3

Random closed sets I – The Boolean model

3.1 Introduction and basic properties

3.1.1 Model description

The *Boolean model* (also known as *Boolean scheme*, *Poisson germ–grain model*, *Poissonian penetrable grain model*, *fully penetrable grain system*, or *homogeneous system of overlapping particles*) is an important and relatively simple example of a random closed set. It is both flexible and amenable to calculation.

Before a general definition is given it may be helpful to consider one of the simplest possible examples of a planar Boolean model. Suppose points are scattered in the plane according to a homogeneous Poisson process of intensity λ. On each of these points a disc of fixed radius r is placed. The union of all of these discs is an example of a Boolean model; a particular realisation is displayed in Figure 3.1 on p. 65. In the terminology to be introduced the points of the Poisson process are the *germs* of the model while the discs are the *(primary) grains*. (Objects that could be called 'secondary grains' are below called 'clumps'.)

This construction can be generalised to produce the general stationary Boolean model. The discs are replaced by independent realisations of a random compact set. In this introduction the notions of random compact sets and random closed sets have been used without a careful treatment merited by the concept. Chapter 6 gives a more detailed discussion. For this chapter an intuitive understanding will suffice.

Suppose $\Phi = \{x_1, x_2, \ldots\}$ is a homogeneous Poisson process in $\mathbb{R}^d$ of intensity λ. Note that the x_i just list the points of Φ in some (arbitrary) manner. Let $\Xi_1, \Xi_2, \ldots$ be a sequence of independent identically distributed random compact sets in $\mathbb{R}^d$ that are independent of

Stochastic Geometry and its Applications, Third Edition.
Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall and Joseph Mecke.

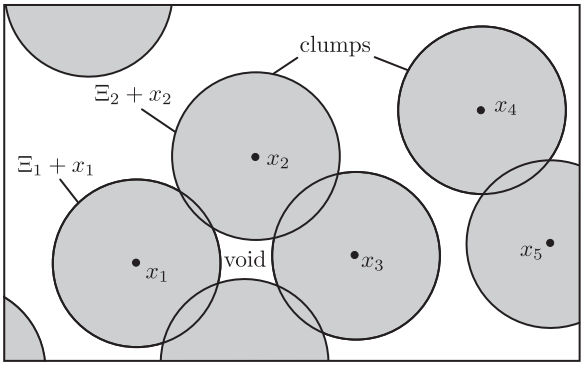


Figure 3.1 Diagrammatic image of a planar Boolean model. The empty space enclosed by four of the grains is a 'void'. Key: germ x_n, •; grain, Ξ_n.

the Poisson process Φ and satisfy a technical condition given below. The *Boolean model* Ξ is constructed by using the *germs* x_n and the *grains* Ξ_n as follows:

$$\Xi = \bigcup_{n=1}^{\infty} (\Xi_n + x_n) = (\Xi_1 + x_1) \cup (\Xi_2 + x_2) \cup \cdots . \tag{3.1}$$

The technical condition mentioned above is that

$$\mathbf{E}\big(\nu_d(\Xi_0 \oplus K)\big) < \infty \qquad \text{for all compact } K. \tag{3.2}$$

Here, and in the following, Ξ_0, called the *typical grain*, denotes a further random compact set of the same distribution as the grains Ξ_n but independent of both the grains and the germ process Φ. The union Ξ is a random closed set called *Boolean model with typical grain* Ξ_0.

The technical condition (3.2) ensures that only finitely many of the grains $\Xi_n + x_n$ meet any given compact set. Thus, in particular, it ensures that the property of being a closed set is inherited by Ξ from the grains. The condition is further motivated by Formula (3.7) below for the probability of Ξ hitting a compact set.

A simpler but more restrictive condition than (3.2) is as follows: if R is the radius of the ball circumscribing Ξ_n then

$$\mathbf{E}(R^d) < \infty, \tag{3.3}$$

see Heinrich (1992b).

The construction of Ξ is illustrated in Figure 3.1, while Figure 3.2 gives an example of a simulated realisation of a Boolean model.

A more theoretical and less constructive point of view defines the Boolean model as the union of sets arising from a Poisson process in the space $\mathbb{K}'$ of all non-empty compact subsets of $\mathbb{R}^d$. Treatments using this approach can be found in Matheron (1975) and Schneider and Weil (2008). The technical condition mentioned above is then naturally expressed as requiring the governing measure of the Poisson process to be a σ-finite Borel measure on $\mathbb{K}'$, using a suitable topology on $\mathbb{K}'$.

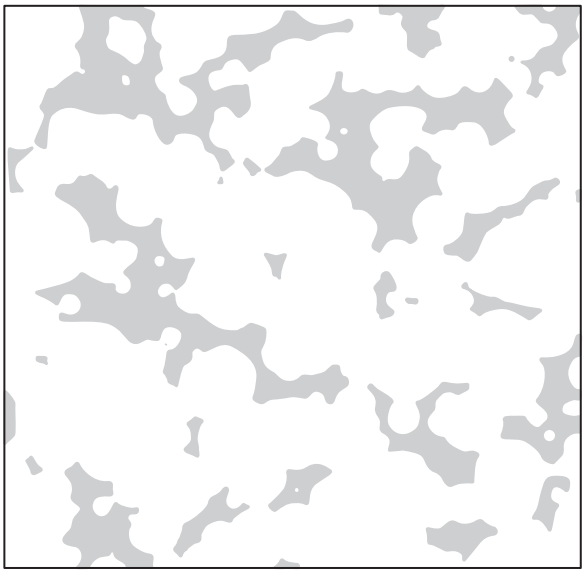


Figure 3.2 Computer simulation of a Boolean model with white discoidal grains. The parameters of the simulation are those fitted to the pattern of a section of a potassium deposit. The white areas correspond to potassium-bearing regions; see Ohser and Stoyan (1980).

Typical possibilities for the grains in the planar case include: discs of random radius (generalising the simple example above), various random polygons, segments of random length and orientation, and random finite clusters of points. The grains are characterised by their distribution M. As they are random sets this distribution is a probability measure on $\mathbb{K}'$. It is the *mark distribution* of the marked point process $\{[x_n; \Xi_n]\}$ as in Section 4.2 and Chapter 6. In the case of a Boolean model the marks are independent.

If the grains are convex then statistical averages of various numerical measures of convex sets play an important rôle in the theory. Their notation and explanation are given in Table 3.1 on p. 67. In the case of nonconvex grains several of the quantities still have meaning and in such cases the same symbols will be used.

The Boolean model has been thoroughly studied by Matheron, Serra and their colleagues; see Matheron (1967, 1975) and Serra (1982) and references therein. More recent references are Hall (1988), Cressie (1993), Stoyan and Mecke (2005), Weil (2007), Schneider and Weil (2008) and Baccelli and Błaszczyszyn (2009a).

3.1.2 Applications

The Boolean model is a basic model in stochastic geometry. As Hadwiger and Giger (1968) show, it also arises from classical assumptions of independence and uniformity in a manner similar to some characterisations of the Poisson process.

There are two general trends of application for the Boolean model. In the first place, it is a natural model for sparse systems of particles distributed at random. Here the sparse nature of the system is modelled by a low value for the intensity λ of the Poisson process of germs.

Table 3.1 Statistical averages of numerical measures of convex sets.

In general Euclidean space $\mathbb{R}^d$:	
$\overline{V}_k$	mean of k^{th} intrinsic volume
For spatial sets, $d = 3$:	
$\overline{V}$	mean volume, $\overline{V} = \overline{V}_3$
$\overline{S}$	mean surface area, $\overline{S} = 2\overline{V}_2$
$\overline{M}$	mean of integral of mean curvature, $\overline{M} = \pi\overline{V}_1$
$\overline{\overline{b}}$	mean average breadth, $\overline{M} = 2\pi\overline{\overline{b}}$, and $\overline{\overline{b}} = \frac{1}{2}\overline{V}_1$
For planar sets, $d = 2$:	
$\overline{A}$	mean area, $\overline{A} = \overline{V}_2$
$\overline{L}$	mean boundary length, $\overline{L} = 2\overline{V}_1$

Note: $\overline{\overline{b}}$ is the statistical mean of a geometric average taken over all possible directions.

If the part of the space covered by Ξ is small, then grains will not often overlap and so Ξ will consist in the main of separated particles. With increasing λ the number of overlappings increases. Of course such overlappings do occur in nature; consider for example pores in cheese or areas of weeds in agricultural fields.

In the second place, the Boolean model may provide a good description for an irregular pattern observed in nature; see for example Figure 3.2. As can be seen from the figure (and in contrast to what is often the case for sparse systems), the grains and their locations at germs generally fail to have operational interpretations. The rôle of the Boolean model in such a case is to provide irregular random sets of a suitable form, and its method of construction needs not correspond to any physical reality. This point is returned to in Section 3.6.

The Boolean model has been used, at least for the special case of spherical grains, since the middle of 19$^{\text{th}}$ century, for example by the physicist Clausius (1858); see Guttorp (2007) and Guttorp and Thorarinsdottir (2012). The following list gives some examples (in alphabetical order of the authors of the earliest papers) drawn mainly from the natural and engineering sciences:

- Armitage (1949): random clumping of dust or powder particles;
- Cahn (1956): kinetics of nucleation at grain boundaries;
- Kallmes and Corte (1960), Corte and Kallmes (1962), Räisänen *et al.* (1997), Niskanen *et al.* (1998) and Åström *et al.* (2000): microstructure of paper; the grains are paper fibres, modelled as long thin rectangles or line segments;
- Diggle (1981): distribution of heather in a forest; see also Møller and Helisová (2010);
- Jacod and Joathon (1971): form of geological structures generated by sedimentation; see also Ohser and Stoyan (1980): form of geological deposits of potassium, Figure 3.2 shows a simulated realisation of a Boolean model using as model parameters those

Figure 3.3 Canopy gaps (in grey) in a forest mapped from aerial photographs. Here the tree crowns (not shown) may be considered as standing for the grains, while the gaps stand for the complement of the random set. Reprinted from Nuske *et al.* (©2009) with permission of Elsevier.

obtained by fitting to the geological data, and the white zones in the figure correspond to areas that are rich in potassium; by the way, a figure in Nuske *et al.* (2009), reproduced in Figure 3.3, showing the canopy gaps in a forest, is quite similar to Figure 3.2;

- Hermann (1991): various examples of the application of the Boolean model in materials science;
- Jeulin *et al.* (2001): microstructure of plaster made of needle-shaped gypsum crystals; the grains are modelled as Poisson polyhedra or random parallelepipeds;
- Kolmogorov (1937), Hermann (1998) and Capasso (2003): modelling the crystallisation in metals and polymers;
- Kopp-Schneider *et al.* (1998) and Groos and Kopp-Schneider (2010): modelling the progression of altered hepatocytes in cancer research; the spherical grains stand for cells forming focal lesions;
- Mack (1954, 1956): corrections for clumping in estimation of number of particles; see also Kellerer (1983);
- Nutting (1913) and Marchant and Dillon (1961): patterns in photographic emulsion; see also Frieden (1983);
- Serra (2009): modelling the spread of forest fire by a Markov chain of Boolean models;
- Thovert *et al.* (2001) and Arns *et al.* (2002): Fontainebleau sandstone, a porous medium; in the former paper the grains are balls, in the latter ellipsoids;

- Thovert and Adler (2004) and Adler *et al.* (2005): fracture networks, where the grains are portions of plane surfaces;
- Widom and Rowlinson (1970): systems of water droplets in the study of liquid-vapour phase transitions.

The grains of a Boolean model are not required to be connected sets. For example, they may be sets of discrete points, in which case the Boolean model is a point process, in fact, a Neyman–Scott point process; see Section 5.3.

The so-called Boolean functions can be viewed as generalisations of Boolean models. Here a number of random functions are randomly translated and combined using the supremum-operation; see Serra (1982, 1988, 1989). This model is applied to modelling random surfaces; see Jeulin (1987, 2000, 2002), Serra (1987), Goulard *et al.* (1994), Heinrich and Molchanov (1994), Chadœuf *et al.* (1996) and Dominguez and Torres (1997), Another model closely related to the Boolean model is the shot-noise process; see Section 5.6.

Bombing problems form a very natural application context for the Boolean model. Robbins (1944, 1945) and Neyman and Scott (1972) studied random closed sets with such applications in view. Here the germs are the points of impact of the bombs, and the grains are the areas covered by the splinters from the corresponding bomb explosions. Robbins' account was confined to theoretical issues; Neyman and Scott briefly discussed details of one particular application concerning the clearance of landmines from the landing beaches in Normandy. Generally in such problems the Boolean model seems plausible for representing the total area covered by splinters, while the shot-noise model is more suitable for considerations of intensity of covering. Hall (1988, Section 1.6) gives a short review of further modelling of the effect of salvoes of weapons fired at a target. Typically, the Boolean model is inhomogeneous in such applications, that is, the intensity of the Poisson process of germs is location dependent, given by an intensity function $\lambda(x)$, which should have its maximum in the target's centre; see Figure 3.4 on p. 70.

In the one-dimensional case, when the grains are intervals, the Boolean model is sometimes described as an $M/G/\infty$ queue. This is the D. G. Kendall notation for a queueing model in which the customer arrivals ('germs') are according to a Poisson process, the service times (intervals of those lengths are the 'grains') are independent and identically distributed, and there are infinitely many servers. In passing, note that Baccelli and Błaszczyszyn (2009a) use Kendall's notation also for shot-noise processes and related models.

The Boolean model has a three-fold value in the description of given samples and situations in nature. First, a relatively parsimonious description is available for the random set in question, by means of the intensity λ of the germ process and the various characteristics of the grains. Second, it is at least possible that the model assumptions may be suggestive of the process of formation of the structure, though it will be emphasised later that this is by no means invariably the case; see Section 3.6. Third, the formulae to be derived for the Boolean model may be used for estimation of quantities not available for direct measurements.

3.1.3 Stationarity and isotropy

From the stationarity of the Poisson process Φ of germs and the identical distribution of the grains, it follows that the Boolean model as defined above is stationary. That is to say, its distribution is translation-invariant; see Section 6.1. As a consequence of more general

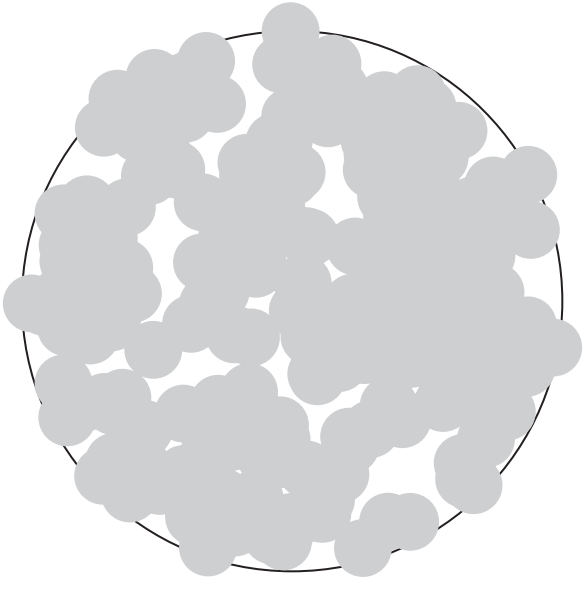


Figure 3.4 A simulated sample of an inhomogeneous Boolean model with discoidal grains and intensity function having a constant value within a disc and vanishing outside.

considerations Nguyen and Zessin (1979a) showed that the Boolean model is *ergodic* and even *mixing*; see Section 6.1 for a discussion of these concepts and Schneider and Weil (2008) for more details. A particular consequence of ergodicity is that

$$\lim_{r\to\infty} \frac{v_d\big(\Xi \cap B(o,r)\big)}{v_d\big(B(o,r)\big)} = p, \tag{3.4}$$

where $p = \mathbf{P}(o \in \Xi)$ is the volume fraction of the Boolean model, that is, that fraction of the $\mathbb{R}^d$ that is covered by Ξ. Note that (3.4) is not the definition of p — this is given by Formula (3.13). Similar relations hold for many other functionals replacing v_d in the numerator of (3.4), for example the intrinsic volumes.

