



CENTRE FOR **STOCHASTIC GEOMETRY**  
AND ADVANCED **BIOIMAGING**



Ó. Thórisdóttir, A.H. Rafati and M. Kiderlen

## **Estimating the surface area of non-convex particles from central planar sections**

No. 08, September 2013

# Estimating the surface area of non-convex particles from central planar sections

Ó. Thórisdóttir<sup>1</sup>, A.H. Rafati<sup>2</sup> and M. Kiderlen<sup>1</sup>

<sup>1</sup>Department of Mathematics, Aarhus University

<sup>2</sup>Stereology and EM Research Laboratory, Aarhus University

## Abstract

In this paper, we present a new surface area estimator in local stereology. This new estimator is called the “Morse type surface area estimator” and is obtained using a two-stage sampling procedure. First a plane section through a fixed reference point of a three-dimensional structure is taken. In this section plane a modification of the area tangent count method is used. The Morse type estimator generalizes Cruz-Orive’s pivotal estimator for convex objects to non-convex objects. The advantages of the Morse type estimator over existing local surface area estimators are illustrated in a simulation study. The Morse type estimator is well suited for computer assisted confocal microscopy and we demonstrate its practicability in a biological application: the surface area estimation of the nuclei of giant-cell glioblastoma from microscopy images. We also present an interactive software that allows the user to efficiently obtain the estimator.

*Keywords:* surface area, local stereology, area tangent count, invariator principle, Morse type surface area estimator

## 1 Introduction

In local stereology, statistical inference about geometric characteristics like volume and surface area is made by taking random sections through a fixed reference point of a spatial structure of interest. This branch of stereology is tailor-made for applications e.g. in biology, where a typical example is optical sectioning of a cell through its nucleus; see the monograph by Jensen (1998) on local stereology. The nucleator derived by Gundersen (1988) is a well-established local stereological estimator for volume. Until recently, the surfactor was the only local stereological estimator available for estimating surface area; see Jensen and Gundersen (1987) and Jensen (1998, Section 5.6). Unlike the nucleator, the surfactor has not become a standard estimation tool. This might be due to the fact that the surfactor requires measuring

---

Corresponding author: Ó. Thórisdóttir, olofth@imf.au.dk, phone: +45-87155802  
Postal Address: Ny Munkegade 118, 8000 Aarhus C, Denmark

angles in a section of the object and angle measurements can be rather cumbersome in practice.

Cruz-Orive (2005) derived a new surface area estimator, the invariator estimator. It is an unbiased estimator for the surface area of the boundary of a spatial structure that only requires counting in a two-dimensional isotropic random plane. The estimator is obtained by combining the classical Crofton formula with the invariator principle. The invariator principle states how a line in an isotropic random plane must be chosen such that it is IUR in three-dimensions. The principle has been applied widely and we refer to Thórisdóttir and Kiderlen (Submitted, 2013) for an overview of invariator related results. An unbiased estimator for volume is also obtained by combining the invariator principle with Crofton’s formula but the invariator estimator for volume does not seem to enjoy any advantages over the nucleator. We therefore restrict attention to surface area estimation.

When the spatial structure is convex, improved versions of the invariator estimator have been suggested: the flower estimator and a discretization of it called the pivotal estimator; see Cruz-Orive (2005) and Cruz-Orive (2011). These estimators require more workload than the invariator estimator but adding this extra effort results in considerable variance reduction. Some clinical experts convey concern about the convexity assumption and claim that not many objects in practice can be assumed to be convex in shape. The new rotational Crofton formulae derived in Thórisdóttir and Kiderlen (Submitted, 2013, Theorems 6 and 7) tackle this problem. They present an unbiased estimator for surface area, which is better (with respect to variance) than the invariator estimator but still works for non-convex objects. The aforementioned convex estimators, the flower estimator and discretizations of it, are special cases of this new estimator. The new estimator can be obtained using a modification of the *area tangent count* method by DeHoff (1967) on a section of the structure. For a given section profile and a given direction in the section plane, a sweeping line is used and all tangents to the section profile are recorded, together with their type (if they represent a positive or a negative tangent) and their distance from the origin. We call the distance of a tangent from the origin a *critical value* and the type an *index*. The new estimator can be written entirely in terms of these indices and critical values. The estimator was originally derived using classical Morse theory and is therefore called the *Morse type surface area estimator*. It is well suited for computer assisted confocal microscopy and the main goal of this paper is to illustrate its feasibility using an expert-assisted procedure.

After introducing some notation in Section 2 we give a detailed overview of the sampling designs for the different surface area estimators discussed in this paper. The motivation for deriving the new rotational Crofton formulae came from studying the variance of the invariator estimator. The estimator is obtained by choosing three random variables: an IR section plane and both of the polar coordinates (a direction and a distance) of a uniformly distributed point in the section plane. In Section 3.1 we show how the variance of the invariator estimator can be decomposed according to these three different variables. These different variance contributions are calculated numerically when the object of interest is an ellipsoid as discussed in Section 3.2. It turns out that the variance from choosing the distance of the uniformly distributed point from the origin is by far the biggest variance contributor.

This variance contribution is not present in the new Morse type surface area estimator which is what makes it much more efficient than the invariator estimator. Instead of choosing a uniformly distributed point, only a uniform direction in the plane is chosen and all tangents are found at this direction. A further variance reduction is achieved by applying angular systematic sampling in the section plane, that is, by not only finding tangents for one direction but for several, as discussed in Section 3.3.

In Section 3.4 we give recommendations concerning the implementation of the Morse type estimator and compare it in a simulation study to the invariator grid estimator in Cruz-Orive (2005, Section 3.1). Finding tangents for four directions in the section plane gives a very precise estimator but using only two directions also results in a good estimator. The Morse type estimator should always be preferred to the invariator grid estimator, at least when the expert-assisted approach, derived in this paper, is used. The application of the Morse type surface area estimator is illustrated in a study of giant-cell glioblastoma in Section 4. We present an expert-assisted procedure to obtain the estimator and use the procedure to estimate the surface area of the nuclei of giant-cell glioblastoma from microscopy images. The results are presented in Section 4.4. We conclude the paper with a brief discussion.

## 2 Theoretical background

Before giving an overview of the different surface area estimators, we introduce the notation and recall some important concepts. Throughout,  $\mathbb{R}^n$  denotes the  $n$ -dimensional Euclidean space,  $O$  its origin,  $\|\cdot\|$  the Euclidean norm and  $\langle \cdot, \cdot \rangle$  the Euclidean scalar product. We restrict consideration to three-dimensional space although many of the results can be generalized to arbitrary dimension. We let  $RB_3 = \{x \in \mathbb{R}^3 \mid \|x\| \leq R\}$  be the 3-dimensional ball of radius  $R$ , centered at  $O$ , and  $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$  be the unit sphere. For  $z \in \mathbb{R}^2$  we write  $z^\perp$  for the line through  $O$  that is orthogonal to the axis joining  $z$  with  $O$  and  $F_z = z + z^\perp$  for the line that is parallel to  $z^\perp$  and passes through the point  $z$ .

We use  $\mathcal{K}$  for the family of all *convex bodies* (compact, convex sets with nonempty interior) of  $\mathbb{R}^3$ . A non-degenerate ellipsoid is an example of a convex body. If the axes of an ellipsoid  $E_3 \subseteq \mathbb{R}^3$  are parallel to the usual coordinate axes it is of the form

$$E_3 = \left\{ x \in \mathbb{R}^3 \mid \sum_{i=1}^3 \left( \frac{x_i - z_i}{a_i} \right)^2 \leq 1 \right\}, \quad (2.1)$$

where  $a_1, a_2, a_3 > 0$  are the lengths of the semiaxes and  $z = (z_1, z_2, z_3) \in \mathbb{R}^3$  is its center. For  $X \in \mathcal{K}$  its *support function*,  $h_X$ , is given by

$$h_X(u) = \max_{x \in X} \langle u, x \rangle, \quad u \in S^2.$$

The value  $h_X(u)$  is the signed distance from  $O$  to the supporting hyperplane to  $X$  with outer unit normal vector  $u$ .

We let  $\mathcal{L}_2$  be the family of all two-dimensional linear (passing through  $O$ ) planes. A random plane  $L_2 \in \mathcal{L}_2$  is called *isotropic random* (IR) if its distribution is rotation

invariant on  $\mathcal{L}_2$ . We introduce the surface area estimators for a fixed spatial structure  $X \subseteq \mathbb{R}^3$ . The spatial structure can either be a polyconvex set (composed of a finite union of convex bodies) or a compact set with smooth boundary. If  $X$  is a compact set with smooth boundary we have to add the technical requirement that  $X \cap L_2$  has again a smooth boundary, for almost all  $L_2 \in \mathcal{L}_2$ . This assumption is discussed in Thórisdóttir and Kiderlen (Submitted, 2013) and typically imposes no restrictions in practice. In Section 4 we extend the setting to a random particle  $X$ . We write  $\partial X$  for the boundary of  $X$ ,  $S(\partial X)$  for the surface area of its boundary,  $V(X)$  for its volume and  $\chi(X)$  for its Euler characteristic. If  $X \subseteq \mathbb{R}^1$  is compact,  $\chi(X)$  is the number of connected components of  $X$ . An overview of the notation of the different surface area estimators that will be discussed, their name and where they originate from can be found in Table 1.

**Table 1:** An overview of the surface area estimators relevant for the present paper.

Notation	Name	Reference
$\hat{S}_{\text{inv}}$	invariant estimator	Cruz-Orive (2005)
$\hat{S}_{\text{grid}}$	invariant grid estimator	Cruz-Orive (2005)
$\hat{S}_N^+$	pivotal estimator	Cruz-Orive (2005)
$\hat{S}_{\text{flo}}$	generalized flower estimator	Thórisdóttir and Kiderlen (Submitted, 2013)
$\hat{S}_N$	Morse type estimator	Thórisdóttir and Kiderlen (Submitted, 2013)

All these estimators are unbiased and are based on measurements of  $X$  in central two-dimensional sections. In a typical application  $X$  could be a biological cell with a reference point (for instance a nucleus or nucleolus) which we identify with the origin. The data necessary for the estimators is obtained from an isotropic two-dimensional section through  $X$ .

## 2.1 The invariant estimator for surface area

Combining the classical Crofton formula and the invariant principle, the surface area of  $\partial X$  can be written as

$$S(\partial X) = \mathbb{E} \left[ 4 \int_{L_2} \chi(X \cap L_2 \cap F_z) dz^2 \right] \quad (2.2)$$

where the expectation is with respect to the IR distribution of the section plane  $L_2 \in \mathcal{L}_2$ ; see Cruz-Orive (2005, first eq. (2.10)). Recall that  $F_z$  is the line in the section plane passing through  $z$  and orthogonal to the axis connecting  $z$  with  $O$ . Introducing polar coordinates in the section plane, we have

$$S(\partial X) = \mathbb{E} \left[ 4 \int_{S^2 \cap L_2} \int_0^\infty \chi(X \cap L_2 \cap F_{ru}) r dr du \right]. \quad (2.3)$$

The line  $F_{ru}$ , where  $r$  is its distance from  $O$ , is referred to as an  $r$ -weighted line in Cruz-Orive (2005).

