Image Anal Stereol 2011;30:11-18
Original Research Paper

TESTING HISTOLOGICAL IMAGES OF MAMMARY TISSUES ON COMPATIBILITY WITH THE BOOLEAN MODEL OF RANDOM SETS

TOMÁŠ MRKVIČKA[1] AND TORSTEN MATTFELDT[2]

[1]Institute of Mathematics and Biomathematics, Faculty of Science, University of South Bohemia, Branišovská 31, 37005 České Budějovice, Czech Republic; [2]Institute of Pathology, University of Ulm, Oberer Eselsberg M23, D-89081 Ulm, Germany
e-mail: mrkvicka@prf.jcu.cz, Torsten.Mattfeldt@uniklinik-ulm.de
(Accepted March 1, 2011)

ABSTRACT

Methods for testing the Boolean model assumption from binary images are briefly reviewed. Two hundred binary images of mammary cancer tissue and 200 images of mastopathic tissue were tested individually on the Boolean model assumption. In a previous paper, it had been found that a Monte Carlo method based on the approximation of the envelopes by a multi-normal distribution with the normalized intrinsic volume densities of parallel sets as a summary statistics had the highest power for this purpose. Hence, this method was used here as its first application to real biomedical data. It was found that mastopathic tissue deviates from the Boolean model significantly more strongly than mammary cancer tissue does.

Keywords: Boolean model, breast cancer, goodness-of-fit test, pathology, random sets, stereology.

INTRODUCTION

Benign alterations and malignant tumours originating from glandular tissues (*e.g.*, mammary, prostatic, pancreatic tissue) are important diseases in human pathology. Structural characteristics of such tissues may be characterized in a descriptive manner, or it may be attempted to obtain objective quantitative data from the microscopic images. Such data may be used, *e.g.*, for statistical group comparisons, for correlation analyses, for parametric modelling and pattern recognition. Here one should keep in mind that the tissues are in fact three-dimensional; in histopathology we are faced with planar sections of them, an aspect which is often overlooked. *Stereology* means that data obtained from sections are extrapolated to the three-dimensional properties of a structure using mathematical methods.

In previous investigations, it has been shown that the texture of mammary tissue, as seen at low magnification, may be characterized quantitatively in terms of stereology (Mattfeldt *et al.*, 1993, 1996, 2007; Mattfeldt, 2003). Basically, glandular tissue may be subdivided into three phases, namely the epithelial cells (the tumour cells), the lumina, and the stroma, which together account for 100% of the tumour tissue. These three phases may be understood as random closed sets (RACS) with positive volume fraction (volume processes). Applying methods of spatial statistics to digitized images, or by simple manual counting methods, it is possible to characterize these three phases quantitatively in terms of area fraction A_A, boundary length density L_A and Euler number per unit tissue area χ_A (see, *e.g.*, Mattfeldt *et al.*, 2007). The three aforementioned specific intrinsic volumes have a clear stereological interpretation, hence they can be used for the estimation of stereological model parameters:

$$\hat{V}_V = A_A \,, \tag{1a}$$

$$\hat{S}_V = \frac{4}{\pi} L_A \,, \tag{1b}$$

$$\hat{M}_V = 2\pi\chi_A \,, \tag{1c}$$

where V_V is the volume fraction, S_V is the mean surface area per unit reference volume, and M_V is the curvature density (integral of mean curvature per unit volume); by $\hat{V}_V$, $\hat{S}_V$ and $\hat{M}_V$ we denote the estimators of these quantities.

However, a RACS is not uniquely characterized by the specific intrinsic volumes. They inform basically about the amount of the phases per unit volume, but not about the pattern in which the features are arranged ('histological texture, architecture'). A useful nonparametric way to describe the tissue texture in this sense consists in the estimation of second-order statistics of the RACS (Mattfeldt *et al.*, 1993). The blend of stereology with second-order methods of stochastic geometry was called *second-order stereology* (Cruz-Orive, 1989; Jensen *et al.*, 1990; Mattfeldt *et al.*, 1993, 1996, 2003). Estimates of the covariance $C(r)$, of the radial distribution function $RDF(r)$, of the correlation function $k(r)$, of the pair correlation function $g(r)$, and of the reduced

second moment function $K(r)$ of the epithelial volume may be obtained, as reported before (Mattfeldt *et al.*, 1993, 2003). These summary statistics provide a quantitative characterization of the inner order of the structure in terms of attraction (clustering) and repulsion as a function of distance, without specifying any particular stochastic model. The specific intrinsic volumes computed from parallel sets of RACS (*e.g.*, the Minkowski sum of the RACS and the disk with radius r) with varying radius of the parallel set r (Mrkvička, 2009) reflect also the inner order of the structure and furthermore these statistics are based on the specific intrinsic volumes.