If the distribution of the grains is isotropic (distribution invariant under rotations about the origin o), then the Boolean model is also isotropic in distribution. However, a Boolean model can be isotropic even if the grains are not isotropic; see Molchanov and Stoyan (1994). For example, consider the spatial case, with Ξ_0 a ball with south pole at the origin; see also Remark (3) on p. 73. Isotropy and stationarity together imply that the Boolean model has a distribution invariant under rigid motions.

Some applications consider *inhomogeneous* Boolean models. An important particular case is where the grains are i.i.d. as in the stationary case but the Poisson process of germs is inhomogeneous with intensity function $\lambda(x)$. This model is used in the military applications mentioned above; it is also related to Cressie's tumor model (see Cressie and Hulting, 1992). Figure 3.4 shows a simulated set of this kind. In the case of the figure

$$\lambda(x) = \lambda \mathbf{1}_A(x)$$

with $A = B(o, R)$ and $\Xi_0 = B(o, r)$. Another application is discussed in Hahn *et al.* (1999), where $\lambda(x)$ follows a vertical gradient. The theory of inhomogeneous Boolean models is presented in Schneider and Weil (2008, Chapter 11, especially Theorem 11.1.3).

3.1.4 Simulation

The simulation of Boolean models is not difficult for readers who know how to simulate Poisson processes, homogeneous as well as inhomogeneous. (See Lantuéjoul, 2002, for a very brief sketch.) The second step, the simulation of grains, is usually not very difficult. Of course, for particular grains, for example Poisson polyhedra, particular simulation methods are necessary. There is only one nontrivial problem, that of *edge-effects*.

If the grains are bounded, that is, there is a radius R such that $\Xi_0 \subset B(o, R)$ with probability one, then it suffices to simulate the Boolean model in an enlarged window: instead of the original window W the larger window $W_{\oplus R}$ is used to obtain a sample which is correct also at the boundary of W.

For the case of grains for which such a bound does not exist, methods of 'exact simulation' are useful. In principle the ideas in Brix and Kendall (2002) for the simulation of Poisson cluster processes can be applied also for Boolean models. Another, simpler, possibility is to generalise the method described above, using enlargement on window based on the enclosing radius R: the grains are binned by size and for each bin the R-method is used separately and the patterns obtained are superposed.

Exact simulation can be also used for conditional simulation of Boolean models: simulation of Boolean models such that the given locations $y_1, \ldots, y_n$ are covered by grains; see Thönnes (2001), Cai and Kendall (2002) and Lantuéjoul (2002).

The simulation can be organised so that the data generated are either vector data, that is, coordinates of germs and parameters of grains, as needed when the aim is to determine model parameters numerically, or pixel data, which are needed in the context of statistics of pixel data. Then one starts with an empty image, where all pixels are set to 0. In turn for all pixels it is checked whether these are overlapped by at least one grain. If so, the pixel value is set to 1.

3.1.5 The capacity functional

The distribution of a Boolean model is uniquely determined by its *capacity functional* or hitting distribution T_Ξ:

$$\begin{aligned} T_\Xi(K) &= \mathbf{P}(\Xi \cap K \text{ is not empty}) \\ &= \mathbf{P}(\Xi \cap K \neq \emptyset) \qquad \text{for all compact sets } K. \end{aligned} \tag{3.5}$$

Note that in the same form as (3.5), the capacity functional is also defined for any random closed set Ξ. The theory of random closed sets asserts that the functional uniquely determines the distribution of a random closed set; see Section 6.1.

In the case of the Boolean model, $T_\Xi(K)$ can be expressed in a relatively simple form, in pleasant contrast to the complexity of expressions for many other summary characteristics connected with this model. By application of the Poisson assumption, $T_\Xi(K)$ can be shown to take the form

$$T_\Xi(K) = 1 - \exp\bigl(-\lambda \mathbf{E}\bigl(\nu_d(\check{\Xi}_0 \oplus K)\bigr)\bigr) \tag{3.6}$$

or

$$T_\Xi(K) = 1 - \exp\bigl(-\lambda \mathbf{E}\bigl(v_d(\Xi_0 \oplus \check{K})\bigr)\bigr). \tag{3.7}$$

The proof of Formula (3.7) is of methodological interest.

Proof. Let K be an arbitrary fixed compact set. From the original germ process Φ a *thinned* process Φ_K can be produced. The thinning takes place by deleting x_n from Φ if $(\Xi_n + x_n) \cap K$ is empty. Thus

$$\Phi_K = \{x_n \in \Phi : (\Xi_n + x_n) \cap K \neq \emptyset\}.$$

Whether or not a germ x_n is deleted by this thinning procedure is independent of thinning happening to other germs. This follows from the independence properties of the Boolean model: deletion depends only on the location of the germ and on the corresponding grain. This independence implies that Φ_K is also a Poisson process (see Chapter 2), but now inhomogeneous with intensity function $\lambda_K(x)$ and only finitely many points. Here

$$\lambda_K(x) = \lambda p(x) \tag{3.8}$$

and $p(x) = \mathbf{P}\bigl((\Xi_0 + x) \cap K \neq \emptyset\bigr)$. Note that $p(x)$ is the probability that a germ placed at x is not deleted (the so-called retention probability). See Section 5.1 for a discussion of thinnings of point processes.

The total number of points of Φ_K has a Poisson distribution with mean μ_K,

$$\mu_K = \lambda \int_{\mathbb{R}^d} p(x)\, \mathrm{d}x. \tag{3.9}$$

Thus the probability that Φ_K is empty is

$$\mathbf{P}(\Phi_K \text{ has no points}) = \exp(-\mu_K).$$

Consequently the value $T_\Xi(K)$ is obtained as

$$T_\Xi(K) = 1 - \exp(-\mu_K). \tag{3.10}$$

So it suffices to evaluate μ_K.

The probability $p(x)$ is given by

$$p(x) = \mathbf{P}(x \in \check{\Xi}_0 \oplus K)$$

since $(\Xi_0 + x)$ meets K precisely when $x \in \check{\Xi}_0 \oplus K$. Thus

$$\begin{aligned}
\mu_K &= \lambda \int_{\mathbb{R}^d} \mathbf{P}(x \in \check{\Xi}_0 \oplus K)\mathrm{d}x \\
&= \lambda \int_{\mathbb{R}^d} \mathbf{E}\bigl(\mathbf{1}_{\check{\Xi}_o \oplus K}(x)\bigr)\mathrm{d}x \\
&= \lambda \mathbf{E}\bigl(v_d(\check{\Xi}_0 \oplus K)\bigr),
\end{aligned}$$

where in the last step the expectation is interchanged with the integral. This last expectation is finite, by the imposition of the technical condition required in the definition of the Boolean model.

Thus Formula (3.6) is established:

$$\begin{aligned} T_\Xi(K) &= 1 - \exp(-\mu_K) \\ &= 1 - \exp\bigl(-\lambda \mathbf{E}\bigl(\nu_d(\check{\Xi}_0 \oplus K)\bigr)\bigr), \end{aligned}$$

and Formula (3.7) follows by noting that

$$(\check{\Xi}_0 \oplus K) = -(\Xi_0 \oplus \check{K}). \qquad \square$$

Remarks (1) It is useful to introduce the notation

$$\psi(K) = -\ln\bigl(T_\Xi(K)\bigr). \tag{3.11}$$

Then

$$\psi(K) = \lambda \mathbf{E}\bigl(\nu_d(\Xi_0 \oplus \check{K})\bigr). \tag{3.12}$$

(2) The proof suggests that similar formulae hold for inhomogeneous Boolean models, such as the one that produces Figure 3.4.

(3) From Formula (3.7), one can see that there is an arbitrary element in the choice of location for each grain Ξ_n. Suppose that $z_1, z_2, \ldots,$ are points produced by a construction depending on the corresponding grain, so that z_n depends on Ξ_n alone and not on any other grains nor on any other germs. Then a new Boolean model Ξ^* can be constructed using the original Φ for the germs and $\Xi_n^* = \Xi_n + z_n$ for the grains. For example, take $\Xi_n = B(o, R_n)$ where R_n is a random radius, and then pick

$$z_n = (0, 0, \ldots, 0, R_n).$$

In any case, by the translation-invariance of Lebesgue measure

$$\begin{aligned} \psi^*(K) &= \lambda \mathbf{E}\bigl(\nu_d(\Xi_0^* \oplus \check{K})\bigr) \\ &= \lambda \mathbf{E}\bigl(\nu_d((\Xi_0 \oplus \check{K}) + z_0)\bigr) \\ &= \lambda \mathbf{E}\bigl(\nu_d(\Xi_0 \oplus \check{K})\bigr) \\ &= \psi(K). \end{aligned}$$

Hence, the two Boolean models Ξ and Ξ^* have the same capacity functional and so must have the same distribution.

3.1.6 Basic characteristics

Formulae (3.6) and (3.7) and the Poisson process property related to μ_K provide means of calculating some important characteristics of the Boolean model.

The volume fraction

The volume fraction p is the mean fraction of volume occupied by Ξ in a region B of unit volume,

$$p = \mathbf{E}\big(\nu_d(\Xi \cap B)\big) \qquad \text{for } \nu_d(B) = 1. \tag{3.13}$$

This value does not depend on the choice of the region B, by virtue of stationarity of Ξ. Note that p is defined without a limit operation, Formula (3.4) is a consequence of (3.13) and ergodicity. A further consequence of stationarity is that

$$p = \mathbf{P}(o \in \Xi), \tag{3.14}$$

a result that holds for general stationary random sets; it is proved in Section 6.3.1.

The formula for p is obtained by putting $K = \{o\}$ in Formula (3.7). Then

$$\mathbf{P}(o \in \Xi) = 1 - \exp\big(-\lambda \mathbf{E}\big(\nu_d(\Xi_0)\big)\big), \tag{3.15}$$

where Ξ_0 is as before the typical grain of Ξ. Hence

$$p = 1 - \exp(-\lambda \overline{V}), \tag{3.16}$$

in which $\overline{V} = \mathbf{E}\big(\nu_d(\Xi_0)\big)$ denotes the mean volume of the typical grain. Of course, if the grains of the Boolean model all have zero volume (for example, if they are points or line segments in the plane, or points, segments or flat objects in space), then the volume fraction will be zero.

Note that even if the volume fraction is unity it need not follow that $\Xi = \mathbb{R}^d$; see Hall (1988, Theorem 3.3) and Chiu (1995c). If $\Xi \neq \mathbb{R}^d$, then there can be a relation to fractals.

The covariance or two-point probability function

If a random set has volume fraction $p > 0$, then important second-order properties are summarised in its covariance or *two-point probability function*, as discussed in Chapter 6.

The (noncentred) *covariance* is defined as

$$C(\mathbf{r}) = \mathbf{P}(o \in \Xi \text{ and } \mathbf{r} \in \Xi) = \mathbf{P}(\{o, \mathbf{r}\} \subset \Xi) \qquad \text{for } \mathbf{r} \in \mathbb{R}^d. \tag{3.17}$$

A simple manipulation reveals that $C(\mathbf{r})$ is the volume fraction of $\Xi \cap (\Xi - \mathbf{r})$. Though this intersection is not in general a Boolean model, this interpretation is important for statistical determination of $C(\mathbf{r})$.

If Ξ is a Boolean model with typical grain Ξ_0, then

$$C(\mathbf{r}) = 2p - 1 + (1-p)^2 \exp\big(\lambda \mathbf{E}\big(\gamma_{\Xi_0}(\mathbf{r})\big)\big), \tag{3.18}$$

where $\gamma_{\Xi_0}(\mathbf{r}) = \nu_d\big(\Xi_0 \cap (\Xi_0 - \mathbf{r})\big)$ is the set covariance as in Section 1.7.2.

Proof. Formula (3.18) can be proved by considering the following evaluation of $C(\mathbf{r})$:

$$\begin{aligned} C(\mathbf{r}) &= \mathbf{P}\big(o \in \Xi \cap (\Xi - \mathbf{r})\big) \\ &= 1 - \big(\mathbf{P}(o \notin \Xi) + \mathbf{P}(o \notin \Xi - \mathbf{r}) - \mathbf{P}(o \notin \Xi \cup (\Xi - \mathbf{r}))\big) \\ &= 2p - 1 + \mathbf{P}\big(o \notin \Xi \cup (\Xi - \mathbf{r})\big). \end{aligned}$$

An application of (3.7) shows that

$$\mathbf{P}(\Xi \cap \{o, \mathbf{r}\} \neq \emptyset) = 1 - \exp\bigl(-\psi(\{o, \mathbf{r}\})\bigr), \tag{3.19}$$

where ψ is defined by (3.11). On evaluation of the Minkowski sum,

$$\begin{aligned}\psi(\{o, \mathbf{r}\}) &= \lambda \mathbf{E}\bigl(v_d\bigl(\check{\Xi}_0 \cup (\check{\Xi}_0 + \mathbf{r})\bigr)\bigr) \\ &= \lambda\bigl(\mathbf{E}\bigl(v_d(\check{\Xi}_0)\bigr) + \mathbf{E}\bigl(v_d(\check{\Xi}_0 + \mathbf{r})\bigr) - \mathbf{E}\bigl(v_d(\check{\Xi}_0 \cap (\check{\Xi}_0 + \mathbf{r}))\bigr)\bigr) \\ &= \lambda\bigl(2\mathbf{E}\bigl(v_d(\Xi_0)\bigr) - \mathbf{E}\bigl(v_d(\Xi_0 \cap (\Xi_0 - \mathbf{r}))\bigr)\bigr),\end{aligned}$$

and the proof of (3.18) is concluded by noting that

$$1 - p = \exp\bigl(-\lambda \mathbf{E}(v_d(\Xi_0))\bigr). \qquad \square$$

As discussed in Section 6.3.2, the covariance of a random set is closely related to the concepts of covariance function and correlation function for random fields. Note, however, that the covariance defined above differs from the conventional covariance function in that the covariance is not centred about the mean.

If the random set Ξ is also isotropic, then $C(\mathbf{r})$ depends only on the length $r = \|\mathbf{r}\|$ of the vector $\mathbf{r}$. In such a case a convenient abuse of notation is to consider the covariance as a function of length r alone and to use the symbol $C(r)$.

