Optimal bandwidth of the "Minkowski content"-based estimator of the mean density of random closed sets: theoretical results and numerical experiments

Federico Camerlenghi[1] and Elena Villa[2]

[1]Dept. of Mathematics, Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy
[2]Dept. of Mathematics, Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy

March 20, 2015

Abstract

The estimation of the mean density of random closed sets in $\mathbb{R}^d$ with integer Hausdorff dimension $n < d$ is a problem of interest from both a theoretical and an applicative point of view. In literature different kinds of estimators are available, mostly for the homogeneous case. Recently the non homogeneous case has been faced by the authors; more precisely, two different kinds of estimators, asymptotically unbiased and weakly consistent, have been proposed: in [9] a kernel-type estimator generalizing the well-known kernel density estimator for random variables, and in [29] an estimator based on the notion of Minkowski content of a set. The study of the optimal bandwidth of the "Minkowski content"-based estimator has been left as an open problem in [29, Section 6] and in [30, Remark 14], and only partially solved in [9, Section 4], where a formula is available in the particular case of homogeneous Boolean models. We give here a solution of such an open problem, by providing explicit formulas for the optimal bandwidth for quite general random closed sets (i.e. not necessarily Boolean models or homogeneous germ-grain models). We also discuss a series of relevant examples and corresponding numerical experiments to validate our theoretical results.

Keywords: density estimator, random closed set, stochastic geometry, Hausdorff dimension, Minkowski content

AMS Classification 2010: Primary 62G07; Secondary 60D05, 28A75

1 Introduction

The problem of the evaluation and estimation of the mean density of random closed sets in $\mathbb{R}^d$ with integer Hausdorff dimension n less than d, such as fiber processes, boundaries of germ-grain models, n-facets of random tessellations, and surfaces of full dimensional random sets, has been of great interest in many different scientific and technological fields over the last decades (see [9] and references therein). We remind that, given a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a *random closed set* Θ in $\mathbb{R}^d$ is a measurable map

$$\Theta : (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\mathbb{F}, \sigma_{\mathbb{F}}),$$

where $\mathbb{F}$ denotes the class of the closed subsets in $\mathbb{R}^d$, and $\sigma_{\mathbb{F}}$ is the σ-algebra generated by the so called *Fell topology*, or *hit-or-miss topology*, that is the topology generated by the set system

$$\{\mathcal{F}_G : G \in \mathcal{G}\} \cup \{\mathcal{F}^C : C \in \mathcal{C}\}$$

1

where $\mathcal{G}$ and $\mathcal{C}$ are the system of the open and compact subsets of $\mathbb{R}^d$, respectively (e.g., see [22]). Throughout the paper we shall denote by Θ_n any random closed set in $\mathbb{R}^d$ with Hausdorff dimension n; we mean that a random closed set satisfies a certain property if it satisfies that property $\mathbb{P}$-a.s.

Recently, in [9], [29] and [30], two different kinds of estimators, asymptotically unbiased and weakly consistent, have been proposed by the authors: in [9] a kernel-type estimator generalizing the well-known kernel density estimator for random variables; in [29] and [30] (for Boolean models and more general germ-grain processes, respectively) an estimator based on the notion of Minkowski content. (See also [10] for a survey containing also some numerical experiments validating the theoretical results available so far).

In particular, although the "Minkowski content"-based estimator reveals its benefits in applications in the non-stationary cases, so far general optimal bandwidths have been identified only for the kernel-type estimator. Indeed the study of the optimal bandwidth of the "Minkowski content"-based estimator has been left as an open problem in [29, Section 6] and in [30, Remark 14], and partially solved in [9, Section 4], where a formula is available only in the particular case in which the random set Θ_n is a homogeneous Boolean model.

So, this is the main goal of the present paper: to provide optimal bandwidths for the "Minkowski content"-based estimator for more general random closed sets, not necessarily homogeneous or Boolean models. In view of possible applications this might be of great importance for two main reasons: the first one is that Boolean models are characterized by independence of the grains and it is quite intuitive that such an assumption is often not fulfilled in real applications; the second one is that the "Minkowski content"-based estimator turns out to be much more feasible than the kernel-type estimator. Once we solve the open problem regarding the optimal bandwidth for the "Minkowski content"-based estimator, we provide here a series of numerical experiments for some relevant situations, in order to both validate the theoretical results and compare all the proposed estimators.

More precisely, the paper is organized as follows: in Section 2 basics and notation on germ-grain representation of random closed sets in $\mathbb{R}^d$ are provided; the definition of the "Minkwoski-content"-based estimator and the related known results on its statistical properties are also recalled. Moreover, all the additional assumptions useful throughout the paper are listed at the end of the section for the reader's convenience. In Section 3 the main theorems of the paper are stated, giving an asymptotical expression of the bias of the "Minkwoski-content"-based estimator; as a byproduct general formulas for the optimal bandwidth of the estimator are obtained. In Section 4 some relevant particular cases are discussed; we point out that already known results for the cases in which the random set Θ is actually a random variable, or a stationary Boolean models are reobtained here as particular case. In Section 5 a series of examples and related numerical experiments and remarks are given; in particular it will be evident how much easier is to apply the "Minkowski content"-based estimator with respect to other kinds of estimators of the mean densities. This is due to the fact that, as it is clear from its definition, the estimate of the mean density of Θ_n in any point $x \in \mathbb{R}^d$ does not require any particular calculation, except for counting how many elements of the random sample of Θ_n have at least one point in the ball centered in x; this can be generally easily done by checking whether the pixel corresponding to x belongs to Θ_n or not in its digital image. This is the reason why in the example discussed in Section 5.2.2 we are not able to compute the estimate of the mean density by the other proposed estimators, but only for the "Minkowski content"-based estimator.

To make the paper more readable we collected all the proofs of the main results in Section 6, and useful definitions on sets with positive reach and related properties and curvature measures in the

2

Appendix.

2 Basics and notation

To lighten the presentation we shall use similar notation to previous works [9], [29], [30]; in particular, for the reader's convenience, we refer to [30, Section 2] (and references therein) for the mathematical background and more details on the Minkowski content notion and marked point process theory, while we refer to [9, Appendix] (and references therein) for the classical density estimation theory for random variables.

Here $\mathcal{H}^n$ is the n-dimensional Hausdorff measure, $\mathrm{d}x$ stands for $\mathcal{H}^d(\mathrm{d}x)$, and $\mathcal{B}_{\mathcal{X}}$ is the Borel σ-algebra of any topological space $\mathcal{X}$. $B_r(x)$ and b_n will denote the closed ball with centre x and radius $r > 0$ and the volume of the unit ball in $\mathbb{R}^n$, respectively. By means of marked point processes in $\mathbb{R}^d$ with marks in the class of compact subsets of $\mathbb{R}^d$, every random closed set Θ in $\mathbb{R}^d$ can be represented as a *germ-grain model* as follows

$$\Theta(\omega) = \bigcup_{(x_i, s_i) \in \Psi(\omega)} x_i + Z(s_i), \qquad \omega \in \Omega, \tag{1}$$

where $\Psi = \{(\xi_i, S_i)\}_{i \in \mathbb{N}}$ is a marked point processes in $\mathbb{R}^d$ with marks in a suitable mark space $\mathbf{K}$ so that $Z_i = Z(S_i)$, $i \in \mathbb{N}$, is a random set containing the origin (i.e., $Z : \mathbf{K} \to \mathbb{F}$). Note that even if it would be possible to directly consider marked point processes with mark space $\mathbb{F}$, we prefer to use the same approach and notation of previous works by the authors on this subject, both for the reader's convenience and because many applications can be handled in a simpler way by considering a suitable parametrization of the grains Z_i.

Throughout the paper we assume that Ψ has intensity measure $\Lambda(\mathrm{d}(x,s)) = f(x,s)\mathrm{d}xQ(\mathrm{d}s)$ and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,s,y,t)\mathrm{d}x\mathrm{d}y\,Q_{[2]}(\mathrm{d}(s,t))$ [14, for general theory of point processes],[20], [11]. We also recall that whenever Ψ is a marked Poisson point process, Θ is said to be a *Boolean model*; in such a case $\nu_{[2]} = \Lambda \otimes \Lambda$, so that $g(x,s,y,t) = f(x,s)f(y,t)$.

For any function f, disc$f$ will denote the set of its discontinuity points.

Let Θ_n be a set of locally finite $\mathcal{H}^n$-measure, as in (1); then it induces a random measure $\mu_{\Theta_n}(A) := \mathcal{H}^n(\Theta_n \cap A)$, $A \in \mathcal{B}_{\mathbb{R}^d}$, and the corresponding expected measure $\mathbb{E}[\mu_{\Theta_n}](A) := \mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)]$, $A \in \mathcal{B}_{\mathbb{R}^d}$. (For a discussion of the measurability of the random variables $\mu_{\Theta_n}(A)$, we refer to [5, 31].)

Whenever the measure $\mathbb{E}[\mu_{\Theta_n}]$ is absolutely continuous with respect to the measure $\mathcal{H}^d$, its density (i.e. its Radon-Nikodym derivative) with respect to $\mathcal{H}^d$ has been called *mean density* of Θ_n, and denoted by λ_{Θ_n}. It has been proved that ([30, Proposition 5])

$$\lambda_{\Theta_n}(x) = \int_{\mathbf{K}} \int_{x - Z(s)} f(y,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s), \tag{2}$$

for $\mathcal{H}^d$-a.e $x \in \mathbb{R}^d$, where $-Z(s)$ is the reflection of $Z(s)$ at the origin.

In the sequel we will assume that an i.i.d. random sample $\Theta_n^1, \ldots, \Theta_n^N$ is available for the random closed set Θ_n, with mean density λ_{Θ_n}. The issue of estimation of $\lambda_{\Theta_n}(x)$ may be solved by (2) via the estimation of f and Q, which unfortunately is in general a difficult task in applications. Therefore, being the Lebesgue measure much more robust and computable than $\mathcal{H}^n$, an approximation of the mean density based on the $\mathcal{H}^d$-measure of the Minkowski enlargement of the random set in question has been provided in [30]. More precisely, we recall that the parallel set

(or, equivalently, the Minkowski enlargement) of $A \subset \mathbb{R}^d$ at distance $r > 0$, is the set so defined $A_{\oplus r} := \{x \in \mathbb{R}^d : \mathrm{dist}(x, A) \leq r\}$; we also remind that a compact set $A \subset \mathbb{R}^d$ is called *countably $\mathcal{H}^n$-rectifiable* if there exist countably many n-dimensional Lipschitz graphs $\Gamma_i \subset \mathbb{R}^d$ such that $A \setminus \cup_i \Gamma_i$ is $\mathcal{H}^n$-negligible (e.g., see [3] and references therein for a more exhaustive treatment). The following approximation holds:

Theorem 1 *[30, Theorem 7] Let Θ_n be as in (1) such that the following assumptions are satisfied:*

(A1) for any $(y, s) \in \mathbb{R}^d \times \mathbf{K}$, $y + Z(s)$ is a countably $\mathcal{H}^n$-rectifiable and compact subset of $\mathbb{R}^d$, such that there exists a closed set $\Xi(s) \supseteq Z(s)$ such that $\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) Q(\mathrm{d}s) < \infty$ and

$$\mathcal{H}^n(\Xi(s) \cap B_r(x)) \geq \gamma r^n \qquad \forall x \in Z(s), \ \forall r \in (0, 1)$$

for some $\gamma > 0$ independent of s;

(A2) for any $s \in \mathbf{K}$, $\mathcal{H}^n(\mathrm{disc}(f(\cdot, s))) = 0$ and $f(\cdot, s)$ is locally bounded such that for any compact $K \subset \mathbb{R}^d$

$$\sup_{x \in K_{\oplus \mathrm{diam}(Z(s))}} f(x, s) \leq \tilde{\xi}_K(s)$$

for some $\tilde{\xi}_K(s)$ with $\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \tilde{\xi}_K(s) Q(\mathrm{d}s) < \infty$;

(A3) for any $(s, y, t) \in \mathbf{K} \times \mathbb{R}^d \times \mathbf{K}$, $\mathcal{H}^n(\mathrm{disc}(g(\cdot, s, y, t))) = 0$ and $g(\cdot, s, y, t)$ is locally bounded such that for any compact $K \subset \mathbb{R}^d$ and $a \in \mathbb{R}^d$,

$$\mathbf{1}_{(a - Z(t))_{\oplus 1}}(y) \sup_{x \in K_{\oplus \mathrm{diam}(Z(s))}} g(x, s, y, t) \leq \xi_{a, K}(s, y, t)$$

for some $\xi_{a, K}(s, y, t)$ with

$$\int_{\mathbb{R}^d \times \mathbf{K}^2} \mathcal{H}^n(\Xi(s)) \xi_{a, K}(s, y, t) \mathrm{d}y Q_{[2]}(\mathrm{d}s, \mathrm{d}t) < \infty$$

Then

$$\lambda_{\Theta_n}(x) = \lim_{r \downarrow 0} \frac{\mathbb{P}(x \in \Theta_{n \oplus r})}{b_{d-n} r^{d-n}}, \qquad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d.$$

As a byproduct, given an i.i.d. random sample $\{\Theta_n^i\}_{i \in \mathbb{N}}$ of Θ_n, the following *"Minkowski content"-based estimator* of $\lambda_{\Theta_n}(x)$ was proposed in [30]:

$$\widehat{\lambda}_{\Theta_n}^{\mu, N}(x) := \frac{\sum_{i=1}^{N} \mathbf{1}_{\Theta_n^i \cap B_{r_N}(x) \neq \emptyset}}{N b_{d-n} r_N^{d-n}}.$$

For a discussion on the above assumptions we refer to [30, Sec. 3.1]. Here we only point out that

- whenever Θ_n is a Boolean model, $(A3)$ is an immediate consequence of $(A1)$ and $(A2)$.

- Given a subset A of $\mathbb{R}^d$ and an integer n with $0 \leq n \leq d$, the *n-dimensional Minkowski content* of A is defined as

$$\mathcal{M}^n(A) := \lim_{r \downarrow 0} \frac{\mathcal{H}^d(A_{\oplus r})}{b_{d-n} r^{d-n}},$$

4

whenever the limit exists and is finite; assumption (A1) can be seen as the stochastic version of the assumption in [3, Theorem 2.104] which guarantees $\mathcal{M}^n(A) = \mathcal{H}^n(A)$. We say that Θ_n admits *mean local n-dimensional Minkowski content* if

$$\lim_{r \downarrow 0} \frac{\mathbb{E}[\mathcal{H}^d(\Theta_{n_{\oplus r}} \cap A)]}{b_{d-n}r^{d-n}} = \mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)]$$

for all $A \in \mathcal{B}_{\mathbb{R}^d}$ such that $\mathbb{E}[\mu_{\Theta_n}](\partial A) = 0$. By noticing that the above equality can be written equivalently in this way

$$\lim_{r \downarrow 0} \int_A \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r}})}{b_{d-n}r^{d-n}} \mathrm{d}x = \int_A \lambda_{\Theta_n}(x)\mathrm{d}x, \tag{3}$$

$(A1) - (A3)$ imply the exchange between limit and integral in (3). This is the reason why we named $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ as "Minkowski content"-based estimator.

We recall that the *mean square error* $MSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$ of $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$, defined as usual by

$$MSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) := \mathbb{E}[(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x) - \lambda_{\Theta_n}(x))^2],$$

can be equivalently written as

$$MSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))^2 + Var(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)). \tag{4}$$

If r_N is such that

$$\lim_{N \to \infty} r_N = 0 \quad \text{and} \quad \lim_{N \to \infty} Nr_N^{d-n} = \infty,$$

it follows that $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ is asymptotically unbiased and weakly consistent, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$: in fact it is easy to check (see [9], [29, Proposition 6.1]) that, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned} Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) &= \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r_N}})}{b_{d-n}r_N^{d-n}} - \lambda_{\Theta_n}(x) &(5) \\ Var(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) &= \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r_N}})(1 - \mathbb{P}(x \in \Theta_{n_{\oplus r_N}}))}{N(b_{d-n}r_N^{d-n})^2} \\ &= \frac{\lambda_{\Theta_n}(x)}{Nr_N^{d-n}b_{d-n}} + o\Big(\frac{1}{Nr_N^{d-n}}\Big). &(6) \end{aligned}$$

By the classical kernel density estimation theory (for the reader's convenience we refer to [9, Appendix A1] and references therein), we define the optimal bandwidth $r_N^{o,\mathrm{AMSE}}(x)$ of $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ associated to a point $x \in \mathbb{R}^d$ as

$$r_N^{o,\mathrm{AMSE}}(x) := \arg\min_{r_N} AMSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)), \tag{7}$$

where $AMSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$ is the *asymptotic* approximation of the *mean square error*. In order to do this, a Taylor series expansion for $Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$ (equivalently for $\mathbb{P}(x \in \Theta_{n_{\oplus r_N}})$) is required and left as open problem in Section 4 and Section 6 of [9]. We provide here sufficient conditions which give a solution to this problem.

