

## UNBIASED ESTIMATORS OF SPECIFIC CONNECTIVITY

JEAN-PAUL JERNOT<sup>1</sup>, PATRICIA JOUANNOT<sup>1</sup> AND CHRISTIAN LANTUÉJOUL<sup>2</sup>

<sup>1</sup>CRISMAT, Ensicaen/CNRS, 6 boulevard du maréchal Juin, 14050 Caen, France; <sup>2</sup>Ecole des Mines, 35 rue Saint-Honoré, 77305 Fontainebleau, France

e-mail: jean-paul.jernot@ensicaen.fr, patricia.jouannot@ensicaen.fr, christian.lantuejoul@ensmp.fr  
 (Accepted November 7, 2007)

### ABSTRACT

This paper deals with the estimation of the specific connectivity of a stationary random set in  $\mathbb{R}^d$ . It turns out that the “natural” estimator is only asymptotically unbiased. The example of a boolean model of hypercubes illustrates the amplitude of the bias produced when the measurement field is relatively small with respect to the range of the random set. For that reason unbiased estimators are desired. Such an estimator can be found in the literature in the case where the measurement field is a right parallelotope. In this paper, this estimator is extended to apply to measurement fields of various shapes, and to possess a smaller variance. Finally an example from quantitative metallography (specific connectivity of a population of sintered bronze particles) is given.

Keywords: Boolean model, edge effects, Euler-Poincaré characteristic, specific connectivity, unbiased estimation.

### INTRODUCTION

A few years ago, one of us was reviewing a paper in which an asymptotically unbiased estimator was used exactly as an unbiased one. As we made reservations about the way of handling this type of estimator, the authors reply “This estimator is UFAPP (unbiased for all practical purposes)”. Although humorous and witty, this reply should not hide the difficulty of using asymptotically unbiased estimators and interpreting the results they produce.

This paper is mainly devoted to the estimation of the specific connectivity of a stationary random set. It turns out that the “natural” estimator is only asymptotically unbiased. The example of the boolean model illustrates the importance of the bias produced when the measurement field is relatively small with respect to the range of the random set. In the case where this measurement field is a right parallelotope, literature shows that this natural estimator can be made unbiased by adding a correcting term, *e.g.*, Mecke *et al.* (1990) or Prasad *et al.* (1990). The unbiased estimator thus obtained can actually be generalized to apply to measurement fields of various shapes, and to possess a smaller variance. This generalisation is presented here only in two dimensions, but there is no doubt that it extends to spaces of any dimension.

It should be pointed out that the bias studied in this paper originates from a support problem (limited size of the measurement field), not a discretisation one. The reader interested in the latter problem can consult Serra (1982) or Ohser *et al.* (2003).

### A UNIDIMENSIONAL EXAMPLE

Let  $X$  be a stationary, unidimensional boolean model of segments. It is shown in Appendix 1 that the mean connectivity of  $X$  on  $[0, a]$  can be written as

$$E\{N(X \cap [0, a])\} = \theta a e^{-\theta m} + 1 - e^{-\theta m}, \quad (1)$$

where  $\theta$  is the Poisson intensity and  $m$  is the mean segment length. Starting from Eq. 1, the specific connectivity of  $X$  is obtained first by dividing by  $a$  and then letting  $a$  tend to infinity

$$N_L = \lim_{a \rightarrow \infty} \frac{E\{N(X \cap [0, a])\}}{a} = \theta e^{-\theta m}. \quad (2)$$

Eqs. 1 and 2 show that  $N(X \cap [0, a])/a$  is a biased estimator of  $N_L$ , although an asymptotically unbiased one. In order to quantify the importance of the bias, it is natural to consider the difference

$$\frac{E\{N(X \cap [0, a])\}}{a} - N_L = \frac{1 - e^{-\theta m}}{a}.$$

Up to the factor  $a$ , the right hand side member is nothing but the probability that a point belongs to the boolean model (Matheron, 1975), or equivalently, the expected connectivity of  $X$  at one point. Accordingly

$$\widehat{N}_L(0) = \frac{1}{a} [N(X \cap [0, a]) - N(X \cap \{0\})],$$

$$\widehat{N}_L(a) = \frac{1}{a} [N(X \cap [0, a]) - N(X \cap \{a\})],$$

are unbiased estimators of  $N_L$ . It is actually better to consider their average,

$$\widehat{N}_L = \frac{1}{a} \left[ N(X \cap [0, a]) - \frac{1}{2}N(X \cap \{0\}) - \frac{1}{2}N(X \cap \{a\}) \right], \quad (3)$$

which is also unbiased and has a smaller variance than each of them. Indeed, for  $x = 0$  or  $x = a$ , we have

$$\text{Var}\{\widehat{N}_L(x)\} - \text{Var}\{\widehat{N}_L\} = \frac{1}{4a^2} \text{Var}\{1_{0 \in X} - 1_{a \in X}\} \geq 0.$$

## SPACE OF ARBITRARY DIMENSION

Let  $X$  be a stationary, ergodic random set in  $\mathbb{R}^d$ . Suppose that a realization of  $X$  is available in a domain  $Z$ . Exactly as in one dimension, the natural estimator that is obtained by taking the connectivity of  $X \cap Z$  divided by the volume  $|Z|$  of  $Z$  is only asymptotically unbiased. In order to illustrate the importance of the bias incurred when the domain  $Z$  is too small, we have considered a  $d$ -dimensional stationary boolean model with intensity  $\theta$  and unit hypercubic objects. In the case where  $Z$  is a hypercube with edges parallel to those of the objects, the expected value of this natural estimator, or *reduced mean connectivity* (RMC), is equal to

$$\frac{E\{N(X \cap Z)\}}{|Z|} = \frac{1}{a^d} - e^{-\theta} \sum_{i=0}^d \binom{d}{i} a^{i-d} \sum_{j=0}^i S(i, j) (-\theta)^j. \quad (4)$$