If the typical grain Ξ_0 is isotropic, then $\gamma_{\Xi_0}(r\boldsymbol{u})$ can be replaced by the isotropised set covariance $\overline{\gamma}_{\Xi_o}(r)$ introduced in Section 1.7.2. If the mean $\mathbf{E}\bigl(\overline{\gamma}_{\Xi_0}(r)\bigr)$ is denoted as $\overline{\overline{\gamma}}_{\Xi_0}(r)$, then Formula (3.18) takes the form

$$C(r) = 2p - 1 + (1 - p)^2 \exp\bigl(\lambda \overline{\overline{\gamma}}_{\Xi_0}(r)\bigr). \tag{3.20}$$

For some random convex bodies Ξ_0 the mean isotropised set covariance $\overline{\overline{\gamma}}_{\Xi_0}(r)$ is known. For a three-dimensional ball of random radius, with radius probability density function $f_R(r)$, it is given by

$$\overline{\overline{\gamma}}_{\Xi_0}(r) = \int_{r/2}^{\infty} \frac{4}{3}\pi x^3 \left(1 - \frac{3r}{4x} + \frac{r^3}{16x^3}\right) f_R(x)\,\mathrm{d}x \qquad \text{for } r \geq 0, \tag{3.21}$$

and for a disc in the plane, it is

$$\overline{\overline{\gamma}}_{\Xi_0}(r) = \int_{r/2}^{\infty} \left(2x^2 \cos^{-1}\frac{r}{2x} - \frac{r}{2}\sqrt{4x^2 - r^2}\right) f_R(x)\,\mathrm{d}x \qquad \text{for } r \geq 0. \tag{3.22}$$

In the literature further formulae can be found for the typical cells of the Voronoi tessellation (see Gilbert, 1962; Brumberger and Goodisman, 1983), the Poisson polyhedron (the typical cell of the Poisson plane tessellation, see Section 9.5.2) and the typical cell of the dead leaves model (see Gille, 2002).

Typically there exists an r_0 that $\overline{\overline{\gamma}}_{\Xi_0}(r) = 0$ for $r > r_0$. If it is the case, then $C(r) = p^2$ for $r > r_0$, and the Boolean model is said to have the finite range of correlation r_0.

Sometimes only the correlation function $\kappa(r)$, given by

$$\kappa(r) = \frac{C(r) - p^2}{p(1 - p)} \qquad \text{for } r \geq 0, \tag{3.23}$$

is known, resulting, for example, from small-angle scattering analysis. If so, then it may be of value to use the following formula, which yields p:

$$p = \lim_{r \to 0} \frac{\kappa''(r)}{\left(\kappa'(r)\right)^2}. \tag{3.24}$$

A useful asymptotic form relating $\kappa(r)$ to $\overline{\overline{\gamma}}_{\Xi_0}$ is given by Sonntag *et al.* (1981). As the intensity λ of the process Φ tends to zero, the grains remaining fixed, so p tends to zero and

$$\lim_{\lambda \to 0} \kappa(r) = \frac{\overline{\overline{\gamma}}_{\Xi_0}(r)}{\overline{\overline{\gamma}}_{\Xi_0}(0)}.$$

Thus for a 'dilute' Boolean model the correlation function $\kappa(r)$ is asymptotically the same as the normalised isotropised set covariance of Ξ_0.

Complicated formulae exist for the n-point probability functions for the case of Boolean models with identical balls, see Torquato (2002, pp. 122–4).

Coverage characteristics

The Boolean model is a union of grains and hence the following characteristics may be of interest:

- the mean number $\overline{c}$ of grains that contain a fixed point (it suffices to consider the origin o), and
- the volume fraction p_k of the random set of all points that are contained in more than k grains.

The proof of Formula (3.7) has shown that the random number of grains hitting a fixed compact set K follows a Poisson distribution with mean μ_K given by Formula (3.9). For the case $K = \{o\}$ this simplifies to

$$\mu_0 = \lambda \mathbf{E} \nu_d(\Xi_0), \tag{3.25}$$

which yields

$$\overline{c} = \mu_0. \tag{3.26}$$

Similar to Formula (3.14), due to stationarity the volume fraction p_k is equal to the probability that the origin o has been covered by more than k grains, that is,

$$p_k = 1 - \sum_{i=0}^{k} \frac{\overline{c}^{\,i}}{i!} \mathrm{e}^{-\overline{c}} \qquad \text{for } k = 1, 2, \ldots. \tag{3.27}$$

3.1.7 Contact distribution functions

In its full generality as a function of compact sets K, the capacity functional $T_\Xi(K)$ is not suitable for practical work. Matheron had the idea to use families of standard sets K, which led to the notion of contact distribution functions.

Let B be a convex body of $\mathbb{R}^d$ containing the origin o, which is called in this context *test set*, *gauge set* or *structuring element*. Then

$$T_\Xi(rB) = \mathbf{P}(\Xi \cap rB \neq \emptyset)$$

is a function of the scaling factor r, which could be graphically presented and yield structural-distributional information on Ξ. However, more elegant is the use of the *contact distribution function* $H_B(r)$ given by

$$H_B(r) = 1 - \frac{\mathbf{P}(\Xi \cap rB = \emptyset)}{1-p} \qquad \text{for } r \geq 0, \tag{3.28}$$

where p is the volume fraction of Ξ. This function has all properties of a distribution function. Clearly, it is

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

which shows the close relationship between the contact distribution function and the capacity functional.

The contact distribution function with a suitably chosen structuring element is valuable in particular in situations where identification of particles or of single vacant regions is impossible or inappropriate. Delfiner (1972) gave an excellent discussion of this using the closely related concept of size distribution functions.

In passing let it be noted that the relationship between contact distribution and size distribution, while close, is not explicit save in one special case. Matheron (1975, p. 53) discussed the essential ideas here and established an explicit connection for the special case when the structuring element is a line segment. His discussion concerns granulometries, from which the notion of size distribution is derived.

Since B contains the origin, $\Xi \cap rB$ being empty implies that Ξ does not contain the origin. So the value of $H_B(r)$ is the conditional probability that the scaled set rB hits (or makes contact with) Ξ, conditional on Ξ not containing o.

As another interpretation, $H_B(r)$ is the conditional distribution function of R:

$$H_B(r) = \mathbf{P}(R \leq r | R > 0), \tag{3.29}$$

where

$$R = \inf\{s : \Xi \cap sB \neq \emptyset\},$$

that is, R is the distance from o to Ξ measured with respect to B.

The shape of the structuring element B plays an important rôle. Two particularly important cases are the extremes where on the one hand, B is a line segment of unit length leading to the *linear contact distribution function* $H_l(r)$, and on the other hand B is the unit ball $B(o, 1)$ leading to the *spherical contact distribution function* $H_s(r)$. Two other cases appear on p. 83. Of course in anisotropic cases the orientation of the line segment is crucial for the values of $H_l(r)$. Section 6.3.5 discusses briefly the use of the linear contact distribution function in orientation analysis.

Note that $H_B(r)$ is produced in the first instance from the *complement* of Ξ rather than Ξ. While $H_l(r)$ is related to the distribution function of chord lengths *outside* of Ξ by (6.66), the chord lengths *in* Ξ are difficult to analyse, see p. 86.

The spherical contact distribution function $H_s(r)$ may be interpreted as the distribution function of the distance from a point chosen randomly outside Ξ, measured to the nearest point of Ξ. Hence, $H_s(r)$ is referred to as 'the law of first contact'.

In general a formula for the contact distribution function follows directly from (3.7) and (3.15):

$$H_B(r) = 1 - \exp\bigl(-\lambda\bigl(\mathbf{E}(\nu_d(\Xi_0 \oplus rB)) - \mathbf{E}(\nu_d(\Xi_0))\bigr)\bigr) \qquad \text{for } r \geq 0. \tag{3.30}$$

3.2 The Boolean model with convex grains

In practice a frequent assumption is that the Boolean model has isotropic and convex grains. This has the advantage of considerable simplification of various formulae. Moreover, practical experience shows that the resulting structures are still sufficiently flexible to be useful in many statistical problems; see also pp. 106–7.

3.2.1 The simplified formula for the capacity functional

If Ξ_0 is isotropic and convex, then $\psi(K)$, which gives the capacity functional by (3.11), can be calculated for convex K by means of the generalised Steiner formula (6.29) on p. 216:

$$\psi(K) = \lambda\mathbf{E}\bigl(\nu_d(\Xi_0 \oplus \check{K})\bigr) = \frac{\lambda}{b_d}\sum_{k=0}^{d}\frac{b_k b_{d-k}}{\binom{d}{k}}\overline{V}_k V_{d-k}(K), \tag{3.31}$$

where $\overline{V}_k$ denotes the mean of the k^{th} intrinsic volume of Ξ_0, that is,

$$\overline{V}_k = \mathbf{E}\bigl(V_k(\Xi_0)\bigr) \qquad \text{for } k = 0, 1, \ldots, d,$$

and b_k the volume of the k-dimensional unit ball.

Thus an important simplification is obtained:

Theorem 3.1. *For convex bodies K, the value of the capacity functional $T_\Xi(K)$ of a Boolean model Ξ with isotropic convex grains depends only on λ and the means of the intrinsic volumes of the grains.*

As an example, let Ξ_0 be a ball $B(o, R)$ of random radius R with $\mathbf{E}(R^d) < \infty$. Then, for convex bodies K, by the Steiner formula (1.23) in the planar case,

$$\psi(K) = \lambda\left(A(K) + L(K)\mathbf{E}(R) + \pi\mathbf{E}(R^2)\right), \tag{3.32}$$

and by Formula (1.24) in the spatial case,

$$\psi(K) = \lambda\Bigl(V(K) + S(K)\mathbf{E}(R) + 2\pi\bar{b}(K)\mathbf{E}(R^2) + \frac{4}{3}\pi\mathbf{E}(R^3)\Bigr). \tag{3.33}$$

By means of the generalised Steiner formula (6.29) both formulae can be extended to the case of isotropic grains.

Note that the distribution of Ξ is of course not determined by the values of $\psi(K)$ for convex bodies K alone, as the distribution of radii is not determined by the first two or three moments.

3.2.2 Intensities or densities of intrinsic volumes

From a Boolean model various random measures can be derived. For example a three-dimensional Boolean model Ξ yields the *surface measure* S_Ξ, defined by

$$S_\Xi(B) = h_2(B \cap \partial\Xi) \qquad \text{for all Borel sets } B, \tag{3.34}$$

where the right hand side denotes the surface area of the boundary $\partial\Xi$ of the Boolean model Ξ in the set B; h_2 is the two-dimensional Hausdorff measure. The random measure S_Ξ inherits stationarity from the Boolean model Ξ. Therefore the mean measure $\mathbf{E}\big(S_\Xi(B)\big)$ is invariant under translations and so is a multiple of Lebesgue measure:

$$\mathbf{E}\big(S_\Xi(B)\big) = S_V \nu_3(B). \tag{3.35}$$

The constant S_V is the *intensity* of the random measure S_Ξ. It is called the *specific surface (area)* because S_V can be interpreted as the *mean surface area of* Ξ *per unit volume*.

An analogous quantity for a planar Boolean model is L_A, the *mean boundary length per unit area*. Similar intensities can be defined also with respect to the other intrinsic volumes; see Section 7.3.4.

Some readers may wonder whether this approach is over-complicated. However it generalises well, for example to the nonstationary case, where Radon–Nikodym densities are used. The alternative density approach provides an easier route:

$$S_V = \lim_{r\to\infty} \frac{\mathbf{E}\big(S_\Xi(B(o,r))\big)}{\nu_3\big(B(o,r)\big)} = \lim_{r\to\infty} \frac{2\mathbf{E}\big(V_2(\Xi \cap B(o,r))\big)}{\nu_3\big(B(o,r)\big)}. \tag{3.36}$$

For the k^{th} intrinsic volume in $\mathbb{R}^d$ the *density* v_k is

$$v_k = \lim_{r\to\infty} \frac{\mathbf{E}\big(V_k\big(\Xi \cap B(o,r)\big)\big)}{\nu_d\big(B(o,r)\big)} \qquad \text{for } k = 0, 1, \ldots, d. \tag{3.37}$$

Instead of $B(o,r)$, the set rW can be used, where W is a convex body with positive volume. Since the Boolean model is ergodic, the '$\mathbf{E}$' can be omitted in the equations above (see Theorem 6.2 on p. 210). More specific notations for the v_k are, in the planar case

$$A_A = v_2, \tag{3.38}$$
$$L_A = 2v_1, \tag{3.39}$$
$$N_A = v_0, \tag{3.40}$$

and in the spatial case

$$V_V = v_3, \tag{3.41}$$
$$S_V = 2v_2 \tag{3.42}$$
$$M_V = \pi v_1, \tag{3.43}$$
$$N_V = v_0. \tag{3.44}$$

The quantities N_A and N_V are called *specific connectivity numbers* and M_V is the *specific integral of mean curvature*. The intensities N_A, N_V and M_V belong to signed measures and can be negative.

There are 'positive' counterparts N_A^+, N_V^+, called *specific convexity numbers*, and M_V^+. The quantity N_A^+ is the intensity of lower convex tangent points of Ξ in $\mathbb{R}^2$. Each grain Ξ_n has a lower convex tangent point l_n, which is the point with minimum second coordinate in Ξ_n. (If this minimum is attained by more than one point, then the point having the minimum first coordinate is chosen, i.e. the lexicographic minimum point.) The points $x_n + l_n$ that are not covered by other grains $\Xi_m + x_m$ ($m \neq n$) are the lower convex tangent points of Ξ. The analogous intensity in the spatial case is N_V^+.

The densities or intensities satisfy the following equations, often called *Miles' formulae*:

$$V_V = p = 1 - \exp(-\lambda \overline{V}), \tag{3.45}$$

$$S_V = \lambda(1-p)\overline{S} = \lambda \exp(-\lambda \overline{V})\overline{S}, \tag{3.46}$$

$$M_V = \lambda(1-p)\left(\overline{M} - \frac{\pi^2 \lambda \overline{S}^2}{32}\right), \tag{3.47}$$

$$N_V = \lambda(1-p)\left(1 - \frac{\lambda \overline{M}\,\overline{S}}{4\pi} + \frac{\pi \lambda^2 \overline{S}^3}{384}\right), \tag{3.48}$$

$$N_V^+ = \lambda(1-p) = \lambda \exp(-\lambda \overline{V}), \tag{3.49}$$

and

$$A_A = p = 1 - \exp(-\lambda \overline{A}), \tag{3.50}$$

$$L_A = \lambda(1-p)\overline{L} = \lambda \exp(-\lambda \overline{A})\overline{L}, \tag{3.51}$$

$$N_A = \lambda(1-p)\left(1 - \frac{\lambda \overline{L}^2}{4\pi}\right), \tag{3.52}$$

$$N_A^+ = \lambda(1-p) = \lambda \exp(-\lambda \overline{A}), \tag{3.53}$$

see Miles (1976), Weil (2007, p. 211) and Schneider and Weil (2008, p. 389).

Remarks (1) In the case $p = 0$ (for example in the case of a segment process) Formulae (3.51) and (3.46) should be replaced by

$$2L_A = \lambda \overline{L} \qquad \text{and} \qquad 2S_V = \lambda \overline{S}.$$

This change is necessary since in this case overlappings do not reduce the boundary lengths or surface areas of grains, and for a convex body K of volume zero, the perimeter $L(K)$ and the surface area $S(K)$ should be twice the length and the area of K, respectively; see p. 12.

(2) Many of the formulae from (3.45) to (3.53) above are true also in the anisotropic case. Isotropy is needed only for the formulae which contain a 'π'.

Sketch of the proof of (3.46) and (3.53). Formula (3.46) is a direct consequence of (3.103), which implies that for every grain the ratio of 'mean surface area not covered by other grains' to 'total mean surface area of grain' is $1 - p$. Summing this relation over all grains yields (3.46).

