ACTA STEREOL 1998; 17/2: 265-272
ORIGINAL SCIENTIFIC PAPER

A NEW DISTANCE METHOD FOR THE QUANTITATIVE ANALYSIS OF POINT PATTERNS IN SPACE

Yang Zhengwei [1], Liu Guoquan [2], and Nigel G. Wreford [3]

[1] Department of Histology and Embryology, North Sichuan Medical College, Nanchong, Sichuan 637007, China. E-mail: zwyang@nc-public.sc.cninfo.net
[2] Department of Engineering and Materials, University of Science and Technology Beijing, Beijing 100083, China
[3] Department of Anatomy, Monash University, Melbourne, Victoria 3168, Australia

ABSTRACT

A simple way to analyze the spatial distribution of a given point set in space is to first measure the nearest neighbor distance (the distance from a random point sampled to its nearest neighbor point). Then a completely random point pattern with the same point density is generated by computer simulation (a Poisson point process) and the nearest neighbor distances measured in the same way. By comparing the sizes and distributions of the distances, the pattern [clustered, random, or systematic (regular)] of the given point set is determined. It is demonstrated in this paper that the complex and time-consuming procedure, computer simulation of a random point set, can be replaced by a simpler measurement of another nearest distance - the *nearest star distance.* This is the distance from a random *test point* "thrown" into the space to its nearest neighbor point from the given point set. A mean nearest neighbor distance which is smaller than, the same as or bigger than the mean nearest star distance, in combination with their distribution, will suggest a clustered, random or systematic point pattern, respectively.

Key words: point pattern, spatial distribution, stereology.

INTRODUCTION

In the morphological study of particles in metallurgical, geological and biological microscopic structures, a frequently asked but difficult question is related to their spatial distribution: are they distributed in a clustered, independent random or systematic (regular) way; if they are not random how systematic or clustered are they and how far away are they from one another? The general approach is to assign to each particle a unique point (Miles, 1978), and then the analysis problem relates to a point set. There are a few methods available for the quantitative study of spatial distributions (Aherne and Diggle, 1978; Diggle, 1983; Pedro et al, 1984; Baddeley et al, 1987; Brændgaard and Gundersen, 1988). A relatively simple and straightforward method is to first measure the nearest neighbor distance, a distance from a random point sampled to its nearest point; then a completely random point set with the same point density is generated by computer simulation (a Poisson point process) and the nearest neighbor distances measured in the same way. Then the two distances, including their frequency distributions, are compared to judge the spatial distribution for the original point set (Diggle, 1983; Brændgaard and Gundersen, 1988). In this paper a simpler manual measurement of the *nearest star distance* is introduced to achieve the same analysis, making computer simulation unnecessary.

MEASUREMENT OF NEAREST DISTANCES

Nearest neighbor distance

Sample a group of points from a given point set using UR (uniform random) unbiased 2D (two-dimensional) frames (Gundersen, 1977) in 1D or 2D reference space or using 3D *disectors* (Sterio, 1984) in 3D reference space, and measure the distance from each of the points sampled to its nearest point of the set. This distance is called the nearest neighbor distance (Brændgaard and Gundersen, 1988), see Figs. 1 and 2. For the point set distributed along a 1D line, this distance is from the point sampled to one of the two neighbor points on its two sides, and the distance should be measured along the line, not along a straight line as should be the case in 2D or 3D space. The uniform sampling of the points from a given point set in 1D or 2D reference space can be realized by superimposing a rectangular frame and then sampling the points inside the frame (Figs. 1 and 2a). The uniform sampling of points in 3D space should be performed in 3D space by *inserting* a spatial grid [e.g. a disector (Sterio, 1984)] into the space. The manual measurement of the distance in 3D (microscopic) space is almost impossible, and the measurement may be obtained by taking the coordinates first and then calculating the distance from the coordinates (Baddeley et al, 1987).

