TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

USE OF TWO NEW FORMULAE TO ESTIMATE THE POISSON INTENSITY OF A BOOLEAN MODEL

Christian LANTUEJOUL*, Michel SCHMITT**

* **Centre de Géostatistique**, 35 rue Saint-Honoré, 77305 Fontainebleau, France
** **Thomson CSF, L.C.R.**, Domaine de Corbeville, 91404 Orsay Cedex, France

Résumé. Cet article porte sur l'estimation de la densité Poissonnienne d'un modèle Booléen. Trois méthodes sont passées en revue: la méthode classique de Steiner qui suppose les grains primaires convexes, et deux nouvelles méthodes reposant sur deux formules récentes valides pour des grains primaires convexes (Weil) ou bornés (Schmitt). On compare les performances de ces trois méthodes, et l'on regarde ce qui se passe lorsque celles-ci sont utilisées au dela de leur seuil de validité.

Abstract. This paper deals with the estimation of the Poisson intensity of a Boolean model. Three different methods are reviewed: the classical Steiner method which holds for convex primary grains, and two new methods based on recent formulae which require the primary grains to be convex (Weil) or bounded (Schmitt). We compare the performance of these three methods, and we also consider what happens if these methods are used beyond their range of applicability.

1. INTRODUCTION

Over the last few years, there has been a surge of interest in the modelization of textures using Boolean models and their variations. Examples can be found in many areas such as mining industry [1], oil industry [2,3,4], geology [5], pedology [6], forestry [7], metallurgy [8,9,10], physics [11], medecine [12,13], to name some of them. The reason for the interest in such a random set model is at least threefold. Firstly, the construction of a Boolean model is genetic. It expresses the idea of a population of objects located at random. Secondly, the Boolean model possesses strong stability properties [14,15]. The union of two independent Boolean models is Boolean. A cross-section of a Boolean model is Boolean. Thirdly, explicit calculations can be made for the Boolean model, and in particular, a formula for its distribution is available.

A Boolean model is the union of independent, equally distributed, compact random sets (called *primary grains*) implanted at the constitutive points (called *germs* of a Poisson point process).

Modelling textures using a Boolean model unavoidably raises the problem of the statistical inference of its parameters. Such a problem can be addressed in two different stages. Are the experimental data compatible with a Boolean model? If so, how are the parameters estimated? In preliminary work [16], it has been shown that this second question makes sense, in that given the distribution of a Boolean model, there exists a unique value for the Poisson intensity and a unique distribution for the shape of primary grains that give the Boolean distribution. In this paper, we shall be concerned with the estimation of the Poisson intensity. Various methods can be found in the literature, such as the method of moments [17], the maximum likelihood method [18], the contact distribution function [19]... In what follows, we are going to compare the most used method [20], referred to as "Steiner's method", with two new methods derived from formulae which have recently appeared in the literature [16,21], called here "Weil's method" and "Schmitt's method" for ease of reference. We shall compare the performance of these three methods in the Boolean case. We shall also see what happens when the underlying model departs from the Boolean one.

For the rest of the paper, we shall adopt the following notation:
$\check{A} = \{-x, x \in A\}$ the reflection of the set A in the origin.
$A \oplus B = \{x+y, x \in A, y \in B\}$ the Minkowsky sum of the two sets A and B.

2. THE THREE METHODS

In [15], it has been shown that the distribution of a closed random set X is totally characterized by the mapping Q which associates to any compact set K the probability $Q(K)$ that K is disjoint from X

$$Q(K) = P\{K \cap X = \emptyset\}$$

In the case where X is a two-dimensional Boolean model of Poisson intensity θ and primary grain A, $Q(K)$ is given by the formula

$$Q(K) = e^{-\theta\, \overline{a}(A \oplus \check{K})}$$

where $\overline{a}(A \oplus \check{K})$ denotes the mean area of the set $A \oplus \check{K}$, usually termed A *dilated* by K.