In the case of (3.53) the starting point is the definition of N_A^+ above. Without loss of generality it can be assumed that $l_n = o$ for all n; the grains $\Xi_n^* = \Xi_n - l_n$ have this property, and by Remark (3) on p. 73 the Boolean model Ξ^* formed by the system of the grains $\{\Xi_n^*\}$ and the original germs $\{x_1, x_2, \ldots\}$ is still a Boolean model, with the same distribution as Ξ.

The intensity of the point process of lower convex tangent points of Ξ^* is $p^*\lambda$, where p^* is the probability that the typical point of the germ process is not covered by grains belonging to other germs. By the Slivnyak–Mecke theorem the union of the other grains has the distribution of the original Boolean model and p^* is the probability that the origin o is not covered. Thus $p^* = 1 - p$. □

By the way, the lower convex tangent points of the Boolean model form a stationary point process, which can be interpreted as resulting from a (dependent) thinning of the original germ process. Because tangents parallel to the x-axis play the main rôle in the process construction, the thinned process is in general not isotropic, not even in the case of spherical grains. Its second-order characteristics can be computed; see Molchanov and Stoyan (1994).

The proof of the formulae for v_0 in the cases $d = 2$ and $d = 3$ (which up to a factor are the specific connectivity numbers N_A and N_V), and v_1 in the case $d = 3$ (which up to a factor is the specific integral of mean curvature M_V) are much more complicated.

For the particular case that the grains are random segments of constant length ℓ, in the planar case it holds

$$N_A = \lambda - N_3 \tag{3.54}$$

with

$$N_3 = \lambda^2 \ell^2 / \pi.$$

The term N_3 is linked to the topological concept of Betti numbers; see Robins (2002).

3.2.3 Contact distribution functions

By means of (3.31) the general Formula (3.30) for contact distribution functions can be simplified for a convex compact structuring element B. It becomes

$$H_B(r) = 1 - \exp\left(-\frac{\lambda}{b_d} \sum_{k=0}^{d-1} \frac{b_k b_{d-k}}{\binom{d}{k}} \overline{V}_k V_{d-k}(B) r^{d-k} \right) \qquad \text{for } r \geq 0. \tag{3.55}$$

Note that the last term, corresponding to $k = d$, of the original summation cancels with the $-\lambda \overline{V}_d$ or $-\lambda \mathbf{E}\big(v_d(\Xi_0)\big)$ term of the exponent in (3.30).

The important planar and spatial cases are as follows.

Planar case ($d = 2$):

$$H_B(r) = 1 - \exp\Big(-\lambda\Big(\frac{\overline{L}L(B)}{2\pi}r + A(B)r^2\Big)\Big)$$
$$= 1 - \exp\Big(-\frac{1}{1-A_A}\Big(\frac{L_A L(B)}{2\pi}r + N_A^+ A(B)r^2\Big)\Big) \qquad \text{for } r \geq 0. \tag{3.56}$$

Spatial case ($d = 3$):

$$H_B(r) = 1 - \exp\left(-\lambda\Big(\frac{\overline{S}M(B)}{4\pi}r + \frac{\overline{M}S(B)}{4\pi}r^2 + V(B)r^3\Big)\right)$$
$$= 1 - \exp\left(-\frac{1}{1-V_V}\Big(\frac{S_V M(B)}{4\pi}r + \frac{M_V^+ S(B)}{4\pi}r^2 + N_V^+ V(B)r^3\Big)\right) \qquad \text{for } r \geq 0. \tag{3.57}$$

In the formulae for $H_B(r)$ the physical dimensionality should be noted. For example, in (3.56) it seems to be natural to give $\overline{L}$ the right physical dimension of length. However, if r is given the dimension length, as it seems to be reasonable, then $L(B)$ and $A(B)$ should be dimension-free. The corresponding numerical values should be chosen such that $rL(B)$ and $r^2A(B)$ have the right values in comparison to $\overline{L}$ and $\overline{A}$, respectively.

Linear contact distribution function $H_l(r)$

This is given by

$$H_l(r) = 1 - \exp\left(-\lambda\frac{2b_{d-1}\overline{V}_{d-1}}{db_d}r\right) = 1 - \exp(-\lambda_1^{(d)}r) \qquad \text{for } r \geq 0. \tag{3.58}$$

Note that this is an exponential distribution of parameter $\lambda_1^{(d)}$. The parameter $\lambda_1^{(d)}$ is defined by its occurrence here, and also by Formula (3.68), which comes from (3.66).

Example 3.1. Consider the *range of vision in a forest* where the set of tree cross sections can be modelled (after Pólya, 1918) as a planar Boolean model realisation, with discoidal grains. If the mean radius of a grain (radius of a typical trunk) is $\mathbf{E}(R)$ then

$$H_l(r) = 1 - \exp\big(-2\lambda r\mathbf{E}(R)\big) \qquad \text{for } r \geq 0.$$

Taking for example the values $\mathbf{E}(R) = 0.20\,\text{m}$ and $\lambda = 0.01\,\text{m}^{-2}$, then

$$H_l(r) = 1 - \exp(-0.004r),$$

and hence $H_l(500) = 1 - 0.135$. That is to say that in a forest which can be described by such a Boolean model, with parameters as above, an observer standing at the typical point and looking in an arbitrary direction would have a probability of 13.5% of being able to see more than 500 m.

Further visibility properties are summarised in the *star*, or visible volume outside of Ξ, which is the set

$$\mathrm{star}_x(\Xi) = \{y \in \mathbb{R}^d : [x, y) \cap \Xi = \emptyset\}$$

defined for x not in Ξ; see also Serra (1982), Wieacker (1985), Yadin and Zacks (1985, 1988), Schneider (1987), Molchanov (1994) and Zacks (1994). The mean volume of the star in the planar case is

$$\mathbf{E}\big(\mathrm{star}_o(\Xi)\big) = \frac{2}{\pi\lambda_1^{(2)}}, \tag{3.59}$$

and in the spatial case

$$\mathbf{E}\big(\mathrm{star}_o(\Xi)\big) = \frac{8}{\pi\lambda_1^{(3)}}. \tag{3.60}$$

Spherical contact distribution function $H_s(r)$

This is given by

$$H_s(r) = 1 - \exp\left(-\lambda \sum_{k=1}^{d} b_k \overline{V}_{d-k} r^k\right) \qquad \text{for } r \geq 0. \tag{3.61}$$

Consider, for example, a Boolean model with spherical grains in three-dimensional space, that is, random balls whose centres are distributed by a Poisson process in $\mathbb{R}^3$. The spherical contact distribution function is the distribution function of the distance to the nearest ball from a point chosen at random but lying outside all balls. It is

$$H_s(r) = 1 - \exp\left(-\frac{4}{3}\lambda\pi r(3\mathbf{E}(R^2) + 3r\mathbf{E}(R) + r^2)\right). \tag{3.62}$$

Quadrat contact distribution function $H_q(r)$

In statistical applications with $d = 2$, this function is useful. Here B is a square of unit area and

$$H_q(r) = 1 - \exp\left(-\lambda r\Big(\frac{2\overline{U}}{\pi} + r\Big)\right) \qquad \text{for } r \geq 0. \tag{3.63}$$

Discoidal contact distribution function $H_d(r)$

Analogously, in statistical applications with $d = 3$ and planar sections, this function is useful. Here B is a disc of radius 1 in $\mathbb{R}^3$ and

$$H_d(r) = 1 - \exp\left(-\lambda\pi r\Big(\frac{\overline{S}}{4} + \overline{\overline{b}}r\Big)\right) \qquad \text{for } r \geq 0. \tag{3.64}$$

This function was introduced in Hug *et al.* (2006) and called there 'disc contact distribution'; instead of a circular disc, general convex flats were considered. That paper shows that the linear or polynomial form of $\ln\bigl(1 - H_B(r)\bigr)$ for the linear and discoidal contact distribution functions of stationary Boolean models is valid only for convex grains. However, note that there are non-Boolean random set models for which the logarithm above is also linear or polynomial.

3.2.4 Morphological functions

In 1991 Klaus Mecke suggested that one could transform the intrinsic volume densities to functions, making them powerful tools in random set statistics; see Mecke (1994, 2000) and Stoyan and Mecke (2005). In his papers he speaks about 'Minkowski functions' and denotes these as $W_k(r)$, following the notation for the Minkowski functionals. The exposition here uses instead the intrinsic volumes, and the functions are called *Mecke's morphological functions*. The following explains the idea for S_V; the same can be done for M_V and N_V. The case of V_V can be seen as covered by the spherical contact distribution function.

The function $S_V(r)$ is simply the specific surface of the parallel set $\Xi \oplus B(o, r)$. Since for a Boolean model Ξ the set $\Xi \oplus B(o, r)$ is the Boolean model with the same intensity λ and typical grain $\Xi_0 \oplus B(o, r)$, it is easy to give a formula for $S_V(r)$ by means of the Steiner formula and its counterparts for surface and average breadth:

$$S_V(r) = \lambda(\overline{S} + 4\pi\overline{\overline{b}}r + 4\pi r^2) \exp\left(-\lambda\left(\overline{V} + \overline{S}r + 2\pi\overline{\overline{b}}r^2 + \frac{4}{3}\pi r^3\right)\right) \qquad \text{for } r \geq 0. \quad (3.65)$$

It is derived by means of (1.47), (1.48) and (3.46).

Similarly, the formulae for $M_V(r)$ and $N_V(r)$ can be derived by means of (1.47), (1.48), (3.47) and (3.48). These functions can also be defined for negative values of r, as the specific surface of $\Xi \ominus B(o, r)$, see Stoyan and Mecke (2005).

3.2.5 Intersections with linear subspaces

Let L be an affine linear subspace of $\mathbb{R}^d$ (so not necessarily passing through the origin) of dimension l. For Ξ a Boolean model in $\mathbb{R}^d$ consider the intersection $\Xi_L = \Xi \cap L$. Then Ξ_L is again a Boolean model on L, when L is identified with $\mathbb{R}^l$. That Ξ_L is a Boolean model follows from considering its capacity functional. If the grains of Ξ are convex, then so of course are those of Ξ_L.

In the case of isotropy, when Ξ has convex grains, the induced Boolean model Ξ_L is also isotropic. Moreover, the intensity of its germ process is $\lambda_l^{(d)}$ with

$$\lambda_l^{(d)} = \lambda \frac{b_{d-l} b_l}{b_d} \frac{\overline{V}_{d-l}}{\binom{d}{d-l}}, \quad (3.66)$$

and the mean value $\overline{V}_k^{(l)}$ of the k^{th} intrinsic volume of the grain in $\mathbb{R}^l$ is

$$\overline{V}_k^{(l)} = \frac{b_{d-l+k}}{b_k b_{d-l}} \binom{d}{d-l} \frac{\overline{V}_{d-l+k}}{\overline{V}_{d-l}} \qquad \text{for } k \leq l. \quad (3.67)$$

These formulae were proved by Matheron (1975, p. 146), where there is also a discussion of the anisotropic case and the second moments of the volume of the induced grain; see also Schneider and Weil (2008). In the particular case of intersection of a planar Boolean model with a line ($l = 1$) the proof of (3.66) and (3.67) is as follows.

Proof. Suppose Ξ is a Boolean model in the plane which is isotropic and has convex grains. Let L be a line and let K be an arbitrary compact subset of L. Clearly $\mathbf{P}(\Xi \cap K = \emptyset) = \mathbf{P}(\Xi_L \cap K = \emptyset)$ thus $\psi(K) = \psi_L(K)$, which shows that indeed Ξ_L is a Boolean model, where $\psi_L(K) = -\ln\bigl(T_{\Xi_L}(K)\bigr)$.

Let K be a bounded closed interval in L of length $l(K)$. Let $\lambda_1^{(2)}$ be the intensity of the germ process of Ξ_L. Then Formula (3.12) yields

$$\psi_L(K) = \lambda_1^{(2)}\bigl(l(K) + \overline{\ell}\bigr),$$

where $\overline{\ell}$ is the mean length of the grains, which are segments, of Ξ_L. On the other hand

$$\psi(K) = \lambda \mathbf{E}\bigl(A(\Xi_0 \oplus \check{K})\bigr) = \lambda\left(\overline{A} + \frac{\overline{L}\, l(K)}{\pi}\right)$$

from the planar case ($d = 2$) of (3.31). Equating the coefficients of $l(K)$ gives

$$\lambda_1^{(2)} = \lambda\frac{\overline{L}}{\pi} = \lambda\frac{b_1}{b_2}\overline{V}_1 \qquad \text{as } \overline{L} = 2\overline{V}_1$$

and

$$\overline{\ell} = \lambda\frac{\overline{A}}{\lambda_1^{(2)}} = \frac{b_2}{b_1}\frac{\overline{V}_2}{\overline{V}_1} \qquad \text{as } \overline{A} = \overline{V}_2. \qquad \square$$

The case of planar sections of three-dimensional Boolean models is considered in Section 10.3.1, Example 10.1.

Distribution function of chord lengths for Boolean models

Let Ξ be a Boolean model of convex grains and assume that Ξ is intersected with a line L. The intersection $\Xi_L = \Xi \cap L$ yields two alternating sequences of chords or intervals on L; one sequence forms Ξ_L and the other the exterior $L \setminus \Xi_L$. Because of the independence assumptions for the Boolean model and convexity of the grains, these intervals are independent, both kinds of chords follow fixed distributions and the resulting structure is called in the mathematical literature an 'alternating renewal process'; see for example Cox and Isham (1980).

The determination of the two types of chord length distributions differs greatly. It is easy to show that the *exterior* chords in the complement of Ξ follow an exponential distribution. If furthermore the Boolean model Ξ is isotropic, then the parameter of this exponential distribution is $\lambda_1^{(d)}$, where

$$\lambda_1^{(d)} = \lambda\frac{2db_{d-1}}{b_d}\overline{V}_{d-1}, \tag{3.68}$$

which is the special case of (3.66) for $l = 1$. Then the chord length probability density function $f_{\mathrm{ex}}(l)$ is

$$f_{\mathrm{ex}}(l) = \lambda_1^{(d)} \exp(-\lambda_1^{(d)} l) \qquad \text{for } l \geq 0. \tag{3.69}$$

The mean chord length is

$$\overline{\ell}_{\mathrm{ex}} = 1/\lambda_1^{(d)}. \tag{3.70}$$

Clearly in the anisotropic case the parameter depends on the direction of the line.

