

## ON THE NUMBER OF CLUMPS RESULTING FROM THE OVERLAP OF RANDOMLY PLACED FIGURES IN A PLANE

A. M. KELLERER,\* *University of Würzburg*

### Abstract

When two-dimensional figures, called laminae, are randomly placed on a plane domains result that can either be aggregates or individual laminae. The intersection of the union,  $U$ , of these domains with a specified field of view,  $F$ , in the plane is considered. The separate elements of the intersection are called clumps; they may be laminae, aggregates or partial laminae and aggregates. A formula is derived for the expected number of clumps minus enclosed voids. For bounded laminae homeomorphic to a closed disc with isotropic random direction the formula contains only their mean area and mean perimeter, the area and perimeter of  $F$ , and the intensity of the Poisson process.

GEOMETRIC PROBABILITY; PARTICLE COUNTING; OVERLAP CORRECTION; EDGE CORRECTION; RANDOM COVERAGE

### 1. Introduction

In visual or automatic optical counting of particles in a plane the need for appropriate corrections arises if the particle density is high enough that the probability of overlap cannot be disregarded and if the counting procedure does not permit the resolution of aggregates into separate particles. Approximate solutions of the problem are known. In the following an exact relation will be reported.

In agreement with earlier terminology the term lamina will be used for the area covered by a particle. The expression clump will be used for a connected aggregate of laminae. Single non-overlapped particles will also be called clumps.

Motivated by a problem of particle counting Armitage (1949) derived formulae for the number of clumps per lamina in the plane that are suitable if the nominal coverage, i.e. the mean total area of laminae per unit area of the plane, is small. Mack (1948–1954) has given a formula for the mean number of clumps per lamina with arbitrary convex laminae (see also Kendall and Moran (1963)

---

Received 17 May 1982.

\* Postal address: University of Würzburg, Versbacher Str. 5, D-8700 Würzburg, W. Germany.  
Research supported by EURATOM Contract BIO-286-81 D.

and Santaló (1976)) that was first considered as exact but was later recognized as an approximation (Mack (1956), see also Roach (1968)).

Dosimetric studies with heavy ions have recently led us to the overlap problem in connection with the automatic counting of particle tracks that are made visible as etched pits on the surface of exposed samples (see for example François et al. (1980)). The formulae of Armitage or Mack and approximate edge corrections are satisfactory in these applications at nominal coverages less than 0.2 as they are routinely encountered. However, it appeared desirable to obtain an exact relation that links the counts in a finite field of observation with the unknown intensity of the Poisson process. Such a formula will be reported.

## 2. Concepts and definitions and properties of the Poisson process

The subsequent considerations will deal with domains, i.e. laminae, that result from a Poisson process in the plane. The union of these laminae will be termed  $U$ , and the intersection,  $I$ , of  $U$  with a specified field of observation,  $F$ , will be examined.

The field  $F$  will usually be rectangular or circular, but it is merely required to be a sufficiently regular closed domain. This includes, for instance, all polygons. The area of  $F$  is designated by  $A$ , its perimeter by  $S$ .

Cases of pragmatic interest may involve convex or even circular laminae of identical shape and size. Mack has treated the more general case of convex laminae that may differ in shape and size. Even less stringent assumptions are required for the subsequent considerations. The laminae may vary in size and shape and need not be convex nor homeomorphic to a disc. As with  $F$  it will merely be assumed that the laminae are sufficiently regular closed domains. The formula to be derived will contain the mean area,  $a$ , the mean perimeter,  $s$ , and the mean Euler characteristic,  $\chi$ , of the laminae. The Euler characteristic is 1 for laminae homeomorphic to a disc and equals  $\alpha - \beta$  for laminae that consist of  $\alpha$  separate domains and  $\beta$  enclosed voids. The usual case will be that of simple laminae without holes, i.e.  $\chi = 1$ . The perimeter is the total length of all contours that form the boundary of a lamina.

