

## SET-VALUED ESTIMATORS FOR MEAN BODIES RELATED TO BOOLEAN MODELS

I. S. MOLCHANOV

*Centrum voor Wiskunde en Informatica, Amsterdam*

*(Received 25 February 1993; in final form 6 June 1995)*

**Summary.** The stationary Boolean model in the Euclidean space with a convex typical grain is considered. Relations between the mean (or the Aumann expectation) of the difference body of the grain and other characteristics of the Boolean model are discussed, and set-valued estimators for the mean difference body are suggested.

*AMS 1980 subject classifications:* Primary 60D05; Secondary 62M30.

*Key words:* Random sets, Boolean models, Aumann expectation, mean body, linear contact distribution function, intensity estimator.

### 1. INTRODUCTION

The Boolean model is the most popular and best investigated random set model with a wide spectrum of applications, see Serra (1982), Stoyan, Kendall and Mecke (1987), Cressie (1991). The Boolean model  $\Xi$  in the Euclidean space  $\mathbf{R}^d$  is defined to be the union

$$\Xi = \bigcup_{i=1}^{\infty} (x_i + \Xi_i) \quad (1.1)$$

of iid random compact sets  $\Xi_1, \Xi_2, \dots$ , shifted by the points  $x_i$  belonging to a stationary Poisson point process in  $\mathbf{R}^d$  with intensity  $\lambda$ . The *typical grain* of the Boolean model is a random closed set  $\Xi_0$  with the same distribution as the  $\Xi_i$ . For rigorous definitions the reader is referred to Matheron (1975, pp.61, 139) and Stoyan *et al.* (1987, pp. 66,168).

The first step of an application of the Boolean model to real data is estimating the model parameters using observations of the union set  $\Xi$ . These parameters of  $\Xi$  are the intensity  $\lambda$  of the underlying Poisson point process and the distribution of the typical grain  $\Xi_0$ .

Until now the most investigations have dealt with estimators of  $\lambda$  and the expected values of the Minkowski functionals of the typical grain  $\Xi_0$ , see Serra (1982), Stoyan *et al.* (1987), Cressie and Laslett (1987), Hall (1988), Weil (1988), Cressie (1991), and Schmitt (1991). While the distribution of the stationary Poisson point process is determined by the value of its intensity  $\lambda$ , the distribution of  $\Xi_0$ , in

general, cannot be retrieved by known mean values of the Minkowski functionals. In order to reconstruct the distribution of  $\Xi_0$  we have to know more about the shape of this set. Some shape characteristics and the corresponding estimation techniques were discussed in Molchanov (1990, 1991, 1994) for special typical grains (deterministic or consisting of a finite number of points). Here we deal with the estimation of the mean body of the difference body  $\tilde{\Xi}_0 = \Xi_0 \oplus \check{\Xi}_0 = \{x - y: x, y \in \Xi_0\}$  of the grain. Note that  $\tilde{\Xi}_0/2$  is the central symmetrization of the set  $\Xi_0$ . We mostly work with planar Boolean models, i.e., we assume  $d = 2$ . In the anisotropic case the mean body provides information about the shape of the typical grain, while, unfortunately, for any isotropic typical grain the mean is a ball centred at the origin. Other estimators for the mean body (derived by means of different techniques) were discussed in Weil (1994, 1995) and Molchanov (1995).

The paper is organized as follows. Section 2 provides the definition of the mean body. In Section 3 several relations between the mean body and other characteristics of a Boolean model are derived. These formulae are the base for the corresponding estimators proposed in Section 4. Section 5 contains numerical examples.

## 2. THE MEAN BODY OF A RANDOM COMPACT SET

Let us recall the definition of the expectation (or mean body) of a random compact set, see Vitale (1988). It is known that any compact convex set  $F$  can be described by means of the corresponding support function  $h(F, u)$  for  $u$  running through the unit sphere in  $\mathbf{R}^d$ . The support function  $h(F, u)$  is defined to be the maximum of scalar products  $\langle u, x \rangle$  for  $x \in F$ . For a random compact convex set  $A$  the expected value of the support function  $\mathbf{E}h(A, u)$  (if everywhere finite) is again a support function of a certain convex set  $\mathbf{E}A$ , which is called the (Aumann) expectation or the *mean body* of  $A$ .

Throughout the paper we suppose that the typical grain  $\Xi_0$  is *convex*. Also we suppose that the radius of the circumscribed ball of  $\Xi_0$  has a finite  $d$ th moment. This is necessary to have a nontrivial union set  $\Xi$  in (1.1), see Heinrich (1992).

The mean body of the typical grain,  $\mathbf{E}\Xi_0$ , is an important characteristics of the Boolean model. The corresponding estimation problem is discussed in Weil (1995). Here we propose another approach which leads to estimators of the set

$$\tilde{\Xi}_0 = \mathbf{E}(\Xi_0 \oplus \check{\Xi}_0).$$

Here  $\oplus$  is the Minkowski addition and  $\check{\Xi}_0 = \{-x: x \in \Xi_0\}$ . Note that the set  $\tilde{\Xi}_0 = \Xi_0 \oplus \check{\Xi}_0$  is called the *difference body* of  $\Xi_0$ , see Leichtweiß (1980). Its expectation  $\tilde{\Xi}_0$  is said to be *mean difference body* of  $\Xi_0$ . If  $\Xi_0$  is centrally symmetric, then  $\Xi_0 = \frac{1}{2}\tilde{\Xi}_0$ , and we obtain an estimator of  $\mathbf{E}\Xi_0$ . The estimators which will be proposed are set valued. Their consistency is understood with respect to the Hausdorff metric as the sampling window grows infinitely.