To fix the notation, in the sequel $\alpha := (\alpha_1, ..., \alpha_d)$ will be a multi-index of $\mathbb{N}_0^d$; we will denote

$$
\begin{array}{rcl}
|\alpha| & := & \alpha_1 + \cdots + \alpha_d \\
\alpha! & := & \alpha_1! \cdots \alpha_d! \\
y^\alpha & := & y_1^{\alpha_1} \cdots y_d^{\alpha_d} \\
D_y^\alpha f(y,s) & := & \dfrac{\partial^{|\alpha|} f(y,s)}{\partial y_1^{\alpha_1} \cdots \partial y_d^{\alpha_d}}; \\
\mathcal{D}^{(\alpha)}(s) & := & \mathrm{disc}(D_y^\alpha f(y,s))
\end{array}
$$

Note that $D_y^\alpha f(y,s) = f$ if $|\alpha| = 0$.

Here and in the following we denote by $\mathrm{reach}(A)$ the *reach* of a compact set $A \subset \mathbb{R}^d$, and by $\Phi_i(A \ \cdot \ )$, $i = 0, \ldots, n$ its *curvature measures*. To make the paper more readable we list here the assumptions which we shall use in the following, and refer to the Appendix for basics definitions and results on sets with positive reach and on curvature measures which we shall use in the sequel:

(R) For any $s \in \mathbf{K}$, $\mathrm{reach} Z(s) > R$, for some $R > 0$, such that there exists a closed set $\Xi(s) \supseteq Z(s)$ such that $\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) Q(\mathrm{d}s) < \infty$ and

$$
\mathcal{H}^n(\Xi(s) \cap B_r(x)) \geq \gamma r^n \qquad \forall x \in Z(s), \ \forall r \in (0,1)
$$

for some $\gamma > 0$ independent of s;

(M2) For any $s \in \mathbf{K}$, $f(\cdot, s)$ is of class C^1 and, for any α with $|\alpha| \in \{0,1\}$, $D_x^\alpha f(\cdot, s)$ is locally bounded such that for any compact $K \subset \mathbb{R}^d$

$$
\sup_{x \in K_{\oplus \mathrm{diam}(Z(s))}} |D_x^\alpha f(x,s)| \leq \widetilde{\xi}_K^{(\alpha)}(s)
$$

for some $\widetilde{\xi}_K^{(\alpha)}(s)$ with

$$
\int_{\mathbf{K}} |\Phi_i|(\Xi(s)) \widetilde{\xi}_K^{(\alpha)}(s) Q(\mathrm{d}s) < \infty \quad \forall i = 0, ..., n.
$$

$(\overline{A3})$ For any $(s,t) \in \mathbf{K}^2$, the function $g(\cdot, s, \cdot, t)$ is continuous and locally bounded such that for any compact sets $K, \overline{K} \subset \mathbb{R}^d$:

$$
\sup_{y \in \overline{K}_{\oplus \mathrm{diam} Z(t)}} \sup_{x \in K_{\oplus \mathrm{diam} Z(s)}} g(x, s, y, t) \leq \xi_{K, \overline{K}}(s,t)
$$

for some $\xi_{K, \overline{K}}(s,t)$ with

$$
\int_{\mathbf{K}^2} \mathcal{H}^n(\Xi(s)) \mathcal{H}^n(\Xi(t)) \xi_{K, \overline{K}}(s,t) Q_{[2]}(\mathrm{d}(s,t)) < +\infty.
$$

Let us notice that

- the above assumption (R) will play here the role of assumption (A1) of Theorem 1; namely, it is known that a lower dimensional set with positive reach is locally the graph of a function of class C^1 (e.g., see [6, p. 204]), and so the rectifiability condition in (A1) is fulfilled. With regard to the existence of $\Xi(s)$ and γ as in (R), we remind that if a deterministic lower dimensional set is the image of some one-to-one Lipschitz function g, then the constant γ can be written in terms of $\mathrm{Lip}(g^{-1})$ and $\mathrm{Lip}(g)$ (see [3, p. 111]). So condition (R) is generally satisfied; we only point out that, when dealing with random sets, the choice of $\Xi(s)$ and of γ (which has to be independent of s) may depend also on the probability law of the grains (see for instance Example 4 in [1]).

- The assumption reach$Z(s) > R$ (which appears also in [9] in the case of homogenous Boolean models), will play a central role in the application of a Steiner type formula (Eq.(45)) in the proof of the two main theorems.

- $(M2)$ implies $(A2)$, while $(\overline{A3})$ together with (R) imply $(A3)$.

- $(\overline{A3})$ is the same assumption which appears in [9] where the authors provide an optimal bandwidth for the so called *kernel estimator* introduced there.

3 Optimal bandwidth: the general case

In this section we state the main result of the paper: a general formula for the optimal bandwidth of $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ by providing a Taylor series expansion of the bias. To this aim we shall have to distinguish two cases: the case in which the dimension n of Θ_n equals $d-1$, and that one in which $0 < n < d-1$, whereas the particular case $n = 0$ will be discussed in Section 4.2. In both cases the following lemma, which might be considered the generalization of Theorem 3.5 in [29] to product spaces, will be useful.

Lemma 2 *Let $A, B \subset \mathbb{R}^d$ be two countably $\mathcal{H}^n$-rectifiable compact sets, and assume that*

$$
\begin{aligned}
\eta_A(B_r(x)) &\geq \gamma_A r^n, &\forall x \in A, \ \forall r \in (0,1) \\
\eta_B(B_r(x)) &\geq \gamma_B r^n, &\forall x \in B, \ \forall r \in (0,1)
\end{aligned}
$$

holds for some constants $\gamma_A, \gamma_B > 0$ and some finite measures $\eta_A, \eta_B \ll \mathcal{H}^n$ in $\mathbb{R}^d$. Then

$$
\lim_{r \downarrow 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{A_{\oplus r}} \int_{B_{\oplus r}} f(y_1, y_2) \mathcal{H}^d(\mathrm{d}y_1) \mathcal{H}^d(\mathrm{d}y_2)
$$

$$
= \int_A \int_B f(y_1, y_2) \mathcal{H}^n(\mathrm{d}y_1) \mathcal{H}^n(\mathrm{d}y_2) \quad (8)
$$

for any non negative and continuous function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

Proof. See Section 6. $\qquad\square$

(For definitions and related results of the curvature measures Φ_i and the support measures μ_i which appear in the statements of the next theorem, see the Appendix.)

3.1 Case $d - n > 1$

Theorem 3 *Let Θ_n be as in (1) satisfying the assumptions (R), $(M2)$ and $(\overline{A3})$, with $0 < n < d-1$. Then*

$$
Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)) = \big(\mathcal{A}_1(x) - \mathcal{A}_2(x)\big) r_N + o(r_N),
$$

where

$$
\mathcal{A}_1(x) := \frac{b_{d-n+1}}{b_{d-n}} \int_K \int_{Z(s)} f(x - y, s) \Phi_{n-1}(Z(s); \mathrm{d}y) Q(\mathrm{d}s) \quad (9)
$$

$$
\mathcal{A}_2(x) := \frac{d-n}{d-n+1} \sum_{|\alpha|=1} \int_K \int_{N(Z(s))} D^\alpha_x f(x - y, s) u^\alpha \mu_n(Z(s); \mathrm{d}(y, u)) Q(\mathrm{d}s)
$$

7

Proof. See Section 6. □

As a corollary, together with equations (4) and (6), we have the following approximation of $MSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$:

$$MSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))^2 + Var(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$$

$$= \underbrace{(\mathcal{A}_1(x) - \mathcal{A}_2(x))^2 r_N^2 + \frac{\lambda_{\Theta_n}(x)}{N b_{d-n} r_N^{d-n}}}_{=: AMSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))}$$

$$+ o(r_N^2) + o\Big(\frac{1}{N r_N^{d-n}}\Big) \quad \text{as } N \to \infty,$$

from which it is easy to get the optimal bandwidth

$$r_N^{o,AMSE}(x) \overset{(7)}{=} \Big(\frac{(d-n)\lambda_{\Theta_n}(x)}{2N b_{d-n}(\mathcal{A}_1(x) - \mathcal{A}_2(x))^2}\Big)^{\frac{1}{d-n+2}}. \tag{10}$$

3.2 Case $d - n = 1$

We are going to proceed along the same lines as in the previous case. In order to do this, referring to the proof of the previous theorem, let us observe that Eq. (29) does not hold whenever $d-n=1$, and that the role of $\sum_{k \geq 2} \mathbb{P}(Y_r \geq k)$ in (32) will be now played by $\sum_{k \geq 3}(2k-1)\mathbb{P}(Y_r \geq k)$. This is the reason why the introduction of a condition on the third factorial moment measure $\nu_{[3]}$ of Ψ, similar to those one on $\nu_{[2]}$ in $(\overline{A3})$, seems to be needed in this case.

Theorem 4 *Let the assumptions of Theorem 3 be satisfied, with $0 < n = d-1$. Moreover let us assume that Ψ has third factorial moment measure*

$$\nu_{[3]}(\mathrm{d}(y_1, s_1, y_2, s_2, y_3, s_3))$$

$$= h(y_1, s_1, y_2, s_2, y_3, s_3)\mathrm{d}y_1 \mathrm{d}y_2 \mathrm{d}y_3 Q_{[3]}(\mathrm{d}(s_1, s_2, s_3))$$

such that:

(M4) for any $(y_1, s_1, s_2, s_3) \in \mathbb{R}^d \times \mathbf{K}^3$, the function $h(y_1, s_1, \cdot, s_2, \cdot, s_3)$ is continuous and locally bounded such that for any compact sets $K, \overline{K} \subset \mathbb{R}^d$ and $a \in \mathbb{R}^d$:

$$\mathbf{1}_{(a-Z(s_1))\oplus 1}(y_1) \sup_{y_2 \in K_{\oplus \mathrm{diam}(Z(s_2))}} \sup_{y_3 \in \overline{K}_{\oplus \mathrm{diam}(Z(s_3))}}$$

$$h(y_1, s_1, y_2, s_2, y_3, s_3) \leq \overline{\xi}_{a,K,\overline{K}}(s_1, y_1, s_2, s_3)$$

for some $\overline{\xi}_{a,K,\overline{K}}(s_1, y_1, s_2, s_3)$ with:

$$\int_{\mathbb{R}^d \times \mathbf{K}^3} \mathcal{H}^n(\Xi(s_2))\mathcal{H}^n(\Xi(s_3))\overline{\xi}_{a,K,\overline{K}}(s_1, y_1, s_2, s_3)$$

$$\mathrm{d}y_1 Q_{[3]}(\mathrm{d}(s_1, s_2, s_3)) < +\infty. \tag{11}$$

Then

$$Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \big(\mathcal{A}_1(x) - \mathcal{A}_3(x)\big)r_N + o(r_N),$$

where

$$\mathcal{A}_1(x) \overset{(9)}{=} \frac{\pi}{2} \int_K \int_{Z(s)} f(x-y,s)\Phi_{d-2}(Z(s);\mathrm{d}y)Q(\mathrm{d}s),$$

$$\mathcal{A}_3(x) := \int_{K^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))}$$
$$g(y_1,s_1,y_2,s_2)\mathcal{H}^{d-1}(\mathrm{d}y_2)\mathcal{H}^{d-1}(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1,s_2)).$$

Proof. See Section 6. □

Analogously to the case $d - n > 1$, as a corollary we have the following approximation of $MSE(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x))$:

$$MSE(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x)) = Bias(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x))^2 + Var(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x))$$

$$= \underbrace{\big(\mathcal{A}_1(x) - \mathcal{A}_3(x)\big)^2 r_N^2 + \frac{\lambda_{\Theta_{d-1}}(x)}{2Nr_N}}_{=:AMSE(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x))}$$

$$+ o(r_N^2) + o\Big(\frac{1}{Nr_N}\Big) \qquad \text{as } N \to \infty,$$

from which it is easy to get the optimal bandwidth

$$r_N^{o,AMSE}(x) = \Big(\frac{\lambda_{\Theta_{d-1}}(x)}{4N\big(\mathcal{A}_1(x) - \mathcal{A}_3(x)\big)^2}\Big)^{\frac{1}{3}}, \qquad (12)$$

Remark 5 *Even if the assumption (M4) may seem quite technical, it is generally satisfied in applications. Indeed it is easily checked for inhomogeneous Boolean models (see Section 4.3), and it is not difficult to show that it is fulfilled, for instance, when Ψ is an independent marking of a Matèrn cluster process in $\mathbb{R}^2$. We give here a sketch of the proof.*
Let $\widetilde{\Psi}$ be Matèrn cluster process in $\mathbb{R}^2$ (e.g., see [4, p. 19]) with parent Poisson point process Υ having intensity α, and whose clusters consist of $\mathcal{M} \sim \text{Poisson}(m)$ points independently and uniformly distributed in the ball $B_R(x)$, where x is the center of the cluster; then it is well known that its intensity function $f_{\widetilde{\Psi}}$ and its second moment density $g_{\widetilde{\Psi}}$ are given by $f_{\widetilde{\Psi}} = m\alpha$ and $g_{\widetilde{\Psi}}(x,y) = \alpha^2 m^2 + \alpha m^2 \mathcal{H}^2(B_R(x) \cap B_R(y))/(\pi^2 R^4)$ (e.g., see [4, p. 35]). By arguing similarly to the proof of the expression for $g_{\widetilde{\Psi}}$ it is possible to prove that the third moment density $h_{\widetilde{\Psi}}$ of $\widetilde{\Psi}$ is given by

$$h_{\widetilde{\Psi}}(x_1,x_2,x_3) =$$

$$(m\alpha)^3 + \frac{\alpha m^3}{\pi^3 R^6}\mathcal{H}^2(B_R(x_1) \cap B_R(x_2) \cap B_R(x_3))$$

$$+ \frac{\alpha^2 m^3}{\pi^2 R^4}\Big(\mathcal{H}^2(B_R(x_1) \cap B_R(x_2)) + \mathcal{H}^2(B_R(x_1) \cap B_R(x_3))$$

$$+ \mathcal{H}^2(B_R(x_2) \cap B_R(x_3))\Big). \quad (13)$$

Indeed, by definition, the third moment factorial measure $\widetilde{\nu}_{[3]}$ of $\widetilde{\Psi}$ can be written as follows $\forall A, B, C \in \mathcal{B}_{\mathbb{R}^2}$:

$$\widetilde{\nu}_{[3]}(A \times B \times C)$$

$$= \mathbb{E}\Big[\sum_{\substack{x_1, x_2, x_3 \in \widetilde{\Psi} \\ x_1 \neq x_2, x_1 \neq x_3, x_2 \neq x_3}} \mathbf{1}_A(x_1)\mathbf{1}_B(x_2)\mathbf{1}_C(x_3) \Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{\substack{x_1, x_2, x_3 \in \widetilde{\Psi} \\ x_1 \neq x_2, x_1 \neq x_3, x_2 \neq x_3}} \mathbf{1}_A(x_1)\mathbf{1}_B(x_2)\mathbf{1}_C(x_3) \Big| \Upsilon \Big]\Big].$$

The expectation above can be handled by dividing the sum in five cases: x_1, x_2, x_3 belong to the same cluster; only two points among x_1, x_2, x_3 belong to the same cluster and the other point to a different one; x_1, x_2, x_3 belong to different clusters. Finally, simple applications of the Campbell theorem and the independence between clusters yield the following final expression for $\widetilde{\nu}_{[3]}$:

$$\widetilde{\nu}_{[3]}(A \times B \times C) =$$

$$\int_{A \times B \times C} \Big[(m\alpha)^3 + \frac{\alpha m^3}{\pi^3 R^6} \mathcal{H}^2(B_R(x_1) \cap B_R(x_2) \cap B_R(x_3))$$

$$+ \frac{\alpha^2 m^3}{\pi^2 R^4} \Big(\mathcal{H}^2(B_R(x_1) \cap B_R(x_2)) +$$

$$\mathcal{H}^2(B_R(x_1) \cap B_R(x_3)) + \mathcal{H}^2(B_R(x_2) \cap B_R(x_3)) \Big) \Big] \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3,$$

and so (13).

