RANDOM CLOSED SET MODELS: ESTIMATING AND SIMULATING BINARY IMAGES

ÁNGELES M GALLEGO AND AMELIA SIMÓ

Department of Mathematics, University Jaume I, 12071 Castellón, Spain.
e-mail: simo@mat.uji.es

ABSTRACT

In this paper we show the use of the Boolean model and a class of RACS models that is a generalization of it to obtain simulations of random binary images able to imitate natural textures such as marble or wood. The different tasks required, parameter estimation, goodness-of-fit test and simulation, are reviewed. In addition to a brief review of the theory, simulation studies of each model are included.

Keywords: Boolean model, germ-grain model, parameter estimation, simulation.

INTRODUCTION

There are many practical situations in which the imitation of natural textures such as marble or wood is of great interest. In other words, apparently equal but random binary or gray level images must be obtained. Examples include textile and tile manufacture. Stochastic Geometry is a branch of Applied Probability that provides very powerful tools for this task. In particular, Random Closed Sets (RACS) models can be used. The most widely known and used Random Closed Set model is the Boolean model. In this paper we show how the Boolean model and a class of RACS models that is a generalization of it, can be used to obtain these simulations. With the exception of one case, we consider only binary images because the corresponding mathematical techniques are simpler and better developed. In any case, a grey level image can be transformed to a family of binary images by means of a family of thresholds.

If we want to use a RACS model to simulate similar images to a natural one these steps should be followed. A model that fits the data should first be found. Then the goodness of this fit should be checked. Thirdly the parameters of this model should be estimated from the observed image. Finally the simulation can be carried out. All these steps are considered in this paper. A brief review of the theory of the Boolean model and the most important generalizations of it is provided. These include germ-grain models, three phased models and Poisson-Boolean models. Parameter estimation methods and simulation algorithms are proposed for each of these models. In order to evaluate the goodness of the fit, the Monte Carlo test should be used and this is also briefly reviewed in this paper.

We have written a library of functions to be used with MATLAB[1], to carry out the different tasks that we refer to: parameter estimation, goodness-of-fit test and simulation. This library is available at http://www3.uji.es/~simo.

There is a vast literature about Random closed Set Models, mainly about Boolean Model. But many of the papers are focused on theoretical properties and most of them are devoted to a particular model. The goal of this paper is to put together the most useful and interesting models and to emphasize their practical aspects, doing them easy to use. Although the statistical methods proposed are mainly well-known, specially those of section "The Boolean model", as far as we know they never have been used in other kind of models like those of Section "Non-homogeneous germ-grain model" and "Three-phase Boolean model or sinter textures".

The rest of the paper is organized as follows. The next section begins with the definition of RACS and general properties are reviewed. The second section is devoted to the Boolean model: definition, basic properties, parameter estimation and simulation. In the third section the same is carried out for the Boolean model generalizations. Finally, conclusions and future research are given in the fourth section.

RANDOM CLOSED SETS

DEFINITIONS AND ESTIMATORS

Random closed sets are mathematical models for irregular random area patterns whose formal definition

---

[1]MATLAB is a trademark of The MathWorks Inc.

was provided by Matheron (1974). A random closed set (RACS) Ξ is a random variable taking values in $(\mathscr{F}, \sigma_f)$, the class of closed subsets in the Euclidean space, $I\!R^2$, with the σ-algebra generated by the Hit or Miss topology.

The RACS models considered here will be, with the exception of one case, stationary and ergodic. Stationarity means that its probability distribution is invariant against translations. A RACS is said to be ergodic if its mean characteristics can be obtained from spatial averages of the functional of this RACS.

The parameters of the probability distribution of a RACS can be classified into two types: aggregate and individual. The former are directly observable and can thus be easily estimated by using the ergodic property. Their definitions and estimators are common to all the RACS models. The latter are not directly observed, although they are of greatest interest when fitting a model to real data and they are specific to each model. Individual parameters have to be estimated through estimates of aggregate parameters and their estimation is very difficult except in the Boolean case. In this section the most important aggregate parameters of a general stationary RACS model are studied.

The probability distribution of any general RACS is uniquely determined by its capacity functional (Matheron, 1974);

$$T_{\Xi}(K) = P(\Xi \cap K \neq \emptyset) . \tag{1}$$

where K is any compact subset of $I\!R^2$. If Ξ is observed in a window W, an unbiased estimator of T will be given by:

$$\hat{T}_{\Xi}(K) = \frac{A((\Xi \oplus \check{K}) \cap (W \ominus K))}{A(W \ominus K)} ,$$

where $\Xi \oplus \check{K}$ is the dilation of Ξ with K, $W \ominus K$ the erosion and for $B \subset I\!R^2$, $A(B)$ denotes the area of B. We will call $\hat{T}_{\Xi}$ *the empirical capacity functional* of Ξ and it is a consistent estimator of T_{Ξ}.

The simplest aggregate parameter of a RACS is the area fraction that is defined as the mean of the area of Ξ in a unitary window W:

$$p = E(A(\Xi \cap W)) = P(0 \in \Xi) .$$

An unbiased estimator of p is:

$$\hat{p} = \frac{A(\Xi \cap W)}{A(W)} . \tag{2}$$

A further, very useful aggregate functional parameter is the contact distribution function, which provides a kind of measurement of 'size' when identification of single particles is inappropriate. The definition of the contact distribution function depends on the choice of a structuring element B. If B is a compact set containing the origin, then the contact distribution function is defined by:

$$H_B(r) = 1 - \frac{1 - T_{\Xi}(rB)}{1 - p} \tag{3}$$

for $r \geq 0$. The cases of particular practical importance are the linear contact distribution function where B is a segment of unit length and the spherical contact distribution function where $B = B(0,1)$ is a unit disk. The contact distribution function can be estimated by combining the estimators of the capacity functional and the area fraction.

The last three reviewed aggregate parameters are less important, they are defined only for Random Closed Sets that fulfill certain conditions of regularity (convexity, smooth boundary,. . .) and will be used only in the Boolean case.

The specific convexity number, N_A^+, of a stationary and isotropic RACS is defined as the average number of exposed tangent points in a given direction. The estimation of the specific convexity number can be performed as follows. We choose an arbitrary direction u and we count the number of tangent points in the direction u inside the window W, $N(N^+(u) \cap W)$. Thus

$$\hat{N}_A^+ = \frac{N(N^+(u) \cap W)}{A(W)}$$

is a strong consistent estimator of N_A^+.