The *interior* chord lengths for the grain phase have a complicated distribution depending on the distribution of the grains. Tools such as the Fourier transform have been used to obtain results, see Torquato and Lu (1993) and Gille (2011). The curves in Figure 3.5 for the probability density function $f_{\mathrm{in}}(l)$ in the case of identical balls of diameter 1 were obtained numerically by inverting the Fourier transform. Nevertheless, for the first two moments formulae are known. The mean chord length $\overline{\ell}_{\mathrm{in}}$ can be obtained from

$$p = \frac{\overline{\ell}_{\mathrm{in}}}{\overline{\ell}_{\mathrm{in}} + \overline{\ell}_{\mathrm{ex}}}, \tag{3.71}$$

and the second moment is

$$\overline{\ell^2}_{\mathrm{in}} = \frac{\overline{\ell}_{\mathrm{in}}^{\,2} \cdot l_c}{\overline{\ell}_{\mathrm{ex}} \cdot p}, \tag{3.72}$$

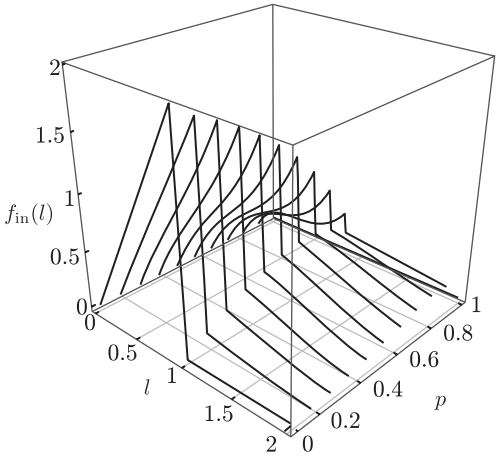


Figure 3.5 Probability density functions $f_{\mathrm{in}}(l)$ of chord lengths for a Boolean model with identical spherical grains of diameter 1 in dependence of volume fraction p. For small p isolated balls dominate and so $f_{\mathrm{in}}(l)$ looks like the chord length probability density function for a ball as in Section 1.7. For larger values of p clumping plays a rôle and the chords become longer. Courtesy of W. Gille.

with

$$l_c = 2\int_0^\infty \kappa(r)\mathrm{d}r \tag{3.73}$$

and $\kappa(r)$ is given in Equation (3.23), as shown by Gille (2011).

3.2.6 Formulae for some special Boolean models with isotropic convex grains

Two-dimensional models:

(1) Grains = *random segments* with uniform (or isotropic) orientation and of mean length $\overline{\ell}$. The linear contact distribution is an exponential distribution with parameter

$$\lambda_1^{(2)} = \frac{2\lambda\overline{\ell}}{\pi}. \tag{3.74}$$

The spherical contact distribution function is

$$H_s(r) = 1 - \exp\bigl(-\lambda r(2\ell + \pi r)\bigr) \qquad \text{for } r \geq 0. \tag{3.75}$$

(2) Grains = *random discs* with random radius R. The area fraction is given by

$$p = 1 - \exp\left(-\lambda\pi\mathbf{E}(R^2)\right). \tag{3.76}$$

Moreover

$$\lambda_1^{(2)} = 2\pi\lambda\mathbf{E}(R), \tag{3.77}$$

$$H_s(r) = 1 - \exp\bigl(-\lambda\pi r\,(2\mathbf{E}(R) + r)\bigr) \qquad \text{for } r \geq 0. \tag{3.78}$$

(3) Grains = *Poisson polygons* with parameter ϱ (see Section 9.5.1). The expected area of a grain is then

$$\overline{A} = \frac{4}{\pi\varrho^2}. \tag{3.79}$$

The other formulae are:

$$p = 1 - \exp\left(\frac{-4\lambda}{\pi\varrho^2}\right), \tag{3.80}$$

$$\lambda_1^{(2)} = \frac{4\lambda}{\pi\varrho}, \tag{3.81}$$

$$H_s(r) = 1 - \exp\left(-\lambda r\Bigl(\frac{4}{\varrho} + \pi r\Bigr)\right) \qquad \text{for } r \geq 0, \tag{3.82}$$

$$C(r) = 2p - 1 + (1-p)^2 \exp\left(\frac{4\lambda}{\pi\varrho^2}\exp(-\varrho r)\right) \qquad \text{for } r \geq 0. \tag{3.83}$$

Three-dimensional models:

(4) Grains = *random segments* with uniform orientations of mean length $\overline{\ell}$. The discoidal and spherical contact distribution functions are

$$H_d(r) = 1 - \exp(-\lambda\pi^2 r^2 \overline{\ell}) \qquad \text{for } r \geq 0, \tag{3.84}$$

$$H_s(r) = 1 - \exp\left(-\lambda\pi r^2\left(\overline{\ell} + \frac{4r}{3}\right)\right) \qquad \text{for } r \geq 0. \tag{3.85}$$

(5) Grains = *random discs* with uniform orientation and random radius R. Then

$$\lambda_1^{(3)} = \frac{\lambda\pi\mathbf{E}(R^2)}{2}, \tag{3.86}$$

$$H_d(r) = 1 - \exp\left(-\lambda\pi r\left(\frac{\pi\mathbf{E}(R^2)}{2} + \frac{\pi\mathbf{E}(R)r}{2}\right)\right) \qquad \text{for } r \geq 0, \tag{3.87}$$

$$H_s(r) = 1 - \exp\left(-\lambda\pi r\left(2\mathbf{E}(R^2) + \pi r\mathbf{E}(R) + \frac{4r^2}{3}\right)\right) \qquad \text{for } r \geq 0. \tag{3.88}$$

(6) Grains = *random balls* with random radius R. Then

$$p = 1 - \exp\left(-\frac{4}{3}\lambda\pi\mathbf{E}\left(R^3\right)\right), \tag{3.89}$$

$$\lambda_1^{(3)} = \lambda\pi\mathbf{E}(R^2), \tag{3.90}$$

$$H_d(r) = 1 - \exp\left(-\lambda\pi r\left(\pi\mathbf{E}(R^2) + 2\mathbf{E}(R)r\right)\right) \qquad \text{for } r \geq 0, \tag{3.91}$$

$$H_s(r) = 1 - \exp\left(-\frac{4}{3}\lambda\pi r\left(3\mathbf{E}(R^2) + 3r\mathbf{E}(R) + r^2\right)\right) \qquad \text{for } r \geq 0. \tag{3.92}$$

(7) Grains = *Poisson polyhedra* with parameter ϱ; see Section 9.5.2. (The first paragraph of Section 3.3.5 has an explanation of why the Poisson polyhedron grain can serve as a good model in a wide variety of applications.) The expected volume of a grain is then

$$\overline{V} = \frac{6}{\pi\varrho^3}.$$

The other formulae are:

$$p = 1 - \exp\left(\frac{-6\lambda}{\pi\varrho^3}\right), \tag{3.93}$$

$$\lambda_1^{(3)} = \frac{6\lambda}{\pi\varrho^2}, \tag{3.94}$$

$$H_d(r) = 1 - \exp\left(-\lambda\pi r\left(\frac{6}{\pi\varrho^2} + \frac{3}{2\varrho}r\right)\right) \qquad \text{for } r \geq 0, \tag{3.95}$$

$$H_s(r) = 1 - \exp\left(-\lambda r\left(\frac{24}{\pi\varrho^2} + \frac{3\pi}{\varrho}r + \frac{4}{3}\pi r^2\right)\right) \qquad \text{for } r \geq 0, \tag{3.96}$$

$$C(r) = 2p - 1 + (1-p)^2 \exp\left(\frac{6\lambda\exp(-\varrho r)}{\pi\varrho^3}\right) \qquad \text{for } r \geq 0. \tag{3.97}$$

The isotropised set covariance of the typical grain is

$$\overline{\gamma}_{\Xi_0}(r) = \frac{6\exp(-\varrho 2r)}{\pi\varrho^3} \qquad \text{for } r \geq 0. \tag{3.98}$$

The case of cylindrical grains is considered in Peyrega *et al.* (2009).

3.3 Coverage and connectivity

3.3.1 Coverage probabilities

Coverage probabilities are quantities of the form

$$\mathbf{P}(K \subset \Xi),$$

where K is any compact subset of $\mathbb{R}^d$. The volume fraction $p = \mathbf{P}(o \in \Xi)$ can be written in the form

$$p = \mathbf{P}(\{o\} \subset \Xi)$$

and is therefore a coverage probability with $K = \{o\}$. Also the covariance $C(r)$ can be interpreted as a coverage probability, that for the two-point set $\{o, \mathbf{r}\}$.

Clearly, it holds for general compact K

$$\mathbf{P}(K \subset \Xi) \leq \mathbf{P}(K \cap \Xi \neq \emptyset); \tag{3.99}$$

that is, typically the coverage probability is smaller than the hitting probability.

As might be expected, the calculation of coverage probabilities is difficult. Hall (1988, p. 180) discussed the coverage problem in detail. For the particular case $d = 2$, $K = [0, 1]^2$ and $\Xi_0 = B(o, r)$ (r deterministic) he showed that

$$1 - 3\min\{1,\ (1 + \lambda^2\pi r^2)\mathrm{e}^{-\lambda\pi r^2}\} < \mathbf{P}(K \subset \Xi) < 1 - 0.05\min\{1,\ (1 + \lambda^2\pi r^2)\mathrm{e}^{-\lambda\pi r^2}\} \tag{3.100}$$

if $\lambda \geq 1$ and $r \leq 0.5$. This is a rather rough bound for the coverage probability.

Janson (1986) found asymptotic relationships for coverage probabilities for large λ and small grains; see also Baccelli and Błaszczyszyn (2009a, p. 327).

The coverage of the whole space $\mathbb{R}^d$ is also of interest. As shown by Hall (1988, p. 130), the condition given in (3.2) ensures that with probability one the uncovered part of $\mathbb{R}^d$ has infinite Lebesgue measure. If $\mathbf{E}\big(\nu_d(\Xi_0)\big) = \infty$, which violates (3.2), then with probability one almost all of $\mathbb{R}^d$ is covered by Ξ. However, $\Xi = \mathbb{R}^d$ is not necessarily true.

Now consider the coverage or hitting of parts of Ξ by the rest of Ξ.

Let x_* be the typical point of the germ point process, and let Ξ_* be the corresponding grain. This grain has the same distribution as the typical grain Ξ_0 and can be interpreted heuristically as a grain chosen at random in such a way that every grain is equally likely to be chosen. Thus, it is also called the typical grain (in the Palm sense). The relationship between Ξ_* and the rest of the Boolean model is of interest. Let

$$\Xi' = \bigcup\{\Xi_i + x_i : x_i \neq x_*\}$$

be the union of the remaining grains. Then the following quantities are of interest:

$$p_0 = \mathbf{P}(x_* \text{ is covered by } \Xi'),$$
$$p_G = \mathbf{P}(\Xi_* + x_* \text{ intersects } \Xi'),$$
$$p_s = \text{fraction of boundary of } \Xi_* + x_* \text{ which is covered by } \Xi'.$$

Of course $p_0 > 0$ only if $\mathbf{E}\big(v_d(\Xi_0)\big) > 0$, and p_s only makes sense if the grains of Ξ have finite nonzero surface area. The value $1 - p_G$ is the probability that the typical grain is isolated. If it is positive, then there are even infinitely many isolated grains. Clearly, the mean number of isolated grains with germ points in a given compact set K is $\lambda(1 - p_G)v_d(K)$.

These probabilities satisfy

$$p_0 = p, \tag{3.101}$$
$$p_G = 1 - \mathbf{E}\big(\exp(-\psi(\Xi_0))\big), \tag{3.102}$$
$$p_s = p, \tag{3.103}$$

with $\psi(\Xi_0)$ as defined in equation (3.11). In the particular case of the grains being balls of fixed radius r,

$$p_G = 1 - \exp\big(-\lambda v_d(B(o, 2r))\big) = 1 - \exp(-2^d \lambda b_d r^d). \tag{3.104}$$

In the two-dimensional case, with grains being discs of fixed radius, it is

$$p_G = 1 - (1 - p)^4, \tag{3.105}$$

independently of the particular values of λ and r.

Formula (3.101) can be established as follows. By an analogue of the theorem of Slivnyak–Mecke (see Sections 2.3.4 and 4.4.4) the grain $\Xi_* + x_*$ may be taken to be independent of the other grains, and $\Xi' \stackrel{d}{=} \Xi$. Thus, by taking $x_* = o$, $p_0 = \mathbf{P}(o \in \Xi) = p$.

The proofs of Formulae (3.102) and (3.103) are similar. In the same way the mean number n_0 of grains that hit the typical grain can be obtained:

$$n_0 = \lambda \mathbf{E}\big(v_d(\Xi_0 \oplus \check{\Xi}_0)\big). \tag{3.106}$$

The corresponding distribution is a Poisson distribution. Here $\check{\Xi}_0$ denotes an independent random compact set with the same distribution as $-\Xi_0$. The quantity $v_d(\Xi_0 \oplus \check{\Xi}_0)$ is called *excluded* volume. Note that in the case of isotropy the generalised Steiner formula (6.29) yields formulae for the mean excluded volume $\overline{V}_{\text{ex}} = \mathbf{E}\big(v_d(\Xi_0 \oplus \check{\Xi}_0)\big)$:

$$\overline{V}_{\text{ex}} = 2\overline{A} + \frac{\overline{L}^2}{2\pi} \qquad \text{for } d = 2, \tag{3.107}$$

$$\overline{V}_{\text{ex}} = 2\overline{V} + \frac{\overline{M}\,\overline{S}}{2\pi} \qquad \text{for } d = 3. \tag{3.108}$$

In the case of fibrous grains n_0 is called the 'mean number of fibre crossings per fibre' (Sampson, 2004). If Ξ_0 is a randomly oriented flat convex object, Formula (3.108) becomes

$$\overline{V}_{\text{ex}} = \frac{\overline{A}\,\overline{L}}{2}, \tag{3.109}$$

where $\overline{A}$ is the mean area and $\overline{L}$ the mean perimeter of Ξ_0. Such grains play an important rôle in the context of fracture networks, see Thovert and Adler (2004).

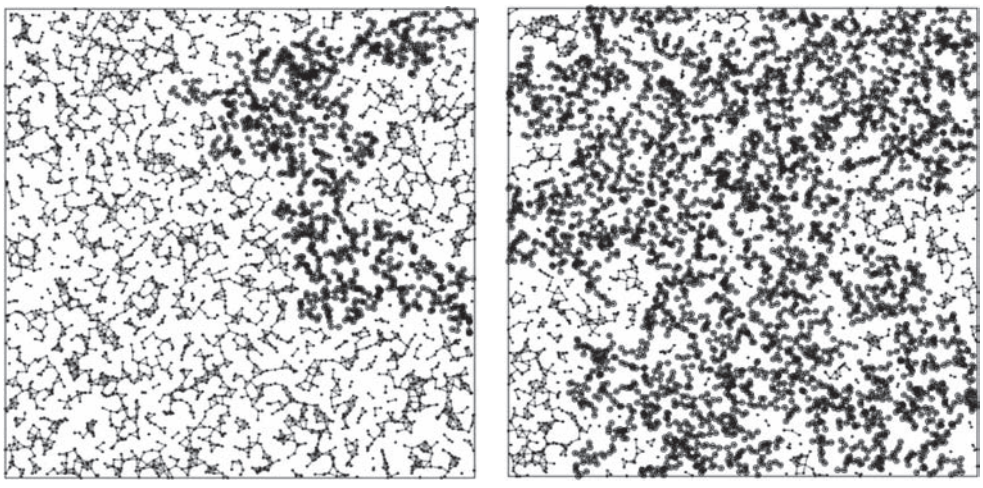
3.3.2 Clumps

A clump is a maximal connected cluster of overlapping grains. Its size is defined as the number of grains constituting it; it may be infinite. Hall (1988, Chapter 4) gave properties of clumps for sparse models. They include asymptotics for the number of n-clumps (formed by n overlapping grains) in a bounded region and the expected number of clumps per unit volume. Błaszczyszyn *et al.* (1999) considered the moments of clump size.

3.3.3 Connectivity

Two germs x_i and x_j are said to connected if $(\Xi_i + x_i) \cap (\Xi_j + x_j) \neq \emptyset$. This leads to the *Boolean connectivity graph* or, in the case of identical spherical grains, *random geometric graph* or *Gilbert graph*; see Figure 3.6. The name 'Gilbert graph' is used to honour E. Gilbert, who studied this graph in his pioneering paper Gilbert (1961).