Being Ψ an independent marking of $\widetilde{\Psi}$, $Q_{[3]}(\mathrm{d}(s_1, s_2, s_3)) = Q(\mathrm{d}s_1)Q(\mathrm{d}s_2)Q(\mathrm{d}s_3)$, and its third factorial moment density h does not depend on (s_1, s_2, s_3) and coincides with $h_{\widetilde{\Psi}}$. By noticing that

$$h(x_1, x_2, x_3) \leq (m\alpha)^3 + \frac{\alpha m^3}{\pi^2 R^4} + \frac{3\alpha^2 m^3}{\pi R^2} =: c \in \mathbb{R}_+,$$

we can choose $\overline{\xi}_{a,K,\overline{K}}(s_1, y_1, s_2, s_3) = c\mathbf{1}_{(a-Z(s_1))_{\oplus 1}}(y_1)$ in (M4); then condition (11) is directly implied by the assumption (R).

We refer to [30, Example 2, 3] as hint for the construction of other examples satisfying (M4).

4 Optimal bandwidth: particular cases

In this section we show how the general formulas for the optimal bandwidth of $\widehat{\lambda}_{\Theta_n}^{\mu, N}(x)$ obtained in the previous section specialize in some particular cases of interest.

4.1 Θ_n stationary

The case in which Θ_n is stationary has been extensively studied in literature, and so some other different methods may be used in such a case; for instance by making use of sufficiently large observation windows, or estimating one or more characteristic quantities of the model by means of stereological methods. We refer to [7, 11] and references therein, for more exhaustive discussions about this; here we just want to observe how the general expressions for $\widehat{\lambda}_{\Theta_n}^{\mu, N}(x)$ and its optimal bandwidth simplify in the stationary case.

Let us assume that $\Lambda(\mathrm{d}(x,s)) = c\mathrm{d}xQ(\mathrm{d}s)$; i.e. $f(x,s) \equiv c$ for any $(x,s) \in \mathbb{R}^d \times \mathbf{K}$; then $\Psi = \{(x_i,s_i)\}_{i\in\mathbb{N}}$ is an independent marking of the marginal process $\{x_i\}_{i\in\mathbb{N}}$, which is stationary, and so Θ_n is a stationary random closed set as well. Then as a corollary of Theorem 3 and Theorem 4, the following statement is easily proved.

Corollary 6 *Let Θ_n be a germ-grain model as in (1), with intensity measure $\Lambda(\mathrm{d}(y,s)) = c\mathrm{d}yQ(\mathrm{d}s)$, and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,s,y,t)\mathrm{d}x\mathrm{d}yQ_{[2]}(\mathrm{d}(s,t))$, satisfying (R), $\overline{(A3)}$, and such that $\mathbb{E}[\Phi_i(Z)] < \infty$ for all $i = 0,\ldots,n-1$.*

a) If $d - n > 1$, then, for any $x \in \mathbb{R}^d$, the optimal bandwidth is given by

$$r_N^{o,AMSE} = \Big(\frac{b_{d-n}(d-n)\mathbb{E}_Q[\mathcal{H}^n(Z)]}{2Nb_{d-n+1}^2 c\mathbb{E}_Q[\Phi_{n-1}(Z)]^2} \Big)^{\frac{1}{d-n+2}}. \tag{14}$$

b) If $d - n = 1$, and moreover Ψ has third factorial moment measure

$$\nu_{[3]}(\mathrm{d}(y_1,s_1,y_2,s_2,y_3,s_3))$$
$$= h(y_1,s_1,y_2,s_2,y_3,s_3)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3 Q_{[3]}(\mathrm{d}(s_1,s_2,s_3))$$

satisfying the assumption (M4), then, for any $x \in \mathbb{R}^d$, the optimal bandwidth is given by

$$r_N^{o,AMSE} = \Big(\frac{c\mathbb{E}[\mathcal{H}^{d-1}(Z)]}{D} \Big)^{1/3}$$

with

$$D := N\Big(\pi c\mathbb{E}[\Phi_{d-2}(Z)] - 2 \int_{\mathbf{K}^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))}$$
$$g(y_1,s_1,y_2,s_2)\mathcal{H}^{d-1}(\mathrm{d}y_2)\mathcal{H}^{d-1}(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1,s_2)) \Big)^2$$

Proof. See Section 6. □

Remark 7 (Boolean model) *Corollary 6 applied to the Boolean case coincides with Proposition 18 in [9].*

4.2 Case $n = 0$

In this section we specialize the "Minkowski content"-based estimator for the case $n = 0$, which includes relevant situations already treated in literature, namely: random vectors [23, 24] (for a general treatment see e.g. [26, 27, 8]), point processes, cluster point processes (see e.g. [16], [13, page 629]), and the recent paper [28]). We point out that the case $n = 0$ requires further expansion in the asymptotic approximation of the bias of $\widehat{\lambda}_{\Theta_0}^{\mu,N}(x)$.

In Theorem 3 and Theorem 4, the grains $Z(s)$ of the germ-grain model Θ_n as in (1) were assumed to be $\mathcal{H}^n$-rectifiable and compact, with positive reach. By observing that countably $\mathcal{H}^0$-rectifiable sets correspond to countable sets, and that a compact countable set with positive reach is a finite union of points, throughout this section Θ_0 will be a locally finite union of points, and so a point process, say $\widetilde{\Psi}$.

In order to proceed as in the previous sections, let us remind (see, e.g. [9, Corollary 13]) that any point process $\widetilde{\Psi}$ in $\mathbb{R}^d$ with intensity $f_{\widetilde{\Psi}}$ and second factorial moment density g, may be seen as a trivial germ-grain model Θ_0 as in (1), with Ψ marked point process in $\mathbb{R}^d$ with mark space $\mathbf{K} = \mathbb{R}^d$ ($Z(s) \equiv s$) having intensity measure $\Lambda(\mathrm{d}(y,s)) = f_{\widetilde{\Psi}}(y)\mathrm{d}y\delta_0(s)\mathrm{d}s$, and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,y)\mathrm{d}x\mathrm{d}y\delta_0(s)\delta_0(t)\mathrm{d}s\mathrm{d}t$, where δ_0 is the usual Dirac-delta in 0. Analogously for the third factorial moment measure. We also recall that, in such a case, $\lambda_{\Theta_0} \equiv f_{\widetilde{\Psi}}$, and

$$\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x) = \widehat{\lambda}_{\Theta_0}^{\mu,N}(x) :=$$

$$\frac{\sum_{i=1}^N \mathbf{1}_{\Theta_0^i \cap B_{r_N(x)} \neq \emptyset}}{N b_{d-n} r_N^{d-n}} = \frac{\sum_{i=1}^N \mathbf{1}_{\widetilde{\Psi}(B_{r_N(x)})>0}}{N b_{d-n} r_N^{d-n}}.$$

Throughout this section let us assume that $\widetilde{\Psi}$ is a point process in $\mathbb{R}^d$ with intensity $f_{\widetilde{\Psi}}(y)$, second factorial moment density g, and third factorial moment density h.
In order to provide an optimal bandwidth $r_N^{o,AMSE}(x)$ for $\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)$, we have to distinguish three cases: $d = 1$, $d = 2$, and $d > 2$.

4.2.1 Case $d = 1$

In order to proceed along the same lines of Theorem 4, let us further assume that $\widetilde{\Psi}$ has third factorial moment density h continuous. Then the third moment measure associate to the marked point process Ψ is of the type

$$\nu_{[3]}(\mathrm{d}(y_1, s_1, y_2, s_2, y_3, s_3)) =$$

$$h(y_1, y_2, y_3)\mathrm{d}y_1\mathrm{d}_2\mathrm{d}y_3\delta_0(s_1)\delta_0(s_2)\delta_0(s_3)\mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3.$$

It is easy to see that the assumptions of Theorem 4 are trivially satisfied (take $\Xi(s) \equiv Z(s)$); by proceeding along the same lines of its proof it is easy to check that now $Bias(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) = -g(x,x)r_N + o(r_N)$, while $Var(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) = f_{\widetilde{\Psi}}(x)/2Nr_N + o(1/Nr_N)$. It follows then

$$r_N^{o,AMSE}(x) = \sqrt[3]{\frac{f_{\widetilde{\Psi}}(x)}{4Ng(x,x)^2}}. \tag{15}$$

(which is just Eq. (12) with $d = 1$ and $\mathcal{A}_1(x) = 0$).

4.2.2 Case $d \geq 2$

Throughout this section assume that $f_{\widetilde{\Psi}} \in C^2$ and that g is continuous.
By the proof of Theorem 3 we have:

$$\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r}) = \mathbb{P}(x \in \Theta_{0 \oplus r}) \overset{(32)}{=} \mathbb{E}[Y_r] - \sum_{k \geq 2} \mathbb{P}(Y_r \geq k),$$

where Y_r is the random variable defined in (30), and

$$\sum_{k \geq 2} \mathbb{P}(Y_r \geq k) \overset{(31)}{\leq} \int_{B_r(x)} \int_{B_r(x)} g(y_1, y_2)\mathrm{d}y_1\mathrm{d}y_2 \overset{(29)}{=} O(r^{2d}).$$

Therefore,

$$\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r}) = \int_{B_r(x)} f_{\widetilde{\Psi}}(y)\mathrm{d}y + O(r^{2d})$$

$$= r^d \int_{B_1(0)} f_{\widetilde{\Psi}}(x + rz)\mathrm{d}z + O(r^{2d}). \quad (16)$$

Whereas in the general case with $n > 0$ a first order Taylor expansion is enough, here a second order one is needed, being $\int_{B_1(0)} z_k = 0$ for every $k = 1, ..., d$; in fact:

$$\int_{B_1(0)} f_{\widetilde{\Psi}}(x + rz)\mathrm{d}z = \int_{B_1(0)} \Big(f_{\widetilde{\Psi}}(x) +$$

$$\sum_{|\alpha|=1} \frac{1}{\alpha!} D_x^\alpha f_{\widetilde{\Psi}}(x) z^\alpha r + \sum_{|\alpha|=2} \frac{1}{\alpha!} D_x^\alpha f_{\widetilde{\Psi}}(x + \theta r z) z^\alpha r^2 \Big)\mathrm{d}z$$

with $\theta \in (0, 1)$. Since $f_{\widetilde{\Psi}} \in C^2$, it is easily observed that

$$\lim_{r \downarrow 0} \int_{B_1(0)} \sum_{|\alpha|=2} \frac{1}{\alpha!} D_x^\alpha f_{\widetilde{\Psi}}(x + \theta r z) z^\alpha \mathrm{d}z$$

$$= \sum_{|\alpha|=2} \frac{1}{\alpha!} D_x^\alpha f_{\widetilde{\Psi}}(x) \int_{B_1(0)} z^\alpha \mathrm{d}z := \mathcal{A}_4(x); \quad (17)$$

therefore we get

$$\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r}) \stackrel{(16)}{=} b_d r^d f_{\widetilde{\Psi}}(x) +$$

$$r^{d+2} \mathcal{A}_4(x) + o(r^{d+2}) + O(r^{2d}). \quad (18)$$

Thus we have to distinguish two cases:

1) Case $d > 2$
Then $O(r^{2d}) = o(r^{d+2})$, and so

$$Bias(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) \stackrel{(5)}{=} \frac{\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r_N})}{b_d r_N^d} - f_{\widetilde{\Psi}}(x)$$

$$= \frac{1}{b_d} \mathcal{A}_4(x) r_N^2 + o(r_N^2) \qquad \text{as } N \to \infty,$$

and

$$MSE(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) \stackrel{(4)}{=} \frac{1}{b_d^2} \mathcal{A}_4(x)^2 r_N^4 + \frac{f_{\widetilde{\Psi}}(x)}{b_d N r_N^d}$$

$$+ o\Big(\frac{1}{N r_N^d}\Big) + o(r_N^4) \quad \text{as } N \to \infty.$$

Finally we get following expression for the optimal bandwidth of $\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)$ at the point $x \in \mathbb{R}^d$:

$$r_N^{o,AMSE}(x) = \sqrt[4+d]{\frac{d b_d f_{\widetilde{\Psi}}(x)}{4 N \mathcal{A}_4^2(x)}}. \quad (19)$$

13

2) Case $d = 2$

Then $O(r^{2d}) \neq o(r^{d+2})$, and so a further Taylor expansion of $\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r})$ is needed in (18). Under the assumption $f_{\widetilde{\Psi}} \in C^2$, g and h continuous, by mimicking the proof of Theorem 4, it is easy to observe

$$\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r}) \stackrel{(38)}{=} \frac{3}{2}\mathbb{E}[Y_r] - \frac{1}{2}\mathbb{E}[Y_r^2] + \sum_{k \geq 3}(k-2)\mathbb{P}(Y_r \geq k),$$

with

- $\sum_{k \geq 3}(k-2)\mathbb{P}(Y_r \geq k) = o(r^4)$,

- $\mathbb{E}[Y_r] = r^2 \int_{B_1(0)} f_{\widetilde{\Psi}}(x + rz)\mathrm{d}z \stackrel{(18)}{=} \pi r^2 f_{\widetilde{\Psi}}(x) + r^4 \mathcal{A}_4(x) + o(r_N^4)$,

- $\mathbb{E}[Y_r^2] = \mathbb{E}[Y_r] + \pi^2 r^4 g(x,x) + o(r^4)$. Therefore

$$Bias(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) = \frac{\mathbb{P}(x \in \widetilde{\Psi}_{0 \oplus r_N})}{\pi r_N^2} - f_{\widetilde{\Psi}}(x)$$

$$= \big(\frac{\mathcal{A}_4(x)}{\pi} - \frac{\pi}{2}g(x,x)\big)r_N^2 + o(r^4), \qquad \text{as } N \to \infty,$$

and

$$MSE(\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)) = \big(\frac{\mathcal{A}_4(x)}{\pi} - \frac{\pi}{2}g(x,x)\big)^2 r_N^4 + \frac{f_{\widetilde{\Psi}}(x)}{\pi N r_N^2}$$

$$+ o\Big(\frac{1}{N r_N^2}\Big) + o(r_N^4) \quad \text{as } N \to \infty.$$

Finally we get following expression for the optimal bandwidth of $\widehat{f}_{\widetilde{\Psi}}^{\mu,N}(x)$ at the point $x \in \mathbb{R}^d$:

$$r_N^{o,AMSE}(x) = \sqrt[6]{\frac{2\pi f_{\widetilde{\Psi}}(x)}{N\big(2\mathcal{A}_4(x) - \pi^2 g(x,x)\big)^2}}. \tag{20}$$

4.2.3 Random Variables

An absolutely continuous random vector X in $\mathbb{R}^d$ admits a probability density function $f_X(x)$, $x \in \mathbb{R}^d$. So, as a random closed set, its mean density λ_X coincides with f_X. The random vector has also a germ-grain representation, which is characterized by the marked point process $\Psi = \{(X,s)\}$ in $\mathbb{R}^d$ with mark space $\mathbf{K} = \mathbb{R}^d$ and intensity measure $\Lambda(\mathrm{d}(y,s)) = f_X(y)\mathrm{d}y\delta_0(s)\mathrm{d}s$. Assuming $f_X \in C^2$ all the required assumptions are trivially satisfied and, if $d \geq 2$, the expressions of the optimal bandwidth (19) and (20) coincide, since $g \equiv 0$, and we have

$$r_N^{o,AMSE}(x) = \sqrt[4+d]{\frac{db_d f_X(x)}{4N(\sum_{|\alpha|=2}\frac{1}{\alpha!}\int_{B_1(0)} z^\alpha \mathrm{d}z D^\alpha f_X(x))^2}}. \tag{21}$$

Note that if $d = 1$ formula (15) does not apply, because the denominator equals zero. By finding additional terms in the bias expansion (as for the case $d \geq 2$), it can be easily checked that (21) is true also if $d = 1$, and it yields:

$$r_N^{o,AMSE}(x) = \sqrt[5]{\frac{9 f_X(x)}{2N(f_X''(x))^2}}.$$

The above expression coincides with the well known expression for the optimal bandwidth of the kernel density estimator with kernel $k(y) = \frac{1}{2}\mathbf{1}_{[-1,1]}(y)$ (see Eq. (A.2) in [9]); this is in accordance with the fact that these two estimators coincide in this case.

4.3 Inhomogeneous Boolean models

(For the homogeneous case see Remark 7.)

Boolean models are germ-grain processes as in (1) characterized by Ψ being a Poisson point process; this implies the independence of the grains, and, as a consequence, they are computationally much easier to handle. This is the reason why there is a lot of literature on Boolean models, in particular on homogeneous Boolean models. The inhomogeneous case is of course of major interest, but results on this direction are available only in the most recent literature. It is worth noting that the generalization from the homogeneous case to the inhomogeneous one is often anything but trivial, as discussed (and left as open problem which we try to solve here) in [9, Section 4].

