

# CHARACTERIZATION OF MAMMARY GLAND TISSUE USING JOINT ESTIMATORS OF MINKOWSKI FUNCTIONALS

TORSTEN MATTFELDT<sup>1</sup>, DANIEL MESCHENMOSER<sup>2</sup>, URSA PANTLE<sup>2</sup> AND VOLKER SCHMIDT<sup>2</sup>

<sup>1</sup>Institute of Pathology, University of Ulm, 89081 Ulm, Germany, <sup>2</sup>Institute of Stochastics, University of Ulm, 89081 Ulm, Germany

e-mail: torsten.mattfeldt@medizin.uni-ulm.de, daniel.meschenmoser@uni-ulm.de, ursula.pantle@uni-ulm.de, volker.schmidt@uni-ulm.de

(Accepted March 13, 2007)

## ABSTRACT

A theoretical approach to estimate the Minkowski functionals, *i.e.*, area fraction, specific boundary length and specific Euler number in 2D, and their asymptotic covariance matrix proposed by Spodarev and Schmidt (2005) and Pantle *et al.* (2006a;b) is applied to real image data. These two-dimensional images show mammary gland tissue and should be classified automatically as tumor-free or mammary cancer, respectively. The estimation procedure is illustrated step-by-step and the calculations are described in detail. To reduce dependencies from chosen parameters, a least-squares approach is considered as recommended by Klenk *et al.* (2006). Emphasis is placed on the detailed description of the estimation procedure and the application of the theory to real image data.

Keywords: asymptotic covariance matrix, breast cancer, mammary carcinoma, mammary gland tissue, Minkowski functionals, random closed set, specific intrinsic volumes.

## INTRODUCTION

Breast cancer is the most frequent malignant tumor in women. In routine diagnostics, it is usual to perform a histopathological grading, which is based on a three-tiered scheme with grades I, II, and III (Ellston and Ellis, 1991; Mattfeldt *et al.*, 2004). As the reproducibility of tumor grading is unknown for individual cases, many attempts have been made to arrive at an objective and quantitative grading of tumor structure. Let us consider here the tumor texture, which reflects the degree of differentiation of the tumor. The tissue may be conceived as a random set with different phases, which all possess a positive volume fraction. This means consideration of the tumor tissue as a volume process (Mattfeldt and Fleischer, 2005; Mattfeldt *et al.*, 2006). It consists of three phases: tumor cells, stroma and lumina, which altogether account for 100% of the tissue.

In diagnostic pathology we deal with histological sections, *i.e.*, very thin slices, onto which windows, usually of rectangular or quadratic shape, are placed for evaluation under microscopical view. Hence we are faced in practice with random closed sets in 2D, which may be quantified in terms of the three Minkowski functionals:  $A_A$ , the mean area of the interesting phase per unit reference area (area fraction);  $B_A$ , the mean boundary length of the interesting phase per unit reference area; and  $\chi_A$ , the mean Euler number of the interesting phase per unit reference area. Notably

all these quantities have a stereological interpretation, hence they can be used for the estimation of stereological model parameters:

$$V_V = A_A \quad (1)$$

$$S_V = \frac{4}{\pi} B_A \quad (2)$$

$$M_V = 2\pi \chi_A \quad (3)$$

where  $V_V$  is the volume fraction,  $S_V$  is the mean surface area per unit reference volume, and  $M_V$  is the mean curvature density (Stoyan *et al.*, 1995). Eqs. 1–3 are all fundamental stereological formulae and hold for random closed sets under the conditions of isotropy and stationarity for arbitrary sections. Recently a new approach has been developed which allows a joint estimation of all three Minkowski functionals for a given image (Schmidt and Spodarev, 2005; Spodarev and Schmidt, 2005). It provides not only point estimates of  $V_V$ ,  $S_V$ , and  $M_V$ , but also estimates of their asymptotic variances and covariances. Up to now the aforementioned estimator has only been applied to simulated images, but not yet to real image material. As a first application to real images in a simple situation, we decided to compare mammary cancer tissue to normal (tumor-free) mammary tissue, see also earlier publications of our group (Mattfeldt *et al.*, 1993; 1996; 2000; Mattfeldt and Stoyan, 2000).

The paper is organized as follows. In Section ‘Mathematical Methods’ the notation used throughout

the paper is introduced. The specific intrinsic volumes are defined by means of Steiner's formula. The method given in Spodarev and Schmidt (2005) and Pantle *et al.* (2006a;b) to estimate these quantities and their asymptotic covariance matrix is described where some technical details are omitted. Section 'Application to image data' deals with the estimation of specific intrinsic volumes from real image data showing mammary tissue. The procedure for the two-dimensional case is described in detail. With the estimated quantities a statistical test is considered to classify an image as tumor-free or as mammary cancer, respectively. The paper ends with a discussion and an outlook to further projects.

## MATHEMATICAL METHODS

First we introduce some notation. For some fixed  $d \geq 2$ , denote the family of convex bodies, *i.e.*, compact convex sets, in  $\mathbb{R}^d$  by  $\mathcal{K}$  and let  $\mathcal{R} = \{K \subset \mathbb{R}^d : K = \cup_{i=1}^n K_i, K_i \in \mathcal{K}, n \in \mathbb{N}\}$  be the convex ring, *i.e.*, the family of all polyconvex sets in  $\mathbb{R}^d$ . By  $\mathcal{S} = \{M \subset \mathbb{R}^d : M \cap K \in \mathcal{R} \forall K \in \mathcal{K}\}$  we denote the extended convex ring. Then it holds  $\mathcal{K} \subset \mathcal{R} \subset \mathcal{S}$ . Let  $B_r(x)$  be the closed ball in  $\mathbb{R}^d$  with radius  $r > 0$  centered at  $x$  and let  $o \in \mathbb{R}^d$  be the origin. Further,  $k_j$  denotes the volume of the  $j$ -dimensional unit ball for  $j = 0, \dots, d$ . For two sets  $A, B \subset \mathbb{R}^d$  the Minkowski sum  $A \oplus B$  and the Minkowski difference  $A \ominus B$  are defined by  $A \oplus B = \{a + b : a \in A, b \in B\}$  and  $A \ominus B = \{x \in \mathbb{R}^d : \check{B} + x \subset A\}$ , respectively, where  $\check{B} = \{x \in \mathbb{R}^d : -x \in B\}$  denotes the set  $B$  reflected at the origin.