The specific boundary length, L_A of Ξ is its expected boundary length per unit area. The following is an unbiased estimator (Weil and Weaieacker, 1984):

$$\hat{L}_A = \frac{U(\Xi \cap W) - U(\Xi \cap \delta^+ W)}{A(W)} ,$$

where $\delta^+ W$ is the upper-right boundary of W and $U(\Xi \cap \delta^+ W)$ is the length of the corresponding intersection. Simpler estimators, as for example, the length of the boundary of Ξ inside W divided by the area of W are strong consistent but biased.

Finally, the specific connectivity number χ_A of a RACS is defined as the expected connectivity number per unit area, roughly speaking the mean of the difference between the number of clumps and the number of holes per unit area. An unbiased estimator of the specific connectivity number is given by

$$\hat{\chi}_A = \frac{\chi(\Xi \cap W) - \chi(\Xi \cap \delta^+ W)}{A(W)} ,$$

where for $B \subset I\!R^2$, $\chi(B)$ denotes the connectivity number of B.

MONTE CARLO GOODNESS-OF-FIT TEST

Different procedures can be found in the literature to test whether a realization of a RACS can be assumed that follows a particular model. In Molchanov (1997) some of them are proposed for the Boolean model. If we are interested in a general method, Monte Carlo methods should be used (see Diggle, 1983).

A simple exact Monte Carlo test based on the contact distribution function can be constructed as follows.

Let us firstly assume that $H_B(r)$ is known under the null hypothesis.

Let $\hat{H}_B^{(1)}(r)$ be the empirical distribution function of the observed Ξ and let $\hat{H}_B^{(i)}(r)$, $i = 2, 3, \ldots s$, be the corresponding to each of the $s - 1$ independent simulations of the RACS model that we are assuming under the null hypothesis, we define u_i some measure of the discrepancy between $\hat{H}_B^{(i)}(r)$ and $H_B(r)$ over the whole range of r, for example

$$u_i = \int (\hat{H}_B^{(i)}(r) - H_B(r))^2 dr \,,$$

and proceed to a test based on the rank of u_1, i.e. for a significance level $\alpha = 1 - k/s$, the Boolean model hypothesis is rejected if u_1 ranks kth largest or higher.

Another possibility is to use a graphical test. The graphical procedure consists of plotting $\hat{H}_B^{(1)}(r)$, $H_u(r)$ and $H_l(r)$ against $H_B(r)$, being $H_u(r)$ and $H_l(r)$ the upper and lower envelopes of $\hat{H}_B^{(i)}(r)$, $i = 2, 3, \ldots s$. If $\hat{H}_B^{(1)}(r)$ lies close to $H_B(r)$ and it is between both envelopes, there is no evidence to reject the null hypothesis.

If the theoretical distribution function $H_B(r)$ is unknown then both, the exact test and the graphical test, can still be carried out if $H_B(r)$ is replaced by

$$\bar{H}_B^{(i)}(r) = \frac{1}{s-1} \sum_{i \neq j} \hat{H}_B^{(j)}(r) \,.$$

THE BOOLEAN MODEL

DEFINITION

This section is devoted to the study of the Boolean model, the most important and relatively simple example of RACS. It is both flexible and amenable to calculations.

The formal definition of Boolean model is as follows.

Definition 1 *Suppose* $\Phi_\lambda = \{x_1, \ldots, x_n, \ldots\}$ *is a stationary Poisson point process in* $\mathbb{R}^2$ *of intensity* λ. *Let* $\Xi_1, \Xi_2, \ldots$ *be a sequence of independent identically distributed random compact sets in* $\mathbb{R}^2$ *that are independent of the Poisson process* Φ_λ *and satisfy* $EA(\Xi_0 \oplus \check{K}) < +\infty$ *for all compacts K, where* Ξ_0 *is a random compact set of the same distribution as* Ξ_n. *The Boolean model* Ξ *is:*

$$\Xi = \bigcup_i (x_i + \Xi_i) \,. \qquad (4)$$

The points x_i are called *germs*, the sets Ξ_n are known as *grains* and the random set Ξ_0 is said the *typical grain* of the Boolean model.

The value of parameter λ is said to be the *intensity* of the Boolean model.

A more in-depth study of this model can be found in Stoyan *et al.* (1995), Molchanov (1997), Cressie (1993), Ayala (1988) and Serra (1982). Boolean model applications to real images can be found in Stoyan *et al.* (1995), Serra (1982), Plaza (1991), Margalef (1974) and Lyman (1972).

Individual parameters of a Boolean model are the intensity of the germ process and parameters of the distribution of the primary grain.

PROPERTIES

In the Boolean model, the expression of the aggregate parameters defined in the previous section are known and relatively simple. They are reviewed in this section, and a theoretical study on them can be found in Stoyan *et al.* (1995) and Molchanov (1997).

The expression of the capacity functional for a Boolean model is:

$$T_\Xi(K) = 1 - \exp\{-\lambda EA(\Xi_0 \oplus \check{K})\} \,.$$

If the primary grains are convex, the generalized Steiner formula gives for convex K:

$$T_\Xi(K) = 1 - \exp\{-\lambda(\bar{A} + \frac{1}{2\pi} U(K)\bar{U} + A(K))\} \,,$$

where $\bar{U}$ is the mean of the perimeter of Ξ_0 and $\bar{A}$ is the mean of its area. If we substitute K for the origin, we obtain the expression of the area fraction: $p = 1 - \exp(-\lambda\bar{A})$.

Taking into account Eq. 3, the contact distribution function of a Boolean model has the following expression

$$H_B(r) = 1 - \exp\{-\lambda E[A(\Xi_0 \oplus r\check{B})] - \bar{A}\} \,. \qquad (5)$$

Finally can easily be shown that

$$N_A^+ = \lambda (1 - p),$$
$$L_A = \lambda (1 - p)\bar{U},$$

and if Ξ_0 is convex and isotropic:

$$\chi_A = (1 - p)(\lambda - \frac{\lambda^2}{4\pi^2}\bar{U}^2).$$

PARAMETER ESTIMATION

In this section we provide a brief review of methods for the estimation of numerical individual parameters of the Boolean model. As mentioned above, they are the intensity and the parameters of the probability distribution of the primary grain. A great number of methods are available which can be classified into two general types: Minimum Contrast Methods and Moment Methods. In practice we will need to combine some of them to obtain the estimation of all the parameters of the model.

Minimum Contrast Method

This method was first introduced for contact distribution functions (Dupac, 1980; Diggle, 1981) although it can be used for other aggregate functions.