We assume now that  $X$  is contained in a known reference set, which we take to be a ball  $RB_3$  of radius  $R > 0$ . The invariator estimator for surface area proposed by Cruz-Orive (2005, first eq. (2.12)) can be expressed as

$$\hat{S}_{\text{inv}} = 4\pi R^2 \chi(X \cap L_2 \cap F_Z), \quad (2.4)$$

where  $L_2 \in \mathcal{L}_2$  is an IR plane and  $Z \sim \text{unf}(RB_3 \cap L_2)$  is a uniformly distributed point on the disk  $RB_3 \cap L_2$ . From (2.2) it follows that  $\hat{S}_{\text{inv}}$  is an unbiased estimator for  $S(\partial X)$ . It can be obtained using the following sampling procedure:

1. Choose  $R > 0$  such that  $X \subseteq RB_3$ .
2. Choose an IR  $L_2$  (for instance by parametrizing  $S^2$  and choosing an isotropic unit vector  $V \in S^2$ , being a unit normal of  $L_2$ ).
3. Choose a uniformly distributed point  $Z$  on the disk  $RB_3 \cap L_2$ .
4. Determine the estimator by counting the number of connected components of  $X \cap L_2 \cap F_Z$ .

It is important to notice that the estimator is unbiased for any given reference set containing the object.

### 2.1.1 Invariator grid estimator

To reduce the variance of the invariator estimator, it was suggested in Cruz-Orive (2005, Section 3.1) to use systematic random sampling in the IR section plane as described in the following. Apply the first two steps in the sampling procedure described in Section 2.1. For a given IR section plane  $L_2$  a twice periodic point grid of grid distance  $d > 0$ , with  $O$  chosen uniformly at random in a square of area  $d^2$ , is thrown onto the section plane; see Cruz-Orive et al. (2010, Fig.1) for illustration. For any grid point  $z$  in the reference space  $RB_3$  consider the test line  $F_z$ . Then

$$\hat{S}_{\text{grid}} = 4d^2 \sum_z \chi(X \cap L_2 \cap F_z), \quad (2.5)$$

where the sum is over all grid points  $z$  in the reference space, is an unbiased estimator for  $S(\partial X)$ . The application of this estimator was illustrated by Cruz-Orive et al. (2010) on a group of rat brains.

## 2.2 Morse type surface area estimator

In this paper we promote another improvement of the invariator estimator, which also reduces variance. It uses the fact that the Euler characteristic in (2.3) only changes value (for  $u \in S^2 \cap L_2$  fixed and  $r \in [0, \infty)$  varying), if the line  $F_{ru}$  is tangent to the section profile  $X \cap L_2$ . We call the distance  $r$  of a tangent line from  $O$  a critical value. The innermost integral in (2.3) can be calculated explicitly in terms of the critical values of the section profile. This presents surface area estimators that can be written entirely in terms of critical values in a section plane. These estimators were derived by Thórisdóttir and Kiderlen (Submitted, 2013) in arbitrary dimension using Morse theory. Morse theory studies the topology of manifolds in terms of

functions defined on the manifolds; see the famous monograph by Milnor (1963). In our setting, the manifold is the boundary of the object of interest. Here we will not discuss Morse theory but rather give an intuitive explanation of critical values based on a procedure that is known in stereology under the name of *area tangent count*; see DeHoff (1967). We refer to Thórisdóttir and Kiderlen (Submitted, 2013) for a more mathematically rigorous derivation of the estimators.

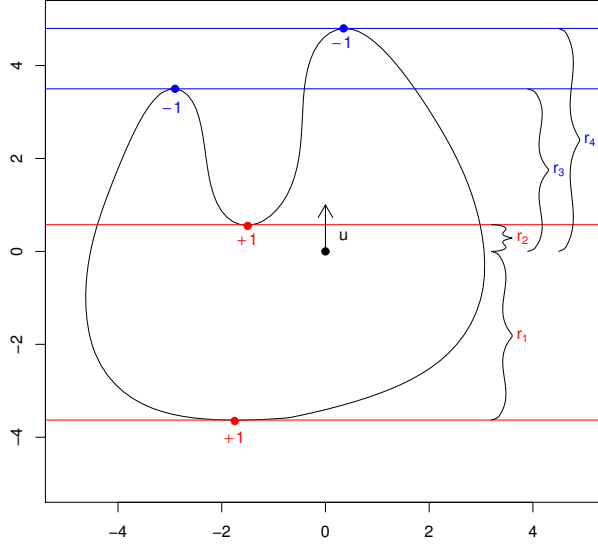
### 2.2.1 A modification of the area tangent count method

For an IR section plane  $L_2 \in \mathcal{L}_2$  let  $Y = X \cap L_2$  be the section profile and  $u \in S^2 \cap L_2$  a given direction in the section plane. The idea of the area tangent count method is to sweep the line  $u^\perp$  through the section profile and find all translates of the line that are tangent to  $Y$ . We distinguish between *positive tangents* (+) and *negative tangents* (-). When the sweeping line passes a tangent there is either an increase or a decrease in the number of connected components (of the sweeping line section with the profile). If the tangent represents an increase in the number of connected components we say that it is a positive tangent and a negative tangent if it represents a decrease in the number of connected components.

In the classical use of the area tangent count method, e.g. in DeHoff (1967) where the integral mean curvature in a structure is estimated, only the number of tangents and their type are registered. As we are interested in calculating the weighted integral in (2.3), we also record the distance of each tangent from  $O$ . If the line  $ru + u^\perp$  is tangent to  $Y$  we call the distance  $r = r(u)$  a *critical value* of  $Y$ . Moreover we let  $\iota_u(r) \in \{-1, 1\}$  be the type of the tangent, which we also call its *index*. A positive tangent (a tangent with index 1) can occur in two ways: either the sweeping line enters a new connected component of  $Y$ , like for the lowest critical value  $r_1$  of Figure 1, or a connected component splits up, like for the critical value  $r_2$  in Figure 1. A negative tangent (a tangent with index  $-1$ ) indicates that the sweeping line leaves a part of  $Y$ , like for the remaining two critical levels in Figure 1, or that two components of  $Y$  intersected with the line melt together (not represented in Figure 1).

The above interpretation of critical values also leads to an intuitive protocol for finding them. Given a direction  $u$  in the section plane, find all tangents to the section profile  $Y$  that are parallel to  $u^\perp$ . For each of these tangents with critical value  $r_0$ , say, determine its index by checking if the number of connected components  $\chi(Y \cap F_{ru})$  is increasing or decreasing when sweeping over value  $r_0$ .

A positive tangent is sometimes referred to as a “convex” tangent and a negative one as a “concave” tangent, see e.g. Baddeley (1984). Baddeley (1984) remarked that tangent counting is derived for ideal smooth objects and that it can therefore be unstable when it is applied to blurred images. This is not a severe practical limitation in our setting as we do not only use the number of tangents for a given section profile in a given direction, but rather their distances from  $O$ , the critical values.



**Figure 1:** Critical levels and indices for a given direction  $u$  in a section profile. Positive tangents are colored red while negative tangents are colored blue.

### 2.2.2 Surface area in terms of critical values

If  $m = m(u)$  is the number of critical values of a section profile  $Y$  in direction  $u$ , enumerated such that  $r_1 < r_2 < \dots < r_m$ , the Euler characteristic in (2.3) can be written as

$$\chi(Y \cap F_{ru}) = \sum_{i:r_i \leq r} \iota_u(r_i) \quad (2.6)$$

for almost all  $r$ ; see Thórisdóttir and Kiderlen (Submitted, 2013, Section 4.3). This says that the number of connected components can be given entirely in terms of the tangents. For instance, for the example in Figure 1, we have

$$\chi(Y \cap F_{2u}) = 1 + 1 = 2, \quad \chi(Y \cap F_{4u}) = 1 + 1 - 1 = 1.$$

Inserting the expression for the Euler characteristic into (2.3), the innermost integral can be calculated explicitly, and we obtain

$$S(\partial X) = \mathbb{E} \left[ \int_{S^2 \cap L_2} M(X \cap L_2, u) du \right] \quad (2.7)$$

where

$$M(Y, u) = \sum_{k=2}^m (r_k |r_k| - r_{k-1} |r_{k-1}|) \sum_{i=1}^{k-1} \iota_u(r_i) \quad (2.8)$$

depends on all the critical values  $r_1 < r_2 < \dots < r_m$  of  $Y = X \cap L_2$  in direction  $u \in S^2 \cap L_2$ . The function  $M$  may look complicated but it is only a linear combination of the squared critical values. For the example in Figure 1,

$$M(Y, u) = (r_2^2 + r_1^2) + (r_3^2 - r_2^2)2 + (r_4^2 - r_3^2) = r_1^2 - r_2^2 + r_3^2 + r_4^2.$$



It follows from (2.7) that the estimator

$$\hat{S}_1 = 2\pi M(X \cap L_2, U), \quad (2.9)$$

where  $L_2 \in \mathcal{L}_2$  is IR and  $U \sim \text{unf}(S^2 \cap L_2)$ , is an unbiased estimator for the surface area of  $\partial X$ . The estimator can be obtained using the following sampling procedure:

1. Choose an IR  $L_2$ .
2. Choose a uniformly distributed direction  $U \sim \text{unf}(S^2 \cap L_2)$  in the section plane.
3. For a given direction  $U = u$  find all tangents of the section profile and record their critical values and indices.
4. Calculate the linear combination given by (2.8).

When  $X \in \mathcal{K}$ , there are only two critical values for every given direction and using the definition of the support function, the function  $M$  simplifies

$$M(Y, u) = h_Y(u)|h_Y(u)| + h_Y(-u)|h_Y(-u)|. \quad (2.10)$$

If furthermore  $O \in X$ , the support function is non-negative and

$$M(Y, u) = h_Y^2(u) + h_Y^2(-u), \quad (2.11)$$

so then the estimator given by (2.9) becomes

$$\hat{S}_1 = 2\pi(h_{X \cap L_2}^2(U) + h_{X \cap L_2}^2(-U)).$$

### 2.2.3 Generalized flower estimator

If it is possible to find tangents for *all* directions in the section plane, the unbiased estimator for surface area

$$\hat{S}_{\text{flo}} = \int_{S^2 \cap L_2} M(X \cap L_2, u) du, \quad (2.12)$$

where  $L_2 \in \mathcal{L}_2$  is IR, can be obtained. In other words

$$\hat{S}_{\text{flo}} = \mathbb{E}[\hat{S}_{\text{inv}} | L_2]. \quad (2.13)$$

When  $X \in \mathcal{K}$  and  $O \in X$ , equation (2.12) can be written as

$$\hat{S}_{\text{flo}} = 2 \int_{S^2 \cap L_2} h_{X \cap L_2}^2(u) du, \quad (2.14)$$

where the reflection invariance of the Hausdorff measure has been used. This is the flower estimator given in Cruz-Orive (2005).