To make this paper more self-contained recall also that the distribution of the Boolean model (1.1) is determined by the capacity functional

$$T(K) = \mathbf{P}\{\Xi \cap K \neq \emptyset\} = 1 - \exp\{-\lambda \mathbf{E}\mu_d(\Xi_0 \oplus \check{K})\}, \quad (2.1)$$

for  $K$  running through the class  $\mathcal{K}$  of all compact subsets of  $\mathbf{R}^d$ , see Matheron (1975, p. 62), Stoyan *et al.* (1987, p. 69). Here  $\mu_d$  is the  $d$ -dimensional Lebesgue measure in  $\mathbf{R}^d$ . Note that  $p = T(\{o\})$  is the volume fraction of  $\Xi$ .

### 3. THE MEAN DIFFERENCE BODY AND OTHER CHARACTERISTICS OF THE BOOLEAN MODEL

#### 3.1. The Mean Difference Body and the Linear Contact Distribution Function

The linear contact distribution function of a random closed set  $\Xi$  is defined as

$$H_u(r) = 1 - \mathbf{P}\{\Xi \cap [o, ru] = \emptyset\} / (1 - p), \quad r \geq 0,$$

where  $u$  is a unit vector. The quantity  $H_u(r)$  is the conditional probability that the segment  $[o, ru]$  hits  $\Xi$ , conditional on  $\Xi$  not containing the origin  $o$ , see Stoyan *et al.* (1987, p. 74).

It follows from (2.1) that for a Boolean model  $\Xi$

$$H_u(r) = 1 - \exp\{-\lambda \mathbf{E} \mu_d(\Xi_0 \oplus [o, -ru])\} \exp\{\lambda \mathbf{E} \mu_d(\Xi_0)\}.$$

Clearly,

$$\mu_d(\Xi_0 \oplus [o, -ru]) = \mu_{d-1}(\Pi_{u^\perp} \Xi_0) r + \mu_d(\Xi_0),$$

where  $\Pi_{u^\perp} \Xi_0$  is the projection of  $\Xi_0$  in the hyperplane orthogonal to  $u$ .

In the planar case  $\Pi_{u^\perp} \Xi_0$  is the segment of the length  $h(\Xi_0, v) + h(\Xi_0, -v) = h(\tilde{\Xi}_0, v)$  for the unit vector  $v$  orthogonal to  $u$  (in the following we always suppose that  $u$  and  $v$  are related in this way). Thus,

$$\mu_1(\Pi_{u^\perp} \Xi_0) = h(\tilde{\Xi}_0, v).$$

Hence

$$H_u(r) = 1 - \exp\{-\lambda r \mathbf{E} h(\tilde{\Xi}_0, v)\}, \quad r \geq 0. \tag{3.1}$$

Thus, in the planar case ( $d = 2$ ) the mean difference body  $\bar{\Xi}_0 = \mathbf{E} \tilde{\Xi}_0$  has the following support function

$$h(\bar{\Xi}_0, v) = -\frac{1}{\lambda r} \log(1 - H_u(r)), \tag{3.2}$$

where  $\langle u, v \rangle = 0$  and  $\|v\| = 1$ . Note that the right-hand side of (3.2) does not depend on  $r$ .

It should be noted that formula (3.2) for the support function of the mean difference body is valid only in the planar case. In the following we will consider only the planar case ( $d = 2$ ) unless otherwise stated.

### 3.2. The Mean Difference Body and Star-Shaped Expected Sets

The linear contact distribution function is the conditional distribution function of the random variable

$$\zeta_u = \sup \{r: [o, ru] \cap \Xi = \emptyset\},$$

conditional on  $o \notin \Xi$ . It follows from (3.1) that the corresponding conditional expectation is given by

$$\mathbf{E}(\zeta_u | o \notin \Xi) = (\lambda \mathbf{E}h(\bar{\Xi}_0, v))^{-1} = (\lambda h(\bar{\Xi}_0, v))^{-1}. \quad (3.3)$$

For each  $x \notin \Xi$  let  $S_x(\Xi)$  denote that star (or visible set), i.e.

$$S_x(\Xi) = \{y \in \mathbf{R}^d: [x, y] \cap \Xi = \emptyset\}, \quad (3.4)$$

see Stoyan *et al.* (1987, p. 81), Serra (1982, p. 472) and Yadin and Zacks (1985). Clearly,  $S_x(\Xi)$  is a star-shaped random set. Its distribution does not depend on  $x$ .

The star-shaped set  $S_x(\Xi)$  is determined by its radius-vector function  $\zeta_u$  defined on the unit circle. The *radius-vector expectation*  $\mathbf{E}_r S_x(\Xi)$  of  $S_x(\Xi)$  is defined in Stoyan and Stoyan (1994, p.111) to be the star-shaped set whose radius-vector function is equal to  $\mathbf{E}\zeta_u$ . The conditional expectation of  $\zeta_u$  is given by (3.3). Thus, the *conditional radius-vector expectation* of  $S_o(\Xi)$  (i.e., for  $x = o$ ) is equal to

$$\mathbf{E}_r(S_o(\Xi) | o \notin \Xi) = \bigcup_{\|u\|=1} \{[o, (\lambda h(\bar{\Xi}_0, v))^{-1}u]: \|v\|=1, \langle u, v \rangle = 0\}.$$

Hence the support function of  $\bar{\Xi}_0$  is given by

$$h(\bar{\Xi}_0, v) = (\lambda \sup \{r: ru \in \mathbf{E}_r(S_o(\Xi) | o \notin \Xi)\})^{-1}, \quad \|u\| = \|v\| = 1, \quad u \perp v. \quad (3.5)$$

Note again that, unfortunately, for  $d \geq 3$  there is no such an interpretation of the mean difference body in terms of the linear contact distribution function and radius-vector averages.