In spatial statistics, much work has been done in the field of the statistical analysis and modelling of spatial point patterns (see, *e.g.*, Illian *et al.*, 2008). When studying planar point patterns, the usual first step is to check whether the pattern is compatible with the null model of a stationary Poisson point process. Analogously, when it comes to stochastic modelling of a RACS with positive volume, one would proceed by testing its compatibility with a suitable null model. The *Boolean model*, which represents the classical, "purely random" germ-grain model of stochastic geometry, may be used instrumentally as a null model for a volume process (Molchanov, 1997). This means that one has to develop a test whether a given image is compatible with the Boolean model of random sets, or whether this hypothesis should be rejected.

In the present investigation, the images of the mammary cases studied in a previous publication (Mattfeldt *et al.*, 1996) were reexamined (20 cases of mastopathy and 20 cases of mammary cancer, each with 10 images). In the previous paper, it had been explored how strongly the images deviated from an appropriately parametrized Boolean model in terms of contact distribution functions. A direct test of the images on compatibility with the Boolean model was outlined in that paper, but was not yet performed. In a further paper (Mattfeldt, 2003) these images were tested on compatibility with the Boolean model using Laslett's theorem (see section Laslett's theorem). Since the results of Laslett's test did not demonstrate a significant deviation of most of these images from the Boolean model hypothesis, we used now a more powerful method to prove this difference (Mrkvička, 2009). Furthermore, we compared how strongly the mastopathy images and mammary cancer images deviated from the Boolean model. Finally, we compare the results obtained in the paper (Mattfeldt, 2003) with the results obtained in this study.

THEORETICAL PART

REVIEW OF BASIC CONCEPTS

Throughout this paper, we consider realizations of stationary and isotropic ergodic RACS Ξ_V with positive volume fraction V_V in an unbounded three-dimensional reference space. Such RACS with the property $V_V > 0$ are denoted as *volume processes*, for convenience (Cruz-Orive, 1989; Mattfeldt *et al.*, 1993). A RACS Ξ_V is called stationary and isotropic if Ξ_V has the same distribution after arbitrary translations and rotations, *i.e.*, if the distribution of Ξ_V is invariant under all rigid motions in space. The property of ergodicity may be summarized as follows: statistical averages can be expressed by limits of arithmetic or spatial averages (Stoyan *et al.*, 1987, p. 170–171). Consider germs which are distributed according to a stationary Poisson process with intensity λ_V in space, and primary grains which are independent identically distributed random compact sets of mean volume $\bar{V}$ and mean surface area $\bar{S}$, with their centers at the origin. Then the Boolean model is the union of a shifted version of the primary grains, namely of sets $x+K$, where x is a germ point and K is a primary grain (Stoyan *et al.*, 1987). The shifted primary grains may overlap. In addition, here we assume that the primary grains are convex and isotropic (*i.e.*, the distribution of the primary grains is invariant under rotations about the origin). A planar section of Ξ_V is a Boolean model Ξ_A with intensity λ_A in the two-dimensional Euclidean plane, where the planar primary grains have mean area $\bar{A}$ and mean boundary length $\bar{L}$ (Stoyan *et al.*, 1987, p.85). Therefore, testing for a Boolean model was throughout our work restricted to a study in the plane, which provides an example for stereological inference from sections to 3D space (*i.e.*, rejection of the Boolean model hypothesis in 2D implies rejection of the Boolean model hypothesis in 3D). The three parameters λ_A, $\bar{A}$ and $\bar{L}$ of Ξ_A can be estimated from the image, through the following three equations (Stoyan *et al.*, 1987):

$$A_A = 1 - \exp(-\lambda_A \bar{A}) \tag{2a}$$

$$L_A = \lambda_A (1 - A_A) \bar{L} \tag{2b}$$

$$N_A^+ = \lambda_A (1 - A_A) \tag{2c}$$

In these equations, A_A denotes the mean area fraction, L_A denotes the mean boundary length per unit area, and N_A^+ denotes the specific convexity number of Ξ_A, which can be estimated unbiasedly from the images by manual counting techniques or with image analyzers.