As shown by Cruz-Orive (2005, (2.19)), the flower estimator can also be written as four times the area of a so-called flower set, which is the set whose radial function is the support function of  $Y = X \cap L_2$ ; see also Thórisdóttir and Kiderlen (Submitted, 2013, p. 3). When  $Y$  is a planar polygon, the flower set is a union of finitely many

disks and resembles slightly a flower, which explains the terminology; see the figure by Cruz-Orive (2011, Fig.5). The estimator  $\hat{S}_{\text{flo}}$  in (2.12) can be referred to as the generalized flower estimator. There exists a simple computational formula for the generalized flower estimator when  $X$  is a simply connected polytope with interior points. The formula can be found in Thórisdóttir and Kiderlen (Submitted, 2013, Corollary 10) and only requires a list of the vertices of the polygon  $Y = X \cap L_2$ , where  $L_2 \in \mathcal{L}_2$  is IR. When  $X$  is a convex polytope containing  $O$ , alternative formulae can be found in Cruz-Orive (2011, Proposition 3) and Cruz-Orive (2012, Corollary 2).

#### 2.2.4 Morse type surface area estimator

When  $X$  is not a polytope, a good compromise between accuracy and effort might be cyclic systematic sampling of finitely many unit vectors in  $S^2 \cap L_2$  for a discrete approximation of the integral in (2.12). Let  $N \in \mathbb{N}$  be the number of directions to be sampled. Then the following unbiased estimator is obtained

$$\hat{S}_N = \frac{2\pi}{N} \sum_{l=0}^{N-1} M(X \cap L_2, u_{\alpha_0 + l \frac{\pi}{N}}), \quad (2.15)$$

where  $L_2 \in \mathcal{L}_2$  is IR,  $\alpha_0 \sim \text{unf}[0, \pi/N)$  and  $u_\alpha$  is the unit vector in  $L_2$  making an angle  $\alpha$  with a fixed axis in the section plane. This estimator is referred to as the *Morse type surface area estimator* as, like already mentioned, it was originally derived using the theory of Morse. When  $N = 1$ , the estimator (2.9) is obtained. For convex  $X$

$$\hat{S}_N = \frac{2\pi}{N} \sum_{l=0}^{N-1} (h_{X \cap L_2}^2(u_{\alpha_0 + l \frac{\pi}{N}}) + h_{X \cap L_2}^2(-u_{\alpha_0 + l \frac{\pi}{N}})). \quad (2.16)$$

For  $N = 2$ , equation (2.16) can be found in Cruz-Orive (2005, Eq. (3.2)). Dvořák and Jensen (2013) showed that  $\hat{S}_{\text{flo}}$  and  $\hat{S}_2$  are identical when  $X$  is a three-dimensional ellipsoid with  $O$  in its interior. To avoid confusion we remark that the number of sampled directions  $N$  in (2.16) equals half the number of sampled rays in Dvořák and Jensen (2013, Eq. (8)). More specifically, if rays are used and  $O$  is contained in the interior of  $X$ , the estimator

$$\hat{S}_N^+ = \frac{4\pi}{N} \sum_{l=0}^{N-1} h_{X \cap L_2}^2(u_{\alpha_0 + l \frac{2\pi}{N}}), \quad (2.17)$$

where now  $\alpha_0 \sim \text{unf}[0, 2\pi/N)$ , is an unbiased estimator for the surface area of the boundary of  $X$ . This estimator is referred to as the pivotal estimator in Cruz-Orive (2008) and Dvořák and Jensen (2013, Eq. (8)). We note that  $\hat{S}_N = \hat{S}_{2N}^+$ . In practice it is more natural to sweep a line at a given direction through the whole section profile and not stop when the line hits  $O$ . Therefore we introduced the Morse type estimator (2.15) using the  $M$  function given by (2.8) instead of using  $2M^+$  given by

$$M^+(Y, u) = \sum_{k: r_k > 0} (r_k^2 - \max(0, r_{k-1})^2) \sum_{i=1}^{k-1} \iota_u(r_i), \quad (2.18)$$

where  $M^+(Y, u) = 0$  if  $r_k < 0$  for all  $k$ .

### 3 Variance

The variance of the surface area estimators in Table 1 has only been studied to a very limited extent and the purpose of this section is to analyse the variance in greater detail. The main result is a decomposition of the variance of the invariator estimator  $\hat{S}_{\text{inv}}$ . As these different variance contributions cannot be evaluated easily for general sets, we discuss the different variance contributions for ellipsoids. The decomposition can be used to express the variances of the Morse type- and the generalized flower estimator. A recommendation concerning the choice of  $N$  when applying  $\hat{S}_N$  is given and it is compared to the invariator grid estimator  $\hat{S}_{\text{grid}}$  in a simulation study.

#### 3.1 Variance decomposition

To evaluate the quality of the invariator estimator  $\hat{S}_{\text{inv}}$ , we studied the different sources of variation separately. The variance of the estimator can be decomposed into three different contributions according to the integration variables in (2.3):

$$V_{\text{dist}} = \mathbb{E} \text{Var}(\hat{S}_{\text{inv}} | L_2, u), \quad (3.1)$$

which is the variance contribution from choosing the distance of the line  $F_z$  from  $O$ ,

$$V_{\text{orient}} = \mathbb{E} \text{Var}(\mathbb{E}[\hat{S}_{\text{inv}} | L_2, u] | L_2), \quad (3.2)$$

which is the variance contribution from choosing the orientation of the line  $F_z$  and

$$V_{\text{plane}} = \text{Var} \mathbb{E}[\hat{S}_{\text{inv}} | L_2], \quad (3.3)$$

which is the variance contribution from choosing the IR section plane  $L_2$ . This is made more explicit in the following proposition.

**Proposition 3.1.** *The variance of  $\hat{S}_{\text{inv}}$  in (2.4) can be decomposed into three parts*

$$\text{Var}(\hat{S}_{\text{inv}}) = V_{\text{dist}} + V_{\text{orient}} + V_{\text{plane}}, \quad (3.4)$$

where

$$V_{\text{dist}} = 16\pi^2 \mathbb{E} \left[ M^+(X \cap L_2, u) (R^2 \sum_{k:r_k > 0} \sum_{i=1}^{k-1} \iota_u(r_i) - M^+(X \cap L_2, u)) \right], \quad (3.5)$$

$$V_{\text{orient}} = 16\pi^2 \mathbb{E} \text{Var}(M^+(X \cap L_2, u) | L_2), \quad (3.6)$$

$$V_{\text{plane}} = 4\pi^2 \text{Var} \mathbb{E}[M(X \cap L_2, u) | L_2]. \quad (3.7)$$

Here the function  $M$  is given by (2.8) and  $M^+$  by (2.18).

*Proof.* The variance of  $\hat{S}_{\text{inv}}$  is finite and hence, by the law of total variance

$$\text{Var}(\hat{S}_{\text{inv}}) = \mathbb{E} \text{Var}(\hat{S}_{\text{inv}} | L_2) + \text{Var} \mathbb{E}[\hat{S}_{\text{inv}} | L_2].$$

Let  $L_2 \in \mathcal{L}_2$  be a given IR section plane and  $(r, u)$  be the polar coordinates of a uniformly distributed point in  $RB_3 \cap L_2$ , with  $r \in [0, \infty)$  and  $u \in S^2 \cap L_2$ . Using the conditional version of the law of total variance in Bowsheer and Swain (2012), we get

$$\text{Var}(\hat{S}_{\text{inv}}|L_2) = \mathbb{E}[\text{Var}(\hat{S}_{\text{inv}}|L_2, u)|L_2] + \text{Var}(\mathbb{E}[\hat{S}_{\text{inv}}|L_2, u]|L_2).$$

The variance of  $\hat{S}_{\text{inv}}$  can therefore be decomposed into the three contributions given in (3.1)–(3.3). In the following we calculate these different contributions. Inserting the estimator  $\hat{S}_{\text{inv}}$  into (3.3), introducing polar coordinates in the section plane and then using the expression (2.6) for the Euler characteristic we find

$$\begin{aligned} V_{\text{plane}} &= \text{Var} \left( 4\pi R^2 \frac{1}{\pi R^2} \int_{L_2 \cap RB_3} \chi(X \cap L_2 \cap F_z) dz^2 \right) \\ &= 16\pi^2 \text{Var} \left( R^2 \frac{1}{2\pi} \int_{S^2 \cap L_2} \frac{1}{R^2} \int_{-\infty}^{\infty} \chi(X \cap L_2 \cap F_{ru}) |r| dr du \right) \\ &= 16\pi^2 \text{Var} \left( \frac{1}{2\pi} \int_{S^2 \cap L_2} \sum_{k=2}^m \sum_{i=1}^{k-1} \iota_u(r_i) \int_{r_{k-1}}^{r_k} |r| dr du \right). \end{aligned}$$

Integrating the inner integral explicitly and using the notation (2.8), we find

$$V_{\text{plane}} = 4\pi^2 \text{Var} \mathbb{E}[M(X \cap L_2, u)|L_2].$$

By applying similar arguments and the notation (2.18), we obtain

$$\begin{aligned} V_{\text{orient}} &= 16\pi^2 \mathbb{E} \text{Var} \left( 2 \int_0^\infty \chi(X \cap L_2 \cap F_{ru}) r dr \mid L_2 \right) \\ &= 16\pi^2 \mathbb{E} \text{Var}(M^+(X \cap L_2, u)|L_2). \end{aligned}$$

Furthermore, using that

$$V_{\text{dist}} = \mathbb{E}[\hat{S}_{\text{inv}}^2] - \mathbb{E}[\mathbb{E}[\hat{S}_{\text{inv}}|L_2, u]^2]$$

we get from (2.6) and (2.18) that

$$V_{\text{dist}} = 16\pi^2 \mathbb{E} \left[ R^2 M^+(X \cap L_2, u) \sum_{k:r_k > 0} \sum_{i=1}^{k-1} \iota_u(r_i) \right] - 16\pi^2 \mathbb{E}[(M^+(X \cap L_2, u))^2],$$

which simplifies to (3.5).  $\square$

Proposition 3.1 implies in particular, that only the variance contribution (3.5) depends on the size  $R$  of the reference set, and it is increasing with  $R$ . More importantly, (2.13) and (3.3) show that

$$V_{\text{plane}} = \text{Var}(\hat{S}_{\text{flo}})$$

is the variance of the generalized flower estimator. Moreover, as all the surface area estimators  $\hat{S}$  in Table 1 satisfy  $\mathbb{E}[\hat{S}|L_2] = \hat{S}_{\text{flo}}$  a variance decomposition similar to (3.4) for  $\text{Var}(\hat{S})$  implies that  $\hat{S}_{\text{flo}}$  has the lowest possible variance among them all. This is discussed further in Section 3.3.

