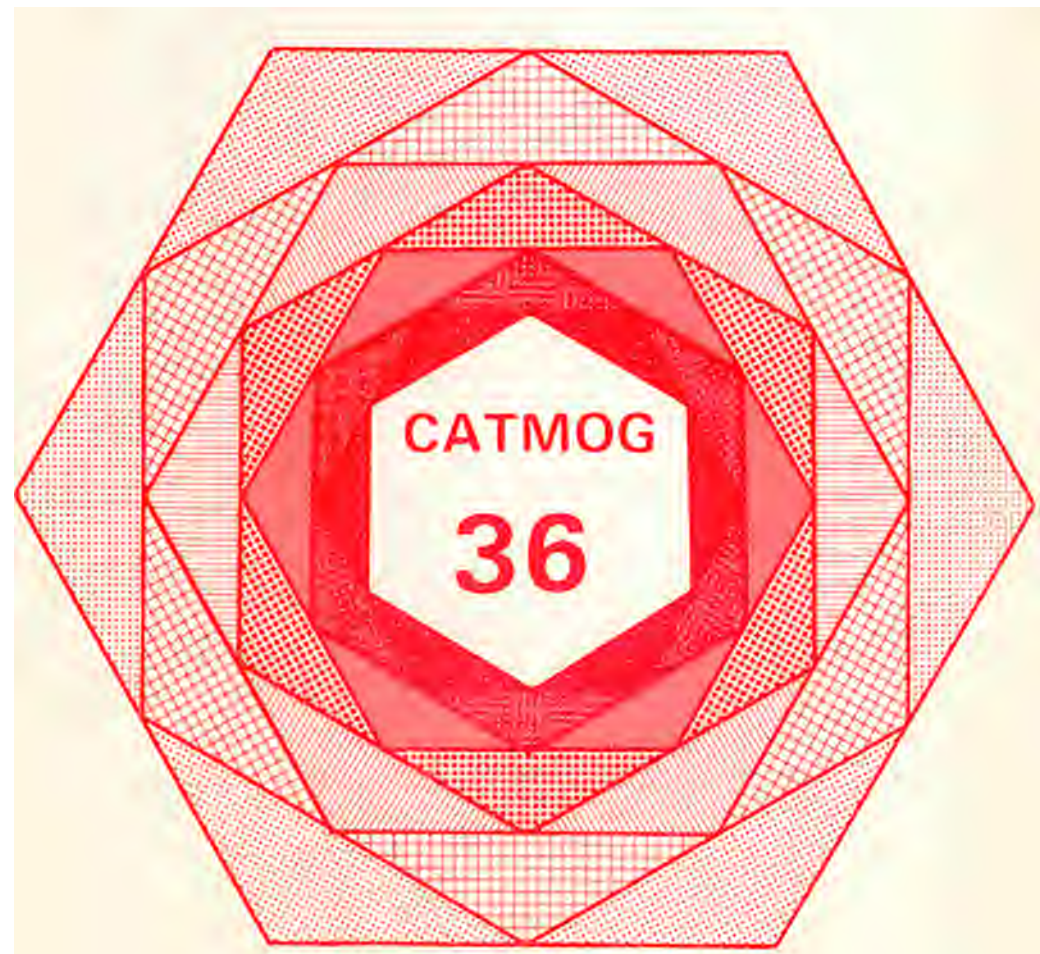
ORDER-NEIGHBOUR ANALYSIS

Graeme **Aplin**



ISSN 0306-6142

ISBN 0 86094 126 4

Graeme Aplin 1983

Published by Geo Books, Norwich - Printed by Headley Brothers Ltd, Kent

CATMOG - Concepts and Techniques in Modern Geography

CATMOG has been created to fill in a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

I. Introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analysis - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
0. Introduction to the use of logit models in geography - N. Wrigley
I. Linear programming: elementary geographical applications of the transportation problem - A. Hay
2. An introduction to quadrat analysis (2nd edition) - R.W. Thomas
3. An introduction to time-geography - N.J. Thrift
4. An introduction to graph theoretical methods in geography - K.J. Tinkler
5. Linear regression in geography - R. Ferguson
6. Probability surface mapping. An introduction with examples and FORTRAN programs - N. Wrigley
7. Sampling methods for geographical research - C.B. Dixon & B. Leach
8. Questionnaires and interviews in geographical research- C.B. Dixon & B. Leach
9. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one dimensional series: an introduction - K.S. Richards
24. Linear programming: the Simplex method with geographical applications - James E. Killen
25. Directional statistics - G.L. Gaile & J.E. Burt
26. Potential models in human geography - D.C. Rich
27. Causal modelling: the Simon-Blalock approach - D.G. Pringle
28. Statistical forecasting - R.J. Bennett
29. The British Census - J.C. Dewdney
30. The analysis of variance - J. Silk
31. Information statistics in geography - R.W. Thomas
32. Centrographic measures in geography - Aharon Kellerman
33. An introduction to dimensional analysis for geographers - Robin Haynes
34. An introduction to Q-analysis -J. Beaumont & A Gatrell
35. The agricultural census - United Kingdom and United States G. Clark
36. Order-neighbour analysis -Graeme Aplin

This series is produced by the Study Group in Quantitative methods, of the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, I Kensington Gore, London SW7.

The series is published by:
Geo Books, Regency House, 34 Duke Street, Norwich NR3 3AP, England, to whom all other enquiries should be addressed.

ORDER-NEIGHBOUR ANALYSIS

by

Graeme Aplin
(Macquarie University)

CONTENTS

Page

I INTRODUCTION 3

II ORDER-NEIGHBOUR ANALYSIS 4

(i) The R statistic 4
(ii) First neighbour level - R(1) 4
(iii) R(1) values for random patterns 5
(iv) R(1) values for non-random patterns 6
(v) Worked example 8
(vi) Higher neighbour levels 11
(vii) Conclusion 13

III THE BOUNDARY EFFECT 13

(i) The boundary effect and approaches to it 13
(ii) Critical values of R(K) in bounded random patterns 14
(iii) Correction equations 15
(iv) Conclusion and example 16

IV CUMULATIVE FREQUENCY DISTRIBUTION 17

(i) Introduction 17
(ii) Bounded patterns 18
(iii) Examples 20
(iv) Lowering the ambiguity level 23

V APPLICATIONS OF ORDER-NEIGHBOUR METHODS 24

(i) Introduction 24
(ii) Selected urban activities, Melbourne 25
(iii) Comparative studies, Sydney and Melbourne 29
(iv) Other published studies 30

VI CONCLUSION 33

(i) Summary of order-neighbour method 33

(ii) Comparisons between patterns 34

(iii) Comparisons with other methods and conclusions 35

BIBLIOGRAPHY 36

ACKNOWLEDGEMENT

Many thanks are due to Adrienne Morgan of the University of California, Berkeley, for drawing the diagrams.

I INTRODUCTION

Geographers have shown understandable interest in point pattern analysis, a group of techniques originating in plant ecology. Many objects of geographical concern can be mapped as sets of points in two-dimensional space and, once a point pattern has been so derived, many questions concerning its nature arise. Point pattern analyses attempt to answer such questions. But while the various methods can provide valuable assistance in doing so, they have been frequently abused or misunderstood. This outline introduces a small subset of the available approaches - order-neighbour analysis - and points out both the values and the pitfalls of its use.

One way of simplifying the complexity that so often confronts us is to abstract phenomena as points, lines and surfaces. Once this has been done, questions of pattern, relative location and spacing become important, and various mathematical and statistical techniques can be brought to bear on them. Such abstraction is not without its own problems, however. In the context of point pattern analysis, it is important to consider whether, at the scale of the study, the areal extent of the original objects can be ignored, and whether the objects can be treated as equal. Section V will treat these and related questions in the context of particular published studies that have used point pattern analysis.

Three basic types of point patterns are commonly recognized: clustered, random and dispersed. These three types are often viewed as a continuum ranging between two extremes; complete clustering and entirely regular spacing. Not all authors agree that such a concept is either valid or useful (e.g. Dacey, 1973). Experience certainly indicates that point patterns are much more complex than a one-dimensional continuum implies. It is thus ultimately desirable that a range of techniques reflecting different pattern characteristics be used. Order-neighbour analysis, as outlined below, is only one of many available approaches.

A random pattern is characterized by two essential conditions: all locations are equally likely to receive a point; and the points are independently located. Strictly speaking, the locations of the points are the results of purely chance occurrences. Whilst this is unlikely to be true for objects studied by geographers, processes do at times interact to produce patterns indistinguishable from mathematically random ones. Most methods of point pattern analysis, first and foremost, provide means of testing the hypothesis of randomness.