2.1 Steiner's method

If A is *convex*, it is possible to derive an explicit formula for A dilated by K depending upon the choice of K. For instance, if $K = K_r$ is a disk of radius r, then

$$a(A \oplus \check{K}_r) = a(A) + rp(A) + \pi r^2$$

which involves the area $a(A)$ and the perimeter $p(A)$ of the primary grain A. If $K = K_r$ is a line segment of length r, then

$$a(A \oplus \check{K}_r) = a(A) + rd(A)$$

which also involves Feret diameter $d(A)$ of A in the direction of K. Steiner's method rests on such formulae. For instance, if K is circular, the experimental procedure is the following one:
i) Computation of $Q(K_r)$ for several r values starting from the experimental data.
ii) Since $\ln Q(K_r)$ is a polynomial of degree 2 in r, the corresponding experimental curve is fitted to a quadratic polynomial.
iii) The estimated value for θ is the quadratic coefficient of the fitted polynomial up to the factor π.

2.2 Weil's method

In principle, the probability $Q(K)$ would be known with perfect accuracy if a realization of the Boolean model was avail-

able in the whole space. In practice however, this is never the case. What is usually experimentally accessible is a realization of the model within a bounded field, say D. Let us again assume that the primary grain is *convex*. Then, the Boolean model within D is a finite union of convex sets. It is known in stereology [22] that any measurement which is additive

$$m(A \cup B) + m(A \cap B) = m(A) + m(B),$$

continuous and invariant under planar displacements is necessarily a linear combination of three basic measurements, namely the *area*, the *perimeter* and the *Euler-Poincaré characteristic*, which is the number of connected components minus the number of holes within them. It turns out that Weil and Wieacker [21,23] have obtained the explicit formula for the mean of these three measurements in the Boolean case

$$\begin{aligned}
\overline{a}(X \cap D) &= a(D)\left(1 - e^{-\theta\,\overline{a}(A)}\right) \\
\overline{p}(X \cap D) &= a(D)\, e^{-\theta\,\overline{a}(A)}\theta\,\overline{p}(A) + p(D)\left(1 - e^{-\theta\,\overline{a}(A)}\right) \\
\overline{k}(X \cap D) &= a(D)\, e^{-\theta\,\overline{a}(A)}\left(\theta - \theta^2\,\overline{a}(A, \check{A})\right) + \\
&\quad 2\, e^{-\theta\,\overline{a}(A)}\theta\,\overline{a}(D, \check{A}) + 1 - e^{-\theta\,\overline{a}(A)}
\end{aligned}$$

where the quantities $a(A, B)$ stand for $\frac{1}{2}a(A \oplus B) - a(A) - a(B)$. Note that an explicit formula can be obtained for $a(A, B)$ in only a limited number of cases that we are not going to detail here. Let us simply mention that $a(A, \check{A}) = a(A)$ if A is symmetric around the origin.

Starting from these formulae, a procedure can be proposed to estimate the Poisson intensity of the Boolean model:
i) Experimental measurement of $\overline{a}(X \cap D)$. Using the first formula, one gets an estimate for $\theta\ a(A)$.
ii) Experimental measurement of $\overline{p}(X \cap D)$. By replacing $\theta\ a(A)$, by its estimate, the second formula yields an estimate for $\theta\ p(A)$.
iii) Experimental measurement of $\overline{k}(X \cap D)$. By replacing $\theta\ a(A)$ and $\theta\ p(A)$ by their estimates, the third formula can be applied to give an estimate for θ.

2.3 Schmitt's method

Here, the primary grains are assumed to be *bounded*. In particular, they are not supposed to be convex, nor even connected. All of them are contained in a disk of a fixed diameter, say λ. Under this assumption, it is possible to associate to each primary grain implanted at a Poisson point a marker that is the bottom left vertex of the smallest rectangle enclosing the primary grain, with sides parallel to the coordinate axes (see Figure 1).

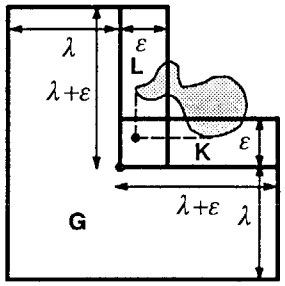

Figure 1: Construction of Schmitt's estimator

Now, it can be observed that a marker of a grain is located within a square of side ε if and only if the grain hits the domains K and L, but is disjoint from G. Starting from this remark, it is possible to derive an exact formula for θ [16], namely

$$\theta = \frac{1}{\varepsilon^2}\ \ln\ \frac{Q(G)\,Q(G \cup K \cup L)}{Q(G \cup K)\,Q(G \cup L)}$$

Insofar as the four Q quantities are experimentally accessible, this formula can be used to estimate θ.