Let Θ_n be a Boolean model with integer Hausdorff dimension $0 < n < d$, where the associated Poisson marked point process Ψ has intensity measure $\Lambda(\mathrm{d}(x,s)) = f(x,s)\mathrm{d}xQ(\mathrm{d}s)$, satisfying the assumptions (R) and $(M2)$

By remembering that in such a case $\nu_{[2]} = \Lambda \otimes \Lambda$, and $\nu_{[3]} = \Lambda \otimes \Lambda \otimes \Lambda$, and so as a consequence, $g(x_1,s_1,x_2,s_2)$
$= f(x_1,s_1)f(x_2,s_2)$, $Q_{[2]}(\mathrm{d}(s_1,s_2)) = Q(\mathrm{d}s_1)Q(\mathrm{d}s_2)$, $h(x_1,s_1,x_2,s_2,x_3,s_3) = f(x_1,s_1)f(x_2,s_2)f(x_3,s_3)$, and $Q_{[3]}(\mathrm{d}(s_1,s_2,s_3)) = Q(\mathrm{d}s_1)Q(\mathrm{d}s_2)Q(\mathrm{d}s_3)$, it is easy to check that the assumptions $(\overline{A3})$ and $(M4)$ are trivially implied by $(M2)$; therefore Eqs. (10) and (12) hold. In particular note that the quantity $\mathcal{A}_3(x)$ defined in Theorem 4 simplifies now as follows:

$$\mathcal{A}_3(x) = \int_{\mathbf{K}^2}\int_{(x-Z(s_1))}\int_{(x-Z(s_2))}$$
$$f(x_1,s_1)f(x_2,s_2)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_1)Q(\mathrm{d}s_1)Q(\mathrm{d}s_2)$$
$$\overset{(2)}{=} \left(\lambda_{\Theta_{d-1}}(x)\right)^2.$$

This answers to the problem risen in [9, Sect. 4], providing a general formula for the optimal bandwidth in the inhomogeneous case.

As a simple example of applicability of such formulas, let us consider the inhomogeneous Boolean model of segments discussed in [9, Sect. 5]:

Example 1 *Let Θ_1 be an inhomogeneous Boolean model of segments in $\mathbb{R}^2$ with random length L and uniform orientation, so that the mark space is $\mathbf{K} = \mathbb{R}_+ \times [0,2\pi]$; for all $s = (l,\alpha) \in \mathbf{K}$, let $Z(s) := \{(u,v) \in \mathbb{R}^2 : u = \tau\cos\alpha, \ v = \tau\sin\alpha, \ \tau \in [0,l]\}$ be the segment with length $l \in \mathbb{R}_+$, and orientation $\alpha \in [0,2\pi]$. Denoted by $\mathbb{P}_L(\mathrm{d}l)$ the probability law of the random length L, we assume that $\int_{\mathbb{R}_+} l^3\mathbb{P}_L(\mathrm{d}l) < \infty$. Finally we assume that the segment process Θ_1, represented as in (1), is driven by the marked Poisson process Ψ in $\mathbb{R}^2 \times \mathbf{K}$ having intensity measure $\Lambda(\mathrm{d}y \times \mathrm{d}s) = f(y)\mathrm{d}yQ(\mathrm{d}s)$, with $f(y) = f(y_1,y_2) = c(y_1^2 + y_2^2)$, where $c > 0$, and $Q(\mathrm{d}s) = \frac{1}{2\pi}\mathrm{d}\alpha\mathbb{P}_L(\mathrm{d}l)$. It is easy to check that*

$$\lambda_{\Theta_1}(x_1,x_2) = c(x_1^2 + x_2^2)\mathbb{E}[L] + \frac{1}{3}c\mathbb{E}[L^3],$$

for $\mathcal{H}^2$-a.e. $x = (x_1,x_2) \in \mathbb{R}^2$.
Since $d - n = 1$ the optimal bandwidth is given by (12).

By denoting $O = (0,0)$ and $P(s) = P((l, \alpha)) = (l \cos \alpha, l \sin \alpha)$ the endpoints of the segment $Z(s)$, we may observe that

$$\Phi_0(Z(s); \mathrm{d}y) \stackrel{(50)}{=} \frac{1}{2} \big(\delta_O(y) + \delta_{P(s)}(y) \big) \mathrm{d}y,$$

and so we get

$$\mathcal{A}_1(x) = \frac{\pi}{2} \int_K \int_{Z(s)} f(x-y, s) \Phi_0(Z(s); \mathrm{d}y) Q(\mathrm{d}s)$$

$$= \frac{\pi}{4} \int_K \big(f(x) + f(x - P(s)) Q(\mathrm{d}s)$$

$$= \frac{c\pi}{4} \int_0^\infty \int_0^{2\pi} \big(2(x_1^2 + x_2^2) + l^2 - 2lx_1 \cos \alpha - 2lx_2 \sin \alpha \big) \mathrm{d}\alpha$$

$$\frac{1}{2\pi} \mathbb{P}_L(\mathrm{d}l) = \frac{c\pi}{2}(x_1^2 + x_2^2) + \frac{c\pi}{4} \mathbb{E}[L^2].$$

Thus,

$$r_N^{o, AMSE}(x) \stackrel{(12)}{=} \left(\frac{c(x_1^2 + x_2^2)\mathbb{E}[L] + \frac{1}{3} c \mathbb{E}[L^3]}{D} \right)^{1/3},$$

with

$$D :=$$

$$4N \left(\frac{c\pi}{2}(x_1^2 + x_2^2) + \frac{c\pi}{4} \mathbb{E}[L^2] - \big(c(x_1^2 + x_2^2)\mathbb{E}[L] + \frac{1}{3} c \mathbb{E}[L^3] \big)^2 \right)^2$$

in accordance with the result showed in [9, Sect. 5].

5 Concluding remarks and numerical experiments in comparison with other estimators

As mentioned in the Introduction, a kernel-type estimator for the mean density $\lambda_{\Theta_n}(x)$ has been introduced in [9], so defined:

$$\widehat{\lambda}_{\Theta_n}^{\kappa, N}(x) := \frac{1}{N r_N^d} \sum_{i=1}^N \int_{\Theta_n^i} k\left(\frac{x-y}{r_N} \right) \mathcal{H}^n(dy),$$

where $k : \mathbb{R}^d \to \mathbb{R}$ is a multivariate kernel. Under suitable regularity assumptions on Θ_n, analogous to those ones introduced here, explicit formulas for the associated optimal bandwidth have been provided. In particular, by choosing the kernel function

$$k(y) = \frac{1}{b_d} \mathbf{1}_{B_1(0)}(y),$$

the above kernel-type estimator coincides with the so-called *natural estimator* $\widehat{\lambda}_{\Theta_n}^{\nu, N}(x)$ (also this introduced in [9], and arising as a "natural" consequence of Besicovitch derivation theorem), so defined:

$$\widehat{\lambda}_{\Theta_n}^{\nu, N}(x) := \frac{1}{N b_d r_N^d} \sum_{i=1}^N \mathcal{H}^n(\Theta_n^i \cap B_{r_N}(x)).$$

16

In [10] some numerical experiments comparing these estimators are provided. Of course the "Minkowski content"-based estimator was not completely discussed there, because of the lack of general formulas for the optimal bandwidth proved here.

So, in this section, we want to investigate mainly similarity and/or dissimilarity of these three estimators ($\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$, $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ and $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$). First of all it is worth noting that, by their definition, it is much more difficult to compute $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ and $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ than $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$. Indeed it is easily observed that, given an i.i.d. random sample $\{\Theta_n^i\}_{i\in\mathbb{N}}$ of Θ_n, in order to compute $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ it is enough to count how many Θ_n^i of the sample have not void intersection with $B_r(x)$; since usually in applications $\{\Theta_n^i\}$ is a sequence of digital images given in terms of pixels, $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ is easily computed by checking if any pixel of Θ_n^i belongs also to $B_r(x)$, however complicated Θ_n^i is. On the other hand, the computation of the integral in the definition of $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ and of the measure $\mathcal{H}^n(\Theta_n^i \cap B_{r_N}(x))$ in the definition of $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ might be computationally difficult.

In this section we will discuss some simple relevant examples of random sets in $\mathbb{R}^2$: the estimation of the intensity of an inhomogeneous and of a homogeneous point process (Section 5.1), and the estimation of the mean density of a homogenous and of an inhomogeneous segment process (Section 5.2). We point out that the intensity function of a point process will be estimated through $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$, $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ and $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$, without much effort because a point process is a random closed set with Hausdorff dimension $n = 0$ and so $\mathcal{H}^0$ in the definition of $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ and $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ is easily computed. This is not the case of the examples regarding segment processes, that is random sets with Hausdorff dimension $n = 1$. Even if in the inhomogeneous case discussed in Section 5.2.1 we consider the simpler case of random segments with horizontal orientation, we do not provide estimates of the mean density by the kernel-type estimator because they would require non trivial computation. Instead, due to the fact that we choose segments parallel to the x-axis, the length of their intersection with the ball $B_{r_N}(x)$ in the definition of $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ may be easily obtained by multiplying the number of the pixels corresponding to each segment in the ball by the length of a single pixel (in all the examples discussed here 1pixel $= 0.0029$). We can not state the same in the case of segments with random orientation discussed in Section 5.2.2, where we estimate the mean density of the random closed set only by means of $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$. Thus, even if the natural estimator results to be more stable than the "Minkowski content"-based estimator (it provides a quite good estimation only near to the optimal bandwidth), it is evident that this last estimator is the most feasible.

Therefore, possible open problems of interest for future work could be to find upper and lower bounds for r_N in terms of the estimation error. We are also aware that, as in the classical kernel density estimation theory for random variables (e.g., see [26]), the optimal bandwidths of all the proposed estimators depend on the unknown quantities f, g and Q of the marked point process Ψ describing Θ_n; their estimation could be investigated for example by means of plug-in methods as discussed in [18].

5.1 Point processes

We apply here the results shown in Section 4.2 to two different types of point processes $\widetilde{\Psi}$ in $\mathbb{R}^2$, comparing the three different estimators mentioned above. First we consider an inhomogeneous Poisson point process (Section 5.1.1), then a homogeneous Matèrn cluster process (Section 5.1.2). An i.i.d. random sample $\widetilde{\Psi}_1, \ldots, \widetilde{\Psi}_N$ of size N for $\widetilde{\Psi}$ is assumed to be available. Both the Poisson

and the Matèrn cluster point process have been simulated in the square $[-3, 3]^2$, while the estimates of their intensities have been evaluated in the points of a grid with step size 0.2 in the compact window $W := [-2, 2]^2$ in order to avoid edge effects.

5.1.1   The inhomogeneous case

Let $\widetilde{\Psi}$ be an inhomogeneous Poisson point process in $\mathbb{R}^2$ with intensity function $f_{\widetilde{\Psi}}(x) = x_1^2 + x_2^2$. Such a process has been discussed in [9], where the optimal bandwidth associated to $\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}$ and $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}$ is provided. Figure 1 and Figure 2 shows the estimates of $f_{\widetilde{\Psi}}$ by means of the three estimators $\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}$, $\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}$ and $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}$, in the case of a sample with size $N = 1000$ and $N = 10000$, respectively. We used the Epanechnikov kernel, i.e.

$$k(t) = \begin{cases} \frac{2}{\pi}(1 - x_1^2 - x_2^2) & \text{if } (x_1, x_2) \in B_1(0) \\ 0 & \text{otherwise,} \end{cases}$$

for the kernel-type estimator $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}$. At each point of the grid the theoretical optimal bandwidth associated to each estimator has been evaluated (by (20) and [9, Eq. (21)]):

$$r_N^{o,AMSE}(x) = \begin{cases} \sqrt[6]{\dfrac{2(x_1^2 + x_2^2)}{N\pi(1 - \pi(x_1^2 + x_2^2)^2)^2}} & \text{for } \widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}(x) \\[3ex] \sqrt[6]{\dfrac{6(x_1^2 + x_2^2)}{N\pi}} & \text{for } \widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}(x) \\[3ex] \sqrt[6]{\dfrac{2(x_1^2 + x_2^2)}{N\pi}} & \text{for } \widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}(x) \end{cases}$$

Figure 1, Figure 2 and Table 1 show that the rate of convergence of $\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}$ and $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}$ is higher than $\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}$, and that the estimation error decreases as N increases, as expected, for all the three estimators.

N	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}(x) - f_{\widetilde{\Psi}}(x)|$	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}(x) - f_{\widetilde{\Psi}}(x)|$	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}(x) - f_{\widetilde{\Psi}}(x)|$
1000	1.820291	0.324669	0.293807
10000	0.760128	0.169820	0.161852

Table 1: Maximum distance between the intensity of the process and its estimates at the grid points in $W = [-2, 2]^2$

5.1.2   The homogeneous case

As an example of a non Poisson point process, let us consider the Matèrn cluster process $\widetilde{\Psi}$ in $\mathbb{R}^2$ introduced in Remark 5; then the optimal bandwidth associated to $\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}(x)$ is given by

$$r_N^{o,AMSE} \stackrel{(20)}{=} \sqrt[6]{\frac{2R^4 m\alpha}{N\pi(\pi R^2 \alpha^2 m^2 + \alpha m^2)^2}}; \tag{22}$$

of course it does not depend on x, since the process is stationary. We recall that $r_N^{o,AMSE} = +\infty$ for $\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}(x)$ and $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}(x)$ (see [9]). Figures 3 and 4 compare the estimation of the intensity of $\widetilde{\Psi}$ with $\alpha = 5$, $m = 5$ and $R = 0.1$, in the compact window $W = [-2,2]^2$, based on a sample with size $N = 1000$ and $N = 10000$, respectively. With regard to $\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}(x)$, the theoretical optimal value of $r_N^{o,AMSE}$ given in (22) has been employed, whereas $r_N^{o,AMSE} = 1$ has been chosen for $\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}(x)$ and $\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}(x)$ in order to avoid edge effects.

As for the previous example, from the figures and from Table 2 we may conclude that the rate of convergence of kernel-type estimators is higher than that one of the "Minkowski content"-based estimator. But as we previously observed, this last one estimator is easier to compute, as we will emerge in the next examples.

N	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\mu,N}(x) - f_{\widetilde{\Psi}}(x)|$	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\nu,N}(x) - f_{\widetilde{\Psi}}(x)|$	$\max_{x \in W} |\widehat{\lambda}_{\widetilde{\Psi}}^{\kappa,N}(x) - f_{\widetilde{\Psi}}(x)|$
1000	13.731913	0.582247	0.603758
10000	6.226284	0.183946	0.192639

Table 2: Maximum distance between the intensity of the process and its estimates at the grid points in $W = [-2,2]^2$.

5.2 Segment processes

We consider now segment processes Θ_1 in $\mathbb{R}^2$: in Section 5.2.1 a non Boolean germ-grain model of horizontal segments, and in Section 5.2.2 the segment Boolean model introduced in Example 1. In this last case we provide estimates of $\lambda_{\Theta_1}(x)$ only by the "Minkowski content"-based estimator, because, as already mentioned before, the other kinds of estimators would require non-trivial calculations. The observation window $W := [0,1]^2$ has been chosen in all the proposed simulations.

5.2.1 Horizontal segments

As an example of a stationary non Boolean germ-grain model, let us consider a random closed set Θ_1 in $\mathbb{R}^2$ driven by Ψ which is an independent marking of a Matèrn cluster process $\widetilde{\Psi}$, as introduced in Remark 5; the grains $Z(s)$ are assumed to be segments parallel to the x-axis, with random length L. Therefore Ψ is a marked point process in $\mathbb{R}^2$ with marks in $\mathbb{R}_+$, and with intensity measure $\Lambda(\mathrm{d}(x,l)) = m\alpha \mathrm{d}x Q(\mathrm{d}l)$ and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = \left(\alpha^2 m^2 + \alpha m^2 \frac{\mathcal{H}^2(B_R(x) \cap B_R(y))}{\pi^2 R^4}\right) \mathrm{d}x \mathrm{d}y Q(\mathrm{d}s) Q(\mathrm{d}t)$, where $Q(\mathrm{d}l)$ is the probability law of L.

In our numerical experiments we assumed $\alpha = 20$, $m = 5$, $R = 0.2$ and $L \sim U(0,R)$ (see Figure 5 for a realization of Θ_1). By applying the general formula in Corollary 6 (b), it follows

$$r_N^{o,AMSE} =$$

$$\left(\frac{m\alpha R}{N(4\pi m\alpha - 2\alpha^2 m^2 R^2 - \frac{2\alpha m^2}{9\pi^2}(2\pi - 27\sqrt{3} + 64))^2}\right)^{\frac{1}{3}}. \quad (23)$$

Estimates of $\lambda_{\Theta_1}((0.5,0.5)) = \alpha m \mathbb{E}[L] = 10$, based on samples with size $N = 10$ and $N = 100$, by the "Minkowski content"- based estimator and by the natural estimator for different values

of the bandwidth r are depicted in Figure 6 and Figure 7, respectively. The distances between the theoretical value of $\lambda_{\Theta_1}((0.5, 0.5))$ and its estimate by these two estimators, evaluated in the corresponding optimal bandwidth, are shown in Table 3. We point out that, since in this case the optimal bandwidth associated to the natural estimator is $+\infty$ (see [9]), we have chosen $r = 105$ pixels to avoid edge effects.

