dependence in spatial extremes beyond the mathematically elegant but rigid framework of asymptotically stable dependence in max-stable processes and other regularly varying processes. We have opted for the median as a summary parameter that allows for simple interpretation as a level that is on average exceeded in half of all cases of marginal exceedance at the reference location. It further allows for relatively robust estimation since it is less influenced by pixellization biases arising for small extremal ranges at lower quantiles and by boundary effects arising for large extremal ranges at higher quantiles. The proposed quantile regression approach further offers the possibility to in-

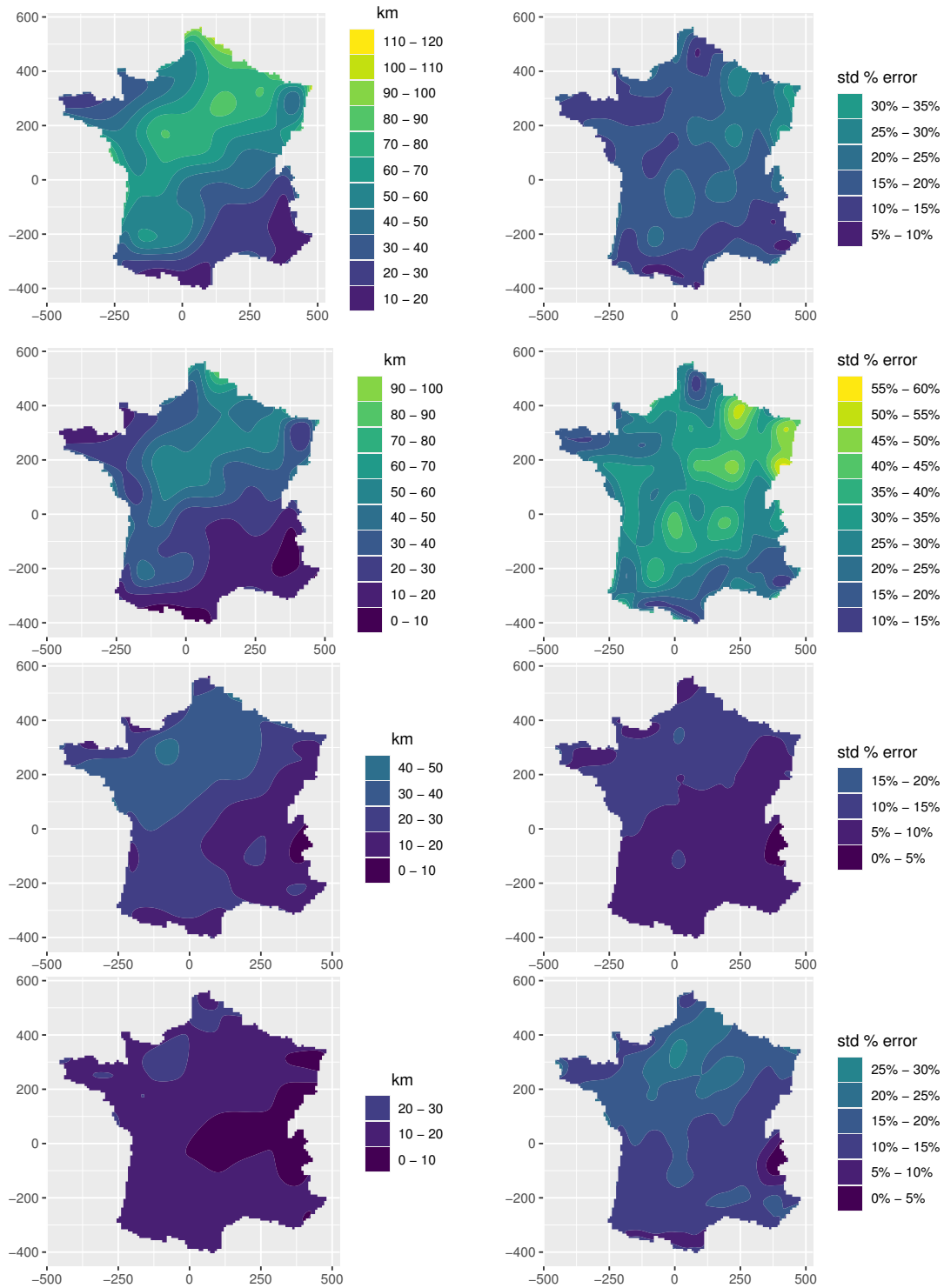


Figure 5.4: Estimates of the $MER(s; p)$ with relative standard errors (in %) for the SAFRAN re-analysis dataset for $p = 0.95$ (first row) and $p = 1 - 1/92 \approx 0.989$ (second row, corresponding to a one-year return level). The same quantities are computed for the IPSL-WRF simulated temperatures for $p = 0.95$ (third row) and $p = 1 - 1/92 \approx 0.989$ (fourth row). Color scales are the same across datasets and quantile levels.

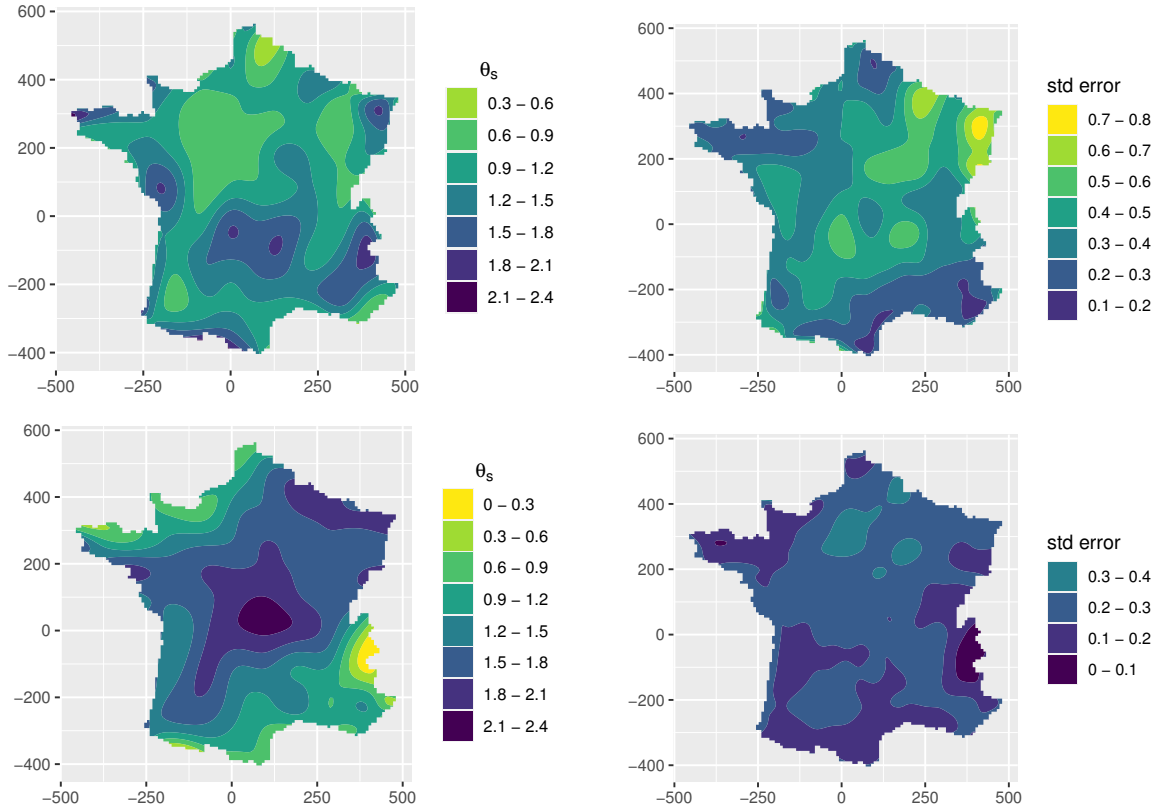


Figure 5.5: Estimates of the tail decay rate θ_s and its standard error for the SAFRAN reanalysis data (first row) and the IPSL-WRF simulated data (second row). The estimate is the negative slope of the model line in Figure 5.3. The larger θ_s , the less asymptotic dependence there is at the site s ; see Section 5.4.3. Color scales are the same for the two datasets.

clude other covariates to model and explain non-stationary extremal dependence, such as temporal trends due to climate change. Our estimation methods scale well even for large gridded datasets with hundreds of thousands of pixels. By contrast with the common bivariate statistics defined as a function of spatial distance, our approach avoids the rather arbitrary and heuristic selection of location pairs to be considered for inferential purposes.

Data on regular grids at relatively high resolution, such as climate model output, provide a good approximation of the framework of continuous geographic space underpinning the concept of the extremal range. If data are not in this form, interpolation techniques such as kriging, piecewise linear interpolation based on Delaunay triangulation, or more sophisticated basis-function methods, could be applied first before conducting estimation of the extremal range.