Important tools to test whether a stationary and isotropic ergodic RACS Ξ is a Boolean model are the

contact distribution functions $H_B(r)$ with

$$H_B(r) = \frac{P(\Xi \cap rB \neq \emptyset) - P(o \in \Xi)}{1 - P(o \in \Xi)} , \qquad (3)$$

for $r \geq 0$, where B belongs to a certain class of structuring elements; $rB = \{rx : x \in B\}$, and o is the origin (Stoyan *et al.*, 1987). This definition means that, for a given value $r > 0$, the value $H_B(r)$ denotes the conditional probability that the r-fold enlargement rB of the structuring element B — here: a quadrat of sidelength 2 pixels — intersects Ξ, provided that the reference point of the enlarged structuring element (usually its centre) lies in the pore space, outside Ξ. Theoretically, the direction of the quadrat B may be arbitrarily selected when the structure is isotropic; in digitized images one will usually select the vertical or the horizontal direction. As its name implies, $H_B(r)$ is a cumulative distribution function. General equations have been derived where the contact distribution functions were given for the Boolean model in terms of the intrinsic volumes (which are also known under another normalization as Minkowski functionals or Quermaßintegrale) of the primary grains and the structuring element B (Matheron, 1975; Schneider and Weil, 1992); here we restrict ourselves to the familiar specifications for the plane. Let us denote by $H_{BM}(r)$ and $H_{QM}(r)$ the exact contact distribution function for general and quadratic structuring elements B, respectively, under the condition that Ξ_A is a Boolean model. For general structuring elements B with area $A(B)$ and perimeter $U(B)$, which are subsequently enlarged r-fold, we have the following general contact distribution function:

$$H_{BM}(r) = 1 - \exp\left\{\frac{1}{A_A - 1}\left[\frac{U(B) r L_A}{2\pi} + A(B) r^2 N_A^+\right]\right\} . \qquad (4a)$$

For the aforementioned structuring element B, *i.e.*, a quadrat of sidelength 2 pixels, we obtain the following contact distribution function:

$$H_{QM}(r) = 1 - \exp\left\{\frac{1}{A_A - 1}\left(\frac{4 r L_A}{\pi} + 4 r^2 N_A^+\right)\right\} . \qquad (4b)$$

Using image analysis, we have asymptotically unbiased estimators of the contact distribution functions at our disposal, which are suitable irrespectively whether the assumption of a Boolean model is fulfilled or not. Let us denote by $H_{BI}(r)$ and $H_{QI}(r)$ the contact distribution functions for general and quadratic structuring elements B, respectively, as measured by image analysis without the assumption of a Boolean model. The general contact distribution function can be estimated from the image by

$$H_{BI}(r) = \{A_A(\mathrm{dil}[r]) - A_A\}/(1 - A_A) , \qquad (5)$$

(Stoyan *et al.*, 1987; Bindrich and Stoyan, 1991), where A_A is the area fraction before dilation, and where $A_A(\mathrm{dil}[r])$ is the area fraction after dilation by the r-fold enlargement of the structuring element B.

TESTING FOR THE BOOLEAN MODEL

The detailed description of the tests can be found in Mrkvička (2009). Here we review the basic idea of the tests only.

The graphical method

A first check for presence of a Boolean model usually proceeds as follows (*e.g.*, Bindrich and Stoyan, 1991). A series of successive dilations of the grain phase is performed, then $-\log\{1 - H_{QI}(r)\}/r$ is plotted as a function of r. If the image stems from a Boolean model, the data points should lie near a straight line in the resulting plot, apart from random error. This behaviour is a direct consequence of Eq. 4b.

Laslett's theorem

Another test method may proceed from the following theorem (Laslett *et al.*, 1985). If Ξ_A is a Boolean model with intensity λ_A and isotropic convex primary grains in the plane, for which a unique marker point is defined on their boundary, then the exposed marker points of Ξ_A (called the induced point process) form, after a certain transformation, a homogeneous Poisson process of intensity λ_A (see also Cressie, 1991; Černý, 2006). After the transformation we may leave the domain of RACS and apply tests for homogeneous Poisson *point* processes to the new image.

A Monte Carlo test for the Boolean model based on the quadratic contact distribution function

The equations discussed above provide an opportunity for a *Monte Carlo test* on the presence of a Boolean model. The equations (2a–c) provide estimates $\hat{\lambda}_A$, $\hat{\bar{A}}$, and $\hat{\bar{L}}$ of the parameters λ_A, $\bar{A}$, and $\bar{L}$ from a quadratic window of area h^2, say. Now the Boolean model with estimated parameters is simulated in a given window. For each simulated image a series of linear and quadratic dilations is performed at the same step sizes that were used for the original images, and for each simulation $H_{QI}(r)$ are determined according to Eq. 5. With respect to a certain notion of distance (*e.g.*, sum of squared differences), the deviations between the theoretical contact distribution

function $\hat{H}_{QM}(r)$ and the simulated contact distribution functions $H_{QI}(r)$ are ordered by size. If the deviation of the function $\hat{H}_{QI}(r)$, that had been determined directly from the original image, from $\hat{H}_{QM}(r)$ is among 5% largest deviations computed from simulations, the hypothesis of a Boolean model is rejected.

A general Monte Carlo test with envelopes approximated by the multinormal distribution

The null model assumption here can be the Boolean model or any other model.