### 3.1.1 Variance decomposition for convex bodies

In view of (2.10) the different variance contributions in Proposition 3.1 simplify when the particle  $X$  is convex. Using (2.11) and the rotation invariance of the Hausdorff measure, the expressions are particularly simple when the reference point is contained in  $X$ .

**Remark 3.2.** When  $X \in \mathcal{K}$  and  $O \in X$  the variance contributions (3.5)–(3.7) simplify to

$$\begin{aligned} V_{\text{dist}} &= 16\pi^2 \mathbb{E}[h_{X \cap L_2}^2(u)(R^2 - h_{X \cap L_2}^2(u))], \\ V_{\text{orient}} &= 16\pi^2 \mathbb{E} \text{Var}(h_{X \cap L_2}^2(u)|L_2), \\ V_{\text{plane}} &= 16\pi^2 \text{Var} \mathbb{E}[h_{X \cap L_2}^2(u)|L_2]. \end{aligned}$$

When  $X$  is a three-dimensional ellipsoid, the section profile  $Y = X \cap L_2$  is an ellipse that can be expressed with respect to an orthonormal basis of the section plane  $L_2$  as shown by Thórisdóttir (2010, Proposition 10). Hence, the support function of  $Y$  can be obtained and using elementary but tedious calculations the different variance contributions in Proposition 3.1 can be made more explicit. As the expressions are quite involved they are deferred to Proposition A.1 in the appendix. For illustration we give explicit analytic expressions for the different variance contributions when  $X$  in Proposition 3.1 is a three-dimensional ball of radius  $r$ . We assume without loss of generality that the ball is centered at  $O' = (0, 0, z)$ ,  $z \geq 0$ , and assume that the ball contains  $O$ .

**Proposition 3.3.** *Let  $X = O' + rB_3$ , with  $r > 0$ ,  $O' = (0, 0, z)$ ,  $z \in [0, r]$ , and assume that  $X \subseteq RB_3$ . Then the different variance contributions (3.5)–(3.7) in Proposition 3.1 can be expressed as*

$$\begin{aligned} V_{\text{dist}} &= 16\pi^2 r^2 \left( R^2 - r^2 - \frac{4}{3}z^2 \right), \\ V_{\text{orient}} &= 16\pi^2 z^2 \left( \frac{4}{3}r^2 - \frac{1}{5}z^2 \right), \\ V_{\text{plane}} &= \frac{16\pi^2 z^4}{5}. \end{aligned}$$

The proof uses Pythagoras' theorem and elementary but tedious calculations and can be found in Thórisdóttir (2010, p. 36–38). Adding up the different variance contributions in Proposition 3.3, we obtain the total variance of the estimator when  $X$  is a ball of radius  $r$

$$\text{Var}(\hat{S}_{\text{inv}}) = 16\pi^2 r^2 (R^2 - r^2); \quad (3.8)$$

see also Cruz-Orive (2008).

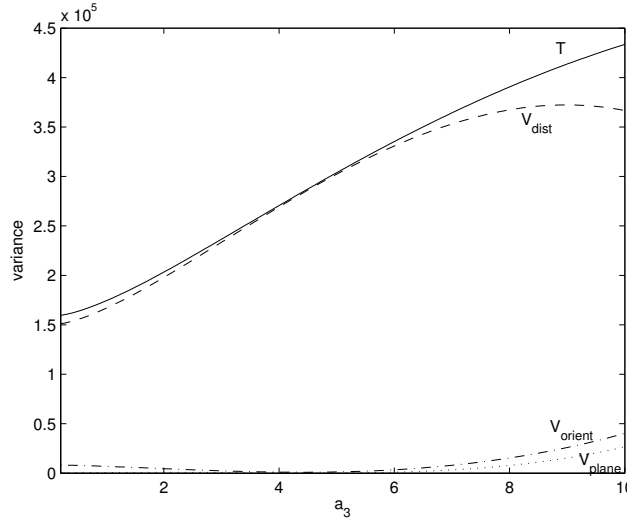
## 3.2 Numerical results for ellipsoids

As mentioned earlier, Proposition A.1 in the appendix gives more explicit expressions for the different variance contributions of the estimator  $\hat{S}_{\text{inv}}$  when  $X$  is a three-dimensional ellipsoid. It is quite involved, if at all possible, to derive explicit analytic formulas for these expressions. We therefore turned to numerical methods

in the language and interactive environment **Matlab** for calculating the different contributions (A.2)–(A.4). At the home page [home.imf.au.dk/olofth](http://home.imf.au.dk/olofth) programs for calculating the variance contributions in **Matlab** can be found. In the following we briefly describe the most important modules of this implementation.

We assume without loss of generality that the ellipsoid is of the form (2.1). In **calculations.m** the user reports the center  $(z_1, z_2, z_3)$  of the ellipsoid under study, the lengths,  $a_1, a_2$  and  $a_3$ , of the ellipsoidal axes, and the radius  $R$  of the reference ball. The output are numerical estimates of the respective variance contributions  $V_{\text{dist}}$  (the variance from choosing the distance of the line  $F_z$  from  $O$ ),  $V_{\text{orient}}$  (the variance due to the choice of the orientation of the line),  $V_{\text{plane}}$  (the variance due to choosing the IR section plane  $L_2$ ), the total variance obtained by adding up the different contributions **Total variance**  $= V_{\text{dist}} + V_{\text{orient}} + V_{\text{plane}}$ , and the total variance obtained from implementing the theoretical expression (A.6). We used the theoretical expression for the total variance (A.6) involving elliptic integrals to compare with the total variance obtained by adding up the different variance contributions. The comparison study included numerous different values of the parameters and implied that the algorithm is very stable and precise.

The function **dblquad** in **Matlab** was used to calculate numerically the double integrals in (A.2)–(A.4) and the program **elliptic12.m**, written by Moiseev Igor<sup>1</sup>, was used to calculate the incomplete elliptic integrals of first and second kind to obtain (A.6). In Figure 2 the contribution  $V_{\text{dist}}$  (dashed curve),  $V_{\text{orient}}$  (dashed-dotted



**Figure 2:** The different variance contributions for ellipsoids centered at  $(0.4, 0.2, 0.2)$ , with axes of lengths  $a_1 = 5$ ,  $a_2 = 4$  and  $a_3$  varying from 0.2 to 10.

curve),  $V_{\text{plane}}$  (dotted curve) and the total variance  $T = V_{\text{dist}} + V_{\text{orient}} + V_{\text{plane}}$  (solid curve) can be seen for ellipsoids embedded in a ball of radius 10.5 and given by

$$\left\{ x \in \mathbb{R}^3 \mid \frac{(x_1 - 0.4)^2}{5^2} + \frac{(x_2 - 0.2)^2}{4^2} + \frac{(x_3 - 0.2)^2}{a_3^2} \leq 1 \right\}, \quad (3.9)$$

where  $a_3$  varies from 0.2 to 10.

<sup>1</sup>The **Matlab** program was downloaded from: <http://www.mathworks.com/matlabcentral/fileexchange/8805-elliptic-integrals-and-functions>.

Further discussion on the different variance contributions for ellipsoids can be found in Thórisdóttir (2010, Section 3.5), where the shapes of the different curves, comparison between them and the role of the size of the reference space are discussed. The general conclusion is that the variance contribution  $V_{\text{dist}}$  (the variance contribution from choosing the distance of the closest point on the line  $F_z$  from  $O$ ), is very large relative to both  $V_{\text{orient}}$  and  $V_{\text{plane}}$ . Hence, to improve the estimator it is most efficient to concentrate on reducing  $V_{\text{dist}}$ .

### 3.3 Variance of the Morse type surface area estimator

The study of three-dimensional ellipsoids suggests strongly that the variance of  $\hat{S}_{\text{inv}}$  primarily comes from choosing the distance of the line  $F_z$  in (2.4) from  $O$ . We believe this to hold for more general sets, too. As this variance contribution is not present in the Morse type estimator  $\hat{S}_N$ , it is highly recommended to add the extra effort needed to find tangents, as a dramatic variance reduction is obtained.

As an example of the variance reduction obtained, the standard error is reduced by a factor of 2.9 for the ellipsoid given by (3.9), with  $a_3 = 10$  and  $R = 10.5$ , when the pivotal estimator  $\hat{S}_1^+$  is used instead of the invariator estimator  $\hat{S}_{\text{inv}}$ . When  $X$  is a ball of radius  $r > 0$ , centered at  $(0, 0, z)$ , we obtain from Proposition 3.3 that the coefficient of variation of  $\hat{S}_1^+$  is

$$CV(\hat{S}_1^+) = \frac{\sqrt{\text{Var}(\hat{S}_1)}}{\mathbb{E}\hat{S}_1} = \frac{2z}{\sqrt{3}r}. \quad (3.10)$$

This has been reported earlier by Cruz-Orive (2008, Eq. (6)) when  $X$  is a unit ball. Cruz-Orive (2008) compared this coefficient of variation to the one obtained for the surfactor estimator derived by Jensen and Gundersen (1987). For the unit ball,  $\hat{S}_1^+$  had smaller coefficient of variation than the surfactor.

As already mentioned after Proposition 3.1, the law of total variance immediately gives  $\text{Var}(\hat{S}_{\text{flo}}) \leq \text{Var}(\hat{S}_N)$  for all  $N \in \mathbb{N}$  but as  $\hat{S}_{\text{flo}}$  requires finding tangents in all directions in the section plane it is usually not feasible in practice (unless the section profile is a polygon or if its boundary can be approximated by a polygon using automated segmentation as briefly discussed in Section 5.2). We therefore restrict attention to  $\hat{S}_N$  in what follows.

