

Received 24 September 2014,

Accepted 12 March 2015

Published online 7 April 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6497

Fractal and stochastic geometry inference for breast cancer: a case study with random fractal models and Quermass-interaction process

Philipp Hermann,^a Tomáš Mrkvíčka,^b Torsten Mattfeldt,^c Mária Minárová,^d Kateřina Helisová,^e Orietta Nicolis,^f Fabian Wartner^a and Milan Stehlík^{a,g,*†}

Fractals are models of natural processes with many applications in medicine. The recent studies in medicine show that fractals can be applied for cancer detection and the description of pathological architecture of tumors. This fact is not surprising, as due to the irregular structure, cancerous cells can be interpreted as fractals. Inspired by Sierpinski carpet, we introduce a flexible parametric model of random carpets. Randomization is introduced by usage of binomial random variables. We provide an algorithm for estimation of parameters of the model and illustrate theoretical and practical issues in generation of Sierpinski gaskets and Hausdorff measure calculations. Stochastic geometry models can also serve as models for binary cancer images. Recently, a Boolean model was applied on the 200 images of mammary cancer tissue and 200 images of mastopathic tissue. Here, we describe the Quermass-interaction process, which can handle much more variations in the cancer data, and we apply it to the images. It was found out that mastopathic tissue deviates significantly stronger from Quermass-interaction process, which describes interactions among particles, than mammary cancer tissue does. The Quermass-interaction process serves as a model describing the tissue, which structure is broken to a certain level. However, random fractal model fits well for mastopathic tissue. We provide a novel discrimination method between mastopathic and mammary cancer tissue on the basis of complex wavelet-based self-similarity measure with classification rates more than 80%. Such similarity measure relates to Hurst exponent and fractional Brownian motions. The R package FractalParameterEstimation is developed and introduced in the paper. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: box-counting dimension; breast cancer; Hausdorff measure; pathology; Quermass-interaction process

1. Introduction

Benign alterations and malignant tumors originating from glandular tissues (e.g., mammary, prostatic, or 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 an objective model of it. Here, one should keep in mind that tissues are in fact three-dimensional, but in histopathology, we are faced with planar sections of them, and we will model those planar sections in this paper. In previous

^aDepartment of Applied Statistics, Johannes Kepler University Linz, Linz, Austria

^bDepartment of Applied Mathematics and Informatics, University of South Bohemia in České Budějovice, České Budějovice, Czech Republic

^cInstitute of Pathology, Universitätsklinikum Ulm, Ulm, Germany

^dDepartment of Mathematics, Slovak University of Technology, Bratislava, Slovak Republic

^eDepartment of Mathematics, Czech Technical University in Prague, Prague, Czech Republic

^fInstitute of Statistics, University of Valparaíso, Valparaíso, Chile

^gDepartamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile

*Correspondence to: Milan Stehlík, Departamento de Matemática, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile.

†E-mail: mlnstehlik@gmail.com, Milan.Stehlik@jku.at

investigations, it has been shown that the texture of mammary tissue, as seen at low magnification, may be characterized quantitatively in terms of stereology [1–3]. Basically, glandular tissue may be subdivided into three phases, namely epithelial cells (tumor cells), lumina, and stroma, which together account for 100% of the tumor tissue. These three phases may be understood as random closed sets with positive volume fraction [4].

The term fractal came from the Latin adjective *fractus* meaning fragmented or broken and was originally given to highly irregular sets by the French mathematician Benoit Mandelbrot in 1975 [5]. He characterized a fractal as ‘a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole’. Mandelbrot introduced fractals as a new class of mathematical objects that represent nature. Many areas of science, such as physics, chemistry, and biology, soon recognized how powerful his ideas were. In mathematics, a new area, namely fractal geometry, came up and developed very quickly on the basis of geometric measure theory, harmonic analysis, dynamical systems, and ergodic theory. Fractal geometry can be viewed as an extension of the classical Euclidean geometry. Fractal dimension is a number associated with a fractal that can be used to compare one fractal to another. Every fractal has its own dimension, and this dimension can be measured by means of experiments. There are many ways to define the fractal dimension. Such terms as Hausdorff, Sandbox, box-counting and packing dimensions are widely used. Hausdorff dimension plays an important role in developing general mathematical theory of fractal dimension. In this paper also Hausdorff measure (HM) is applied and we discuss its tight connection with Hausdorff dimension. Box-counting dimension can be easily implemented in practice and calculated for the classical Sierpinski carpet as

$$\frac{\log(N)}{\log(3^r)}, \quad (1)$$

where N equals to the number of boxes filled with 1 and r stands for the level of ramification. These computations of the box-counting dimension for the Sierpinski carpet yield $D = \frac{\ln(8)}{\ln(3)} = 1.892789$. For the classical Sierpinski carpet, this dimension does not change, while the ramification is increased step by step. In [6], it was found in their study with planar tissue preparations in mice remarkably consistent scaling exponents (fractal dimensions) for tumor vasculature even among tumor lines that have quite different vascular densities and growth characteristics. These dimensions reveal important aspects of underlying mechanism of vascular growth in tumors. They find that fractals which model tumor vessels yield shapes with fractal dimensions of 1.89 ± 0.04 , whereas normal arteries and veins yield dimensions of 1.7 ± 0.03 . An extensive study of cancer risk assessment on simulated and real data for fractal-based cancer is given in [7]. Therein, we realized that deterministic fractals are too artificial models for tissue growth. Therefore, we shall concentrate on models of random carpets and Quermass-interaction processes, introduced in this paper in Sections 3 and 4. Indeed, it has been shown in previous investigations that the texture of mammary tissue, as seen at low magnification, may be characterized quantitatively in terms of stereology (see [2] and references therein). In [8], images of mammary cases were reexamined. In [9], a construction of a statistical test is given, which enables us to distinguish between the two groups. The distribution of box-counting dimension from histological images of mammary cancer and malignant tissues has been modeled in [10]. Fractal analysis was conducted concerning normality, depth, heterogeneity, and complexities of distributions for modeling fractal dimension of these images. Therefore, several tests have been used in order to have a better understanding of the data and its distributional deviations from normal, Weibull, or gamma samples. Kiselák *et al.* [11] discuss the relation between tissue growth and the mathematical model of diffusion. It is clear that a deterministic fractal is not an appropriate model for cancer growth (see also [7]). Stochastic fractals (in our area of applications, relevant references are, e.g., [12, 13], which relate to fractional Brownian motions; for more details see Section 6) are more appropriate; however, a validation measure should be developed for a better comparability with advanced stochastic geometry models, for example, Quermass-interaction processes. A relation between temperature and geometry of the tissue is studied in [11]. Therein, partial results on nonstandard diffusions for benign alterations and malignant tumors originating from glandular tissues (e.g., mammary, prostatic, or pancreatic) are given. For standard diffusions, fair approximation is provided by analytical models based on convective heat transfer in infinite tissues volume (e.g., model given by Perl in [14]).

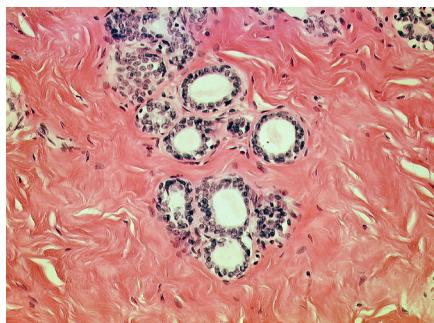
A simple and flexible model for tissue growth is to the best knowledge of the authors under this line of thought still missing. The other relevant contributions are given by, for example, Luebeck and Moolgavkar ([15–17], among others), or Kopp-Schneider ([18–20], among others). The particular aim of this paper is to fill this gap. The paper is organized as follows: the case study motivates the fractal

analysis as a tool for cancer research in Section 2. In Section 3, we introduce classical fractals and a flexible parametric model of random carpets. Randomization is introduced by usage of binomial random variables. Such a model has been developed to fit the typical slices from mammary cancer and mastopathy. We provide an algorithm for estimation of parameters of the model and analyze the goodness of fit by means of a simulation study. In Section 4, we search for the model in the area of stochastic geometry, and thus, we decide to model breast cancer by the Quermass-interaction process. This process is an extension of the Boolean model in the sense that the particles interact among each others. The interaction here is based on the Quermass integrals of the image, more precisely on volume, perimeter, and Euler Poincaré characteristic. Section 5 introduces an alternative method of simulation of random carpets. We illustrate theoretical and practical issues in generation of Sierpinski gaskets and HM calculations. Impact of ramification level is addressed. We also calculated the HM for both mastopathic and mammary cancer tissues. Mammary cancer discrimination via two-dimensional (2D) complex wavelet-based self-similarity measures is introduced in Section 6. Therein, again the relation between fractional Brownian motion, Hurst exponent, and Hausdorff dimension is recalled for relevant stochastic fractals. Section on Summary concludes the paper. Moreover, we introduce the R package FractalParameterEstimation (Appendix A), and we justify the usage of $[p-p-p-q]$ -model in contrast to other combinations of model parameters for four ramification steps among others as $[p-q-p-p]$ (Appendix B).

2. Motivating case study

2.1. Cases and sampling

Forty cases of human mammary tumors submitted for histopathological diagnosis were investigated. Twenty cases were fibrous mastopathies, that is, 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 (Figure 1(a), 1(b)). These were compared to twenty cases of invasive ductal mammary cancer, the most frequent type of breast cancer in humans (Figure 2(a), 2(b)). One paraffin section per case with a nominal thickness of 4 µm from the center 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 \times 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 tumor tissue were evaluated in the group of carcinomas at the same magnification by the same sampling strategy, that is, systematic sampling with a random start. The selected visual fields were transmitted to the image analysis system Kontron IBAS 2000 (Kontron Electronic Group, München, Germany) with a black-and-white charge-coupled device (CCD) camera. The result was a gray-level image with a resolution of 512 \times 512 pixels at a final magnification of 430 \times on the screen. By segmentation, a binary image was produced, which consisted of two phases only (Figure 17(a), 17(b)). All images were interactively segmented by the same person by tracing the epithelial formations. The

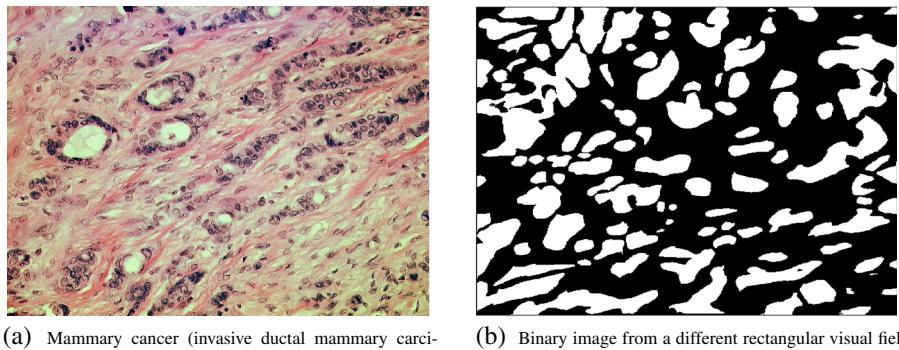


(a) Mastopathic tissue



(b) Binary image from a different rectangular visual field from a mastopathy

Figure 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) Binary image of a different rectangular visual field from a histological section from a mastopathy. The gray-level image has been reduced to a binary image by manual segmentation. White: epithelial cells (grain phase), black: gland openings, stroma, vessels, and so on (pore phase).



(a) Mammary cancer (invasive ductal mammary carcinoma)
(b) Binary image from a different rectangular visual field from a mammary carcinoma

Figure 2. (a) Mammary cancer (invasive ductal mammary carcinoma): the normal orderly glandular architecture has been replaced by irregular epithelial blocks with few stroma in between hematoxylin–eosin stain. (b) Binary image of a different rectangular visual field from a histological section from a mammary carcinoma. The gray-level image has been reduced to a binary image by manual segmentation. White: epithelial cells (grain phase), black: gland openings, stroma, vessels, and so on (pore phase).

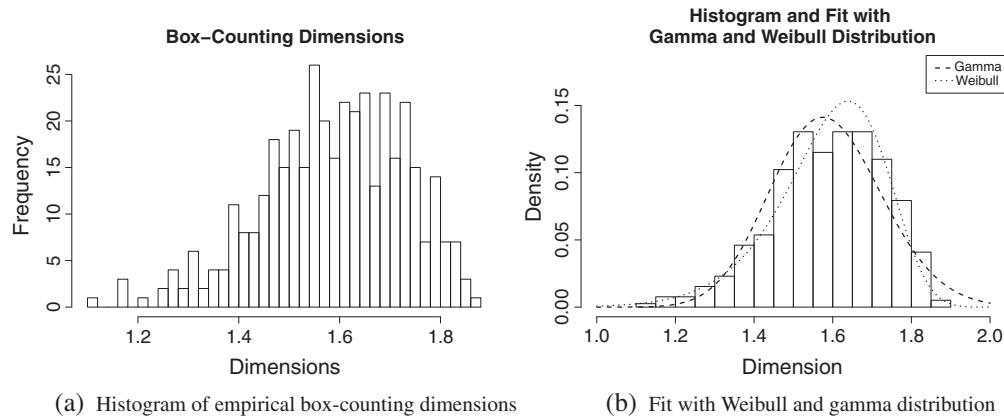


Figure 3. (a) Histogram of empirical box-counting dimensions and (b) fit with Weibull and gamma distributions.

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, and so on—was shown as black.