A pattern may be non-random in either of two basic ways. Points may be grouped together to form a clustered pattern, the extreme form of such clustering being the occurrence of all points at a single location. On the other hand, the points may effectively repel, rather than attract, each other, giving a highly dispersed pattern. The extreme form of dispersion leads to a pattern of equally and regularly spaced points forming the triangular lattice arrangement familiar from central place theory. Figure 1 shows these three basic pattern types.

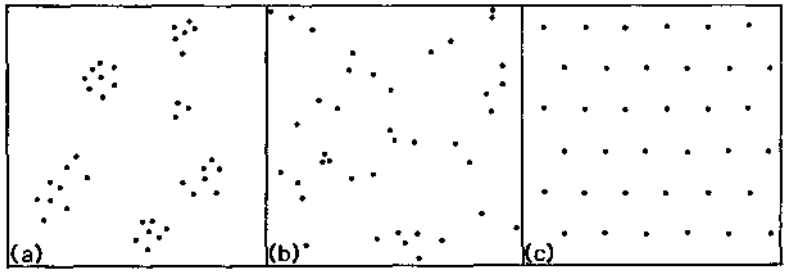


Figure 1. The three major pattern types: (a) clustered; (b) random; and (c) regular.

The various techniques of point pattern analysis attempt to provide a greater degree of objectivity than can be obtained through visual perusal and verbal description. All attempt to increase the precision and replicability of description and analysis. None of them altogether succeeds, and all of them need to be handled with care and understanding.

II ORDER-NEIGHBOUR ANALYSIS

(i) The R statistic

The R statistic method is one of a number of distance-based approaches to point pattern analysis. It involves taking measurements from a set of population points to the first, second, ..., Kth nearest neighbouring points, regardless of the direction of those measurements. It also involves calculation of the density of points in order to be able to establish expected values. Clark and Evans (1954) introduced the first neighbour case (see Section II(ii)), whilst Thompson (1956) extended it to higher neighbour levels (see Section II(vi)).

(ii) First neighbour level - R(1)

The order-neighbour method depends on the calculation of a ratio that relates the mean of the observed point-to-point distances to the corresponding mean expected for random patterns. The calculated value of R(1), i.e. the R statistic at the first neighbour level, is then compared to a value or range of values known to characterize random patterns. This method is, first and foremost, a means of testing the hypothesis of randomness. The full derivations of the reference values and equations listed below are not given here, but it is essential to appreciate the assumptions on which they are based.

The main assumptions and conditions are as follows. First, the study area must be uniform or homogeneous in the sense that the density could, theoretically, be uniform throughout. In other words, we need to be sure that any observed pattern results from some characteristic of the phenomena represented by the points, and not from variations in the environment in

which the locational behaviour takes place. Secondly, the area is infinitely extensive or, at the very least, there is not a boundary that restricts the directions in which distance measurements can be made. Thirdly, a sample of points is taken as the set of origins from which distances are measured.

For the moment, we will assume that the first and second conditions do hold true. The implications of a boundary that does restrict measurements, and some possible ways of dealing with that situation, are treated in Section III. The question of homogeneity is discussed in Section V. In practice, research workers have not been particularly concerned if they have used all available data instead of a sample in many different statistical contexts. The standard argument is that the data available represent a sample from some hypothetical population. In the present context, it is argued that a pattern of points is one of many possible patterns for the group of objects represented. The actual distance measurements are thus a sample from the population of all distance measurements from all patterns. If the number of points in a pattern is large, then a sample of them can be used. Otherwise, it is often thought better to use all of a small data base than to limit it further.

In a random pattern, the mean point-to-point distance has been shown to be:

$$\rho(1) = 0.5\ d^{-0.5} \quad , \qquad (1)$$

where d is the density of points, i.e.,

$$d = N / A \quad , \qquad (2)$$

if N is the total number of points in the study area (whether or not a sample is taken) and A is the area. Equation 1 can also be written as:

$$\rho(1) = 1 / (2\sqrt{d}) \quad . \qquad (1A)$$

An equivalent observed value, F(1), can be found for any empirical pattern by measuring the distance from each (sampled) point to its nearest neighbouring point. Values are then summed, and the mean found. The two mean values are finally compared in ratio form to give:

$$R(1) = \bar{r}(1) / \rho(1) \quad . \qquad (3)$$

(iii) R(1) values for random patterns

As the expected value, p(1), is derived from known characteristics of random patterns, one might expect an empirical random pattern to be characterized by $\bar{r}(1) = \rho(1)$, and hence R(1) = 1.0. It is not quite that simple, however. For any finite value of N, there is no single random pattern: the locations of the points in such patterns are themselves determined by chance. As there is no single random pattern for a given value of N, there can be no corresponding single, definitive value of R. There is, instead, a statistically known and mathematically defined range of values centred on the theoretical value of R(1) = 1.0. In other words, $\bar{r}(1) \approx \rho(1)$, and $R(1) \approx 1.0$.