1. A summary statistic $S(\varepsilon)$ of RACS is chosen. The summary statistic is a function $S(\varepsilon)$ of $\varepsilon > 0$ and it is estimated from the data in n different points $\varepsilon_1, \ldots, \varepsilon_n$. Denote these estimates $\hat{S}^D_{\varepsilon_1}, \ldots, \hat{S}^D_{\varepsilon_n}$.
2. The parameters θ of the assumed null model are estimated by an estimator $\hat{\theta}$.
3. N independent samples of the null model with estimated parameters are simulated and the summary statistics $\hat{S}^i_{\varepsilon_1}, \ldots, \hat{S}^i_{\varepsilon_n}$, $i = 1, \ldots, N$ are computed.
4. To compute the p-level of this test, the summary statistics $S_{\varepsilon_1}, \ldots, S_{\varepsilon_n}$ are approximated by the random vector $\mathbf{X}$ with multinormal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ and the covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})_{i,j=1,\ldots,n}$ computed from N simulations. The upper envelope is constructed as $UE(s) = (\mu_1 + s\sqrt{\Sigma_{11}}, \ldots, \mu_n + s\sqrt{\Sigma_{nn}})^T$ and the lower envelope is constructed as $LE(s) = (\mu_1 - s\sqrt{\Sigma_{11}}, \ldots, \mu_n - s\sqrt{\Sigma_{nn}})^T$. The width of the envelopes depends on the parameter $s > 0$.
5. An integer number K is chosen. This is the maximum number of values $\hat{S}^D_{\varepsilon_1}, \ldots, \hat{S}^D_{\varepsilon_n}$ which fall outside of the envelopes and for which the test does not reject. The widest envelope when the test rejects is then given by

$$s_{\sup} = \sup_s(\#\{\hat{S}^D_{\varepsilon_i} \notin (LE(s)_i, UE(s)_i), \\ i = 1, \ldots, n\} > K) \,. \quad (6)$$

6. The p-value is then the probability that more than K components of the random vector $\mathbf{X}$ fall outside of the envelopes $(LE(s_{\sup}), UE(s_{\sup}))$. Since $\mathbf{X}$ is usually high dimensional, the probability is computed by Monte Carlo methods.

Normalized intrinsic volumes densities

The last method is described for the general summary statistic $S(\varepsilon)$. It is possible to use quadratic contact distribution function in this test as it was used in the Monte Carlo test described in previous section. Or one can use as the summary statistic the following normalized intrinsic volumes densities of the ε-parallel sets.

The intrinsic volumes densities $\overline{V}_0(\Xi)$, $\overline{V}_1(\Xi)$, $\overline{V}_2(\Xi)$ of RACS Ξ in the plane are the mean Euler number density, one half of the circumference density of border $\partial\Xi$ and the area density (*i.e.*, $\overline{V}_2(\Xi) = A_A$, $\overline{V}_1(\Xi) = L_A/2$).

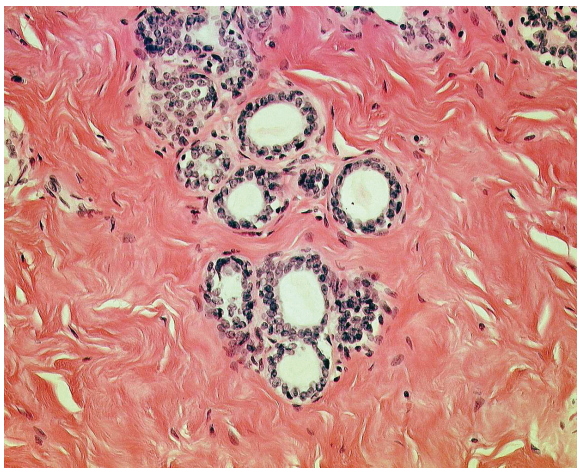
We considered the intrinsic volumes densities $\overline{V}_k(\Xi_{\varepsilon_i})$, $k = 0,1,2$, $i = 1, \ldots, n$ of the ε_i-parallel sets. The parallel set Ξ_{ε_i} is produced as dilation of the set Ξ by a disc with a radius ε_i. We chose $n = 24$ different discs with radii $\varepsilon_1, \ldots, \varepsilon_{24}$ evenly spanned between 1 and 25 pixels of the image. When the envelopes are made from the simulations, the estimated intrinsic volumes densities $\overline{V}_k(\Xi_{\varepsilon_1})$ vary a lot for different simulations thus the envelopes are wide. Therefore we fixed first point ε_1 and chose as summary statistics for the proposed test the normalized intrinsic volumes of parallel sets:

$$\overline{V}_k(\Xi_{\varepsilon_i})/\overline{V}_k(\Xi_{\varepsilon_1}), \quad i = 2, \ldots, 24, \quad k = 0,1,2. \quad (7)$$