We may notice that the "Minkowski content"- based estimator decreases as the function $\frac{1}{2r}$, when $r \rightarrow +\infty$, in accordance with its definition; furthermore it is more sensitive to the choice of the bandwidth; so the importance of having general theoretical formulas for this is evident. Furthermore, even if the natural estimator is more stable, and in this case it has been of easy computation, it might be difficult to apply in general, for instance in the next example where the segments have random orientation.

N	$|\widehat{\lambda}_{\Theta_1}^{\mu,N}(0.5, 0.5) - \lambda_{\Theta_1}(0.5, 0.5)|$	$|\widehat{\lambda}_{\Theta_1}^{\nu,N}(0.5, 0.5) - \lambda_{\Theta_1}(0.5, 0.5)|$
10	1.769231	0.797556
100	0.545000	0.206940

Table 3: Distances between the theoretical mean density at the point $(0.5, 0.5)$ and its estimate, evaluated with $r = r_N^{o,AMSE}$ for the "Minkowski content"- based estimator, and $r = 105$ pixels for the natural estimator.

5.2.2 Segments with random orientation

Let Θ_1 be the Boolean model of segments described in Example 1, with $f(y) \equiv 700(y_1^2 + y_2^2)$ and $L \sim U(0, 0.2)$ (see Figure 8 for a realization of Θ_1 in the observation window $[0, 1]^2$). In this section we provide estimates of its mean density only by means of the "Minkowski content"-based estimator, because the two other estimators would require non-trivial calculations, due to the random orientation of the segments.

In Figure 9, estimates of $\lambda_{\Theta_1}(x)$ at $x = (0.5, 0.5)$ based on samples with size $N = 10$ (a) and $N = 100$ (b) are depicted for different values of the bandwidth. We may notice that the numerical experiments and the theoretical results are in accordance, by observing that there is a good approximation around the theoretical optimal bandwidth $r_N^{o,\text{AMSE}}$.

6 Proofs of the main results

Proof of Lemma 2

In Theorem 3.5 in [29] is proved that, for any set A as in the hypotheses,

$$\lim_{r \downarrow 0} \frac{\int_{A_{\oplus r}} g(y) \mathcal{H}^d(\mathrm{d}y)}{b_{d-n} r^{d-n}} = \int_A g(y) \mathcal{H}^n(\mathrm{d}y) \tag{24}$$

for any non-negative and locally bounded function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\mathcal{H}^n(\mathrm{disc} g) = 0$. Therefore, if f is of the type $f(y_1, y_2) = f_1(y_1) f_2(y_2)$, then (8) trivially follows by (24). Otherwise we proceed

as follows.

Since $r \downarrow 0$, we can assume $r \leq 1$; then f is bounded on $A_{\oplus 1} \times B_{\oplus 1}$, and by the Stone-Weierstrass approximation theorem it follows (e.g., see [12, p. 148]) that, for each $\varepsilon > 0$, there exist $f_1^{(1)}, ..., f_1^{(N(\varepsilon))}$, continuous functions on $A_{\oplus 1}$, and $f_2^{(1)}, ..., f_2^{(N(\varepsilon))}$, continuous functions on $B_{\oplus 1}$, such that

$$\sup_{(y_1, y_2) \in A_{\oplus 1} \times B_{\oplus 1}} |f(y_1, y_2) - \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2)| \leq \varepsilon. \qquad (25)$$

Moreover, by (24), for each $\tilde{\varepsilon} > 0$ there exists $r(\tilde{\varepsilon}) > 0$ such that

$$\left| \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{A_{\oplus r}} \int_{B_{\oplus r}} \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^d(dy_1) \mathcal{H}^d(dy_2) \right.$$

$$\left. - \int_A \int_B \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^n(dy_1) \mathcal{H}^n(dy_2) \right| < \tilde{\varepsilon} \qquad \forall r < r(\tilde{\varepsilon}). \quad (26)$$

Then for any $\varepsilon > 0$, we have that for all $r < \min\{r(\varepsilon), r_N\}$, with r_N such that $N(\varepsilon) = [1/r_N]$,

$$\left| \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{A_{\oplus r}} \int_{B_{\oplus r}} f(y_1, y_2) \mathcal{H}^d(dy_1) \mathcal{H}^d(dy_2) - \int_A \int_B f(y_1, y_2) \mathcal{H}^n(dy_1) \mathcal{H}^n(dy_2) \right| \leq$$

$$\leq \frac{1}{b_{d-n}^2 r^{2(d-n)}} \left| \int_{A_{\oplus r}} \int_{B_{\oplus r}} f(y_1, y_2) \mathcal{H}^d(dy_1) \mathcal{H}^d(dy_2) - \int_{A_{\oplus r}} \int_{B_{\oplus r}} \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^d(dy_1) \mathcal{H}^d(dy_2) \right| +$$

$$+ \left| \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{A_{\oplus r}} \int_{B_{\oplus r}} \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^d(dy_1) \mathcal{H}^d(dy_2) - \int_A \int_B \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^n(dy_1) \mathcal{H}^n(dy_2) \right| +$$

$$+ \left| \int_A \int_B \sum_{i=1}^{N(\varepsilon)} f_1^{(i)}(y_1) f_2^{(i)}(y_2) \mathcal{H}^n(dy_1) \mathcal{H}^n(dy_2) - \int_A \int_B f(y_1, y_2) \mathcal{H}^n(dy_1) \mathcal{H}^n(dy_2) \right| \leq$$

$$\overset{(25),(26)}{\leq} \varepsilon \frac{\mathcal{H}^d(A_{\oplus r})}{b_{d-n} r^{d-n}} \frac{\mathcal{H}^d(B_{\oplus r})}{b_{d-n} r^{d-n}} + \varepsilon + \varepsilon \mathcal{H}^n(A) \mathcal{H}^n(B)$$

$$\leq \varepsilon \left(\frac{2^{2n} 4^{2d} b_d^2 \eta_A(\mathbb{R}^d) \eta_B(\mathbb{R}^d)}{\gamma_A \gamma_B b_{d-n}^2} + 1 + \mathcal{H}^n(A) \mathcal{H}^n(B) \right),$$

where the last inequality follows by recalling (from the proof of Lemma 7 in [1]) that

$$\frac{\mathcal{H}^d(A_{\oplus r})}{b_{d-n} r^{d-n}} \leq \frac{2^n 4^d b_d \eta_A(\mathbb{R}^d)}{\gamma_A b_{d-n}}$$

holds for any $r < 2$ (and the same for B). $\qquad \square$

Proof of Theorem 3

By [30, Remark 4] we know that (R) guarantees that

$$\mathcal{H}^d(Z(s)_{\oplus R}) \leq \begin{cases} \mathcal{H}^n(\Xi(s)) \gamma^{-1} 2^n 4^d b_d R^{d-n} & \text{if } R < 2 \\ \mathcal{H}^n(\Xi(s)) \gamma^{-1} 2^n 4^d b_d R^n & \text{if } R \geq 2 \end{cases}. \qquad (27)$$

21

Let us observe that the Dominated Convergence Theorem implies

$$\lim_{r \downarrow 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{\mathbf{K}^2} \int_{(x-Z(s_1))_{\oplus r}} \int_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathrm{d}y_2 \mathrm{d}y_1 Q_{[2]}(\mathrm{d}(s_1, s_2)) =$$
$$\int_{\mathbf{K}^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2) \mathcal{H}^n(\mathrm{d}y_2) \mathcal{H}^n(\mathrm{d}y_1) Q_{[2]}(\mathrm{d}(s_1, s_2)), \quad (28)$$

being

$$\lim_{r \downarrow 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{(x-Z(s_1))_{\oplus r}} \int_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathrm{d}y_2 \mathrm{d}y_1 =$$
$$\int_{(x-Z(s_1))} \int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2) \mathcal{H}^n(\mathrm{d}y_2) \mathcal{H}^n(\mathrm{d}y_1) \quad \forall (s_1, s_2) \in \mathbf{K} \times \mathbf{K},$$

by Lemma 2, and for any $r < 2$

$$\frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{(x-Z(s_1))_{\oplus r}} \int_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathrm{d}y_2 \mathrm{d}y_1 \leq$$
$$\leq \frac{1}{b_{d-n}^2 r^{2(d-n)}} \sup_{(x-Z(s_1))_{\oplus r}} \sup_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathcal{H}^d((x-Z(s_1))_{\oplus r}) \mathcal{H}^d((x-Z(s_2))_{\oplus r}) \leq$$
$$\overset{(27)}{\leq} \frac{2^{2n} 4^{2d} b_d^2}{\gamma^2 b_{d-n}^2} \mathcal{H}^n(\Xi(s_1)) \mathcal{H}^n(\Xi(s_2)) \sup_{(x-Z(s_1))_{\oplus 2}} \sup_{(x-Z(s_2))_{\oplus 2}} g(y_1, s_1, y_2, s_2) \leq$$
$$\overset{(\overline{A3})}{\leq} \frac{2^{2n} 4^{2d} b_d^2}{\gamma^2 b_{d-n}^2} \mathcal{H}^n(\Xi(s_1)) \mathcal{H}^n(\Xi(s_2)) \xi_{B_2(x), B_2(x)}(s_1, s_2) \quad \forall (s_1, s_2) \in \mathbf{K} \times \mathbf{K},$$

with

$$\int_{\mathbf{K}^2} \frac{2^{2n} 4^{2d} b_d^2}{\gamma^2 b_{d-n}^2} \mathcal{H}^n(\Xi(s_1)) \mathcal{H}^n(\Xi(s_2)) \xi_{B_2(x), B_2(x)}(s_1, s_2) Q_{[2]}(\mathrm{d}(s_1, s_2)) \overset{(\overline{A3})}{<} +\infty.$$

Therefore we may claim that

$$\int_{\mathbf{K}^2} \int_{(x-Z(s_1))_{\oplus r}} \int_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathrm{d}y_2 \mathrm{d}y_1 Q_{[2]}(\mathrm{d}(s_1, s_2)) = O(r^{2(d-n)}) = o(r^{d-n+1}).$$
$$(29)$$

Let Y_r be the random variable counting the number of enlarged grains (i.e. the parallel grains at distance r) which cover the point x:

$$Y_r := \sum_{(y_i, s_i) \in \Psi} \mathbf{1}_{(y_i + Z(s_i))_{\oplus r}}(x), \quad (30)$$

and W_r be the random variable counting the number of pairs of different enlarged grains which cover the point x, i.e.

$$W_r = \begin{cases} 0 & \text{if } Y_r = 0, 1 \\ \frac{Y_r(Y_r - 1)}{2} & \text{if } Y_r \geq 2 \end{cases} \quad .$$

Note that

$$
\begin{aligned}
\mathbb{E}[W_r] &= \sum_{w \in \mathrm{Range}(W_r)} w\mathbb{P}(W_r = w) = \sum_{k=2}^{\infty} \frac{k(k-1)}{2}\mathbb{P}(Y_r = k) = \sum_{k=1}^{\infty} \frac{k(k+1)}{2}\mathbb{P}(W_r = \frac{k(k+1)}{2}) \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{k} n\mathbb{P}(W_r = \frac{k(k+1)}{2}) = \sum_{n=1}^{\infty} n \sum_{k=n+1}^{\infty} \mathbb{P}(W_r = \frac{k(k-1)}{2}) \\
&\geq \sum_{n=1}^{\infty} \mathbb{P}(W_r \geq \frac{n(n+1)}{2}) = \sum_{k \geq 2} \mathbb{P}(W_r \geq \frac{k(k-1)}{2});
\end{aligned}
$$

since $W_r \geq \frac{k(k-1)}{2}$ if and only if $Y_r \geq k$, when $k \geq 2$, we may conclude that

$$
\mathbb{E}[W_r] \geq \sum_{k \geq 2} \mathbb{P}(Y_r \geq k).
$$

Thus the following holds:

$$
\sum_{k \geq 2} \mathbb{P}(Y_r \geq k) \leq \mathbb{E}[W_r] \leq \mathbb{E}\Big[\sum_{\substack{(y_i, s_i), (y_j, s_j) \in \Psi, \\ y_i \neq y_j}} \mathbf{1}_{(y_i + Z(s_i))_{\oplus r} \cap (y_j + Z(s_j))_{\oplus r}}(x) \Big] =
$$

$$
= \int_{\mathbf{K}^2} \int_{(x - Z(s_1))_{\oplus r}} \int_{(x - Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2) \mathrm{d}y_2 \mathrm{d}y_1 Q_{[2]}(\mathrm{d}(s_1, s_2)) \overset{(29)}{=} o(r^{d-n+1}). \quad (31)
$$

We are now ready to consider:

$$
\mathbb{P}(x \in \Theta_{n \oplus r}) = \mathbb{P}(Y_r \geq 1) = \sum_{k \geq 1} \mathbb{P}(Y_r \geq k) - \sum_{k \geq 2} \mathbb{P}(Y_r \geq k) = \mathbb{E}[Y_r] - \sum_{k \geq 2} \mathbb{P}(Y_r \geq k)
$$

$$
= \int_{\mathbf{K}} \int_{(x - Z(s))_{\oplus r}} f(y, s) \mathrm{d}y Q(\mathrm{d}s) + o(r^{d-n+1}). \quad (32)
$$

Let us observe that for any $r_N \leq R$

$$
\int_{Z(s)_{\oplus r_N}} f(x - y, s) \mathrm{d}y = \int_{\mathbb{R}^d \setminus Z(s)} \mathbf{1}_{Z(s)_{\oplus r_N}}(y) f(x - y, s) \mathrm{d}y
$$

$$
\overset{(45)}{=} \sum_{i=0}^{d-1} b_{d-i}(d-i) \int_0^{\infty} \int_{N(Z(s))} t^{d-1-i} \mathbf{1}_{[0, \delta(Z(s), y, u))}(t) \mathbf{1}_{Z(s)_{\oplus r_N}}(y + tu) f(x - (y + tu), s) \mu_i(Z(s); \mathrm{d}(y, u)) \mathrm{d}t
$$

$$
= \sum_{i=0}^{d-1} b_{d-i}(d-i) \int_0^{r_N} \int_{N(Z(s))} t^{d-1-i} f(x - (y + tu), s) \mu_i(Z(s); \mathrm{d}(y, u)) \mathrm{d}t
$$

$$
= \sum_{i=0}^{d-1} b_{d-i}(d-i) r_N^{d-i} \int_0^1 \int_{N(Z(s))} t^{d-1-i} f(x - y - r_N tu, s) \mu_i(Z(s); \mathrm{d}(y, u)) \mathrm{d}t
$$

$$
= \sum_{i=0}^{n} b_{d-i}(d-i) r_N^{d-i} \int_0^1 \int_{N(Z(s))} t^{d-1-i} f(x - y - r_N tu, s) \mu_i(Z(s); \mathrm{d}(y, u)) \mathrm{d}t, \quad (33)
$$

having $Z(s)$ Hausdorff dimension $n < d$.
To simplify the notation let us define

$$
c_{d-i}(r_N) := \int_{\mathbf{K}} \int_0^1 \int_{N(Z(s))} t^{d-1-i} f(x - y - r_N tu, s) \mu_i(Z(s); \mathrm{d}(y, u)) \mathrm{d}t Q(\mathrm{d}s) \quad \forall i = 0, ..., n. \quad (34)
$$

23

By noticing that

$$|b_{d-i}(d-i)r_N^{d-i}c_{d-i}(r_N)| \le b_{d-i}(d-i)r_N^{d-i}\int_{\mathbf{K}}\int_0^1 \sup_{\xi \in x - Z(s) - r_N t B_1(0)} f(\xi,s)|\Phi_i|(Z(s))\mathrm{d}tQ(\mathrm{d}s)$$

$$\le b_{d-i}(d-i)r_N^{d-i}\int_{\mathbf{K}}\widetilde{\xi}_{B_R(x)}(s)|\Phi_i|(\Xi(s))Q(\mathrm{d}s) \overset{(M2)}{<} \infty \quad \forall i=0,...,n,$$

and putting (33) in (32), it follows then

$$\mathbb{P}(x \in \Theta_{n \oplus r_N}) = \sum_{i=0}^n b_{d-i}(d-i)r_N^{d-i}c_{d-i}(r_N) + o(r_N^{d-n+1})$$

$$= (d-n)b_{d-n}r_N^{d-n}c_{d-n}(r_N) + (d-n+1)b_{d-n+1}r_N^{d-n+1}c_{d-n+1}(r_N) + o(r_N^{d-n+1}) \quad (35)$$