One of the chief applications of the R statistic is in testing the hypothesis of randomness. Stated in precise, statistical terms, we wish to

be able to either accept or reject the null-hypothesis

$$H_0 : R(1) = 1.0 \quad .$$

This shorthand statement is interpreted as saying that the null-hypothesis (H_0) is that the empirical value of R(1) is not significantly different from 1.0. But how different can it be without that difference becoming statistically significant?

It is known that the distribution of R(1) values from random patterns approximates a normal distribution with a mean of 1.0 and a standard deviation of:

$$\sigma[R(1)] = 0.5228\ N^{-0.5} \quad . \qquad (4)$$

The approximation is, in fact, very close for moderately large values of N, but becomes less so as N decreases. It is recommended, for this reason, that the interpretation of the results of an order-neighbour analysis be treated with considerable caution if N<20. Once the type of distribution and its key parameter values are known, critical values for a two-tailed test can be established as:

$$1.0 \pm z \ . \ \sigma[R(1)] \quad , \qquad (5)$$

where z is the standard deviate corresponding to the required level of significance. Its value can be found from any table of the areas under the normal curve: the two most frequently used levels of significance give z = 1.96 for a = 0.05, and z = 2.58 for a = 0.01. Although treated here as a two-tailed test, it is also possible to test specifically for a deviation from randomness in one direction only, in which case a one-tailed version of the test is used.

Once the critical values have been established, H_0 can be either accepted or rejected. If the calculated value of R(1) falls between the upper and lower critical values, H_0 is accepted and the pattern is taken, on the basis of the present evidence, to be a random one. If, on the other hand, the R(1) value is outside the range defined by the critical values, H_0 is rejected, and the pattern cannot be taken to be random. As with so many similar situations in statistics, the size of the range decreases as N increases. In other words, the greater the number of points, the less is the likelihood that extreme values of R(1) will arise by chance in random patterns. Such a relationship is indicated by the forms of Equations 4 and 5.

Critical values for selected values of N and for $\alpha = 0.01$ and $\alpha = 0.05$ are listed in the first section of Table 1. Values for other N can be found by interpolation or by using Equations 4 and 5.

(iv) R(1) values for non-random patterns

If H_0 is rejected, that merely leads to the conclusion that the pattern is likely, on the basis of this test, to be non-random. As geographers, however, we are likely to be particularly concerned with the form of that non-randomness. As geographers, we may well 'find little of interest in the truly random and unconnected occurrences of theoretical, mathematical randomness' (Gould, 1970, 443). In fact, patterns of geographical interest are

Table 1. Theoretical critical values for R(K) in unbounded random patterns

K	N	Lower critical values		Upper critical values	
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.05$	$\alpha = 0.01$
1	20	0.6984	0.7709	1.2291	1.3016
	30	0.7538	0.8129	1.1871	1.2462
	40	0.7868	0.8380	1.1620	1.2132
	50	0.8093	0.8551	1.1449	1.1907
	60	0.8259	0.8677	1.1323	1.1741
	70	0.8388	0.8775	1.1225	1.1612
	80	0.8492	0.8855	1.1145	1.1508
	90	0.8578	0.8920	1.1080	1.1422
	100	0.8651	0.8975	1.1025	1.1349
	110	0.8714	0.9023	1.0977	1.1286
	120	0.8769	0.9065	1.0935	1.1231
	150	0.8898	0.9163	1.0837	1.1102
	200	0.9045	0.9275	1.0725	1.0955
2	20	0.7905	0.8409	1.1591	1.2095
	30	0.8290	0.8701	1.1299	1.1710
	40	0.8519	0.8875	1.1125	1.1481
	50	0.8675	0.8994	1.1006	1.1325
	60	0.8791	0.9081	1.0919	1.1209
	70	0.8880	0.9149	1.0851	1.1120
	80	0.8953	0.9204	1.0796	1.1047
	90	0.9013	0.9250	1.0750	1.0987
	100	0.9063	0.9288	1.0712	1.0937
	110	0.9107	0.9322	1.0678	1.0893
	120	0.9145	0.9350	1.0650	1.0855
	150	0.9236	0.9420	1.0580	1.0764
	200	0.9337	0.9496	1.0504	1.0663
3	20	0.8303	0.8711	1.1289	1.1697
	30	0.8615	0.8948	1.1052	1.1385
	40	0.8800	0.9089	1.0911	1.1200
	50	0.8927	0.9185	1.0815	1.1073
	60	0.9020	0.9256	1.0744	1.0980
	70	0.9093	0.9311	1.0689	1.0907
	80	0.9152	0.9356	1.0644	1.0848
	90	0.9200	0.9392	1.0608	1.0800
	100	0.9241	0.9424	1.0576	1.0759
	110	0.9277	0.9450	1.0550	1.0723
	120	0.9307	0.9474	1.0526	1.0693
	150	0.9381	0.9530	1.0470	1.0619
	200	0.9463	0.9592	1.0408	1.0537