For convex bodies  $K \in \mathcal{K}$  it can be proven that there exist  $d + 1$  functionals  $V_j : \mathcal{K} \rightarrow [0, \infty)$  for  $j = 0, \dots, d$ , such that the volume of the so called parallel body  $K \oplus B_r(o)$  for  $r > 0$  is given by Steiner's formula

$$|K \oplus B_r(o)| = \sum_{j=0}^d r^{d-j} k_{d-j} V_j(K). \quad (4)$$

A proof of this formula can be found, *e.g.*, in chapter 2 of Schneider and Weil, 1992. The functionals  $V_j$  are called intrinsic volumes. They are related to the Minkowski functionals  $W_j : \mathcal{K} \rightarrow [0, \infty)$  in the following way:  $W_j(K) = k_j / \binom{d}{j} V_{d-j}(K)$  for all  $j = 0, \dots, d$ . The intrinsic volumes are not restricted to convex bodies. There is a unique additive extension to the convex ring  $\mathcal{R}$  given by the inclusion-exclusion-formula. For any polyconvex set  $K \in \mathcal{R}$ , any  $n \in \mathbb{N}$  and any convex bodies  $K_1, \dots, K_n \in \mathcal{K}$  with  $K = K_1 \cup \dots \cup K_n$  it holds

$$V_j(K) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} V_j(K_{i_1} \cap \dots \cap K_{i_k}) \quad (5)$$

for  $j = 0, \dots, d$ . Notice that the value of  $V_j(K)$  does not depend on the particular representation of  $K$  as the union of convex sets  $K_i$ . The proof of existence and uniqueness of this extension can be found in Schneider, 1993. The formula itself can be shown by induction using the fact that the intrinsic volumes are additive, *i.e.*,  $V_j(\emptyset) = 0$  and for all  $K_1, K_2 \in \mathcal{K}$  with  $K_1 \cup K_2 \in \mathcal{K}$  it holds  $V_j(K_1 \cup K_2) = V_j(K_1) + V_j(K_2) - V_j(K_1 \cap K_2)$ . Some of the intrinsic volumes have a nice geometric interpretation:  $V_d(K)$  is the usual volume of  $K$ ,  $dV_{d-1}(K)$  is the surface area of  $K$  and  $V_0(K)$  is the Euler-Poincaré characteristic of  $K$ .

In the following let  $\Xi$  be a stationary random closed set in  $\mathbb{R}^d$  with values in the extended convex ring  $\mathcal{S}$  almost surely. Let  $\{W_n\}$  be a monotonically increasing sequence of compact convex observation windows

$$W_n = nW \quad (6)$$

with

$$W \in \mathcal{K}, |W| > 0 \text{ and } o \in \text{int}(W). \quad (7)$$

Under appropriate assumptions, the expectation  $\mathbb{E}V_j(\Xi \cap W_n)$  is well defined and the limit

$$\bar{V}_j(\Xi) = \lim_{n \rightarrow \infty} \frac{\mathbb{E}V_j(\Xi \cap W_n)}{|W_n|} \quad (8)$$

exists for all  $j = 0, \dots, d$ , see, *e.g.*, Schneider and Weil, 2000. The functionals  $\bar{V}_j(\Xi)$  are called specific intrinsic volumes of  $\Xi$ . In the two-dimensional case they are well-known under the notation  $A_A = \bar{V}_2$ ,  $B_A = 2\bar{V}_1$  and  $\chi_A = \bar{V}_0$  and are connected to the stereological model parameters  $V_V$ ,  $S_V$  and  $M_V$  by Eqs. 1–3.

To estimate the specific intrinsic volumes from a binary image we use a method developed in Spodarev and Schmidt (2005). It makes use of the local Euler-Poincaré characteristic, which is defined as the expected Euler number of  $\Xi$  in a neighborhood of a point  $x$ , *i.e.*, as  $\mathbb{E}V_0(\Xi \cap B_r(x))$ . It can be shown that for any  $r > 0$  and for any  $x \in \mathbb{R}^d$  it holds

$$\mathbb{E}V_0(\Xi \cap B_r(x)) = \sum_{j=0}^d r^{d-j} k_{d-j} \bar{V}_j(\Xi). \quad (9)$$

A proof of Eq. 9 can be found in chapter 5 of Schneider and Weil (2000), see also Spodarev and Schmidt (2005). Since this formula holds for any  $r > 0$  we can plug in  $d + 1$  pairwise different radii  $0 < r_0 < \dots < r_d$ , where we have to take care that the edge-corrected observation window  $W \ominus B_{r_j}(o)$  has positive volume for  $j = 0, \dots, d$ . Since the radii are numbered in ascending order this holds if  $|W \ominus B_{r_d}(o)| > 0$ . From

Eq. 9 we get the following system of  $d + 1$  linear equations

$$Av = y, \quad (10)$$

where

$$A = \begin{pmatrix} r_0^d k_d & r_0^{d-1} k_{d-1} & \cdots & r_0 k_1 & 1 \\ r_1^d k_d & r_1^{d-1} k_{d-1} & \cdots & r_1 k_1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_d^d k_d & r_d^{d-1} k_{d-1} & \cdots & r_d k_1 & 1 \end{pmatrix}, \quad (11)$$