Each lamina is taken to have a reference point that can be called centre. The centres form a Poisson process in the plane, i.e., they are uniformly and independently distributed with an intensity of  $\lambda$  centres per unit area. The laminae are assumed to have isotropically and independently distributed directions. The case of anisotropic orientation will be briefly considered in Section 7.

Certain properties of the Poisson process will be utilized and will be listed in the following without explicit derivation.

The probability of a point in  $F$  not to be covered is  $\exp(-\Psi)$ , where  $\Psi = a\lambda$  is termed the nominal coverage. The expected covered area of  $F$ , i.e., the mean area of  $I = U \cap F$  is

$$(2.1) \quad a_l = A(1 - \exp(-\Psi))$$

where  $A$  is, as stated, the area of  $F$ .

The expected total boundary length of laminae in  $F$  is  $s\lambda A$ . The probability of a differential line element to be not covered by a lamina is  $\exp(-\Psi)$ . Therefore the mean uncovered boundary length of laminae, i.e. the mean contour length of  $U$  in  $F$ , is  $s_1 = s\lambda A \exp(-\Psi)$ . The boundary,  $\partial I$ , of  $I$  contains also the covered part of the boundary,  $\partial F$ , of  $F$ ; this covered part has the mean length  $s_2 = S(1 - \exp(-\Psi))$ , where  $S$  is, as stated, the perimeter of  $F$ . Accordingly the mean length of  $\partial I$  is

$$(2.2) \quad s_l = s_1 + s_2 = s\lambda A \exp(-\Psi) + S(1 - \exp(-\Psi)).$$

The mean number of crossings of a line segment of length  $t$  with boundaries of laminae is, under the condition of isotropy,  $2s\lambda t/\pi$ , and the probability of a crossing not to be covered by other laminae is  $\exp(-\Psi)$  (for similar considerations see Morton (1966) and Gilbert (1965)). The expected number of uncovered crossings in  $F$  between boundaries of laminae is therefore

$$(2.3) \quad n_1 = \exp(-\Psi)A\lambda^2 s^2/\pi.$$

The expected number of uncovered crossings of  $\partial F$  with boundaries of laminae is

$$(2.4) \quad n_2 = \exp(-\Psi)2\lambda sS/\pi.$$

### 3. Formula for the expected number of clumps minus voids

The subsequent considerations will utilize the notion of the total curvature,  $c$ , of a line,  $L$ , in the plane:

$$(3.1) \quad c = \int_L \kappa ds = \int_L d\tau$$

where  $\kappa$  is the curvature and  $s$  is the arc length of  $L$ ;  $\tau$  is the angle of the oriented tangent of  $L$  with the  $x$ -axis, and  $\kappa$  equals  $d\tau/ds$ . The second integral includes the turning angles at any corners, and shows that  $c$  is defined even for lines with corners although the curvature there is infinite.

For a closed curve the total curvature is  $\pm 2\pi$  where the sign depends on the orientation of the curve (see e.g. Santaló (1976)). Fixing the orientation in the usual way so that the domain lies on the left over each contour, one concludes that the average total curvature for a lamina is  $2\pi\chi$ , i.e., the average curvature on the boundary of laminae is  $2\pi\chi/s$ . The probability of a differential line element to be uncovered is independent of its curvature, therefore  $2\pi\chi/s$  is also the average curvature of the uncovered boundaries of laminae.

One can now compute the mean total curvature  $c_I$  on  $\partial I$  and thus obtain  $\chi_I$ , i.e. the mean number of clumps minus enclosed voids. Figure 1 illustrates the situation by a realization of a Poisson process where discs of varying radii intersect a rectangular field of view. The uncovered boundary of the field of view is indicated by the broken line; the intersection,  $I$ , is marked by shading. There are nine clumps and one void in this example.