3. COMPARISON OF THE THREE METHODS ON A BOOLEAN MODEL

In order to compare these three methods, We have considered a Boolean model of intensity 0.1, and whose primary grain is a square of side 2. 100 independent simulations have been carried out in a square field of side 50. Figure 2 shows an example of such a simulation.

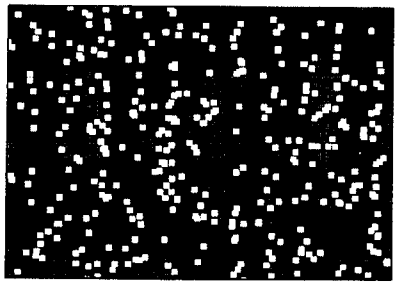

Figure 2: Realization of a Boolean model with square primary grains

Steiner's method has been applied by taking squares of side $0, 0.6, 1.2, 1.8, 2.4$ for K. Schmitt's method has been used with $\lambda = 2.$ and $\varepsilon = 0.5$. In order to facilitate the comparison between the three estimates, the three scattergrams have been plotted.

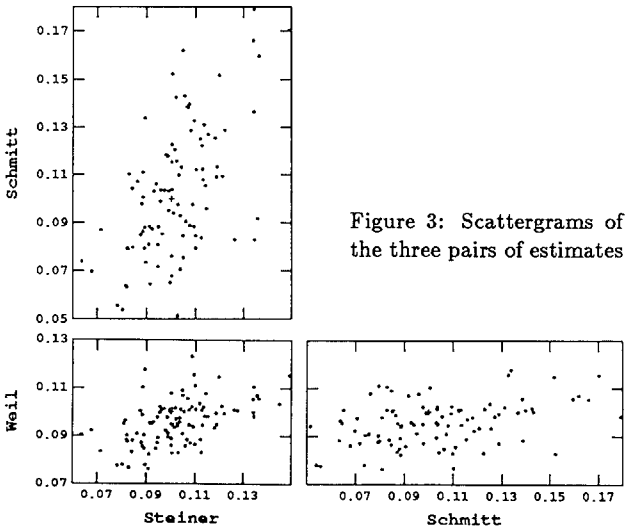

Figure 3: Scattergrams of the three pairs of estimates

These scattergrams suggest several comments. Firstly, the fact that the three estimators are not unbiased, but only asymptotically unbiased, seems to be of minor importance in practice. Secondly, Schmitt's estimator appears to be more dispersed than the other two, especially that of Weil. In order to understand this difference in dispersion, we have carried out another estimation using Steiner's method, with larger sides for K $(0., 1., 2., 3., 4.,)$. This has resulted in Steiner's estimator being more dispersed than Schmitt's one. So the dispersion of an estimator appears to be directly related to the required domain to perform the estimation. Things become clearer if we plot the standard deviation of the area of the part of the field not occupied by the grains versus the side of the square primary grains (cf. Figure 4), after noticing that the quantity $Q(K) = \exp\{-\theta\ \overline{a}(A \oplus \check{K})\}$ can be interpreted as the probability that a point is not covered by the grains of a Boolean model of intensity θ and primary grain $A \oplus \check{K}$. In the case of a small-sized primary grain, the part of the field not occupied by the grains is almost the whole field itself and its area has a negligible standard deviation. In the case of a large-sized primary grain, all of the points of the field are likely to be covered by a grain, and the standard deviation is again negligible. On the contrary, in the case of a primary