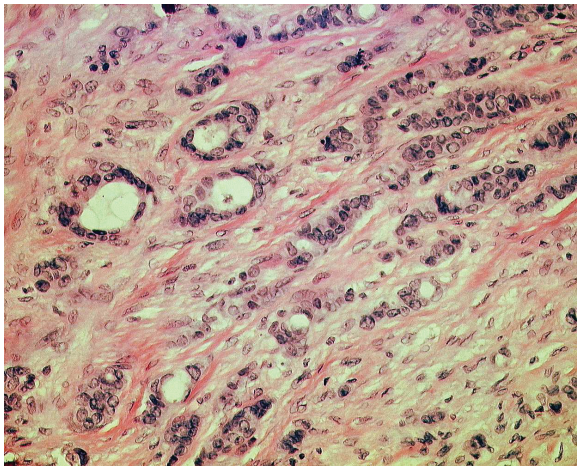
This normalized intrinsic volumes reflect the interactions among the grains and are unaffected by the number and size of the grains. The estimation of $\overline{V}_2(\Xi_{\varepsilon_i})$ is performed by the standard unbiased point counting estimator and the estimation of $\overline{V}_0(\Xi_{\varepsilon_i}), \overline{V}_1(\Xi_{\varepsilon_i})$ is performed by the unbiased estimator described in Mrkvička and Rataj (2008; 2009).

Choosing the testing method

All the previously described methods were compared in Mrkvička (2009) by extensive simulations. The powers of the tests with respect to two alternatives (regular and clustered structure) were compared. The results of this study were as follows. Using normalized intrinsic volume densities for the summary statistics was more powerful than using quadratic contact distribution functions. Furthermore, the Monte Carlo test with envelopes approximated by a multinormal distribution was more powerful than all the other tests described there. Moreover, it was shown in Mrkvička (2009) that the aforementioned test was sensitive to the grain interaction only. It was not sensitive on an eventually wrong choice of the random prototype of the Boolean model (the primary grain of the Boolean model). Therefore, the Monte Carlo test with envelopes approximated by a multinormal distribution with normalized intrinsic volumes densities as the summary statistics was adopted for testing the Boolean assumption in the present study.



(a)



(b)

Fig. 1. *(a) Mastopathic tissue: the mammary parenchyma shows an increase of stroma (fibrous tissue) and sometimes dilatation of ductules. The general ductulo-lobular architecture is however preserved. (b) Mammary cancer (invasive ductal mammary carcinoma). The normal orderly glandular architecture has been replaced by irregular epithelial blocks with few stroma in between. Haematoylin-Eosin stain.*

MATERIALS AND METHODS

CASES AND SAMPLING

Forty cases of human mammary tumours submitted for histopathological diagnosis were investigated. Twenty cases were fibrous mastopathies, *i.e.*, benign lesions where the glandular architecture of the mammary tissue within the lobules was fully preserved and the main changes consisted in an increase of fibrous tissue and a microcystic dilatation of the glandular lumina (Fig. 1a). These were compared to 20 cases of invasive ductal mammary cancer, the most frequent type of breast cancer in humans (Fig. 1b). One paraffin section per case with a nominal thickness of 4 μm from the centre of the lesion was stained with hematoxylin and eosin. Ten visual fields per case from the lobular parenchyma were evaluated in the group of mastopathies at 10× primary magnification at the level of the objective of the light microscope by systematic random sampling. Ten visual fields per case from non-necrotic invasive tumour tissue were evaluated in the group of carcinomas at the same magnification by the same sampling strategy, *i.e.*, systematic sampling with a random start. The selected visual fields were transmitted to the image analysis system Kontron IBAS 2000 with a black-and-white CCD camera. The result was a gray level image with a resolution of 512×512 pixels at a final magnification of 430× on the screen. By segmentation a binary image was produced, which consisted of two phases only (Figs. 2a,b). All images were interactively segmented by the same person by tracing the epithelial formations. The epithelial component – the union of the primary grains – was shown as white, whereas the whole non-epithelial remainder of the tissue – the pore space, consisting of fibrous stroma, blood vessels, nerves, gland lumina, *etc.* – was shown as black.