While our focus was on descriptive statistics here, the extremal range could also be useful for generative modeling based on spatial random fields, for example by using $\text{MER}(s; p)$ as a goodness-of-fit diagnostics or as a covariate in the dependence structure.

The extremal range is a tool at the intersection of spatial EVT, stochastic geometry and topological data analysis. Further research in this area could help foster a high-dimensional statistical learning toolbox for studying complex structures in large volumes of climate data.

Appendix

5.A Technical definitions and examples

5.A.1 Positive reach

Definition 5.9. (Federer, 1959) The reach of a set $S \subseteq \mathbb{R}^2$ is given by $\sup\{r \in \mathbb{R} : \forall x \in S_r, \exists! s \in S \text{ nearest to } x\}$. A subset of \mathbb{R}^2 is termed *positive reach* if its reach is positive.

Recall from Federer (1959) that a closed set is convex if and only if its reach is infinite. Therefore, the empty set is trivially a positive reach set.

5.A.2 A case where asymptotic dependence is not captured by the extremal range

Let $E \sim \text{Exp}(1)$, $\theta \in \text{Unif}([0, 2\pi])$, and $U \sim \text{Unif}([0, 1]^2)$ be independent, and consider the stationary, isotropic random field $\{X(t)\}_{t \in \mathbb{R}^2}$ defined by

$$X(t) = E \left(1 - \sum_{s \in \mathbb{Z}^2} \mathbb{1}_{\{\|s+U-ER_\theta(t)\| < 3^{-E}\}} \right),$$

where $R_\theta(t)$ is the image of t rotated by an angle θ about the origin. Notice that X satisfies Assumption 5.1, for if the excursion set is not empty, it is the complement of the union of disks of

radius 3^{-E} with centers on a randomly oriented square grid with spacing E^{-1} . If for some $p \in (0, 1)$ we have $X(0) > u_p$, then $E > u_p$ and $X(R_{-\theta}(U)/E) = 0 \leq u_p$. Thus, $\tilde{R}^{(u_p)}(0) < \|U\|/E < \sqrt{2}/u_p$ which tends to 0 as $p \rightarrow 1$. Thus, $R_0^{(u_p)} \xrightarrow[p \rightarrow 1]{\mathbb{P}} 0$. It is not hard to check that

$$\frac{C_2^*(E_X(u_p))}{C_1^*(E_X(u_p))} \xrightarrow[p \rightarrow 1]{} \infty,$$

contrary to the behaviour of $R_0^{(u_p)}$.

5.A.3 Regularly varying random fields

The process $X|_T$ is said to be *regularly varying* with exponent α and spectral measure σ on \mathcal{S} if there exists a function $a : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $a(u) \rightarrow \infty$ and

$$u\mathbb{P}\left(\frac{1}{\|X\|_T}X|_T \in A, \|X\|_T > ra(u)\right) \rightarrow r^{-\alpha}\sigma(A)$$

as $u \rightarrow \infty$, for all $r > 0$ and $A \in \mathcal{B}(\mathcal{S})$ satisfying $\sigma(\partial A) = 0$ (see Definition 1 in [Dombry and Ribatet \(2015\)](#)). If $X|_T$ is regularly varying with exponent α and spectral measure σ , we write $X|_T \in \text{RV}_{\alpha, \sigma}(\mathcal{C}_0)$.

5.B Proofs and auxiliary results

5.B.1 Proofs for Section 5.3

Proof of Proposition 5.1. Since $R_0^{(u)}$ is almost surely non-negative, it suffices to check that (5.1) holds for $r \geq 0$. Note that by the continuity of X , the excursion set $E_X(u)$ is open, and so the events $\{\tilde{R}_0^{(u)} > r\}$ and $\{B(0, r) \subset E_X(u)\}$ are equal. Furthermore,

$$\begin{aligned} \mathbb{P}(R_0^{(u)} > r) &= \mathbb{P}(B(0, r) \subset E_X(u) \mid X(0) > u(0)) = \frac{\mathcal{L}(T)\mathbb{P}(B(0, r) \subset E_X(u))}{\mathcal{L}(T)\mathbb{P}(X(0) > u(0))} \\ &= \frac{\mathbb{E}\left[\int_T \mathbb{1}_{\{B(t, r) \subset E_X(u)\}} dt\right]}{\mathbb{E}\left[\int_T \mathbb{1}_{\{X(t) > u(t)\}} dt\right]} = \frac{\mathbb{E}[\mathcal{L}(\{t \in T : B(t, r) \subset E_X(u)\})]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]} \\ &= \frac{\mathbb{E}[\mathcal{L}(E_X(u)_{-r} \cap T)]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]}. \end{aligned}$$

□

Lemma 5.2. *Suppose that X and the corresponding threshold function u_p for some $p \in (0, 1)$ satisfy Assumption 5.1. Then Equation (5.1) holds with u_p in place of u for any compact $T \subset \mathbb{R}^2$ with positive Lebesgue measure, and any $r \geq 0$.*

Proof. The proof corresponds exactly with that of Proposition 5.1. \square

Proof of Lemma 5.1. Let $r > 0$ and let $T \subset \mathbb{R}^2$ be compact with positive Lebesgue measure. Define $A := \{s \in \mathbb{R}^2 : \text{dist}(s, E_X(u)^c) = r\}$, and let $x \in A$. For each $\epsilon \in (0, r)$, the ball $B(x, \epsilon)$ contains an open ball \tilde{B} of radius $\epsilon/2$ such that for all $y \in \tilde{B}$, one has $\text{dist}(y, E_X(u)^c) < r$. In particular, $\tilde{B} \cap A = \emptyset$, and so the Lebesgue density of A at x cannot exceed $3/4$. By the Lebesgue differentiation theorem, the Lebesgue density of A at s is 1 for almost every $s \in A$. Since there are no elements of A for which this holds, $\mathcal{L}(A) = 0$.

Suppose the statement of the Lemma is false. Then, there exists an $r > 0$ such that $\mathbb{P}(\tilde{R}^{(u)}(0) = r \mid X(0) > u) > 0$, or equivalently, $\mathbb{P}(\tilde{R}^{(u)}(0) = r) > 0$. By stationarity and Fubini's theorem,

$$0 < \mathbb{P}(\tilde{R}^{(u)}(0) = r) = \mathbb{E} \left[\int_T \mathbf{1}_{\{\tilde{R}^{(u)}(s)=r\}} \, ds \right] = \mathbb{E}[\mathcal{L}(A)].$$

Hence, we have arrived at a contradiction, as $\mathbb{P}(\mathcal{L}(A) > 0) = 0$. \square

The following lemmas are useful in the proof of Theorem 5.1 given below.

Lemma 5.3. *Let $A \subseteq \mathbb{R}^2$ and let $T \subset \mathbb{R}^2$ be compact, and let $r > 0$. The following identities hold.*

- $\mathcal{L}(A^c \cap T) = \mathcal{L}(T) - \mathcal{L}(A \cap T)$,
- $(A \cap T)_{-r} = A_{-r} \cap T_{-r} = ((A^c \cap T)_r)^c \cap T_{-r}$.

Proof. The first item holds by the additivity of the Lebesgue measure, and the two facts $(A^c \cap T) \cap (A \cap T) = \emptyset$ and $(A^c \cap T) \cup (A \cap T) = T$.

Now to prove the second item, note that $(S^c)_r = (S_{-r})^c$ for any set $S \subseteq \mathbb{R}^2$ (Cotsakis, 2023, Lemma 1). Now, by De Morgan's laws,

$$\begin{aligned} ((A \cap T)_{-r})^c &= (A^c \cup T^c)_r = \bigcup_{x \in A^c \cup T^c} B(x, r) = \left(\bigcup_{x \in A^c} B(x, r) \right) \cup \left(\bigcup_{y \in T^c} B(y, r) \right) \\ &= (A^c)_r \cup (T^c)_r = (A_{-r})^c \cup (T_{-r})^c = (A_{-r} \cap T_{-r})^c. \end{aligned}$$

This proves the first equality by taking complements. For the second equality, write

$$((A^c \cap T)_r)^c \cap T_{-r} = (A \cup T^c)_{-r} \cap T_{-r} = ((A \cup T^c) \cap T)_{-r} = (A \cap T)_{-r}.$$

\square

Lemma 5.4. *Let $T \subset \mathbb{R}^2$ be compact. It holds for $r > 0$ that*

$$\mathcal{L}(E_X(u) \cap T_{-r}) - \mathcal{L}((E_X(u) \cap T)_{-r}) \leq \frac{9\pi r}{4} \left(\ell(\partial E_X(u) \cap T) + N_T^{(u)} r \right), \quad (5.11)$$

where $N_T^{(u)}$ is as defined in Definition 5.4.