In this expression,  $a$  is the edge length of the hypercubic field, and the  $S(i, j)$ 's are the Stirling numbers of the second kind (*cf.* Appendix 2). This gives

$$\begin{aligned} & \frac{1}{a} - e^{-\theta} \left( \frac{1}{a} - \theta \right) \\ & \frac{1}{a^2} - e^{-\theta} \left[ \frac{1}{a^2} - \theta \left( \frac{2}{a} + 1 \right) + \theta^2 \right] \\ & \frac{1}{a^3} - e^{-\theta} \left[ \frac{1}{a^3} - \theta \left( \frac{3}{a^2} + \frac{3}{a} + 1 \right) + \theta^2 \left( \frac{3}{a} + 3 \right) - \theta^3 \right] \end{aligned}$$

in one, two and three dimensions respectively. Fig. 1 shows this expected value as a function of the proportion of space ( $p(\theta) = 1 - e^{-\theta}$ ) occupied by the objects, for various field sizes ranging from 1 to  $\infty$ . The two-dimensional case is considered here in order to understand the shape of the curves produced. For small  $p$ , the objects are isolated and the RMC is nothing but the mean number of objects per unit area. As  $p$  increases, objects start to overlap and the

RMC keeps on growing, but not so fast. Letting  $p$  growing again, the objects become numerous enough to create holes and the RMC starts decreasing. These holes contribute effectively to the RMC insofar as they are totally included within the field, which explains why their influence is more notable for large fields than for small ones. When  $p$  becomes close to 1, the holes become gradually filled by the arrival of new objects and the RMC starts again growing. Finally, when  $p = 1$ , the whole field is covered with objects, in which case the RMC is equal to  $1/a^d$ .

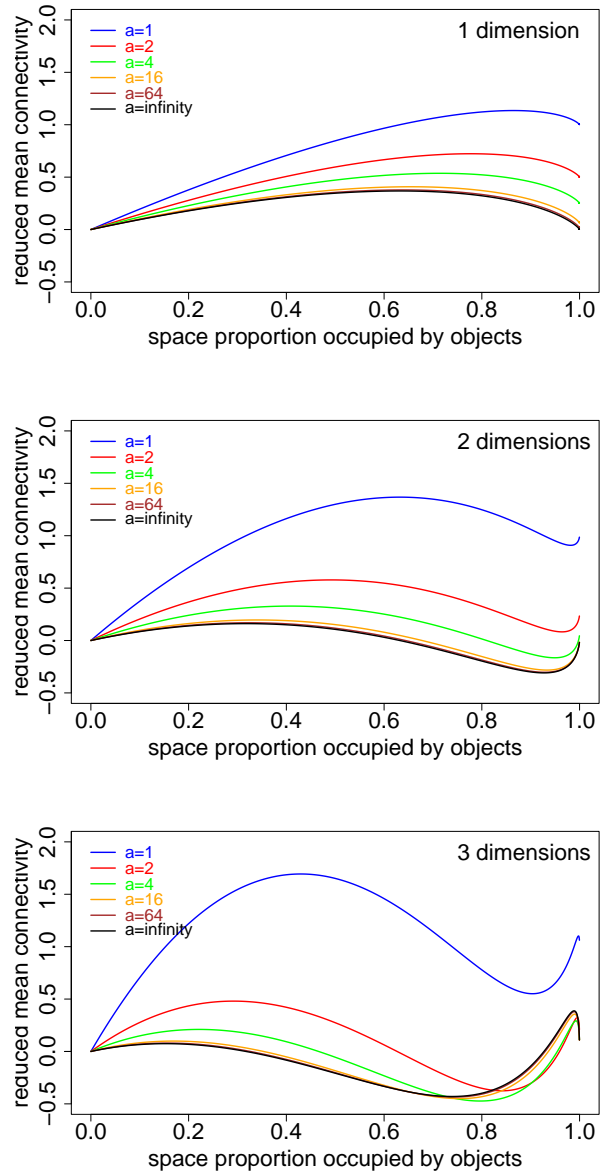


Fig. 1. *Plotting the expected value of the “natural estimator” versus the proportion of space occupied by the objects in one, two and three dimensions, and for various field sizes. The specific connectivity curves are obtained for an infinite size field. They are represented in black.*

Let us come back to our general problem. Unbiased estimators of the specific connectivity of  $X$  can be found in the literature when a realization of  $X$  is available in a right parallelotope  $Z$ , e.g., Mecke *et al.* (1990). They derive from the expression

$$N_V = \frac{1}{|Z|} [E\{N(X \cap Z)\} - E\{N(X \cap \cup_{F \ni v} F)\}], \quad (5)$$

where the union is taken over all faces of  $Z$  (vertices, edges... ) containing the vertex  $v$ . In this equation, the mean connectivity of  $X$  in  $Z$  is amended by a term referred to as the “shell correction” in Prasad *et al.* (1990). Of course, Eq. 5 is valid whatever the vertex  $v$ . By averaging Eq. 5 over all vertices of  $Z$ , we obtain a new expression that can be written as

$$N_V = \frac{1}{|Z|} \sum_{F \in \mathcal{F}} (-2)^{\dim F - d} E\{N(X \cap F)\}, \quad (6)$$

where  $\mathcal{F}$  is the family of all faces of  $Z$ , including  $Z$  itself which is conventionally considered as a face of dimension  $d$  (cf. Appendix 3).