$c_I$  is the sum of four terms  $c_1$  to  $c_4$ . The first term  $c_1$  is the contribution on uncovered boundaries of laminae

$$(3.2) \quad c_1 = s_1 2\pi\chi / s = \exp(-\Psi) 2\pi\chi\lambda A.$$

The second term is the contribution on the covered part of  $\partial F$ . The probability of a differential line element on  $\partial F$  to be covered is independent of its curvature, i.e., the mean curvature on the covered part of  $\partial F$  is equal to the mean curvature on  $\partial F$ . The total curvature on  $\partial F$  is  $2\pi X$ , where  $X$  is the Euler characteristic of  $F$ . Therefore

$$(3.3) \quad c_2 = s_2 2\pi X / S = (1 - \exp(-\Psi)) 2\pi X.$$

The third contribution is from uncovered crossings between boundaries of laminae in  $F$ . The probability of the crossings to be uncovered is independent of their aperture angles, and the mean turning angle of the uncovered crossings is therefore equal to  $-\pi/2$ , i.e. the mean turning angle of all crossings under the condition of isotropy. Accordingly one has with Equation (2.3):

$$(3.4) \quad c_3 = -n_1 \pi / 2 = -\exp(-\Psi) A \lambda^2 s^2 / 2.$$

Analogous considerations apply to the angles formed in the uncovered intersections of boundaries of laminae with  $\partial F$ , except that the possible angles vary

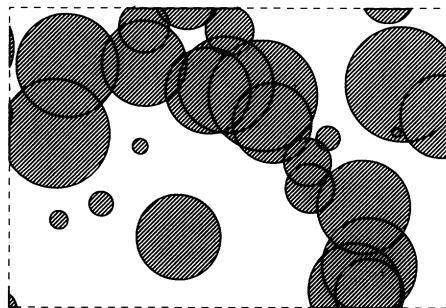


Figure 1. Poisson process in the plane with discs of radius uniformly distributed between 0 and 0.4. The field of view,  $F$ , is taken to be a rectangle of side lengths 2 and 3; its boundary is indicated by the broken line. The intersection,  $I = U \cap F$ , of the discs with  $F$  is represented by the shaded areas. This particular random event has been obtained with nominal coverage  $\Psi = 0.6$ ; there are nine clumps (shaded areas) and one void (enclosed blank area)

between 0 and  $\pi$  and their mean value is  $\pi/2$ . With Equation (2.4) one obtains therefore the last contribution to the mean total curvature of  $\partial I$ :

$$(3.5) \quad c_4 = n_2\pi/2 = \exp(-\Psi)sS\lambda.$$

From the sum of the four terms and with  $\chi_l = c_l/2\pi$  one obtains the main result. The mean Euler characteristic of  $I$ , i.e., the mean number of clumps minus entrapped voids in  $F$  is

$$(3.6) \quad \begin{aligned} \chi_l &= \exp(-\Psi)(\lambda(\chi A + sS/2\pi) - \lambda^2 As^2/4\pi - X) + X \\ &= \exp(-\Psi)(\Psi(\chi A/a + sS/2\pi a) - \Psi^2 As^2/4\pi a^2 - X) + X. \end{aligned}$$

In cases of pragmatic importance one will usually have simply connected laminae and a simply connected field of view, i.e.,  $\chi = 1 = X$ .

It is of interest that common image-analysis systems determine, in their counting mode, the number of clumps minus entrapped voids, rather than the number of clumps. With the additional automatic determination of the area,  $a_l$ , and the contour length,  $s_l$ , the Poisson intensity,  $\lambda$ , can therefore be estimated from Equations (2.1), (2.2) and (3.6), even if  $a$  and  $s$  are unknown. Knowledge of the mean Euler characteristic of the laminae is required; but in most applications one will have laminae homeomorphic to a disc, i.e.  $\chi = 1$ .