Let B be a structuring element; we consider the contact distribution function $H_B(r)$ (Eq. 5). Its corresponding logarithmic transform (henceforth referred to as the logarithmic contact distribution function) is equal to:

$$H_B^l(r) = -log(1 - H_B(r)) = -\lambda E[A(\Xi_0 \oplus r\check{B})] - \bar{A},$$

and on applying the Steiner formula (Matheron, 1974):

$$H_B^l(r) = -\lambda (A(B) + \frac{1}{2\pi}\bar{U}U(B)).$$

For typical structuring elements such as disks, lines or squares, $H_B^l(r)$ is a polynomial function. The essence of the minimum contrast method lies in finding the polynomial $f(r)$ which best approximates the empirical logarithmic contact distribution function. The estimators of individual parameters can be obtained from their coefficients.

As an example, we can examine the case of the spherical contact distribution function. If the structuring element B is the unit disk,

$$H_B^l(r) = \lambda \pi r^2 + \lambda r\bar{U}$$

and

$$\frac{H_B^l(r)}{r} = \lambda \pi r + \lambda \bar{U}$$

is linear, if $f(r) = a + br$ is fitted to it then:

$$\hat{\lambda} = \hat{b}/\pi,$$
$$\hat{\bar{U}} = \hat{a}\pi/\hat{b}.$$

METHOD OF MOMENTS

This method is similar to the method of moments in classical statistics, namely, the estimators of parameters for the Boolean model are chosen to match the empirical values of aggregate parameters. Individual parameters are estimated with the relations given in the previous section being taking into account:

$$\hat{p} = 1 - \exp - \hat{\lambda}\hat{\bar{A}}, \tag{6}$$

$$\hat{L}_A = \hat{\lambda}(1 - \hat{p})\hat{\bar{U}}, \tag{7}$$

$$\hat{\chi}_A = (1 - \hat{p})(\hat{\lambda} - \frac{\hat{\lambda}^2}{4\pi}\hat{\bar{U}}^2), \tag{8}$$

$$\hat{N}_A^+ = \hat{\lambda}(1 - \hat{p}). \tag{9}$$

It is worthwhile to note that the determination of $\hat{\chi}_A$ and $\hat{N}_A$ may be difficult in a discrete context. With respect to the first one, it is usual that popular software package of mathematical computations, like MATLAB, has implemented efficient functions to calculate it. With respect to the second one, our experience in the practical image analysis, tell us that good results are obtained by considering as a lower tangent point, any configuration belonging to the one of the types shown in Fig. 7.

BOOLEAN MODEL SIMULATION

The simulation of a Boolean model is very simple. Let W be a rectangular window of size $m \times n$, λ the germ intensity and Ξ_0 the typical grain. The algorithm is:

Algorithm 1 *1. Generate $k = \mathrm{Po}(\lambda A(W))$.*

2. For $i = 1$ to k:

   (a) Generate $x_1^{(i)} = \mathrm{Un}(0,n)$, $x_2^{(i)} = \mathrm{Un}(0,m)$, $x_i = (x_1^{(i)}, x_2^{(i)})$

   (b) Generate Ξ_i following the distribution of Ξ_0. Draw $\Xi_i + x_i$

Po() and Un() denote the Poisson and the uniform distribution respectively. In the case when $\lambda A(W)$ is large then some form of rejection technique must be used in step 1 (Ripley, 1987). In order to correct the edge effect, this algorithm should be applied to a larger window W^* such that the probability of a grain centered in the boundary of W^* intersects W is depreciable. In Fig. 1 we can see various simulations of a Boolean model in a 512×512 window. The primary grains are random balls, random line segments and random squares respectively.

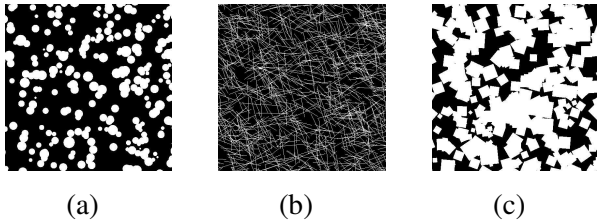| (a) | (b) | (c) |

Fig. 1. *Three simulations of a Boolean model in a 512×512 window. (a) Balls with random radii $R \sim N(10,2)$ and $\lambda = 0.001$. (b) Line segments with random length $L \sim N(30,10)$ and $\lambda = 0.005$. (c) Squares with random side $l \sim N(30,10)$.*

SIMULATION STUDY

In this section we carry out a simulation study in order to compare the performance of the different parameter estimation procedures. We simulated 50 Boolean models in a 512×512 window whose primary grains are random discs with Gaussian radii. Thus the parameters to be estimated are the intensity and the first and second moment $(E(R), E(R^2))$ of the random radius. (We have to note that in all our simulation studies we will use $E(R^2)$ instead of the variance because most estimation methods provide the direct estimation of this parameter). Two different intensity values were used for each experiment: 0.001 and 0.0005. For each intensity value, two values for first and second moments were used: (20, 449) and (10, 109).

As explained in section "Parameter estimation" a great number of methods could be used by combining the different types previously explained. We used the three following methods:

1. The first method is based on the minimum contrast method to estimate the intensity and the mean of the radius and Eq. 6 to estimate the second moment of the radius. The minimum contrast method is applied to the contact distribution function with a square as structuring element (we used a square instead of a disk because its evaluation is better in a discrete screen). We call this method 1.

2. The second method is the method of moments using Eqs. 6, 8 and 9 (in tables, method 2).

3. The last method is identical to the second one but we use Eqs. 6, 7 and 8 (method 3).

Table 1. *Results of the experimental study of Section 3.6: sample means (standard deviation) of the estimates of the different parameters for each one of the methods for a Boolean model with $\lambda = 0.001$.*

		method 1	method 2	method 3
$E(R) = 20$	$\hat{\lambda}$	$1.2e^{-3}(6.7e^{-4})$	$9.1e^{-4}(1.2e^{-4})$	$1.5e^{-3}(2.5e^{-4})$
$E(R^2) = 449$	$\hat{E(R)}$	26.1(30.4)	23.8(1.9)	17.3(2.3)
	$\hat{E(R^2)}$	598.6(672.3)	529.7(59.8)	325(51.3)
$E(R) = 10$	$\hat{\lambda}$	$1.1e^{-3}(1.8e^{-4})$	$9.0e^{-4}(6.3e^{-5})$	$1.2e^{-3}(9.9e^{-5})$
$E(R^2) = 109$	$\hat{E(R)}$	9.9(2.4)	10.5(0.9)	11.2(0.16)
	$\hat{E(R^2)}$	114.1(23.5)	133.9(6.3)	100.0(15.5)