unlikely to be strictly of that sort at all. However, as Curry says, '.... men, motivated by various ideas (may) act so that from the point of view of the locational structure as a whole, their actions appear random' (Curry, 1964, 145). As a discipline, geography is, though, more likely to be concerned with the discernment of interdependence and non-random patterning. How can order-neighbour analysis help?

First, and most obviously, the value of R(1) can be taken to indicate the broad form of non-randomness present. If R(1) < 1.0, and particularly if R(1) falls below the lower critical value, a tendency to clustering is suggested. On the other hand, if R(1) > 1.0, and particularly if R(1) is larger than the upper critical value, a tendency to dispersion is indicated. In either case, the greater the divergence of R(1) from 1.0, the stronger is the appropriate tendency.

Secondly, R(1) values have been established for the two extreme forms of non-randomness (Section I). If all points are coincident, all r(1) = 0, and hence R(1) = 0. If the points are located on a triangular lattice - the extreme form of dispersion - it can be shown that R(1) = 2.1491. Unfortunately, the R statistic does not serve nearly as well in defining either clustered or dispersed patterns in a broader sense. Should all values of R outside the critical values be taken to indicate either clustered or dispersed patterns? Or are there patterns in which the non-randomness is so weak that they should not be classified as clustered, random, or dispersed?

There are no obvious upper limits to the values of R for clustered patterns or obvious lower limits for dispersed patterns. There are no statistical distributions upon which to base critical values analogous to those for randomness, and hence any limits that are set must be based on other criteria, particularly the context in which the definitions are to be used. Any limit chosen, if its choice is justified, is likely to be an improvement on vaguely defined terms such as 'highly clustered' or 'moderately regular', however.

(v) Worked example

A random point pattern (Figure 2) was constructed by locating points according to pairs of x and y co-ordinates derived from a table of random numbers. The critical values discussed above (Section II(iii)) are based on the assumption, amongst others, that there is not a boundary that restricts measurements in any way. In practical terms, we do have to limit the study area. One frequent approach is to use only points within an inner area as origins, but to allow measurements from them to those points in a surrounding 'buffer zone' if such are closer than any in the inner area. In this example, points were located within the larger area at random until 20 points occurred within the inner area. Their co-ordinates are given in Columns 1 and 2 of Table 2. All 20 points were used as origins, rather than a sample being taken.

The order-neighbour statistic, R(1), is calculated as follows:

1) Each point in the inner area is taken in turn and the distance to its nearest neighbour either measured or calculated (Column 3). Figure 2 shows the measurements involved. N.B. If a pair of points are nearest neighbours to one another (e.g. 71 80 and 76 79), the distance between them must be listed twice.

Table 2. Order-neighbour analysis of Figure 2

Point co-ordinates		Point-to-point distances r(K)*					
		K= 1		K= 2		K= 3	
X	Y	A	B	A	B	A	B
1	2	3	4	5	6	7	8
20	50	2		2	6	6	12
22	49	2		4	5	5	10
24	45	4		6		8	
24	67	4		14		16	18
26	64	4		11		14	16
32	47	4		8		10	
34	50	4		9		11	
36	59	9		11		13	
42	71	10		13		16	
52	72	9		10		13	
55	80	9		16		16	
57	21	9		15	23	16	24
57	56	17		21		21	
63	27	8		13	16	16	
71	80	5		16		18	21
74	38	8		13		13	
76	79	5		13	21	18	25
77	51	8		9	13	13	21
79	26	11	13	13	16	16	18
80	44	8		9		15	18
Er(K)		140	142	226	257	274	320
		7.0	7.1	11.3	12.9	13.7	16.0
p(K)		6.68		10.02		12.53	
R(K)		1.05	1.06	1.13	1.29	1.09	1.28

*r(K) values are given to the nearest millimetre. In each case, Col.A includes distances measured to points in the 'buffer zone' when necessary; Col.B includes only those distances substituted if such measurements are not allowed.