$v = (\bar{V}_0(\Xi), \dots, \bar{V}_d(\Xi))^\top$  and  $y = (\mathbb{E}V_0(\Xi \cap B_{r_0}(0)), \dots, \mathbb{E}V_0(\Xi \cap B_{r_d}(0)))^\top$ . The matrix  $A$  is regular because the radii  $r_0, \dots, r_d$  are pairwise different and it can be computed without problems. With an appropriate estimator  $\hat{y}$  of  $y$  we now get an estimator  $\hat{v}$  of  $v$  by

$$\hat{v} = A^{-1}\hat{y}. \quad (12)$$

Since the local Euler characteristic, *i.e.*, the vector  $y$ , can be estimated from one single image we also can estimate the vector of specific intrinsic volumes, *i.e.*, the vector  $v$ , from one single image. To estimate the vector  $y$  of local Euler characteristics for different radii we consider the stationary random field  $Y_j = \{Y_j(x), x \in \mathbb{R}^d\}$  with

$$Y_j(x) = V_0(\Xi \cap B_{r_j}(x)) \quad (13)$$

for  $j = 0, \dots, d$ . The stationarity of  $Y_j$  follows directly from the stationarity of  $\Xi$ . An unbiased estimator  $\hat{y}_j$  of  $y_j = \mathbb{E}Y_j(o)$  is given by

$$\hat{y}_j = \int_{W \ominus B_{r_d}(o)} Y_j(x) \mu(dx), \quad (14)$$

where  $\mu$  is an arbitrary probability measure concentrated on the reduced observation window  $W \ominus B_{r_d}(o)$  to avoid edge effects. If  $\mu$  is the normalized Lebesgue measure, *i.e.*,

$$\mu(\cdot) = \frac{|\cdot \cap W \ominus B_{r_d}(o)|}{|W \ominus B_{r_d}(o)|}, \quad (15)$$

the estimator  $\hat{y}_j$  is given by

$$\hat{y}_j = \frac{1}{|W \ominus B_{r_d}(o)|} \int_{W \ominus B_{r_d}(o)} Y_j(x) dx. \quad (16)$$

To study the variance of the estimator  $\hat{v}$  of the specific intrinsic volumes we consider a sequence of observation windows  $\{W_n\}$ , which satisfies conditions Eq. 6 and Eq. 7. For each  $j =$

$0, \dots, d$  and for each  $n \in \mathbb{N}$  we can estimate  $y_j$  on  $W_n$  by

$$\hat{y}_{n,j} = \frac{1}{|W_n \ominus B_{r_d}(o)|} \int_{W_n \ominus B_{r_d}(o)} Y_j(x) dx. \quad (17)$$

Under appropriate assumptions, the covariances

$$\text{Cov}_{ij}(x) = \text{Cov}(Y_i(o), Y_j(x)) \quad (18)$$

exist for all  $i, j = 0, \dots, d$ , see, *e.g.*, Schmidt and Spodarev (2005). If the covariances are absolutely integrable and some further assumptions are fulfilled (*cf.* Pantle *et al.*, 2006a), then the random vector  $\sqrt{|W_n|}(\hat{y}_{0,n} - y_0, \dots, \hat{y}_{d,n} - y_d)^\top$  is asymptotically normally distributed with mean vector  $o$  and covariance matrix  $\Sigma = (\sigma_{ij})_{i,j=0}^d$  with  $\sigma_{ij} = \int_{\mathbb{R}^d} \text{Cov}_{ij}(x) dx$ . Therefore the random vector  $\sqrt{|W_n|}(\hat{v}_{0,n} - v_0, \dots, \hat{v}_{d,n} - v_d)^\top$  is also asymptotically normally distributed with zero mean vector and covariance matrix  $\Sigma_{\hat{v}} = A^{-1}\Sigma(A^{-1})^\top$ .

The values of these estimators  $\hat{v}_j$  of the specific intrinsic volumes depend heavily on the choice of the radii  $r_0, \dots, r_d$ , see also the discussion in Klenk *et al.* (2006). To reduce this dependence a least-squares approach is considered. Let  $0 < r_0 < \dots < r_{k-1}$  be  $k > d + 1$  pairwise different radii. Similar to Eq. 10 we get a system of  $k$  linear equations with the difference that the vector  $y$  is  $k$ -dimensional and that  $A$  is not a squared matrix any more because it has  $k$  rows and  $d + 1$  columns. Anyhow, the minimization problem

$$|\hat{y} - Av^*| = \min_{x \in \mathbb{R}^{d+1}} |\hat{y} - Ax| \quad (19)$$

has a unique solution given by

$$v^* = (A^\top A)^{-1} A^\top \hat{y}. \quad (20)$$

The estimator  $v^*$  of the vector of specific intrinsic volumes does not depend on the choice of radii as much as the estimator  $\hat{v}$  in Eq. 12. Furthermore, the random vector  $\sqrt{|W_n|}(v_{0,n}^* - v_0, \dots, v_{d,n}^* - v_d)^\top$  is asymptotically normally distributed with zero mean vector and covariance matrix  $\Sigma_{v^*} = \tilde{A}\Sigma\tilde{A}^\top$  where  $\tilde{A} = (A^\top A)^{-1}A^\top$ .