Let  $Y = X \cap L_2$ ,  $L_2 \in \mathcal{L}_2$ , be the section profile. Applying systematic sampling in the section plane typically reduces the variance. More specifically, using that  $M(Y, U) = M(Y, U + \pi)$ ,  $U \sim \text{unf}(S^2 \cap L_2)$ , and Cauchy-Schwarz inequality, we obtain

$$\text{Var}(\hat{S}_{2N}) \leq \text{Var}(\hat{S}_N) \leq \text{Var}(\hat{S}_1) \quad (3.11)$$

for all  $N \in \mathbb{N}$ . This was shown for  $X \in \mathcal{K}$  in Dvořák and Jensen (2013, Section 2.2). We only mention here that using a result of Gual-Arnau and Cruz-Orive (2000, Eq. (2.4)), the variance of the Morse type estimator  $\hat{S}_N$  can be expressed as

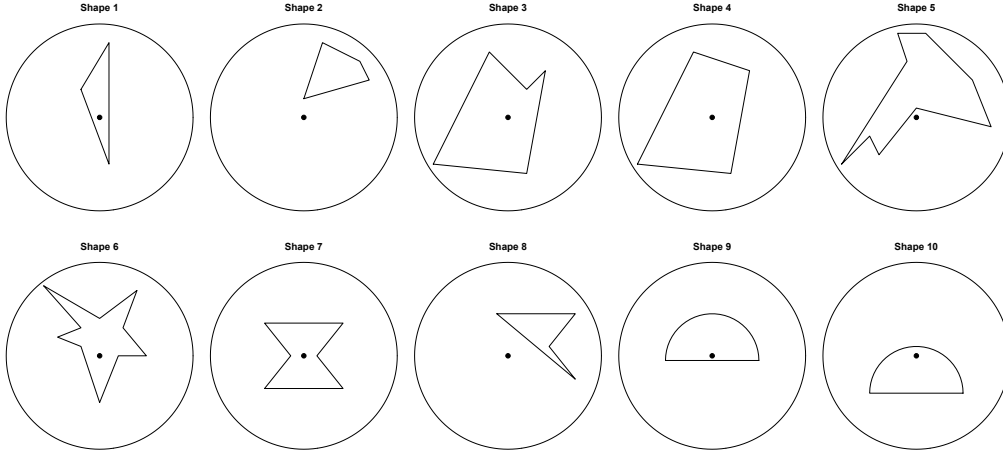
$$\text{Var}(\hat{S}_N) = V_{\text{plane}} + \mathbb{E} \left[ \frac{2\pi}{N} \sum_{l=0}^{N-1} g(u_{l\frac{\pi}{N}}) - \int_{S^2 \cap L_2} g(\omega) d\omega \mid L_2 \right],$$

where the function  $g$  is the circular covariogram of  $M$  on the section profile

$$g(\omega) = \int_{S^2 \cap L_2} M(Y, u) M(Y, u + \omega) du, \quad \omega \in S^2 \cap L_2.$$

### 3.4 Efficiency of the Morse type surface area estimator

We now discuss how the number  $N$  of directions in  $L_2$  should be chosen so that the variance of  $\hat{S}_N$  becomes small while still keeping a reasonable workload of the sampling procedure in  $L_2$ . As the variance contribution  $V_{\text{plane}}$  is common for all surface area estimators considered here, we ignore this contribution in the following discussion and consider the variance of  $\hat{S}_N$  given  $L_2$ . Clearly, the function  $f(N) = \text{Var}(\hat{S}_N | L_2)$  depends crucially on the shape of the underlying object. We therefore carried out a simulation study with a variety of different shapes shown in Figure 3. The simulations were done using R.



**Figure 3:** Polygons and half-circles.

In the next section, we determine the empirical CV's of  $\hat{S}_N$ ,  $N \in \{1, 2, 3, 4, 5\}$ , from independent Monte Carlo replications and derive an empirical recommendation for the choice of  $N$ . In Section 3.4.2 we compare empirically the precision and workload required for the new estimator  $\hat{S}_N$  and the well-known estimator  $\hat{S}_{\text{grid}}$ .

#### 3.4.1 Choice of $N$

Although the variance of  $\hat{S}_N$  for a given section profile does not increase when the number of orientations sampled is doubled, as shown in (3.11), it is not a non-increasing function of  $N$ . Table 2 presents estimates of the CV's for the shapes in Figure 3, which are all contained in a reference disk of radius one, centered at  $O$ . To recommend a value of  $N$ , we suggest that a CV of not more than 2.5 % is acceptable. Hence, the smallest  $N$  for a given shape where the empirical CV is 2.5 % or less, is our recommendation for the choice of  $N$ . We only need to find tangents in two directions for shapes 1, 5, 6 and 9, in three directions for shapes 2, 3 and 4 and in four directions for shapes 7, 8 and 10. For all simulated ellipses, two orientations give



**Table 2:** CV for the shapes in Figure 3.

CV	N = 1	N = 2	N = 3	N = 4	N = 5
Shape 1	0.6388	<b>0.0242</b>	0.0151	0.0085	0.0055
Shape 2	0.3410	0.0848	<b>0.0146</b>	0.0141	0.0123
Shape 3	0.2652	0.1048	<b>0.0249</b>	0.0087	0.0120
Shape 4	0.2587	0.1487	<b>0.0246</b>	0.0149	0.0090
Shape 5	0.1376	<b>0.0188</b>	0.0239	0.0084	0.0049
Shape 6	0.1970	<b>0.0228</b>	0.0065	0.0080	0.0029
Shape 7	0.2217	0.0747	0.0456	<b>0.0167</b>	0.0072
Shape 8	0.3025	0.0422	0.0331	<b>0.0204</b>	0.0082
Shape 9	0.2783	<b>0.0129</b>	0.0037	0.0023	0.0013
Shape 10	0.2542	0.0908	0.0255	<b>0.0162</b>	0.0091

a CV of less than 2.5 %. This is even true for ellipses not containing the reference point  $O$ . This is not surprising as it follows from Dvořák and Jensen (2013) that  $\hat{S}_2$  and  $\hat{S}_{\text{flo}}$  are identical when  $Y$  is an ellipse with  $O$  in its interior, as already mentioned at the end of Section 2.2 for three-dimensional ellipsoids. When the reference point is moved away from the center of an ellipse, the CV increases. This is evident for a disk from (3.10). The elongation of the ellipse does not seem to influence the CV when  $N > 1$ . For  $N = 1$ , the CV increases slightly for increased elongation of the ellipse.

A number of simulations were carried out and they suggest strongly that taking  $N = 4$  is sufficient for obtaining a CV of less than 2.5 % for a large class of shapes. Although the CV can increase as  $N$  increases, as is the case for Shape 5 with  $N = 2$  and  $N = 3$ , it is our experience that this increase is rare, and, when it occurs, it is not substantial. Therefore taking  $N = 4$  instead of  $N = 3$  should typically not decrease the precision of the estimator. When the object of interest resembles an ellipse, we recommend using  $N = 2$ . We note that these recommendations are only based on the variance of the sampling procedure. When there are other sources of variation that are much larger than the variance due to the sampling procedure,  $N = 2$  is typically adequate, see Section 4.4.

### 3.4.2 Comparison of $\hat{S}_N$ and $\hat{S}_{\text{grid}}$

We still consider a given section profile  $Y = X \cap L_2$  and fix  $L_2 \in \mathcal{L}_2$ . An alternative to the Morse type estimator  $\hat{S}_N$  is the invariator grid estimator  $\hat{S}_{\text{grid}}$  given by (2.5). We compare the efficiency of these two estimators by finding the amount of workload needed to estimate  $\mathbb{E}[\hat{S}_{\text{flo}}|L_2]$  at a given precision. Define the complexity numbers

$$C_{\hat{S}_N} = \text{total number of tangents for the } N \text{ directions,}$$

$$C_{\hat{S}_{\text{grid}}} = 2 \sum_z \chi(Y \cap F_z) + \sum_z \mathbf{1}_{\{(Y \cap F_z) = \emptyset\}}.$$

Hence,  $C_{\hat{S}_{\text{grid}}}$  is the total number of points in the intersection of the boundary of the section profile with the test lines plus the number of test lines that do not hit the section profile. Both complexity numbers are motivated by the number of mouse

clicks in an interactive microscopy software to determine the estimator; see also Section 4.3. The workload in obtaining these numbers is typically not equivalent, as determining a critical point is most likely more difficult than to determine if a line hits the boundary of the object or not. But given the ratio  $Q = C_{\hat{S}_{\text{grid}}}/C_{\hat{S}_N}$ , a clinical expert can decide if it is more feasible to use  $\hat{S}_N$  or  $\hat{S}_{\text{grid}}$ . The simulation study shows strong evidence for that  $\hat{S}_N$  should be preferred to  $\hat{S}_{\text{grid}}$ , at least when  $N > 1$ .

For a given polygon we generated 1000 independent replications of  $\hat{S}_N$  for a given  $N$  and of  $\hat{S}_{\text{grid}}$  for a given grid distance  $d$ . From these observations we calculated an empirical estimate for the variance of  $\hat{S}_N$ ,  $\hat{S}_{\text{grid}}$ , respectively, and recorded the complexity constants  $C_{\hat{S}_N}$ . We then used a normal kernel with smoothness parameter  $s$  to smooth  $\widehat{\text{Var}}(\hat{S}_{\text{grid}})$  as a function of the grid distance  $d$ . We denote the smoothed estimate of the variance by  $\widehat{\text{Var}}_s(\hat{S}_{\text{grid}})$ . For a given  $N$  we found  $d$  such that  $\hat{S}_{\text{grid}}$  has essentially the same precision as  $\hat{S}_N$ , that is, such that

$$|\widehat{\text{Var}}_s(\hat{S}_{\text{grid}}) - \widehat{\text{Var}}(\hat{S}_N)| \leq 10^{-6}.$$

Given this  $d$  we then again generated 1000 independent replications of  $\hat{S}_{\text{grid}}$  and calculated  $C_{\hat{S}_{\text{grid}}}$ . For a given  $N$  and the corresponding  $d$  we obtained an unbiased estimate of the ratio  $Q$

$$\hat{Q} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{C_{\hat{S}_{\text{grid}}}}{C_{\hat{S}_N}}.$$

In Table 3 results for the shapes in Figure 3 can be seen, where the smoothness parameter  $s = 0.1$  has been used. When  $N > 2$ ,  $d$  is close to zero and  $\hat{Q}$  very large (much greater than 100) for all the shapes in Figure 3. This means that  $\hat{S}_{\text{grid}}$  only attains the efficiency of  $\hat{S}_N$  when unrealistically many test lines are used, making the complexity number of  $\hat{S}_{\text{grid}}$  large compared to the one of  $\hat{S}_N$ . For ellipses we find that  $\hat{Q} > 2$  when  $N = 1$  and that it is very large when  $N = 2$ . The ratio  $Q$  typically decreases with increased elongation of an ellipse.

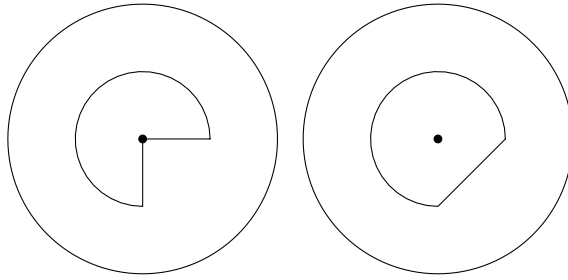
The simulations imply that  $Q$  is always greater than 2. Hence,  $\hat{S}_N$  should at least be preferred to  $\hat{S}_{\text{grid}}$  if the workload needed to identify a tangent, as well as its type and critical value, is not more than twice of that needed to determine if a line hits the boundary of an object or not. It should be noted that this factor 2 is a worst-case scenario. The ratio  $Q$  can be much larger than two even when  $N = 1$ , as can be seen in Table 3. When the software, described in Section 4.3, is used, the estimator  $\hat{S}_N$  can be obtained very efficiently and should therefore always be preferred to  $\hat{S}_{\text{grid}}$ .