### 3.3. The Mean Difference Body and Sections of the Boolean Model

In this subsection the dimension  $d$  can be arbitrary,  $d \geq 2$ . For a unit vector  $u$  let  $u^\perp$  be the  $(d-1)$ -dimensional hyperplane orthogonal to  $u$ . It was shown in Matheron (1975, p.144) that  $\Xi \cap u^\perp$  is again a Boolean model with intensity

$$\lambda(u) = \lambda \mathbf{E}\mu_1(\Pi_u \Xi_0),$$

where  $\Pi_u(\Xi_0)$  is the projection of  $\Xi_0$  in the direction of  $u$ . Hence

$$\lambda(u) = \lambda \mathbf{E}h(\bar{\Xi}_0, u), \quad (3.6)$$

i.e. the mean difference body  $\bar{\Xi}_0$  can be found through intensities of the corresponding  $(d-1)$ -dimensional sections of the Boolean model  $\Xi$  divided by the intensity of  $\Xi$ . If  $d = 2$ , then the corresponding sections are one-dimensional.

3.4. *The Mean Body and the Boundary of the Boolean Model*

Let us consider again the planar Boolean model  $\Xi$ . The boundary  $\partial\Xi$  of  $\Xi$  can be considered to be a fibre process in the sense of Mecke and Stoyan (1980). The rigorous definition of such a process can be found also in Stoyan *et al.* (1987, p.233). Roughly speaking, a planar fibre process is a random set composed from curve pieces placed randomly on the plane.

For a unit vector  $u$  define  $l_u$  to be the line  $\{xu: x \in \mathbf{R}\}$ . Then  $l_u \cap \partial\Xi$  is a point process. Its intensity is denoted by  $P_L(u)$ , and the function  $P_L(\cdot)$  is said to be the *rose of intersections* of  $\partial\Xi$ . It follows from Stoyan *et al.* (1987, p.78) that

$$P_L(u) = 2\lambda(u)(1 - p), \tag{3.7}$$

where  $\lambda(u)$  is the intensity of the Boolean model  $\Xi \cap l_u$ , and  $p$  is the area fraction of  $\Xi$ .

It is known (see Stoyan *et al.*, 1987, p.236) that  $P_L(u)/2$  is the support function of the compact convex set  $\mathfrak{G}$  called the *Steiner compact* of the fibre process  $\partial\Xi$ . The Steiner compact is of use to describe anisotropy of the fibre process, see Rataj and Saxl (1989, 1992) for further discussion.

On the other hand,  $\lambda(u)/\lambda$  is the support function of the set  $\Xi_0$  rotated by the angle  $\pi/2$ , see (3.6). Thus

$$\Xi_0 = \frac{\omega_{\pi/2} \mathfrak{G}}{\lambda(1 - p)}, \tag{3.8}$$

where  $\omega_{\pi/2} \mathfrak{G}$  is the rotation of  $\mathfrak{G}$  to  $\pi/2$ . Note also that the intensity of the fibre process  $\partial\Xi$  coincides with the mean boundary length per unit area (specific boundary length)  $L_A$  for the Boolean model  $\Xi$ . This parameter is given by

$$L_A = \lambda(1 - p)EU(\Xi_0),$$

where  $U(\Xi_0)$  is the perimeter of  $\Xi_0$ , see Stoyan *et al.* (1987, p.88). On the other hand,  $L_A = U(\mathfrak{G})/2$ . Thus we obtain the following result

$$\frac{E\Xi_0}{EU(\Xi_0)} = \frac{2\omega_{\pi/2} \mathfrak{G}}{U(\mathfrak{G})}. \tag{3.9}$$

If  $\Xi_0$  is centrally symmetric, then we get  $\Xi_0 = 2\Xi_0$ , and

$$\frac{E\Xi_0}{EU(\Xi_0)} = \frac{\omega_{\pi/2} \mathfrak{G}}{U(\mathfrak{G})}, \tag{3.10}$$

see also Weil (1993).

4. ESTIMATION OF THE MEAN BODY

4.0. *Estimation of the Intensity  $\lambda$*

Formulae (3.2), (3.5), (3.6) and (3.8) of Section 3 give possible foundations for the estimation of the mean body  $\Xi_0$ . All these formulae involve the intensity  $\lambda$  of the Boolean model  $\Xi$ ; (3.6) includes also intensities of  $(d - 1)$ -dimensional sections.

For *isotropic* Boolean models, the intensity can be estimated by means of one of the methods, which have been reviewed in Lantuejoul and Schmitt (1991). These methods are

- the method based on the Steiner formula (the *minimum contrast method* using the corresponding contact distribution function, see Cressie and Laslett, 1987, Heinrich, 1993). It works for convex grains only;
- *the method of intensities*, where parameters of the Boolean model are chosen to match the empirical values of the intensities of Minkowski measures, see Weil (1988) and Stoyan *et al.* (1987). This method is sometimes called also Weil's or the intensity method. It is designed for convex grains, although its newest variant allows to work with non-convex grains and anisotropic Boolean models as well, see Weil (1995);
- *Schmitt's method*, see Schmitt (1991).