2.2. Goodness of fit

Among the 400 images, there were nine without epithelial phase (Figure 17(a)); for these images, no evaluation is possible. All remaining 391 histological images of mammary and mastopathic tissues have been used to calibrate the distribution of box-counting dimension. We have used IMAGEJ software (Appendix A.5) to determine the box-counting dimension of these images. Cancer data contain the observation number, box-counting dimensions, and the characteristics (mammary cancer or mastopathy). Within this sample, there are 199 observations with mastopathic and 192 with mammary cancer tissue. Naturally, the box-counting dimensions of these 391 binary images are between 1 and 2. The mean is 1.590, and standard deviation is 0.142. The lowest dimension is 1.104, and the highest dimension is 1.872.

The histogram plotted in Figure 3(a) shows a skewed to the left distribution, which is also shown by the value of skewness calculated as -0.47. The difference between the first and the third quartile (1.497 and 1.693) is another indication that the main part of the data lies within a small range of the dimensions. We assume gamma or Weibull distribution to fit for box-counting dimensions. Estimation of gamma distribution gives the shape parameter $\hat{\lambda} = 125.379$ and scale $\hat{\beta} = 78.98$. For Weibull distribution, shape parameter $\hat{\lambda} = 13.68$ and scale $\hat{\beta} = 1.648$ were obtained (by the method of moments). The fit of Weibull and gamma distribution on the box-counting dimension can be seen in Figure 3(b). The dashed line of Weibull distribution provides a slightly better fit compared with the dotted line of gamma distribution.

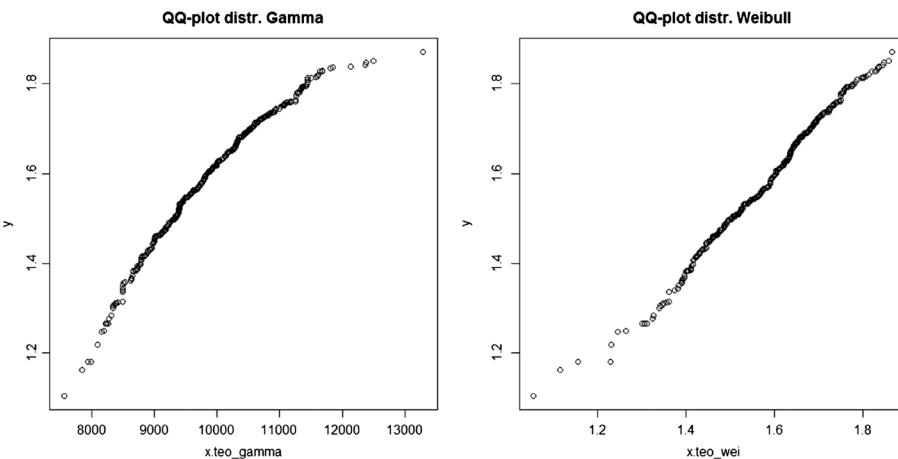


Figure 4. Q–Q plot for gamma and Weibull distributions.

Table I. Lower percentiles of empirical and Weibull distributions.

Lower percentiles	Empirical distribution		Weibull distribution	
	Value	Percentage of cancer	Value	Percentage of cancer
10th percentile:	1.3977	74.36	1.3987	75.00
Ninth percentile:	1.3868	74.29	1.3875	74.29
Eighth percentile:	1.3824	70.97	1.3750	72.41
Seventh percentile:	1.3656	70.37	1.3611	76.00
Sixth percentile:	1.3504	78.26	1.3454	77.27
Fifth percentile:	1.3139	73.68	1.3271	73.68
Fourth percentile:	1.3072	86.67	1.3051	84.62
Third percentile:	1.2750	90.91	1.2775	90.91
Second percentile:	1.2486	85.71	1.2397	100
First percentile:	1.1791	100	1.1780	100

Moreover, we applied a Kolmogorov–Smirnov test to investigate the goodness of fit [21]. Gamma distribution is not acceptable on a 95% significance level. The Q–Q plots show the results for the two distributions, plotted in Figure 4. The points in the Q–Q plot of Weibull distribution are nearly on a 45° degree line; in contrast to that, the result of the Q–Q plot of gamma distribution shows a slight curve and is above a 45° degree line.

2.3. Discriminating between tissues based on order statistics

Together with rank statistics, there are the order statistics among the most fundamental tools in nonparametric statistics and inference. For usage of order statistics, data are sorted in ascending order, the sample percentiles being the important cases. We have fitted the empirical cumulative distribution function to the data via percentages in Excel (Microsoft, Redmond, WA, USA). For Weibull distribution, certain percentiles are used. The results of the discrimination between mastopathic and cancer tissues are listed in Table I for lower and in Table II for upper percentiles.

In both cases, the percentage of cancer is high in lower percentiles and low in higher percentiles. By employing a generic observation for typical tissues in cancer research (e.g., [6]), we may conclude that in our study, the risk of cancer is higher for images with fractal dimension higher or equal to 1.75 than for images with fractal dimension of 1.39 and lower. Simple discrimination on the basis of box-counting dimension was performed in [10], which led to obvious differences between cancer and non-cancer tissues. Moreover, the distribution of box-counting dimension was analyzed separately for these two groups. Here, we observed that box-counting dimensions of mastopathy were on average higher compared with mammary cancer. We can conclude that one should be very careful in automatic application of the classical assumption that ‘the higher the dimension, the more cancerous the tissue is’. We can see that mastopathy provides higher complexity than cancer (see also Section 4 on Quermass-interaction process and Section 6 on complex wavelet-based self-similarity measures approach). Testing for normality with

Table II. Upper percentiles of empirical and Weibull distributions.

Upper percentiles	Empirical distribution		Weibull distribution	
	Value	Percentage of cancer	Value	Percentage of cancer
90th percentile:	1.7598	10.00	1.7524	13.04
91st percentile:	1.7744	8.33	1.7581	11.63
92nd percentile:	1.7802	9.38	1.7643	8.11
93rd percentile:	1.7897	10.71	1.7709	8.33
94th percentile:	1.7934	12.50	1.7783	8.57
95th percentile:	1.7982	15.00	1.7864	10.34
96th percentile:	1.8135	6.25	1.7958	13.04
97th percentile:	1.8173	8.33	1.8071	5.88
98th percentile:	1.8294	0.00	1.8216	0.00
99th percentile:	1.8424	0.00	1.8435	0.00

outlier detection or truncation of data (with respect to skewness and kurtosis) has shown obvious differences. Normal distribution of box-counting dimension of cancer tissue was not rejected in contrast to normal distribution of box-counting dimension of mastopathic tissue. These results motivate necessity for deeper analysis of the data.

3. Fractals: Sierpinski carpet

Sierpinski carpet will be the pattern model throughout the paper, especially for calculation of box-counting dimensions and for estimation of parameters. Sierpinski carpet is a fractal, which was firstly developed by the mathematician Waclaw Sierpinski in 1916. This fractal can have different structures and is built upon a special algorithm. The basis of this carpet is a square. Each side of the square window is divided into three equal parts, and by joining the opposite points, nine equal (smaller) squares arise. Afterwards, the middle square is removed, and the previously described procedure is applied on the other eight remaining squares, whereas the middle field is always removed. This mathematical model can then be used to model porous bodies, like cancerous tissue [22]. Mandelbrot [5] was the first who introduced fractals as a new class of mathematical objects. He claims that these objects have a rough or fragmented geometric shape, which can be split into parts, and are self-similar to each other. In other words, it could also be named a reduced-size copy of the whole object. Self-similarity plays an important role in these circumstances, but it differs from instance to instance. Per mathematical definition of fractals, the characteristic of self-similarity can be repeated infinitely [23]. In the beginning of cancer research, the disease was seen as a chaotic, poorly regulated growth. Therefore, the traditional Euclidean geometry is not fitting well for such shapes. One can try to measure the irregularity of tumor growth with the aid of fractal geometry as described in [6]. In [24], authors constitute that objects may not be defined as fractals, if self-similarity only holds over a limited range. Scientists coincide that this is an appropriate criticism, because the requirement of fulfilling fractals may then be used as a model.

3.1. Procedure and formation of Sierpinski carpets

At each level, also called ramification (r), a field is divided into nine commensurate parts, where $r \in N$. In further consequence, this already mentioned field, more precisely a matrix, needs to be specified more thoroughly. A program on the basis of recursion and arrays was developed in Java (Oracle Corporation500 Oracle Parkway, M/S 5op7Redwood Shores, CA 94065 United States) to create Sierpinski carpets and in further consequence, randomized Sierpinski carpets. Figure 5(a) displays the pattern for the first three ramifications. Previously named recursion is already visible in these three levels. Thus, the program works well, and the structure of Sierpinski carpet is visible in Figure 5(b) after five ramifications.

3.2. Dimensions of the matrices

The basis is one square from which a (3×3) matrix was created. Applying this procedure in the second step on all nine boxes of the existing matrix results in nine matrices with a dimension of (3×3) . In short, this means that after a process consisting of two steps, a dimension of (9×9) has been created from the origin. The mathematical equation for the dimension of rows and columns of the matrix behind this procedure is simply written as $\text{dim} = 3^r$. Potency is a positive integer, and the basis is constant, results

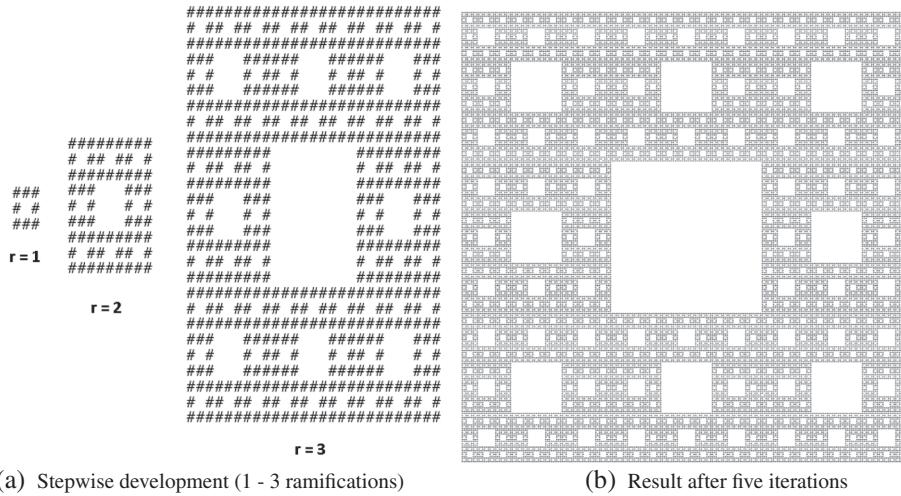


Figure 5. (a) Development of Sierpinski carpet for one to three ramifications and (b) the result after five steps.



Figure 6. Representations with $p \in \{0.4, 0.6\}$.

that the dimension increases with ascending ramification rapidly, for example, five and six ramifications lead to matrices with dimensions of (243×243) and (729×729) , respectively.

3.3. Random fractals

Random fractals do not completely fulfill the condition of self-similarity in order to be defined as a classical fractal. Nevertheless, the nonuniform appearance of such random fractals is quite similar to natural phenomena like coastlines or topographical surfaces. Falconer [25] claims that random fractals should display randomness at all scales. This means that a random element is part of the construction at every step. A very important fact in this definition is that a fractal can be named statistically self-similar only if it fulfills the requirement that ‘enlargements of small parts have the same statistical distribution like the whole set’[25, p.244]. The aim of the simulations in this section is to create a modified version of the Sierpinski carpet with the help of random variables. The basic matrix is identical to the original one, but the conditions of the single boxes are different. Indeed, the middle box of every matrix remains free as for the classical Sierpinski carpet. The random numbers result from a binomial distribution by usage of the inversion method. Thereby, the values of p stand for the probabilities of receiving zeros in the following described setting:

- (1) Generate a uniform distributed random number.
- (2) Compare it with a fixed value of p .
- (3) Let random variable = 1; if the value of the uniform-distributed random number is greater than the value of p , then let random variable = 0.

For the sake of demonstration of how this process is accomplished by the program, images for the first and second ramification with values of p of 0.4 and 0.6 are displayed in Figure 6.

Figure 7 presents the results of the simulation for five ramifications, given by a set of values of $p \in \{0.01, 0.05, 0.1, 0.2, 0.5, 0.8\}$, which clearly show the structure of the adapted Sierpinski carpet. Instead of the value 1, the symbol ‘#’ is shown, and all boxes that include zero remain free (white).