No corresponding formulae exist for the numbers of clumps and voids, separately. Their derivation requires Monte Carlo computations. Figure 2 gives numerical results that correspond to Figure 1. A rectangular field of view with side lengths 2 and 3 is assumed and random discs are generated with radii  $]0, 0.4]$  uniformly distributed.  $\chi_l$  from Equation (3.6), i.e., the mean number of clumps minus voids, is represented by the solid curve. The dots with standard errors are

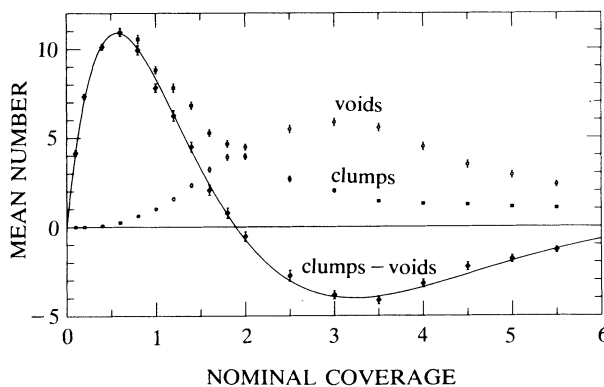


Figure 2. The curve gives as a function of nominal coverage,  $\Psi$ , the mean number of clumps minus voids,  $\chi_l$ , for the process exemplified in Figure 1. The points and standard errors are results for the mean number of clumps (■), the mean number of voids (○), and the mean number of clumps minus voids (◆) obtained by 100 or more Monte-Carlo simulations each for different values of  $\Psi$

results of Monte Carlo simulations. Values for  $\chi_l$  and for the number of clumps and of voids are given separately. The data are obtained by computer analysis of all uncovered contours resulting in each random image and by determination of their total curvatures that are either  $2\pi$  or  $-2\pi$ .

#### 4. Limit case of the infinite plane

Consider  $n = \chi_l / \lambda A$ , i.e. the mean number of clumps minus voids divided by the mean number of laminae (centres) in  $F$ , and let  $F$  be expanded by the linear factor  $r$ , i.e.  $A \rightarrow r^2 A$  and  $S \rightarrow rS$ . For  $r \rightarrow \infty$  one obtains then from Equation (3.6) the mean number of clumps minus voids per lamina:

$$(4.1) \quad m = \exp(-\Psi) \left( \chi - \frac{s^2}{4\pi a} \Psi \right) = \exp(-\Psi)(\chi - f\Psi).$$

For discs of identical size one has  $f = 1$ ; for identical regular convex polygons with  $n$  sides one obtains  $f = tg(\pi/n)n/\pi$ . In both cases one has  $\chi = 1$ .

Equation (4.1) is a lower bound in the limit case for the number of clumps,  $m_c$ , per lamina. Mack's earlier result gives an upper bound (Mack (1956)); for identical laminae with  $\chi = 1$  it takes the form

$$(4.2) \quad m_c \leq \exp \left( -\Psi \left( 1 + \frac{s^2}{4\pi a} \right) \right) = \exp(-\Psi(1 + f)).$$

For laminae of varying shapes and size the relation is more complicated and depends on additional parameters other than the mean area and perimeter of the laminae.

#### 5. A modified relation and the problem of complete coverage

Equation (3.6) applies to the part,  $I = U \cap F$ , of the field of view,  $F$ , that is covered by at least one lamina, i.e., that has a multiplicity of overlap  $k > 0$ . Analogous results can be obtained with different conditions for  $k$ . This will be exemplified by the simple case  $k = 0$ .

The condition  $k = 0$  defines the uncovered part,  $\bar{U} \cap F$ , of  $F$ . Let  $J$  be the closure of  $\bar{U} \cap F$ , this region is depicted in Figure 3 in analogy to the representation of  $I$  in Figure 1 for the same random event. For  $J$  one may speak of gaps rather than clumps (shaded areas in Figure 3) and isles rather than voids (enclosed blanks in Figure 3). There are four gaps and four isles in this example.