Proof. By Lemma 5.3, the LHS of Equation (5.11) is equal to the Lebesgue measure of $(E_X(u) \cap T_{-r}) \setminus E_X(u)_{-r}$. Remark that for $r > 0$,

$$(E_X(u) \cap T_{-r}) \setminus E_X(u)_{-r} \subseteq (\partial E_X(u))_r \cap T_{-r} \subseteq (\partial E_X(u) \cap T)_r.$$

By following the construction in (Cotsakis et al., 2023c, Lemma 1), let $\gamma^{(i)}$ be the i^{th} connected component of $\partial E_X(u) \cap T$. It is possible to write

$$\gamma^{(i)} = \bigcup_{j=1}^{\lfloor \ell(\gamma^{(i)})/r \rfloor + 1} \beta^{(i,j)},$$

with $\ell(\beta^{(i,j)}) \leq r$ for $j \in \{1, \dots, \lfloor \ell(\gamma^{(i)})/r \rfloor + 1\}$. Now,

$$\partial E_X(u) \cap T = \bigcup_{i=1}^{N_T^{(u)}} \bigcup_{j=1}^{\lfloor \ell(\gamma^{(i)})/r \rfloor + 1} \beta^{(i,j)},$$

and so

$$(\partial E_X(u) \cap T)_r \subseteq \bigcup_{i=1}^{N_T^{(u)}} \bigcup_{j=1}^{\lfloor \ell(\gamma^{(i)})/r \rfloor + 1} \beta_r^{(i,j)}.$$

With the length of each $\beta^{(i,j)}$ is bounded by r , it follows that each $\beta_r^{(i,j)}$ is contained in a ball of radius $3r/2$. Therefore,

$$\begin{aligned} \mathcal{L}\left(\bigcup_{i=1}^{N_T^{(u)}} \bigcup_{j=1}^{\lfloor \ell(\gamma^{(i)})/r \rfloor + 1} \beta_r^{(i,j)}\right) &\leq \left(\frac{\ell(\partial E_X(u) \cap T)}{r} + N_T^{(u)}\right) \pi(3r/2)^2 \\ &= \frac{9\pi r}{4} \left(\ell(\partial E_X(u) \cap T) + N_T^{(u)} r\right). \end{aligned}$$

□

Proof of Theorem 5.1. Let $T \subset \mathbb{R}^2$ be a compact, convex set, and fix $r > 0$. By Lemma 5.3, for all $r > 0$,

$$\begin{aligned} \mathcal{L}((E_X(u) \cap T)_{-r}) &= \mathcal{L}(T_{-r}) - \mathcal{L}((E_X(u)^c \cap T)_r \cap T_{-r}) \\ &= \mathcal{L}(T_{-r}) - \mathcal{L}((E_X(u)^c \cap T)_r) + \mathcal{L}((\partial T)_r \cap A^{(r,T)}), \end{aligned}$$

for some $A^{(r,T)} \subset \mathbb{R}^2$. Now, $E_X(u)^c \cap T$ has positive reach almost surely, and for each positive r

smaller than the reach of $E_X(u)^c \cap T$, there is a corresponding set $A^{(r,T)}$ such that

$$\begin{aligned} \mathcal{L}((E_X(u) \cap T)_{-r}) &= \mathcal{L}(T_{-r}) - \left(\mathcal{L}(E_X(u)^c \cap T) + \ell(\partial(E_X(u)^c \cap T))r + \chi(E_X(u)^c \cap T)\pi r^2 \right) \\ &\quad + \mathcal{L}((\partial T)_r \cap A^{(r,T)}) \\ &= \mathcal{L}(T_{-r}) - \mathcal{L}(T) + \mathcal{L}(E_X(u) \cap T) - \mathcal{L}(E_X(u) \cap T_{-r}) + \mathcal{L}(E_X(u) \cap T_{-r}) \\ &\quad - \ell(\partial(E_X(u) \cap T))r - \chi(E_X(u)^c \cap T)\pi r^2 + \mathcal{L}((\partial T)_r \cap A^{(r,T)}), \end{aligned}$$

almost surely. Therefore,

$$\limsup_{r \rightarrow 0} \left| \frac{\mathcal{L}(E_X(u) \cap T_{-r}) - \mathcal{L}((E_X(u) \cap T)_{-r})}{r} - \ell(\partial(E_X(u) \cap T)) \right| \leq 2\ell(\partial T), \quad (5.12)$$

almost surely. Together, Assumption 5.1 and Lemma 5.4 verify the hypotheses of the reverse Fatou lemma applied to (5.12), and so one achieves

$$\limsup_{r \rightarrow 0} \left| \frac{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})] - \mathbb{E}[\mathcal{L}((E_X(u) \cap T)_{-r})]}{r} - \mathbb{E}[\ell(\partial(E_X(u) \cap T))] \right| \leq 2\ell(\partial T),$$

and by Lemma 5.2,

$$\limsup_{r \rightarrow 0} \left| \frac{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})] \mathbb{P}(R_0^{(u)} \leq r)}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]r} - \frac{\mathbb{E}[\ell(\partial(E_X(u) \cap T))]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]} \right| \leq \frac{2\ell(\partial T)}{\mathcal{L}(T)\mathbb{P}(X(0) > u)},$$

by a division of $\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]$. Recall that T is arbitrary and so can be taken arbitrarily large such that its perimeter length is negligible to its area. This implies

$$\lim_{T \nearrow \mathbb{R}^2} \lim_{r \rightarrow 0} \frac{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})] \mathbb{P}(R_0^{(u)} \leq r)}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]r} = \frac{2C_1^*(E_X(u))}{C_2^*(E_X(u))}. \quad (5.13)$$

The result holds since, for any T ,

$$\frac{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T)]} = \frac{\mathcal{L}(T_{-r})}{\mathcal{L}(T)} \xrightarrow{r \rightarrow 0} 1.$$

□

5.B.2 Proof of Proposition 5.2

In Kac and Slepian (1959), it is shown that for a one-dimensional centered Gaussian process $\{Y(t)\}_{t \in \mathbb{R}}$ having the stationary covariance function in (5.5), it holds that

$$u(Y(t/u) - u) \mid \{Y(0) = u, Y'(0) > 0\} \xrightarrow[u \rightarrow \infty]{d} -\frac{\alpha}{2}t^2 + \xi_\alpha t,$$

where the convergence holds for the finite dimensional distributions of the process (in t), and ξ_α is some random variable that depends on α and the sense of the conditioning on the event $Y(0) = u$, but not on t . Therefore, $u(Y(t/u) - u) \mid Y(0) > u$ converges in the same sense to $-\frac{\alpha}{2}t^2 + \xi_1 t + \xi_2$, where ξ_1 and ξ_2 are random variables, and $\xi_2 > 0$ almost surely.

Now, focusing on the two-dimensional random field X , we see that X evaluated on any one-dimensional affine linear subspace of \mathbb{R}^2 is a Gaussian process, and so the analysis in the preceding paragraph applies to these processes. Therefore, seen as a process in t ,

$$u(X(t/u) - u) \mid X(0) > u \xrightarrow[u \rightarrow \infty]{d} -\frac{\alpha}{2}\|t\|^2 + \langle \tilde{\xi}_\alpha, t \rangle + \xi, \quad (5.14)$$

for some random vector $\tilde{\xi}_\alpha$ that depends on α but not on t , and for some almost surely positive random variable ξ . One can show that $\xi \sim \text{Exp}(1)$ independently of α and t .

Now,

$$\begin{aligned} uR_0^{(u)} &\stackrel{d}{=} \sup \left\{ r \in \mathbb{R}^+ : B(0, r/u) \subset E_X(u) \right\} \mid X(0) > u \\ &\stackrel{d}{=} \sup \left\{ r \in \mathbb{R}^+ : B(0, r) \subset uE_X(u) \right\} \mid X(0) > u \\ &\stackrel{d}{=} \sup \left\{ r \in \mathbb{R}^+ : B(0, r) \subset E_{X(\cdot/u)}(u) \right\} \mid X(0) > u. \end{aligned}$$

Equation (5.14) implies that $E_{X(\cdot/u)}(u) \mid X(0) > u$ converges to a random disk containing the origin as $u \rightarrow \infty$, which finishes the proof. □

5.B.3 Proofs and definitions for Section 5.4.1

Definition 5.10 (Definition 4 in [Dombry and Ribatet \(2015\)](#)). The random field $W : \Omega \times T \rightarrow \mathbb{R}$ is an ℓ -Pareto random field with exponent $\alpha \in \mathbb{R}^+$ and spectral measure σ_ℓ if

- $\mathbb{P}(W \in \mathcal{C}_0) = 1$,
- $\mathbb{P}(\ell(W) > u) = u^{-\alpha}$, for all $u > 1$,
- $W/\ell(W)$ and $\ell(W)$ are independent, and
- $\sigma_\ell(A) = \mathbb{P}(W/\ell(W) \in A)$ for $A \in \mathcal{B}(\mathcal{C}_0)$.