From the estimation standpoint, Eqs. 5 and 6 yield the unbiased estimators

$$\widehat{N}_V(v) = \frac{1}{|Z|} [N(X \cap Z) - N(X \cap (\cup_{F \ni v} F))], \quad (7)$$

$$\widehat{N}_V = \frac{1}{|Z|} \sum_{F \in \mathcal{F}} (-2)^{\dim F - d} N(X \cap F). \quad (8)$$

By construction,  $\widehat{N}_V$  and the  $\widehat{N}_V(v)$ 's are related by the expression

$$\widehat{N}_V = \frac{1}{2^d} \sum_v \widehat{N}_V(v),$$

and if the  $\widehat{N}_V(v)$ 's are identically distributed, then the distribution of  $\widehat{N}_V$  is less dispersed than that of the  $\widehat{N}_V(v)$ 's (cf. Appendix 4).

More generally, Eq. 6 can be written as

$$N_V = \frac{1}{|Z|} \sum_{F \in \mathcal{F}} (-1)^{d - \dim F} \alpha_F E\{N(X \cap F)\}, \quad (9)$$

with  $\alpha_F = 2^{\dim F - d}$ . It turns out that this type of equation is valid for a wide range of polytopes. In what follows, we are going to make precise the role played by the coefficients  $\alpha_F$  in two dimensions, in order to facilitate their assessment.

## BIDIMENSIONAL SPACE

### Right triangle

Suppose at first that  $Z = T_{abc}$  is a right triangle with vertices  $a, b$  and  $c$ , as in Fig. 2. Let  $T_{bcd}$  the triangle deduced from  $T_{abc}$  by symmetry with respect to the midpoint of  $[b, c]$ . The union of both triangles  $R = T_{abc} \cup T_{abd}$  is a rectangle on which Eq. 6 applies.

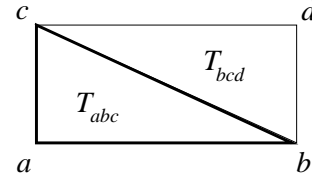


Fig. 2. Case of a right triangle.

In order to get more concise expressions, we put

$$\chi_F = E\{N(X \cap F)\}$$

from now onwards. Because  $\chi_{ab} = \chi_{cd}$ ,  $\chi_{ac} = \chi_{bd}$  as well as  $\chi_a = \chi_b = \chi_c = \chi_d$ , we can write

$$N_A = \frac{1}{|R|} [\chi_R - \chi_{ab} - \chi_{ac} + \chi_a].$$

But we also have  $\chi_R = \chi_{T_{abc}} + \chi_{T_{bcd}} - \chi_{bc}$ . Moreover  $\chi_{T_{abc}} = \chi_{T_{bcd}}$  if the distribution of  $X$  is symmetric w.r.t. the origin. Under this latter assumption, we obtain

$$N_A = \frac{1}{|T_{abc}|} \left[ \chi_{T_{abc}} - \frac{1}{2} (\chi_{ab} + \chi_{ac} + \chi_{bc}) + \frac{1}{2} \chi_a \right]. \quad (10)$$

### Scalene triangle

More generally, let us take for  $Z$  a scalene triangle  $T_{abc}$ . Let  $d$  be the projection of  $a$  onto the line that supports  $b$  and  $c$ . Eq. 10 applies on both right triangles  $T_{abd}$  and  $T_{acd}$ :

$$2|T_{abd}|N_A = 2\chi_{abd} - \chi_{ab} - \chi_{ad} - \chi_{bd} + \chi_a, \quad (11)$$

$$2|T_{acd}|N_A = 2\chi_{acd} - \chi_{ac} - \chi_{ad} - \chi_{cd} + \chi_a. \quad (12)$$

At this stage, two cases must be distinguished, depending on whether  $d$  lies inside or outside  $[b, c]$  (cf. Fig. 3):

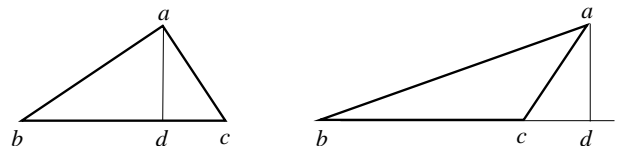


Fig. 3. Case of a scalene triangle.

- if  $d \in [b, c]$ , then  $T_{abc}$  is the union of  $T_{abd}$  and  $T_{acd}$ . By adding Eqs. 11 and 12, and owing to  $\chi_{abd} + \chi_{acd} = \chi_{abc} + \chi_{ad}$  as well as  $\chi_{bd} + \chi_{cd} = \chi_{bc} + \chi_a$ , we arrive at Eq. 10.

- if  $d \notin [b, c]$ , then  $T_{abc}$  is the set difference between  $T_{abd}$  and  $T_{acd}$ . By subtracting Eqs. 11 and 12, and owing to  $\chi_{abd} - \chi_{acd} = \chi_{abc} - \chi_{ac}$  as well as  $\chi_{bd} - \chi_{cd} = \chi_{bc} - \chi_a$ , we end up again with Eq. 10.