By analogy with Equations (3.2) to (3.5) one obtains the following four terms for the mean total curvature,  $c_J$ , on  $\partial J$ :

$$(5.1) \quad c_1 = -\exp(-\Psi)2\pi\chi\lambda A$$

$$(5.2) \quad c_2 = \exp(-\Psi)2\pi X$$

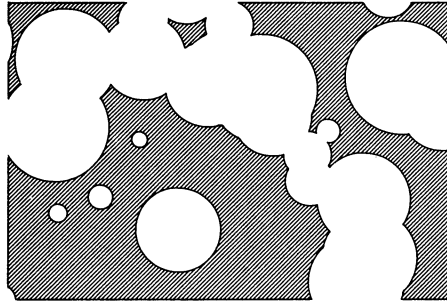


Figure 3. Diagram of the uncovered part,  $J = \bar{U} \cap F$ , of the field of view,  $F$ , for the same random event as in Figure 1. There are four gaps (shaded areas) and four isles (enclosed blank areas)

$$(5.3) \quad c_3 = \exp(-\Psi)A\lambda^2s^2/2$$

$$(5.4) \quad c_4 = \exp(-\Psi)sS\lambda.$$

The mean number of gaps minus isles,  $\chi_I = c_I/2\pi$ , is therefore

$$(5.5) \quad \begin{aligned} \chi_I &= \exp(-\Psi)(\lambda(sS/2\pi - \chi A) + \lambda^2As^2/4\pi + X) \\ &= \exp(-\Psi)(\Psi(sS/2\pi a - \chi A/a) + \Psi^2As^2/4\pi a^2 + X). \end{aligned}$$

Equation (5.5) has some relation to the problem of complete coverage (see e.g. Kendall and Moran (1963), Moran and Fazekas de St Groth (1962), and Gilbert (1965)) of a domain  $F$ .

With  $\chi_I$  from Equation (5.5) one has the inequality for the mean number,  $m_g$ , of gaps on  $F$ :

$$(5.6) \quad m_g \geq \chi_I$$

and with Equation (2.1) one obtains an inequality for the mean area,  $a_g$ , of a gap:

$$(5.7) \quad a_g \leq \exp(-\Psi)A/\chi_I = a/(\Psi(sS/2\pi A - \chi) + \Psi^2s^2/4\pi a + aX/A).$$

Whenever no lamina can be fully contained in  $F$  the equality sign applies in (5.6) and (5.7).

One may surmise that, under certain conditions on the shape and size of the laminae, the gaps tend to be Poisson distributed for  $\Psi \gg 1$ . If, furthermore, the probability for isles can be disregarded, one obtains the approximation  $\exp(-\chi_I)$  for the probability of complete coverage. Figure 4 gives numerical data for a unit radius disc being covered by circular laminae of radius  $r$ , a problem earlier treated by Gilbert (1965). The solid lines for  $1 - \exp(-\chi_I)$  are compared to probabilities of incomplete coverage obtained by Monte Carlo simulations. For  $r \geq 1$  one cannot have more than two gaps, and it is evident that the Poisson distribution cannot apply; for smaller values of  $r$  it appears that  $1 - \exp(-\chi_I)$  is a fair approximation of the probability for incomplete coverage.

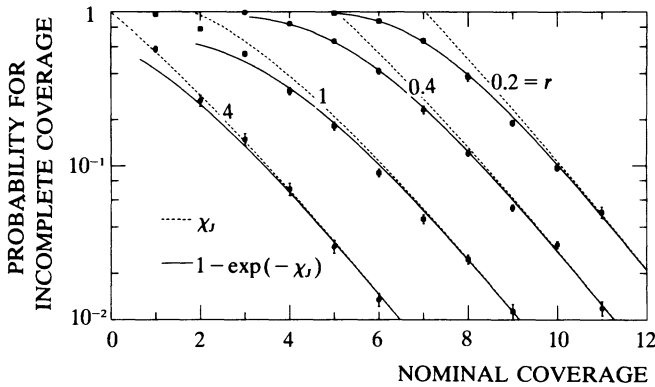


Figure 4. The broken lines give the mean number of gaps minus isles,  $\chi_I$ , resulting in the random coverage of a unit radius disc by discs of specified radius,  $r$ , as a function of the nominal coverage  $\Psi = r^2\pi\lambda$ . The solid lines represent the quantity  $1 - \exp(-\chi_I)$ . The points and standard errors give the probability for incomplete coverage obtained in Monte-Carlo simulations with 100 or more events of incomplete coverage for each point