All these methods work in the space  $\mathbf{R}^d$  of general dimension  $d$ . Schmitt's method is of particular value for Boolean models with bounded and non-convex typical grains. It works also in the anisotropic case. It was shown by means of simulation experiments that, for Boolean models with convex grains, the method of intensities provides better results in comparison with the two other methods, see Lantuejoul and Schmitt (1991) and Schröder (1992).

For *anisotropic* Boolean models (this is the usual case in this paper) we estimate the intensity by means of the following simple method proposed in Serra (1982, p.492) and Stoyan *et al.* (1987, pp.78,201). We associate with each grain its tangent point with respect to a fixed direction (in the plane it can be viewed as its lower-left tangent point). Some of these tangent points are covered by other grains, other are exposed. The intensity of the point process of the exposed tangent points is denoted by  $N_A^+$ . In other words,  $N_A^+$  is the specific convexity number of  $\Xi$  considered in Serra (1982, p.142). It is known that

$$N_A^+ = \lambda(1 - p).$$

Let us consider a sequence of expanding windows  $W_t, t > 0$ . A typical choice is  $W_t = tW$  with a certain convex compact set  $W$  of positive Lebesgue measure and containing the origin as an interior point. The intensity  $N_A^+$  can be estimated by counting the number of the exposed tangent points inside the window  $W_t$  divided by the Lebesgue measure (or area in the planar case) of  $W_t$ :

$$\hat{N}_{A,t}^+ = N_A^+(W_t) / \mu_d(W_t).$$

The volume fraction  $p$  can be estimated by the quotient

$$\hat{p}_t = \mu_d(\Xi \cap W_t) / \mu_d(W_t). \quad (4.1)$$

Then an estimator of  $\lambda$  is given by

$$\hat{\lambda}_t = \hat{N}_{A,t}^+ / (1 - \hat{p}_t). \quad (4.2)$$

This estimator is strong consistent and asymptotically normal at  $t \rightarrow \infty$ . Its asymptotic properties were studied in Molchanov and Stoyan (1994b). Note that its form does not depend on the dimension  $d$ , but the convexity of the typical grain is required.

The next step in the estimation of the mean difference body  $\Xi_0 = \mathbb{E}\tilde{\Xi}_0$  depends on the choice of the formula which relates this mean body with other characteristics of the Boolean model  $\Xi$ .

4.1. Empirical Capacities and the Empirical Linear Contact Distribution Function

This method is based on the uniform law of large numbers for empirical capacity functionals, see Molchanov (1988,1990). It is known that the stationary Boolean model  $\Xi$  given by (1.1) is ergodic, see Stoyan *et al.* (1987, p.171). Similarly to Section 3.1 we consider here only the planar case. Take a sequence of expanding planar windows  $W_t = tW, t > 0$ . It follows from Mase (1982) that, for any given compact  $K_0$ ,

$$\hat{T}_t(K) = \frac{\mu_2((\Xi \oplus \check{K}) \cap (W_t \ominus K_0))}{\mu_2(W_t \ominus K_0)}, \quad K \in \mathcal{X}, \quad (4.3)$$

is a strongly consistent estimator of  $T(K), K \subset K_0$ , as  $t \rightarrow \infty$ . Here  $\ominus$  denotes the Minkowski subtraction, i.e.  $W_t \ominus K_0 = \{x: x + K_0 \subseteq W_t\}$ . The values of  $\hat{T}_t(K)$  can be determined by measurements of area fractions (see Ripley, 1986), since  $\hat{T}_t(K)$  is obviously a (minus-sampling) estimator of the area fraction of the random set  $\Xi \oplus \check{K}$ . For general dimensions,  $\hat{T}_t(K)$  can be evaluated by (4.3) with the Lebesgue measure  $\mu_d$  instead of the area  $\mu_2$ .

Suppose that the typical grain  $\Xi_0$  has an a.s. non-empty interior. Since the Boolean model  $\Xi$  is a stationary set, it follows from the result of Molchanov (1988) that  $\hat{T}_t(K)$  is a *locally uniformly* strong consistent estimator of  $T(K)$ , i.e. for each compact  $K_0$

$$\sup_{K \subseteq K_0, K \in \mathcal{X}} |\hat{T}_t(K) - T(K)| \rightarrow 0 \quad \text{a.s. as } t \rightarrow \infty. \quad (4.4)$$

Formula (4.3) leads to the following estimator of the linear contact distribution function

$$\hat{H}_{u,t}(r) = 1 - (1 - \hat{T}_t([0, ru])) / (1 - \hat{p}_t), \quad (4.5)$$

where  $\hat{p}_t = \hat{T}_t(\{o\})$ , see also (4.1). It follows from (4.4) that, for any  $c > 0$ , the estimator  $\hat{H}_{u,t}(r)$  converges to the linear contact distribution function (3.1) uniformly for  $r \in [0, c]$ , as  $t \rightarrow \infty$ . Thus, (3.2) suggests the following estimator of the support function of  $\Xi_0$

$$\hat{h}_{t,r}(\Xi_0, v) = -\frac{1}{\hat{\lambda}_{t,r}} \log(1 - \hat{H}_{u,t}(r)), \quad \|v\| = 1, \quad (4.6)$$

where  $\|u\| = 1, \langle u, v \rangle = 0$ , and  $\hat{\lambda}_t$  is a strongly consistent estimator of the intensity  $\lambda$ , say given by (4.2) for  $d = 2$ . Then an estimator  $\hat{\Xi}_{t,r}$  of  $\Xi_0$  can be constructed by the values of the support function (4.6)

$$\hat{\Xi}_{t,r} = \{x: \langle x, v \rangle \leq \hat{h}_{t,r}(\Xi_0, v) \text{ for all } v, \|v\| = 1\}. \quad (4.7)$$