The asymptotic covariance matrix  $\Sigma$  can be estimated from the observation of the stationary random fields  $Y_j$  defined in Eq. 13. For each  $n \in \mathbb{N}$  let  $W_{n0}, \dots, W_{nk-1} \subset W_n$ . Let  $\{U_{nij}\}$  be a monotonously increasing sequence of bounded sets with  $U_{nij} \subset W_{ni} \oplus \check{W}_{nj}$  and  $|U_{nij}| > 0$  for all  $n \in \mathbb{N}$  and for all  $i, j = 0, \dots, k - 1$ . Additionally let  $\lim_{n \rightarrow \infty} U_{nij} =$

$\text{supp}(\text{Cov}_{ij})$  and let the sets  $U_{nij}$  grow smaller in comparison to  $W_{nij}$ , *i.e.*,

$$\lim_{n \rightarrow \infty} \frac{|U_{nij}|^2}{|W_{nij}|} = 0. \quad (21)$$

Furthermore, assume that

$$\lim_{n \rightarrow \infty} \frac{\min_{x \in U_{nij}} |W_{nij} \cap (W_{nij} + x)|}{|W_{nij}|} = 1. \quad (22)$$

For each  $n \in \mathbb{N}$  and  $i, j = 0, \dots, k-1$  we consider the estimator

$$\hat{\sigma}_{nij} = \int_{U_{nij}} \widehat{\text{Cov}}_{nij}(x) \frac{|W_{nij} \cap (W_{nij} - x)|}{|W_{nij}|} dx \quad (23)$$

of  $\sigma_{ij}$  with

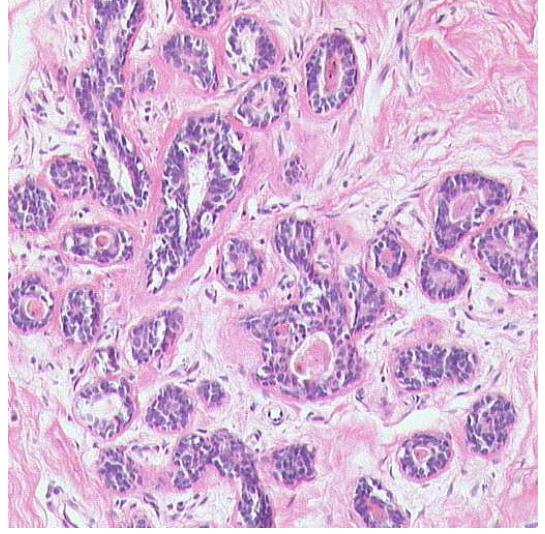
$$\widehat{\text{Cov}}_{nij}(x) = \int_{W_{nij} \cap (W_{nij} - x)} \frac{Y_i(y) Y_j(x+y)}{|W_{nij} \cap (W_{nij} - x)|} dy - \hat{y}_{i,n} \hat{y}_{j,n}. \quad (24)$$

This sequence  $\hat{\Sigma}_n = (\hat{\sigma}_{nij})$  of estimators of  $\Sigma$  is asymptotically unbiased, *i.e.*,  $\lim_{n \rightarrow \infty} \|\mathbb{E} \hat{\Sigma}_n - \Sigma\| = 0$ , see, *e.g.*, Schmidt and Spodarev (2005); Pantle *et al.* (2006b), where  $\|\Sigma\| = \sqrt{\sum_{i,j=0}^d \sigma_{ij}^2}$  denotes the matrix norm. Under additional integrability conditions,  $\hat{\Sigma}_n$  is  $L^2$  consistent for  $\Sigma$ , *i.e.*, it holds  $\lim_{n \rightarrow \infty} \mathbb{E} \|\hat{\Sigma}_n - \Sigma\|^2 = 0$ . If  $\Xi$  is the Boolean model with primary grain  $M_0$  for example, the integrability conditions are fulfilled if  $\mathbb{E}|M_0 \oplus B_{r_d}|^2 < \infty$ . From the estimators  $\hat{\Sigma}_n$  described in Eq. 23 and Eq. 24 we get a sequence of estimators  $\hat{\Sigma}_n^*$  of the asymptotic covariance matrix  $\Sigma_{v^*}$  by

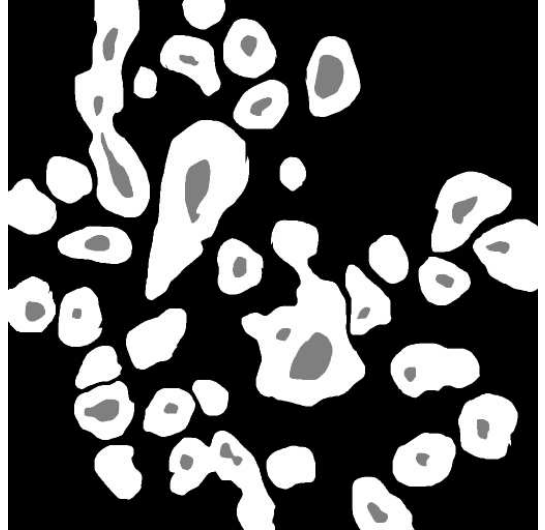
$$\hat{\Sigma}_n^* = \tilde{A} \hat{\Sigma}_n \tilde{A}^\top \text{ with } \tilde{A} = \left( A^\top A \right)^{-1} A^\top. \quad (25)$$

## APPLICATION TO IMAGE DATA

Now we are ready to apply the statistical approach explained in the previous section to estimate the specific intrinsic volumes of real image data showing mammary tissue. There were ten cases of ductal mammary cancer tissue and ten cases showing normal, *i.e.*, cancer-free, mammary tissue. From each case, a sample of  $3 \times 3 = 9$  contiguous quadratic images was evaluated, where the first image was selected at random. Each image had a size of  $510 \times 510$  pixels. The concatenation led to a large quadratic image with  $1530 \times 1530$  pixels, which is needed for the estimation of the asymptotic variances. This means that the final observation window is given by the rectangle



(a) Original image of tumour-free mammary tissue.

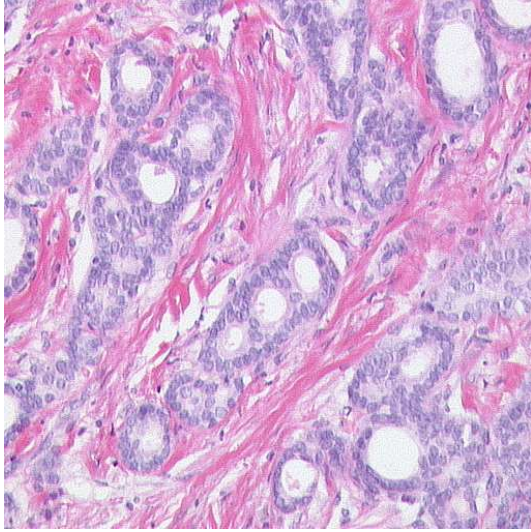


(b) Segmentation of Fig. 1(a) leads to this image, which contains three phases: white—epithelial cells, gray—lumen, black—stroma.