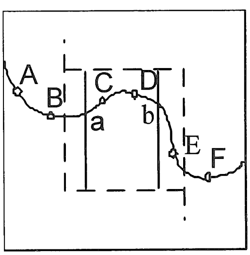

Figure 1

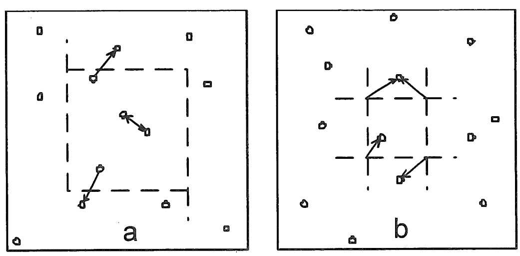

Figure 2

Fig. 1 In this field a segment of a linear structure, along which particles (small circles) A to F are distributed, is illustrated. Superimposed on the field is a sampling frame (dashed). Thus particles C, D and E are sampled and the nearest neighbor distances are measured along the length of the arcs C-D, D-C and E-F. Two straight test lines (solid) are also superimposed to produce test points (the intersections ***a*** *and* ***b*** *between the two test lines and the linear structure), from which the nearest star distances (the length of the arcs a-C and b-D) are measured.*

Fig. 2 In this figure two fields containing particle sets sampled from a reference space are illustrated. (a) A dashed frame is superimposed inside the field and four particles inside it are sampled, thus four nearest neighbor distances (arrows) can be measured. (b) A square grid (dashed lines) is superimposed in the field and the four intersections between the lines are regarded as test points, thus four nearest star distances (arrows) can be measured. In practice the nearest neighbor distances and the nearest star distances may be measured in the same field, but for clarity they are illustrated in separate fields (a and b).

Nearest star distance

"Throw" some UR *test points* into a given point set, and measure the distance from each of the *test points* to its nearest point in the given point set. This distance is called the *nearest star distance*, see Figs. 1 and 2b. In 1D space, *test points* can be generated by superimposing an IUR (isotropic uniform random) straight *test line* grid (2D), the intersections between the 1D structure and the *test lines* representing the *test points* (Fig. 1). In 2D space, UR *test points* are generated by superimposing a UR point grid or a square grid (Fig. 2b). As, in practice, the distance between two points in 3D space may be calculated from their coordinates, the coordinates of *test points* can be arbitrarily predetermined, and the actual *insertion* of a spatial grid into space is unnecessary.

Interpretation of the measurements

Point patterns can be generally divided into three categories: clustered, independent random and systematic (regular). The coordinates of the points in a random pattern are determined by independent uniform random numbers, i.e. the pattern can be generated by a completely random process (statistically a Poisson point process). In other words, the points are distributed irregularly without apparent attraction or repulsion of one another. Clustered or systematic patterns can be generally regarded as varying degrees of deviations from a random one. The points in a clustered pattern tend to attract one another so that their nearest neighbor distances have smaller values and tend to have a smaller (coefficient of) variation in their distribution compared to the nearest star distances. However, the points in a systematic pattern tend to reject one another so that their nearest neighbor distances have larger values and tend to have a smaller variation compared to their nearest star distances.

The nearest neighbor distance reflects the distance between two *neighboring* points, especially in a systematic pattern; *neighbors* are defined here as those between which the nearest neighbor distance is measured. The nearest star distance reflects the size of the *empty space* between points. Imagine a circle in 2D space or a sphere in 3D space with a radius of the nearest star distance, the *test point* from which the distance is measured being the center of the circle or sphere, then the distance just reflects (does not equate to) the star area or star volume (Serra, 1982; Gundersen et al, 1988) of the *empty space* in the point set.

The overall pattern of a given point set is reflected by both the nearest distances and their variation. In a completely random pattern, the mean nearest neighbor distance is expected to equate to the mean nearest star distance, and both are expected to have the same variation. A coefficient of distribution may be defined as the ratio between the two distances:

$$Dc = \overline{d(n)} \div \overline{d(s)} \tag{1}$$

where

Dc = coefficient of distribution

$\overline{d(n)}$ = mean nearest neighbor distance

$\overline{d(s)}$ = mean nearest star distance

The Dc value is expected to be 1 in a random pattern (see below), whereas in a clustered and systematic pattern, the Dc values are expected to be <1 (>0) and >1, respectively. The smaller the Dc value the greater the degree of clustering, whilst larger values of Dc correspond to more even spacing of the points. The Dc value in a systematic pattern has an upper limit, which is 4 for 1D point patterns, 2.849 for 2D point patterns and ~2.255 for 3D point patterns (see Appendix).