It should be pointed out that formula (Eq. 10) is not fully satisfactory because it does not assign the same role to the vertices  $a$ ,  $b$  and  $c$ . Yet, we can resort to the fact that the angles of a triangle add up to  $\pi$  and write

$$\frac{1}{2}\chi_a = \alpha_a\chi_a + \alpha_b\chi_b + \alpha_c\chi_c,$$

where  $\alpha_v$  denotes the internal angle of vertex  $v$  in the triangle divided by  $2\pi$ .

This interpretation is interesting because it is applicable to all faces of the triangle and to the triangle itself. The factor  $1/2$  associated with an edge can be written as  $\pi/(2\pi)$ , and  $\pi$  is precisely the internal angle at any inner point of that edge (cf. Fig. 4). Similarly, the coefficient 1 associated with the triangle can be written as  $(2\pi)/(2\pi)$ , and  $2\pi$  is precisely the internal angle at any inner point of the triangle.

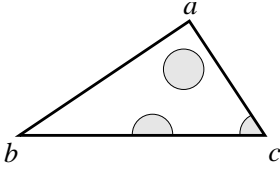


Fig. 4. *Internal angles in a triangle.*

The function  $\alpha$  can be therefore defined on all faces of  $T_{abc}$  including  $T_{abc}$  itself.  $\alpha_F$  is called the *internal angular proportion* of face  $F$ . Using this function, formula (Eq. 9) is valid for triangles.

### Polygon

Let  $Z$  be a simple polygon (polygon limited by a simple Jordan curve). How to estimate  $N_A$  starting from measurements in  $Z$ ?

Let  $(T_i, i \in I)$  be a triangulation of  $Z$ . This means that the  $T_i$ 's have disjoint interiors and that their union is exactly  $Z$  (cf. Fig. 5).

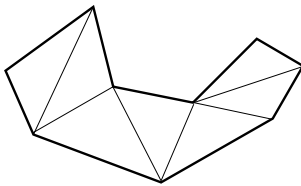


Fig. 5. *Triangulation of a polygon.*

Resorting to a triangulation is advantageous because formula (Eq. 9) can be applied to each element of the triangulation. The results obtained can then be

combined via the inclusion-exclusion formula. This finally gives

$$N_A = \frac{1}{|Z|} \sum_{F \in \mathcal{F}} (-1)^{\dim F} \alpha_F \chi_F, \quad (13)$$

which once again involves the internal angular proportion of all faces of  $Z$  (cf. Fig. 6). Details of the proof are reported in Appendix 5.

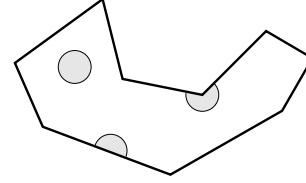


Fig. 6. *Internal angles in a polygon.*

It should be pointed out that this approach is applicable whenever the field  $Z$  can be triangulated. Accordingly, formula (Eq. 13) is valid not only for simple polygons, but also for polygonal domains that will be examined below.

There is another way of writing formula (Eq. 13). Let  $e_1, \dots, e_n$  be the edges of  $Z$ , and let  $v$  be an arbitrary point (vertex or not). As each edge has an internal angular proportion of  $1/2$ , and because the internal angular proportions of each vertex add up to  $(n-2)\pi/2\pi = (n-2)/2$ , this formula becomes

$$N_A = \frac{1}{|Z|} \left( \chi_Z - \frac{1}{2} \sum_{i=1}^n \chi_{e_i} + \frac{(n-2)}{2} \chi_v \right).$$

Note also that the inclusion-exclusion formula relates the sum of all edge contributions to the mean connectivity associated with the boundary  $\partial Z$  of  $Z$

$$\sum_{i=1}^n \chi_{e_i} = \chi_{\partial Z} + n\chi_v.$$

Replacing the edge contributions by their expression, a second formula is obtained for  $N_A$

$$N_A = \frac{1}{|Z|} \left( \chi_Z - \frac{1}{2} \chi_{\partial Z} - \chi_v \right). \quad (14)$$

### Finite union of pairwise disjoint polygons

Here  $Z$  is the union of a finite family  $(Z_i, i \in I)$  of pairwise disjoint polygons. Formula (Eq. 14) applies to each polygon

$$|Z_i|N_A = \chi_{Z_i} - \frac{1}{2} \chi_{\partial Z_i} - \chi_v, \quad i \in I.$$

But  $\sum_i \chi_{Z_i} = \chi_Z$  and  $\sum_i \chi_{\partial Z_i} = \chi_{\partial Z}$ . By summation over all polygons and division by  $|Z|$ , one obtains

$$N_A = \frac{1}{|Z|} \left( \chi_Z - \frac{1}{2} \chi_{\partial Z} - |I|\chi_v \right). \quad (15)$$

### Polygonal domain

Here  $Z$  is a domain limited by a union of simple polygonal curves. Suppose that  $Z$  has  $|I|$  connected components and  $|J|$  holes. Note that  $\chi_Z$  and  $\chi_{\partial Z}$  are not affected by a shift of the connected components of  $Z$  as long as they remain mutually pairwise disjoint. Accordingly, it can be assumed without loss of generality that  $Z$  has no connected component located within one of its holes, as in Fig. 7. Let  $H$  be the closure of the union of the holes. In what follows,  $H$  is supposed to be contained within the interior of  $Z \cup H$ .