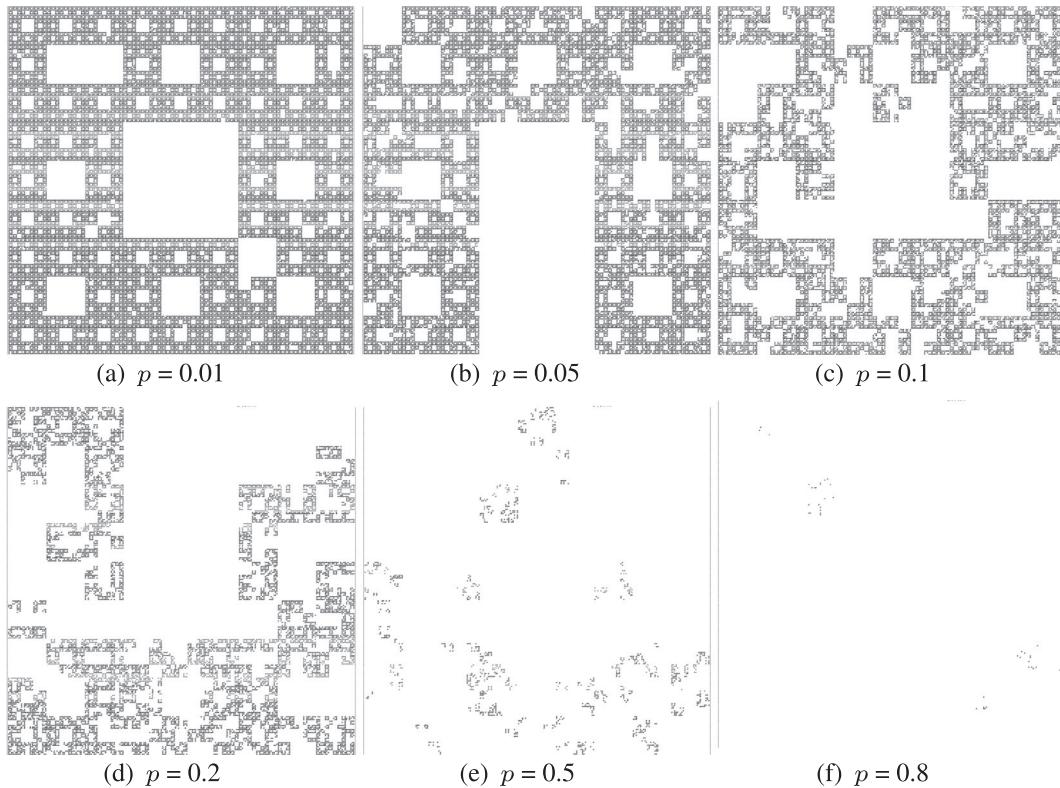


Figure 7. Comparison of random Sierpinski carpet with values of $p \in \{0.01, 0.05, 0.1, 0.2, 0.5, 0.8\}$.

By simple calculations, how many parts of the first matrix are expected to be filled with zeros is investigated. In this case, the expectation $E(X_i) = \sum_{i=1}^r x_i p(x_i)$ is used, where p represents the fixed probability of the simulation as previously introduced and $E(X_1)$, for example, equals the number of expected deleted fields at first ramification. One box (in addition to the middle box) of the matrix has to remain white (equal to 0) for given probability between 0.1 and 0.2. The results of these expected values for $\{0.01, 0.05, 0.1, 0.2, 0.5, 0.8\}$ are $\{0.08, 0.4, 0.8, 1.6, 4, 6.4\}$. Obviously, this calculation is an approximation, but the quality of simulation and random number computation in Java can be tested, which shows that the values mostly coincide. The deviations between the expected and actually deleted boxes in first ramification are $\{0.08, 0.6, 0.8, 0.4, 2, 0\}$. From the results of this short test, it can be assumed that the random numbers have the requested effect. Indeed, randomness is responsible for a deletion of at least two-ninths (the middle square is always removed) of the total area at a given value of p equal to 0.05. The opposite effect can be observed in the third plot of the first row in Figure 7 where 0 was not simulated, such that no box is deleted in the first step. Obviously, the impact of zeros is stronger because of the deletion of subjacent boxes. Only few boxes are represented with the symbol '#' in fifth ramification for higher values of p , such that computations with values of p higher than 0.8 are left out because of a (almost completely) white area.

3.4. Simulation with two parameters

The ramification was fixed to four, yielding dimensions for the matrices of (81×81) , and the set of possible values for the parameters p and q is $\{0.1, 0.3, 0.5, 0.6\}$. The structure of the simulation is the same, although in this setting, the random numbers are calculated three times with a value of p and once with value of q . Furthermore, the structure of the matrices differs, because all nine boxes of the matrices are filled with random numbers. A possible area of application of this simulation is medicine, where development of tissue of cancer patients can be simulated. With these two parameters, the following structure is given for the upcoming plots: $[p-p-p-q]$. In Appendix B, we discuss why this model is preferred to other combinations of the parameters p and q such as $[q-p-p-p]$. All 4×4 configurations (for instance, $[0.1-0.1-0.1-0.6]$ or $[0.5-0.5-0.5-0.1]$, ...) are possible, which enables in total 16 versions. The effect of this modification is visible in Figure 8. A new simulation was started for every plot, and therefore, the basic matrices differ. The increase of the parameter q from 0.1 to 0.6 for a given p of 0.1 is shown in

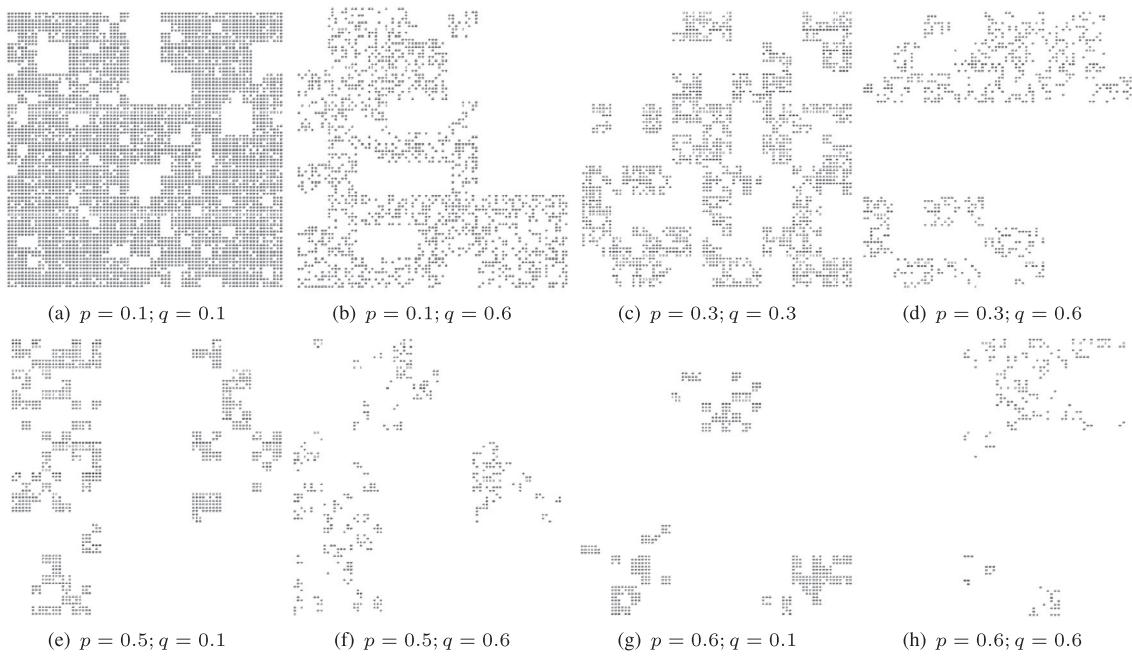


Figure 8. Comparison of random Sierpinski carpets with values of $p \in \{0.1, 0.3, 0.5, 0.6\}$ and $q \in \{0.1, 0.3, 0.6\}$.

the plots in Figure 8(a) and 8(b). This increase leads to an evidently higher rate of deleted boxes. The higher probability for occurrence of 0 causes that in the last level, fewer boxes are filled with ones, and consequently, the thickness of the symbols ‘#’ is different between the two plots.

By comparing the plots in Figure 8(c) and 8(d) in the first row, the impact of a change of q from 0.3 to 0.6 for a given value of p of 0.3 can be seen. The difference was stronger in the previous example, but the main effect of q is also visible. Hence, the value of p is fixed with 0.5, and the image for the two values of q , 0.1 and 0.6, is displayed in Figure 8(e) and 8(f) in row two. Because of the relatively high value of p , a big part of the matrix is deleted, and the additional increase of the q value from 0.1 to 0.6 yields a matrix with few entries. The last two plots in Figure 8(g) and 8(h) were done with a value of p of 0.6 and already known q values 0.1 and 0.6. Because of the high value of p , only a manageable number of ‘#’ is remaining. The additional increment of q leads to a dissolution of the groups, and as a result, only little isolated black areas can be seen. Estimators of the parameters p and q are not independent; this is intuitively clear from their geometrical interpretation, and it can be empirically checked, for example, by correlation coefficient.

3.5. Parameter p and q estimation

The aim of this section is to develop a method and a program to estimate the parameters p and q . A modification of the estimated values is facilitating with known data. The values of the parameters p and q result from the set $\{0.1, 0.3, 0.5, 0.6\}$. The program for the estimation has been implemented in software R [26], and the package FractalParameterEstimation (Appendix A) was developed.

3.5.1. Description of the program. First of all, the ramification of the original matrix has to be computed, which equals 4 in our setup. In order to estimate the parameters independently, a matrix with a size of lower ramification has to be generated and filled. In the next step, the dimension of the smaller matrix can be calculated and created with this value. Two loops run through every box of the two matrices and count the quantity of 1. If this value is at least 1, the cell of the smaller matrix will be filled with 1, otherwise with 0. This procedure is performed until every box of the smaller matrix contains either 1 or 0. Two further loops enable to set the previously defined value in every box of the smaller matrix. If one part of the matrix is 0, then every box of the resulting matrix will also be zero. Exactly, this feature proves that if the value 1 occurred in one box of the matrix, the matrix ‘above’ (smaller matrix) would also have to be 1. Otherwise, no random numbers were calculated, and every box is set to 0.

3.5.2. Computing estimators \hat{p} and \hat{q} . After imputing values in the smaller matrix, estimators for the parameters p and q can be calculated. Equation (2) shows how this estimation is performed, where '#' stands for frequency, m defines the matrix, and the subscript r represents the ramification of the matrix of interest:

$$\hat{p} = \frac{\# 0 \text{ of } m_{r=3} - \# 0 \text{ of } m_{r=2} \cdot 9}{\# 1 \text{ of } m_{r=2} \cdot 9} \quad \hat{q} = \frac{\# 0 \text{ of } m_{r=4} - \# 0 \text{ of } m_{r=3} \cdot 9}{\# 1 \text{ of } m_{r=3} \cdot 9} \quad (2)$$

In order to test the quality of the estimation, all combinations of the values of p and q were used, and their estimators were compared with the true value of the parameters. A simulation study with 50 computed images for every combination of the parameters has been performed. The differences between estimators and the original values were computed. We observed small deviations between the estimators and their true values, not exceeding 0.15 for in total 800 ($50 \cdot 16$) simulations. The study has shown that estimates are closer to the true value of the parameter by only observing the impact of one ramification. Another important aspect is that estimates are less precise when estimated from earlier ramification, that is, for [p - p - p - q] model, estimates of p for third ramification are better than those of second ramification. Figure 9 gives a precise overview of the differences, where one can see that with increasing value of either p or q , the deviation increases too.

In Table III, estimates for the parameters of interest, p and q , are listed for each of the 16 possible combinations with the set of values $\{0.1, 0.3, 0.5, 0.6\}$, computed on the basis of (2) for [p - p - p - q] model. The mean estimates for p (column 2) and q (column 5) are on average very close to the true value, displayed on their right-hand side. The largest discrepancy can be observed in the setup for p and q both

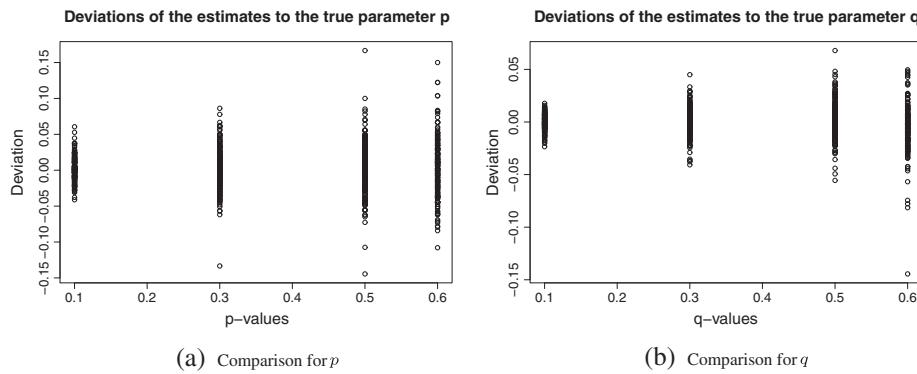


Figure 9. Plot of mean estimator against original values of p (a) and q (b).

Table III. Simulation study for [p - p - p - q] with given parameter estimates and dimensions.

p	Mean (\hat{p})	Δ_p	q	Mean (\hat{q})	Δ_q	Mean (dim)	SD (dim)
0.10	0.0991	0.0009	0.10	0.0996	0.0004	1.9048	0.0105
0.10	0.0980	0.0020	0.30	0.3008	-0.0008	1.8481	0.0099
0.10	0.1028	-0.0028	0.50	0.4994	0.0006	1.7683	0.0155
0.10	0.1085	-0.0085	0.60	0.5946	0.0054	1.7159	0.0151
0.30	0.2998	0.0002	0.10	0.1005	-0.0005	1.7321	0.0255
0.30	0.3029	-0.0029	0.30	0.3002	-0.0002	1.6720	0.0239
0.30	0.3027	-0.0027	0.50	0.5014	-0.0014	1.5949	0.0283
0.30	0.3089	-0.0089	0.60	0.5963	0.0037	1.5400	0.0434
0.50	0.5065	-0.0065	0.10	0.1010	-0.0010	1.4929	0.0325
0.50	0.5066	-0.0066	0.30	0.3009	-0.0009	1.4351	0.0401
0.50	0.5050	-0.0050	0.50	0.5069	-0.0069	1.3482	0.0909
0.50	0.4992	0.0008	0.60	0.5908	0.0092	1.3091	0.0841
0.60	0.5972	0.0028	0.10	0.0992	0.0008	1.3534	0.0525
0.60	0.6036	-0.0036	0.30	0.2989	0.0011	1.2851	0.0579
0.60	0.6074	-0.0074	0.50	0.4984	0.0016	1.1953	0.0907
0.60	0.6086	-0.0086	0.60	0.5944	0.0056	1.1449	0.0976

SD, standard deviation.

equal to 0.6, but the mean difference is not larger than 0.009 in absolute terms for any combination. Although larger variation in the estimates can be observed for higher parameter values, the mean estimates are very close to the true parameter without exceptions.