In this case, we write $W \sim P_{\alpha, \sigma_\ell}^\ell$

The link between regularly varying random fields and ℓ -Pareto random fields is made by the following lemma.

Lemma 5.5 (Theorem 3 in [Dombry and Ribatet \(2015\)](#)). Let $X|_T \in \text{RV}_{\alpha, \sigma}(\mathcal{C}_0)$ with exponent $\alpha > 0$ and spectral measure σ on \mathcal{S} . Let $\ell : \mathcal{C}_0 \rightarrow [0, \infty)$ be a homogeneous cost functional that is

continuous at the origin, and is nonzero on a subset of \mathcal{S} with positive σ measure. Let $W \sim \mathbb{P}_{\alpha, \sigma}^\ell$ where

$$\sigma_\ell(A) := \frac{1}{c} \int_{\mathcal{S}} \ell(x)^\alpha \mathbf{1}_{\{x/\ell(x) \in A\}} \sigma(dx), \quad A \in \mathcal{B}(\mathcal{C}_0),$$

with $c := \int_{\mathcal{S}} \ell(x)^\alpha \sigma(dx)$. Then,

$$\mathbb{P}(u^{-1}X|_T \in A \mid \ell(X|_T) > u) \xrightarrow{u \rightarrow \infty} \mathbb{P}(W \in A), \quad A \in \mathcal{B}(\mathcal{C}_0).$$

Proof of Proposition 5.3. We begin by showing the second equality in (5.6). By the continuity of X , the excursion set $E_X(u)$ is open, and so the events $\{\tilde{R}_0^{(u)} > r\}$ and $\{B(0, r) \subset E_X(u)\}$ are equal. Also, $E_X(u) = E_{X/u}(1)$, and so for $r < r_T$,

$$\begin{aligned} \mathbb{P}(R_0^{(u)} > r) &= \mathbb{P}(B(0, r) \subset E_X(u) \mid X(0) > u) = \mathbb{P}(B(0, r) \subset E_{X/u}(1) \mid X(0) > u) \\ &\xrightarrow{u \rightarrow \infty} \mathbb{P}(B(0, r) \subset E_{Y_0}(1)) = 1 - \mathbb{P}(\exists t \in B(0, r) : Y_0(t) \leq 1). \end{aligned} \quad (5.15)$$

The convergence in (5.15) holds by Lemma 5.5. To show the first equality in (5.6), remark that $(E_{Y_T}(u) \cap T)_{-r} = E_{Y_T}(u)_{-r} \cap T_{-r}$. Also, recall from Proposition 5.1 that

$$\begin{aligned} \mathbb{P}(R_0^{(u)} > r) &= \frac{\mathbb{E}[\mathcal{L}(E_X(u)_{-r} \cap T_{-r})]}{\mathbb{E}[\mathcal{L}(E_X(u) \cap T_{-r})]} = \frac{\mathbb{E}[\mathcal{L}((E_{X/u}(1) \cap T)_{-r}) \mid \|X\|_T > u]}{\mathbb{E}[\mathcal{L}(E_{X/u}(1) \cap T_{-r}) \mid \|X\|_T > u]} \\ &\xrightarrow{u \rightarrow \infty} \frac{\mathbb{E}[\mathcal{L}((E_{Y_T}(1) \cap T)_{-r})]}{\mathbb{E}[\mathcal{L}(E_{Y_T}(1) \cap T_{-r})]} = \frac{\mathbb{E}[\mathcal{L}((E_{Y_T}(1) \cap T)_{-r})]}{\mathcal{L}(T_{-r})\mathbb{P}(Y_T(0) > 1)}. \end{aligned}$$

□

5.B.4 Proof of Propostion 5.6

We start by decomposing

$$\hat{\theta}_s^{(n)}(p_0, p_n) = \frac{A_n - B_n}{C - D_n},$$

where

$$\begin{aligned} A_n &= \log(\widehat{q}_{50\%}^{(n)}(R_s^{(u_{p_n})})), & B_n &= \log(\widehat{q}_{50\%}^{(n)}(R_s^{(u_{p_0})})) \\ C &= \log(-\log(1 - p_0)), & D_n &= \log(-\log(1 - p_n)). \end{aligned}$$

We begin by showing

$$-\frac{A_n}{D_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta_s. \quad (5.16)$$

Under Assumption 5.2, for any $\alpha \in (0, 1)$, the function $h_\alpha(x) := q_\alpha(R_s^{(u_{p(x)})})$ with $x > 0$ and $p(x) = 1 - e^{-x}$ is regularly varying with index $-\theta_s$. Thus, by the Karamata characterization

theorem, $x^{\theta_s} h_\alpha(x)$ is slowly varying. For $\epsilon > 0$, and $\alpha \in (0, 1)$, there exists $x_{\alpha, \epsilon} \in \mathbb{R}^+$ such that

$$-\epsilon \log x < \theta_s \log x + \log h_\alpha(x) < \epsilon \log x,$$

for all $x > x_{\alpha, \epsilon}$. Equivalently,

$$\frac{\log h_\alpha(x)}{\log x} \in (-\theta_s - \epsilon, -\theta_s + \epsilon),$$

for all $x > x_{\alpha, \epsilon}$. Let $n \in \mathbb{N}$, and let \mathcal{A}_n denote the event $\{\tilde{R}_i^{(u_{p_n})}(s) > 0, \text{ for at least one } i = 1, \dots, n\}$. By the assumption that $n(1 - p_n) \rightarrow \infty$ as $n \rightarrow \infty$, we have $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$. Choose $\underline{\alpha} \in (0, 1/2)$ and $\bar{\alpha} \in (1/2, 1)$. Notice that for $x = -\log(1 - p_n) = \exp(D_n)$, one has $p(x) = p_n$, and

$$\begin{aligned} \bar{\alpha} - \underline{\alpha} &= \mathbb{P}(h_{\underline{\alpha}}(x) < R_s^{(u_{p_n})} < h_{\bar{\alpha}}(x)) \\ &\leq \mathbb{P}(h_{\underline{\alpha}}(x) < \widehat{q_{50\%}}^{(n)}(R_s^{(u_{p_n})}) < h_{\bar{\alpha}}(x) \mid \mathcal{A}_n) \\ &\leq \frac{\mathbb{P}(h_{\underline{\alpha}}(x) < \widehat{q_{50\%}}^{(n)}(R_s^{(u_{p_n})}) < h_{\bar{\alpha}}(x))}{\mathbb{P}(\mathcal{A}_n)} \\ &= \frac{\mathbb{P}\left(\frac{\log h_{\underline{\alpha}}(x)}{\log(x)} < \frac{A_n}{D_n} < \frac{\log h_{\bar{\alpha}}(x)}{\log(x)}\right)}{\mathbb{P}(\mathcal{A}_n)}. \end{aligned}$$

As seen previously, for n (or equivalently x) large enough, one has

$$\mathbb{P}\left(\frac{\log h_{\underline{\alpha}}(x)}{\log(x)} < \frac{A_n}{D_n} < \frac{\log h_{\bar{\alpha}}(x)}{\log(x)}\right) \leq \mathbb{P}\left(\frac{A_n}{D_n} \in (-\theta_s - \epsilon, -\theta_s + \epsilon)\right).$$

The left-hand side is bounded below by $\mathbb{P}(\mathcal{A}_n)(\bar{\alpha} - \underline{\alpha})$ which can be made arbitrarily close to 1 by choice of $\bar{\alpha}$, $\underline{\alpha}$, and n . This proves (5.16).

Since $D_n \rightarrow \infty$ as $n \rightarrow \infty$, the statement of the proposition holds since B_n converges to a constant $\log(q_{50\%}(R_s^{(u_{p_0})}))$ in probability as $n \rightarrow \infty$. \square

Chapter 6

Perspectives

Here, we provide some outlook on possibilities for this work to be expanded upon in future research. The numerous tools and concepts that we have introduced in this manuscript have largely been studied independently of each other. Here, we provide several suggestions for how these works may be bridged together.

For a stationary random field X , the quantiles of the extremal range at the level $u \in \mathbb{R}$ are seen to behave like the ratio of the expected area of the excursion set $E_X(u)$ to the expected perimeter of $E_X(u)$ (see Corollary 5.1). This relationship is not leveraged in the case study of French temperature data, where instead, the parameters of the model for the median of the extremal range at a point s are inferred by measuring the distance from s to the excursion boundary on days that exhibit a threshold exceedance at s . To leverage the established relationship between the intrinsic volumes of the excursion set and the extremal range, one may estimate the perimeter and area densities locally using the pseudo-local counting algorithm in Definition 3.4, or by using a local counting algorithm (see Definition 2.3) and correcting for the bias (see Theorem 2.2). The ratio of these estimated local densities may also be used to infer the model parameters in (5.10) that describe the asymptotic behavior of the median extremal range.