### 6. Linear structures

Linear structures can be treated as laminae that are narrow bands of vanishing area and of perimeter twice their length. One obtains then from Equation (3.6) the simplified expression for the mean number of clumps minus enclosed voids:

$$(6.1) \quad \chi_I = \lambda(\chi A + lS/\pi) - \lambda^2 A l^2/\pi = \Phi(\chi A/l + S/\pi) - \Phi^2 A/\pi$$

where  $l$  is the mean length of the structures and  $\Phi = \lambda l$  is the mean total length per unit area. (This quantity and its analogue in three-dimensional space are commonly termed fluence.) The mean Euler characteristic,  $\chi$ , is equal to the average number of separate pieces minus the average number of enclosed voids per structure. For single open line segments or branched structures one has  $\chi = 1$ , for ring-like structures  $\chi = 0$ .

Instead of Equation (4.1) one obtains the relation for the mean number of clumps minus entrapped voids per lamina in the infinite plane:

$$(6.2) \quad m = \chi - \lambda l^2/\pi = \chi - \Phi l/\pi.$$

As a further illustration one may consider the modified form of Equations (5.5) and (5.6):

$$(6.3) \quad \begin{aligned} m_g \cong \chi_I &= \lambda^2 l^2 A/\pi + \lambda(lS/\pi - \chi A) + 1 \\ &= \Phi^2 A/\pi + \Phi(S/\pi - \chi A/l) + 1. \end{aligned}$$

If none of the structures can be contained in  $F$ , the equality sign applies, i.e.  $\chi_I$  is the mean number of cells into which  $F$  is subdivided.

It is equally straightforward to obtain relations that apply if  $F$  is a linear structure; these results need therefore not be listed.



### 7. Anisotropic orientation of laminae

Results analogous to Equation (3.6) can be obtained also for anisotropic orientation of the laminae. However, the relations are then of less general nature and can depend on a multitude of parameters.

$\chi_l$  is again the sum of four terms as in Equations (3.2) to (3.5). The terms  $c_1$  and  $c_2$  remain unchanged. For  $c_3$  one needs the mean number,  $n_1 = \exp(-\Psi)A\lambda^2\nu_1$ , of uncovered crossings between laminae and the mean turning angle,  $\tau_1$ , at these crossings. For  $c_4$  one requires the mean number,  $n_2 = \exp(-\Psi)\lambda\nu_2$ , of uncovered crossings on the boundary of  $F$  and their mean turning angle,  $\tau_2$ . The mean number of clumps minus enclosed voids is then

$$(7.1) \quad \chi_l = \exp(-\Psi)(\lambda(\chi A + \nu_2\tau_2/2\pi) + \lambda^2 A\nu_1\tau_1/2\pi - X) + X.$$

Instead of Equation (4.1) one obtains the mean number of clumps minus voids per lamina:

$$(7.2) \quad m = \exp(-\Psi)(\chi - \lambda\nu_1\tau_1/2\pi).$$

If  $F$  is circular, the term  $\nu_2\tau_2$  equals  $sS$ , as in Equation (3.6). In general, however, the computation of  $\nu_2\tau_2$  and  $\nu_1\tau_1$  is complex and depends on the particularities of the situation. The problem will, therefore, not be treated.