3.6. Comparison of box-counting-dimension for [p-p-p-q] model of randomized Sierpinski carpets

The fractal dimension of the classical Sierpinski carpet does not change while the ramification is increased step by step. The reason for that is the pattern of the basic matrix, where the middle box always remains free. This pattern is repeated in every following matrix, leading to constant box-counting dimension for any ramification. Hence, it can be calculated for the initial form ((3×3) matrix) and used further. The box-counting dimension of a random fractal differs from the dimension of the classical Sierpinski carpet according to the size of the parameters p and q , because of the introduction of binomial distributed random variables in every ramification. We used the usual set of values as $\{0.1, 0.3, 0.5, 0.6\}$ for this computation, because this range is sufficient in cancer research (compare Figure 8 with Figure 1 or Figure 2). Equation (3) was used to compute box-counting dimension for simulated images as follows:

$$\dim = \frac{\log(3^8 \cdot (1 - \hat{p})^3 \cdot (1 - \hat{q}))}{\log(3^4)}, \quad (3)$$

where the expected number of ‘remaining boxes’ is divided by the total number of boxes as explained by (1) for classical fractals. One obtains the expectation of random fractals by multiplying the total number of boxes, $(3^2)^4$, with the probability of receiving 1 in one ramification step as $(1 - \hat{p})^3$ for three iterations of p respectively $1 - \hat{q}$ for last ramification with q . Furthermore, the effects of changes of the parameters are also noticeable with these values. We have observed from computations that dimension will decrease if the parameter p or q increases. This effect is illustrated for the parameter p as well as q in Table III. Mean and standard deviations are given in columns 7 and 8 for the 50 simulated images per combinations of p and q .

3.7. Mastopathic and mammary cancer tissue estimation results for [p-p-p-q] model

We applied the estimation procedure of [p-p-p-q] model on both mammary cancer and mastopathy images. Therefore, the mean estimates were computed on the basis of the 10 images available for every patient of both groups. The overall means for p and q already show differences in terms of values between the groups. Mammary cancer images were estimated on average with parameters 0.1428 (p) and 0.0554 (q), respectively. The obtained mean parameter estimates for mastopathy are smaller and equal 0.0533 (p) and 0.0191 (q). Figure 10 gives a graphical overview where mean estimates for p are displayed on the left-hand side and q on the right-hand side. Therein, mammary cancer patients are plotted with circles, and mastopathy patients are represented by crosses. Obviously, the estimates are higher for cancer group compared with those of mastopathy for almost all patients. Moreover, we have computed standard deviations for both groups and obtained means of 0.0488 (p) and 0.0212 (q) for mammary cancer and 0.0374 (p) and 0.0142 (q) for mastopathy, respectively. Statistically significant differences are computed with the aid of nonparametric tests as Kolmogorov–Smirnov and Wilcoxon tests. Moreover, a

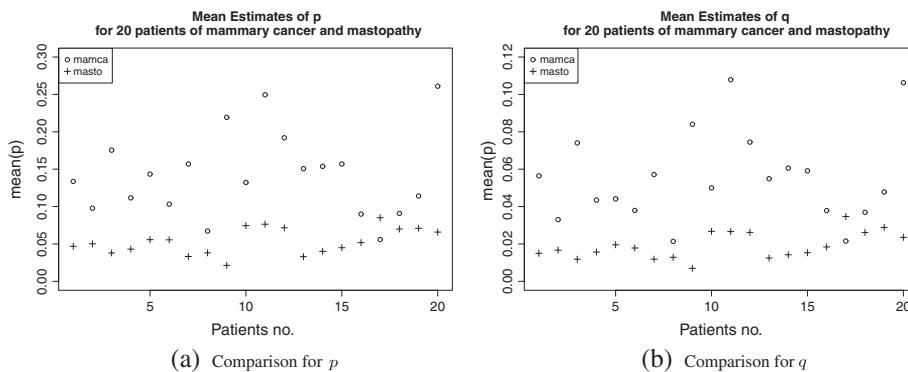


Figure 10. Estimations for real images of mammary cancer and mastopathy for 20 patients, each represented by 10 images, are compared. Circles represent mammary cancer and crosses mastopathy estimates.

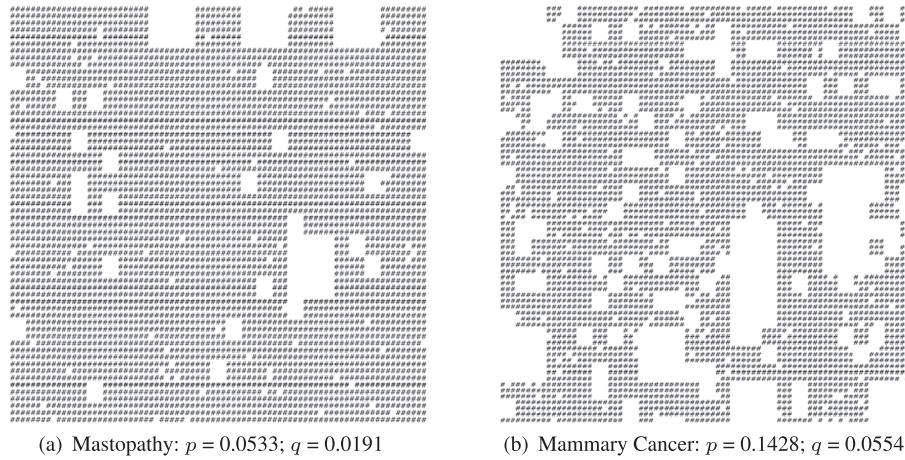


Figure 11. Simulations of mastopathic and mammary cancer tissue image with estimated means of p and q on the basis of $[p-p-p-q]$ -model.

two-sample t -test has been conducted because of the fact that the estimates can be seen as normally distributed (checked by Shapiro test), yielding the same results as the previous tests on the basis of a 95% significance level.

In Figure 11, we provide simulation results with previously estimated means of p and q for both groups of interest. Notice that the differences in the parameters between mastopathy and mammary cancer are obvious in the simulated tissues too. Recall that higher estimated parameters of interest lead to more and stronger white areas of the images, leading to a good graphical differentiation on the basis of estimated mean parameters.

4. Quermass-interaction process

Consider a planar random set \mathbf{X} given by a union of interacting discs with random radii and centers randomly scattered in a bounded planar region $S \subset \mathbb{R}^2$. For any finite configuration $\mathbf{x} = (x_1, \dots, x_n)$ of the discs x_1, \dots, x_n , the set \mathbf{X} is described by a density $f_\theta(\mathbf{x})$ with respect to the probability measure of a stationary random-disc Boolean model with intensity ρ of disc centers and the distribution Q of the disc radii. Suppose that the density is of the form

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta \cdot T(U_{\mathbf{x}})\}, \quad (4)$$

where c_θ is a normalizing constant, $\theta = (\theta_1, \theta_2, \theta_3)$ is a vector of parameters, \cdot denotes the dot product, and $T(U_{\mathbf{x}}) = (A(U_{\mathbf{x}}), L(U_{\mathbf{x}}), \chi(U_{\mathbf{x}}))$ is a vector of geometrical characteristics of the union $U_{\mathbf{x}}$ composed of the discs from the configuration \mathbf{x} , where A is the total area, L the perimeter, and χ the Euler-Poincaré characteristic [27].

The interpretation of the parameters is such that positive values of them force the process \mathbf{X} to have larger values of corresponding geometrical characteristics than the reference Boolean model, while in the case of negative parameters, the corresponding geometrical characteristics are smaller in realizations of \mathbf{X} than in the reference process. The combinations of positive and negative values of all three parameters then allow us to describe a really wide group of random sets. For more details about the model properties, see [28–30].

4.1. Estimating parameters of Quermass-interaction process

Several methods for estimating the parameters were developed. Method of maximum likelihood estimate with usage of MCMC simulations (in the following text denoted as MCMC MLE, see [31]) which is described in the following paragraph was first applied to extended Quermass-interaction processes in [32]. It was then improved for cases with high dimensional parameter by dimension reduction in [30]. In [33], Takacs–Fiksel method was studied, which provides similarly satisfactory results compared with MCMC MLE. Here, we concentrate on MCMC MLE, while we use its basic version [32] because only a three-dimensional parameter is to be estimated.

Observing a union U_x of the discs \mathbf{x} , that is, a realization of the random set \mathbf{X} , the method is based on maximizing the log-likelihood function

$$l(\theta) = \log f_\theta(\mathbf{x}) = \theta_1 A(U_x) + \theta_2 L(U_x) + \theta_3 \chi(U_x) - \log c_\theta.$$

However, c_θ has no explicit expression; therefore, the likelihood ratio f_θ/f_{θ^0} for a fixed θ^0 is maximized instead. Because denoting

$$h_\theta(\mathbf{x}) = \exp\{\theta_1 A(U_x) + \theta_2 L(U_x) + \theta_3 \chi(U_x)\},$$

we can use importance sampling [31] to approximate the log-likelihood ratio as

$$\begin{aligned} \log \frac{f_\theta}{f_{\theta^0}} &= l(\theta) - l(\theta^0) = \log(h_\theta(\mathbf{x})/h_{\theta^0}(\mathbf{x})) - \log(c_\theta/c_{\theta^0}) \\ &\approx \log(h_\theta(\mathbf{x})/h_{\theta^0}(\mathbf{x})) - \log \frac{1}{R} \sum_{i=1}^R h_\theta(\mathbf{z}_i)/h_{\theta^0}(\mathbf{z}_i), \end{aligned} \quad (5)$$

where \mathbf{z}_i are realizations from f_{θ^0} obtained by MCMC simulations and R is a given number of that realizations. Denoting the approximation (5) as

$$l_{\theta^0}(\theta) = \log(h_\theta(\mathbf{x})/h_{\theta^0}(\mathbf{x})) - \log \frac{1}{R} \sum_{i=1}^R h_\theta(\mathbf{z}_i)/h_{\theta^0}(\mathbf{z}_i),$$

the aim is to maximize

$$\begin{aligned} l_{\theta^0}(\theta) &= (\theta_1 - \theta_1^0) A(U_x) + (\theta_2 - \theta_2^0) L(U_x) + (\theta_3 - \theta_3^0) \chi(U_x) \\ &\quad - \log \frac{1}{R} \sum_{i=1}^R \exp \left\{ (\theta_1 - \theta_1^0) A(U_{\mathbf{z}_i}) + (\theta_2 - \theta_2^0) L(U_{\mathbf{z}_i}) + (\theta_3 - \theta_3^0) \chi(U_{\mathbf{z}_i}) \right\}. \end{aligned} \quad (6)$$

For more details about this method, see [30, 32].

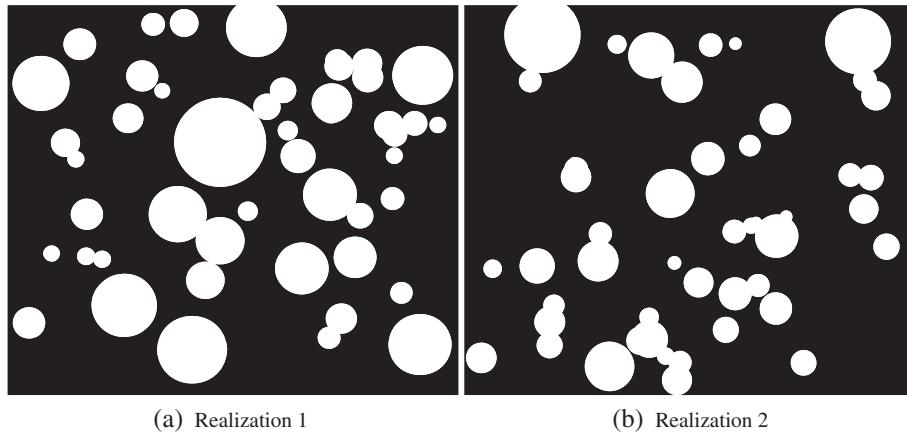
4.2. Numerical results

In [8], Mrkvička and Mattfeldt modeled the images by a Boolean model of discs, and each image was then tested on the compatibility with the Boolean model using the rather powerful method introduced first in [34]. The results of the compatibility with the Boolean model were not satisfactory. Only 14 out of 200 images of mammary cancer and 2 out of 200 images of mastopathy were compatible with the Boolean model on the significance level of 0.05. This clearly showed that cancer tissue is significantly closer to the complete spatial randomness described by Boolean model than mastopathic tissue. In other words, it showed that mastopathic tissue is significantly more structured than cancer tissue. Indeed, for modeling cancer tissue, this result is unsatisfactory; therefore, we will search here for another model of cancer tissue.