In addition, it would be interesting to study how the distribution of the reach of the excursion set $E_X(u)$ intersected with a compact set T behaves as u tends to the upper endpoint of the distribution of X . Alternatively, one may be interested in the setting in which the threshold u is fixed, but T grows to cover all of Euclidean space as described in Section 3.3.2. This would shed light on the asymptotic behavior of $\Lambda_X^T(u)$ in Definition 3.5, which characterizes the applicability of the pseudo-local counting algorithm for perimeter estimation in Definition 3.4. Such studies may be carried out numerically using the theory concerning the reach developed in Chapter 4.

Areas of future research also include:

- **An investigation of the variance of local counting algorithms on point-referenced d -honeycombs.** Theorem 2.2 relates the limiting expected surface area estimate to the true expected surface area for excursion sets observed over the convex, polytopic tessellations in

Definition 2.2. The linearity of the expectation allows one to separate the analysis into the contributions of pairs of neighboring points. As mentioned in Section 1.3.4, a study of the higher moments would be more involved, since this linearity argument would cease to hold. Nevertheless, such an analysis may lead to variance reduction techniques for local counting algorithms.

- **The tessellations for which Theorem 2.2 holds.** Point-referenced d -honeycombs, the tessellations defined in Definition 2.2, restrict the reference points to be contained in each cell. This feature of point-referenced d -honeycombs is not used explicitly in the proof of Theorem 2.2. Moreover, if this property is relaxed, then the resulting class of sets is closed under taking the dual tessellation. The class of tessellations which can be expressed as a power diagram (see, *e.g.*, Aurenhammer (1987)) is a subclass of this larger class of tessellations. It is unclear if there are tessellations that are point-referenceable, but cannot be expressed as a power diagram of a set of points with appropriately associated radii.
- **A pseudo-local counting algorithm for perimeter estimation.** The perimeter estimator introduced in Chapter 3 is studied only on two-dimensional excursion sets (see Definition 3.4). Seeing as how this estimator is inspired by Lemma 1.2, which holds in arbitrary dimension, it is likely that the natural, higher dimensional analogue of the estimator is almost surely multigrid convergent to the true $(d - 1)$ -dimensional surface area. The reason for which our result of almost sure consistency, Theorem 3.1, is stated for two-dimensional excursion sets is because the proof involves a piecewise linear approximation of the boundary of the excursion set, where the pieces are delimited by the square subregions of side length m (in the notation of Definition 3.4). In dimension $d > 2$, similar piecewise linear constructions exist, but for technical reasons, they are more difficult to work with when following the lines of the proof of Theorem 3.1.
- **A non-parametric β -reach extrapolation method.** The β -reach profiles in Section 4.3.3 provide the exact β -reach of a point cloud that approximates a Euclidean subset for β in an interval. For β sufficiently far from 0, the β -reach of the point cloud is close to that of the underlying set. The continuity of the β -reach at $\beta = 0$ (see Theorem 4.2) suggests that the reach of the underlying set might be deduced from the β -reach profile of the approximating point-cloud for values of β away from 0. It would be interesting to investigate how machine learning techniques may be used in relation to the one-dimensional β -reach profile to obtain a new, non-parametric estimate of the reach.
- **A toolbox for studying the extremes of random fields.** As mentioned previously in Section 5.7, the extremal range opens the door for the extremal behavior of random fields to be studied using existing tools in computational geometry. For instance, the GitHub repository <https://github.com/RyanCotsakis/excursion-sets> provides some tools for extracting geometric properties from two-dimensional excursion sets. Integrating these existing tools with

methods for analyzing the extremes of random fields may benefit climate scientists, and other researchers working in spatial statistics. Finally, such analyses may be performed in higher dimensions, as the extremal range has a natural, higher dimensional analogue that obeys the theory developed in Chapter 5.

Bibliography

- Aamari, E., Berenfeld, C., and Levrard, C. (2023). Optimal reach estimation and metric learning. *The Annals of Statistics*, 51(3):1086–1108.
- Aamari, E., Kim, J., Chazal, F., Michel, B., Rinaldo, A., and Wasserman, L. (2019). Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399.
- Aamari, E. and Levrard, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204.
- Aaron, C., Cholaquidis, A., and Fraiman, R. (2022). Estimation of surface area. *Electronic Journal of Statistics*, 16(2):3751–3788.
- Abaach, M., Biermé, H., and Di Bernardino, E. (2021). Testing marginal symmetry of digital noise images through the perimeter of excursion sets. *Electronic Journal of Statistics*, 15(2):6429–6460.
- Abaach, M., Biermé, H., Di Bernardino, E., and Estrade, A. (2023). Local isotropy test based on the oriented perimeter of digitalized images. *Preprint: [hal-04037394](https://hal.archives-ouvertes.fr/hal-04037394)*.
- Ade, P., Aghanim, N., Akrami, Y., Aluri, P., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A., Barreiro, R., et al. (2016). Planck 2015 results: XVI. Isotropy and statistics of the CMB. *Astronomy & Astrophysics*, 594:A16.
- Adler, R. J. and Taylor, J. E. (2007). *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York.
- Adler, R. J. and Taylor, J. E. (2011). *Topological complexity of smooth random functions*, volume 2019 of *Lecture Notes in Mathematics*. Springer, Heidelberg.
- AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasn, O., Mof-takhari, H., Papalexioiu, S. M., Ragno, E., and Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, 48:519–548.
- Angulo, J. and Madrid, A. (2010). Structural analysis of spatio-temporal threshold exceedances. *Environmetrics*, 21:415–438.

- Attali, D. and Lieutier, A. (2015). Geometry-driven collapses for converting a Čech complex into a triangulation of a nicely triangulable shape. *Discrete & Computational Geometry*, 54:798–825.
- Attali, D., Lieutier, A., and Salinas, D. (2013). Vietoris–rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, 46(4):448–465. 27th Annual Symposium on Computational Geometry (SoCG 2011).
- Aurenhammer, F. (1987). Power diagrams: Properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96.
- Azaïs, J. M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. John Wiley and Sons Ltd, United Kingdom.
- Baccelli, F., Klein, M., Lebourges, M., and Zuyev, S. (1997). Stochastic geometry and architecture of communication networks. *Telecommunication Systems*, 7:209–227.
- Baccelli, F., Tchoumatchenko, K., and Zuyev, S. (2000). Markov paths on the poisson-delaunay graph with applications to routing in mobile networks. *Advances in Applied Probability*, 32(1):1–18.
- Baorui, M., Yu-Shen, L., Matthias, Z., and Zhizhong, H. (2022). Surface reconstruction from point clouds by learning predictive context priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Bauer, U. (2021). Ripser: efficient computation of Vietoris-Rips persistence barcodes. *Journal of Applied Computational Topology*, 5(3):391–423.
- Beliaev, D., McAuley, M., and Muirhead, S. (2020). On the number of excursion sets of planar Gaussian fields. *Probability Theory and Related Fields*, 178(3):655–698.
- Bentley, J. L. (1975). Multidimensional Binary Search Trees Used for Associative Searching. *Communications of the ACM*, 18(9):509–517.
- Berenfeld, C., Harvey, J., Hoffmann, M., and Shankar, K. (2022). Estimating the Reach of a Manifold via its Convexity Defect Function. *Discrete & Computational Geometry*, 67:403–438.
- Berman, S. (1971). Excursions above high levels for stationary Gaussian processes. *Pacific Journal of Mathematics*, 36(1):63–79.
- Berman, S. M. (1982). Sojourns and Extremes of Stationary Processes. *The Annals of Probability*, 10(1):1–46.
- Berrendero, J. R., Cuevas, A., and Pateiro López, B. (2012). A multivariate uniformity test for the case of unknown support. *Statistics and Computing*, 22:259–271.