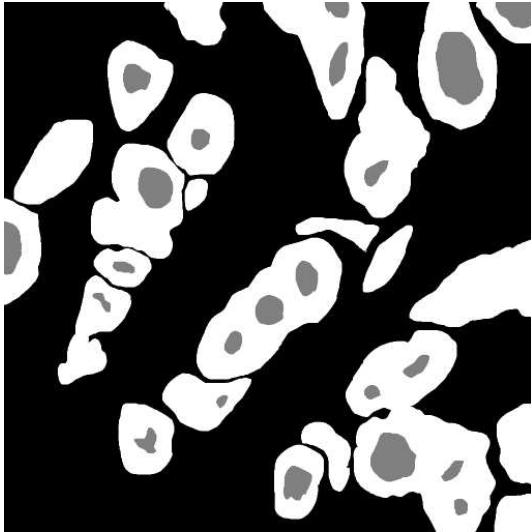
Fig. 1. *Tumour-free mammary tissue. Haematoxylin-Eosin stain (a) and segmented image (b), respectively. The edgelenh of the quadrat corresponds to 0.4 mm at the scale of the specimen at this magnification.*

$[0, 1529] \times [0, 1529]$ . Fig. 1a shows tumour-free mammary tissue and Fig. 2a shows invasive ductal mammary carcinoma. The edgelenh of 510 pixels corresponds to 0.4 mm at the scale of the specimen at this magnification, *i.e.*, Figs. 1a and 2a show only one of the nine contiguous images. The same images are shown in Figs. 1b and 2b, respectively, after interactive segmentation of stroma, epithelium and lumina. The

stroma is represented by the black phase, the grey phase forms the lumina and the white phase stands for the epithelium without the lumina, *i.e.*, the tumor cells.



(a) Original image of invasive ductal mammary carcinoma.



(b) Segmentation of Fig. 2a.

Fig. 2. Invasive ductal mammary carcinoma. In (a), stain and magnification identical with Fig. 1a. In (b), segmentation identical with Fig. 1b.

These three phase images were converted into three binary images by combining two phases. The foreground of the resulting images may consist of one of the three grey phases (*e.g.*, the white phase) or of the union of two of the three phases (*e.g.*, white and grey). We understand these binary images as realizations of stationary and isotropic random closed sets (*cf.*

Mattfeldt and Stoyan, 2000b). Since these images are two-dimensional the specific intrinsic volumes  $\bar{V}_0$ ,  $2\bar{V}_1$  and  $\bar{V}_2$  represent the mean Euler number per unit area, the mean boundary length per unit area, and the area fraction, respectively.

In the following, the procedure to calculate the least-squares estimator given in Eq. 20 is described step by step.

1. Choose the number  $k > 3$  and the values of the radii  $r_0 < \dots < r_{k-1}$ . We put  $k = 15$  and  $r_i = 4.2 + 1.3i$ ,  $i = 0, \dots, 14$ , following the recommendation in Klenk *et al.* (2006). In this particular case, the matrix  $A$  defined in Eq. 11 is given by

$$A = \begin{pmatrix} 17.64\pi & 8.4 & 1 \\ 30.25\pi & 11.0 & 1 \\ \vdots & \vdots & \vdots \\ 445.21\pi & 42.2 & 1 \\ 501.76\pi & 44.8 & 1 \end{pmatrix}. \quad (26)$$

2. Estimate the local Euler characteristic in the reduced observation window  $W \ominus B_{r_{14}}(o) = [23, 1506] \times [23, 1506]$  for all radii  $r_0, \dots, r_{14}$  by computing  $V_0(\Xi \cap B_{r_i}(x))$  for all pixels  $x \in W \ominus B_{r_{14}}(o)$  and averaging over all pixels. That means the estimator  $\hat{y}$  from equation Eq. 16 is given in discretized form by

$$\hat{y}_j = \frac{1}{1484^2} \sum_{\substack{x=(x_1, x_2) \\ x_1, x_2 \in \{23, \dots, 1506\}}} V_0(\Xi \cap B_{r_j}(x)). \quad (27)$$

An algorithm to estimate the local Euler characteristic for all radii simultaneously is given in Klenk *et al.* (2006).

3. Now, the estimation of the specific intrinsic volumes is straightforward by computing the least-squares estimator  $v^* = (A^\top A)^{-1} A^\top \hat{y}$ .
4. To estimate the asymptotic covariance matrix the theory says we need an unboundedly increasing sequence of observation windows, *cf.* Eq. 6. In practice we have only one concatenated quadratic image with fixed size, and we assume it is large enough. So we choose the averaging set  $U = B_{300}(o)$  and estimate the variances and covariances using a discretized version of the estimator given in Eq. 23, Eq. 24 and Eq. 25 where the integrals are replaced by sums.

In the present study, the application of the joint estimator described above to the white phase of the images led to the following mean values for the stereological model parameters Eqs. 1–3.