Table 2. *Results of the experimental study of Section 3.6: sample means (standard deviation) of the estimates of the different parameters for each one of the methods for a Boolean model with $\lambda = 0.0005$.*

		method 1	method 2	method 3
$E(R) = 20$	$\hat{\lambda}$	$6.0e^{-4}(2.6e^{-4})$	$5.1e^{-4}(6.4e^{-5})$	$6e^{-4}(1.0e^{-4})$
$E(R^2) = 449$	$\hat{E(R)}$	18.2(5.5)	21.1(1.7)	19.3(1.6)
	$\hat{E(R^2)}$	424.5(125)	465.3(44.9)	357(41.1)
$E(R) = 10$	$\hat{\lambda}$	$5.5e^{-4}(1.4e^{-4})$	$4.6e^{-4}(4.4e^{-5})$	$5.7e^{-4}(8.1e^{-4})$
$E(R^2) = 109$	$\hat{E(R)}$	9.6(2.7)	8.4(2.2)	11.8(4.8)
	$\hat{E(R^2)}$	112.6(25.5)	128.6(8.1)	106.1(13.0)

The results are shown in Tables 1 and 2. These tables show the sample means and standard errors of the estimates of the different parameters for each one of the previous methods. It can be seen that all of them provide good estimations: The sample mean is in all cases very close to the real value and the variance is not too big except in method 1, this method is less efficient.

BOOLEAN MODEL GENERALIZATIONS

In this section, we study some generalizations or modifications of the Boolean model that may be used for the simulation of structures of greater complexity.

GERM-GRAIN MODELS

The Boolean model presupposes very strong assumptions: homogeneous Poisson process of germs and independence between germs and grains. Unfortunately, a great number of real images cannot assume these premises. The first type of generalization of a Boolean model studied in this paper attempts to relax some of these assumptions. The relaxation of these assumptions leads to the class of germ-grain models, a general definition of which can be found in Stoyan *et al.* (1995) and is as follows:

Definition 2 *Suppose* $\Phi_\lambda = \{x_1, \ldots, x_n, \ldots\}$ *is a point process in* $\mathbb{R}^2$. *Let* $\Xi_1, \Xi_2, \ldots$ *be a sequence of random compact sets in* $\mathbb{R}^2$. *A germ-grain model* Ξ *is:*

$$\Xi = \bigcup_{i:x_i \in \Phi_\lambda} (x_i + \Xi_i) . \tag{10}$$

The points x_i are called the *germs* and the sets Ξ_n are known as the *grains* of the model.

In this section we will study three particular cases of germ-grain models: cluster germ-grain model, non-homogeneous germ-grain model and Gibbs process of non-intersecting grains.

CLUSTER GERM-GRAIN MODEL

This model is defined as a germ-grain model (10) whose germ process is a cluster Poisson point process independent of the grains, and the grains are independent and identically distributed as Ξ_0. Note that this definition is fairly different from the definition of Boolean cluster model given in Saxl and Rataj (1996) and Rataj and Saxl (1997).

A cluster point process (Neyman, 1939; Neyman and Scott, 1979; Diggle, 1983) is generated as follows: first, a Poisson point process of intensity ρ is generated. This will be called the *parent process*. A random non negative integer is associated to each parent, the number of offsprings. The offspring of a given parent is located around its parent independently and according with a given probability distribution. The final point process is composed only of the different offspring, i.e., the parent process is not considered.

Some formulas for Cluster germ-grain models can be found in Last and Holtmann (1999).

Simulation

In order to simulate the cluster germ-grain model we simply have to substitute the germ process generation step in the Boolean model simulation algorithm 1 by:

1. Generate $k = \text{Po}(\rho A(W))$

2. For $i = 1$ to k:
   (a) Generate $y_1^{(i)} = \text{Un}(0, n)$, $y_2^{(i)} = \text{Un}(0, m)$, $y_i = (y_1^{(i)}, y_2^{(i)})$
   (b) Generate l_i following a univariate discrete distribution $f(l)$
   (c) For $j = 1$ to l_i. Generate $x_{ij} = u_j + y_i$ with u_j following a bivariate continuous distribution $h(u)$
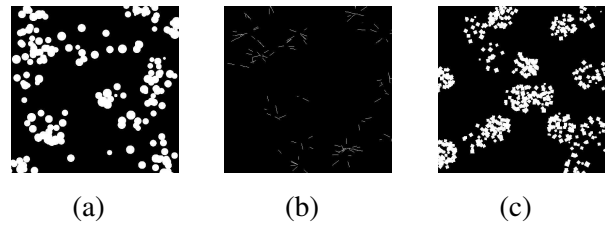

(a)   (b)   (c)

Fig. 2. *Three simulations of a cluster germ-grain model in a* 512×512 *window with* $\rho = 0.0001$, *the position of the offsprings with respect to their parents are uniform in the ball of radius* 40 *and the number of offsprings per parent follows a discrete uniform distribution on* $\{0, \ldots, 10\}$. *(a) balls with random radii* $R \sim N(10, 4)$, *(b) line segments with random length* $L \sim N(10, 4)$ *and (c) squares with random side* $l \sim N(10, 4)$.

Fig. 2 shows three simulations of this model where parents have intensity $\rho = 0.0001$, the positions of the offsprings with respect to their parents are uniformly distributed on the ball of radius 40 and the number of offsprings per parent follows a discrete

uniform distribution on $\{0,\ldots,10\}$. In the first image, the grains are balls with random radii following a Gaussian distribution with mean, $E(R) = 10$, and second moment, $E(R^2) = 104$. In the second image they are random line segments of random length with the same distribution. The last image shows the realization of a model whose grains are random squares, again with the same distribution for their sides.

NON-HOMOGENEOUS GERM-GRAIN MODEL

A non-homogeneous germ-grain model is a germ-grain model (Eq. 10) whose germ process is a non-homogeneous Poisson point process independent of the grains and the grains are independent and identically distributed as Ξ_0.

A non-homogeneous Poisson process is obtained when the constant intensity of the Poisson process is substituted by a general intensity measure $\Lambda(A)$, for $A \subset \mathbb{R}^2$, usually $\Lambda(A) = \int_A \lambda(x)\,dx$.

It can easily be shown that the capacity functional of this germ-grain model is given by:

$$T_X(K) = 1 - \exp\left[-E\left\{|\check{\Xi}_0 \oplus \check{K}|_\lambda\right\}\right], \quad (11)$$

with

$$|\check{\Xi}_0 \oplus K|_\lambda \equiv \int_{\mathbb{R}^2} I_{\check{\Xi}_0 \oplus K}(x)\lambda(x)\,dx.$$

It should be noted that this model is not stationary.

Example

Let us consider $\lambda((x_1,x_2)) = \lambda_0 x_1$, i.e. the intensity at point x is proportional to its abscisa, and $\Xi_0 = B(0,R)$ with random R. A similar type of heterogeneity is considered in Hahn *et al.* (1999). This model is not stationary and the definitions of the area fraction and the contact distribution function will depend on $x = (x_1,x_2)$.