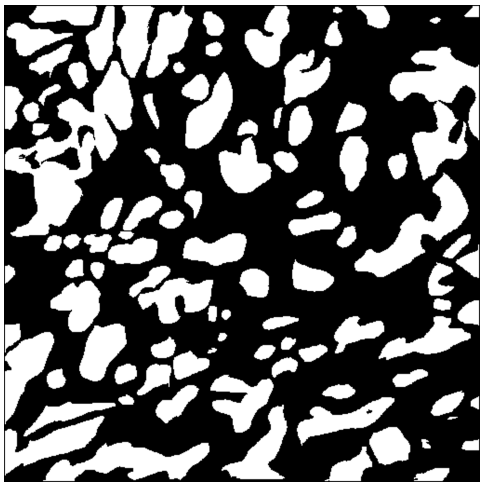
MONTE CARLO TESTS

Each black-white image with the resolution of 512×512 pixels was tested by the chosen test (the Monte Carlo test with envelopes approximated by multinormal distribution with a normalized intrinsic volumes densities as the summary statistic). The chosen test was performed with the following settings: $n = 69$ (23 points for 3 normalized intrinsic volumes densities), $N = 99$ simulations, $K = 3$. The parameters of the Boolean model were estimated using empirical intrinsic volume densities $\overline{V}_0(\Xi), \overline{V}_1(\Xi), \overline{V}_2(\Xi)$ (Molchanov, 1997, p. 81–83).

RESULTS

The results of the chosen test of Boolean assumption are displayed on (Figs. 3a,b,c) for one image of the mastopathic tissue (the image is shown in Fig. 2a). The resulted p-value of this test is $p = 0.00067$. Figs. 4a,b,c shows the same for one image of the mammary cancer tissue (the image is shown in Fig. 2b). The resulting p-value of this test is $p = 0.11687$. The results for other images vary a lot, but usually the significance is proven by $\overline{V}_0(\Xi)$ or $\overline{V}_1(\Xi)$.

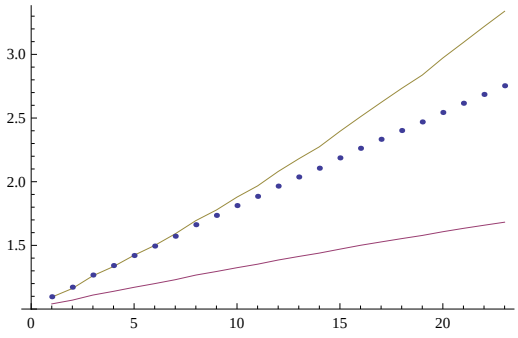
(a)



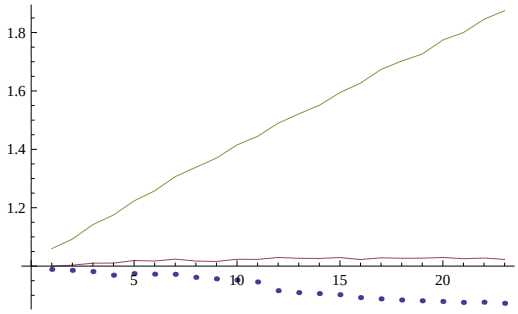
(b)

Fig. 2. *(a) Binary image from a visual field from a histological section from a mastopathy. A gray-level image has been reduced to a binary image by manual segmentation. White: epithelial cells (grain phase), black: gland openings, stroma, vessels etc. (pore phase). (b) Binary image from a visual field from a histological section of a case of mammary cancer.*

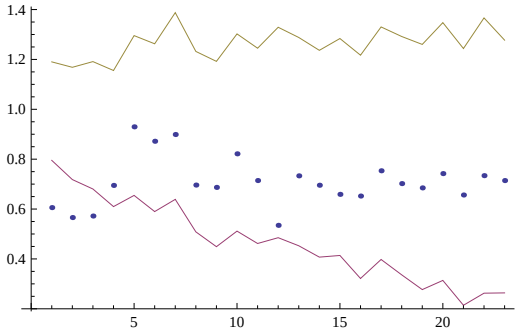
All the *p*-values are summarized in the histograms in the (Figs. 5a,b). We can see that the mammary cancer tissues have generally higher *p*-values than the mastopathic tissues. This means that mammary cancer tissues reveal less deviation from the Boolean model than the mastopathic tissues. Furthermore, the Boolean model assumption was not rejected for 6% mammary cancer images and for 1% mastopathic images with significance level 0.05. To test the hypothesis that the mammary cancer tissues reveal the same deviation from the Boolean model as the mastopathic tissues, we performed a nonparametric Mann-Whitney U Test, where we compared the samples of computed *p*-values. This hypothesis was rejected with the *p*-value 0.000159.



(a)



(b)



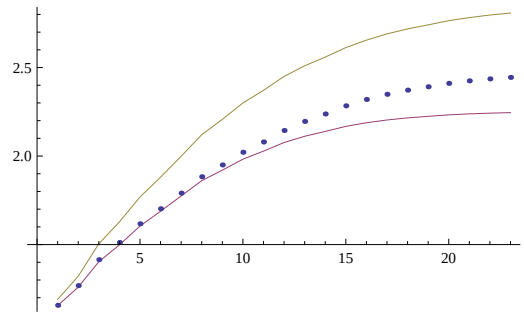
(c)