THEORETICAL BACKGROUND

The critical point for the distance method is that the nearest neighbor distances and the nearest star distances are expected to have the same values and distribution in a completely random point pattern. This can be readily demonstrated as following:

Suppose there are 100 random points in a point set to be generated in space (1D, 2D or 3D). Generate a random point *RP* first in the space, and then generate the other 99 random points by a random process, and then a nearest neighbor distance, d(n), is measured from the *RP*. Now consider a nearest star distance. Generate a random test point *TP* in the space, and then generate 99 points for the random point set, thus a nearest star distance, d(s)', is measured from the *TP*. In both situations, there are 100 random points in space, therefore, d(n) is statistically expected to equate to d(s)'. However, in the second random process, only 99 points are produced for the given point set. That is, one more point should be *thrown* into the point set to measure the real nearest star distance, d(s), from the *TP*. d(s) might not equate to d(s)'; d(s) is smaller than d(s)' if the last point *thrown* is closer to the *TP* than the other 99 points. However, the probability of this is negligible. For a point pattern in 1D space, the probability would be ~1%, whereas it would be even smaller for a point pattern in 2D or 3D space. In practice, the number of points in a given set is almost always much bigger than 100. If there are only a few points in a point set, then it may not be possible to determine what the pattern is. Therefore it is concluded that the nearest neighbor distances and the nearest star distances would be equal and have the same variation for a completely random point pattern.

The Dc value is expected to be 1 for a random point pattern and it should be smaller than 1 for a clustered pattern and bigger than 1 for a systematic pattern by definition of the patterns (see above).

MODEL STUDIES

A random point pattern in 3D space

Suppose the reference space is a cube with a volume of 100^3 mm^3 in an (x, y, z) coordinate system. Determine a random point inside the cube by choosing three independent random numbers between 0 and 99 from a random number table. A total of 64 points were determined in this way to generate a random point pattern.

63 distances from each point [coordinates (x_1, y_1, z_1)] to each of the other 63 points [coordinates (x_i, y_i, z_i)] were calculated using a spreadsheet (Microsoft Excel 7.0) and the following equation:

$$d = [(x_1 - x_i)^2 + (y_1 - y_i)^2 + (z_1 - z_i)^2]^{1/2} \tag{2}$$

The smallest value of the 63 distances is the nearest neighbor distance measured from the point (x_1, y_1, z_1). 64 nearest neighbor distances were determined in this way from all 64 points in the set.

For convenience a set of 64 systematic *test points* in the cube were predetermined as (12.5, 12.5, 12.5), (12.5, 37.5, 12.5), (12.5, 62.5, 12.5), (12.5, 87.5, 12.5) ...... (87.5, 87.5, 87.5), the numbers representing the coordinates (x, y, z) in mm. 64 distances from each *test point* to each of the 64 points in the given point set were calculated using Equation (2) and the smallest value is the nearest star distance measured from the test point. 64 nearest star distances were determined this way.

For this random point pattern in 3D space, the mean nearest neighbor distance (mm) is found to be 14.43 ± 6.95 (SD) with a coefficient of variation (CV) of 48%, and the mean nearest star distance 14.88 ± 6.34 with a CV of 43%. The Dc value is 0.97, very close to the theoretical value 1, and the distributions of both nearest distances (Fig. 3) are similar, consistent with the theoretical analysis for a random point pattern described above. This model study therefore supports the proposition that the nearest neighbor distance is expected to equate to the nearest star distance for a random point pattern in 3D space.

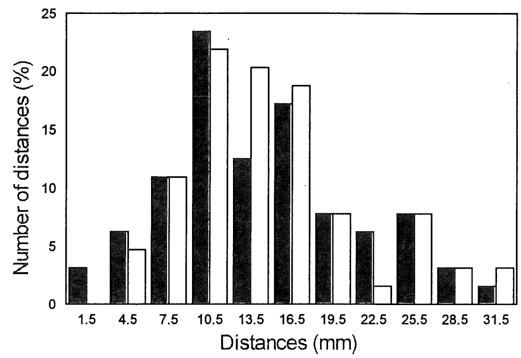

Figure 3 The distributions of the nearest neighbor distances (shaded) and the nearest star distances (empty) measured for the 3D random point pattern described in the text.