Every black–white image with the resolution of 512×512 pixels was fitted by the Quermass-interaction model. The parameters of the model were estimated by MCMC MLE as described earlier. The compatibility of images with Quermass-interaction process was tested by the same method as in [34]. The compatibility was not rejected for 55 images of mammary cancer and for 12 images of mastopathic tissue with a significance level of 0.05. According to the high power of the chosen test, it seems that the Quermass-interaction process is capable to model many variations of mammary cancer images. We can also observe that mammary cancer tissues reveal less deviation from Quermass-interaction process than mastopathic tissues. Table IV shows the averages and standard deviations of estimated parameters of Quermass-interaction process for 55 mammary images, which were proven to be compatible with the Quermass-interaction process. Notice that values of volume, perimeter and Euler number have no dimension. The interpretation of resulted averaged parameters is the following: by the negative volume parameter says that the image prefers smaller volume of union of particles—which means more overlapping than in Boolean model. The positive perimeter parameter says that the image prefers greater perimeter of the union of particles—which means more structured clusters of particles than in Boolean

Table IV. The averages and standard deviations of estimated parameters for selected 55 images of mammary cancer.

Statistics	Estimated volume $\hat{\theta}_1$	Estimated perimeter $\hat{\theta}_2$	Estimated Euler number $\hat{\theta}_3$
Mean	-1.16	0.74	-2.15
Standard deviation	1.23	0.85	1.15

**Figure 12.** Two realizations of the Quermass-interaction process with average estimated parameters of mammary cancer tissue.

model. The negative Euler number parameter says that the image prefers smaller Euler number—which means more holes in the union of the particles than in Boolean model. Figure 12 shows two realizations of the Quermass-interaction process with average estimated parameters.

5. Hausdorff measures of Sierpinski gaskets

In this section, we introduce the HM calculation both on random fractals and real tissues. The term HMs [35] is used for a class of outer measures (introduced by Hausdorff for the first time in [36]) on subsets of a generic metric space (X, d) , or for their restrictions to the corresponding measurable sets.

Let (X, d) be a metric space. For any subset $E \subset X$, we will denote $\text{diam}(E)$ as the diameter of E .

Definition 1

For any $E \subset X$, any $\delta \in [0, \infty)$, and any $\alpha \in [0, \infty)$, we consider the outer measure

$$\mathcal{H}_{\delta}^{\alpha}(E) := \omega_{\alpha} \inf \left\{ \sum_{i=1}^{\infty} (\text{diam } E_i)^{\alpha} : E \subset \bigcup_i E_i \quad \text{and} \quad \text{diam}(E_i) < \delta \right\}, \quad (7)$$

where $\omega_{\alpha} := \frac{\pi^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2} + 1)}$ is a positive factor. When α is a (positive) integer, ω_{α} equals to Lebesgue measure. However, some authors set $\omega_{\alpha} = 1$ (for instance, [37]).

The $\mathcal{H}_{\delta}^{\alpha}(E)$ defined in the preceding text are outer measures, and they are called Hausdorff premeasures by some authors. Moreover, the infimum can be taken over open or closed coverings without changing the result.

The map $\delta \mapsto \mathcal{H}_{\delta}^{\alpha}(E)$ is monotone nonincreasing, and thus, we can define the Hausdorff α -dimensional measure (or Hausdorff α -dimensional outer measure) of E as

$$\mathcal{H}^{\alpha}(E) := \lim_{\delta \downarrow 0} \mathcal{H}_{\delta}^{\alpha}(E), \quad (8)$$

In this paper, we applied Hausdorff dimension of fractals in its simplest computational form, that is, box-counting dimension. It is important to recall that one of several possible equivalent definitions of the Hausdorff dimension directly relates to HM, that is,

$$\dim_{\text{Haus}}(S) = \inf\{d \geq 0 : H^d(S) = 0\} = \sup(\{d \geq 0 : H^d(S) = \infty\} \cup \{0\}),$$

where we take $\inf \emptyset = \infty$.

In this section, we firstly gauge the Sierpinski gasket with edge division $r = 2$, that is, compression ratio = 1/2, with the random choice loaded. The diameter stipulation of such a kind of set R within the text is based on the genetic algorithm. Considering a metric as the Euclidean distance between two points of a set R , we have well-defined diameter of the set. It is easy to see that the diameter of Sierpinski gasket without random choice (Figure 13) equals to the length of the source triangle.

5.1. What is a genetic algorithm?

The genetic algorithm, a self-adaptive and global optimization probability search algorithm inspired by evolution, was introduced and investigated by John Holland (1975) [38] and his students. Let us denote by F_i the Sierpinski gasket of the i -th ramification, being F_0 an equilateral triangle in the Euclidean plane with the edge length $l = 1$. It is clear that $\dots \subset F_n \subset \dots \subset F_2 \subset F_1 \subset F_0$. Moreover, the nonempty set $\cap F_i$ is the Sierpinski gasket of the Hausdorff dimension $\log_2 3 \approx 1.58$.

The particular numbers are allocated to the triangles in a counterclockwise direction within the ramification (Figure 14). Encoding and decoding of the position within the gasket of arbitrary ramification level is performed by using the so-called position mappings. S_1 , S_2 , and S_3 , pointing to the position of the smaller gasket triangle within the greater one (the one of the previous iteration), can be written as follows:

$$S_1(x, y) = (x, y), S_2(x, y) = (x, y) + (l/2, 0), S_3(x, y) = (x, y) + (l/4, l\sqrt{3}/4). \quad (9)$$

So, the mapping S_1 points to the lower-left triangle, the mapping S_2 points to the lower-right triangle, and the mapping S_3 points to the upper triangle of the particular ramification. Accordingly, the position

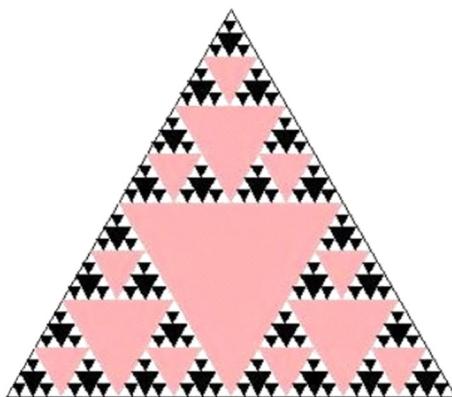


Figure 13. Sierpinski gasket: the fifth ramification with 121 triangles.

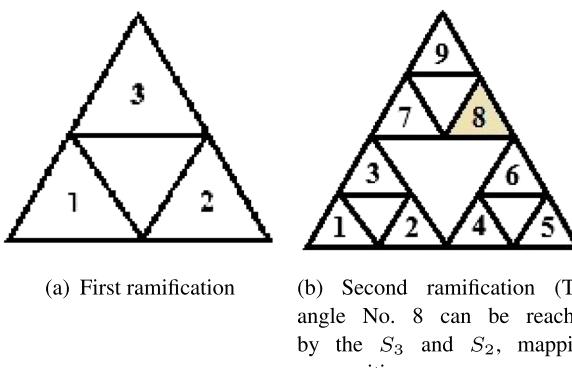


Figure 14. Numbering of triangles within the first and second ramification (counterclockwise). (a) First ramification. (b) Second ramification (triangle no. 8 can be reached by the S_3 and S_2 , mapping composition).

chain of the small triangle in the gasket is given by the mapping composition, that is, triangle no. 8 in Figure 14(b) can be reached by $S_3 \circ S_2$, triangle No. 4 by $S_2 \circ S_1$, triangle No. 5 by $S_2 \circ S_2$, and so on. The number of mappings in the composition is equal to the ramification.

5.1.1. Encoding/decoding. Finally, every small triangle of the m -th ramification either is or is not chosen in each random choice. This fact is represented by assigning 1 for chosen and 0 for non-chosen triangle. A simulation on this example for a representation of a tissue can be seen in Figure 15. Consequently, the entire code of the gasket evolved to the m -th ramification level consists of 3^m digits, therefore 3^m zeros or ones, belonging to each related triangle. We denoted such a set by R . After the decoding procedure, the diameter of R is stipulated, and the HM of R is calculated by using

$$H(R) = \min_{R \in F_m^R} \text{diam}(R)^s / \mu(R), \quad (10)$$

where F_m^R is a set of elements R matching with the gaskets of m -th ramification on which the random choice was applied, $\mu(R)$ is the density of the R gasket (numbers of ones and numbers of all digits ratio), and s is the Hausdorff (box-counting) dimension.

There were 3^{m+1} measurements, that is, 3^{m+1} random choices performed on each ramification from 1 to 13 for the sake of gaining appropriate amount of data. An additional parameter p for probability of occurrence of 1 within the random choice was involved to make the model more realistic. The results for HM together with related CPU times are given in Table V. CPU time increases rapidly with the level of ramification, wherefore computation on level 13 lasts about 10 h.

5.2. Hausdorff measure calculations

5.2.1. Computations of the random Sierpinski carpet. We have used the simulated images in order to compute HM for different parameters of p and q . Therefore, we provide the calculations of obtained HMs

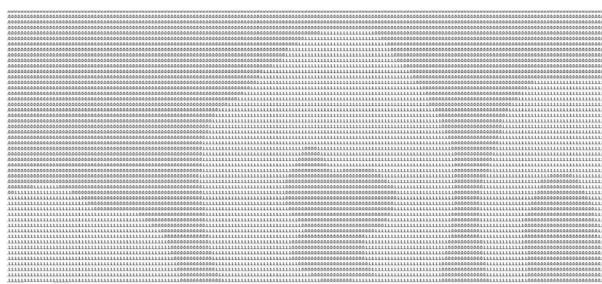


Figure 15. Picture of a tissue represented by zeros and ones.

Table V. Hausdorff measures and diameters for the Sierpinski gasket (second to 13th ramification) with random choices of probability p .

r	CPU(s)	HM, $p = 1/2$	$d(U), p = 1/2$	HM, $p = 3/4$	$d(U), p = 3/4$	HM, $p = 7/8$	$d(U), p = 7/8$
2	0.8593	0.8093	0.8750	0.8093	0.8750	0.8093	0.8750
3	0.9218	0.9028	0.9375	0.8191	0.8817	0.9028	0.9375
4	1.1406	0.9052	0.9391	0.8899	0.9052	0.9509	0.9688
5	1.4531	0.9515	0.9691	0.9531	0.9034	0.9267	0.9531
6	2.9063	0.9304	0.9555	0.9153	0.9457	0.9388	0.9609
7	7.9687	0.9692	0.9805	0.9634	0.9768	0.9570	0.9727
8	25.406	0.9846	0.9903	0.9663	0.9786	0.9724	0.9825
9	83.515	0.9877	0.9922	0.9802	0.9875	0.9693	0.9805
10	265.51	0.9846	0.9902	0.9784	0.9863	0.9800	0.9873
11	862.75	0.9838	0.9897	0.9877	0.9922	0.9896	0.9934
12	3269.2	0.9944	0.9965	0.9975	0.9984	0.9882	0.9926
13	37145	0.9955	0.9971	0.9962	0.9962	0.9894	0.9933

HM, Hausdorff measure.

Table VI. Hausdorff measure of random Sierpinski carpets of fourth ramification [p - p - p - q] with various p and q values.

p	q	$\text{diam}(U)$	$\mu(U)$	Haus. dimension	Haus. Measure
0.1	0.1	0.736	0.446	1.9041	1.2508
0.1	0.3	0.978	0.465	1.8469	2.0640
0.1	0.5	1.375	0.742	1.7703	2.3683
0.1	0.6	1.414	0.763	1.7196	2.3778
0.3	0.1	1.414	0.651	1.7325	2.7995
0.3	0.3	1.389	0.657	1.6753	2.6394
0.3	0.5	1.414	0.814	1.5988	2.1375
0.3	0.6	1.414	0.846	1.5480	2.0208
0.5	0.1	1.414	0.912	1.5028	1.8453
0.5	0.3	1.414	0.901	1.4456	1.8136
0.5	0.5	1.414	0.943	1.3691	1.7045
0.5	0.6	1.414	0.970	1.3183	1.6272
0.6	0.1	1.414	0.978	1.3505	1.6326
0.6	0.3	1.414	0.975	1.2933	1.6055
0.6	0.5	1.414	0.973	1.2167	1.5662
0.6	0.6	1.414	0.970	1.1660	1.5445

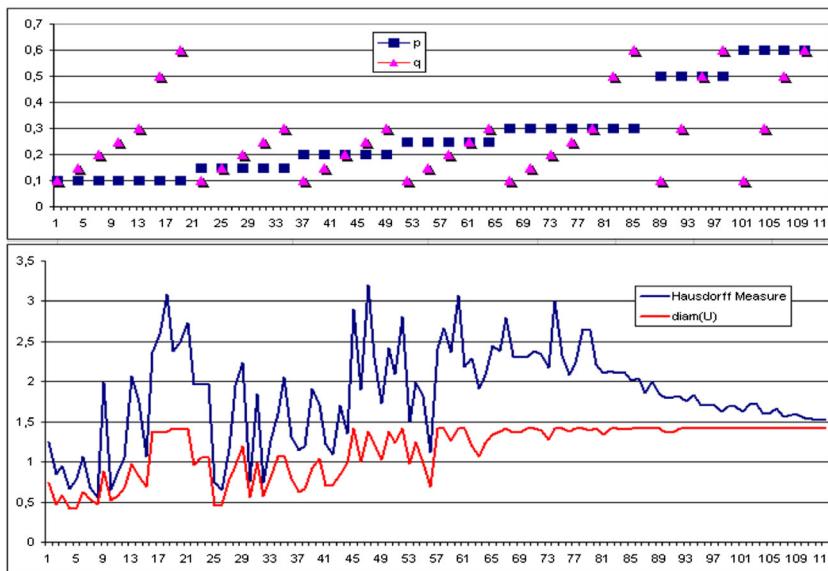


Figure 16. Hausdorff measure calculations on simulations of [p - p - p - q] model and parameter values of p and q from $\{0.1, 0.15, 0.2, 0.25, 0.3, 0.5, 0.6\}$.

for values of p and q given in Table VI. HM of the random Sierpinski carpet depends on the diameter and the density of the domain (as shown on Gasket in Section 5). But both of them (for a random Sierpinski carpet of the fourth ramification of [p - p - p - q] type) are influenced by the values of p and q (Figures 7 and 8). As the value of p and q increase, the ascendant trend of the HM (in accordance with diameter) can be observed. While the diameter reaches its maximum D_m (or the values oscillate slightly below D_m), Hausdorff dimension decreases towards 1; because of continuing increase of the density values, the descending of the HM initiates. HM tends to D_m as visible in Figure 16.