### 3.4.3 Convex hull

Rather surprisingly it is not necessarily better (with respect to variance) to find  $\hat{S}_N$  for the convex hull of an object than for the true object. We show this by an example. The object in Figure 4 to the left is a disk of radius  $r$ , where one fourth has been cut off, and the one to the right is its convex hull. In Table 4 estimates for the CV's based on 1000 independent replications of  $\hat{S}_N$  for  $N \in \{1, 2, 3\}$  and each of the two objects can be seen.

**Table 3:** Comparison for the shapes in Figure 3.

		Var	$d$	$\hat{Q}$
Shape 1	N = 1	3.52	0.81	2.9
	N = 2	$4.8 \cdot 10^{-3}$	$\simeq 0$	$\gg 100$
Shape 2	N = 1	$6.7 \cdot 10^{-2}$	0.64	4.9
	N = 2	$7.2 \cdot 10^{-3}$	0.21	22.6
Shape 3	N = 1	$2.56 \cdot 10^{-1}$	0.65	3.7
	N = 2	$3.84 \cdot 10^{-2}$	0.32	7.9
Shape 4	N = 1	$2.08 \cdot 10^{-1}$	0.68	4.9
	N = 2	$7.2 \cdot 10^{-2}$	0.46	5.3
Shape 5	N = 1	1.23	0.48	5.0
	N = 2	$2.08 \cdot 10^{-2}$	$\simeq 0$	$\gg 100$
Shape 6	N = 1	1.47	0.51	3.2
	N = 2	$1.6 \cdot 10^{-2}$	$\simeq 0$	$\gg 100$
Shape 7	N = 1	$5.92 \cdot 10^{-1}$	0.36	8.6
	N = 2	$6.72 \cdot 10^{-2}$	0.10	52.1
Shape 8	N = 1	1.33	0.58	4.4
	N = 2	$1.34 \cdot 10^{-2}$	0.04	$\gg 100$
Shape 9	N = 1	$4.06 \cdot 10^{-1}$	0.43	10.0
	N = 2	$8.30 \cdot 10^{-4}$	$\simeq 0$	$\gg 100$
Shape 10	N = 1	$4.19 \cdot 10^{-1}$	0.40	12.0
	N = 2	$5.41 \cdot 10^{-2}$	0.16	35.0



**Figure 4:** A disk of radius  $r$ , where one fourth has been cut off, and its convex hull.

**Table 4:** CV for the objects in Figure 4

CV	$N = 1$	$N = 2$	$N = 3$
Fig. 4 left	<b>0.001140</b>	0.000165	0.000357
Fig. 4 right	0.075878	0.040691	<b>0.020190</b>

We note that the CV's for the convex hull, the object to the right in Figure 4, are considerably larger than those for the object to the left. This can be explained by considering the integrand in the generalized flower estimator, the function  $M$ . The true value  $\mathbb{E}[\hat{S}_{\text{flo}}|L_2]$  can be written as

$$\mathbb{E}[\hat{S}_{\text{flo}}|L_2] = 2 \int_0^{2\pi} f(\alpha) d\alpha.$$

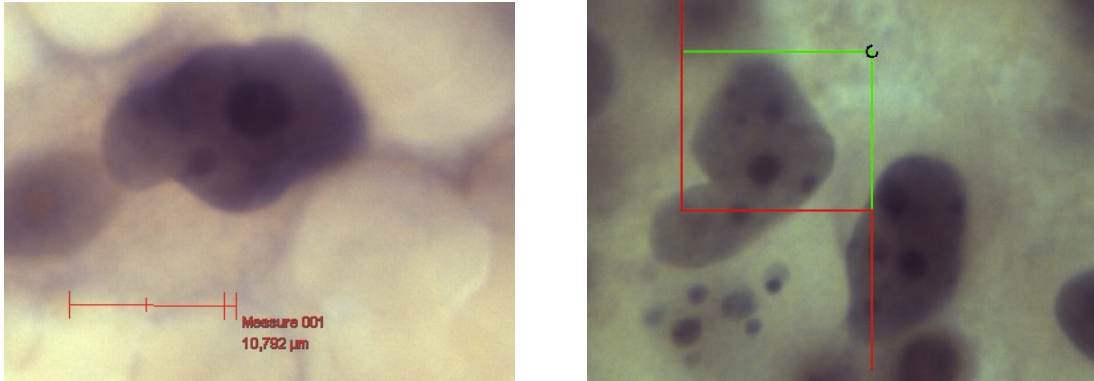
where for the object to the left the function  $f$  is a constant,  $f(\alpha) = r^2$  for all  $\alpha \in [0, 2\pi)$ , while for the convex hull

$$f(\alpha) = \begin{cases} r^2, & \alpha \in [0, 3\pi/2) \\ r^2 \sin^2 \alpha, & \alpha \in [3\pi/2, 7\pi/4) \\ r^2 \cos^2 \alpha, & \alpha \in [7\pi/4, 2\pi). \end{cases}$$

This explains the greater variance of  $\hat{S}_N$  for the convex hull.

## 4 Application of the Morse type surface area estimator to giant-cell glioblastoma

We illustrate the application of the Morse type estimator  $\hat{S}_N$  in a study of giant-cell glioblastoma. The goal is to estimate the average surface area of the nucleus of giant-cell glioblastoma from microscopy images. Giant-cell glioblastoma is a rare brain neoplasm which accounts for 1 % of glioblastomas. It is formerly known as monstro-cellular brain tumour. Histologically, the tumour is characterized by bizarre-looking, variable sized and shaped multinucleated giant-cells with abundant eosinophilic cytoplasm; see Ohgaki et al. (2000). Giant-cell glioblastoma is a subtype of glioblastoma multiforme, however the rarity of giant-cell glioblastoma has prevented further characterization. Giant-cell glioblastoma is more prevalent in young-male patients and tends to occur more in parietal and temporal lobe of brain. We chose to illustrate the estimator for the nuclei of giant-cell glioblastoma as they are typically non-convex in shape, see Figure 5.



**Figure 5:** Section profiles through a nucleoli of the nuclei of two giant-cell glioblastoma.

## 4.1 Model-based setting

In the preceding sections we have assumed that the structure of interest is deterministic and the section plane is random, that is we have worked in a design-based setting. We will now adopt the common model-based approach, where the particles, here nuclei of giant-cell glioblastoma, are random isotropic and the plane is deterministic. This isotropy assumption allows us to avoid a complicated, time consuming protocol, where each particle is sectioned physically or optically by a new isotropically generated plane. Hence we assume that the nuclei are realizations of an ergodic isotropic random process, meaning that nuclei far away from each other have independent orientations. We then can apply a model-based version of the estimator  $\hat{S}_N$ . The estimator is identical to the one given by (2.15), but now the particle  $X$  is assumed to be random and the section plane  $L_2$  fixed. For all random nuclei  $X$  we assume that  $\mathbb{E}S(X) = \bar{S}$ , where the expectation is with respect to the random particle process and  $\bar{S}$  is the average surface area of the typical particle.  $\bar{S}$  is the target quantity. The unbiasedness of the model-based version of the estimator then follows from the unbiasedness of the design-based estimator and the strong law of large numbers.

## 4.2 Materials and preparation methods

From ten small pieces of tissue biopsies sized  $1\text{ mm}^3$  to  $2\text{ mm}^3$ , one piece was randomly chosen, embedded in glycol methacrylate (GMA) for preparation of a plastic block. Two  $40\text{ }\mu\text{m}$ -thick plastic sections were serially sectioned on a microtome (HM355, Microm, DMK & Michelsen, Denmark) and stained with hematoxylin and eosin stain for light microscopy. Data-acquisition was performed using a stereological microscopy system (BX-51 microscope, Olympus, Denmark) equipped with the NewCast Version 4.1 software package (Visiopharm, Hoersholm, Denmark) and mounted with a digital camera (Olympus DP72) on top of the microscope to project live views of tissue sections on a monitor. The  $z$ -axis of the microscope was monitored with a Heidenhain electronic microcator and the  $x$ - $y$  position was monitored by motorized stage system (ProscanTM, Prior Scientific Instruments Ltd., Cambridge, U.K).

The image of the tumour region was captured and delineated by navigator tool of NewCast with  $4\times$  objective. An unbiased counting frame ( $X \times Y$ ,  $15 \times 13 = 195\text{ }\mu\text{m}^2$ ) was used to sample randomly a total of  $n = 51$  nuclei. The red and green “rectangle” on the images in Figure 6 is a counting frame. The images of nuclei were captured by optical disector with an oil objective ( $100\times$ , NA: 1.25).

As already mentioned in Section 2, any point in the nucleus can serve as reference point as long as it is easily identifiable in any section direction. We chose the nucleoli in the selected nuclei as reference points. If a nucleus contained several nucleoli, we chose one out of all the admissible nucleoli (with respect to the counting frame) with uniform probability. In order to decrease the variance of the estimator systematic sampling was applied in the tissue by using a counting frame with a step-length of  $400\text{ }\mu\text{m}$  in both  $x$  and  $y$  direction. This assures that the nuclei can be assumed to have independent orientations. In other words, we can assume that the  $n$  sampled nuclei are realizations of independent, isotropic random particles  $X_1, X_2, \dots, X_n$  satisfying  $\mathbb{E}S(X_i) = \bar{S}$  for  $i = 1, \dots, n$ .

### 4.3 Implementation of the Morse type surface area estimator

The estimator  $\hat{S}_N$  was used to estimate the surface areas of the sampled nuclei. In Section 3.4.1 we recommended to use  $N = 4$ , that is to find tangents for four directions in the section plane ( $N = 2$  when the object of interest resembles an ellipsoid), to obtain a good precision of the surface area. In the present implementation we used one, two and four directions in the section plane in order to be able to compare the performance of the estimators. The estimation was done using an expert-assisted procedure which was implemented using `Matlab`. It is based on a program derived in Kallemose (2012) for estimating the perimeter of planar geometric structures. The interactive software is available at the home page `home.imf.au.dk/olofth`. We give a brief description of it in the following.

The present implementation is an off-line expert-assisted procedure, where the sampled section profiles – microscopy images in JPEG format – are used as input. The user, preferably a clinical expert, can choose the number  $N$  of directions that are to be sampled. As previously stated we recommend using  $N = 4$  ( $N = 2$  for ellipsoids) but we describe the procedure for an arbitrary  $N$ . Microscopy images are often stored with a scale, see for instance Figure 5 to the left. If this is the case, the user should indicate the scale before running the procedure. If a scale is reported, the calculated surface area estimate is given on this scale, otherwise it is given in pixel units. We assume in the following that a scale is given.

When the procedure is run, the image pops up with a short explanation of what the user should do next. The measurement procedure consists of three steps which are illustrated in Figure 6.

**Scale** The user left clicks on the endpoints of the scale, see the image in the upper left corner in Figure 6. This tells the program that the length of the red “measure” line segment is 10.706  $\mu\text{m}$ .