This graph has been extensively studied in the literature, see Penrose (2003). The probability that the typical vertex of the Gilbert graph is isolated is $1 - p_G$. The degree of the typical vertex has a Poisson distribution with mean n_0 given by (3.106). Furthermore, an asymptotic result (for $\lambda \to \infty$) for *connectivity in a compact set*, that is, for the connectivity of the finite random graph formed by the germs in a given compact set K, is shown in Baccelli and Błaszczyszyn (2009a, Section 13.1.2), using Penrose (1997).



(a) (b)

Figure 3.6 Planar random geometric graph with (a) $\lambda = 0.98\lambda_c$ and (b) $\lambda = 1.02\lambda_c$ and the radius such that the grain area is 1; the critical intensity λ_c is believed to be 1.1281. The largest component is highlighted in black. Courtesy of B. Błaszczyszyn.

3.3.4 Percolation

The problem of the occurrence of giant clumps or of *continuum percolation* is of particular interest. One says that a Boolean model *percolates* if there exists an infinite connected component. (Such a component consists of infinitely many overlapping grains, and it is almost surely an unbounded set.) Percolation in the Boolean model is a practically important question since it often serves as a model for porous media. Fluid flow is impossible in a porous medium with isolated grains. Furthermore, 'continuum percolation is becoming an important tool in the investigation of various telecommunication networks' (van der Hofstad, 2010a).

Percolation is studied by considering that the distribution of the typical grain is fixed, while the intensity λ acts as a variable model parameter. For $d = 1$, the Boolean model does not percolate if the mean length of the typical grain is finite (Meester and Roy, 1996, Theorem 3.1). Thus, only the case $d \geq 2$ is discussed here.

Two percolation characteristics are considered:

(a) the probability p_∞ that the typical grain $\Xi_* + x_*$ of the Boolean model is part of a clump of infinite order, that is, there are infinitely many $\Xi_k + x_k$ which are connected with $\Xi_* + x_*$ by overlapping;

(b) the expected order o_∞ of the clump (= number of its grains) containing the typical grain $\Xi_* + x_*$.

For the case of spherical grains, Hall (1985b) showed that if and only if $\mathbf{E}\big(\nu_d(\Xi_0)^2\big) < \infty$, there is a positive finite constant λ_c' such that $o_\infty < \infty$ for $\lambda < \lambda_c'$ and $o_\infty = \infty$ for $\lambda > \lambda_c'$; if $\mathbf{E}\big(\nu_d(\Xi_0)^{2-1/d}\big) < \infty$, there exists another positive finite λ_c such that $p_\infty = 0$ for $\lambda < \lambda_c$ and $p_\infty > 0$ for $\lambda > \lambda_c$. (Note that $p_\infty = 1$ does not hold, since isolated grains always exist. Nevertheless, since there are infinitely many grains, whenever $p_\infty > 0$, the Boolean model does percolate. Thus, λ_c and λ_c' are the *critical intensities* for the so-called *phase transition*.) The two different moment conditions lead to the interesting case that when $\mathbf{E}\big(\nu_d(\Xi_0)^2\big) = \infty$ and $\mathbf{E}\big(\nu_d(\Xi_0)^{2-1/d}\big) < \infty$, then $\lambda_c > \lambda_c' = 0$, that is, if a Boolean model has a sufficiently small intensity, it does not percolate but the expected number of grains in the clump containing the origin is infinite. Clearly in general $0 \leq \lambda_c' \leq \lambda_c \leq \infty$. Meister and Roy (1994) showed that the Boolean model with spherical grains can have at most one infinite clump.

Other percolation characteristics have also been studied. An example is the existence of a path of overlapping grains from one facet of a cube to the opposite facet, see Zuyev and Sidorenko (1985a,b).

Local percolation probabilities, used by Hilfer (1991, 2000) to characterise the connectivity in porous media, give the fraction of cells where percolation is from one side to the opposite.

It is characteristic of this field that there is a great gap between what mathematicians could prove until now and what physicists 'know' by simulations.

Many authors believe that except for pathological cases, $\lambda_c = \lambda_c'$. The critical intensity λ_c' is easier to determine, usually as the asymptote of the graph of mean clump size against λ. For $\Xi_0 = B(o, r)$ (r chosen so that disc area and ball volume are 1, respectively) simulations yielded

$$\lambda_c' \approx 1.1281 \quad (d = 2) \qquad \text{and} \qquad \lambda_c' \approx 0.3419 \quad (d = 3), \tag{3.110}$$

see Figure 3.6 on p. 91. The corresponding values of p_∞ are

$$p_\infty \approx 0.6763 \quad (d = 2) \qquad \text{and} \qquad p_\infty \approx 0.2896 \quad (d = 3), \tag{3.111}$$

see Lorenz and Ziff (2001) for the spatial case by developing a cluster growth algorithm, and Quintanilla and Ziff (2007) for the planar case by employing the continuum frontier-walk method, using inhomogeneous Boolean models; the latter approach has been theoretically justified by Zuyev and Quintanilla (2003). Note that

$$p_\infty = 1 - \mathrm{e}^{-\lambda_c}. \tag{3.112}$$

Various authors have studied the influence of the radius distribution on λ_c' for the case of random discs or balls; see Torquato (2002, p. 253). Meester *et al.* (1994) conjectures that the monodisperse case minimises λ_c'. Torquato (2012) studies the influence of dimensionality d.

Physicists have suggested various approximations for λ_c' in cases of nonspherical grains; see Bretheau and Jeulin (1989), Mecke and Wagner (1991) and Torquato (2002). The following exposition follows Stoyan and Mecke (2005).

A simple lower bound uses the mean excluded volume $\overline{V}_{\mathrm{ex}}$ given by (3.107) and (3.108); see Balberg *et al.* (1984). From (3.106), the equality

$$\lambda_c' \overline{V}_{\mathrm{ex}} = 1 \tag{3.113}$$

yields an approximation, based on the idea that on the average in a percolating Boolean model each grain should be connected with at least one other grain.

In the particular case of segments of length l as grains of planar Boolean models Formula (3.113) yields the approximation $\lambda_c' = \alpha/l^2$ with $\alpha = \pi/2$. Simulations showed that probably $\lambda_c' = 5.7/l^2$, that is, the true coefficient α is larger than that predicted by (3.113). Nevertheless, the qualitative behaviour of λ_c' seems to be well predicted.

A more precise approach uses a topological argument. It considers the specific connectivity number N_V. For small λ the grains are mainly isolated and N_V is positive. For large λ the Boolean model becomes sponge-like with many small holes and N_V is negative. Therefore, it is natural to assume that the smallest zero of the function $N_A(\lambda)$ in (3.52) or $N_V(\lambda)$ in (3.48) provides a good approximation for λ_c'. The zeros are

$$\lambda_0 = \frac{4\pi}{\overline{L}^2} \qquad \text{for } d = 2, \tag{3.114}$$

$$\lambda_0 = \frac{48\overline{M}}{\pi^2 \overline{S}^2} \left(1 - \left(1 - \frac{\pi^3 \overline{S}}{6\overline{M}^2} \right)^{1/2} \right) \qquad \text{for } d = 3. \tag{3.115}$$

Mecke and Wagner (1991), Mecke and Seyfried (2002) and Stoyan and Mecke (2005) demonstrate for various grain shapes the good quality of the approximation

$$\lambda_c' = \lambda_0. \tag{3.116}$$

For the case of discs of area 1 in $\mathbb{R}^2$ Formula (3.114) yields the approximation $\lambda_c' = 1$ and for balls in $\mathbb{R}^3$ (3.115) gives the value 0.38.

Formula (3.116) together with (3.114) and (3.115) shows that the shape of the grains has a great influence on the value of λ_0. For long and thin grains λ_0 is clearly smaller than for spherical grains; thus percolation happens for such grains earlier than for balls.

Oriented grains have also been considered, for example the cases of squares ($d = 2$), cubes ($d = 3$) (see Torquato, 2002, p. 254) and cylinders (see Jeulin and Moreaud, 2007).

The field of percolation is far wider than is represented in the above account. The interested reader may look into Hughes (1996), Meester and Roy (1996), Grimmett (1999), Torquato (2002, Chapters 9 and 10), Penrose (2003), Hunt (2005), Bollobás and Riordan (2006) and van der Hofstad (2010a).

3.3.5 Vacant regions

Sometimes also the complement Ξ^{c} of a Boolean model Ξ is of interest. Its closure $(\Xi^{\mathrm{c}})^{\mathrm{cl}}$ is also a random closed set. (Since the Boolean model Ξ is a closed set, its complement is not closed – the boundaries belong to Ξ and not to Ξ^{c}. Therefore, it is necessary to take the closure.) Each connected component of $(\Xi^{\mathrm{c}})^{\mathrm{cl}}$ is called a *vacancy*. In the case of a volume fraction p close to 1 and not very small grains, the vacancies are randomly scattered and each resembles a convex polygon or polyhedron, or a union of such shapes; see Figures 3.2 and 3.3. Thus a Boolean model with grains that are Poisson polygons or polyhedra may be often a good approximation for $(\Xi^{\mathrm{c}})^{\mathrm{cl}}$.

This heuristic idea can be partly made mathematically rigorous; see Hall (1988, Chapter 3), Aldous (1989, Chapter H), Molchanov (1996), Michel and Paroux (2003) and Calka *et al.* (2009).

A particular case is as follows: Consider in $\mathbb{R}^d$ a sequence of Boolean models with the same typical grain Ξ_0, which is isotropic, and with intensities λ_n with $\lambda_n \to \infty$. Denote by $\mathcal{V}_n$ the vacancy containing the origin o given that o is not covered. Then it holds

$$\lambda_n \mathcal{V}_n \to C_0 \qquad \text{for } n \to \infty, \tag{3.117}$$

where C_0 is the zero cell (the cell containing the origin) of the Poisson hyperplane tessellation of intensity ϱ (which is S_V for $d = 3$ and L_A for $d = 2$) equal to the mean surface area or mean perimeter of Ξ_0; see Molchanov (1996). (In practical application, some length scale is chosen and fixed during all calculations, in which physical dimensions are ignored.) The convergence of the random compact sets is almost-sure convergence with respect to the Hausdorff metric in $\mathbb{K}'$; see Calka *et al.* (2009), who also considered the convergence of the difference between $\lambda_n \mathcal{V}_n$ and C_0 in fixed direction. (These types of convergence are explained in Molchanov, 2005.)

The convergence in the 'local' limit theorem (3.117) is still true, at least in distribution, for grains of more general shapes, see Molchanov (1996) and Michel and Paroux (2003).

A 'global' limit theorem says that, for $d = 2$, the mean number of vacancies of a Boolean model with $\Xi_0 = B(o, r)$ per unit area is approximately

$$N_A = \lambda^2 \pi r^2 \mathrm{e}^{-\lambda \pi r^2};$$

see Hall (1985a) and Aldous (1989, p. 149).

The problem of connectivity of unbounded components of Ξ^{c} has been studied by Meester and Roy (1996); see also Torquato (2002, p. 253).

For a Boolean model Ξ with typical grain whose volume is of finite second moment, Hall (1988, pp. 142 and 190) gives asymptotic expressions, as $\lambda \to \infty$, for the variance $\mathbf{var}\big(\nu_d(\Xi^{\mathrm{c}} \cap W)\big)$ of the volume of the union of vacant regions in a window W and shows that it is minimised when the grains are uniformly rotated. Rau and Chiu (2011) study the

continuity properties of this asymptotic variance, when the grains are subject to both nonuniform rotations and random shape distortions, where the latter belong to a certain class of linear transformations that are not necessarily volume-preserving.

3.4 Statistics

3.4.1 General remarks

This section discusses in detail statistical analysis for the Boolean model, both because of its intrinsic interest and as an example of parametric statistics for random closed sets. The general nonparametric statistics for random sets will be presented in Section 6.4. There it will be explained how summary characteristics such as the intrinsic volume densities (e.g. p or S_V) and functions such as $C(r)$ or the various contact distribution functions can be estimated.

The main objective of statistical analysis for the Boolean model is the determination of model parameters such as λ, $\overline{L}$, $\overline{A}$, $\overline{\overline{b}}$, $\overline{S}$ and $\overline{V}$, and the distribution of the typical grain, in particular the probability density function of the radii in the case of balls. A further and also very important problem concerns the decision as to whether or not the Boolean model is appropriate in describing a particular data set, that is, testing the Boolean model hypothesis.

These problems present less difficulty if it is possible to discriminate clearly between the grains. Such a discrimination can be carried out, albeit with some difficulty, in the case of lower dimensional grains, that is, in the cases of segments ($d = 2$ or 3) or of convex flats ($d = 3$). In this case the germs can be identified and measured individually, leading to separate studies of the grain distributions and of the point pattern of germs. Study of the grains is a problem in random-compact-set statistics while study of the germs involves point-process statistics.

The statistics become more complicated when the grains have full dimension and overlap. Then the individual grains cannot be identified and the statistics must go indirect ways, using the summary characteristics mentioned in the first paragraph of this section. Since in the corresponding formulae parameters, such as λ and the means of the intrinsic volumes (e.g. $\overline{V}$ or $\overline{S}$), appear and they can be estimated – usually not by maximum likelihood method but by method of moments and related procedures.

When an empirical two-phase structure is given, sometimes it is not clear which phase of the image should be interpreted as the realisation of a Boolean model. For example, when a porous medium is modelled by a Boolean model, either the system of pores or its complement may be fitted by a Boolean model. Thus it may make sense to perform some steps of statistical analysis in parallel, comparing results obtained from assuming first that one and then the other component forms a Boolean model. See also the remark in the first paragraph of Section 3.3.5 about the possibility of approximating the complement of a Boolean model by another Boolean model. A final decision in favour of one component may result from considering empirical contact distribution functions.

This section mainly presents the theory and methods for the stationary and isotropic case. The key references for the anisotropic case are Weil (1995) and Schneider and Weil (2008). For nonparametric estimation of the intensity function of a nonstationary or partially stationary Boolean model, see Molchanov and Chiu (2000).

The data are usually two- or three-dimensional pixel images, the planar images may belong to true planar structures or to planar sections through three-dimensional structures. Also data

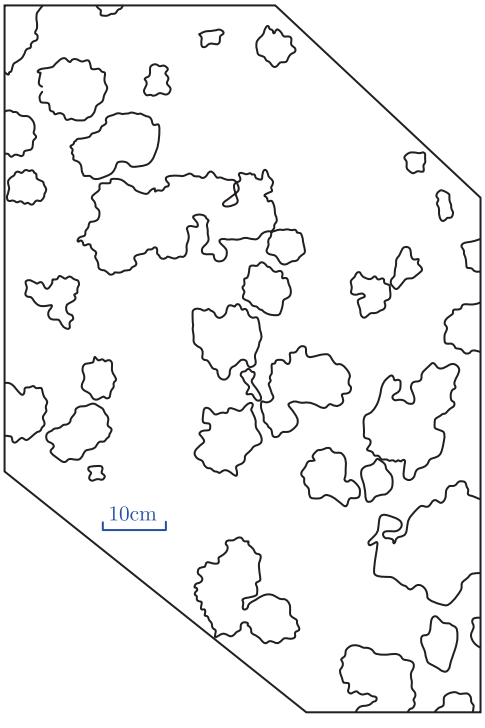


Figure 3.7 A pattern of lichen on a stone. It is in some sense similar to a sample from a Boolean model with convex grains if the boundary roughnesses are ignored.

such as Figure 3.7 are usually digitised for the determination of most summary characteristics. Such situations require the use of numerically stable and robust methods that work well for digital data and automatic analysis, and hence exclude methods that use for example tangent points or curvatures in boundary points. In this context the 'precision hierarchy' of the intrinsic volume density estimators (see p. 99) should be considered. One should try to get along with V_V and S_V and, if possible, to renounce the use of M_V and N_V.