Three point patterns in 2D space

Clustered, random and systematic point patterns were generated in three 2D squares (each with area 100^2 mm^2), with 64 points in each square (Fig 4). For convenience a set of 64 UR *test points* were predetermined as (6.25, 6.25), (18.75, 6.25), (31.25, 6.25) ...... (93.75, 93.75), the numbers representing the coordinates (x, y) in mm.

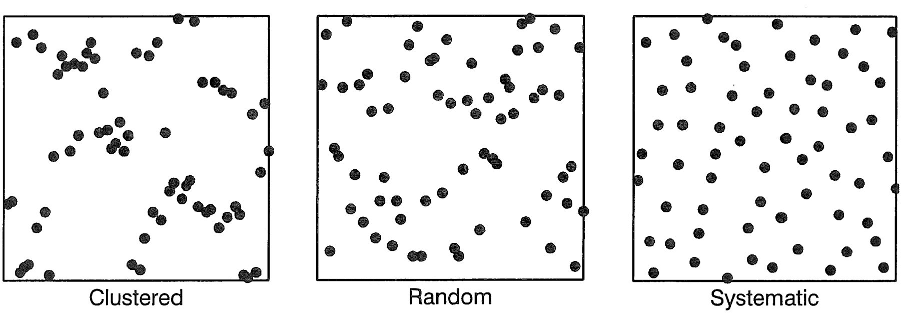

Figure 4 Three point patterns in 2D space (squares) with 64 points (dots) in each pattern. The original size of each square was 100^2 mm^2 in area. The clustered and systematic patterns were arbitrarily drawn first on a piece of coordinate paper and then the coordinates directly read. The coordinates of the points in the random pattern were the (x, y) coordinates of the 64 random points in 3D space generated as above, i.e. the 2D random point pattern being the vertical projection of the random point pattern in 3D space described in the previous part of the text.

Using the equation for distance between two points in a 2D plane, 64 nearest neighbor distances and 64 nearest star distances were calculated from the coordinates of the points as described above. The results show that for the random point pattern, the distribution of the nearest neighbor distances is similar to that of the nearest star distances (Fig. 5); the mean value and CV of the nearest neighbor distances are in the same order of those of the nearest star distances, 6.19 vs 6.35 (mm) in the mean and 57% vs 52% in CV, respectively (Table 1). For the clustered point pattern, however, the mean of the nearest neighbor distances is 51% smaller than that of the nearest star distances. In contrast to the clustered pattern the systematic pattern has a mean nearest neighbor distance 97% larger whilst the CV is 43% smaller than the nearest star distance. Therefore this model is consistent with the proposition that the nearest neighbor distance is expected to equate to the nearest star distance for a random point set in 2D space.

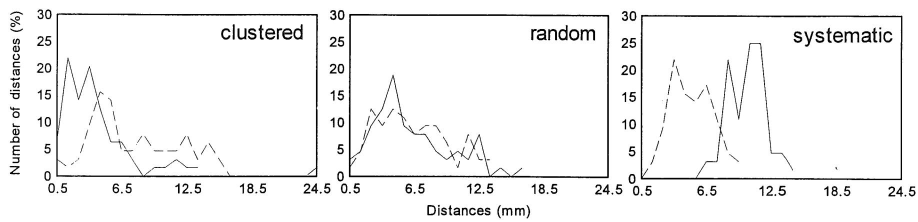

Figure 5 Distributions of the nearest neighbor distances (solid lines) and the nearest star distances (dashed lines) measured for the 2D point patterns (clustered, random and systematic) shown in Fig. 4.

Table 1 The nearest neighbor distances and star distances for the clustered, random and systematic point patterns in 2D space shown in Fig. 4 (mean ± SD, CV in brackets)

patterns	$\overline{d(n)}$	$\overline{d(s)}$	Dc
clustered	4.00 ± 2.91 (73%)	7.89 ± 4.52 (57%)	0.51
random	6.19 ± 3.51 (57%)	6.35 ± 3.30 (52%)	0.97
systematic	10.20 ± 1.64 (16%)	5.17 ± 1.91 (37%)	1.97

$\overline{d(n)}$, the mean nearest neighbor distances; $\overline{d(s)}$, the mean nearest star distance; Dc, $\overline{d(n)}$ / $\overline{d(s)}$.