If we consider $K = B(x,t)$ the exponent in Eq. (11) has the expression

$$E \int_{(\check{\Xi}_0 \oplus K)} \lambda(y)dy = E\left[\int_{B(x,R+t)} \lambda_0 y_1 dy\right] = E\left[\lambda_0 x_1(R+t)^2\pi\right].$$

For $t = 0$ the expression of the area fraction is:

$$p(x) = 1 - \exp(-E[\lambda_0 R^2 x_1 \pi]), \quad (12)$$

and the contact distribution function is:

$$H_{B(x,1)}(t) = 1 - \exp-(\lambda_0 x_1 \pi t(t + 2E(R))). \quad (13)$$

Parameter estimation

In this section we study parameter estimation in the particular non-homogeneous germ-grain model given in the example of the last section.

The capacity functional $T_\Xi(K)$ and the area fraction of this model, and in general of any non stationary model, can be estimated by using a kernel estimator. Let us consider $\{y^{(1)},\ldots,y^{(m)}\}$ a grid of points in the observation window W, then we estimate the capacity functional at $K = B(x,t)$ by:

$$\hat{T}_\Xi(B(x,t)) = \frac{\sum_{y^{(i)}\in\Xi\oplus B(0,t)} k(x - y^{(i)})}{\sum_{i=1}^m k(x - y^{(i)})}, \quad (14)$$

with $k()$ denoting a kernel function.

Note that if $t = 0$, we have the estimation of the area fraction. But for the type of heterogeneity considered in this example (vertical linear trend), one may estimate the capacity functional by the length fraction of $\Xi \oplus \check{K}$ along the horizontal line with abcisa x_1. Hence no bivariate kernel is needed.

Taking into account Eq. (13), we can estimate individual parameters of this model using a minimum contrast method. If we assume that radii are Gaussian, we have to estimate λ_0, $E(R)$ and $E(R^2)$. The expression of the logarithmic contact distribution function is:

$$-\frac{\log\{1 - H_{B(x,1)}(t)\}}{t} = \lambda_0 x_1 \pi(t + 2E(R)).$$

Given x we can estimate λ_0 and $E(R)$ by fitting a linear function to the empirical logarithmic contact distribution function. However, a better estimation will be obtained by taking a set of points $\{x^{(1)},\ldots,x^{(n)}\}$, estimating $f_i(t) = -\log\{1 - H_{B(x,1)}(t)\}/tx_1^{(i)}$, taking their sample mean, $\bar{f}(t)$ and fitting a linear function to $\bar{f}(t)$. $E(R^2)$ can be estimated from the expression of the area fraction (12).

Simulation

The simulation of this model in a rectangular window W is again very simple. The germ process step generation in the Boolean model algorithm 1 is substituted by (see Diggle, 1983):

1. $\lambda_0 = \max_{x\in W}\lambda(x)$. Generate $k = \text{Po}(\lambda_0)$

2. For $i = 1$ to k:

   (a) generate $x_1^{(i)} = \text{Un}(0,n)$, $x_2^{(i)} = \text{Un}(0,m)$, $x_i = (x_1^{(i)}, x_2^{(i)})$

(b) generate $p = \text{Un}(0,1)$

- if $p < \lambda(x_i)/\lambda_0$, generate Ξ_i following the distribution of Ξ_0 and draw $\Xi_i + x_i$
- else x_i is deleted

Fig. 3 shows some simulations of the example of the non-homogeneous germ-grain model given in this section with Gaussian radii.
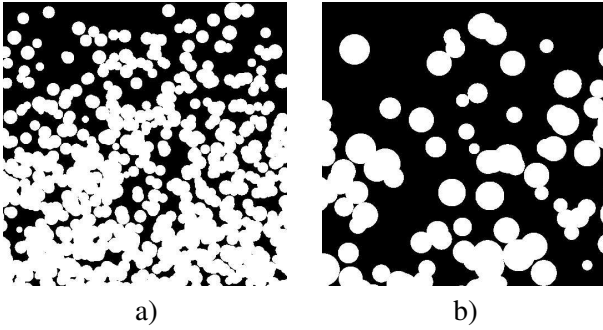

<div align="center">a)   b)</div>

Fig. 3. *Two simulations of a non-homogeneous germ-grain model whose intensity is proportional to the abcisa coordinate. Grains are random balls with Gaussian radii and parameters: a)* $\lambda_0 = 0.00001$, $E(R) = 10$ *and* $E(R^2) = 104$, *b)* $\lambda_0 = 0.000001$, $E(R) = 20$ *and* $E(R^2) = 425$.

Simulation study

A simulation study was again carried out in this section. 50 realizations of the example of the previously studied non-homogeneous germ-grain model were generated in a 512×512 window. The grains are random balls with Gaussian radii and all of them have parameters $\lambda_0 = 2e^{-6}$, $E(R) = 20$ and $E(R^2) = 449$. We estimated the parameters by using the method explained in Section "Parameter estimation". The results are shown in Table 3 and, as can be seen, they were quite acceptable.

$\hat{\lambda}$	$E(\hat{R})$	$E(\hat{R}^2)$
$3e^{-6}(1.03e^{-6})$	14.3(7.4)	344.5(134.1)

Table 3. *Results of the experimental study of Section 4.3.4: sample mean (standard deviance) of the estimates of* 50 *realizations with* $\lambda_0 = 2e^{-6}$, $E(R) = 20$ *and* $E(R^2) = 449$.

Gibbs process of non-intersecting grains

This model relaxes the assumption of independence between germs and grains in such a way that the grains do not overlap. Their formal definition is as follows (see Stoyan *et al.*, 1995).

Definition 3 *Let* $\Omega = \mathbb{R}^2 \otimes \mathbb{K} = \{w = (z,K) : z \in \mathbb{R}^2; K \in \mathbb{K}\}$
Let $\Lambda_0 = \{w_n\}_{n \geq 1}$ *be a point process in* Ω *where:*

1. $\{z_n\}_{n \geq 1}$ *is a stationary point process in* $\mathbb{R}^2$.

2. $\{K_n\}_{n \geq 1}$ *i.i.d* K_0.