- Berzin, C. (2021). Estimation of local anisotropy based on level sets. *Electronic Journal of Probability*, 26:1–72.
- Berzin, C., Latour, A., and León, J. R. (2017). *Kac-Rice formulas for random fields and their applications in: random geometry, roots of random polynomials and some engineering problems*. Ediciones IVIC.
- Beutler, F. J. and Leneman, O. A. Z. (1966). The theory of stationary point processes. *Acta Mathematica*, 116:159–190.
- Biermé, H. and Desolneux, A. (2020). Mean geometry for 2D random fields: Level perimeter and level total curvature integrals. *The Annals of Applied Probability*, 30(2):561–607.
- Biermé, H. and Desolneux, A. (2021). The effect of discretization on the mean geometry of a 2D random field. *Annales Henri Lebesgue*, 4:1295–1345.
- Biermé, H., Di Bernardino, E., Duval, C., and Estrade, A. (2019). Lipschitz-Killing curvatures of excursion sets for two-dimensional random fields. *Electronic Journal of Statistics*, 13(1):536–581.
- Bleau, A. and Leon, L. J. (2000). Watershed-based segmentation and region merging. *Computer Vision and Image Understanding*, 77(3):317–370.
- Blum, H. (1967). A transformation for extracting new descriptions of shape. *Models for the perception of speech and visual form*, pages 362–380.
- Boissonnat, J.-D., Devillers, O., and Hornus, S. (2009). Incremental construction of the delaunay triangulation and the delaunay graph in medium dimension. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*. ACM.
- Boissonnat, J.-D., Lieutier, A., and Wintraecken, M. (2019). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 3(1):29–58.
- Boissonnat, J.-D., Wormser, C., and Yvinec, M. (2008). Locally uniform anisotropic meshing. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*. ACM.
- Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106.
- Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10(4):1047–1050.
- Boos, D. D. (1985). A converse to Scheffe’s theorem. *The Annals of Statistics*, pages 423–427.
- Boulin, A., Di Bernardino, E., Laloë, T., and Toulemonde, G. (2023). Identifying regions of concomitant compound precipitation and wind speed extremes over europe.

- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability surveys*, 2:107–144.
- Bulinski, A. (2010). Central limit theorem for random fields and applications. In *Advances in Data Analysis*, pages 141–150. Springer.
- Bulinski, A. and Shashkin, A. (2007). *Limit theorems for Associated Random Fields and Related Systems*, volume 10. World Scientific.
- Bulinski, A., Spodarev, E., and Timmermann, F. (2012). Central limit theorems for the excursion set volumes of weakly dependent random fields. *Bernoulli*, 18(1):100–118.
- Cabaña, E. M. (1987). Affine Processes: A Test of Isotropy Based on Level Sets. *SIAM Journal on Applied Mathematics*, 47(4):886–891.
- Calka, P. (2019). *Some Classical Problems in Random Geometry*, pages 1–43. Springer International Publishing, Cham.
- Cambanis, S. (1973). On some continuity and differentiability properties of paths of gaussian processes. *Journal of Multivariate Analysis*, 3(4):420–434.
- Camp, R. J., Asing, C. K., Banko, P. C., Berry, L., Brinck, K. W., Farmer, C., and Genz, A. S. (2022). Density surface and excursion sets modeling as an approach to estimating population densities. *The Journal of Wildlife Management*, 87(2).
- Casaponsa, B., Crill, B., Colombo, L., Danese, L., Bock, J., Catalano, A., Bonaldi, A., Basak, S., Bonavera, L., Coulais, A., et al. (2016). Planck 2015 results: Xvi. Isotropy and statistics of the CMB.
- Chazal, F., Cohen-Steiner, D., and Lieutier, A. (2009a). A Sampling Theory for Compact Sets in Euclidean Space. *Discrete & Computational Geometry*, 41:461–479.
- Chazal, F., Cohen-Steiner, D., Lieutier, A., and Thibert, B. (2007). Shape Smoothing Using Double Offset. *Proceedings - SPM 2007: ACM Symposium on Solid and Physical Modeling*.
- Chazal, F., Cohen-Steiner, D., Lieutier, A., and Thibert, B. (2009b). Stability of curvature measures. In *Computer Graphics Forum*, volume 28, pages 1485–1496. Wiley Online Library.
- Chazal, F. and Lieutier, A. (2005a). The “ λ -medial axis”. *Graphical Models*, 67(4):304–331.
- Chazal, F. and Lieutier, A. (2005b). Weak Feature Size and Persistent Homology: Computing Homology of Solids in \mathbb{R}^n from Noisy Data Samples. In *Proceedings of the Twenty-First Annual Symposium on Computational Geometry*, SCG '05, pages 255–262, New York, NY, USA. Association for Computing Machinery.

- Chazal, F. and Lieutier, A. (2008). Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Computational Geometry*, 40(2):156–170.
- Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:108.
- Cheng, D. and Schwartzman, A. (2018). Expected number and height distribution of critical points of smooth isotropic gaussian random fields. *Bernoulli*, 24(4B):3422–3446.
- Chiles, J.-P. and Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley & Sons.
- Cholaquidis, A. (2023). A counter example on a borsuk conjecture. *Applied General Topology*, 24(1):125–128.
- Cholaquidis, A., Fraiman, R., and Moreno, L. (2023). Universally consistent estimation of the reach. *Journal of Statistical Planning and Inference*, 225:110–120.
- Clarke, F. and Wolenski, P. (1995). Proximal Smoothness and the Lower- C^2 Property. *Journal of Convex Analysis*, 2:117–144.
- Coeurjolly, D. and Klette, R. (2002). A comparative evaluation of length estimators. In *2002 International Conference on Pattern Recognition*, volume 4, pages 330–334.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.
- Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339–365.
- Colesanti, A. and Manselli, P. (2007). Geometric and isoperimetric properties of sets of positive reach in \mathbf{E}^d . *Preprint: [arXiv:math/0703634](https://arxiv.org/abs/math/0703634)*.
- Cotsakis, R. (2023). Computable bounds for the reach and r-convexity of subsets of \mathbb{R}^d . *Discrete & Computational Geometry. Accepted Author Manuscript*.
- Cotsakis, R., Di Bernardino, E., and Duval, C. (2023a). Surface area and volume of excursion sets observed on point cloud based polytopic tessellations. *The Annals of Applied Probability. Accepted Author Manuscript*.
- Cotsakis, R., Di Bernardino, E., and Opitz, T. (2023b). A local statistic for the spatial extent of extreme threshold exceedances. *Preprint [arXiv:2310.09075](https://arxiv.org/abs/2310.09075)*.
- Cotsakis, R., Di Bernardino, E., and Opitz, T. (2023c). On the perimeter estimation of pixelated excursion sets of 2D anisotropic random fields. *Scandinavian Journal of Statistics. Accepted Author Manuscript*.

- Cuevas, A. (2009). Set estimation: Another bridge between statistics and geometry. *BEIO, Boletín de Estadística e Investigación Operativa*, 25:71–85.
- Cuevas, A., Fraiman, R., and Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *The Annals of Statistics*, 35:1031–1051.
- Cuevas, A., Fraiman, R., and Pateiro-López, B. (2012). On Statistical Properties of Sets Fulfilling Rolling-Type Conditions. *Advances in Applied Probability*, 44(2):311–329.
- Cuevas, A., Llop, P., and Pateiro-López, B. (2014). On the estimation of the medial axis and inner parallel body. *Journal of Multivariate Analysis*, 129:171–185.
- Cuevas, A. and Rodríguez-Casal, A. (2004). On Boundary Estimation. *Advances in Applied Probability*, 36(2):340–354.
- Dalmao, F., León, J. R., Mordecki, E., and Mourareau, S. (2019). Asymptotic normality of high level-large time crossings of a Gaussian process. *Stochastic Processes and their Applications*, 129(4):1349–1370.
- Das, A. B., Makogin, V., and Spodarev, E. (2021). Extrapolation of stationary random fields via level sets. *Theory of Probability and Mathematical Statistics*.
- de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Springer Berlin Heidelberg.
- de Castro, P. M. M. and Devillers, O. (2017). Expected Length of the Voronoi Path in a High Dimensional Poisson–Delaunay Triangulation. *Discrete & Computational Geometry*, 60(1):200–219.
- de Fondeville, R. and Davison, A. C. (2022). Functional peaks-over-threshold analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1392–1422.
- de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer.
- de Vieilleville, F., Lachaud, J.-O., and Feschet, F. (2007). Maximal digital straight segments and convergence of discrete geometric estimators. *Journal of Mathematical Image and Vision*, 27(2):471–502.
- Debinski, D. M. and Holt, R. D. (2000). A survey and overview of habitat fragmentation experiments. *Conservation Biology*, 14(2):342–355.
- Dedecker, J. (1998). A central limit theorem for stationary random fields. *Probability Theory and Related Fields*, 110(3):397–426.