DISCUSSION

Nearest neighbor distances, in combination with nearest star distances, can be used to describe the spatial distribution of a point set in space. The measurement of the nearest star distance, a simple procedure demonstrated in this paper, is an efficient alternative to computer simulation of random point patterns for the quantitative analysis of spatial distribution (Brændgaard and Gundersen, 1988).

The manual measurement of the two nearest distances for 1D or 2D point patterns is not difficult, however it is almost impossible for 3D point patterns. As distances between points can be calculated from their coordinates and the coordinates of the UR *test points* can be arbitrarily predetermined, it is potentially possible to automate the measurements in image analysis after acquisition of the coordinates using, for example, the tandem-scanning reflected light microscope (Baddeley et al, 1987).

To judge the spatial distribution from the nearest distances, both the average value and their variation (distribution) should be considered. For example, for a systematic (rectangular) point pattern in 2D space shown in Fig. 6c, the nearest neighbor distance is equal to the mean nearest star distance, which may be taken to suggest a random point pattern if not considering the distributions of the distances: the nearest neighbor distance is a constant, i.e. the CV = 0, whereas the CV ≠ 0 for the nearest star distances).

In measuring the distances in practice, there must be sufficient *guard area* around the points (particles) if the natural (not artificial) border is not contained within the sampling space. The two end particles along a segment of linear structure in the sampled reference space should not be sampled to measure the nearest neighbor distances (Fig. 1). Draw a circle (for particles in 2D space) or sphere (for particles in 3D space) with a radius of the nearest neighbor / star distance, the (test) point from which the nearest distance is measured being the center of the circle or sphere, and then the circle or sphere must be inside the reference space sampled to measure the distances (Fig. 2).

APPENDIX

It is readily accepted that there is only one completely regular point pattern in 1D space (along a line), in which the distance between any two neighbor points is a constant, d. The expected nearest star distance is d/4 and therefore its Dc value is 4.

A completely regular point pattern in 2D or 3D space may be defined on the basis of the following two conditions. (i) The nearest neighbor distance to be measured from every point of a given point set is a constant, d. (ii) Draw a circle / sphere with a radius of d/2 at every point in the 2D / 3D point pattern, and then the circles or spheres must meet tangentially to one another in the space, and the empty space not occupied by the circles or spheres must be fully bounded by the circles or spheres which fill it. There are two completely regular point patterns in 2D space: the square pattern (points are at the intersections of a square grid, Fig. 6a) and the right-triangular pattern (points are at the intersections of a rhomboid grid, the rhombus with 4 equal sides, Fig. 6b).

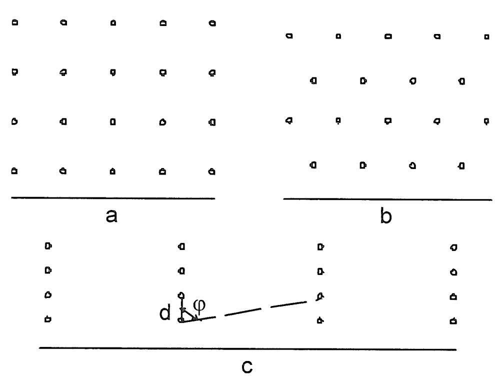

Figure 6 *Three systematic point patterns in 2D space.* ***a****: a square pattern;* ***b****: a right-triangular pattern;* ***c****: a rectangular pattern. The nearest neighbor distances to be measured from each point in each pattern is a constant. Patterns* ***a*** *and* ***b*** *are completely systematic. Pattern* ***b*** *has the biggest coefficient of distribution as defined in Equation (1) for 2D point patterns, 2.849. Pattern* ***a*** *has a coefficient of distribution of 2.614. Pattern* ***c*** *is not completely systematic and has a coefficient of distribution of ~1, with φ equal to ~80°.*