Table 1. Means of the estimated values and the means of their asymptotic standard deviations. The latter means are not equal to the usual standard deviations between the cases within the groups.

	normal group	carcinoma group
$V_V$	0.15	0.43
$\text{Std}(V_V)$	0.037	0.047
$S_V$	$24.01 \text{ mm}^{-1}$	$40.85 \text{ mm}^{-1}$
$\text{Std}(S_V)$	$2.68 \text{ mm}^{-1}$	$1.97 \text{ mm}^{-1}$
$M_V$	$106.91 \text{ mm}^{-2}$	$297.91 \text{ mm}^{-2}$
$\text{Std}(M_V)$	$122.07 \text{ mm}^{-2}$	$180.93 \text{ mm}^{-2}$

In Table 1, the terms  $\text{Std}(V_V)$ ,  $\text{Std}(S_V)$  and  $\text{Std}(M_V)$  denote the means of the asymptotic standard deviations of the white phase per concatenated large image. Thus, they are not identical with the standard deviations of these model parameters ‘between images within cases’ and also not identical with the ordinary standard deviations of  $V_V$ ,  $S_V$  and  $M_V$  ‘between cases within groups’. The latter may be computed using standard statistical formulae even with a table calculator; however, this does not apply to the asymptotic standard deviations. In addition, the covariances between  $V_V$ ,  $S_V$  and  $M_V$  were computed, but these are not reiterated here. In order to see which parameter discriminated best between the groups, the results were visualized graphically. For example, Figs. 3 and 4 show the estimated area fraction of the white phase and the estimated mean curvature density of the white phase per unit area, respectively.

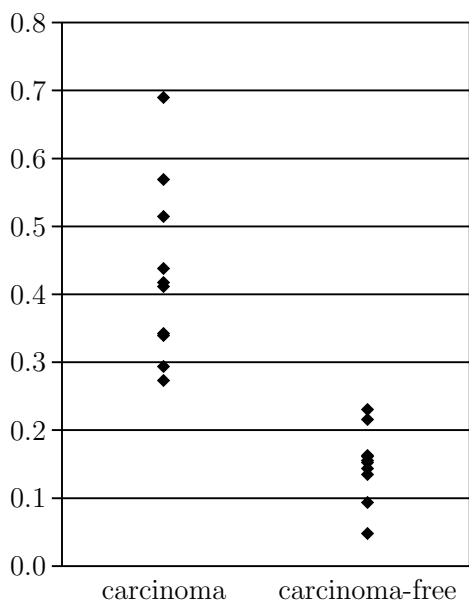


Fig. 3. Estimated area fraction of the white phase.

From Fig. 4 one can see that it is not possible to base the decision whether an image shows mammary carcinoma or not only on the mean curvature density. Also a statistical test of the mean surface area per unit volume alone does not lead to a sufficient discrimination of the groups. However, the area fraction of the white phase (Figure 3) seems to be a better parameter to categorize the images into two groups. Since this paper is focused on explaining the theory and the algorithm of the new estimation approach we will only test the area fraction which yields acceptable results. Anyhow, it is clear that the two groups in Fig. 4 belong to two different settings. Therefore one could combine, e.g., area fraction and Euler number and consider vectorial tests to strengthen the results of a one-dimensional test. With the approach described in the last section this is possible and will be done in a further paper.

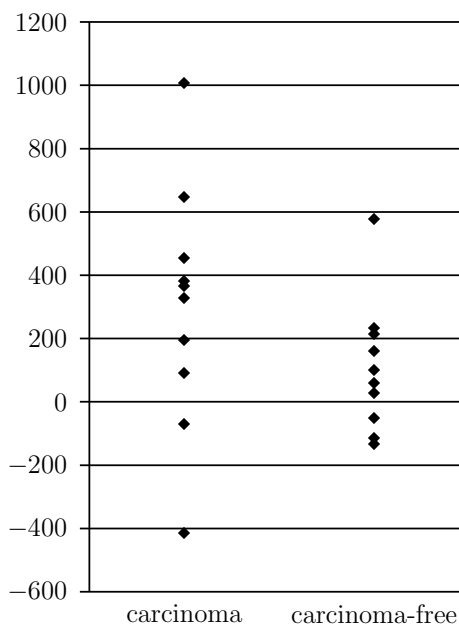


Fig. 4. Estimated mean curvature density  $M_V$  (in  $\text{mm}^{-2}$ ) of the white phase per unit area.

In the following, we write just ‘area fraction’ and omit the words ‘white phase’ for convenience. From the last section we know that the least-squares estimator is asymptotically normally distributed so we can construct a statistical test for the area fraction. Since the images have fixed size we don’t have an unboundedly increasing sequence of observation windows  $W_n$ . Instead we claim that the estimators are approximately Gaussian because our images are large enough. The null hypothesis states that ‘the expected

area fraction in image  $j$  corresponds to the mean area fraction in images showing normal mammary tissue,' where the images 1–10 show normal mammary tissue and the images 11–20 show mammary carcinoma tissue. So if we write  $p_j$  for the expected area fraction in image  $j$  and  $p_0$  denotes the mean area fraction in images showing normal mammary tissue, the null hypothesis reads as  $H_0 : p_j = p_0$ . As we can see from Fig. 3, the estimated values of the area fraction in images showing mammary carcinoma tissue are greater than in images showing tumor-free mammary tissue. That's why we consider the one-sided alternative hypothesis  $H_1 : p_j > p_0$ . The significance level is  $\alpha = 5\%$ . The unknown expected area fraction  $p_j$  in image  $j$  is estimated by the least-squares estimator given in Eq. 20 or, to be precise, by its third entry, and denoted by  $\hat{p}_j$ . When calculating the mean area fraction  $p_0$  one has to distinguish, if the image whose area fraction we want to test shows mammary cancer or normal tissue. In the first case we define  $p_0$  as the arithmetic mean of the estimated area fractions of the ten images showing tumor-free tissue, *i.e.*,

$$p_0 = \frac{1}{10} \sum_{i=1}^{10} \hat{p}_i. \quad (28)$$

But if the considered image shows tumor-free tissue, we have to exclude it from the calculation of the mean, so we define  $p_0$  by

$$p_0 = \frac{1}{9} \sum_{\substack{i=1 \\ i \neq j}}^{10} \hat{p}_i. \quad (29)$$

Now we can calculate the test statistic

$$T_j = \sqrt{|W|} \frac{\hat{p}_j - p_0}{\hat{\sigma}_j} \quad (30)$$

in image  $j$ , where  $\hat{\sigma}_j^2$  denotes the estimated variance of  $\sqrt{|W|}(\hat{p}_j - p_0)$ , see equations Eq. 23 and Eq. 24 and subsequent lines. In fact, it holds that  $\sqrt{|W|} = 1530$  in the considered case, because all images are squares with sidelength 1530 pixel. Slutsky's theorem yields that  $T_j$  is approximately standard Gaussian. With the 95%-quantile  $z_{0.95} = 1.64$  the critical range is  $(1.64, \infty)$ , so the null hypothesis for image  $j$  is rejected if the value  $T_j$  is greater than 1.64. The results are shown in Table 2.