Furthermore, the following Taylor expansion for $c_{d-n}(r_N)$ holds (with $\theta \in (0,1)$):

$$c_{d-n}(r_N) = \int_{\mathbf{K}}\int_0^1\int_{N(Z(s))} t^{d-n-1}f(x-y-r_N tu, s)\mu_n(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s) \qquad (36)$$

$$= \int_{\mathbf{K}}\int_0^1\int_{N(Z(s))} t^{d-n-1}\Big[f(x-y,s) - \sum_{|\alpha|=1}\frac{1}{\alpha!}D_x^\alpha f(x-y-\theta r_N tu, s)u^\alpha t r_N\Big]$$

$$\mu_n(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

$$= \int_{\mathbf{K}}\int_0^1 t^{d-n-1}\int_{Z(s)} f(x-y,s)\mathcal{H}^n(\mathrm{d}y)\mathrm{d}tQ(\mathrm{d}s) -$$

$$\int_{\mathbf{K}}\int_0^1 t^{d-n}\int_{N(Z(s))}\sum_{|\alpha|=1}\frac{1}{\alpha!}D_x^\alpha f(x-y-\theta r_N tu, s)u^\alpha r_N\mu_n(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

$$\overset{(2)}{=} \frac{\lambda_{\Theta_n}(x)}{(d-n)} - \int_{\mathbf{K}}\int_0^1 t^{d-n}\int_{N(Z(s))}\sum_{|\alpha|=1}\frac{1}{\alpha!}D_x^\alpha f(x-y-\theta r_N tu, s)u^\alpha r_N\mu_n(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

where we used (46), $\Phi_n(Z(s);\cdot) = \mathcal{H}^n_{|Z(s)}(\cdot)$ (see [17]), and

$$\int_{N(Z(s))} f(x-y,s)\mu_i(Z(s);\mathrm{d}(y,u)) = \int_{\partial Z(s)} f(x-y,s)\Phi_i(Z(s);\mathrm{d}y) \quad \forall i=0,...,n.$$

Thus the bias of $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ is given by

$$Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \frac{\mathbb{P}(x \in \Theta_{n \oplus r_N})}{b_{d-n}r_N^{d-n}} - \lambda_{\Theta_n}(x)$$

$$\overset{(35)}{=} -(d-n)r_N\int_{\mathbf{K}}\int_0^1 t^{d-n}\int_{N(Z(s))}\sum_{|\alpha|=1}\frac{1}{\alpha!}D_x^\alpha f(x-y-\theta r_N tu, s)u^\alpha\mu_n(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

$$+(d-n+1)\frac{b_{d-n+1}}{b_{d-n}}r_N\int_{\mathbf{K}}\int_0^1\int_{N(Z(s))} t^{d-n}f(x-y-r_N tu, s)\mu_{n-1}(Z(s);\mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)+o(r_N).$$

Thanks to the assumption (M2), it easy to check that

$$|t^{d-n}f(x-y-r_N tu, s)| \le \widetilde{\xi}_{B_R(x)}^{(0)}(s)$$

24

and

$$\Big| \sum_{|\alpha|=1} \frac{1}{\alpha!} D_x^\alpha f(x - y - \theta r_N tu, s) u^\alpha t^{d-n} \Big| \le \sum_{|\alpha|=1} \frac{1}{\alpha!} \widetilde{\xi}_{B_R(x)}^{(\alpha)}(s)$$

with $t \in (0,1), (y,u) \in N(Z(s))$, besides we have

$$\int_{\mathbf{K}} \int_0^1 \int_{N(Z(s))} \widetilde{\xi}_{B_R(x)}^{(0)}(s) |\mu_{n-1}|(Z(s); \mathrm{d}(y,u)) \mathrm{d}t Q(\mathrm{d}s) = \int_{\mathbf{K}} \widetilde{\xi}_{B_R(x)}^{(0)}(s) |\Phi_{n-1}|(\Xi(s)) Q(\mathrm{d}s) < \infty,$$

and

$$\int_{\mathbf{K}} \int_0^1 \int_{N(Z(s))} \sum_{|\alpha|=1} \frac{1}{\alpha!} \widetilde{\xi}_{B_R(x)}^{(\alpha)}(s) |\mu_n|(Z(s); \mathrm{d}(y,u)) \mathrm{d}t Q(\mathrm{d}s) = \sum_{|\alpha|=1} \frac{1}{\alpha!} \int_{\mathbf{K}} \widetilde{\xi}_{B_R(x)}^{(\alpha)}(s) |\Phi_n|(\Xi(s)) Q(\mathrm{d}s) < \infty,$$

respectively. By remembering that $f(\cdot, s) \in C^1$, the Dominated Convergence Theorem (for signed measures) allows us to claim that

$$\lim_{N \to \infty} \frac{Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))}{r_N} = \frac{b_{d-n+1}}{b_{d-n}} \int_{\mathbf{K}} \int_{Z(s)} f(x - y, s) \Phi_{n-1}(Z(s); \mathrm{d}y) Q(\mathrm{d}s)$$

$$- \frac{d-n}{d-n+1} \sum_{|\alpha|=1} \int_{\mathbf{K}} \int_{N(Z(s))} D_x^\alpha f(x - y, s) u^\alpha \mu_n(Z(s); \mathrm{d}(y,u)) Q(\mathrm{d}s) := \mathcal{A}_1(x) - \mathcal{A}_2(x), \quad (37)$$

and so the assertion. $\qquad\qquad\square$

Proof of Theorem 4

Let Y_r be as in (30). First, we want to obtain a kind of *second order expansion* of the probability $\mathbb{P}(x \in \Theta_{n \oplus r})$. It is worth noticing that

$$\mathbb{E}(Y_r^2) = \sum_{k \ge 1} \mathbb{P}(Y_r^2 \ge k) = \mathbb{P}(Y_r \ge 1) + \sum_{k \ge 2} \mathbb{P}(Y_r^2 \ge k) = \mathbb{P}(Y_r \ge 1) + \sum_{k \ge 2} \mathbb{P}(Y_r \ge \sqrt{k})$$

$$= \mathbb{P}(Y_r \ge 1) + \sum_{k \ge 2}(k^2 - (k-1)^2)\mathbb{P}(Y_r \ge k) = \mathbb{P}(Y_r \ge 1) + \sum_{k \ge 2}(2k-1)\mathbb{P}(Y_r \ge k),$$

which yields

$$\mathbb{P}(x \in \Theta_{n \oplus r}) = \mathbb{P}(Y_r \ge 1) = \mathbb{E}(Y_r^2) - \sum_{k \ge 2}(2k-1)\mathbb{P}(Y_r \ge k). \qquad (38)$$

Solving the previous equation with respect to $\mathbb{P}(Y_r \ge 2)$, we get

$$\mathbb{P}(Y_r \ge 2) = \frac{1}{3}\Big(\mathbb{E}Y_r^2 - \sum_{k \ge 3}(2k-1)\mathbb{P}(Y_r \ge k) - \mathbb{P}(x \in \Theta_{n \oplus r})\Big). \qquad (39)$$

Substituting the above expression for $\mathbb{P}(Y_r \ge 2)$ in (32), and then solving with respect to $\mathbb{P}(x \in \Theta_{n \oplus r})$, we finally obtain

$$\mathbb{P}(x \in \Theta_{n \oplus r}) = \frac{3}{2}\mathbb{E}Y_r - \frac{1}{2}\mathbb{E}Y_r^2 + \sum_{k \ge 3}(k-2)\mathbb{P}(Y_r \ge k). \qquad (40)$$

25

Let us introduce the random variable T_r, which counts the number of triples of different enlarged grains which cover the point x, in order to estimate $\sum_{k \geq 3}(k-2)\mathbb{P}(Y_r \geq k)$:

$$T_r := \#\{(i,j,k);:\ i<j<k \text{ and } x \in (y_i+Z(s_i))_{\oplus r} \cap (y_j+Z(s_j))_{\oplus r} \cap (y_k+Z(s_k))_{\oplus r}\} = \begin{cases} 0 & \text{if } Y_r = 0,1,2 \\ \binom{Y_r}{3} & \text{if } Y_r \geq 3 \end{cases}$$

By observing that T_r can assume only values of the type $\frac{k(k-1)(k-2)}{6}$, with $k \in \mathbb{N}$, we have that

$$
\begin{aligned}
\mathbb{E}[T_r] &= \sum_{k=3}^{\infty} \frac{k(k-1)(k-2)}{6} \mathbb{P}(T_r = \frac{k(k-1)(k-2)}{6}) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} \frac{n(n+1)}{2} \mathbb{P}(T_r = \frac{k(k+1)(k+2)}{6}) \\
&= \sum_{n=1}^{\infty} \frac{n(n+1)}{2} \sum_{k=n+2}^{\infty} \mathbb{P}(T_r = \frac{k(k-1)(k-2)}{6}) = \sum_{n=1}^{\infty} \frac{n(n+1)}{2} \mathbb{P}(T_r \geq \frac{n(n+1)(n+2)}{6}) \\
&= \sum_{k=3}^{\infty} \frac{(k-2)(k-1)}{2} \mathbb{P}(T_r \geq \frac{k(k-1)(k-2)}{6}) \geq \sum_{k \geq 3}(k-2)\mathbb{P}(Y_r \geq k). \qquad (41)
\end{aligned}
$$

being $T_r \geq \frac{k(k-1)(k-2)}{6}$ if and only if $Y_r \geq k$ when $k \geq 3$.

Note that

$$\lim_{r \downarrow 0} \frac{\mathbb{E}[T_r]}{4r^2} \leq \lim_{r \downarrow 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \mathbb{E}\Big[\sum_{\substack{(y_i,s_i),(y_j,s_j),(y_k,s_k)\,\in\,\Psi, \\ y_i \neq y_j, y_i \neq y_k, y_k \neq y_j}} \mathbf{1}_{(y_i+Z(s_i))_{\oplus r} \cap (y_j+Z(s_j))_{\oplus r} \cap (y_k+Z(s_k))_{\oplus r}}(x)\Big]$$

$$= \lim_{r \downarrow 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{\mathbb{R}^d \times \mathbf{K}^3} \mathbf{1}_{(x-Z(s_1))_{\oplus r}}(y_1) \Big[\int_{(x-Z(s_2))_{\oplus r} \times (x-Z(s_3))_{\oplus r}} h(y_1,s_1,y_2,s_2,y_3,s_3)\mathrm{d}y_2\mathrm{d}y_3 \Big]$$
$$\mathrm{d}y_1 Q_{[3]}(\mathrm{d}(s_1,s_2,s_3)) = 0, \quad (42)$$

where the last equation follows as an application of the Dominated Convergence Theorem, after having observed that

- by Lemma 2, for all $(s_1,s_2,s_3) \in \mathbf{K}^3$ and $y_1 \in \mathbb{R}^d$,

$$\lim_{r \to 0} \frac{1}{b_{d-n}^2 r^{2(d-n)}} \int_{(x-Z(s_2))_{\oplus r} \times (x-Z(s_3))_{\oplus r}} h(y_1,s_1,y_2,s_2,y_3,s_3)\mathrm{d}y_2\mathrm{d}y_3 =$$
$$\int_{(x-Z(s_2)) \times (x-Z(s_3))} h(y_1,s_1,y_2,s_2,y_3,s_3)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_3);$$

- being the dimension of $Z(s)$ less than d, $\forall s_1 \in \mathbf{K}$,

$$\lim_{r \to 0} \mathbf{1}_{(x-Z(s_1))_{\oplus r}}(y_1) = 0 \quad \mathcal{H}^d\text{--a.e. } y_1 \in \mathbb{R}^d;$$

- for any $r \leq 1$

$$\frac{1}{b_{d-n}^2 r^{2(d-n)}} \mathbf{1}_{(x-Z(s_1))_{\oplus r}}(y_1) \int_{(x-Z(s_2))_{\oplus r} \times (x-Z(s_3))_{\oplus r}} h(y_1,s_1,y_2,s_2,y_3,s_3)\mathrm{d}y_2\mathrm{d}y_3$$

$$\leq \frac{\mathbf{1}_{(x-Z(s_1))_{\oplus r}}(y_1)}{b_{d-n}^2 r^{2(d-n)}} \sup_{y_2 \in (x-Z(s_2))_{\oplus r}} \sup_{y_3 \in (x-Z(s_3))_{\oplus r}} h(y_1,s_1,y_2,s_2,y_3,s_3)\mathcal{H}^d((x-Z(s_2))_{\oplus r})\mathcal{H}^d((x-Z(s_3))_{\oplus r})$$

$$\overset{(27),(M4)}{\leq} \frac{4^{2d+n}b_d^2}{b_{d-n}^2\gamma^2}\mathcal{H}^n(\Xi(s_2))\mathcal{H}^n(\Xi(s_3))\overline{\xi}_{x,B_1(x),B_1(x)}(s_1,y_1,s_2,s_3),$$

26

whose integral on $\mathbb{R}^d \times \mathbf{K}$ with respect to $\mathcal{H}^d \otimes Q_{[3]}$ is finite by (11).

Thus, by (41) and (42), $\mathbb{P}(x \in \Theta_{n \oplus r}) \overset{(40)}{=} \frac{3}{2}\mathbb{E}Y_r - \frac{1}{2}\mathbb{E}Y_r^2 + o(r^2)$. By observing that

$$\mathbb{E}[Y_r^2] = \mathbb{E}\Big[\Big(\sum_{(y_i, s_i) \in \Psi} \mathbf{1}_{(y_i + Z(s_i))_{\oplus r}}(x)\Big)^2\Big] = \mathbb{E}Y_r + \mathbb{E}\Big[\sum_{\substack{(y_i, s_i),(y_j, s_j) \,\in\, \Psi, \\ y_i \neq y_j}} \mathbf{1}_{(y_i + Z(s_i))_{\oplus r} \cap (y_j + Z(s_j))_{\oplus r}}(x)\Big],$$

it follows that

$$\mathbb{P}(x \in \Theta_{n \oplus r}) = \int_{\mathbf{K}}\int_{(x-Z(s))_{\oplus r}} f(y,s)\mathrm{d}yQ(\mathrm{d}s) - \frac{1}{2}\int_{\mathbf{K}^2}\int_{(x-Z(s_1))_{\oplus r}}\int_{(x-Z(s_2))_{\oplus r}} g(y_1, s_1, y_2, s_2)\mathrm{d}y_2\mathrm{d}y_1 Q_{[2]}(\mathrm{d}(s_1, s_2)) +$$

$$\overset{(33),.(28)}{=} 2rc_1(r) + 2\pi\, r^2 c_2(r) - 2r^2 \int_{\mathbf{K}^2}\int_{(x-Z(s_1))}\int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1, s_2)) + o(r^2),$$

$$(43)$$

with c_i defined as in (34).

Therefore we have that

$$Bias(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x)) \overset{(5)}{=} c_1(r_N) + \pi r_N c_2(r_N) - r_N \int_{\mathbf{K}^2}\int_{(x-Z(s_1))}\int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1, s_2))$$

$$+ o(r_N) - \lambda_{\Theta_{d-1}}(x)$$

$$\overset{(36),.(34)}{=} -r_N\int_{\mathbf{K}}\int_0^1\int_{N(Z(s))} \sum_{|\alpha|=1} \frac{1}{\alpha!}D_x^\alpha f(x - y - \theta r_N tu, s)u^\alpha t \mu_{d-1}(Z(s); \mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

$$+ \pi r_N\int_{\mathbf{K}}\int_0^1\int_{N(Z(s))} tf(x - y - r_N tu, s)\mu_{d-2}(Z(s); \mathrm{d}(y,u))\mathrm{d}tQ(\mathrm{d}s)$$

$$- r_N\int_{\mathbf{K}^2}\int_{(x-Z(s_1))}\int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1, s_2)) + o(r_N),$$

where $\theta \in (0, 1)$.