It follows from (4.4) that the estimator (4.6) converges uniformly to  $h(\Xi_0, v)$  for all  $v$  belonging to the unit sphere. Since the Hausdorff distance  $\rho_H$  between convex compact sets is equal to the uniform distance between their support functions, we get

$$\sup_{\varepsilon \leq r \leq c} \rho_H(\hat{\Xi}_{t,r}, \Xi_0) \rightarrow 0 \text{ a.s. as } t \rightarrow \infty$$

for all  $\varepsilon$  and  $c$  such that  $0 < \varepsilon < c$ .

#### 4.2. Star-Shaped Averages

This method is based on formula (3.5) for planar Boolean models. Consider the same family of expanding planar windows as in the previous subsections. Let  $Z^2$  be a lattice (grid) in  $\mathbf{R}^2$ . For each  $x \in (Z^2 \cap W_t) \setminus \Xi$  define  $S_x(\Xi)$  to be the star (3.4), constructed by an observation  $\Xi$  of the Boolean model. Each set  $S_x(\Xi)$  is described by its radius-vector function

$$r_x(u) = \sup \{r : x + ru \in S_x\}.$$

Since the Boolean model (1.1) is ergodic, the average

$$\hat{r}_t(u) = \frac{1}{n_t} \sum_{x \in (Z^2 \cap W_t) \setminus \Xi} r_x(u), \quad (4.8)$$

for each  $u$  from the unit circle tends as  $t \rightarrow \infty$  to the radius vector function

$$\bar{r}(u) = \sup \{r : ru \in E_r(S_x(\Xi) | x \notin \Xi)\}$$

of the conditional radius-vector expectation  $E_r(S_x(\Xi) | x \notin \Xi)$ . Here  $n_t$  is the cardinality of the set  $(Z^2 \cap W_t) \setminus \Xi$ .

The "continuous" version of (4.8) is given by

$$\frac{1}{\mu_2(W_t \setminus \Xi)} \int_{W_t \setminus \Xi} r_x(u) dx.$$

**Theorem 1.** *Suppose that  $\Xi_0$  contains a certain neighborhood of the origin. Then the estimator (4.8) tends to  $\bar{r}(u)$  uniformly for all  $u$  belonging to the unit circle.*

*Proof.* Suppose that  $|\hat{r}_t(u_t) - \bar{r}(u_t)|$  does not converge to zero as  $t \rightarrow \infty$  for a certain sequence of points  $u_t$ . Without loss of generality suppose that  $u_t \rightarrow u_0$  as  $t \rightarrow \infty$ . Then, for any neighborhood  $U_0$  of  $u_0$  and sufficiently large  $t$ ,

$$\hat{r}_t(u_t) - \bar{r}(u_0) \leq \frac{1}{n_t} \sum_{x \in (Z^2 \cap W_t) \setminus \Xi} \sup_{u \in U_0} r_x(u) - \bar{r}(u_0).$$

The right-hand side converges to  $E \sup_{u \in U_0} r_x(u) - \bar{r}(u_0)$ . It follows the monotone convergence theorem that this value converges to zero as  $U_0 \downarrow \{u_0\}$ . A similar inequality from below and the continuity property of  $\bar{r}(u)$  finish the proof. ■

**Remark.** The same result is valid for general dimensions, but then radius-vector expectations have no interpretation in terms of mean difference bodies.



It follows from (3.5) that the support function of the estimator  $\hat{\Xi}_t$  of the mean difference body given by

$$h(\hat{\Xi}_t, v) = (\hat{\lambda}_t(u))^{-1}, \quad \langle u, v \rangle = 0, \quad \|u\| = \|v\| = 1, \quad (4.9)$$

is a uniformly strong consistent estimator of the support function of the mean difference body, whence the corresponding set-valued estimator of  $\Xi_0$  is strongly consistent in the Hausdorff metric as  $t \rightarrow \infty$ .

#### 4.3. Intensities of Intersections

This method works for general  $d$  and is based on the estimator for intensities of  $(d - 1)$ -dimensional sections. Namely, replacing  $\lambda(u)$  and  $\lambda$  in (3.6) by their estimators  $\hat{\lambda}_t(u)$  and  $\hat{\lambda}_t$  leads to an estimator of the support function of the mean body  $\Xi_0$ .

Let  $u_1, \dots, u_n$  be an  $\varepsilon$ -net on the unit sphere. This means that each point of the unit sphere lies in the  $\varepsilon$ -neighborhood of a member of this  $\varepsilon$ -net. It was proven in Rataj and Saxl (1989) that the polyhedron

$$\Xi_0(\varepsilon) = \{x: \langle x, u_k \rangle \leq h(\Xi_0, u_k), 1 \leq k \leq n\}$$

tends to  $\Xi_0$  in the Hausdorff metric as  $\varepsilon \downarrow 0$ . On the other hand, the finiteness of the  $\varepsilon$ -net yields

$$\rho_H(\hat{\Xi}_t(\varepsilon), \Xi_0(\varepsilon)) \rightarrow 0 \quad \text{a.s. as } t \rightarrow \infty$$

for the estimator

$$\hat{\Xi}_t(\varepsilon) = \{x: \langle x, u_k \rangle \leq \hat{\lambda}_t(u_k) / \hat{\lambda}_t, 1 \leq k \leq n\}.$$

Thus, for small  $\varepsilon$  the set  $\hat{\Xi}_t(\varepsilon)$  estimates the mean difference body  $\Xi_0$ .