3.4.2 Testing model assumptions

Whether or not the Boolean model assumption is appropriate for given data is usually evaluated by considering functional summary characteristics such as the covariance $C(r)$, contact distribution functions $H_B(r)$ or Mecke's morphological functions. Which function should be used depends mainly on the measurement conditions: if small-angle scattering data are analysed, there is no other choice than to work with $C(r)$, though it is known to be not a powerful test characteristic and the corresponding formulae are rather complicated. The best characteristics for test purposes seem, if applicable, to be the quadrat, discoidal and spherical contact distributions and the morphological functions; see the classical papers by Ohser (1980), Ohser and Stoyan (1980), Diggle (1981), Serra (1982) and Mecke (2000). The linear contact distribution function is not a good mean for such tests since it is also approximately exponential for many random set models that are not Boolean.

Many authors simply show images with empirical summary characteristic diagrams and theoretical counterparts corresponding to the fitted Boolean model. They consider some degree of visual similarity as a proof of a good fit. A correct significance test is a *deviation test* by the parametric bootstrap procedure as already demonstrated in Diggle (1981). It is explained in the following for the case that the summary characteristic used is $H_s(r)$, as also in Diggle (1981).

The fitted Boolean model with the theoretical spherical contact distribution function $H_s(r)$ is simulated k times, and for each pattern the corresponding estimate $\hat{H}_{s,i}(r)$ is determined. Then the deviations

$$\tau_{\text{emp}} = \max_{r \le r_{\max}} \left| H_s(r) - \hat{H}_s(r) \right|, \tag{3.118}$$

where $\hat{H}_s(r)$ is the empirical spherical contact distribution function, and

$$\tau_i = \max_{r \le r_{\max}} \left| H_s(r) - \hat{H}_{s,i}(r) \right| \qquad \text{for } i = 1, 2, \ldots, k \tag{3.119}$$

are determined for a suitable $r_{\max}$. Under the Boolean model hypothesis, τ_{emp} has the same distribution as τ_i and so (if there is no tie) has a rank, when pooled together with all τ_i, uniformly distributed on integers between 1 and $k+1$. A large rank suggests rejection of the Boolean model hypothesis, analogous to the case of testing the Poisson process hypothesis in Section 2.6.4. The (approximate) p-value of this test is given by (2.64).

Instead of using the L^∞-norm, the L^2-norm

$$\tau_{\text{emp}} = \int_0^{r_{\max}} \left(H_s(r) - \hat{H}_s(r) \right)^2 \mathrm{d}r$$

for a suitable $r_{\max}$ and corresponding τ_i can be used to measure deviation. Sometimes *a priori* knowledge may help decide which τ-definition should be used.

This procedure is well explained in Mrkvička and Mattfeldt (2011), albeit in the slightly different context of the quadrat contact distribution function.

If such a test is applied to an image in the form of pixel data, which is usually the case in practice, it is important to digitise the simulated Boolean models in a similar (ideally in the same way) way as the original pixel image, in particular the proportions of pixel size to grain size should be similar. The estimation of $H_s(r)$ has to be carried out as for pixel images and not for images with vector data (which would be possible for simulated data).

When contact distribution functions are used, it is popular to visualise the differences between empirical and theoretical values in the case of convex and isotropic grains for $d = 2$ and 3 as follows.

As a result of Formula (3.55) the logarithm of $1 - H_B(r)$ must be a polynomial of the form:

$$\ln\bigl(1 - H_B(r)\bigr) = -a(B)r - b(B)r^2 - c(B)r^3 \qquad \text{for } r \ge 0, \tag{3.120}$$

with nonnegative coefficients $a(B)$, $b(B)$ and $c(B)$ for each convex B. For the linear contact distribution it is

$$b(B) = c(B) = 0,$$

and for the cases of $d = 2$ and discoidal contact distribution function it is

$$c(B) = 0.$$

Consequently the following informal test suggests itself. For various shapes of B the empirical contact distribution function $\hat{H}_B(r)$ is computed from the given image. If the normalised logarithm

$$\ln\left(1 - \hat{H}_B(r)\right) \Big/ r \tag{3.121}$$

is well approximated by polynomials as in Formula (3.120), for each B considered, then the Boolean model can be considered not inappropriate as a model to fit the data.

The case of nonconvex grains is considered in Chadœuf *et al.* (2008).

3.4.3 Estimation of model parameters

Method of densities

The best method for parameter estimation for a Boolean model is probably the method of densities or intensities, which goes back to Santaló (1976) and Weil (1984). It is based on Formulae (3.50) to (3.53) for $d = 2$ and (3.45) to (3.49) for $d = 3$. The idea is to estimate these intensities and then to use formulae which contain them and model parameters, such as λ and parameters of grain size distributions.

For example, the densities $A_A = p$, L_A, N_A^+ and N_A are connected with model parameters by the four equations known already from p. 80:

$$A_A = 1 - \exp(-\lambda \overline{A}), \tag{3.122}$$

$$L_A = \lambda(1 - A_A)\overline{L}, \tag{3.123}$$

$$N_A^+ = \lambda(1 - A_A), \tag{3.124}$$

$$N_A = \lambda(1 - A_A)\left(1 - \frac{\lambda \overline{L}}{4\pi}\right). \tag{3.125}$$

Consequently, when estimates of the densities are given, the parameters λ, $\overline{L}$ and $\overline{A}$ can be estimated.

The statistical properties of this method are not yet thoroughly investigated. The main problem is the accuracy of the density estimators. Several authors have determined estimation variances by simulation. The asymptotical normality of estimators of A_A, L_A and N_A^+ is known: see Section 6.4 for A_A and Molchanov and Stoyan (1994) for N_A^+ and L_A. The estimator

$$\hat{\lambda} = \frac{\hat{N}_A^+}{1 - \hat{A}_A}$$

for the intensity λ has the property that for a large window W the quantity

$$A(W)^{1/2}(\hat{\lambda} - \lambda) \tag{3.126}$$

follows a Gaussian distribution with zero mean and variance $\lambda/(1-p)$.

The problems that appear in the context of pixel images are discussed in Ohser and Schladitz (2009). An important message of that book is that there is a 'hierarchy in precision': The most precise estimation is possible for A_A and V_V; then comes estimation of L_A and S_V; followed by M_V, N_A^+ and N_V^+; and finally N_A and N_V. Therefore one should try to use in the case of

(i) a *two-parameter model* only Equations (3.122) and (3.123), or (3.45) and (3.46), respectively; important two-parameter models are Boolean models with identical balls (parameters: intensity λ and radius R) and with Poisson polyhedral grains (parameters: λ and ϱ); for example, for the planar model with identical discs, the right-hand terms $\overline{A}$ and $\overline{L}$ can be both expressed by the radius R, and one has two equations for the two unknowns λ and R;

(ii) a *three-parameter model* only Equations (3.122) to (3.125), or (3.45) to (3.49); an important example of a three-parameter Boolean model is a model with balls and a radius distribution with two parameters, for example a gamma (or Schulz), Weibull or lognormal distribution.

In the case of *segments* as grains, using N_A^+ is reasonable. However, unfortunately, the equations (3.123) to (3.125) can yield only the mean segment length. The estimation of the length distribution is nontrivial since usually there are segments that cross the boundary of the window and so their lengths are censored. In this case more sophisticated methods have to be applied; see Chadœuf *et al.* (2000).

The method of densities has a long history. For the case of isotropic grains it first appeared in Santaló (1976) and Kellerer (1983). The general stationary case (with anisotropic grains) was studied in Weil (1988), and Bindrich and Stoyan (1991) used the specific convexity number N_V^+.

Minimum contrast method

The minimum contrast method consists in determining such parameter values that bring a theoretical summary characteristic function as close as possible to its empirical counterpart. In Boolean model statistics the use of the covariance $C(r)$ or correlation function $\kappa(r)$ can be recommended, as it may lead to model parameters different from $\overline{V}$, $\overline{S}$, etc., which could be determined using contact distributions but are better determined by the method of densities. Sometimes minimising contrast is the only method for parameter estimation; see Example 3.3 below.

Let $F(r;\theta_1,\theta_2)$ be a summary characteristic function depending on the unknown parameters θ_1 and θ_2, and let $\hat{F}(r)$ be its empirical counterpart. Then, for example, minimum contrast estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ_1 and θ_2 are those values of ζ_1 and ζ_2 which minimise

$$\int_{r_1}^{r_2} \left(F(r;\zeta_1,\zeta_2) - \hat{F}(r)\right)^2 \mathrm{d}r \tag{3.127}$$

for suitable r_1 and r_2.

Often the integral is replaced by a sum over squared differences of $F(r_i;\zeta_1,\zeta_2)$ and $\hat{F}(r_i)$ for a series of r_i-values.

Heinrich (1993) investigated such estimators for the Boolean model and showed their consistency and asymptotic normality.

Example 3.2. *A pattern of lichen (genus Placodium) on a stone*

The pattern is illustrated in Figure 3.7 on p. 96. Is it possible to fit a planar Boolean model with discoidal grains to these data? Of course, in a strict sense such a model is implausible because of the irregularity of the boundaries of the regions occupied by the lichen – the grains are definitely neither convex nor discoidal. However, if only the large-scale features of the pattern are considered, then the model makes some sense: visual inspection suggests the suitability of a model involving grains. And also biologically, a Boolean model, which is based on the ideas of random germ points and independent grain growth, is not misleading.

The set-theoretic summary characteristics of the pattern of lichen were estimated using an image analyser. The area of the window W shown in Figure 3.7 is 7400 cm^2. Area fraction and mean boundary length per unit area were estimated as

$$\hat{p} = 0.35,$$
$$\hat{L}_A = 0.011\,\mathrm{mm}^{-1}.$$

Counting of lower tangent points for a bit smoothed boundaries led to 38, which gives $\hat{N}_A^+ = 0.0051\,\mathrm{cm}^{-2}$.

Now Formula (3.124) yields $\hat{\lambda} = 0.0079\,\mathrm{cm}^{-2}$, which means that on the average in a window like W, 58 grains are expected.

Formula (3.122) then results in $\overline{A} = 54.5\,\mathrm{cm}^2$ and Formula (3.123) $\overline{L} = 21.4\,\mathrm{cm}$.

For demonstration purposes now the case of a gamma distribution for the radii of discoidal grains is considered. (There is no indication that this distribution is the only candidate; visual inspection of Figure 3.7 suggests that the radius distribution is skewed.) The radius probability density $f_R(r)$ is assumed to be

$$f_R(r) = \frac{\mu^\alpha r^{\alpha-1}}{\Gamma(\alpha)} \exp(-\mu r).$$

Since mean and variance are

$$\mathbf{E}(R) = \frac{\alpha}{\mu} \quad \text{and} \quad \mathbf{var}(R) = \frac{\alpha}{\mu^2},$$

the estimates

$$\hat{\alpha} = 2.03 \quad \text{and} \quad \hat{\mu} = 0.60$$

are obtained.

Example 3.3. *Small-angle scattering analysis of materials*

Small-angle scattering of X-rays or neutrons is able to yield information on random three-dimensional structures at the scale of 1 to 100 nm. The small-angle intensity $I(s)$ is connected with the correlation function $\kappa(r)$ of the random set by

$$\kappa(r) = \frac{1}{2\pi^2 r} \int_0^\infty sI(s) \sin sr \,\mathrm{d}s, \tag{3.128}$$

where $s = \frac{4\pi}{\omega} \sin\theta$, in which 2θ is the scattering angle and ω the wavelength; see Hosemann and Bagchi (1962).

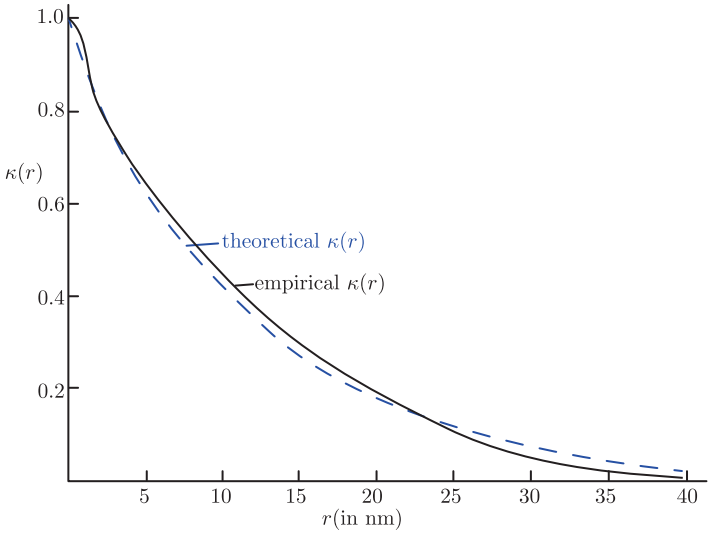


Figure 3.8 The empirical correlation function (solid line) for Co precipitate in a $Co_{75}Fe_5B_{20}$ alloy and the theoretical correlation function (dashed line) of a Boolean model with Poisson polyhedra as grains, in which the parameter of the Poisson polyhedra was fitted by a least squares method.

The first example considers an amorphous $Co_{75}Fe_5B_{20}$ alloy, following Sonntag *et al.* (1981). Figure 3.8 shows the experimental correlation function $\hat{\kappa}(r)$ obtained from small-angle scattering. One possible model for the set Ξ of Co precipitate is a Boolean model of Poisson polyhedra with parameter ϱ as grains. The model correlation function is $\kappa(r)$, where

$$\kappa(r) = \frac{1}{p}\left(\exp\left(6\lambda\frac{\exp(-\varrho r)-1}{\pi\varrho^3}\right) - 1 + p\right) \qquad \text{for } r \geq 0. \tag{3.129}$$

Fitting this to the experimental correlation function $\hat{\kappa}(r)$, via minimisation of

$$\sum_{i=1}^{N}\left(\hat{\kappa}(r_i) - \kappa(r_i)\right)^2$$

using a Marquardt procedure yields the estimates

$$\hat{p} = 0.1,$$
$$\hat{\varrho} = 0.027\,\text{nm}^{-1}.$$

Here the r_i are equidistant values of r for which the experimental correlation function has been calculated from the small-angle scattering data. The dashed curve on Figure 3.8 is the model correlation function $\kappa(r)$ for these parameters.

In order to show that also non-exponential correlation functions are observed, statistical results for a sample of nanoporous silica is investigated. Figure 3.9 shows a TEM image of a nanoporous silica sample. The pores, shown in black, appear to be spherical.