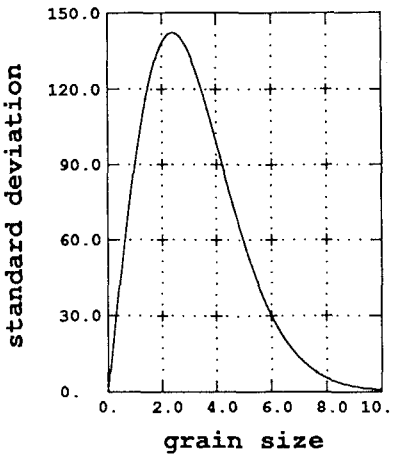

Figure 4: Plot of the standard deviation of the area of the part of the field not occupied by the grains versus the size of the primary grains

grain with an intermediate size, the part of the field which is not occupied by the grains has a highly variable area, and its standard deviation can be very large. This justifies the use of Steiner's procedure using small-sized K's as in [24]. In conclusion, Weil's estimator is the least dispersed for it requires the smallest domain.

4. ON THE VALIDITY RANGE OF THE METHODS

In each of the estimation methods, the primary grains must fulfill some requirements (*Convexity* for Steiner's and Weil's methods, *boundedness* for Schmitt's method). What happens when the methods are not applied according their specific validity domain? To answer such a question, we have considered a Boolean model in a square field of side 50. The Poisson intensity is 0.1, and the primary grains are made up of two points, the second point being uniform over the boundary of a square of side $2r$ centered at the first point. 100 independent simulations have been carried out for various r values. The results are displayed in Figure 5.

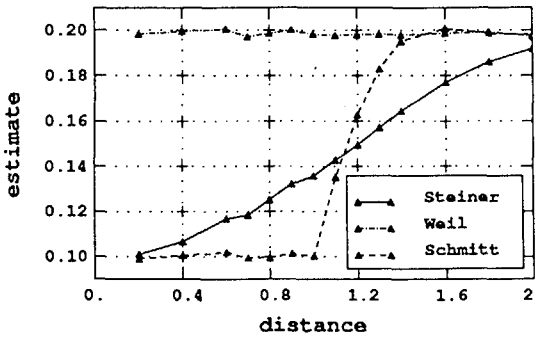

Figure 5: Three different estimates of the Poisson intensity versus the size of the primary grain. The actual Poisson intensity is 0.1.

Since the primary grain is not convex, Weil's method is not suitable. What it gives is not the Poisson intensity, but simply the number of points per unit area. Steiner's method gives a slightly more subtle result. In the case where the two points of a primary grain are close together, the difference between the dilation of the two points and the dilation of the segment between them is not big. As a consequence, Steiner's method gives a reasonable result provided that the two points are rather close. As both points become more distant, the estimate increases to the number of points per unit area. Schmitt's method is even more striking. When the distance between the two points is less than the critical size $\lambda = 1$, we get exactly the expected Poisson intensity. But as soon as the distance becomes larger than λ, the estimate suddenly departs from the Poisson intensity and rapidly increases to give the number of points per unit area.
It is noteworthy that for large distances the three methods give the same result (twice the actual Poisson intensity). Although being conceptually different, the three methods may be self-validating, in the sense that they give consistent but erroneous estimates. This suggests that there may also be some validation problems in the case where two conceptually identical methods are used to perform the estimation. For example, in Figure 6 we have represented the scattergram of the estimates obtained from 100 simulations using the linear and the square Steiner's methods. Here, the primary grains have been taken to be two squares of edge 0.2 located as above with $r = 3$. The scattergram suggests that the Poisson intensity is around 0.16, whereas its actual value is 0.1.

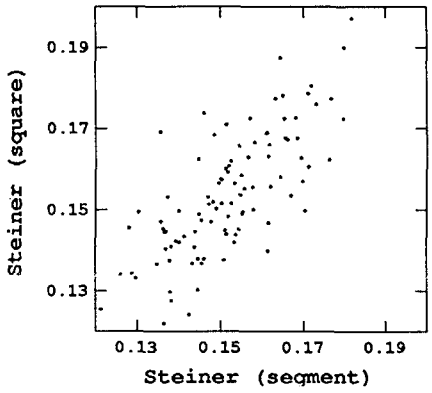

Figure 6: Scattergram of two estimates of the Poisson intensity. The primary grain is not convex, so the methods are not suitable. Both methods suggest that the Poisson intensity is around 0.16, whereas its actual value is 0.1.