5.2.2. HM computations on tissues. Figure 17(a) shows 10 mastopathic tissues in black and white color representation. The black regions represent holes, whereas the white ones are the surroundings. The task is to count the number and stipulate the diameter of the holes in tissue on each slice. The diameter is given as the maximal distance between arbitrary two points of one hole. All holes of each image are measured in this way. The density $\mu(R)$ of mastopathic tissue, as mentioned before, is the ratio of the black color to the whole picture (Table VII). By the same procedure, mammary cancer tissue is treated (Figure 17(b))

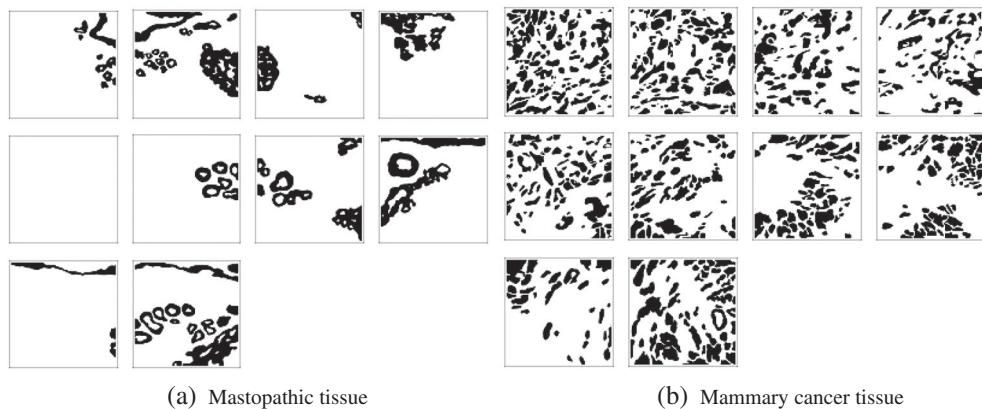


Figure 17. Mastopathic (a) and mammary cancer (b) tissue picture of which the Hausdorff measure is estimated.

Table VII. Hausdorff measure of mastopathic tissues calculated by using (10).

No.	# of holes	diam(U)	$\mu(U)$	Haus. dimension	Haus. Measure
1	12	0.3750	0.0658	1.5636	3.2770
2	33	0.7029	0.2337	1.7669	2.2947
3	16	0.4448	0.0826	1.6000	3.3119
4	11	0.7100	0.1296	1.6722	4.3517
5	—	—	—	—	—
6	11	0.3098	0.0692	1.5716	2.2909
7	19	0.3877	0.1478	1.6934	1.3593
8	13	0.9965	0.2033	1.7444	4.8896
9	2	1.0017	0.0503	1.5204	19.9452
10	29	1.0063	0.2633	1.7860	3.8404

Table VIII. Hausdorff measure of mammary cancer tissues calculated by using (10).

No.	# of holes	diam(U)	$\mu(U)$	Haus. dimension	Haus. Measure
1	103	0.4257	0.3609	1.8365	0.5774
2	94	0.3664	0.3356	1.8249	0.4769
3	63	0.4308	0.2672	1.7883	0.8301
4	62	0.5176	0.2109	1.7504	1.4971
5	78	0.2832	0.2661	1.7876	0.3940
6	61	0.3402	0.2655	1.7873	0.5484
7	59	0.2590	0.2759	1.7935	0.3214
8	65	0.1977	0.2561	1.7815	0.2176
9	30	0.2806	0.1795	1.7245	0.6228
10	58	0.4530	0.3821	1.8457	0.6068

and Table VIII). Tables VII and VIII provide quantitative information about mastopathic (Table VII) and mammary cancer (Table VIII) tissue dissimilation within the monitored 2D section pattern. The measurable entries are performed for the sake of rating the layout of the tissue arising (R) within the parent one: number of coherent circumscribed ‘holes’ in parent tissue filled by the mastopathic/mammary cancer tissue equals the number of elements of R , diameter (diam(R)) in fact performs the diameter of the greatest hole in the picture (herein, the diameter stands for the maximal distance of two arbitrary points within one ‘hole’), density $\mu(R)$ stands for the ratio between mastopathic (or mammary cancer) tissue to the entire image area, and HM of the pattern depicted on particular picture represents the ratio between diam(R) and $\mu(R)$. The remarkable value 19.9452 in Table VII for HM results from the combination of a small amount of little areas (two holes; see first the plot of the third row in Figure 17(a)), a large enough diameter (more than 1) and density which has tiny values because of a long and thin tissue.

6. Mammary cancer classification using complex wavelet-based self-similarity measures

In high-frequency and irregular data collected in real-life settings, a commonly occurring phenomenon is that of regular scaling. Examples of this have been found in a variety of systems and processes including several applications in biology and medicine. The irregular behaviors of these complex structures are difficult or impossible to quantify by standard modeling techniques. However, when observations are inspected at different scales, there is in fact a regular relationship between behavior and scale. This phenomenon has been demonstrated in many medical images, leading to diagnostic use of tools capable to quantify statistical similarity of data patterns at various scales.

Standard measures of scaling and self-similarity are the Hurst exponent H and the fractal dimension D . These measures can be used in the analysis of signals and images as informative summaries. In medical applications, fractal dimension and Hurst exponent are widely used for the analysis, characterization, and classification of mammogram images [39–43].

For self-similar processes, the fractal dimension is related to the Hurst exponent H through the linear relation

$$D + H = n + 1, \quad (11)$$

where n is the dimension of space, and methods for estimating the Hurst exponent can be used for assessing fractal properties. A typical stochastic model for fractals is the fractional Brownian motion, a self-similar process of order $H \in (0, 1)$. [12, 44] use the concept of fractional Brownian motion for estimating the fractal dimension in medical images.

Many techniques for estimating the Hurst exponent (and hence D) exist, and assessing the accuracy of these estimations can be complicated. Wavelet transforms are powerful tools to estimate the Hurst exponent and model statistical similarity at different scales [13, 39, 43, 45, 46].

In this section, we use the planar complex wavelet spectra proposed by [47] for estimating the Hurst exponent in binary images. Hence, we focus on the estimated Hurst exponent and show its ability to differentiate mastopathic from cancerous tissues visible in binary histological images. Moreover, complex wavelet transform produces an additional measure, namely phase information. We also demonstrate the classification power of the phase information and use it as additional modality in the discriminatory analysis. This classifying measure based on binary images of mammary tissues would be a new tool to be used in combination with existing clinical diagnostic tools, thus improving the power of noninvasive diagnostic techniques.

6.1. The complex wavelet spectra

Wavelets and wavelet-based spectra are important tools in the analysis of self-similarity and fractal features [48–50]. In general, the spectra describe the distribution of energies in the signal/image along a range of resolution scales. Orthogonal discrete wavelets are ‘energy preserving’, and as such are natural for defining the power spectrum.

Suppose that a one-dimensional signal y of length n has a wavelet decomposition $d = Wy = (c_{j_0}, d_{j_0}, d_{j_0+1}, \dots, d_j)$, where j_0 is a fixed level smaller than $j = \log_2 n - 1$, c_{j_0} are scaling coefficients, and d_j 's are levels of detail coefficients. The wavelet spectra is defined as

$$S(j) = \log_2 E(d_j^2),$$

where d_j^2 are detail coefficients at level j . In the planar case, different hierarchies of detail coefficients can be considered depending on the tessellation and directions (horizontal, vertical, and diagonal). The natural definition of wavelet spectra involves power spectra corresponding to these hierarchies. Moreover, real and complex wavelets can be used. In this work, we will consider only the detail hierarchy correspondent to the diagonal direction and complex wavelets as in [47].

If we consider a planar fractional Brownian motion, $B_H(\mathbf{u})$, the standard model for self-similar random processes, the complex wavelet coefficients are given by

$$d_{j,k} = 2^j \int B_H(\mathbf{u}) \psi^*(2^j u_1 - k_1, 2^j u_2 - k_2) d\mathbf{u}, \quad (12)$$

where $*$ denotes the complex conjugate of the wavelet ψ . These coefficients are random variables with zero mean and variance [51], which leads to

$$\begin{aligned}\mathbb{E} [|d_{j,k}|^2] &= 2^{2j} \int \psi(2^j u_1 - k_1, 2^j u_2 - k_2) \\ &\quad \times \psi^*(2^j v_1 - k_1, 2^j v_2 - k_2) \mathbb{E}[B_H(\mathbf{u})B_H(\mathbf{v})] d\mathbf{u} d\mathbf{v}.\end{aligned}\quad (13)$$

As in [52], it is assumed here that the coefficients within and across the scales are uncorrelated. From (13), it can be shown that

$$\mathbb{E} [|d_{j,k}|^2] = 2^{-j(2H+2)} V_\psi(H), \quad (14)$$

where $V_\psi(H)$ is an expression depending on ψ and H , but not on the scale j [39]. By taking logarithms in (14)

$$\log_2 \mathbb{E} [|d_{j,k}|^2] = -(2H+2)j + \log_2 V_\psi(H), \quad (15)$$

for $j \in \mathbb{Z}$, the Hurst exponent can be estimated from the slope of the linear equation (15). Finally, the empirical counterpart of (15) is a regression defined on pairs

$$\left(j, \log_2 \frac{1}{n} \sum_{jj} |d_{j,k}|^2 \right), \quad j \in \mathbb{Z}. \quad (16)$$

The slope of the regression would estimate the Hurst exponent, that is, $H = -(\text{slope} + 2)/2$. Although (16) is based on the ordinary least-squares regression, the variance of wavelet coefficients is not constant. Robust approaches that incorporate heteroscedasticity are given in [40]. Moreover, the complex wavelet coefficients $d_{j,k}$ are composed by real and imaginary parts, which can be used to compute the phase information as

$$\angle d_{j,k} = \arctan \left(\frac{\text{Im}(d_{j,k})}{\text{Re}(d_{j,k})} \right)$$

when $|\text{Re}(d_{j,k})| > 0$. Complex wavelet coefficients can be computed by the recursive formula of Mallat's algorithm [53].

6.2. Mastopathy and mammary cancer classification

In this section, we illustrate how the complex wavelet-based spectra and the phase information can be used to classify mastopathic and cancerous tissues. We demonstrate that the spectra slope and phase, as descriptors of mammary tissue, have good discriminatory power.

We performed the discrete complex wavelet transform for every image of size 512×512 by using Daubechies 6-tap filter. After each transform, we estimated the slope of wavelet spectra using traditional ordinary least-squares regression along with the modified harmonic average robust estimation methods proposed by [40]. The application of the complex wavelet transform to binary images rather than digital mammograms constitutes a new diagnostic modality in the field. For each classification method, we randomly selected 67% of the data as a training set to fit the classifier and used the remaining 33% of the data to test performance. The random selection of training and testing sets was repeated 10,000 times, such that the reported prediction errors are averaged over 10,000 runs. Performance of each model was compared in terms of sensitivity, specificity, and overall correct classification rate.

Figure 18 shows the scaling property of the cancerous tissue as an example, and Figure 19 depicts the estimated densities of the fractal dimensions for mastopathic and cancerous tissues obtained from the estimated Hurst exponents using (11). One can see that estimated fractal dimensions of mastopathic tissues tend to be more irregular (with a lower Hurst exponent) than tumor-affected tissue, although the dispersion is higher. Similar results have been obtained by [13] and [39] using digital mammogram images.

Another important element in the classification of mammary images is given by the phase information of the finest detail coefficients. Figure 19 shows how the estimated density of the phase variance from mastopathic images differs from cancerous ones.

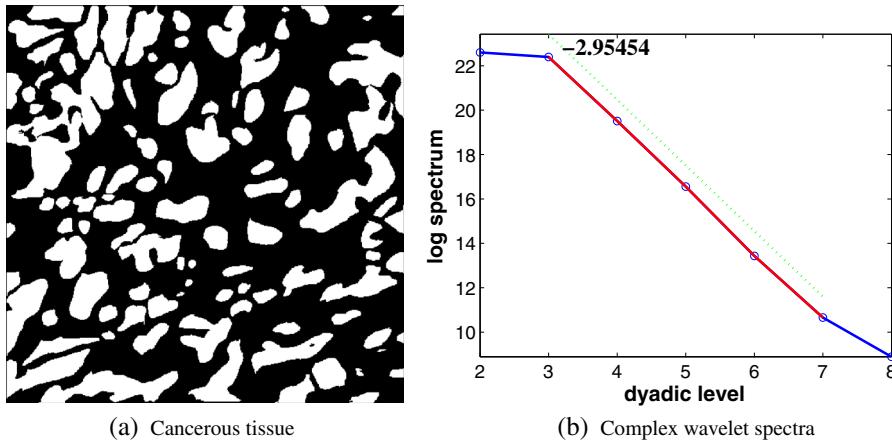


Figure 18. Case of binary image of cancerous tissue and correspondent complex wavelet spectra.