The expected nearest star distance for the patterns in Fig. 6 can be simplified as a distance from a uniform random point inside a right-angular triangle with area (d^2 tgφ / 8) to a vertex (with angle φ, Fig. 6c) of the triangle, and thus obtained by solving a double integral:

$$d(s) = (8\ \mathrm{ctg}\varphi / d^2) \int \int (x^2 + y^2)^{1/2}\, dxdy$$

$$= \int_0^{\varphi} d\varphi \int_0^{(d/2)\sec\varphi} r^2 dr$$

$$= (d\ \mathrm{ctg}\varphi / 6)(\sec\varphi\ \mathrm{tg}\varphi + \ln|\sec\varphi + \mathrm{tg}\varphi|) \qquad (3)$$

where $\varphi = 45°$, $30°$ and $80°$ for the square, right-triangular and rectangular patterns in Figs. 6a, b and c, respectively. The expected nearest star distance is calculated to be 0.383d (Dc = 2.614) for the square pattern (Fig. 6a), 0.351d (Dc = 2.849) for the right-triangular pattern (Fig 6b), and d (Dc = 1) for the rectangular pattern (Fig. 6c).

In 3D space there are also only two completely systematic point patterns, which can be imagined as following: look at the two completely regular patterns in Figs. 6a and b, imagine that the points in the patterns are the centers of a layer of spheres with a constant radius, and then copy the layer of spheres and place them directly onto the first layer in register. Repeat this process. Thus two completely regular patterns in 3D space are achieved – the cuboid pattern and the prismatic pattern. The approximate values of the expected nearest star distance for these two patterns are considered as following:

The unit volume associated with each point in the point set is calculated to be d^3 for the cuboid pattern and $[(3)^{1/2} \cdot d^3 / 2]$ for the prismatic pattern. Now think of two spheres with these volumes, the radii of these two spheres being $\{[3 / (4\pi)]^{1/3} \cdot d\}$ for the cuboid pattern and $[(1/2) \cdot (3)^{1/6} \cdot (3/\pi)^{1/3} \cdot d]$ for the prismatic pattern, respectively. The expected nearest star distance can be approximately regarded as the expected distance from a uniform random point inside the sphere to its center, which is expected to be 3/4 of the radius of the sphere. Thus the mean nearest star distances are calculated to be 0.465d (Dc = 2.149) for the cuboid pattern and 0.443d (Dc = 2.255) for the prismatic pattern. These are very robust estimates for the real values; the approximate Dc values of the completely systematic point patterns in 2D space (Figs. 6a and b) obtained this way have a bias (overestimation) of only 1.7% (for the square pattern) and 0.3% (for the right-triangular pattern) from the real values.

Acknowledgement

The authors wish to thank HJP Gundersen for his reviewing the manuscript.

REFERENCES

Aherne WA, Diggle PJ. The estimation of neuronal population density by a robust distance method. J Microsc 1978; 114: 285-93.

Baddeley AJ, Howard CV, Boyde A, Reid S. Three dimensional analysis of the spatial distribution of particles using the tandem-scanning reflected light microscope. Acta Stereol 1987; 6(Suppl II): 87-100.

Brændgaard H, Gundersen HJG. The impact of recent stereological advances on quantitative studies of the nervous system. J Neurosci Meth 1988; 18: 39-78.

Diggle PJ. Statistical Analysis of Spatial Point Patterns. Academic Press London, 1983.

Gundersen HJG, Bagger P, Bendsten TF, Evans SM, Korbo L, Marcussen N, Mφller A, Nielsen K, Nyengaard JR, Pakkenberg B, Sorensen FB, Versterby A, West MJ. The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. APMIS 1988; 96: 857-81.

Miles RE. The sampling, by quadrats, of planar aggregates. J Microsc 1978; 113: 257-67.

Pedro N, Carmo-Fonseca M, Fernandes P. Quantitative analysis of pore patterns on rat prostate nuclei using spatial statistics methods. J Microsc 1984; 134: 271-80.

Serra J. Image analysis and mathematical morphology. London: Academic Press, 1982: 330-3.

Sterio DC. The unbiased estimation of number and sizes of arbitrary particles using the disector. J Microsc 1984; 134: 127-36.

Accepted: 18 May 1998