3. $\{z_n\}_{n \geq 1}$ *and* $\{K_n\}_{n \geq 1}$ *are independent.*

Let us consider the following neighborhood relation:
$w_1 \sim w_2$ *iff* $\Xi_1 \cap \Xi_2 \neq \emptyset$ *where* $\Xi_n = K_n + z_n$.
Let $\Lambda = \{w_n\}$ *be a Gibbs process with density with respect to* Λ_0:

$$p(w_1, \ldots, w_n) = \frac{1}{z} \prod_{i<j} \exp\{\alpha + \theta(w_i, w_j)\},$$

where

$$\theta(w_i, w_j) = \begin{cases} -\infty & \text{if } w_1 \sim w_2, \\ 0 & \text{otherwise}. \end{cases}$$

The Gibbs process of non-intersecting grains is:

$$\Xi = \bigcup_{w_n \in \Lambda} \Xi_n.$$

This model was studied in Mase (1986) and Stoyan (1989) with circular grains. In Ayala and Simó (1995) this model with elliptical grains was used to model nerve fiber.

Parameter estimation

Individual parameters of this model are α and the parameters of the probability distribution of K_0. Their estimation was studied for the case of circular grains with random radii in Stoyan (1989) and is as follows.

Let λ be the intensity of the Gibbs process, $m(r)$ the density function of the probability distribution of the radii of the Gibbs process while $p(r)$ denotes the probability that a disk centered in the origin and radius r intersects the Gibbs process. All these parameters are directly estimable from the observation window. Let $m_0(r)$ be the density function of the probability distribution of the radius of K_0. We have the following relation:

$$\lambda m(r) = e^{-\alpha} p(r) m_0(r).$$

And the estimation of $m_0(r)$ is obtained:

$$\hat{m}_0(r) = \frac{\hat{\lambda} \hat{m}(r) e^{\hat{\alpha}}}{\hat{p}(r)}$$

with $\hat{\alpha}$ chosen so that:

$$\int_0^\infty \hat{m}_0(r) dr = 1.$$

This method could be generalized, for example, to ellipses or line segments.

Simulation

Different methods of simulating general Gibbs processes can be found in the literature (Ripley, 1981; Van Lieshout, 2000), most of which are based on the so-called spatial birth-and-death (b-and-d) processes. The simulation begins with a start configuration which is then changed step by step, where points disappear ('die') and new ones are generated ('born'). In our experiments, we have used a Metropolis-Hasting (M-H) type algorithm. A M-H algorithm is a discrete time Markov process where the transitions are defined in two steps, a proposal for a new state (in our case a birth or a death) is made that it is subsequently accepted or rejected based on the likelihood ratio of the new state compared to the old one. The spatial b-and-d process is a time continuous Markov process in which all transitions are accepted with probability one, but, the process stay in state x for an exponentially distributed random sojourn time. We have used one of the simplest M-H algorithm: births and deaths are equally likely and sampled uniformly.

Let W be the window where the process is generated.

Algorithm 2 *1. An initial configuration Ξ_0 with k non intersecting disks is generated, with k being arbitrary. Make $n = k$, $\Xi = \Xi_0$.*

2. with probability $\frac{1}{2}$:

(a) Generate a new point x' uniform in W and radius r' from $m_0(r)$. If $\Xi \cap B(x',l') \neq \emptyset$ the new disk is rejected. If not,

   i. If $n \leq e^{-\alpha}A(W) - 1$ the new disk is accepted.

   ii. If $n \geq e^{-\alpha}A(W)$ the new disk is accepted with probability $\frac{e^{-\alpha A(W)}}{n+1}$.

(b) Choose (x,r) at random of Ξ.

   i. If $n \geq e^{-\alpha}A(W)$ this disk dies.

   ii. If $n \leq e^{-\alpha}A(W) - 1$ the disk dies with probability $\frac{n}{e^{-\alpha A(W)}}$.

Although we have choose this algorithm because it is simpler to implement, we have to warn that, in general, this algorithm could result in a low acceptance probability, especially for models exhibiting strong interaction, that would make it inefficient. In this case we should use other kind of proposals or well a spatial b-and-d process. See Clifford and Nicholls (1994) for an excellent comparison.

Another problem here is to asses how long the chain should run in order to achieve the desired approximation. This problem could be solved by using the recently developed perfect or exact simulation (Propp and Wilson, 1996). Fig. 4 shows some simulations of this model.
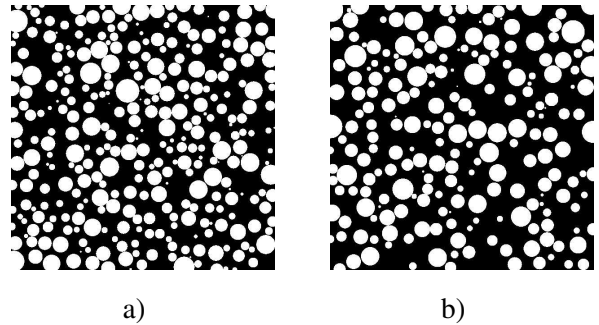

<div align="center">a)   b)</div>

Fig. 4. *Two simulations of a Gibbs process of non-intersecting grains with: a) $\alpha = 3$ and b) $\alpha = 4$, the grains are random balls with radii following a Gaussian distribution $N(20,36)$.*

Simulation study

In this section, a simulation study is again carried out to experimentally test the performance of the previously described parameter estimation procedure. In order to do this, we simulated 20 Gibbs processes of non-intersecting grains with $\alpha = 3$ and 20 with $\alpha = 4$. The grains are in both cases random balls following a Gaussian distribution with first and second moment 20 and 449 respectively. The sample means and standard deviation of the estimates are shown in Table 4. The results are in general fairly good. The results for $\alpha = 4$ are a slightly better, they are less variable and their mean is nearer to the real value.

	$\hat{\alpha}$	$E(\hat{R})$	$E(\hat{R^2})$
$\alpha = 3$	2.7 (1.7)	23.5 (2.9)	572.5 (46.4)
$\alpha = 4$	4.4 (0.67)	22.9 (3.6)	557.41 (38.2)

Table 4. *Results of the experimental study of Section 4.3.8: sample means (standard deviation) of the estimates of 20 realizations with $E(R) = 20$, $E(R^2) = 449$ and $\alpha = 3$ and $\alpha = 4$, respectively.*

THREE-PHASE BOOLEAN MODEL OR SINTER TEXTURES

The departure from the Boolean model studied in this section consists of considering a three-phased Boolean model. This model is defined in a very simple way. Suppose that X_1 and X_2 are independent Boolean models. Then we define:

$$\Xi_1 = X_1 \,, \quad \Xi_2 = X_2 \cap X_1^c \,, \quad \Xi_3 = X_2^c \cap X_1^c \,,$$

and we call (Ξ_1, Ξ_2, Ξ_3) the three-phased Boolean model. Note that the complete Boolean model X_1 forms component 1. Component 2 can be interpreted as a pattern destroyed in part by X_1 and Ξ_3 is the background. This three-phased model was introduced in Serra (1982) and applied to the description of sinter materials.