By the same argument used in the proof of Theorem 3 to obtain (37), we get

$$Bias(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x)) =$$

$$\Big(\frac{\pi}{2}\int_{\mathbf{K}}\int_{Z(s)} f(x-y, s)\Phi_{d-2}(Z(s); \mathrm{d}y)Q(\mathrm{d}s) - \frac{1}{2}\sum_{|\alpha|=1}\int_{\mathbf{K}}\int_{N(Z(s))} D_x^\alpha f(x-y, s)u^\alpha \mu_{d-1}(Z(s); \mathrm{d}(y,u))Q(\mathrm{d}s)$$

$$- \int_{\mathbf{K}^2}\int_{(x-Z(s_1))}\int_{(x-Z(s_2))} g(y_1, s_1, y_2, s_2)\mathcal{H}^n(\mathrm{d}y_2)\mathcal{H}^n(\mathrm{d}y_1)Q_{[2]}(\mathrm{d}(s_1, s_2))\Big)r_N + o(r_N) \quad (44)$$

Let us now observe that $\mathcal{H}^{d-1}(Z(s)) = \mathcal{H}^{d-1}(\partial^2 Z(s))$, being $Z(s)$ a compact set with positive reach with dimension $d-1$ (e.g., see [2, Sect. 4]); therefore $N(Z(s), x) = \{\nu((Z(s), x), -\nu((Z(s), x)\}$ for $\mathcal{H}^{d-1}$-a.e. $x \in Z(s)$ (see the Appendix). Let us denote by $\alpha^{(j)}$ the multi-index with $\alpha_i = \delta_{ij}$; then $D_x^{\alpha^{(j)}} f(x-y, s) = \partial f(x-y, s)/\partial x_j$, and $u^{\alpha^{(j)}} = u_j$. It follows

$$\sum_{|\alpha|=1}\int_{N(Z(s))} D_x^\alpha f(x-y, s)u^\alpha \mu_{d-1}(Z(s); \mathrm{d}(y,u)) = \sum_{j=1}^d\int_{N(Z(s))} \frac{\partial f(x-y, s)}{\partial x_j}u_j\mu_{d-1}(Z(s); \mathrm{d}(y,u)),$$

but, by (49), we have that

$$\int_{N(Z(s))} \frac{\partial f(x-y,s)}{\partial x_j} u_j \mu_{d-1}(Z(s); \mathrm{d}(y,u)) = \frac{1}{2} \int_{\partial^2 Z(s)} \sum_{u \in N(Z(s),y)} \frac{\partial f(x-y,s)}{\partial x_j} u_j \mathcal{H}^{d-1}(\mathrm{d}y)$$

$$= \frac{1}{2} \int_{\partial^2 Z(s)} \frac{\partial f(x-y,s)}{\partial x_j} \big(\nu((Z(s),y)_j - \nu((Z(s),y)_j \big) = 0 \qquad \forall j = 1, \ldots, d.$$

This implies

$$\sum_{|\alpha|=1} \int_{\mathbf{K}} \int_{N(Z(s))} D_x^\alpha f(x-y,s) u^\alpha \mu_{d-1}(Z(s); \mathrm{d}(y,u)) Q(\mathrm{d}s) = 0,$$

and so the assertion:

$$Bias(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x)) = \big(\frac{\pi}{2} \int_{\mathbf{K}} \int_{Z(s)} f(x-y,s) \Phi_{d-2}(Z(s); \mathrm{d}y) Q(\mathrm{d}s)$$

$$- \int_{\mathbf{K}^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))} g(y_1,s_1,y_2,s_2) \mathcal{H}^n(\mathrm{d}y_2) \mathcal{H}^n(\mathrm{d}y_1) Q_{[2]}(\mathrm{d}(s_1,s_2))) r_N + o(r_N).$$

$\square$

Proof of Corollary 6

a) It is easy to check that the hypotheses of Theorem 3 are satisfied, and that (32) simplifies here:

$$\mathbb{P}(x \in \Theta_{n \oplus r}) = c\mathbb{E}_Q[\mathcal{H}^d(Z_{\oplus r})] + o(r^{d-n+1}) \overset{(47)}{=} cr^{d-n} b_{d-n} \mathbb{E}_Q[\mathcal{H}^n(Z)] + cr^{d-n+1} b_{d-n+1} \mathbb{E}_Q[\Phi_{n-1}(Z)] + o(r^{d-n+1})$$

By remembering that in this case $\lambda_{\Theta_n}(x) \equiv c\mathbb{E}_Q[\mathcal{H}^n(Z)]$ for any $x \in \mathbb{R}^d$, it follows then

$$Bias(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \frac{cb_{d-n+1} \mathbb{E}_Q[\Phi_{n-1}(Z)]}{b_{d-n}} r_N + o(r_N) \quad \forall x \in \mathbb{R}^d,$$

and so the asymptotic approximation of the mean square error reduces to

$$AMSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \Big(\frac{cb_{d-n+1}}{b_{d-n}} \mathbb{E}_Q[\Phi_{n-1}(Z)] \Big)^2 r_N^2 + \frac{c\mathbb{E}_Q[\mathcal{H}^n(Z)]}{N b_{d-n} r_N^{d-n}}.$$

Finally, by $r_N^{o,AMSE}(x) := \arg\min_{r_N} AMSE(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$ we get (14).

b) It is easy to check that the hypotheses of Theorem 4 are satisfied, and that (43) simplifies here:

$$\mathbb{P}(x \in \Theta_{n \oplus r}) = 2cr\mathbb{E}_Q[\mathcal{H}^{d-1}(Z)] + cr^2 \pi \mathbb{E}_Q[\Phi_{d-2}(Z)]$$

$$- 2r^2 \int_{\mathbf{K}^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))} g(y_1,s_1,y_2,s_2) \mathcal{H}^n(\mathrm{d}y_2) \mathcal{H}^n(\mathrm{d}y_1) Q_{[2]}(\mathrm{d}(s_1,s_2)) + o(r^2).$$

Then it follows

$$Bias(\widehat{\lambda}_{\Theta_{d-1}}^{\mu,N}(x)) = \frac{\pi}{2} cr_N \mathbb{E}_Q[\Phi_{d-2}(Z)]$$

$$- r_N \int_{\mathbf{K}^2} \int_{(x-Z(s_1))} \int_{(x-Z(s_2))} g(y_1,s_1,y_2,s_2) \mathcal{H}^n(\mathrm{d}y_2) \mathcal{H}^n(\mathrm{d}y_1) Q_{[2]}(\mathrm{d}(s_1,s_2)) + o(r_N^2);$$

28

Being $Var(\widehat{\lambda}^{\mu,N}_{\Theta_{d-1}}(x)) = c\mathbb{E}_Q[\mathcal{H}^{d-1}(Z)]/(2Nr_N) + o(1/(Nr_N))$, the assertion directly follows by $r_N^{o,AMSE}(x) := \underset{r_N}{\arg\min} AMSE(\widehat{\lambda}^{\mu,N}_{\Theta_{d-1}}(x))$. $\qquad\square$

APPENDIX

Positive reach, curvature measures and related results

We summarize here some basic definitions and results on sets with positive reach and associated curvature measures, which might be useful for the non-expert reader for a more readability of the paper. We refer to the existent literature for a more exhaustive treatment of this subejct.

Let $A \subset \mathbb{R}^d$ be a non-empty closed set and $z \in \mathbb{R}^d$, set $\mathrm{dist}(A,z) := \inf\{||a-z|| : a \in A\}$. Denote by $\mathrm{Unp}(A) := \{x \in \mathbb{R}^d : \exists! \, a \in A \text{ such that } \mathrm{dist}(x,A) = ||a-x||\}$ the set of points having a unique projection on A. The definition of $\mathrm{Unp}(A)$ implies the existence of a projection mapping $\xi_A : \mathrm{Unp}(A) \to A$ which assigns to $x \in \mathrm{Unp(A)}$ the unique point $\xi_A(x) \in A$ such that $\mathrm{dist}(z,A) = ||\xi_A(x) - x||$. Then for all $x \in \mathrm{Unp}(A)$ with $\mathrm{dist}(x,A) > 0$ we may define $u_A(x) := (x - \xi_A(x))/\mathrm{dist}(x,A)$. The set of all $x \in \mathbb{R}^d \setminus A$ for which $\xi_A(x)$ is not defined is called the exoskeleton of A and it is denoted by $\mathrm{exo}(A)$. The exoskeleton is a measurable subset of $\mathbb{R}^d$ and $\mathcal{H}^d(\mathrm{exo}(A)) = 0$. Denoted by $\mathbf{S}^{d-1} := \partial B_1(0)$ the unit sphere in $\mathbb{R}^d$, the *normal bundle* of A is the subset of $\partial A \times \mathbf{S}^{d-1}$:

$$N(A) := \{(\xi_A(x), u_A(x)) : x \notin A \cup \mathrm{exo}(A)\}$$

The *reach function* $\delta(A,\cdot) : \mathbb{R}^d \times \mathbf{S}^{d-1} \to [0,\infty]$ is defined by $\delta(A,x,u) := \inf\{t \geq 0 : x+tu \in \mathrm{exo}(A)\}$ if $(x,u) \in N(A)$, $\delta(A,x,u) := 0$ otherwise; then the *reach* of A is defined by

$$\mathrm{reach}(A) := \inf\{\delta(A,x,u) : (x,u) \in N(A)\}.$$

If $\mathrm{reach}(A) > 0$ the set is said to be a *set with positive reach*.
For any non-empty closed set $A \subset \mathbb{R}^d$ there exist uniquely determined signed measures $\mu_0(A;\,\cdot\,), \ldots, \mu_{d-1}(A;\,\cdot\,)$ on $N(A)$, said *support measures of A*, which arise as coefficient measures of a local Steiner formula; namely, in [19, Theorem 2.1] it is proved that

$$\int_{\mathbb{R}^d \setminus A} f(x)\mathcal{H}^d(\mathrm{d}x) = \sum_{i=0}^{d-1} b_{d-i}(d-i) \int_0^\infty \int_{N(A)} t^{d-1-i}$$
$$\mathbf{1}_{[0,\delta(A,x,u))}(t)f(x+tu)\mu_i(A;\mathrm{d}(x,u))\mathrm{d}t, \quad (45)$$

for any measurable bounded function $f : \mathbb{R}^d \to \mathbb{R}$ with compact support.
Furthermore, if $\mathcal{H}^k(\partial A) = 0$ for some $k \in \{1, ..., d-1\}$, $\mu_k(A;\cdot) \equiv 0$ (see [21, Proposition 2.4]).

We also remind that if the closed set A has positive reach, the following relationship between the support measure $\mu_i(A; \cdot)$ and the *curvature measure $\Phi_i(A; \cdot)$ associated with A* introduced in [17] holds:

$$\mu_i(A; \cdot \times \mathbf{S}^{d-1}) = \Phi_i(A; \cdot) \quad \forall i = 0, ..., d-1; \tag{46}$$

by using Federer's notation, $\Phi_i(A) := \Phi_i(A; A)$ is the *total curvature measure of A*. Morover, if reach$(A) > 0$, the following global Steiner formula holds

$$\mathcal{H}^d(A_{\oplus r}) = \sum_{i=0}^{d} r^{d-i} b_{d-i} \Phi_i(A), \quad \forall r < \text{reach}(A); \tag{47}$$

note that $\Phi_i(A) = 0$, for any $i > n$, if $\dim(A) = n$.

Let A be a closed subset of $\mathbb{R}^d$, define

$$\partial^+ A = \{x \in \partial A \ : \ (x, u) \in N(A) \text{ for some } u \in \mathbf{S}^{d-1}\},$$

it is also well-known that $\partial A = \partial^+ A$ if reach$(A) > 0$. For each $x \in \partial^+ A$, we can define

$$N(A, x) := \{u \in \mathbf{S}^{d-1} \ : \ (x, u) \in N(A)\};$$

the *normal cone* of A at x is defined by $n(A, x) := \{\lambda u \ : \ \lambda \geq 0 \, , u \in N(A, x)\}$.
Let

$$\partial^{++} A := \{x \in \partial^+ A \ : \ \dim n(A, x) = 1\}, \tag{48}$$

where $\dim B$ denotes the dimension of the affine hull of $B \subset \mathbb{R}^d$; then it follows that $\partial^{++} A$ is the disjoint union of the sets $\partial^1 A$ and $\partial^2 A$, defined by

$$\partial^i A := \{x \in \partial^{++} A \ : \ \text{card} N(A, x) = i\} \quad i = 1, 2.$$

For $x \in \partial^1 A$, the unique element of $N(A, x)$, say $\nu(A, x)$, is the outer normal to A at x; for $x \in \partial^2 A$ there exist two outer normal ($\nu(A, x)$ and $-\nu(A, x)$ by (48)) to A at x. In particular we recall the following representation for $\mu_{d-1}(A; \cdot)$ (see [19, Proposition 4.1]):

$$\mu_{d-1}(A; \cdot) = \frac{1}{2} \int_{\partial^1 A} \mathbf{1}_{\{(x, \nu(A,x)) \in \cdot\}} \mathcal{H}^{d-1}(\mathrm{d}x) +$$

$$\frac{1}{2} \int_{\partial^2 A} \mathbf{1}_{\{(x, \nu(A,x)) \in \cdot\}} + \mathbf{1}_{\{(x, -\nu(A,x)) \in \cdot\}} \mathcal{H}^{d-1}(\mathrm{d}x) \tag{49}$$

Finally, we also recall that explicit description of the curvature measures of order zero are given in [19, Proposition 4.10]; in particular for any convex set A it holds (see [25, Secton 4.4]) $\forall B \in \mathcal{B}_{\mathbb{R}^d}$

$$\Phi_0(A, B) = \frac{1}{d b_d} \int_{\mathbf{S}^{d-1}} \sum_{x \in B} \mathbf{1}_{N(A)}(x, u) \mathcal{H}^{d-1}(\mathrm{d}u). \tag{50}$$

References

[1] AMBROSIO L., CAPASSO V., VILLA E. (2009). On the approximation of mean densities of random closed sets, *Bernoulli*, **15,** 1222–1242.

[2] AMBROSIO L., COLESANTI A., VILLA E. (2008). Outer Minkowski content for some classes of closed sets and applications to stochastic geometry, *Math. Ann.* **342,** 727–748.

[3] AMBROSIO L., FUSCO N., PALLARA D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*, Clarendon Press, Oxford.

[4] BADDELEY A., BARANY I., SCHNEIDER R., WEIL W. (2007). *Stochastic Geometry*, Lecture Notes in Mathematics **1982**, Springer, Berlin.

[5] BADDELEY A., MOLCHANOV I.S. (1997). On the expected measure of a random set. In: *Proceedings of the International Symposium on Advances in Theory and Applications of Random Sets (Fontainebleau, 1996)*, River Edge, NJ, World Sci. Publishing, 3–20.

[6] BELLETTINI G. (2013). *Lecture notes on mean curvature flow, barriers and singular perturbations*, Scuola Normale Superiore di Pisa (Nuova Serie), **12,** Edizioni della Normale, Pisa.

[7] BENEŠ V. , RATAJ J. (2004). *Stochastic geometry: selected topics*, Kluwer, Dordrecht.

[8] BOSQ D., LECOUTRE J.P (1987). *Theorie De L'estimation Fonctionnelle*, Economica, Paris.

[9] CAMERLENGHI F., CAPASSO V., VILLA E. (2014) On the estimation of the mean density of random closed sets. *J. Multivariate Anal.*, **125,** 65–88.

[10] CAMERLENGHI F., CAPASSO V., VILLA E. (2014). Numerical experiments for the estimation of mean densities of random sets. In: Proceedings of the 11th European Congress of Stereology and Image Analysis. *Image Anal. Stereol.*, **33,** 83–94

[11] CHIU S.N., STOYAN D., KENDALL W.S., MECKE J. (2013). *Stochastic Geometry and its Applications*, 3rd edition, John Wiley & Sons, Chichcester.

[12] CONWAY J.B. (1990). *A course in functional analysis.* Second edition. Graduate Texts in Mathematics **96**. Springer-Verlag, New York.

[13] CRESSIE N. A. C. (1993). *Statistics for Spatial Data.* Revised edition. John Wiley and Sons, New York.

[14] DALEY D.J., VERE-JONES D. (1988). *An Introduction to the Theory of Point Processes*, Springer Series in Statistics, New York.

[15] DEVROYE L., GYÖRFI L. (1985). *Nonparametric density estimation: the L_1 view*, Wiley, New York.

[16] DIGGLE P.J. (1985), A kernel method for smoothing point process data, *Appl. Statist.*, **34,** 138–147.

[17] FEDERER H. (1959), Curvature measures, *Trans. Amer. Math. Soc.*, **93,** 418–491.

[18] HÄRDLE W. (1991). *Smoothing Techniques with Implementation in S*, Springer-Verlag, New York.

[19] HUG D., LAST G., WEIL W. (2004). A local Steiner-type formula for general closed sets and applications. *Math. Z.*, **246,** 237–272.

[20] KARR A.F. (1986). *Point Processes and Their Statistical Inference*, Marcel Dekker, New York.

[21] LAST G. (2006). On mean curvature functions of Brownian paths, *Stochastic Process. Appl.*, **116,** 1876–1891.

[22] MATHERON G. (1975). *Random Sets and Integral Geometry*, John Wiley & Sons, New York.

[23] PARZEN E. (1962). On the estimation of a probability density function and the mode, *Ann. Math. Statist.*, **33,** 1065–1076.

[24] ROSENBLATT M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27,** 832–837.

[25] SCHNEIDER R. (1993). *Convex Bodies: the Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications 44, Cambridge University Press, Cambridge.