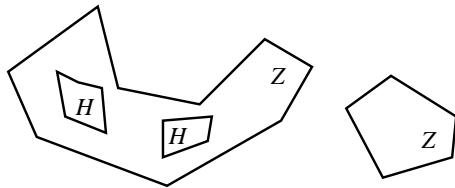


Fig. 7. Example of polygonal domain.

Let us apply formula (Eq. 15) to  $Z \cup H$  and to  $H$

$$|Z \cup H|N_A = \chi_{Z \cup H} - \frac{1}{2}\chi_{\partial(Z \cup H)} - |I|\chi_v,$$

$$|H|N_A = \chi_H - \frac{1}{2}\chi_{\partial H} - |J|\chi_v.$$

By difference, this gives

$$|Z|N_A = \chi_{Z \cup H} - \chi_H - \frac{1}{2}(\chi_{\partial(Z \cup H)} - \chi_{\partial H}) - (|I| - |J|)\chi_v. \tag{16}$$

Now we have

$$\chi_{Z \cup H} = \chi_Z + \chi_H - \chi_{\partial H},$$

because  $Z \cap H = \partial H$ . On the other hand, the boundary of  $Z$  can be split into its “inner” boundary (produced by the holes of  $Z$ ) and its “outer” boundary (obtained by filling the holes of  $Z$ ). Accordingly, we have  $\partial Z = \partial H \cup \partial(Z \cup H)$ , from which we immediately derive

$$\chi_{\partial(Z \cup H)} = \chi_{\partial Z} - \chi_{\partial H}.$$

Replacing  $\chi_{Z \cup H}$  and  $\chi_{\partial(Z \cup H)}$  by their expression in Eq. 16, we obtain

$$|Z|N_A = \chi_Z - \frac{1}{2}\chi_{\partial Z} - (|I| - |J|)\chi_v,$$

and it remains to observe that  $|I| - |J|$  is nothing but the Euler-Poincaré  $N(Z)$  of  $Z$  to conclude

$$N_A = \frac{1}{|Z|} \left( \chi_Z - \frac{1}{2}\chi_{\partial Z} - N(Z)\chi_v \right). \tag{17}$$

### AN EXAMPLE

Classical stereological relationships are often used to characterise microstructures in many scientific fields (biology, geology, materials sciences etc...). The practical application of these relationships rests on the measurement of the specific connectivity. In order to show the difference between its natural estimates and its unbiased ones, as provided by Eqs. 6 and 13, measurements have been carried out on polished sections through a sintered material (Jernot, 1982). The microstructures presented in Fig. 8 correspond to a 60  $\mu\text{m}$  bronze powder, sintered under argon at 730  $^\circ\text{C}$ , 750  $^\circ\text{C}$  and 770  $^\circ\text{C}$  during 2 hrs. The volume fraction of the solid phase is respectively equal to 0.657, 0.716 and 0.790.

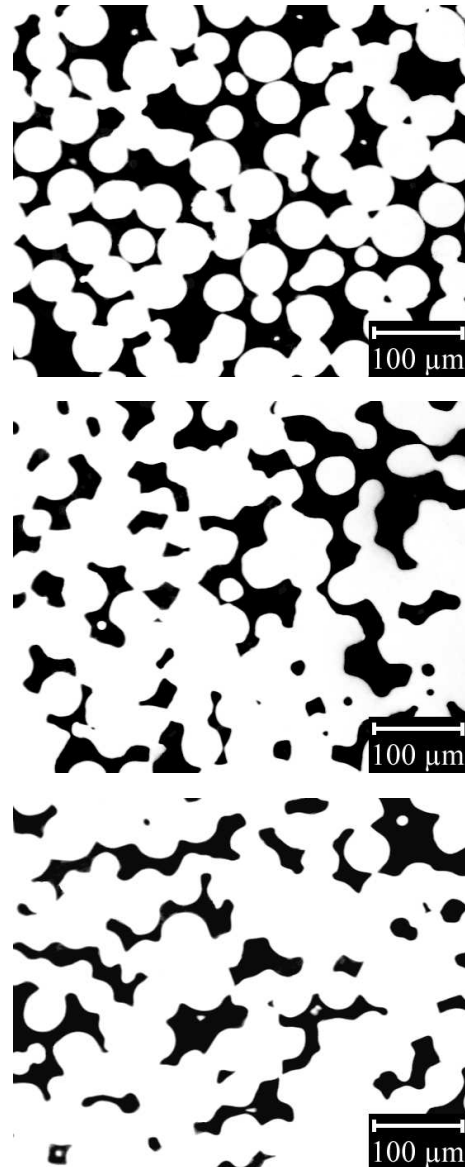


Fig. 8. Three polished sections of bronze particles sintered during 2 hrs at three different temperatures.