It can easily be shown that the area fraction fulfills:

$$p_{X_2} = \frac{p_{\Xi_2}}{1 - p_{X_1}} .$$

Parameter estimation

Since X_1 is a completely observable Boolean model its individual parameters can be estimated using the methods from section "The Boolean Model".

In order to estimate the parameters of X_2, we can use the equation of the capacity functional for union-censoring given in Molchanov (1997):

$$T_{X_2}(K) = \frac{T_{X_1 \cup X_2}(K) - T_{X_1}(K)}{1 - T_{X_1}(K)}. \qquad (15)$$

Using Eq. (15) the contact distribution function of X_2 can be estimated and the parameter of X_2 can be estimated using the minimum contrast method applied to the empirical logarithmic contact distribution function as explained in Section "The Boolean model".

Simulation

The simulation of this model is trivial when its definition and the algorithm 1 are taken into account. Fig. 5 shows a simulation of a three-phased Boolean model, both Boolean models have intensity $\lambda = 0.0005$ and random balls with Gaussian radii with parameters $E(R) = 20$ and $E(R^2) = 449$.
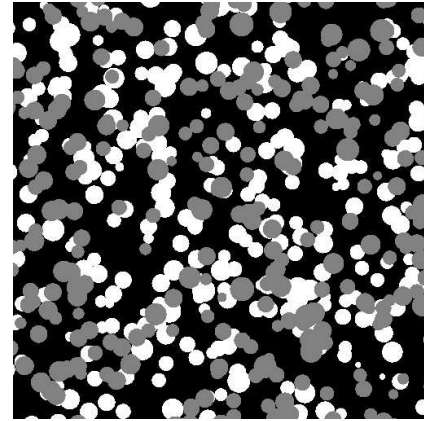

Fig. 5. *A simulation of a three-phased Boolean model. Both Boolean models have intensity $\lambda = 0.0005$ and random balls with Gaussian radii with parameters $E(R) = 20$ and $E(R^2) = 449$.*

SIMULATION STUDY

In this section we study the performance of the parameter estimation procedure, again by means of a simulation study. We applied the previously explained procedure to 20 simulations of a three-phased Boolean model. X_1 and X_2 are both Boolean models with intensity $\lambda = 0.0005$ and random balls with Gaussian radii with parameters $E(R) = 20$ and $E(R^2) = 449$. The results are provided in Table 5. As was expected, the estimates of X_1 parameters are more precise because the model is completely observed.

THE BOOLE-POISSON MODEL

The last Boolean model generalization studied in this paper is obtained by restricting the Boolean realizations to remain inside the polygons generated by a Poisson process of lines. It is called the Boole-Poisson model (Serra, 1982).

Table 5. *Results of the experimental study of Section 4.5: sample means (standard deviation) of the estimates of 20 realizations with $E(R) = 20$, $E(R^2) = 449$ and $\lambda = 0.0005$ for both Boolean models.*

	$\hat{\lambda}$	$E(\hat{R})$	$E(\hat{R^2})$
Boolean 1	0.00056 (0.00023)	22.53 (14.52)	483.8 (313.6)
Boolean 2 (hidden)	0.00068 (0.0005)	26.07 (25.5)	568.14 (629.6)

A line process is a random collection of lines in the plane which is locally-finite; i.e. only a finite number of lines hit each compact planar set. A line process can be seen as a particular case of point process on $\mathbb{R}^2$, because we can parameterize it as: $(p, \phi) \in \mathbb{R} \times$ $(0, 2\pi] \subset \mathbb{R}^2$, with p being the perpendicular distance of the line to the origin and ϕ the angle that the perpendicular makes with the positive x-axis. Under this parameterization, the set G of all the lines in the plane is equivalent to:

$$S = \{(p,\phi) \,:\, 0 \le p < \infty, 0 \le \phi < 2\pi\}.$$

A line process is a point process on S. If K is a compact planar set and G_K the set of lines hitting K, the corresponding subset in S will be called S_K. The measure of these sets is given by the Lebesgue measure considered as subsets of $\mathbb{R}^2$.

A Poisson line process is the line process produced by a Poisson process on S.

Let us now see the definition of Poisson-Boolean model.

Let X_1 be an isotropic Poisson lines process in $\mathbb{R}^2$ with intensity λ. This process generates a random tessellation, each polygon Π of the tessellation is intersected with a realization of a Boolean model X_2 with intensity θ and primary grain X_2^0, in such way that a different realization of X_2 is used for each polygon. The union of all these portions of X_2, together with X_1 make up a set Y that is called the Poisson-Boolean model.

If T_1, T_2 and $T(B)$ denote the capacity functional of X_1, X_2 and Y respectively, we have:

$$T(B) = 1 - \exp\{-\lambda U(B) - \theta E[A(X_2^0 \oplus \check{B})]\}.$$

Applying the Steiner formula and taking the logarithm, we obtain the following expression:

$$-\log\{1 - T(B)\} = \lambda U(B) +$$
$$\theta\left(A(X_2^0 + \frac{1}{2\pi}U(X_2^0)U(B) + A(B)\right). \quad (16)$$

This equation together with the expression of the area fraction:

$$p = 1 - \exp\{\theta A(X_2^0)\}, \quad (17)$$

and will be used in the next section to estimate the parameters of the model.

Parameter estimation

Individual parameters of this model are the individual parameters of the Boolean model. To estimate them we previously need to estimate the intensity of the line process. This is an aggregate parameter and it can be estimated as a spatial mean.

$$\hat{\lambda} = \frac{\text{Number of lines intersecting } W}{A(S_W)}.$$

However, this estimation has added difficulties in practice. The lines in the image have to be counted automatically, which is not a simple task. To count the number of lines in the image we take into account the fact that this number is equal to the number of local minima of the Radon transformation of the image.

Once λ has been estimated, we again use the minimum contrast method to estimate the rest of the parameters of the model. From Eqs. (16) and (17) we obtain the expression of the logarithmic spherical contact distribution function

$$\frac{-\log\{1 - H_B(r)\}}{r} = \lambda 8 + \theta\left(\frac{1}{2\pi}U(X_2^0)8 + 9r\right).$$

The estimates are obtained by fitting a linear function to the empirical function.

SIMULATION

The simulation algorithm of this model has two basic steps. In the first step the lines process is generated and in the second, a Boolean model is simulated in each polygon following algorithm 1. The simulation of the line process step is as follows (Stoyan and Stoyan, 1994):

(i) Let W be the window where the process is to be simulated (To simplify the simulation we assume that W is square). Find the set S_W in S and its Lebesgue measure. (If W is convex, as in this case, this is the measure of the boundary of W)

(ii) Generate a Poisson-distributed random number n with parameter $\lambda A(S_W)$.