5. DEPARTING FROM THE BOOLEAN MODEL

The question addressed in this paragraph is: can the three methods be used to test the compatibility of a random set with a Boolean model? To answer this question would require an extensive and probably unprofitable study. In what follows, we restrict ourselves to considering some random sets selected among the family of non Poisson point processes (see Figures 7-2 to 7-4).

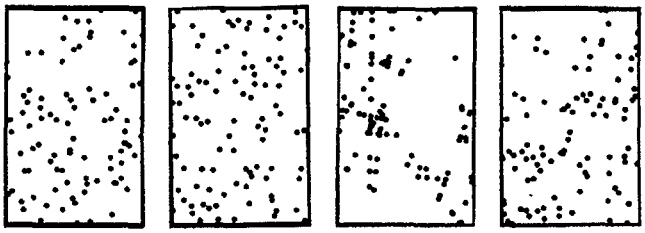

Figure 7: Some examples of point processes. From left to right, Poisson, thinned Poisson, intersection of Poisson lines, Cox process

A typical example is given by the intersections of isotropic Poisson lines [15,17] (see Figure 7-3). The Poisson line intensity μ has been chosen so as to have a number of points per unit area equal to $\theta = \pi\mu^2 = 0.1$. 10 simulations have been carried out in a square field of edge 50. We have applied

Schmitt's method with $\varepsilon = 0.5$ and with values of λ ranging from 0.1 to 1. The experimental results, displayed in Figure 8, show that the estimate is not independent of λ. The assumption of a Boolean model with primary grains of maximal size 1 has to be reconsidered in this case.

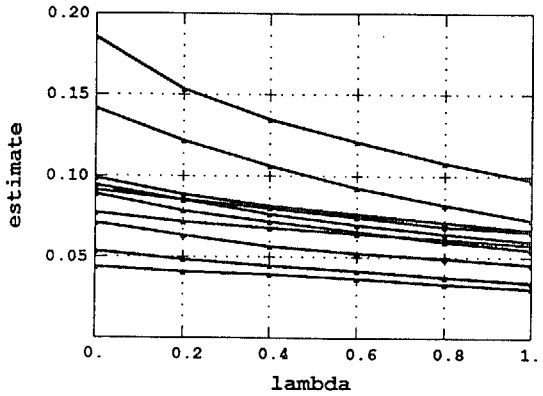

Figure 8: Example of Poisson line intersections. The Schmitt estimate computed from 10 simulations depends on λ. In this case, the assumption of a Boolean model with primary grains of maximal size 1 is not confirmed.

A similar reasoning holds using Steiner's and Weil's methods. Thus the assumption of a Boolean model with convex primary grains is invalidated.

However, there are some cases where these tests are clearly insufficient. Consider for example, a Cox process, that is an inhomogeneous Poisson point process with a stochastic intensity [25]. If the intensity fluctuations are small at the observation scale (i.e. in the field where experimental data are available), a realization of a Cox process does not significantly differ from the one of a Poisson point process. In such a case, the three methods do not invalidate the assumption of a Boolean model. The estimated intensity is just the local Poisson intensity of the Cox process. Note however that if the three methods were applied on another simulation of the Cox process, we should get another estimate with a different value.

To conclude this paragraph, let us simply say that the three methods provide partial tests to validate or to invalidate the assumption of Boolean model.

6. CONCLUSIONS

In this paper, we have presented three methods to estimate the Poisson intensity of a Boolean model: an old one (Steiner) and two new ones (Schmitt and Weil). Based on different assumptions on the primary grains (convexity for Steiner and Weil, boundedness for Schmitt), these three methods have different ranges of applicability. In the case of a Boolean model with convex and bounded primary grains, Weil's method seems to be the most accurate.

When the methods are applied beyond their range of applicability (Boolean model with non convex or non bounded primary grains, or non Boolean model), they may self-validate. This makes it necessary, if not crucial, to design statistical tests to check the compatibility of random sets to Boolean models.

This paper is incomplete in many respects. The influences of the size of the field, of the grain proportion and of the shape of the primary grains to the precision of the estimation have not been investigated. The use of the methods beyond their range of validity would also deserve further examination.