#### 4.4. Fibre Process Method

This method is a particular case (for  $d = 2$ ) of the method described in the previous subsection. In this case  $\lambda(u)$  can be estimated from (3.7) by an estimator  $\hat{P}_{L,t}(u)$  of the rose of intersections. The latter is computed in the usual way for fibre process  $\partial\Xi$ , see Stoyan *et al.* (1987, p.245) and Molchanov and Stoyan (1994a).

It follows from (3.6), (3.7) and (4.2) that the corresponding estimator  $\hat{\Xi}_t$  of the mean difference body has the support function given by

$$h(\hat{\Xi}_t, u) = \frac{\hat{P}_{L,t}(v)}{2\hat{N}_{\lambda,t}^+}, \quad \|u\| = \|v\| = 1, \quad u \perp v. \quad (4.10)$$

### 5. SIMULATION EXPERIMENTS

Let us consider the planar Boolean model  $\Xi$  with the typical grain  $\xi M$ , where  $\xi$  is uniformly distributed on  $[0,1]$ , and  $M$  is a centrally symmetric compact convex set. In this case the mean difference body  $\Xi_0$  is equal to  $2M\mathbf{E}\xi$ , i.e.  $\Xi_0 = M$ . This makes the comparison of empirical and theoretical results especially simple.



Figure 1 A realization of the Boolean model  $\Xi$ .

Let  $M$  be the ellipse with axes 120 and 32 pixels on the digitized computer screen, and  $\lambda = 0.001$ . A realization of  $\Xi$  within the window  $250 \times 200$  pixels is shown in Figure 1 (for the later evaluations larger realizations within the window  $640 \times 350$  pixels were used). The intensity  $\lambda$  was estimated by counting of tangent points as it was described in Section 4.0.

First, we consider estimates of the mean body by means of the fibre processes method and formula (4.10). The estimator  $\hat{P}_{L, \cdot}(\cdot)$  of the rose of intersections of the fibre process  $\partial\Xi$  was determined by counting the number of intersections per unit length between  $\partial\Xi$  and a rotating system of 80 parallel lines. Figure 2a shows an estimate of the mean body obtained in this way. Such set-valued estimates  $\hat{\Xi}_1, \dots, \hat{\Xi}_{100}$  were computed for 100 independent samples of the Boolean model  $\Xi$ . These estimates are considered to be realizations of the corresponding set-valued estimator  $\hat{\Xi}_r$ . The corresponding mean value (Minkowski average)

$$\hat{\Xi} = \frac{1}{100}(\hat{\Xi}_1 \oplus \dots \oplus \hat{\Xi}_{100})$$

is shown in Figure 2b, and some numerical characteristics of the sample of these set-valued estimates, including the variance, are given in Table 1. This Table contains several numerical characteristics of the sample of estimates, which can serve as possible error measures for the estimator  $\hat{\Xi}_r$  of the mean difference body  $\Xi_0$ .

1. The maximum Hausdorff distance between  $\Xi_0$  and its estimates. It characterizes the "worst" behavior of the estimator.
2. The Hausdorff distance between  $\Xi_0$  and the average value  $\hat{\Xi}$  of its estimates. This distance characterizes the bias  $\rho_H(\Xi_0, \mathbf{E}\hat{\Xi}_r)$  of the corresponding set-valued estimator.
3. Maximum of the variance of the estimates of support functions over the unit sphere.

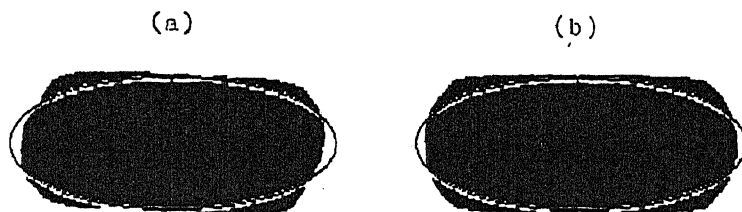


Figure 2 An estimate of the mean body by the fibre processes method (a) and the mean value of 100 independent estimates (b) in comparison with the exact shape of the mean body (two-times enlarged).

Table 1 Numerical characteristics of four series of  $n = 100$  estimates of mean bodies

	Fibre processes method	Star-shaped averages	Corrected star-shaped averages	Linear contact distribution function
$\max_{1 \leq i \leq n} \rho_H(\Xi_0, \hat{\Xi}_i)$	22.9	37.9	26.4	24.6
$\rho_H(\Xi_0, \hat{\Xi})$	5.8	12.2	4.8	5.0
$\sup_{\ u\ =1} [n^{-1} \sum_{i=1}^n h(\hat{\Xi}_i, u)^2 - h(\Xi_0, u)^2]$	49.6	86.5	67.6	47.4
$\hat{m} = n^{-1} \sum_{i=1}^n \rho_H(\hat{\Xi}_0, \hat{\Xi}_i)$	12.5	17.0	11.4	8.8
$n^{-1} \sum_{i=1}^n \rho_H^2(\hat{\Xi}_0, \hat{\Xi}_i) - \hat{m}^2$	8.5	62.9	21.6	14.2

$$\sup_{\|u\|=1} \left[ \frac{1}{n} \sum_{i=1}^{100} h(\hat{\Xi}_i, u)^2 - h(\Xi_0, u)^2 \right],$$

characterizes the variance of the estimator.