**Reference point** The user left clicks on the reference point, see the image in the upper right corner in Figure 6.

**Tangents** The computer generates a random unit vector  $U$  in the section plane including an angle  $\alpha_0$  with the  $x$ -axis that is uniform in  $[0, \pi/N)$ . In the image in the lower left corner in Figure 6, this vector is indicated as a short black line segment attached to the red line orthogonal to  $U$ , passing through the position of the mouse cursor. The line can be translated using the mouse. The user now marks all tangent positions of this line to the section profile, clicking left whenever a positive tangent is located, and clicking right for every negative tangent. If  $N > 1$  this step is repeated with all  $N - 1$  lines with associated angles  $\alpha_0 + l\pi/N$ ,  $l \in \{1, 2, \dots, N - 1\}$ .

When these three steps are completed and the user has found all tangents for the  $N$  sampled directions an estimate  $\hat{S}_N$  of the surface area is calculated. The number of tangents that need to be placed gives the complexity number  $C_{\hat{S}_N}$  defined in Section 3.4.2. It should be noted again that it certainly is more time consuming to place tangents at a given direction than just counting intersection points of the



profile boundary with a given line, as is the case for the invariator estimator. But for an experienced user of the software, the Morse type estimator  $\hat{S}_N$  can be obtained very efficiently and should be preferred to the alternative invariator grid estimator  $\hat{S}_{\text{grid}}$ .

As biological images are blurred, there might be tangents that appear to pass through a point of inflection of the boundary (although this event has probability zero with idealized mathematical particles). The easiest solution is to ignore these tangent lines, and neither left- nor right click at these positions. This is in correspondence with the theory, as a sweeping line passing through this tangent will neither increase nor decrease the associated Euler characteristic. It can also happen in practice, that the sweeping line appears to be tangent to the section profile at more than one point for certain critical values. As the images are blurred, it could for example appear that the sweeping line enters more than one new connected component at a critical value or that it enters a new connected component and leaves a part of the section profile at the same critical value. These events also have probability zero with idealized mathematical particles. To deal with these tangents, we simply suggest making a left- or a right mouse click for every change, without translating the line.

## 4.4 Results

We used the expert-assisted procedure to obtain estimates of the surface area of the  $n = 51$  section profiles. We did this for  $N = 1, 2$  and  $4$ . An unbiased estimator for the average surface area of the nucleus  $X$  of a giant-cell glioblastoma is then given by

$$\hat{S}_N^{\text{ave}}(X) = \frac{1}{51} \sum_{i=1}^{51} \hat{S}_N(X_i \cap L_2),$$

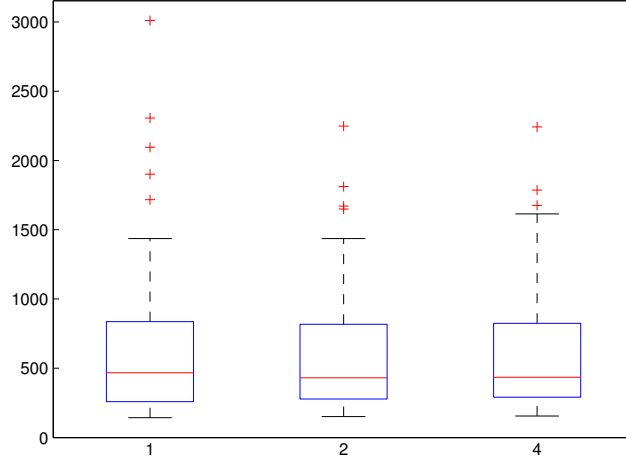
where  $L_2$  is a fixed plane and  $\hat{S}_N$  is the Morse-type estimator given by (2.15). Empirical estimates for the average surface areas can be found in Table 5.

**Table 5:** Empirical estimates for the average surface area of the nucleus of a giant-cell glioblastoma, depending on the number  $N$  of systematic directions used in the section plane.

	$N = 1$	$N = 2$	$N = 4$
$\hat{S}_N^{\text{ave}}(X)$	675.0 $\mu\text{m}^2$	597.8 $\mu\text{m}^2$	599.3 $\mu\text{m}^2$

In Figure 7 boxplots for the surface area estimates of the nuclei for  $N = 1, 2$  and  $4$ , respectively, can be found. The red lines in Figure 7 are the medians, the bottom and the top of the boxes represent the lower quartiles (25th percentiles) and the upper quartiles (75th percentiles), respectively, and the red “pluses” are outliers (more than 1.5 IQR from the upper quartiles). The boxplots in Figure 7 do not suggest that sampling with a higher number  $N$  of directions gives a more precise estimator. This seemingly counter-intuitive result can be explained by looking at the different sources of variation of  $\hat{S}_N^{\text{ave}}(X)$ . The variance of  $\hat{S}_N^{\text{ave}}(X)$  can be decomposed





**Figure 7:** Boxplots for the surface area estimates of the 51 nuclei for  $N = 1, 2$  and  $4$ , respectively.

into the true population variance, that is the variation in surface area among the nuclei, and the mean variance due to the estimation procedure within each nucleus. The variance due to the estimation procedure within each nucleus can then again be decomposed into  $V_{\text{plane}}$  and  $V_{\text{orient}}$ , as shown in Proposition 3.1. The true population variance is unknown and as the data only consist of one section profile for each sampled nuclei we can not assess  $V_{\text{plane}}$ . For a given nucleus,  $V_{\text{orient}}$  can be estimated from the section profile. This was done for each of the estimators in Table 5 for a few of the section profiles. A large decrease in  $V_{\text{orient}}$  was observed for each of the studied nuclei when  $N$  was increased. This decrease in  $V_{\text{orient}}$  was much more pronounced when  $N$  was increased from one to two than when it was increased from two to four. However, this variance contribution is very small compared to the sum of the true population variance and the mean variance due to choosing the plane within each nucleus. This explains why there is no noticeable precision gain in Figure 7 when  $N$  is increased. In the present application the choice  $N = 1$  led to an extreme outlier and  $V_{\text{orient}}$  that is much larger than when  $N = 2$ . As a conclusion, this example suggests that the choice  $N = 2$  is preferable in situations where a relatively large population variance is present. The workload for the choice  $N = 2$  is not very high.

## 5 Discussion

### 5.1 Account of main results

We have presented the Morse type surface area estimator and an expert-assisted protocol to apply it in practice. The Morse type estimator is a generalization of the surface area estimator for convex objects given by Cruz-Orive (2005, Eq. (3.2)) (see also (2.16) with  $N = 2$ ) to non-convex objects.

The surfactor is an alternative surface area estimator. The surfactor does not only require measuring distances in an IR section plane but also angles of the section profile with certain test rays. It was implied by Cruz-Orive (2005) that this requirement to perform angle measurements might be difficult in practice and yield

inaccurate estimates due to the singularity in its representation. In light of this belief and the simplicity of the estimator (2.16) it was claimed by Cruz-Orive (2005) that (2.16) was the only reasonably efficient surface area estimator available at that time.

As mentioned in Section 3.3, the surfactor was compared to the one-ray pivotal estimator, (2.17) with  $N = 1$ , by Cruz-Orive (2008). This variance comparison was extended to ellipsoidal particles by Dvořák and Jensen (2013). The results obtained in Dvořák and Jensen (2013) show that the surfactor performs better than previously thought. In a simulation study involving ellipsoids, the surfactor was shown to be neither much affected by the singularity in its representation nor by inaccuracies in the necessary angle measurements. Dvořák and Jensen (2013) also showed that the surfactor with two orthogonal directions (four rays) sampled in the section plane needs twice as much workload to obtain similar precision as (2.16) with  $N = 2$ . Hence, although the surfactor is a better competitor to (2.16) than previously expected, (2.16) is more efficient and should be preferred when the object of interest is convex. This recommendation can be transferred to the Morse type surface area estimator, which we showed to be the preferred surface area estimator available.

We have shown in Section 3.1 that the variance of the invariator estimator  $\hat{S}_{\text{inv}}$  decomposes into three parts  $V_{\text{plane}}$ ,  $V_{\text{orient}}$  and  $V_{\text{dist}}$  according to the randomization of the section plane, and the orientation and position of the test line, respectively. Among all estimators considered in the present paper, the generalized flower estimator has the lowest variance, as this variance coincides with  $V_{\text{plane}}$ . However, this estimator requires exact knowledge of the boundary of the section profile. We therefore also considered estimators derived from  $\hat{S}_{\text{inv}}$  that still have a reasonable workload but a better variance than  $\hat{S}_{\text{inv}}$ . The detailed analysis of the three variance contributions in Section 3.2 in the case of ellipsoidal particles showed that  $V_{\text{dist}}$  is by far the largest variance contribution. We therefore suggested the Morse type surface area estimator  $\hat{S}_N$  for which this variance contribution vanishes. By choosing the number  $N$  of systematic random test directions, also the variance contribution due to the randomization of the direction can be decreased. However, we have seen that only small values of  $N$  have to be considered, and that values  $N = 2$  or  $N = 4$  are theoretically advisable depending on the regularity of the objects. This also is in agreement with Figure 2 that shows that  $V_{\text{orient}}$  is a relatively small variance contribution (in the case of ellipsoids).

The application example of giant-cell glioblastoma showed that other sources of variation, such as the population variance or the uncertainty of measurements due to blurred or ambiguous images can be of the same order or larger than the theoretical variance contribution of the estimator. This indicates that the choice  $N = 2$  appears to be appropriate in real-world applications.

## 5.2 Automatic and semi-automatic estimation of surface area

Inspired by Cruz-Orive (2012) it was shown by Thórisdóttir and Kiderlen (Submitted, 2013) that the generalized flower estimator  $\hat{S}_{\text{flo}}$  given by (2.12) and the integrated surfactor in Cruz-Orive (2012, Eq. (24)) coincide. The generalized flower

estimator is the optimal estimator with respect to variance (among all the estimators considered in this paper) but as mentioned in Section 3.3 it can typically not be calculated in practice if the section profile is not a polygon. However, if it is possible to identify the boundary of a given section profile by automated segmentation, the boundary can be approximated by a polygon. Using this approximation the generalized flower estimator  $\hat{S}_{\text{flo}}$ , with the true section profile  $Y = X \cap L_2$  replaced by its estimate, can be obtained using Thórisdóttir and Kiderlen (Submitted, 2013, Corollary 10), as mentioned in Section 2.2.3. If the segmentation is flawless, this automatic estimator is unbiased. However, if it is of poor quality, the estimator can be heavily biased. To deal with this problem a semi-automatic procedure for estimating surface area was proposed in Dvořák and Jensen (2013), based on a similar approach for volume estimation in Hansen et al. (2011). In the semi-automatic procedure, a clinical expert supervises the process and determines if the automated segmentation of the boundary of a given section profile is satisfactory or not. If it is determined to be satisfactory the surface area is obtained using  $\hat{S}_{\text{flo}}$  with  $Y$  being the estimate obtained by the automated segmentation. If it is unsatisfactory, the clinical expert intervenes and performs the necessary measurements in the section plane. Dvořák and Jensen (2013) suggested to use (2.16) in the case where all particles are convex. We suggest to use the Morse-type estimator, which does not require any convexity assumption and reduces to (2.16) when the particles are convex. This semi-automatic procedure can reduce the workload substantially.