Table 2. Results of the test  $H_0 : p_j = p_0$  for images showing tumor-free mammary tissue ( $j = 1, \dots, 10$ ) and for images showing mammary cancer tissue ( $j = 11, \dots, 20$ ).

j	$\hat{p}_j$	Rejection of $H_0$
1	0.047224	no
2	0.094392	no
3	0.138055	no
4	0.145576	no
5	0.155753	no
6	0.159763	no
7	0.163755	no
8	0.163812	no
9	0.217735	no
10	0.232209	no
11	0.277096	yes
12	0.297491	yes
13	0.342711	yes
14	0.342963	yes
15	0.418210	yes
16	0.423122	yes
17	0.442027	yes
18	0.520348	yes
19	0.570691	yes
20	0.693262	yes

The null hypothesis is not rejected for the images showing normal mammary tissue, but it is rejected for all ten images showing mammary carcinoma tissue, which means they are classified correctly.

Of course, we can, in a certain sense, exchange null and alternative hypothesis and test the null hypothesis that 'the expected area fraction in image  $j$  corresponds to the mean area fraction in images showing mammary carcinoma tissue' or, shortly,  $H_0 : p_j = \tilde{p}_0$ . Here,  $\tilde{p}_0$  denotes the mean area fraction of images showing mammary cancer tissue and as above there are two definitions for  $\tilde{p}_0$  depending on what type of tissue the considered image shows, *cf.* Eq. 28 and Eq. 29. Again, the alternative is one-sided  $\tilde{H}_1 : p_j < \tilde{p}_0$  and the test statistic is defined by

$$T_j = \sqrt{|W|} \frac{\hat{p}_j - \tilde{p}_0}{\hat{\sigma}_j}. \quad (31)$$

The results of this test are shown in Table 3.

Table 3. Results of the test  $\tilde{H}_0 : p_j = \tilde{p}_0$  for images showing tumor-free tissue ( $j = 1, \dots, 10$ ) and for images showing mammary cancer tissue ( $j = 11, \dots, 20$ ).

j	$\hat{p}_j$	Rejection of $\tilde{H}_0$
1	0.047224	yes
2	0.094392	yes
3	0.138055	yes
4	0.145576	yes
5	0.155753	yes
6	0.159763	yes
7	0.163755	yes
8	0.163812	yes
9	0.217735	yes
10	0.232209	yes
11	0.277096	yes
12	0.297491	yes
13	0.342711	yes
14	0.342963	no
15	0.418210	no
16	0.423122	no
17	0.442027	no
18	0.520348	no
19	0.570691	no
20	0.693262	no

Table 3 shows that the null hypothesis is rejected for all images showing normal tissue. But unfortunately there are three images out of the ten showing mammary carcinoma tissue for which the null hypothesis is rejected although it should not be.

## DISCUSSION

Eqs. 1–3 are well-known fundamental stereological formulae, valid under the conditions of isotropy and stationarity in a model-based approach, and under the condition of IUR sampling from arbitrary structures in a design-based approach. However, there is a practical difference: the estimators of the model parameters in Eqs. 1 and 2 are very easy to implement with an image analyzer, but the estimation of the Euler number is non-elementary even in 2D. While estimation of  $V_V$  and  $S_V$  is already taught in basic courses on stereology, this does not apply for the Euler number. Nevertheless the Euler number is of interest for a quantitative characterization of carcinoma tissue of glandular origin, because the Euler

number is directly linked to fundamental pathological tumor properties such as solid architecture where ideally  $\chi > 0$ , tubular architecture where ideally  $\chi = 0$ , and cribriform architecture where ideally  $\chi < 0$ . These textures may arise in all types of adenocarcinomas. Sometimes a cribriform texture can be found in mammary carcinomas; in practice it is most important to recognize this texture component in prostatic carcinomas, where it is known to be associated with a poorer prognosis as compared to tubular differentiation. Furthermore, the usual stereological approach, even if it encompasses the Euler number, leads merely to point estimates of the model parameters, but does not provide an insight into the covariance matrix, *i.e.*, the asymptotic variances and covariances of  $V_V$ ,  $S_V$ , and  $M_V$  of the three phases remain unknown. Point estimates for  $V_V$  and  $S_V$  of the epithelial phase of tumour-free mammary tissue and mammary carcinomas obtained from conventional stereological methods were previously published (Mattfeldt *et al.*, 2000, see Table 1 therein). The results were very similar to the present study. This shows the good reproducibility (robustness) of the method, in which the images were segmented interactively. Fully automatic segmentation of mammary tissue into epithelium, lumen, and tumour cells by image analysis would be desirable, but this aim is difficult to achieve at the moment. The plausibility of the results was checked by using a theorem of Tomkeieff, which states that the mean length of intercepts through particle profiles,  $l_1$ , is related to the Minkowski functionals of the particles by the equation  $l_1 = 4V_V/S_V$  (see Baddeley and Jensen, 2005, p. 33). According to this relation, one would expect values for  $l_1 \approx 0.025$  mm in the tumour-free group and  $l_1 \approx 0.042$  mm in the carcinoma group, which roughly corresponds to the visual impression, see Figs. 1b and 2b. In contrast to many other methods to estimate the specific intrinsic volumes the approach given in Spodarev and Schmidt (2005) yields not only the estimates, but also the (asymptotic) variances and covariances. All values can be computed from one large concatenated image. The classical approach to estimate the sample variance in each group is inappropriate here because we want to test each image separately. This might be interesting if there is only one image available in a practical application, which may be divided into subwindows.