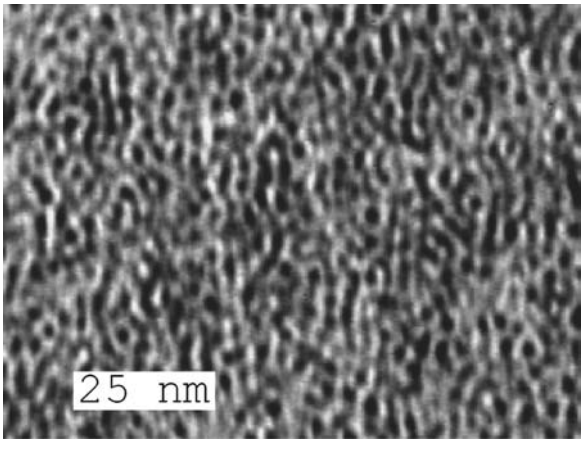


Figure 3.9 TEM image of a nanoporous silica sample. The pores appear in black. Courtesy of H. Hermann.

Methods of image analysis and intuition suggested a truncated power-law distribution for the radii, that is,

$$f_R(r) = \begin{cases} c(r/r_2)^a & \text{for } r_1 \le r \le r_2, \\ 0 & \text{otherwise.} \end{cases}$$

In the paper Hermann *et al.* (2005) the following estimates for the parameters are found:

$$\hat{r}_1 = 1\,\text{nm},$$
$$\hat{r}_2 = 5\,\text{nm},$$
$$\hat{a} = -3.6,$$

and c is a normalising factor.

In Hermann *et al.* (2005) the structure is analysed using methods for hard ball systems, with the comment that in reality the porous phase looks like a system of overlapping balls. Therefore H. Hermann kindly re-analysed the data using a Boolean model with spherical grains. Using the radius distribution above and Formulae (3.18) and (3.22) for $C(r)$ he found a suitable estimate of λ and calculated the dashed curve for $\kappa(r)$ in Figure 3.10. The estimate $\hat{\lambda} = 0.048\,\text{nm}^{-3}$ finally led to the estimate of V_V as 0.7, which is considered by Hermann as a bit too large. (However, in the approach of Hermann *et al.*, 2005, no estimate of V_V can be given.)

Estimation of the radius distribution of spherical grains

Various authors have developed statistical methods for the determination of the radius probability density function $f_R(r)$ in the case of Boolean models with *spherical* (or *discoidal*) *grains*. Many of them seem numerically instable, require difficult measurements and yield in the best case only qualitative information. One of these methods is that in Thovert *et al.* (2001). They consider a grid of test points in Ξ and determine for each of them the largest ball entirely lying in Ξ and covering it, the test point. The sample of the corresponding radii

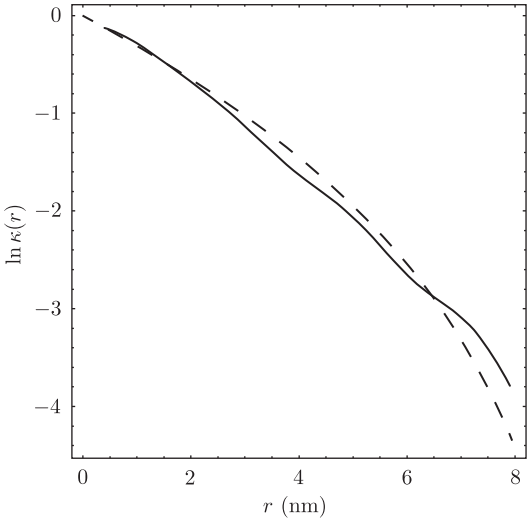


Figure 3.10 The empirical correlation function (solid line) for the silica sample, and the theoretical correlation function (dashed line) of a Boolean model with spherical grains, the radii of which follow a truncated power-law distribution. The ordinate is scaled logarithmically in order to show clearly deviations from exponential curves. Courtesy of H. Hermann.

(which tend to be equal for test points close together) is considered as a sample corresponding to the volume-weighted version of the density $f_R(r)$. The papers Thovert *et al.* (2001) and Thovert and Adler (2011) describe how to determine the radii numerically, based on the covariance, and discuss the accuracy of the method. The authors confess that some '*a posteriori* verification' is necessary. See also Emery *et al.* (2012).

Molchanov and Stoyan (1994) propose a kernel estimator of the diameter probability density, whose asymptotic normality is proved by Heinrich and Werner (2000).

Others

Further methods of statistics for Boolean models are given in Hall (1988), Schmitt (1991), Weil (1995), Molchanov (1997) and Schneider and Weil (2008).

3.5 Generalisations and variations

Some generalisations of the Boolean model may be of use for description of structures of greater complexity. A straightforward generalisation involves the weakening of model assumptions. It seems especially useful to suspend the assumption that the germs constitute a Poisson process. Random structures that are essentially Boolean models, but that use general point processes to provide the germs, are called germ–grain models. They are studied in Chapter 6. Such models may display a greater or a lesser irregularity of structure than the Boolean model, according to the particular properties of the germ process.

Two independent Boolean models $\Xi^{(1)}$ and $\Xi^{(2)}$ considered together can be used to produce patterns of greater irregularity. Clearly, the union $\Xi^{(1)} \cup \Xi^{(2)}$ is again a Boolean model. Three more complicated variations are considered now.

Variation 1. Suppose that $\Xi^{(1)}$ and $\Xi^{(2)}$ have volume fractions $p^{(1)}$ and $p^{(2)}$ and covariances $C^{(1)}(r)$ and $C^{(2)}(r)$. Then the intersection

$$\Xi^{(\mathrm{i})} = \Xi^{(1)} \cap \Xi^{(2)} \tag{3.130}$$

is a closed random set; see Figure 3.11(a) for an illustration. Its basic characteristics are: volume fraction

$$p^{(\mathrm{i})} = p^{(1)} p^{(2)} \tag{3.131}$$

and covariance

$$C^{(\mathrm{i})}(r) = C^{(1)}(r) \cdot C^{(2)}(r) \qquad \text{for } r \geq 0. \tag{3.132}$$

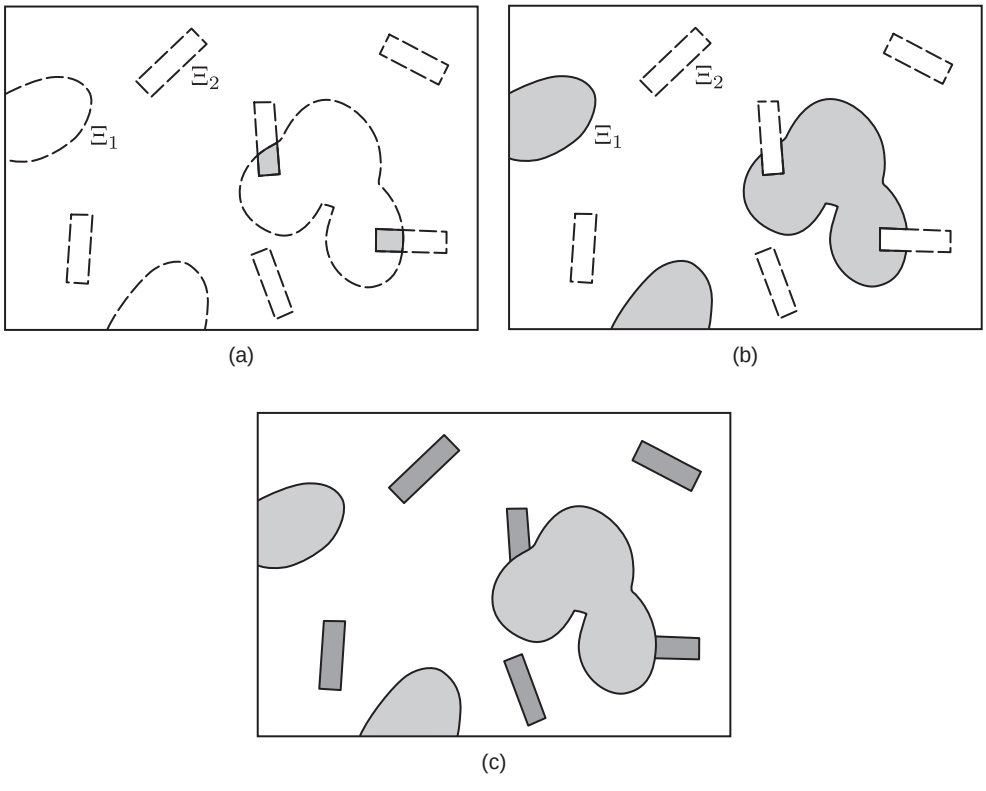


Figure 3.11 Three variations of Boolean models: (a) intersection of two Boolean models; (b) intersection of a Boolean model with the complement of another; (c) a simple three-component model.

The calculation of the capacity functional is complicated. This model appears in Savary *et al.* (1999) and is used in the context of clusters of small grains. Jeulin (2012) speaks about 'multi scale Boolean models'.

Variation 2. Another closed set which can be built from $\Xi^{(1)}$ and $\Xi^{(2)}$ is the intersection of one and the closure of the complement of the other:

$$\Xi^{(\mathrm{ic})} = \Xi^{(1)} \cap \left((\Xi^{(2)})^{\mathrm{c}}\right)^{\mathrm{cl}}. \tag{3.133}$$

Here $\left((\Xi^{(2)})^{\mathrm{c}}\right)^{\mathrm{cl}}$ is the closure of the complement of $\Xi^{(2)}$. This is illustrated in Figure 3.11(b). Such constructions are used in Molchanov (1992) to describe spatial censoring in Boolean model statistics.

The basic characteristics of $\Xi^{(\mathrm{ic})}$ are: volume fraction

$$p^{(\mathrm{ic})} = p^{(1)}(1 - p^{(2)}) \tag{3.134}$$

and covariance

$$C^{(\mathrm{ic})}(r) = C^{(1)}(r)\left(1 - 2p^{(2)} + C^{(2)}(r)\right) \qquad \text{for } r \geq 0. \tag{3.135}$$

Savary *et al.* (1999) also uses $\Xi^{(1)} \cap \Xi^{(2)} \cap \left((\Xi^{(3)})^{\mathrm{c}}\right)^{\mathrm{cl}}$. In a similar way, but extending beyond the province of random closed sets, models can be constructed for multi-component structures. A simple example is the three-component model below, studied by Greco *et al.* (1979) and Serra (1982). Statistical problems related to this are considered in Molchanov (1997).

Variation 3. The sets $\Xi^{(1)}$ and $\Xi^{(2)}$ may be combined to produce a three-phase pattern as in Figure 3.11(c). Its components are:

- component 1: $\Xi^{(1)}$;
- component 2: $\Xi^{(2)} \setminus \Xi^{(1)}$ (the set-difference can be interpreted as a pattern given by $\Xi^{(2)}$ but destroyed in part by $\Xi^{(1)}$);
- component 3: the remainder, $(\Xi^{(1)} \cup \Xi^{(2)})^{\mathrm{c}}$.

This model is used in Greco *et al.* (1979) for the description of sinter metals. In such a case component 1 models the haematites, component 2 the slag, and component 3 the calcium ferrites.

A further model closely related to the Boolean model is Matheron's 'dead-leaves' model; see Serra (1982). Boolean models also appear in the context of birth-and-growth processes; see Section 6.6.4 and Molchanov and Chiu (2000).

Finally, it must be mentioned that in particular physicists also consider discrete Boolean models, defined on lattices. There the germs are random lattice points and the grains are subsets of the lattice; see Hall (1988), Mecke (1994, 1996, 2000) and Stoyan and Mecke (2005).

3.6 Hints for practical applications

Planar case

In the majority of applications a Boolean model with discoidal grains should be a good first choice. The estimators considered above then yield values of λ and of the first two moments of the random radius. If a two-parameter distribution (such as normal, truncated normal, lognormal, gamma or Weibull) is plausible, then the Boolean model is completely specified, and simulations are possible. This may be sufficient if the boundaries are curved. Sometimes this model is also in approximate agreement with physical or biological theories using terms such as 'germs' and 'circular growth'.

If the boundaries are straight, then the Poisson polygon is indicated for use as a grain. In this case the model has only two parameters, intensity λ and polygon parameter ϱ. This is convenient for estimating, but for fitting the model to empirical data this may turn out to be too special. If the area fraction is small, then the Poisson-polygonal Boolean model may be justified by the arguments of Section 3.3.5.

Spatial case

Analogously to the planar case, spherical and Poisson-polyhedric grains are particularly interesting. Usually the real structures are small and/or opaque, and their direct investigation is not possible. In such cases the corresponding parameters can be estimated using data obtained from computerised tomography, X-ray scattering (see Example 3.3) or from planar sections. In Section 10.3 stereological methods for the Boolean model are presented. Clearly, it is then important that planar sections of a Boolean model are again Boolean models.

Two-step statistical procedure

Most known statistical procedures which yield more than the densities and mean values of intrinsic volumes must be considered as numerically unstable. Therefore their results should be considered as preliminary and as a first step of statistics. A typical preliminary result is qualitative information of grain shape and radius probability density function type. Based on the first step then the second step yields the model parameters by the method of densities. It guarantees that, for example in the spatial case, V_V and S_V of the model coincide with the corresponding data values. Careful statisticians will make a fine-tuning based on simulations: the estimated parameters are *a posteriori* so modified that summary characteristics of simulated samples are as close as possible to the empirical ones.

A critical assessment of Boolean models

Finally, the following remarks about difficulties and intrinsic limitations of the Boolean model may be helpful to the interested reader.

- *Interpretation.* In many cases, the suggestive nature of the Boolean model terminology ('germs' as origins of growth, or locations of physical entities) will not coincide with physical reality. This is not strictly speaking a disadvantage of the Boolean model, but more a tendency to be resisted in its interpretation. The Boolean model may fit a data set to a greater or lesser degree, but that of itself does not justify a physical interpretation of the germ–grain mechanism.

Two kinds of interpretation can be distinguished. In the first, which might be called the *genetic* interpretation, the individual grains of the Boolean model correspond to physically identifiable entities. The second interpretation, the 'representation' interpretation, avoids drawing out such a correspondence, but rather uses the Boolean model as a moderately flexible means of producing patterns to compare with the image and to describe it by few parameters only. Example 3.2 on lichen falls in the second category: it is tempting here to identify grains of the Boolean model with individual lichen entities, but in fact such an identification cannot be substantiated without rigorous evidence derived from the botanical context.

The point here is that conclusions appropriate to a genetical interpretation will be misleading, even nonsensical, in a representation context. In an application to dust-particle counting, clearly one is interested in estimating the number of particles. Conceivably one might attempt to estimate lichen 'growth sites' in a similar way, but without evidence favouring a genetic interpretation such an attempt is ill-judged. However, the representation interpretation still provides a possible description of the data set as an image.

- *Interaction between grains.* The previous section has already introduced the generalisation of replacing the Poisson germ process by a more general point process. In many cases, the assumption of independence between the grains is implausible. Consider for example various packing and coverage problems, where intersection of grains is prohibited. In such a case the Boolean model is clearly inadequate. Then germ–grain models form a possible alternative; see Section 6.5. Particularly important models are the 'morphological Gibbs processes'; see Brodatzki and Mecke (2001), Stoyan and Mecke (2005), Dereudre and Lavancier (2009) and Møller and Helisová (2010).

  When some interaction between the grains is merely a possibility to be considered, the Boolean model can serve as a convenient benchmark, so that the strength of the interaction, if any, can be assessed.

- *Nonconvex grains.* Many of the useful formulae for the Boolean model are for the case of convex grains. However, in many data sets arising in practice, such as the one in Example 3.2, the grains are clearly nonconvex. It is indeed the case that larger-scale features of the data set can still be modelled by a Boolean model with convex grains. However, care must be taken over the interpretation of quantities calculated from the model. As shown above, it is possible to derive formulae for the case of nonconvex grains.