[26] SILVERMAN B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

[27] SIMONOFF J.S. (1996). *Smoothing Methods in Statistics.* Springer, New York.

[28] VAN LIESHOUT M.N.M. (2012). On estimation of the intensity function of a point process, *Methodology and Computing in Applied Probability*, **14,** 567-578.

[29] VILLA E. (2010). Mean densities and spherical contact distribution function of inhomogeneous Boolean models, *Stoch. An. Appl.*, **28,** 480–504.

[30] VILLA E. (2014).On the local approximation of mean densities of random closed sets, *Bernoulli,* **20,** 1–27

[31] ZÄHLE M. (1982). Random processes of Hausdorff rectifiable closed sets, *Math. Nachr.*, **108,** 49–72.
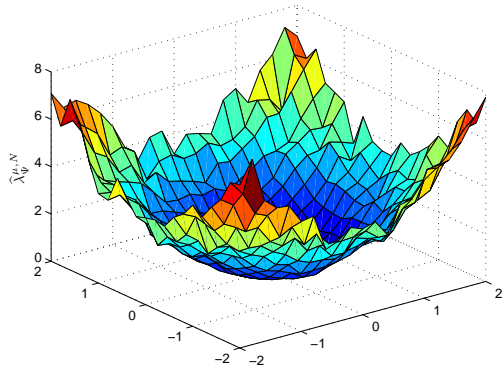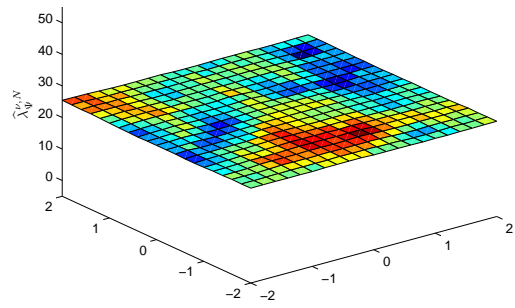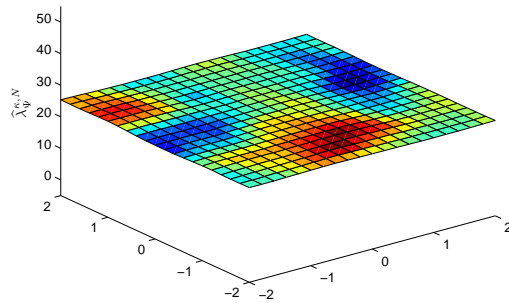
(a)
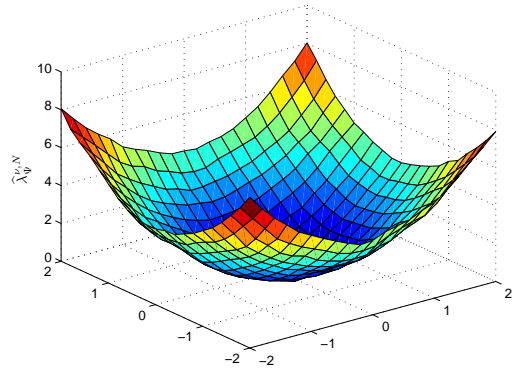
(b)
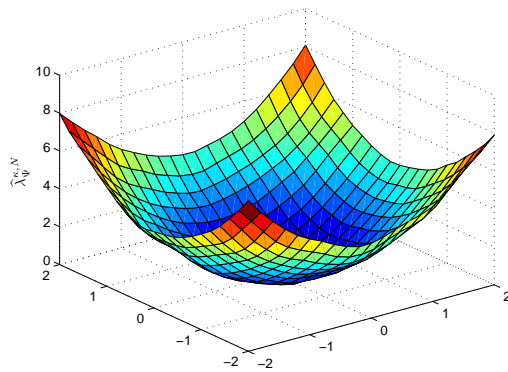
(c)

Figure 1: Estimation of the intensity of an inhomogeneous Poisson process based on a sample with size $N = 1000$, by the "Minkowski content"-based estimator (a), by the natural estimator (b), and by the kernel (Epanechnikov) estimator (c).
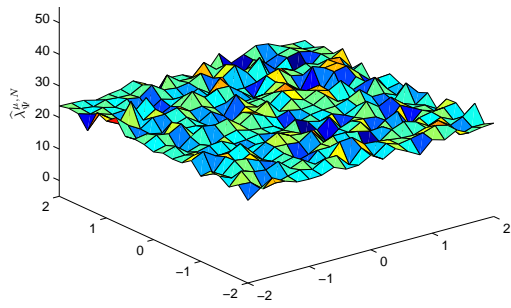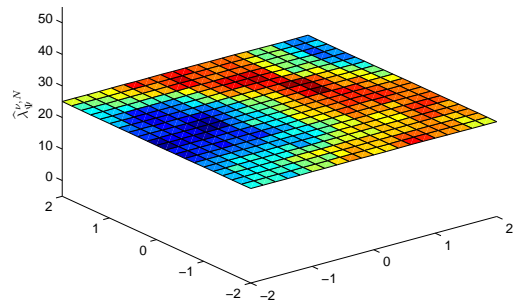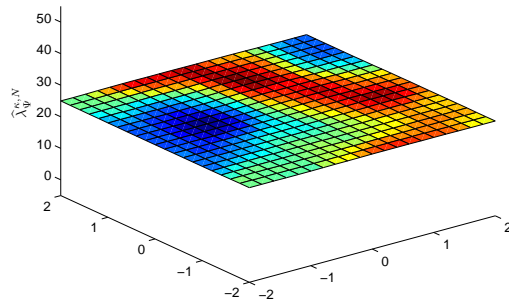
33
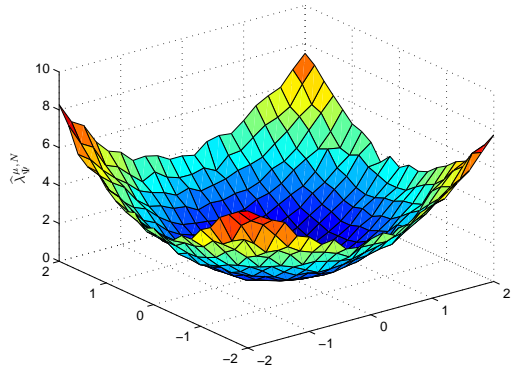
(a)
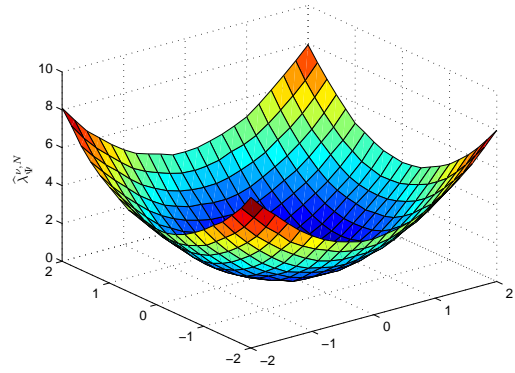
(b)
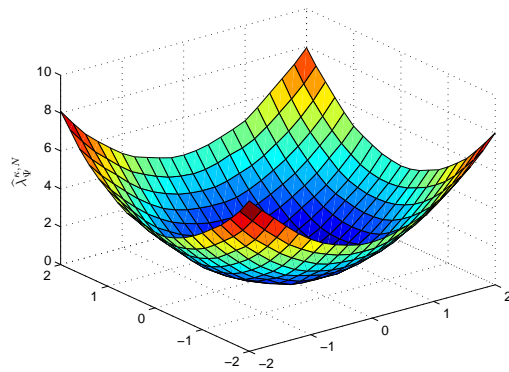
(c)

Figure 2: Estimation of the intensity of an inhomogeneous Poisson process based on a sample with size $N = 10000$, by the "Minkowski content"-based estimator (a), by the natural estimator (b), and by the kernel (Epanechnikov) estimator (c).
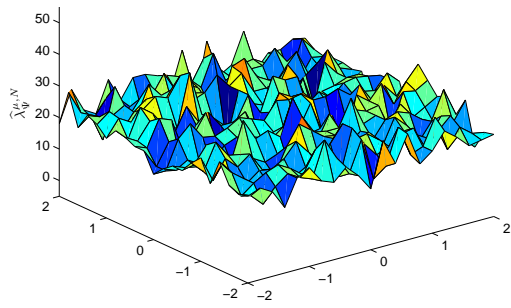
(a)

(b)

(c)

Figure 3: Estimation of the intensity of a Matèrn cluster point process based on a sample with size $N = 1000$, by the "Minkowski content"-based estimator (a), by the natural estimator (b), and by the kernel (Epanechnikov) estimator (c).

(a)

(b)

(c)

Figure 4: Estimation of the intensity of a Matèrn cluster point process based on a sample with size $N = 10000$, by the "Minkowski content"-based estimator (a), by the natural estimator (b), and by the kernel (Epanechnikov) estimator (c).
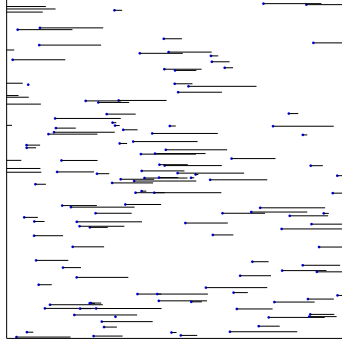
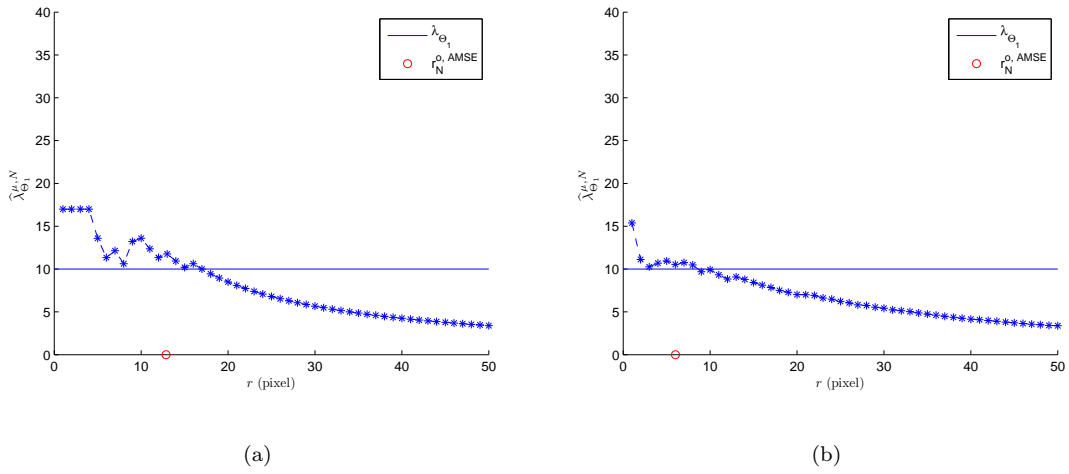Figure 5: A realization of Θ_1 in the observation window $W = [0,1]^2$

(a)

(b)

Figure 6: Estimation of the mean density of the homogeneous segment process Θ_1 at $x = (0.5, 0.5)$ by means of the "Minkowski content"-based estimator for different values of the bandwidth, based on a sample with size $N = 10$ (a), and $N = 100$ (b). For $r_{10}^{\mathrm{o,AMSE}} \approx 13 pixels (0.037873)$ the correspondig estimate is $\widehat{\lambda}_{\Theta_1}^{\mu,N}(0.5, 0.5) = 11.769231$; for $r_{100}^{\mathrm{o,AMSE}} \approx 6 pixels (0.017579)$ the corresponding estimate is $\widehat{\lambda}_{\Theta_1}^{\mu,N}(0.5, 0.5) = 10.545$.
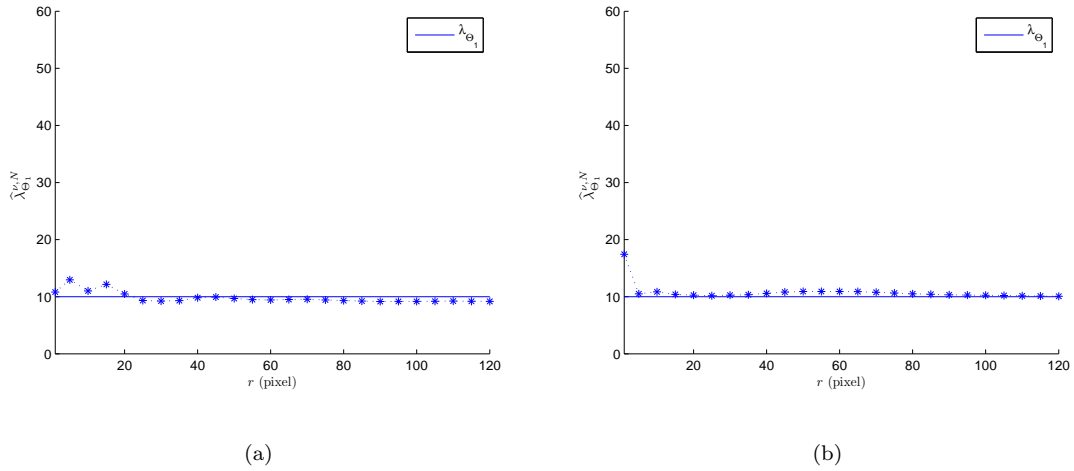
(a)

(b)

Figure 7: Estimation of the mean density of the homogeneous segment process Θ_1 at $x = (0.5, 0.5)$ by means of the natural estimator for different values of the bandwidth, based on a sample with size $N = 10$ (a), and $N = 100$ (b). For $r = 105$ pixels (≈ 0.3), the corresponding estimate is $\widehat{\lambda}_{\Theta_1}^{\nu, N}(0.5, 0.5) = 9.202444$ when $N = 10$, and $\widehat{\lambda}_{\Theta_1}^{\nu, N}(0.5, 0.5) = 10.206940$ when $N = 100$.
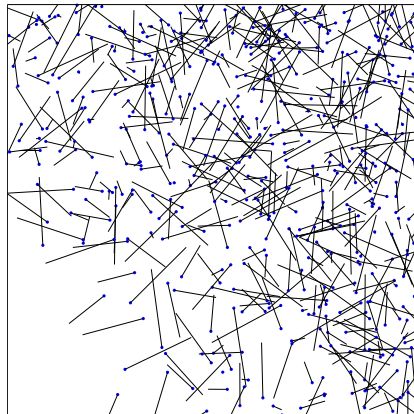

Figure 8: A realization of Θ_1 in the observation window $W = [0, 1]^2$.
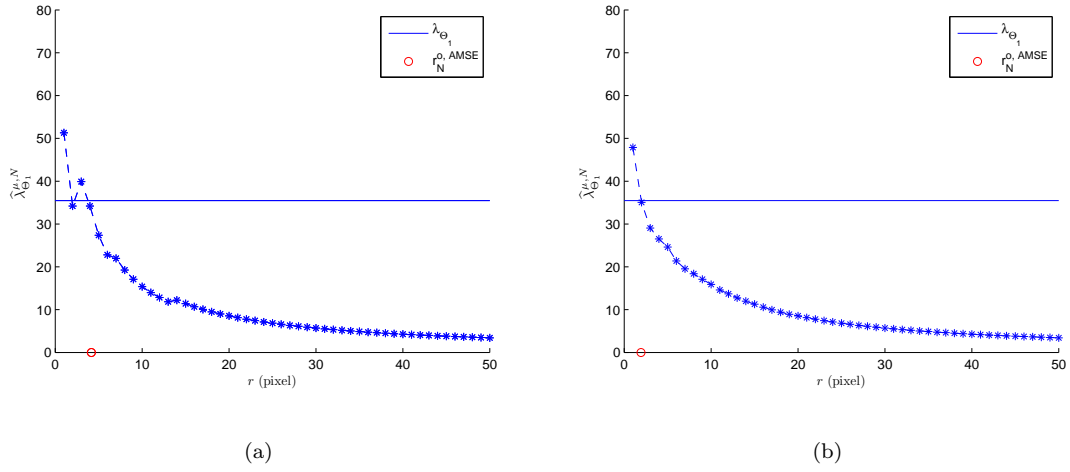
(a)   (b)

Figure 9: Estimation of the $\lambda_{\Theta_1}((0.5, 0.5)) = 35.4667$ by means of the "Minkowski content"-based estimator for different values of the bandwidth, based on a sample with size $N = 10$ (a), and $N = 100$ (b). For $r_{10}^{\mathrm{o,AMSE}} \approx 4 pixels (0.0122)$ the corresponding estimate is $\widehat{\lambda}_{\Theta_1}^{\mu,N}(0.5, 0.5) = 34.2$; for $r_{100}^{\mathrm{o,AMSE}} \approx 2 pixels (0.0057)$ the corresponding estimate is $\widehat{\lambda}_{\Theta_1}^{\mu,N}(0.5, 0.5) = 35.0550$.