(iii) Generate n independent random lines using the following steps (to simplify the notation, assume that W is the unit square):

  1. $u = \text{Un}(0,1)$

  2. If $u > \frac{1}{\sqrt{2}}$ go to (5)

  3. $v = \text{Un}(0,1)$

  4. $p = \frac{u}{\sqrt{2}}$: $\phi = 2\pi v$: go back to (1)

  5. $v = \text{Un}(0,1)$: $\phi = 2\pi v$

  6. $T = \sqrt{2}(sin(\phi) + cos(\phi))$

  7. If $u > T$ then (1)

  8. $p = \frac{u}{\sqrt{2}}$: go back to (1).

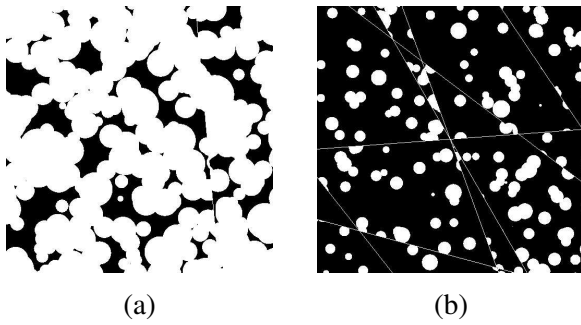Fig. 6 shows two simulations of this model.

(a) (b)

Fig. 6. *Two simulations of a Poisson-Boolean model where the grains are random balls with radii following a Gaussian distribution with parameters (a)* $\lambda = 0.0001$, $\theta = 0.001$, $E(R) = 20$ *and* $E(R^2) = 449$; *(b)* $\lambda = 0.0005$, $\theta = 0.002$, $E(R) = 10$ *and* $E(R^2) = 109$.

Fig. 7. *Pixel patterns used to detect lower tangent points.*

Simulation study

In this section, we again carry out a simulation study to show the performance of the former method. We simulated 20 Poisson-Boolean model realizations with parameters $\lambda = 0.002$, $\theta = 0.00005$. The grains of the Boolean model are balls with random radii following a Gaussian distribution with $E(R) = 20$ and

$E(R^2) = 449$. The results can be seen in Table 6. They are relatively good in the estimation of λ and θ, a slightly less so for the mean of the radius and not so good for the variance.

CONCLUSIONS

In this paper we reviewed the use of Random Closed Sets as a powerful tool in simulating random binary images able to imitate natural textures. The Boolean model and a class of RACS models that is a generalization derived from it have been studied. For each model simulation algorithms and parameter estimation procedures were given.

In addition a library of functions to be used with MATLAB was written to carry out the different tasks that we refer to.

Future work could include other generalizations of Boolean Model with different point processes like Strauss point processes or area-interaction point processes.

ACKNOWLEDGMENTS

Table 6. *Results of the experimental study of Section 4.7.1: sample means (standard deviation) of the estimates of a Poisson-Boolean model with* $\lambda = 0.002$, $\theta = 0.00005$, $E(R) = 20$ *and* $E(R^2) = 449$.

$\hat{\lambda}$	$\hat{\theta}$	$E(\hat{R})$	$E(\hat{R^2})$
0.001 (5.84e^{-7})	0.00008 (3.41e^{-5})	15 (10.39)	130.6 (54.6)

REFERENCES

Ayala G (1988). Inferencia en Modelos Booleanos. Tesis Doctoral. Universidad de Valencia.

Ayala G, Simó A (1995). Random closed sets and nerve fiber degeneration. Adv Appl Probab 27:293–305.

Clifford P, Nicholls G (1994). Comparison of birth-and-death and metropolis-hastings markov chain monte carlo for the strauss process. Tech. rep., Manuscript. Oxford University.

Cressie N (1993). Statistics for spatial data. Wiley Series in Probability and Mathematical Statistics.

Diggle P (1981). Binary mosaics and the spatial pattern of heather. Biometrics 37:531–9.

Diggle P (1983). Statistical analysis of spatial point patterns. Academic Press.

Dupac V (1980). Parameter estimation in the poisson field of disc. Biometrika 67:187–90.

Hahn U, Micheletti A, Pohlink R, Stoyan D, Wendrock H (1999). Stereological analysis and modeling of gradient structures. J Microsc 195:113–24.

Last G, Holtmann M (1999). On the empty space function of some germ-grain models. Pattern Recognition 32:1587–600.

Lyman T (1972). Metals Handbook. Am. Soc. Metals. Ohio.

Margalef R (1974). Ecología. Omega. Barcelona.

Mase S (1986). On the possible form of size distributions for gibbsian processes of mutually non-intersecting balls. Appl Prob :646–59.

Matheron G (1974). Random Sets and Integral Geometry. New York: J. Wiley and Sons.

Molchanov I (1997). Statistics of the Boolean Model for Practitioners and Mathematicians. Wiley Series in Probability and Mathematical Statistics.

Neyman J (1939). On new class of contagious distributions, applicable in entomology and bacteriology. Ann Math Statist 10:35–57.

Neyman J, Scott E (1979). Statistical aproach to problems of cosmology (with discussion). R Statist Soc B 20:1–43.

Plaza M (1991). Contrastes en modelos germen y grano. Tesis Doctoral. Universidad de Valencia.

Propp J, Wilson D (1996). Exact sampling with coupled markov chains and applications to statistical mechanics. Proceedings of the Seventh International Conference on Random Structures and Algorithms 9:223–52.

Rataj J, Saxl I (1997). Boolean cluster models: mean cluster dilations and spherical contact distances. Math Bohem 122:21–36.

Ripley B (1981). Spatial Statistics. J. Wiley and Sons.

Ripley B (1987). Stochastic simulation. J. Wiley and Sons.

Saxl I, Rataj J (1996). Spherical contact and nearest neighbour distances in boolean cluster fields. Acta Stereol 15:91–5.

Serra J (1982). Image analysis and mathematical morphology. Academic Press.

Stoyan D (1989). Statistical inference for a gibbs point process of mutually non-intersecting discs. Biometrik J 31:153–61.

Stoyan D, Kendall W, Mecke J (1995). Stochastic Geometry and its applications. Chichester: J. Wiley and Sons.

Stoyan D, Stoyan H (1994). Fractals, random shapes and point fields. Methods of geometrical statistics. Wiley series in probability and mathematical statistics.

Van Lieshout M (2000). Markov point processes and their applications. Imperial College Press.

Weil W, Weaieacker J (1984). Densities for stationary random sets and point processes. Adv Apply Probab 16:324–46.