- Dey, T. K. and Sun, J. (2006). Normal and feature approximations from noisy point clouds. In Arun-Kumar, S. and Garg, N., editors, *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science*, pages 21–32, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Dey, T. K. and Zhao, W. (2003). Approximating the Medial Axis from the Voronoi Diagram with a Convergence Guarantee. *Algorithmica*, 38(1):179–200.
- Di Bernardino, E. and Duval, C. (2022). Statistics for Gaussian Random Fields with Unknown Location and Scale using Lipschitz-Killing Curvatures. *Scandinavian Journal of Statistics*, 49(1):143–184.
- Di Bernardino, E., Estrade, A., and León, J. R. (2017). A test of Gaussianity based on the Euler characteristic of excursion sets. *Electronic Journal of Statistics*, 11(1):843–890.
- Di Bernardino, E., Estrade, A., and Opitz, T. (2022). Spatial extremes and stochastic geometry for Gaussian-based peaks-over-threshold processes. *Preprint: hal-03825701*.
- Di Bernardino, E., Estrade, A., and Rossi, M. (2020). On the excursion area of perturbed Gaussian fields. *ESAIM Probab. Stat.*, 24:252–274.
- Divol, V. (2021). Minimax adaptive estimation in manifold inference. *Electronic Journal of Statistics*, 15:5888–5932.
- Dombry, C. and Eyi-Minko, F. (2012). Strong mixing properties of max-infinitely divisible random fields. *Stochastic Processes and their Applications*, 122(11):3790–3811.
- Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface*, 8:9–17.
- Dombry, C., Ribatet, M., and Stoev, S. (2018). Probabilities of Concurrent Extremes. *Journal of the American Statistical Association*, 113(524):1565–1582.
- Donchian, P. S. and Coxeter, H. S. M. (1935). 1142. An n-dimensional extension of Pythagoras’ Theorem. *The Mathematical Gazette*, 19(234):206.
- Doukhan, P. (1994). *Mixing*. Springer.
- Dubins, L. E. (1961). On plane curves with curvature. *Pacific Journal of Mathematics*, 11:471–481.
- Edelsbrunner, Letscher, and Zomorodian (2002). Topological Persistence and Simplification. *Discrete & Computational Geometry*, 28(4):511–533.
- Edelsbrunner, H. (2001). *Geometry and Topology for Mesh Generation*. Cambridge University Press.

- Edelsbrunner, H. and Nikitenko, A. (2021). Average and expected distortion of voronoi paths and scapes. *Preprint: arXiv:2012.03350*.
- Edelsbrunner, H. and Pausinger, F. (2016). Approximation and convergence of the intrinsic volume. *Advances in Mathematics*, 287:674–703.
- Estrade, A. and León, J. (2016). A Central Limit Theorem for the Euler characteristic of a Gaussian excursion set. *Annals of Probability*, 44(6):3849–3878.
- Eymard, R., Gallouët, T., and Herbin, R. (2000). *Finite Volume Methods*, volume 7 of *Handbook of Numerical Analysis*. Elsevier.
- Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491.
- Feller, W. (1966). *An Introduction to Probability Theory and its Applications. Vol. II*. John Wiley & Sons Inc., New York.
- Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.
- Fournier, J. (2018). Identification and isotropy characterization of deformed random fields through excursion sets. *Advances in Applied Probability*, 50(3):706–725.
- Friedman, A. (2007). *Advanced Calculus*. Dover Books on Mathematics. Dover Publications.
- Frölicher, T. L., Fischer, E. M., and Gruber, N. (2018). Marine heatwaves under global warming. *Nature*, 560(7718):360–364.
- Gordon, R. D. (1941). Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. *The Annals of Mathematical Statistics*, 12(3):364–366.
- Gott, J. R., Colley, W. N., Park, C.-G., Park, C., and Mugnolo, C. (2007). Genus topology of the cosmic microwave background from the WMAP 3-year data. *Monthly Notices of the Royal Astronomical Society*, 377(4):1668–1678.
- Gott, J. R., Hambrick, D. C., Vogeley, M. S., Kim, J., Park, C., Choi, Y.-Y., Cen, R., Ostriker, J. P., and Nagamine, K. (2008). Genus topology of structure in the sloan digital sky survey: Model testing. *The Astrophysical Journal*, 675(1):16.

- Gott, J. R., Park, C., Juszkievicz, R., Bies, W. E., Bennett, D. P., Bouchet, F. R., and Stebbins, A. (1990). Topology of Microwave Background Fluctuations: Theory. *Astrophysical Journal*, 352:1.
- Gray, S. (1971). Local Properties of Binary Images in Two Dimensions. *IEEE Transactions on Computers*, C-20(5):551–561.
- Guderlei, R., Klenk, S., Mayer, J., Schmidt, V., and Spodarev, E. (2007). Algorithms for the computation of the minkowski functionals of deterministic and random polyconvex sets. *Image and Vision Computing*, 25(4):464–474.
- Harrell, F. E. and Davis, C. (1982). A new distribution-free quantile estimator. *Biometrika*, 69(3):635–640.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.
- Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, 21:166–186.
- Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(1):e1537.
- Iribarren, I. (1989). Asymptotic behaviour of the integral of a function on the level set of a mixing random field. *Probability and Mathematical Statistics*, 10(1):45–56.
- Jarek, S. (2012). *mvnormtest: Normality test for multivariate variables*. R package version 0.1-9.
- Jennings, E., Li, Y., and Hu, W. (2013). The abundance of voids and the excursion set formalism. *Monthly Notices of the Royal Astronomical Society*, 434(3):2167–2181.
- Jurdi, R. E., Petitjean, C., Cheplygina, V., and Abdallah, F. (2021). A surprisingly effective perimeter-based loss for medical image segmentation. In *International Conference on Medical Imaging with Deep Learning*.
- Kac, M. and Slepian, D. (1959). Large Excursions of Gaussian Processes. *The Annals of Mathematical Statistics*, 30(4):1215–1228.
- Kampf, J. (2014). A limitation of the estimation of intrinsic volumes via pixel configuration counts. *Mathematika*, 60(2):485–511.
- Kendall, W. S. and Molchanov, I. (2009). *New Perspectives in Stochastic Geometry*. Oxford University Press.

- Kiderlen, M. (2006). Estimating the Euler Characteristic of a planar set from a digital image. *Journal of Visual Communication and Image Representation*, 17(6):1237–1255.
- Kiderlen, M. and Meschenmoser, D. (2011). Error bounds for surface area estimators based on Crofton’s formula. *Image Analysis and Stereology*, 28(3):165–177.
- Kim, J., Shin, J., Chazal, F., Rinaldo, A., and Wasserman, L. (2019). Homotopy reconstruction via the cech complex and the Vietoris-Rips complex. *Preprint [arXiv:1903.06955](https://arxiv.org/abs/1903.06955)*.
- Klenk, S., Schmidt, V., and Spodarev, E. (2006). A new algorithmic approach to the computation of minkowski functionals of polyconvex sets. *Computational Geometry*, 34(3):127–148.
- Kratz, M. and León, J. R. (2001). Central Limit Theorems for Level Functionals of Stationary Gaussian Processes and Fields. *Journal of Theoretical Probability*, 14:639–672.
- Kratz, M. and Vadlamani, S. (2018). Central Limit Theorem for Lipschitz–Killing Curvatures of Excursion Sets of Gaussian Random Fields. *Journal of Theoretical Probability*, 31(3):1729–1758.
- Kratz, M. F. (2006). Level crossings and other level functionals of stationary Gaussian processes. *Probability Surveys*, 3:230–288.
- Lachaud, J.-O., Coeurjolly, D., Labart, C., Romon, P., and Thibert, B. (2023). Lightweight curvature estimation on point clouds with randomized corrected curvature measures. *Computer Graphics Forum (Proceedings of Symposium on Geometry Processing)*, 42(5).
- Lachaud, J.-O., Romon, P., Thibert, B., and Coeurjolly, D. (2020). Interpolated corrected curvature measures for polygonal surfaces. *Computer Graphics Forum (Proceedings of Symposium on Geometry Processing)*, 39(5).
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer New York.
- Lehmann, G. and Legland, D. (2012). Efficient N-Dimensional surface estimation using Crofton formula and run-length encoding. *The Insight Journal, INRA*.
- Lhotka, O. and Kyselý, J. (2015). Characterizing joint effects of spatial extent, temperature magnitude and duration of heat waves and cold spells over Central Europe. *International Journal of Climatology*, 35(7):1232–1244.
- Lieutier, A. (2004). Any open bounded subset of \mathbb{R}^n has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046.
- Lieutier, A. and Wintraecken, M. (2023). Hausdorff and gromov-hausdorff stable subsets of the medial axis. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 1768–1776, New York, NY, USA. Association for Computing Machinery.