Fig. 3. *The results of the proposed test on the Boolean model assumption for one image of the mastopathic tissue. The points represent the estimates of normalized intrinsic volumes densities of parallel sets (a)* $\overline{V}_2(\Xi_\varepsilon)/\overline{V}_2(\Xi_{\varepsilon_1})$ *, (b)* $\overline{V}_1(\Xi_\varepsilon)/\overline{V}_1(\Xi_{\varepsilon_1})$*, (c)* $\overline{V}_0(\Xi_\varepsilon)/\overline{V}_0(\Xi_{\varepsilon_1})$ *for 23 different radii from the data. The envelopes are constructed from 99 simulations; they correspond to 95% envelopes.*

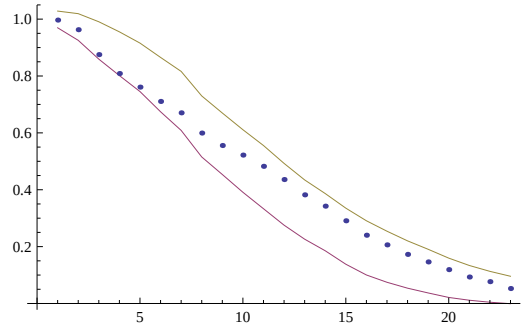
Table 1 shows the average *p*-values computed from the images for each case (20 mastopathic cases and 20 mammary cancer cases). If we take the average *p*-value for a case as the characteristic determining the decision, whether the whole case can be described by Boolean model or not, then 10% of the mammary cancer cases and no mastopathic case at all can be accepted as consonant with the Boolean model.

Table 1. *The average p-values computed from the images for each case.*

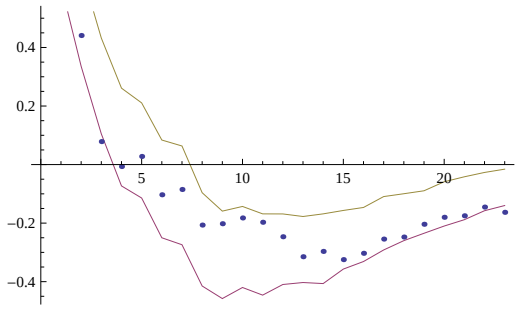
Mammary cancer	0.02844	0	0.00050	0.00681	0.11612	0.02688	0.10948
	0.25630	0.00492	0	0.02463	0	0	0
	0	0.00494	0.00526	0.00011	0	0.03093	
Mastopathic	0.00013	0.06126	0.00005	0.00001	0.00440	0.03554	0.00222
	0.00122	0.00055	0	0	0	0	0
	0	0.00172	0	0	0	0	



(a)

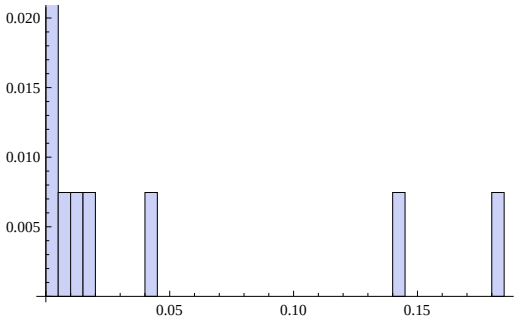


(b)

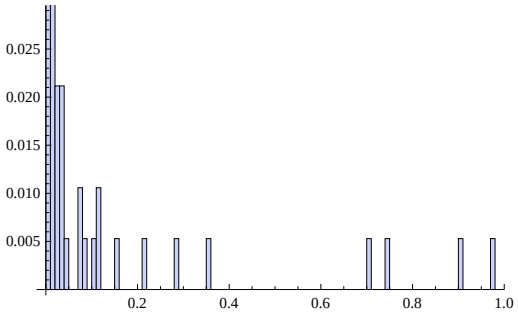


(c)

Fig. 4. *The results of the proposed test on the Boolean model assumption for one image of the mammary cancer tissue. The points represent the estimates of normalized intrinsic volumes densities of parallel sets (a)* $\overline{V}_2(\Xi_\varepsilon)/\overline{V}_2(\Xi_{\varepsilon_1})$ *, (b)* $\overline{V}_1(\Xi_\varepsilon)/\overline{V}_1(\Xi_{\varepsilon_1})$*, (c)* $\overline{V}_0(\Xi_\varepsilon)/\overline{V}_0(\Xi_{\varepsilon_1})$ *for 23 different radii from the data. The envelopes are constructed from 99 simulations; they correspond to 95% envelopes.*



(a)



(b)

Fig. 5. *(a) The histogram of resulting p-values of all mastopathic tissue images. (b) The histogram of resulting p-values of all mammary cancer tissue images. (The scale of the histogram is different.)*