The beginning of the sintering is associated to the growth of the necks between the particles and followed by a progressive closing of the porous network due to the densification of the material. This coalescence of particles is reflected in the microstructure by a simultaneous modification of the area and curvature of the solid/pore interface and, consequently, of its specific connectivity. A quick examination of the results obtained from these images shows that the discrepancies between the natural and the unbiased estimates are far from negligible at every stage of sintering. This is particularly manifest on the two-dimensional results.

|                             | 730 °C | 750 °C  | 770 °C |
|-----------------------------|--------|---------|--------|
| 1D nat. (mm <sup>-1</sup> ) | 15.17  | 11.92   | 9.35   |
| 1D unb. (mm <sup>-1</sup> ) | 13.88  | 10.37   | 7.97   |
| 2D nat. (mm <sup>-2</sup> ) | 74.31  | -83.60  | -46.44 |
| 2D unb. (mm <sup>-2</sup> ) | 11.61  | -125.40 | -84.76 |

## CONCLUSION

The connectivity of a stationary random set  $X$  in a field of measurement, divided by the  $d$ -volume of the field, is only an asymptotically unbiased estimator of the specific connectivity of  $X$ . Explicit calculations on the boolean model show that the bias can be important when the field is relatively small with respect to the size of the objects. In two dimensions, two unbiased estimators have been proposed when the distribution of  $X$  is symmetric w.r.t. the origin, and when the field of measurement is a polygonal domain. The first one (Eq. 13) requires the connectivity of  $X$  in the whole field, along each edge and at each vertex. It involves the internal angular proportion of each face. The second one (Eq. 17) requires the connectivity of  $X$  in the whole field, along its boundary and at a point. It involves the Euler-Poincaré characteristic of the field.

## ACKNOWLEDGEMENTS

The authors want to thank Dr W. Nagel for his careful reading of the paper and his judicious remarks.

## REFERENCES

Aigner M (1997). *Combinatorial theory*. Berlin: Springer.  
 Jernot JP (1982). *Analyse morphologique et modélisation du frittage et des matériaux frittés*. PhD Thesis, Caen University.  
 Lantuéjoul C (2002). *Geostatistical simulation. Models and algorithms*. Berlin: Springer.

Matheron G (1975). *Random sets and integral geometry*. New York: Wiley.  
 Mecke J, Schneider RG, Stoyan D, Weil WRR (1990). *Stochastische Geometrie*. Basel: Birkhäuser.  
 Ohser J, Nagel W, Schladitz K (2003). The Euler number of discretised sets – surprising results in three dimensions. *Image Anal Stereol* 22:11–9.  
 Prasad BP, Lantuéjoul C, Jernot JP (1990). Use of the shell correction for quantification of three-dimensional images. *Trans Roy Microsc Soc* 161-3:387–403.  
 Serra J (1982). *Image analysis and mathematical morphology*. London: Academic Press.

## APPENDICES

### APPENDIX 1: PROOF OF EQ. 1

Let  $X$  be a unidimensional, stationary boolean model of segments with Poisson intensity  $\theta$  and length cumulative distribution function  $F$ . Let also  $a \geq 0$ . The aim of this appendix is to calculate the mean connectivity  $E\{N(X \cap [0, a])\}$  of  $X$  over  $[0, a]$ .

Let us assume here that  $F$  has a probability density function  $f$  as well as a finite mean  $m$ . Following Lantuéjoul (2002),  $X \cap [0, a]$  is a boolean model, but its parameters are not those of  $X$ . More precisely,  $X \cap [0, a]$  is distributed as the union in  $[0, a]$  of a Poisson number  $K$  of independent “typical” segments. The mean of  $K$  is  $\vartheta = \theta(a + m)$ . By typical segments, we mean segments  $[X_k, X_k + L_k]$  with respective probability density functions

$$X_k \sim h(x) = \frac{1 - F(-x \wedge 0)}{a + m} 1_{x \leq a},$$

$$L_k | X_k = x \sim f_x(t) = \frac{f(t)}{1 - F(-x \wedge 0)} 1_{t \geq (-x) \vee 0},$$

( $\vee$  and  $\wedge$  respectively stand for the maximum and the minimum operators). Let  $A_k = [X_k, X_k + L_k] \cap [0, a]$ . Then we can write

$$E\{N(X \cap [0, a])\} = \sum_{k=1}^{\infty} e^{-\vartheta} \frac{\vartheta^k}{k!} E\{N(\cup_{n \leq k} A_n)\}.$$

Now the inclusion-exclusion formula gives

$$E\{N(\cup_{n \leq k} A_n)\} = \sum_{n=1}^k (-1)^{n-1} \binom{k}{n} E\{N(\cap_{i \leq n} A_i)\}.$$

Combining both formulae, we obtain

$$E\{N(X \cap [0, a])\} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\vartheta^n}{n!} E\{N(\cap_{i \leq n} A_i)\}. \tag{18}$$

It therefore remains to calculate  $p_n = E\{N(\cap_{i \leq n} A_i)\}$  for each  $n > 0$ . To do so, observe first that

$$p_n = E\{N(\cap_{i \leq n} [X_i, X_i + L_i])\}.$$

Now,  $\cap_{i \leq n} [X_i, X_i + L_i] \neq \emptyset$  if and only if  $X_m = \vee_{i \leq n} X_i \leq \wedge_{i \leq n} (X_i + L_i)$ . Moreover

$$P\{x_m \leq x_i + L_i\} = \int_{x_m - x_i} f_{x_i}(t) dt = \frac{1 - F_{x_i}((x_m - x_i) \vee (-x_i))}{1 - F(-x_i \wedge 0)}.$$

Accordingly, we can write  $p_n = I_1 + I_2$  with

$$I_1 = \int_{x_m < 0} \prod_{i=1}^n \frac{1 - F(-x_i)}{a + m} dx_1 \cdots dx_n,$$

$$I_2 = \int_{x_m \geq 0} \prod_{i=1}^n \frac{1 - F(x_m - x_i)}{a + m} dx_1 \cdots dx_n,$$