- Lindblad, J. (2005). Surface area estimation of digitized 3d objects using weighted local configurations. *Image and Vision Computing*, 23(2):111–122. Discrete Geometry for Computer Imagery.
- Liu, T., Moore, A. W., and Gray, A. (2006). New Algorithms for Efficient High-Dimensional Nonparametric Classification. *Journal of Machine Learning Research*, 7:1135–1158.
- Longuet-Higgins, M. S. (1957). The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society A*, 249(966):321–387.
- Lévy, P. (1940). Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 7:283–339.
- Mani-Levitska, P. (1993). Characterizations of Convex Sets. In *Handbook of Convex Geometry*, pages 19–41. North-Holland.
- Matheron, G. (1975). *Random Sets and Integral Geometry*. Wiley, NY.
- McGarigal, K. (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*, volume 351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Meschenmoser, D. and Shashkin, A. (2011). Functional central limit theorem for the volume of excursion sets generated by associated random fields. *Statistics & Probability Letters*, 81(6):642–646.
- Meschenmoser, D. and Shashkin, A. (2013). Functional central limit theorem for the measures of level surfaces of the Gaussian random field. *Theory of Probability and its Applications*, 57(1):162–172.
- Meschenmoser, D. and Spodarev, E. (2010). On the computation of intrinsic volumes. *Preprint*.
- Miller, E. (1999). Alternative Tilings for Improved Surface Area Estimates by Local Counting Algorithms. *Computer Vision and Image Understanding*, 74(3):193–211.
- Moloney, N. R., Faranda, D., and Sato, Y. (2019). An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(2):022101.
- Mrkvička, T. and Rataj, J. (2008). On the estimation of intrinsic volume densities of stationary random closed sets. *Stochastic Processes and their Applications*, 118(2):213–231.
- Müller, D. (2017). A central limit theorem for Lipschitz–Killing curvatures of Gaussian excursions. *Journal of Mathematical Analysis and Applications*, 452(2):1040–1081.
- Nagendra, H., Munroe, D. K., and Southworth, J. (2004). From pattern to process: landscape fragmentation and the analysis of land use/land cover change. *Agriculture, Ecosystems & Environment*, 101:111–115.

- Niyogi, P., Smale, S., and Weinberger, S. (2008). Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete & Computational Geometry*, 39:419–441.
- Omohundro, S. M. (1989). *Five balltree construction algorithms*. International Computer Science Institute Berkeley.
- Pateiro López, B. (2008). Set estimation under convexity type restrictions. Ph.D. thesis. [Link to thesis].
- Perkal, J. (1956). Sur les ensembles ϵ -convexes. *Colloquium Mathematicum*, 4:1–10.
- Perkins, S., Alexander, L., and Nairn, J. (2012). Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophysical Research Letters*, 39(20).
- Pham, V.-H. (2013). On the rate of convergence for central limit theorems of sojourn times of Gaussian fields. *Stochastic Processes and their Applications*, 123(6):427–464.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119 – 131.
- Poliquin, R. A., Rockafellar, R. T., and Thibault, L. (2000). Local Differentiability of Distance Functions. *Transactions of the American Mathematical Society*, 352(11):5231–5249.
- Princeton, H. (2023a). Edge-to-edge dual construction. Wikimedia Commons. CC BY-SA 4.0. https://commons.wikimedia.org/wiki/File:Edge-to-Edge_Dual_Construction.svg.
- Princeton, H. (2023b). Edge-to-edge original. Wikimedia Commons. CC BY-SA 4.0. https://commons.wikimedia.org/wiki/File:Edge-to-Edge_Original.svg.
- Qiao, W. and Polonik, W. (2018). Extrema of rescaled locally stationary Gaussian fields on manifolds. *Bernoulli*, 24(3):1834–1859.
- Rataj, J. (2006). Estimation of intrinsic volumes from parallel neighbourhoods. *Suppl. Rend. Circ. Mat. Palermo, Ser. II*, 77:553–563.
- Rataj, J. and Zähle, M. (2003). Normal cycles of lipschitz manifolds by approximation with parallel sets. *Differential Geometry and its Applications*, 19(1):113–126.
- Rataj, J. and Zähle, M. (2019). Sets with positive reach. In *Curvature Measures of Singular Sets*, pages 55–86. Springer.
- Reddy, T. R., Vadlamani, S., and Yogeshwaran, D. (2018). Central limit theorem for exponentially quasi-local statistics of spin models on cayley graphs. *Journal of Statistical Physics*, (173):941–984.

- Rice, S. O. (1944). Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3):282–332.
- Ripley, B. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.
- Rodríguez Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. *ESAIM: Probability and Statistics*, 20:332–348.
- Rogers, C. A. (1998). *Hausdorff measures*. Cambridge University Press.
- Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47.
- Roubin, E. and Colliat, J.-B. (2016). Critical probability of percolation over bounded region in N-dimensional Euclidean space. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(3):033306.
- Santaló, L. A. (1976). *Integral Geometry and Geometric Probability*. Encyclopedia of Mathematics and Its Applications, Vol. 1, Addison-Wesley, Reading, MA.
- Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., and Moreva, O. (2017). *RandomFields: Simulation and Analysis of Random Fields*. R package version 3.1.50.
- Schmalzing, J. and Górski, K. M. (1998). Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps. *Monthly Notices of the Royal Astronomical Society*, 297(2):355–365.
- Schneider, R. and Weil, W. (2008). *Stochastic and integral geometry*. Probability and its Applications. Springer-Verlag, Berlin.
- Serra, J. (1984). *Image Analysis and Mathematical Morphology*. Academic Press.
- Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595.
- Sezgin, M. and Sankur, B. I. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168.
- Shashkin, A. (2013). A functional central limit theorem for the level measure of a Gaussian random field. *Statistics & Probability Letters*, 83(2):637–643.
- Silva, F. and Steele, J. (2014). New methods for reconstructing geographical effects on dispersal rates and routes from large-scale radiocarbon databases. *Journal of Archaeological Science*, 52:609–620.

- Simpson, E. S., Opitz, T., and Wadsworth, J. L. (2023). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *Extremes*, pages 1–45.
- Sommerfeld, M., Sain, S., and Schwartzman, A. (2018). Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate. *Journal of the American Statistical Association*, 113(523):1327–1340.
- Spodarev, E. (2014). Limit theorems for excursion sets of stationary random fields. In *Modern stochastics and applications*, volume 90 of *Springer Optimization and Its Applications*, pages 221–241. Springer, Cham.
- Sullivan, C. and Kaszynski, A. (2019). PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *Journal of Open Source Software*, 4(37):1450.
- Svane, A. M. (2014a). Local digital estimators of intrinsic volumes for boolean models and in the design-based setting. *Advances in Applied Probability*, 46(1):35–58.
- Svane, A. M. (2014b). On multigrid convergence of local algorithms for intrinsic volumes. *J. Math. Imaging Vision*, 49:148–172.
- Svane, A. M. and Vedel Jensen, E. B. (2017). Rotational Crofton formulae for Minkowski tensors and some affine counterparts. *Advances in Applied Mathematics*, 91:44–75.
- Sweeting, T. (1986). On a converse to Scheffé’s theorem. *The Annals of Statistics*, 14(3):1252–1256.
- Taubert, F., Fischer, R., Groeneveld, J., Lehmann, S., Müller, M. S., Rödig, E., Wiegand, T., and Huth, A. (2018). Global patterns of tropical forest fragmentation. *Nature*, 554:519–522.
- Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial statistics*, 28:39–58.
- Thäle, C. (2008). 50 years sets with positive reach—a survey. *Surveys in Mathematics and its Applications*, 3:123–165.
- Thäle, C. and Yukich, J. E. (2016). Asymptotic theory for statistics of the Poisson–Voronoi approximation. *Bernoulli*, 22(4):2372–2400.
- Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.
- Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M. (2010). A 50-year high-resolution atmospheric reanalysis over france with the safran system. *International journal of climatology*, 30(11):1627–1644.

- Wadsworth, J. L. and Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.
- Walther, G. (1999). On a Generalization of Blaschke’s Rolling Theorem and the Smoothing of Surfaces. *Mathematical Methods in the Applied Sciences*, 22:301–316.
- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.
- Worsley, K. J., Evans, A., Marrett, S., and Neelin, P. (1992). A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 12:900–18.
- Wschebor, M. (1985). *Surfaces aléatoires: mesure géométrique des ensembles de niveau*. Lecture Notes in Mathematics, 1147, Springer-Verlag, Berlin.
- Yao, J., Wang, S., Zhu, X., and Huang, J. (2016). Imaging biomarker discovery for lung cancer survival prediction. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., and Wells, W., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 649–657, Cham. Springer International Publishing.
- Yukich, J. E. (2015). Surface order scaling in stochastic geometry. *The Annals of Applied Probability*, 25:177–210.
- Zähle, M. (1987). Curvatures and currents for unions of sets with positive reach. *Geometriae Dedicata*, 23(2):155–171.
- Zhang, Z., Huser, R., Opitz, T., and Wadsworth, J. (2022). Modeling spatial extremes using normal mean-variance mixtures. *Extremes*, 25(2):175–197.
- Ziegel, J. and Kiderlen, M. (2010). Estimation of surface area and surface area measure of three-dimensional sets from digitizations. *Image and Vision Computing*, 28(1):64–77.