[1] Chautru J.M. (1989). "The use of Boolean random functions in geostatistics". In "Geostatistics", Ed. M. Armstrong, Kluwer Acad. Pub., Dordrecht, pp. 201-212.
[2] Conrad F. and Jacquin Ch. (1973). "Représentation d'un réseau bidimensionnel de fractures par un modèle probabiliste". Revue de l'I.F.P., Vol. 28-6, pp. 843-890.
[3] Haldorsen H.H. and MacDonald C.J. (1987). "Stochastic modelling of underground reservoir facies (SMURF)". Society of Petroleum Engineers, Paper 16751.
[4] Dubrule O. (1989). "A review of stochastic models for petroleum reservoirs". In "Geostatistics", Ed. M. Armstrong, Kluwer Acad. Pub., Dordrecht, pp. 493-506.
[5] Chilès J.P. (1989)."Modélisation géostatistique d'un réseau de fractures". In "Geostatistics", Ed. M. Armstrong, Kluwer Acad. Pub., Dordrecht, pp. 57-76.
[6] Chadoeuf J., Monestiez P., Bertuzzi P. and Stengel P. (1989). Parameter estimation in a Boolean model of rough surface: application to soil surfaces". Acta Stereologica, Vol. 8-2, pp. 635-640.
[7] Matern B. (1988). "Spatial variations". Springer-Verlag, Berlin.
[8] Greco A., Jeulin D. and Serra J. (1979). "The use of the texture analyser to study sinter structure". J. of Microscopy, Vol. 116-2, pp. 199-211.
[9] Jeulin D. and Jeulin P. (1981). "Synthesis of rough surfaces by random morphological functions". Acta Stereologica, Vol 3-1, pp. 239-246.
[10] Jeulin D. (1991). "Modèles morphologiques de structures aléatoires et de changement d'échelle". Thèse de Doctorat ès Sciences Physiques, Caen.
[11] Sonntag U., Stoyan D. and Hermann H. (1981). "Random set models in the interpretation of small-angle scattering data". Phys. Stat. Sol. (a), Vol. 68, pp. 281-288.
[12] Préteux F. (1987). "Description et interprétation des images par la morphologie mathématique. Application à l'imagerie médicale". Thèse de Doctorat ès Sciences Mathématiques, Paris 6.
[13] Préteux F. and Schmitt M. (1988). "Boolean Texture Analysis and Synthesis". In "Image Analysis and Mathematical Morphology. Vol. 2: Theroretical Advances" Ed. J. Serra, Academic Press, London, pp. 377-400.
[14] Matheron G. (1967). "Eléments pour une théorie des milieux poreux". Masson, Paris.
[15] Matheron G. (1975). "Randon sets and integral geometry". Wiley, New York.
[16] Schmitt M. (1991). "Estimation of the density in a stationary Boolean model". J. of Appl. Prob., Vol. 28, to appear.
[17] Santalo L.A. (1976). "Integral Geometry and Geometric Probability". Addison-Wesley, Reading (Mass.).
[18] Ayala G., Ferrándiz J. and Montes F. (1990)."Boolean models: maximum likelihood estimation from circular clumps" Biom. J., Vol. 32-1, pp. 73-78.
[19] Stoyan D., Kendall W. S. and Mecke J. (1987). "Stochastic geometry and its applications". Wiley, New York.
[20] Serra J. (1982). "Image Analysis and Mathematical Morphology". Academic Press, London.
[21] Weil W. (1988). "Non-isotropic Boolean models". J. of Microscopy, Vol. 151-3, pp. 235-246.
[22] Hadwiger H. (1957)."Vorselungen über Inhalt, Oberfläche und Isoperimetrie". Springer-Verlag, Berlin.
[23] Weil W. and Wieacker J.A. (1984). "Densities for stationary random sets and point processes". Adv. in Applied Prob., Vol 16-2, pp. 324-346.
[24] Ayala G., Ferrándiz J. and Montes F. (1989). "Two methods of estimation in Booleans models". Acta Stereologica, Vol. 8-2, pp. 629-634.
[25] Diggle P.J. (1983). "Statistical Analysis of spatial points patterns". Academic Press, London.