The calculation of  $I_1$  is straightforward. We obtain

$$I_1 = \left(\frac{m}{a + m}\right)^n. \quad (19)$$

To calculate  $I_2$ , we write

$$I_2 = n! \int \prod_{i=1}^n \frac{1 - F(x_n - x_i)}{a + m} dx_1 \cdots dx_n,$$

the integration domain being  $x_1 < x_2 < \cdots < x_n$  and  $0 \leq x_n \leq a$ . Using the change of variables  $y_i = x_n - x_i$  for  $i = 1, \dots, n - 1$  and  $y_n = x_n$ , the  $n$  integrals separate and we get

$$I_2 = n \left(\frac{m}{a + m}\right)^{n-1} \frac{a}{a + m}. \quad (20)$$

Combining Eqs. 19 and 20, we get

$$p_n = \frac{m^{n-1}}{(a + m)^n} (m + na). \quad (21)$$

It remains to plug the expression of  $p_n$  into Eq. 18. This finally gives

$$E\{N(X \cap [0, a])\} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\vartheta^n}{n!} \frac{m^{n-1}}{(a + m)^n} (m + na),$$

that is

$$E\{N(X \cap [0, a])\} = \theta a e^{-\theta m} + 1 - e^{-\theta m},$$

after summation.

## APPENDIX 2: PROOF OF EQ. 4

Actually, we are going to prove a little more by considering a boolean model whose objects are right parallelotopes with independent and identically

distributed edge lengths (common mean  $m < \infty$ ). The field  $Z$  is a hypercube with edge length  $a$ . Exactly as in Appendix 1,  $X \cap Z$  can be seen as the union of a Poisson number (mean value  $\vartheta = \theta(a + m)^d$ ) of typical objects. The same reasoning leads to

$$E\{N(X \cap Z)\} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\vartheta^n}{n!} P\{\cap_{i=1}^n A_i \neq \emptyset\}, \quad (22)$$

with  $P\{\cap_{i=1}^n A_i \neq \emptyset\} = \left(\frac{im^{i-1}a + m^i}{a + m}\right)^d,$

or equivalently

$$P\{\cap_{i=1}^n A_i \neq \emptyset\} = \left(\frac{\lambda}{\lambda + 1}\right)^{id} \left(\frac{\lambda + i}{\lambda}\right)^d,$$

by putting  $\lambda = m/a$ . Plugging this expression into Eq. 22, we obtain after some calculations

$$E\{N(X \cap Z)\} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!} (\theta m^d)^n \sum_{i=0}^d \binom{d}{i} n^i \lambda^{-i}. \quad (23)$$

To go further, it is convenient to introduce the Stirling numbers of the second kind. They can be defined by their generating function (Aigner, 1997)

$$n^i = \sum_{j=0}^i S(i, j) (n)_j, \quad (24)$$

with the usual convention  $(n)_j = n(n - 1) \cdots (n - j + 1)$  if  $j > 0$  and  $(n)_0 = 1$ . The following table gives the first Stirling numbers

|         | $j = 0$ | $j = 1$ | $j = 2$ | $j = 3$ |
|---------|---------|---------|---------|---------|
| $i = 0$ | 1       |         |         |         |
| $i = 1$ | 0       | 1       |         |         |
| $i = 2$ | 0       | 1       | 1       |         |
| $i = 3$ | 0       | 1       | 3       | 1       |

We now replace  $n^i$  by its generating function in Eq. 23, and because

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!} (\theta m^d)^n (n)_j = (-1)^{j+1} (\theta m^d)^j e^{-\theta m^d},$$

when  $j > 0$ , whereas

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!} (\theta m^d)^n (n)_0 = 1 - e^{-\theta m^d},$$

it follows

$$E\{N(X \cap Z)\} = 1 - e^{-\theta m^d} \sum_{i=0}^d \binom{d}{i} \lambda^{-i} \sum_{j=0}^i S(i, j) (-\theta m^d)^j,$$

which gives Eq. 4 after dividing by  $|Z| = a^d$ . Note that one gets the specific connectivity by letting  $a$  tend to  $\infty$ :

$$N_V = -e^{-\theta m^d} \frac{1}{m^d} \sum_{j=0}^d S(d, j) (-\theta m^d)^j.$$

### APPENDIX 3: PROOF OF EQ. 6

Here we put

$$M(v) = N(X \cap (\cup_{F \ni v} X \cap F)) .$$

Let  $\{F_i, i \in I(v)\}$  the set of all faces of  $Z$  that contain  $v$ . Because  $\cup_{F \ni v} F = \cup_{i \in I(v)} F_i$ , and owing to the distributivity of the intersection over the union, we have

$$M(v) = N(\cup_{i \in I(v)} X \cap F_i) .$$

By the inclusion-exclusion formula, this becomes

$$M(v) = \sum_{\emptyset \neq J \subset I(v)} (-1)^{|J|-1} N(X \cap F_J) ,$$

where  $F_J = \cap_{j \in J} F_j$  and  $|J|$  is the number of elements of  $J$ . Now it can be noticed that for each face  $F$  of  $Z$  that contains  $v$  there exists one  $J$  and only one such that  $F = F_J$ . Moreover  $|J| = d - \dim F$ . Accordingly

$$M(v) = \sum_{F \ni v} (-1)^{d-1-\dim F} N(X \cap F) .$$

We now take the mean over all vertices of  $Z$  and use the fact that each face of  $Z$  has  $2^{\dim F}$  vertices. Then we arrive at

$$\begin{aligned} \frac{1}{2^d} \sum_v M(v) &= \frac{1}{2^d} \sum_F 2^{\dim F} (-1)^{d-1-\dim F} N(X \cap F) \\ &= - \sum_F (-2)^{\dim F - d} N(X \cap F) . \end{aligned}$$

It remains to take the expectation and plug this equality into Eq. 5 that has been averaged over all vertices of  $Z$  to get Eq. 6 as desired.