- The average Hausdorff distance between  $\Xi_0$  and  $\hat{\Xi}_i, 1 \leq i \leq 100$ , represents the mean distance  $E_{\rho_H}(\Xi_0, \hat{\Xi}_i)$ . It is an analog of the first absolute central moment for random variables.
- The empirical variance of the sample  $\rho_H(\Xi_0, \hat{\Xi}_i), 1 \leq i \leq 100$ , approximates the variance  $D\rho_H(\Xi_0, \hat{\Xi}_i)$ .

Then the star-shaped averages method was applied using a grid of 2000 vertices. Figure 3a shows an estimate of the mean body  $\Xi_0$  obtained in this way; the corresponding number of summands  $n_i$  in (4.8) was equal to 705. Figure 3b shows the averaged estimate for the same setting as Figure 2b, see also Table 1.

The figures and the data of Table 1 confirm that the fibre processes method gives better results. Namely, the star-shaped averages method has a greater bias. Also other error measures including the variance are the largest. The bias can be explained by systematic errors, caused by underestimation of the general intensity due to the tangent points method. This underestimation appears due to digitalization of pictures, i.e. the image analyzer cannot distinguish between near points belonging to different grains. By using the fibre processes or intersection methods the support function of the mean body is estimated as the quotient of the intensity of

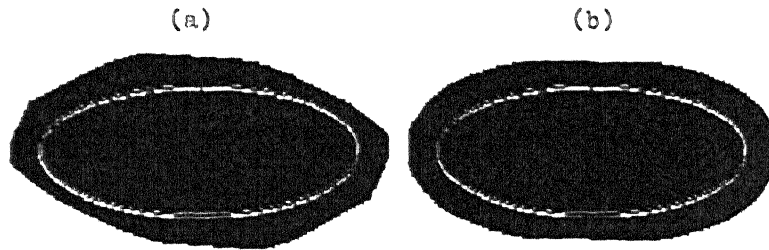


Figure 3 The same as in Figure 2 for the star-shaped averages method.

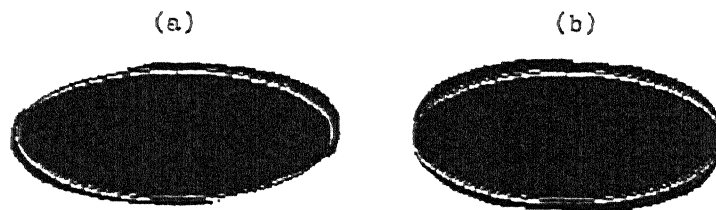


Figure 4 The same as in Figure 3, when the exact value of the intensity is known.

intersections and the general intensity. Both are estimated by the same technique. Thus the errors in the numerator and the denominator may compensate each other. For the star-shaped averages method there is not such a possibility. Moreover, edge effects are essential, especially for Boolean models with small area fractions, since a lot of stars are extended beyond the window of observations.

Figure 4 shows a new estimate obtained for the same data as this in Figure 3a by computing of star-shaped averages using the formula

$$\hat{h}(\hat{\Sigma}, v) = (\lambda \hat{\rho}_i(u))^{-1}, \quad \|u\| = \|v\| = 1, \quad u \perp v.$$

In comparison with (4.9) here the exact value (which in practical applications is unknown) of the intensity has been used. In this case the errors are smaller, see also Table 1.

Computer simulations made for other planar Boolean models with different area fractions show that the star-shaped averages behave better for high area fractions ( $p > 0.9$ ). Otherwise the fibre processes method gives better results.

Table 2 presents simulation results for Boolean models with the same grain as above and different intensities. The results were computed by the fibre process method for 60 independent samples of each Boolean model. These simulations show that the fibre processes methods works better for Boolean models with middle area fractions (around 0.5). Indeed, if  $p$  is small, then we have very few intersections with lines. The same effect appears if the intensity is getting larger but the grain is fixed.

The method based on the evaluation of the linear contact distribution function and the application of (4.6), (4.7) with  $r = 10$  was applied for the same setting as above. The results of Table 1 show that this method has the quality closed to the fibre processes method. However it requires time-consuming evaluations and is

**Table 2** Numerical characteristics of  $n = 60$  estimates of mean bodies by the fibre processes method for different intensities.

intensity, $\lambda, (\times 10^{-3})$	0.35	0.52	0.7	1.04	1.55	2.1
area fraction, $p$	0.290	0.392	0.490	0.659	0.794	0.903
$\max_{1 \leq i \leq n} \rho_H(\Xi_0, \Xi_i)$	26.2	24.5	25.1	22.9	24.9	50.9
$\rho_H(\Xi_0, \Xi)$	7.80	5.96	5.06	6.06	7.6	8.90
$\sup_{ u =1} [n^{-1} \sum_{i=1}^n h(\Xi_i, u)^2 - h(\Xi_0, u)^2]$	78.1	51.4	62.8	48.0	64.5	160
$\hat{m} = n^{-1} \sum_{i=1}^n \rho_H(\Xi_0, \Xi_i)$	16.7	13.5	12.5	12.5	14.7	19.3
$n^{-1} \sum_{i=1}^n \rho_H^2(\Xi_0, \Xi_i) - \hat{m}^2$	17.4	16.0	11.9	8.6	11.2	58.5

more sensible to the discretization of the image (especially, for small  $r$ ). The linear contact distribution function method has been already discussed in Molchanov and Stoyan (1994a) in the relation to the study of the microstructure of paper.