## Acknowledgement

This research was supported by Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation. The authors want to thank Benedicte Parm Ulhøi for providing the tissue biopsies and expert knowledge on histological aspects of tumours. The authors also want to thank Thomas Kallemose for writing the program which the expert-assisted procedure is based on.

## A Appendix: Variance decomposition for ellipsoids

The following proposition gives the different variance contributions in Proposition 3.1 more explicitly when  $X$  is a three-dimensional ellipsoid containing  $O$ . A proof can be found in Thórisdóttir (2010, Theorem 25). As the reference set  $RB_3$  is trivially invariant under all rotations at the origin, and as the sampling procedure is rotation invariant, we may assume without loss of generality that the ellipsoid's main axes are parallel to the standard coordinate axes.

**Proposition A.1.** *Let  $E_3 \subseteq RB_3$  be a non-degenerate ellipsoid given by*

$$E_3 = \left\{ x \in \mathbb{R}^3 \mid \sum_{i=1}^3 \left( \frac{x_i - z_i}{a_i} \right)^2 \leq 1 \right\}, \quad (\text{A.1})$$

where  $a_1, a_2, a_3 > 0$  and  $(z_1, z_2, z_3) \in \mathbb{R}^3$ . If  $O \in E_3$ , the variance of  $\hat{S}_{\text{inv}}$  given by

(2.4), can be expressed as

$$\text{Var}(\hat{S}_{\text{inv}}) = V_{\text{dist}} + V_{\text{orient}} + V_{\text{plane}},$$

where

$$\begin{aligned} V_{\text{dist}} = & 4\pi \int_0^{2\pi} \int_0^{\pi/2} \left( R^2 \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + (z'_1)^2 + (z'_2)^2 \right) - \left( \frac{3}{4\lambda_1^2} + \frac{1}{2\lambda_1\lambda_2} + \frac{3}{4\lambda_2^2} \right. \right. \\ & + \frac{3}{4}((z'_1)^4 + 2(z'_1)^2(z'_2)^2 + (z'_2)^4) + ((z'_1)u'_1 + (z'_2)u'_2)^2 \left( \frac{9}{2\lambda_1} + \frac{3}{2\lambda_2} \right) \\ & \left. \left. + (z'_1 u''_1 + z'_2 u''_2)^2 \left( \frac{3}{2\lambda_1} + \frac{9}{2\lambda_2} \right) \right) \right) \sin \theta d\theta d\phi, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} V_{\text{orient}} = & 4\pi \int_0^{2\pi} \int_0^{\pi/2} \left( \frac{3}{4\lambda_1^2} + \frac{1}{2\lambda_1\lambda_2} + \frac{3}{4\lambda_2^2} + \frac{3}{4}((z'_1)^4 + 2(z'_1)^2(z'_2)^2 + (z'_2)^4) \right. \\ & + (z'_1 u'_1 + z'_2 u'_2)^2 \left( \frac{9}{2\lambda_1} + \frac{3}{2\lambda_2} \right) + (z'_1 u''_1 + z'_2 u''_2)^2 \left( \frac{3}{2\lambda_1} + \frac{9}{2\lambda_2} \right) \\ & \left. - \frac{1}{2} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + (z'_1)^2 + (z'_2)^2 \right)^2 \right) \sin \theta d\theta d\phi, \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} V_{\text{plane}} = & 2\pi \int_0^{2\pi} \int_0^{\pi/2} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + (z'_1)^2 + (z'_2)^2 \right)^2 \sin \theta d\theta d\phi \\ & - \left( \int_0^{2\pi} \int_0^{\pi/2} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + (z'_1)^2 + (z'_2)^2 \right) \sin \theta d\theta d\phi \right)^2. \end{aligned} \quad (\text{A.4})$$

Here  $\lambda_1^{-1/2}$  and  $\lambda_2^{-1/2}$  are the lengths of the semiaxes of the ellipse  $E_3 \cap L_2$ ,  $u' = (u'_1, u'_2)^t$ ,  $u'' = (u''_1, u''_2)^t$  its corresponding principal axes and  $(z'_1, z'_2)^t$  its center, all written with respect to a suitably chosen orthonormal basis of the IR plane  $L_2$ .

Although the decomposition formulae for ellipsoids are quite complicated it is not difficult to derive a theoretical formula for the total variance of  $\hat{S}_{\text{inv}}$  when  $X$  is an ellipsoid. When  $X$  is a convex body,  $\chi(X \cap L_2 \cap F_z)$  is a Bernoulli random variable and hence using the unbiasedness of  $\hat{S}_{\text{inv}}$

$$\text{Var}(\hat{S}_{\text{inv}}) = S(\partial X)(S(\partial(RB_3)) - S(\partial X)). \quad (\text{A.5})$$

As a side note, this immediately gives the coefficient of variation of the estimator, which is

$$CV(\hat{S}_{\text{inv}}) = \left( \frac{4\pi R^2}{S(\partial X)} - 1 \right)^{1/2},$$

in accordance to Cruz-Orive (2008, Eq. (18)), as well as the total variance (3.8) of the estimator when  $X$  is a ball.

Let now  $E_3$  be a non-degenerate ellipsoid given by (A.1). Assume without loss of generality that  $a_1 \geq a_2 \geq a_3$ . Then the surface area of  $E_3$  is given by

$$S(\partial E_3) = 2\pi \left( a_3^2 + \frac{a_2 a_3^2}{\sqrt{a_1^2 - a_3^2}} F(\sigma, m) + a_2 \sqrt{a_1^2 - a_3^2} E(\sigma, m) \right),$$

where

$$\sigma = \arcsin \sqrt{\frac{a_1^2 - a_3^2}{a_1^2}}, \quad m = \frac{a_1^2(a_2^2 - a_3^2)}{a_2^2(a_1^2 - a_3^2)}$$

and  $F(\cdot, \cdot)$  and  $E(\cdot, \cdot)$  are the incomplete elliptic integrals of the first kind and of the second kind, respectively; see Eagle (1958, (12) on p. 281). Inserting this into (A.5), we find

$$\begin{aligned} \text{Var}(\hat{S}_{\text{inv}}) = 4\pi^2 & \left( 2R^2 \left( a_3^2 + \frac{a_2 a_3^2}{\sqrt{a_1^2 - a_3^2}} F(\sigma, m) + a_2 \sqrt{a_1^2 - a_3^2} E(\sigma, m) \right) \right. \\ & \left. - \left( \left( a_3^2 + \frac{a_2 a_3^2}{\sqrt{a_1^2 - a_3^2}} F(\sigma, m) + a_2 \sqrt{a_1^2 - a_3^2} E(\sigma, m) \right) \right)^2 \right). \end{aligned} \quad (\text{A.6})$$

For an oblate ellipsoid (that is when  $a_1 = a_2 > a_3$ ), the surface area has an explicit form and the total variance becomes

$$\begin{aligned} \text{Var}(\hat{S}_{\text{inv}}) = 4\pi^2 & \left( 2R^2 \left( a_1^2 + \frac{a_1 a_3^2}{\sqrt{a_1^2 - a_3^2}} \log \left( \frac{\sqrt{a_1^2 - a_3^2} + a_1}{a_3} \right) \right) \right. \\ & \left. - \left( a_1^2 + \frac{a_1 a_3^2}{\sqrt{a_1^2 - a_3^2}} \log \left( \frac{\sqrt{a_1^2 - a_3^2} + a_1}{a_3} \right) \right)^2 \right); \end{aligned}$$

see Thórisdóttir (2010, p. 34) for a derivation. When  $X$  is an oblate or a prolate ellipsoid, explicit analytic expressions for  $V_{\text{plane}}$  can be found in Dvořák and Jensen (2013, Eq. (2.5) and (2.6)).

## References

- A.J. Baddeley. Stochastic geometry and image analysis. In *Proceedings of the CWI symposium on mathematics and computer science*. CWI Monographs, 1984.
- C.G. Bowsher and P.S. Swain. Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences*, 109(20): E1320–E1328, 2012.
- L.M. Cruz-Orive. A new stereological principle for test lines in three-dimensional space. *Journal of Microscopy*, 219(1):18–28, 2005.
- L.M. Cruz-Orive. Comparative precision of the pivotal estimators of particle size. *Image Analysis & Stereology*, 27(1):17–22, 2008.
- L.M. Cruz-Orive. Flowers and wedges for the stereology of particles. *Journal of Microscopy*, 243(1):86–102, 2011.
- L.M. Cruz-Orive. Uniqueness properties of the invariator, leading to simple computations. *Image Analysis & Stereology*, 31(2):89–98, 2012.
- L.M. Cruz-Orive, M.L. Ramos-Herrera, and E. Artacho-Pérula. Stereology of isolated objects with the invariator. *Journal of Microscopy*, 240(2):94–110, 2010.
- R.T. DeHoff. The quantitative estimation of mean surface curvature. *Transactions of the Metallurgical Society of AIME*, 239:617–621, 1967.

- J. Dvořák and E.B.V. Jensen. On semiautomatic estimation of surface area. *Journal of Microscopy*, 250(2):142–157, 2013.
- A. Eagle. *The Elliptic Functions as they should be: An account, with applications, of the functions in a new canonical form*. Galloway and Porter, Cambridge, 1958.
- X. Gual-Arnau and L.M. Cruz-Orive. Systematic sampling on the circle and on the sphere. *Advances in Applied Probability*, 32(3):628–647, 2000.
- H.J.G. Gundersen. The nucleator. *Journal of Microscopy*, 151(1):3–21, 1988.
- L.V. Hansen, J.R. Nyengaard, J.B. Andersen, and E.B.V. Jensen. The semi-automatic nucleator. *Journal of Microscopy*, 242(2):206–215, 2011.
- E.B.V. Jensen. *Local Stereology*. World Scientific, Singapore, 1998.
- E.B.V. Jensen and H.J.G. Gundersen. Stereological estimation of surface area of arbitrary particles. *Acta Stereologica*, 6:25–30, 1987.
- T. Kallemose. Application of stereological projection formulae for perimeter estimation. Master’s thesis, Aarhus University, 2012.
- J.W. Milnor. *Morse Theory*. Princeton University Press, Princeton, 1963.
- H. Ohgaki, A. Peraud, Y. Nakazato, K. Watanabe, and A. Von Deimling. Giant cell glioblastoma. In *World Health Organization classification of tumours. Pathology and genetics of tumours of the nervous system*, pages 40–41. IARC Press, Lyon, 2000.
- Ó. Thórisdóttir. Variance for surface area estimation based on the invariator. Master’s thesis, Aarhus University, 2010.
- Ó. Thórisdóttir and M. Kiderlen. The invariator principle in convex geometry. Submitted, 2013.