### APPENDIX 4: STATISTICAL COMPARISON OF VARIOUS ESTIMATORS

Suppose that all  $\widehat{N}_V(v)$ 's are identically distributed.

Let  $\dot{v}$  be a vertex selected at random. Then  $\widehat{N}_V(\dot{v})$  shares the same distribution. Moreover

$$\begin{aligned} E\{\widehat{N}_V(\dot{v}) \mid \widehat{N}_V\} &= \frac{1}{2^d} \sum_v E\{\widehat{N}_V(v) \mid \widehat{N}_V\} \\ &= E\{\widehat{N}_V \mid \widehat{N}_V\} \\ &= \widehat{N}_V . \end{aligned}$$

Using Jensen inequality, this implies that we have

$\varphi(\widehat{N}_V) = \varphi(E\{\widehat{N}_V(\dot{v}) \mid \widehat{N}_V\}) \leq E\{\varphi(\widehat{N}_V(\dot{v})) \mid \widehat{N}_V\}$ , for each convex function  $\varphi$ . Then taking expectations, we obtain

$$\begin{aligned} E\{\varphi(\widehat{N}_V)\} &\leq E\{E\{\varphi(\widehat{N}_V(\dot{v})) \mid \widehat{N}_V\}\} \\ &= E\{\varphi(\widehat{N}_V(\dot{v}))\} \\ &= E\{\varphi(\widehat{N}_V(v))\} . \end{aligned}$$

Various inequalities can be obtained depending on the choice of  $\varphi$ . For instance,  $\varphi(t) = t$  gives  $E\{\widehat{N}_V\} \leq E\{\widehat{N}_V(v)\}$ . Similarly,  $\varphi(t) = -t$  gives  $-E\{\widehat{N}_V\} \leq -E\{\widehat{N}_V(v)\}$ . Therefore  $\widehat{N}_V$  and  $\widehat{N}_V(v)$  have the same mean. If we also take  $\varphi(t) = t^2$ , we obtain  $E\{\widehat{N}_V^2\} \leq E\{\widehat{N}_V(v)^2\}$ . In combination with the equality of the means, this gives  $\text{Var}\{\widehat{N}_V\} \leq \text{Var}\{\widehat{N}_V(v)\}$ .

### APPENDIX 5: PROOF OF EQ. 13

It is always possible to choose the triangulation  $(T_i, i \in I)$  in such a way that the vertices of its triangles are vertices of  $Z$ .

The standard inclusion-exclusion formula gives

$$\chi_Z = E\{N(X \cap \cup_{i \in I} T_i)\} = \sum_{\emptyset \neq J \subset I} (-1)^{|J|-1} \chi_{T_J} ,$$

with  $T_J = \cap_{j \in J} T_j$ . The  $T_J$  are either triangles (for all  $J$  with  $|J| = 1$ ), or internal edges (edges of the triangles that are not edges of  $Z$ , for certain  $J$ 's with  $|J| = 2$ ), or even vertices of  $Z$  (for all others  $J$  such that  $|J| = 2$  and all  $J$  satisfying  $|J| \geq 3$ ). It turns out that the contribution of each vertex  $v$  of  $Z$  vanishes. Indeed, assume that  $v$  belongs to  $n$  triangles. Then  $v$  is incident to  $n - 1$  internal edges. Its contribution is therefore

$$(-1)^{2-1} \left[ \binom{n}{2} - (n-1) \right] + \sum_{k=3}^n (-1)^{k-1} \binom{n}{k} = 0 .$$

Accordingly, the inclusion-exclusion formula reduces to a summation over the triangles and the internal edges

$$\chi_Z = \sum_{i \in I} \chi_{T_i} - \sum_{i, j \in I} \chi_{T_i \cap T_j} 1_{\dim(T_i \cap T_j)=1} . \quad (25)$$

Now, let us apply Eq. 9 to all triangles

$$\chi_{T_i} = |T_i| N_A - \sum_{F \in \mathcal{F}_i \setminus \{T_i\}} (-1)^{\dim F} \alpha_F(T_i) \chi_F , \quad i \in I ,$$

and plug them into Eq. 25. The contribution of

- each external edge  $F \subset T_i$  is  $\alpha_F(T_i) = \frac{1}{2}$ . This is the internal angular proportion of  $F$  in  $Z$
- each internal edge  $F = T_i \cap T_j$  is  $\alpha_F(T_i) + \alpha_F(T_j) - 1 = \frac{1}{2} + \frac{1}{2} - 1 = 0$
- each vertex  $v$  is  $\sum_{i \in I} \alpha_v(T_i) 1_{v \in T_i}$ . This is exactly the internal angular proportion of  $v$  in  $Z$

As a consequence, Eq. 25 becomes

$$\chi_Z = |Z| N_A - \sum_{F \in \mathcal{F} \setminus \{Z\}} (-1)^{\dim F} \alpha_F \chi_F ,$$

which is nothing but Eq. 13.