For identical, unidirectional laminae results can be obtained readily. Let the field of view,  $F$ , be a polygon with side lengths  $S_k$  and tangent angles  $\Theta_k$  ( $k = 1, \dots, K$ ) and let the laminae be identical polygons with side lengths  $s_j$  and tangent angles  $\vartheta_j$  ( $j = 1, \dots, J$ ). Simple geometric considerations that are omitted can then provide the following relations:

$$(7.3) \quad \nu_1\tau_1 = \sum_{i>j} s_i s_j \sin(-\vartheta_{ij})\vartheta_{ij}$$

and

$$(7.4) \quad \nu_2\tau_2 = \sum_{k,j} S_k s_j \sin(\Theta_{kj})\Theta_{kj}$$

where  $\vartheta_{ij}$  and  $\Theta_{kj}$  are the turning angles at the crossings of  $s_i$  and  $s_j$ , and of  $S_k$  and  $s_j$ , respectively.

As an example, one obtains from Equation (7.3) for rectangles, or generally for parallelograms, of area  $a$ :

$$(7.5) \quad \nu_1\tau_1 = -2\pi a \quad \text{and, accordingly,} \quad m = \exp(-\Psi)(1 - \Psi).$$

Roach (1968) has, with a different derivation, obtained this result for the case of equal squares aligned parallel.

In the case of triangles one has

$$(7.6) \quad \nu_1\tau_1 = -4\pi a \quad \text{and} \quad m = \exp(-\Psi)(1 - 2\Psi).$$

## 8. Conclusion

Earlier treatments of the overlap problem have led to approximate formulae for the mean number of clumps per lamina. In contrast the present result is exact but holds for the mean number of clumps minus enclosed voids. Such a formula is advantageous because the number of clumps minus voids, rather than the number of clumps, is determined in common image-analysis equipment. A further feature of the solution is that it applies to a specified field of observation, so that it accounts both for overlap of laminae and for edge effects.

The considerations have not been concerned with influences of prior occupancy (see Underwood (1970)) that can violate the properties of the Poisson process.

## Acknowledgement

I am indebted to my brother H. G. Kellerer for extensive discussions and substantial improvements to the manuscript. The work was prompted by my colleagues H. Roos and W. Walter who required the solution for their studies in heavy ion track etching.

## References

- ARMITAGE, P. (1949) An overlap problem arising in particle counting. *Biometrika* **36**, 257–266.
- FRANÇOIS, H., KURTZ, N., MASSUE, J. P., MONNIN, M., SCHMITT, R. AND DURRANI, S. A. (EDS.) (1980) *Solid State Nuclear Track Detectors*. Pergamon Press, Oxford.
- GILBERT, E. N. (1965) The probability of covering a sphere with  $N$  circular caps. *Biometrika* **52**, 323–330.
- KENDALL, M. G. AND MORAN, P. A. P. (1963) *Geometrical Probability*. Griffin, London.
- MACK, C. (1948) An exact formula for  $Q_k(n)$ , the probable number of  $k$ -aggregates in a random distribution of  $n$  points. *Phil. Mag.* (7) **39**, 778–790.
- MACK, C. (1949) The expected number of aggregates in a random distribution of points. *Proc. Camb. Phil. Soc.* **46**, 285–292.
- MACK, C. (1953) The effect of overlapping in bacterial counts of incubated colonies. *Biometrika* **40**, 220–222.
- MACK, C. (1954) The expected number of clumps when convex laminae are placed at random and with random orientation on a plane area. *Proc. Camb. Phil. Soc.* **50**, 581–585.
- MACK, C. (1956) On clumps formed when convex laminae or bodies are placed at random in two or three dimensions. *Proc. Camb. Phil. Soc.* **52**, 246–256.
- MORAN, P. A. P. AND FAZEKAS DE ST GROTH, S. (1962) Random circles on a sphere. *Biometrika* **49**, 389–396.
- MORTON, R. R. A. (1966) The expected number and angle of intersections between random curves in a plane. *J. Appl. Prob.* **3**, 559–562.
- ROACH, S. A. (1968) *The Theory of Random Clumping*. Methuen, London.
- SANTALÓ, L. A. (1976) *Integral Geometry and Geometric Probability*. Addison-Wesley, London.
- UNDERWOOD, E. E. (1970) *Quantitative Stereology*. Addison-Wesley, London.