CONCLUSION

In this paper, the Monte Carlo method based on the approximation of the envelopes by a multi-normal distribution with the normalized intrinsic volume densities of parallel sets as a summary statistics was applied to biological tissue for the first time. In simulation studies, it had been shown to have higher statistical power than alternative methods (Mrkvička, 2009). One of these alternative approaches was the test based on Laslett's theorem (see section Laslett's theorem). This finding could now be corroborated for real specimens. In a previous paper, the same 40 cases had been studied using this test (Mattfeldt, 2003). Using this method, it was found that 25% of the mastopathic cases were compatible with the Boolean

model, whereas 75% of the mammary cancer cases were compatible with the Boolean model (Mattfeldt, 2003, p. 291). However, no mastopathic case at all and only 10% of the mammary cancer cases were accepted as consonant with the Boolean model in the present study. Clearly, this result has to be ascribed to the higher power of the test described here.

Moreover, the present study suggests that the structure of the epithelial component of mammary cancer is less orderly arranged than that of the mastopathic tissue. Although some images of the mammary cancer revealed very small *p*-values, thus revealing a strong dissimilarity to the Boolean model, in global view we found that the mastopathic tissue differed more strongly from the Boolean model than the mammary cancer tissue. This conclusion is biologically plausible as mastopathic tissue is still subject to normal growth regulation mechanisms. However, cancer means that tumour cells have escaped growth control of the organism and are now proliferating autonomously. This means a loss of order in terms of geometry of random sets. Besides, the conclusion that the mastopathic cases differ more strongly from the Boolean model than the mammary cases emerges also from the evaluation using the Laslett test (Mattfeldt, 2003).

The aim of the present study was primarily to find an improved method to test an empirically given volume process on compatibility with the Boolean model, in analogy to a test on complete spatial randomness of point patterns. If it were considered as a pattern recognition tool for tumour diagnosis, it might be asked whether the distance between the studied case and the Boolean model (*i.e.*, the computed average *p*-value) is a suitable criterium for distinguishing whether a case is malignant or not. Table 1 shows, that the average *p*-value per case is generally smaller for mastopathic cases, but there are also cases of mammary cancer, which have an average *p*-value indistinguishable from *p*-values of malignant cases. Thus, the larger its average *p*-value is, the more is it probable that the case is mammary cancer. On the other hand, we have no diagnostic conclusion if the average *p*-value is small. Therefore the average *p*-value per case cannot be used alone for distinguishing whether a case is malignant or not.

ACKNOWLEDGEMENTS

The work was supported by the Grant Agency of Czech Republic, Projects No. P201/10/0472.

REFERENCES

Bindrich U, Stoyan D (1991). Stereology of the pore structure in bread. J Microsc 162:231–9.

Cressie NAC (1991). Statistics for Spatial Data. New York: Wiley.

Cruz-Orive LM (1989). Second-order stereology: estimation of second moment volume measures. Acta Stereol 8:641–6.

Černý R (2006). Laslett's transform for the Boolean model in R-d . Kybernetika 42:569–84.

Illian J, Penttinen A, Stoyan H, Stoyan D (2008). Statistical Analysis and Modelling of Spatial Point Patterns. New York: Wiley.

Jensen EB, Kiêu K, Gundersen HJG (1990). Second-order stereology. Acta Stereol 9:15–35.

Laslett GM, Cressie N, Liow S (1985). Intensity estimation in a spatial model of overlapping particles. Unpublished manuscript, Division of Mathematics and Statistics, CSIRO, Melbourne [quoted from Cressie (1991)].

Matheron G (1975). Random Sets and Integral Geometry. New York: Wiley.

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. A stereological study on benign and malignant glandular tissue using image analysis. J Microsc 183:158–69.

Mattfeldt T (2003). Classification of binary spatial textures using stochastic geometry, nonlinear deterministic analysis and artificial neural networks. Int J Pattern Recogn Artif Intell 17:275–300.

Mattfeldt T, Meschenmoser D, Pantle U, Schmidt V (2007). Characterization of mammary gland tissue using joint estimators of Minkowski functionals. Image Anal Stereol 26:13–22.

Molchanov I (1997). Statistics of the Boolean model for practitionars and mathematicians. New York: Wiley.

Mrkvička T (2009). On testing of general random closed set model hypothesis. Kybernetika 45/2: 293-308.

Mrkvička T, Rataj J (2008). On estimation of intrinsic volume densities of stationary random closed sets. Stoch Proc Appl 118:213–31.

Mrkvička T, Rataj J (2009). On estimation of intrinsic volume densities of stationary random closed sets via parallel sets in the plane. Kybernetika 45:931–45.

Ripley BD (1981). Spatial statistics. Wiley, New York.

Schneider R, Weil W (1992). Integralgeometrie. Stuttgart: Teubner.

Stoyan D, Kendall WS, Mecke J (1987). Stochastic geometry and its applications. Chichester: Wiley.