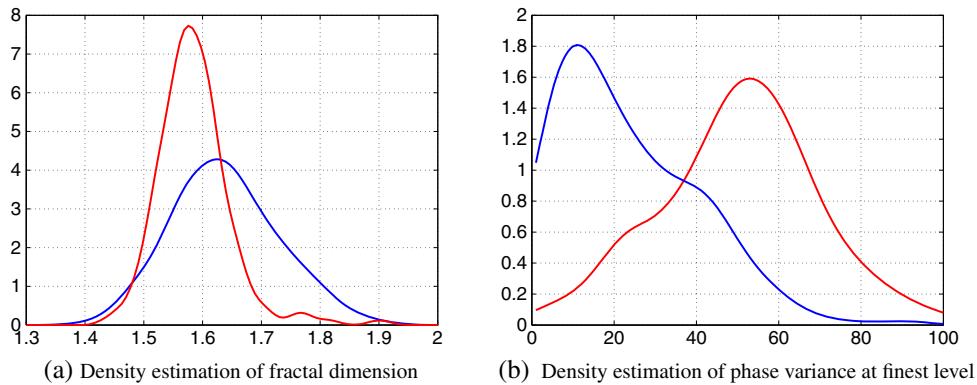


Figure 19. Estimated density of fractal dimension and phase variance at the finest level for mastopathic (blue) and cancerous tissues (red).

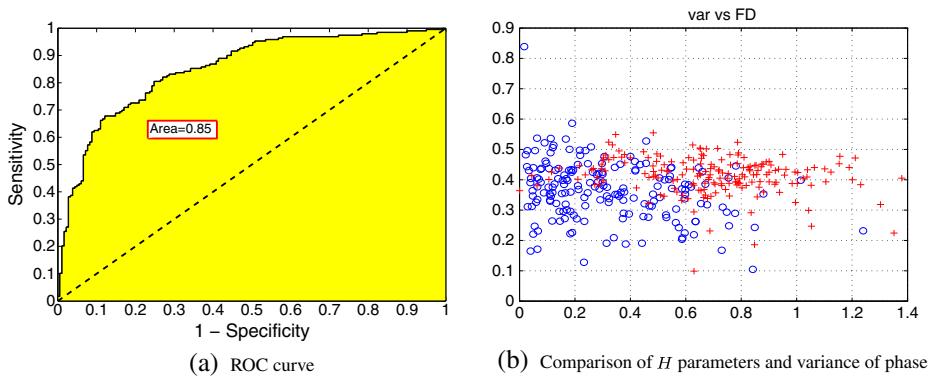


Figure 20. ROC curve of phase variance in differentiating mastopathic versus cancerous cases (a) and estimated H parameters versus variance of phase for mastopathic (circle) and cancerous tissues (crosses) (b).

The discriminatory power of the phase variance of the finest detail is shown in Figure 20(a), which represents the ROC curve of phase variance in differentiating mastopathic versus cancerous cases. The diagonal line represents a test with a sensitivity of 50%, and the area is proportional to the diagnostic accuracy. The plot shows the ROC curve lying significantly to the left of the diagonal where the combination of sensitivity and specificity is high.

Figure 20(b) shows a scatter plot of cases by estimated Hurst parameters versus phase variance, illustrating the differentiation between mastopathic and cancerous cases. Hence, we conducted a classification

Table IX. Classification based on the estimated Hurst exponents and phase variance of the finest detail coefficients; three different classification methods are compared: logistic, linear, and quadratic.

Method	Sensitivity	Specificity	Correct classification
Logistic	0.8317	0.79332	0.77568
Linear	0.82121	0.80318	0.81242
Quadratic	0.80886	0.81435	0.81154

using both Hurst parameters versus phase variances as inputs of three different classification methods: logistic, linear, and quadratic.

As Table IX shows, the linear and quadratic predictors perform the best classification between cancerous and mastopathic images with correct classification rates being more than 80%. The result constitutes a new important contribution in the context of computer-aided classification of mammary tissues.

7. Summary

The main focus of this paper is discrimination of benign and malignant mammary tissues using various approaches (see also [54]). We developed a procedure on the basis of Sierpinski carpet. Randomization was introduced by binomial distributed random variables, which were generated for every box of the matrices except for the middle one. In a further step, the basic matrix was changed, such that for every box, a random number was generated. Therefore, the structure of the Sierpinski carpet was completely dissolved except for the underlying matrix. To be able to adopt these illustrations to the research, black areas instead of random number equal to 1 and white areas instead of 0 are displayed. We have provided a simulation study that shows the best fit of parameters p and q for the model $[p-p-p-q]$ compared with other combinations as $[p-p-q-p]$, $[p-q-p-p]$, or $[q-p-p-p]$. Moreover, the application of the estimation procedure on real data images of mammary cancer and mastopathy allows a good differentiation in terms of values as well as simulated images with these estimators. We also illustrate theoretical and practical issues in the generation of Sierpinski gaskets and HM calculations. Also, HM provides a nice discrimination tool between mammary and mastopathic tissues.

The particular aim of this paper was to find a better model than the Boolean model to describe mammary cancer tissue. It was shown that Quermass-interaction process can serve as such a model. This model describes many variations of mammary cancer images; nevertheless, it did not cover all variations. The Quermass-interaction process can serve for mammary cancer data well; however, random carpet models serve better for more structured mastopathic tissue. In more detail, the Quermass-interaction process can describe mammary cancer tissue whose structure is absolutely or partly broken down up to a certain level. It can be considered as an improvement compared with the Boolean model, which is only able to describe mammary cancer tissue whose structure is absolutely broken. It was also found out that mastopathic tissue deviates graphically significantly stronger from Quermass-interaction process than mammary cancer tissue does. Further studies on this topic can look for the model capable to describe the organized structure of mastopathic tissue or search some variation of the Quermass-interaction process that improves the presented results. For example, other characteristics of the random closed set than volume, perimeter, or Euler number can be taken, or another distribution of the radii Q of the underlying Boolean model can be assumed. However, introduced random carpet model fits well for modeling mastopathic tissue, having only two parameters p and q to be estimated.

The wavelet-based self-similarity measures and also fitted distributions to fractal dimensions show existence of several competing fractal structures (instead of the only one value of dimension). This observation was confirmed also by analysis of box-counting dimension by simple descriptive statistics, for example, by plotting box plots for both mastopathy and mammary cancer groups. To relate only to one dimension establishes ill-posed inverse problems. We need a continuous/dense dimension spectrum, for example, multi-fractal dimension spectra, as already used in breast cancer discrimination by [55]. A multi-fractal system is a generalization of a fractal system in which a single exponent (the fractal dimension) is not enough to describe its dynamics; instead, a continuous spectrum of exponents (the so-called singularity spectrum) is needed. This relates to Tweedie exponential dispersion models, which, as a special case, contain both normal and gamma distributions. This is a further justification for these two simple distributional families: in the case of our empirical data, we have found a strong deviation from normality for mastopathy, and therefore, we used gamma and Weibull distributions. In contrast to that,

mammary cancer data have been adjusted by outliers (for more details, see [56]), which enable usage of normal distribution. Tissue is not a simple structure, and cancer is a complex compound of several diseases. The further systematic study of continuous spectrum of fractal dimension can put more light on several fractal organisms/objects observed in tissues of cancer patients [57–60]. It is clear that more research should be performed, also in microbiological level, to better understand this variability. The theory of lattice-valued bornological vector systems, which has been started in [61], makes another step towards the development of a new technique for cancer research. In practical applications, we can meet a bornological space instead of a metric one. Almeida and Barreira [62] introduced the concept of Hausdorff dimension for convex bornological vector spaces. Having a place for both geometric and algebraic information, bornological systems seem to be more suitable to model certain cases (see also [63]). Several open problems still remain, for example, what is the optimal way of computing non-Euclidean dimension for tissue and which cancers correspond to the paradigm ‘the higher the dimension, the higher the cancer grade’? It is clear that exceptions exist, for example, Wilms tumors [7]. By the cancer grade we mean the quantitative assessment (e.g., by means of fractal dimension analysis) of the degree of abnormality and aggressiveness of cancer tumors. The fractal dimension is an objective and reproducible measure of the complexity of the tissue architecture of the tissue specimen. Hence, we hope that quantitative approaches such as fractal geometry and stochastic geometry may lead to an improved understanding of progression of malignancy [64, 65].

Appendix A. Introduction to the package FractalParameterEstimation

A.1. Using the package

The main function of the package FractalParameterEstimation is *estimationFunction()*, which delivers estimators \hat{p} and \hat{q} . A call of this *estimationFunction()* will result in a list of two values, which stand for estimators of p and q . Data in the form of tables or matrices, where number of rows equals to number of columns, have to be given. Moreover, it has to consist of either 0 or 1 in every cell, which stands in a wider sense for black and white. The reason for this is breast mammography screenings, which deliver images with black and white areas. A simulation of randomized Sierpinski carpets with parameters p and q should in further consequence model mastopathic and malignant tissue. The package FractalParameterEstimation was created in order to estimate parameters \hat{p} and \hat{q} of a $[p-p-p-q]$ model on the basis of random fractals (e.g., Sierpinski carpet). *estimationFunction()* combines calls of the underlying functions, which are *createSmallerMatrix()*, *calcRamification()*, *increment()*, *potence()*, and *fillMatrix()*. The last given functions will not be called by the user but performed by usage of the main function. In the following sections, these functions will be described more precisely.

A.2. Estimation function

As already mentioned, estimators \hat{p} and \hat{q} will be returned by *estimationFunction()* with the aid of a call of *createSmallerMatrix()* and following to *fillMatrix()*. Call of the function is performed by *estimationFunction(data)*, where a table in a defined form is forwarded. In the first step, data are transformed into a matrix, and the impact of q has to be neglected to be able to estimate the parameter p . This is performed by creating a matrix of smaller size by *createSmallerMatrix(data)*. As a consequence, the values inside this matrix need to be filled in a correct way, which is enabled by *fillMatrix()*. On the basis of the computed matrix of the original data and the smaller matrix, the parameters can be estimated by (2).

A.3. Create and fill matrix for $[p-p-p]$ model

The parameters can be estimated more precisely if the impact of p and q in only one ramification of the model is analyzed separately. Indeed, the resulting values of q give information on the values of the cells after three iterations (one step earlier) with fixed parameter p . Hence, a smaller matrix with a size of 3^{r-1} has to be computed. In this special setting from a (81×81) data table ($3^4 = 81$), a square matrix with a dimension of $3^3 = 27$ is computed. *createSmallerMatrix(data)* accomplishes the desired result and sets value 99 in every cell. These values were chosen, so that any mistakes in the computations are detected rapidly; otherwise, further calculations would fail. As mentioned before, only 0 and 1 are allowed in the matrices, wherefore *fillMatrix(data, smallerMatrix)* needs to be proceeded. For this reason, the values in every cell after three ramifications need to be rebuilt. The basic model of the simulation is a Sierpinski

carpet, which has the property of creating in every step (3×3) matrices for every cell. Therefore, matrices of this size of the original data matrix need to be investigated. In total, four *for()* loops enable to run through every cell of the smaller matrix (first two loops) and to count the frequency of 1 in every (3×3) matrix of the original matrix (other two loops). A counter variable was used in this setting, and if this value was larger than zero (there is at least once 1 in a (3×3) matrix), the appropriate cell of the smaller matrix will be set 1, otherwise 0. This can be performed, because occurrence of 0 in any ramification leads to subjacent matrices, which will be completely filled with 0.

A.4. Additional used functions

Further functions were used in order to create the smaller matrix. The value of ramification is returned after calling *createSmallerMatrix()* by usage of function *calcRamification()*. In this computation, the variable r of the equation $\dim = 3^r$ is of interest, and it is calculated with the aid of the dimensions of the original matrix. In the next step, function *potence()* was used, which enables us to calculate values of the form a to the power of b , where in this setting $a = 3$ and $b = \text{calcRamification}(\text{data}) - 1$. Executing these functions leads to the equation 3^{4-1} , and number of rows and number of columns for smaller matrix equal to this value ($\dim = 27$). *increment()* is used to add one to the specific value of the variable, and it returns the resulting variable.

A.5. IMAGEJ software

In order to receive the dimensions of the binary images, IMAGEJ software was used. ImageJ was developed at the National Institute of Health in the USA, and it is a Java-based image-processing program. It is a public domain and designed with an open architecture, which is expandable via Java plug-ins and recordable macros. The plug-in FracLac is for fractal analysis. Within this plug-in, the option *Standard Box Count* stands for estimating box-counting dimension. It is important to define the background of binary images, because this procedure calculates the dimension by examining white pixel in binary images, so the background is defined as black in this case. After these settings, one picture or a folder, to compute several pictures automatically, can be selected. As a result, IMAGEJ delivers box-counting dimension of binary images [66].