#### Acknowledgement

I am grateful to Dietrich Stoyan for helpful discussions and reading of the manuscript. The main idea of this paper originated in one of his lectures on stochastic geometry for students of the Freiberg University of Mining and Technology. The remarks and suggestions of the referees helped me a lot to make a thorough revision and improve the style and readability of the paper. This work was supported by the Alexander von Humboldt-Stiftung (Bonn) and The Netherlands Organization for Scientific Research (NWO).

#### References

- Cressie, N. A. C. and Laslett, G. M. (1987) Random set theory and problems of modeling. *SIAM Review*, 29, 557–574.
- Cressie, N. A. C. (1991) *Statistics for Spatial Data*. Wiley, New York etc.
- Hall, P. (1988) *Introduction to the Theory of Coverage Processes*. Wiley, New York etc.
- Heinrich, L. (1992) On existence and mixing properties of germ-grain models. *Statistics*, 23, 271–286.
- Heinrich, L. (1993) Asymptotic properties of minimum contrast estimators for parameters of Boolean models. *Metrika*, 31, 349–360.
- Lantuejoul, C. and Schmitt, M. (1991) Use of two new formulae to estimate the Poisson intensity of a Boolean model. *Treizième Colloque GRETSI – Juan-Les-Pins du 16 au 20 Septembre 1991*, 1045–1048.
- Leichtweiss, K. (1980) *Konvexe Mengen*. VEB Deutscher Verlag der Wissenschaften, Berlin.
- Mase, S. (1982) Asymptotic properties of stereological estimators of volume fraction for stationary random sets. *J. Appl. Prob.*, 19, 111–126.
- Matheron, G. (1975) *Random Sets and Integral Geometry*. Wiley, New York.
- Mecke, J. and Stoyan, D. (1980) Formulas for stationary planar fibre processes I – general theory. *Math. Operationsf. Statist., ser. statist.*, 12, 267–279.
- Molchanov, I. S. (1988) Convergence of the empirical accompanying functionals of stationary random sets. *Theory Probab. Math. Statist.*, 39, 107–109.
- Molchanov, I. S. (1990) Estimation of the size distribution of spherical grains in the Boolean model. *Biometrical J.*, 32, 877–886.
- Molchanov, I. S. (1991) A consistent estimate of parameters of Boolean models of random closed sets. *Theory Probab. Appl.*, 36, 600–607.
- Molchanov, I. S. (1994) On statistical analysis of Boolean models with non-random grains. *Scand. J. Statist.*, 21, 73–82.

- Molchanov, I. S. (1995) Statistics of the Boolean model: From the estimation of means to the estimation of distributions. *Adv. in Appl. Probab.*, **27**, 63–86.
- Molchanov, I. S. and Stoyan, D. (1994a) Directional analysis of fibre processes related to Boolean models. *Metrika*, **41**, 183–199.
- Molchanov, I. S. and Stoyan, D. (1994b) Asymptotic properties of estimators for parameters of the Boolean model. *Adv. in Appl. Probab.*, **26**, 301–323.
- Rataj, J. and Saxl, I. (1989) Analysis of planar anisotropy by means of the Steiner compact. *J. Appl. Probab.*, **26**, 490–502.
- Rataj, J. and Saxl, I. (1992) Estimation of direction distribution of a planar fibre system. *Acta Stereologica*, **11/Suppl. I**, 631–636.
- Ripley, B. D. (1986) Statistics, images and pattern recognition. *Canadian J. Statist.*, **14**, 83–102.
- Schmitt, M. (1991) Estimation of the density in a stationary Boolean model. *J. Appl. Probab.* **28**, 702–708.
- Schröder, M. (1992) *Schätzer für Boolesche Modelle im  $R^2$  und  $R^3$* . Universität Karlsruhe, Diplomarbeit.
- Serra, J. (1982) *Image Analysis and Mathematical Morphology*. Academic Press, London.
- Stoyan, D., Kendall, W. S. and Mecke, J. (1987) *Stochastic Geometry and its Applications*. Akademie-Verlag, Berlin; Wiley, Chichester.
- Stoyan, D. and Stoyan, H. (1994) *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics*, Wiley, Chichester etc.
- Vitale, R. (1988) An alternate formulation of mean value for random geometric figures. *J. Microscopy* **151**, 197–204.
- Weil, W. (1988) Expectation formulas and isoperimetric properties for non-isotropic Boolean models. *J. Microscopy*, **151**, 235–245.
- Weil, W. (1993) The determination of shape and mean shape from sections and projections. *Acta Stereologica*, **12/2**, 73–84.
- Weil, W. (1994) Support functions on the convex ring in the plane and support densities for random sets and point processes. *Rend. Circ. Mat. Palermo*, **35**, 323–344.
- Weil, W. (1995) The estimation of mean shape and mean particle number in overlapping particle systems in the plane. *Adv. in Appl. Probab.*, **27**, 102–119.
- Yadin, M. and Zacks, S. (1985) The visibility of stationary and moving targets in the plane subject to a Poisson field of shadowing elements. *J. Appl. Prob.*, **22**, 776–786.

I. S. Molchanov,  
 Centrum voor Wiskunde en Informatica (CWI),  
 P. O. Box 94079,  
 1090 GB Amsterdam  
 The Netherlands  
 from 1/Dec/95  
 University of Glasgow  
 Department of Statistics  
 Glasgow G12 8QW  
 UK