In the special case of mammary tissue it turned out that the area fraction of the white phase, *i.e.*, the part of the image that shows epithelial cells, is a good criterion to detect if an image shows mammary cancer tissue or not. The area fraction has the useful property of being independent of the magnification of the image. Although this was not important in our study, it may be relevant for



the analysis of images where the scaling factor is unknown. A practical example for this situation has emerged more and more in the last decade. There are published datasets available in the internet consisting of images of various tumor types, *e.g.*, also mammary and prostatic carcinomas, where expert groups have performed the histopathological grading for reference purposes. Usually, these images are given without any information on the final magnification, and often it will not be possible to retrieve the magnification factor any more. It would be attractive to perform a quantitative meta-analysis of these images by means of spatial statistics. Due to the aforementioned reasons, one will then be restricted to methods which are independent of the magnification. This holds for the  $V_V$  component of the joint estimator described here. In contrast to the usual routine method, *e.g.*, point counting, our method will also provide an estimate of the asymptotic variance, and thus yield valuable additional information. As one can conclude from Tables 2–3, the test on tumor-free tissue yields better results than the test on mammary carcinoma tissue.

The most common tumor types of the female breast are invasive ductal and lobular carcinoma. These designations indicate a tumor differentiation more similar to the ducts or to the lobules of the mammary parenchyma, respectively. For both types of breast cancers, an attempt is made towards grading of malignancy in routine diagnostics. This is important for prognosis prediction and therapy planning. It is intended to apply our method for the characterization of mammary carcinomas of different degrees of malignancy, and eventually to use it for the prediction of the grade of malignancy from spatial data, *i.e.*, for the purpose of pattern recognition. Also it will be interesting to differentiate by this technique between ductal and lobular mammary carcinomas, which may be difficult in some cases (Mattfeldt and Fleischer, 2005). However, before these two more ambitious projects are put into practice, we thought it advisable to implement the methodology first in a simpler setting comparing tumor-free tissue to carcinoma tissue, where the differences between the classes of specimens are more pronounced. For the advanced applications, it may become useful to consider vectorial tests. With the described method it is possible to characterize the tissue high-dimensionally. If only one phase is considered, one obtains 9 instead of 3 numerical values per image (the three point estimates, the three asymptotic variances and the three asymptotic covariances of the Minkowski functionals). This rises to a whole bunch of characteristics if also the other two phases are taken into account. This will be subject of a further paper.

## ACKNOWLEDGMENTS

The authors thank an anonymous reviewer of IAS for constructive comments, which led to substantial improvements in the revised edition.

## REFERENCES

- Baddeley A, Vedel-Jensen EB (2005). *Stereology for Statisticians*. Boca Raton: Chapman & Hall.
- Elston CW, Ellis IO (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow up. *Histopathol* 19:403-10.
- Klenk S, Schmidt V, Spodarev E (2006). A new algorithmic approach to the computation of Minkowski functionals of polyconvex sets. *Computat Geom Theory Appl* 34:127-48.
- Mattfeldt T, Frey H, Rose C (1993). Second-order stereology of benign and malignant alterations of the human mammary gland. *J Microsc* 171:143-51.
- Mattfeldt T, Schmidt V, Reepschläger D, Rose C, Frey H (1996). Centred contact density functions for the statistical analysis of random sets. *J Microsc* 183:158-69.
- Mattfeldt T, Gottfried H-W, Schmidt V, Kestler HA (2000). Classification of spatial textures in benign and cancerous glandular tissues by stereology and stochastic geometry using artificial neural networks. *J Microsc* 198:143-58.
- Mattfeldt T, Stoyan D (2000). Improved estimation of the pair correlation function of random sets. *J Microsc* 200:158-73.
- Mattfeldt T, Kestler HA, Sinn HP (2004). Prediction of the axillary lymph node status in mammary cancer on the basis of clinicopathological data and flow cytometry. *Med Biol Engin Comput* 42:733-9.
- Mattfeldt T, Fleischer F (2005). Bootstrap methods for statistical inference from stereological estimates of volume fraction. *J Microsc* 218:160-70.
- Mattfeldt T, Eckel S, Fleischer F, Schmidt V (2006). Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. *J Microsc* 223:107-19.
- Molchanov I (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians*. Chichester: John Wiley & Sons.
- Pantle U, Schmidt V, Spodarev E (2006a). Central limit theorems for functionals of stationary germ-grain models. *Adv Appl Probab* 38:76-94.

- Pantle U, Schmidt V, Spodarev E (2006b). On the estimation of the integrated covariance function of stationary random fields. Preprint (submitted).
- Schmidt V, Spodarev E (2005). Joint estimators for the specific intrinsic volumes of stationary random sets. *Stochast Processes Appl* 115:959-81.
- Schneider R (1993). Convex Bodies, The Brunn-Minkowski Theory. *Encyclopedia of Mathematics and its Applications*, 44. Cambridge: Cambridge University Press.
- Schneider R, Weil W (1992). *Integralgeometrie*. Stuttgart: Teubner.
- Schneider R, Weil W (2000). *Stochastische Geometrie*. Stuttgart: Teubner.
- Spodarev E, Schmidt V (2005). On the local connectivity number of stationary random closed sets. In: C. Ronse, L. Najman, E. Decenciere, *Mathematical Morphology: 40 Years On*. Proceedings of the 7th International Symposium on Mathematical Morphology, April 18-20, 2005. Springer Series: Computational Imaging and Vision, 343-54.
- Stoyan D, Kendall WS, Mecke J (1995). *Stochastic Geometry and Its Applications*. 2nd Ed. Chichester: John Wiley & Sons.