Appendix B. Combinations of parameters for four ramifications model

An extensive simulation study focusing on combinations of the parameters of a model, consisting of four ramifications, has been conducted in order to differentiate between models by means of goodness of fit of parameters of interest, p and q . For this purpose, the four models were simulated in the first step in Java and their parameters estimated in R. Simulation study with $[p-p-q-p]$ gave similar estimation results as the reference model in the paper, $[p-p-p-q]$. The choice of $[p-p-p-q]$ model is a natural one, because we start to estimate rougher scales of fractal ramification in our model in three steps by parameter p . Subsequently, parameter q estimates very fine scales of ramification (we can call it microergodicity/noise with respect to parameter p). However, estimation on the basis of $[q-p-p-p]$ has shown stronger deviations for parameter q . These discrepancies arise because of the small number of cells in the first ramification, leading to larger variance in the random-number generating process in this step. Moreover, the sum of absolute differences between the true value and the estimator is significantly larger in $[p-q-p-p]$ model in contrast to the reference model, resulting from the observation that estimations of earlier ramification lead to stronger deviations, as can also be seen in $[q-p-p-p]$ model.

Acknowledgements

The first and the corresponding author have been supported by ANR project Desire FWF I 833-N18. The corresponding author acknowledges Proyecto Interno 2015, REGUL. MAT 12.15.33, Modelaciòn del crecimiento de tejidos con aplicaciones a la Investigaciòn del càncer. The third author give thanks to the support of the grant APVV-0184-10. We acknowledge a partial support by Projects 63p22 and 58p5 ‘Aktion Czech Republic-Austria: Testing and modeling of cancer’ and also by the Aktion Project No. 67p5 (Austria–Czech Republic) Algebraic, fuzzy and logical aspects of statistical learning for cancer risk assessment. This work was supported by the Grant Agency of Czech Republic, Project Nos. P201/10/0472 and 13-05466P. First author gratefully thanks Karl Regensburger for his assistance with the JAVA program to model the development of the Sierpinski carpet. Last but not least, the authors are very grateful to the editor and reviewers for their valuable comments.

References

1. Mattfeldt T, Schmidt V, Reepschläger D, Rose C, Frey H. Centred contact density functions for the statistical analysis of random sets. A stereological study on benign and malignant glandular tissue using image analysis. *Journal of Microscopy* 1996; **183**:158–169.
2. Mattfeldt T. Classification of binary spatial textures using stochastic geometry, nonlinear deterministic analysis and artificial neural networks. *International Journal of Pattern Recognition and Artificial Intelligence* 2003; **17**:275–300.
3. Mattfeldt T, Meschenmoser D, Pantle U, Schmidt V. Characterization of mammary gland tissue using joint estimators of Minkowski functionals. *Image Analysis and Stereology* 2007; **26**:13–22.
4. Stoyan D, Kendall WS, Mecke J. *Stochastic Geometry and Its Applications* Second Edition. J. Wiley: Chichester, 1995.
5. Mandelbrot B. *Fractals: Form, Chance and Dimension*. W. H. Freeman and Co.: San Francisco, 1977.
6. Baish JW, Jain RK. Fractals and cancer. *Perspectives in Cancer Research* 2000; **60**:3683–3688.
7. Stehlík M, Giebel SM, Prostakova J, Schenk JP. Statistical inference on fractals for cancer risk assessment. *Pakistan Journal of Statistic* 2014; **30**(4):439–454.
8. Mrkvíčka T, Mattfeldt T. Testing histological images of mammary tissues on compatibility with the Boolean model of random sets. *Image Analysis and Stereology* 2011; **30**:11–18.
9. Stehlík M, Mrkvíčka T, Filus J, Filus L. Recent developments on testing in cancer risk: a fractal and stochastic geometry. *Journal of Reliability and Statistical Studies* 2012; **5**:83–95.
10. Hermann P, Piza S, Ruderstorfer S, Spreitzer-Gröbner S, Stehlík M. Fractal case study for mammary cancer. *Biometrie und Medizinische Informatik - Greifswalder Seminarberichte* 2014; **23**:149–166.
11. Kiselák K, Pardasani R, Adlakha N, Agrawal M, Stehlík M. On some probabilistic aspects of diffusion models for tissue growth. *The Open Statistics and Probability Journal* 2013; **5**:14–21.
12. Pesquet-Popescu B, Lèvy-Véhel J. Stochastic fractal models for image processing. *IEEE Signal Processing Magazine* 2002; **19**(5):48–62.
13. Ramírez P, Vidakovic B. Wavelet-based 2D multifractal spectrum with applications in analysis of digital mammography images. *Computational Statistics and Data Analysis* 2013; **58**:71–81.
14. Perl W. Heat and matter distribution in body tissues and determination of tissue blood flow by local clearance methods. *Journal of Theoretical Biology* 1962; **2**:201–235.
15. Luebeck EG, Moolgavkar SH. Stochastic analysis of intermediate lesions in carcinogenesis experiments. *Risk Anal* 1991; **11**(1):149–157.
16. Luebeck EG, Moolgavkar SH. Simulating the process of malignant transformation. *Mathematical Biosciences* 1994; **123**(2):127–146.
17. Gregori G, Hanin L, Luebeck EG, Moolgavkar SH, Yakovlev A. Testing goodness of fit for stochastic models of carcinogenesis. *Mathematical Biosciences* 2002; **175**(1):13–29.
18. Kopp-Schneider A. Birth-death processes with piecewise constant rates. *Statistics & Probability Letters* 1992; **13**:121–127.
19. Kopp-Schneider A. Carcinogenesis models for risk assessment. *Statistical Methods in Medical Research* 1997; **6**:317–340.
20. Kopp-Schneider A. Biostatistical evaluation of focal hepatic preneoplasia. *Toxicologic Pathology* 2003; **31**:121–125.
21. Wosniok W, Kitsos C, Watanabe K. Statistical issues in the application of multistage and biologically based models. In *Perspectives on Biologically Based Cancer Risk Assessment* 1999, Cogliano VJ, Luebeck EG, Zapponi GA (eds). Springer: New York US; 243–273.
22. Kaye BH. *Characterization of Powders and Aerosols*. John Wiley & Sons: Weinheim, 2008.
23. Peintgen H.O, Juergens H, Saupe D. *Bausteine des Chaos. Fraktale*. Springer-Verlag: Berlin-Heidelberg, 1991.
24. Avnir D, Biham O, Lidar D, Malcai O. Is the geometry of nature fractal. *Science (Washington DC)* 1998; **279**:39–40.
25. Falconer K. *Fractal Geometry. Mathematical Foundations and Applications* Second Edition. John Wiley & Sons: New York, 1998.
26. R Development Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria, 2012. <http://www.R-project.org>.
27. Molchanov I. *Statistics of the Boolean Model for Practitioners and Mathematicians*. John Wiley & Sons: New York, 1997.
28. Kendall WS, van Lieshout MNM, Baddeley AJ. Quermass–interaction processes: conditions for stability. *Advances in Applied Probability* 1999; **31**:315–342.
29. Møller J, Helisová K. Power diagrams and interaction process for unions of discs. *Advances in Applied Probability* 2008; **40**:321–347.
30. Staňková Helisová K, Staněk J. Dimension reduction in extended Quermass–interaction process. *Methodology and Computing in Applied Probability* 2014; **16**:355–368.
31. Møller J, Waagepetersen RP. *Statistical Inference and Simulations for Spatial Point Processes*. Chapman and Hall/CRC: Boca Raton, 2004.
32. Møller J, Helisová K. Likelihood inference for unions of interacting discs. *Scandinavian Journal of Statistics* 2010; **37**(3):365–381.
33. Dereudre D, Lavancier F, Staňková Helisová K. Estimation of the intensity parameter of the germ-grain Quermass–interaction model when the number of germs is not observed. *Scandinavian Journal of Statistics* 2014; **41**(3):809–829.
34. Mrkvíčka T. On testing of general random closed set model hypothesis. *Kybernetika* 2009; **45**:293–308.
35. http://www.encyclopediaofmath.org/index.php/Hausdorff_measure, [Accessed on 15 February 2015].
36. Hausdorff F. Dimension und äußeres Maß. *Mathematische Annalen* 1918; **79**(1–2):157–179. MR1511917.
37. Mattila P. *Geometry of Sets and Measures in Euclidean Spaces. Fractals and Rectifiability*, Vol. 44. Cambridge University Press: Cambridge, 1995. MR1333890.
38. Holland JH. *Adaptation in Natural and Artificial Systems*. University of Michigan Press: Ann Arbor, 1975.

39. Nicolis O, Ramirez P, Vidakovic B. 2-D wavelet-based spectra with applications. *Computational Statistics and Data Analysis* 2011; **55**(1):738–751.
40. Hamilton EK, Jeon S, Ramirez-Cobo P, Lee KS, Vidakovic B. Diagnostic classification of digital mammograms by wavelet-based spectral tools: a comparative study. *The Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine* 2011:384–339.
41. Don S, Chung D. A new approach for mammogram image classification using fractal properties. *Cybernetics and Information Technologies* 2013; **12**(2):69–83.
42. Dobrescu R, Ichim L, Crisan C. Diagnosis of breast cancer from mammograms by using fractal measures. *International Journal of Medical Imaging* 2013; **1**(2):32–38.
43. Qingrui P, Zhuangzhi Y. Mammogram analysis based on wavelets and fractals. *Proceedings of IEEE Engineering in Medicine and Biology Society* 1996; **3**:1159–1162.
44. Parra C, Iftekharuddin K, Rendon T. Wavelet based estimation of the fractal dimension in fBm images. *Presented in the 1st IEEE EMBS conference on Neural Engineering*, Capri Island, Italy, 2003.
45. Kestener P, Saint-Jean P, Arneodo A. Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Anal. Stereol.* 2001; **20**:169–174.
46. Tafti PD, Ville DVDM, Unser M. Innovation modelling and wavelet analysis of fractal processes in bio-imaging. *Proceedings on the 5th IEEE International Symposium on Biomedical Imaging (ISBI'08)*, (Paris, France), 2008:1501–1504.
47. Nicolis O, Jeon S, Vidakovic B. Mammogram diagnostics via 2D complex wavelet-based self-similarity measures. *Working paper*. <http://aisberg.unibg.it/handle/10446/26683>, 2012.
48. Flandrin P. On the spectrum of fractional Brownian motions. *IEEE Transaction on Information Theory* 1989; **35**:197–199.
49. Flandrin P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transaction on Information Theory* 1992; **38**:910–917.
50. Wornell G. *Signal Processing with Fractals: A Wavelet Based Approach*, 1995. Prentice-Hall: Englewood Cliffs, New Jersey, 1996.
51. Heneghan C, Lown SB, Teich MC. Two dimensional fractional Brownian motion: wavelet analysis and synthesis. Image analysis and interpretation. *Proceedings of the IEEE Southwest Symposium* 1996:213–217.
52. Veitch D, Abry P. Wavelet-based joint estimate of the long-range dependence parameters. *IEEE Trans. on Information theory, special issue on Multiscale Statistical Signal Analysis and its Applications* 1999; **45**(3):878–897.
53. Mallat SG. *A Wavelet Tour of Signal Processing*. Academic Press: San Diego, 1998.
54. Kimmel M, Gorlova O, Henschke C. Modeling lung cancer screening. *Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment* 2005:161–175.
55. George LE, Kamal HS. Breast cancer diagnosis using multi-fractal dimension spectra. *IEEE International Conference on Signal Processing and Communications (ICSPC)*, Dubai, 2007, 592–595.
56. Stehlík M, Hermann P. Fractal based cancer modelling. *Current topics on Risk Analysis: ICRA6 and RISK 2015 Conferences Proceedings*, Barcelona, 2015.
57. Chiu Ch C, Groll AH, Walsh TJ. New drugs and novel targets for treatment of invasive fungal infections in patients with cancer. *The Oncologist* 2000; **5**:120–135.
58. Taheri FH, Seyedolmohadesin M, Bayat M, Mahdavi M, Yazdi MH, Eslamifar A, Abolhassani M. The effect of Candida albicans systemic infection on matrix metalloproteinases in breast cancer bearing BALB/c mice. *Iran J Allergy Asthma Immunol* 2013; **12**(1):81–85.
59. Chakravarthi S, Choo ZW, Nagaraja HS. Susceptibility to renal candidiasis due to immunosuppression induced by breast cancer cell lines. *Science World Journal* 2010; **5**(1):5–10.
60. Enby E. A breast cancer tumor consisted of a spore-sac fungus (Ascomycotina). *3rd Millennium Health Care Science* 2013; **18**(1):8–10.
61. Paseka J, Solovyov SA, Stehlík M. On the category of lattice-valued bornological vector spaces. *Journal of Mathematical Analysis and Applications* 2014; **419**:138–155.
62. Almeida J, Barreira L. Hausdorff dimension in convex bornological spaces. *Journal of Mathematical Analysis and Applications* 2002; **268**(2):590–601.
63. Paseka J, Solovyov SA, Stehlík M. Lattice-valued bornological systems. *Fuzzy Sets and Systems* 2015; **259**(15):68–88.
64. Kitsos CP. Estimating the relative risk for the breast cancer. *Biometrical Letters* 2010; **47**(2):133–146.
65. Risch A, Dally H, Edler L. Genetic polymorphisms in metabolising enzymes as lung cancer risk factors. *Recent Advances in Quantitative Methods in Cancer and Human Risk Assessment* 2005:43–62.
66. Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. *Biophotonics International* 2004; **11**(7):36–43.