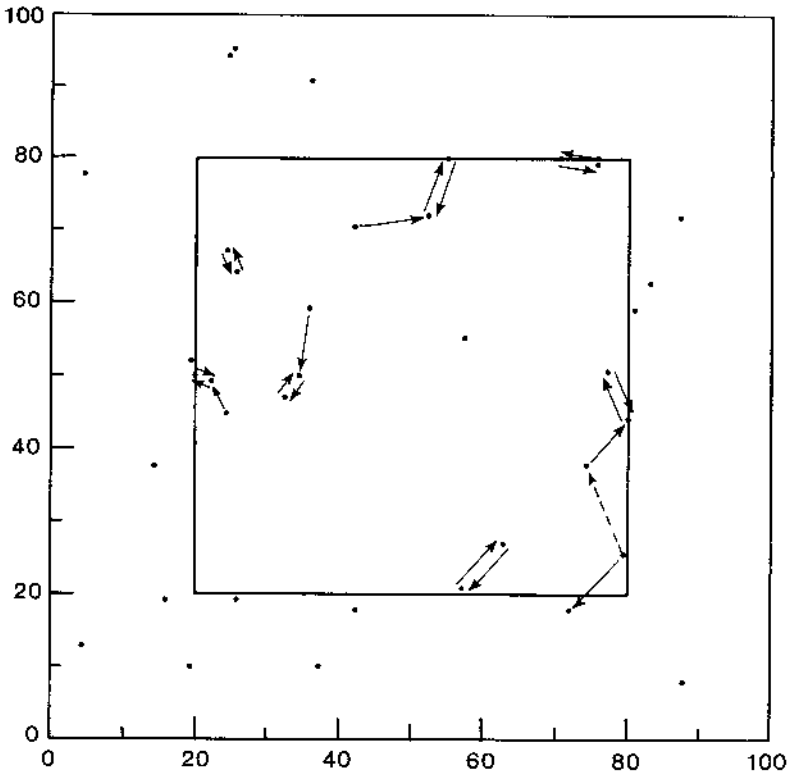


Figure 2. Generated random pattern used as a basis for the worked example of Section II(v).

2) Distances are totalled (140) and the mean found ($\bar{r}(1) = 7.0$).

3) The density of points in the inner area is found:

$$d = 20 / (60 \times 60)$$
$$= 0.0056$$

4) The value of $\sigma(1)$ is calculated from Equation 1 (or 1A):

$$\rho(1) = 0.5\ (0.0056)^{-0.5}$$
$$= 0.5 \times 13.36$$
$$= 6.68$$

5) The observed to expected ratio is calculated:

$$R(1) = 7.0 / 6.68$$
$$= 1.05$$

6) R(1) is well within the range of values defined by the a = 0.05 limits for N = 20 (0.7709 and 1.2291). The null-hypothesis H_o : R(1) = 1.0 is accepted, and the pattern taken to be highly likely to be random.

(vi) Higher neighbour levels

Only limited aspects of a pattern are considered if nearest neighbour distances alone are used. For example, the three patterns in Figure 3 all give R(1) = 0.36. In each case all points are located in pairs, and r(1) values are the same for all patterns. The pairs of points are, however, clustered (A), randomly located (B), and regularly spaced (C) respectively. If R statistics analogous to the R(1) case are calculated for the second and third nearest neighbours, clear differences emerge (Figure 3). This is the basic reason for extending the order neighbour method to higher neighbour levels. It allows analysis and description of patterns at different scales or at different degrees of generalization.

It has been shown (e.g. by Thompson, 1956) that the expression equivalent to Equation 1 is:

$$\rho(K) = [(2K)!\ K] / [(2^K K!)^2\ d^{0.5}]\ , \qquad (6)$$

where K is the neighbour level. More specifically:

$$\rho(2) = 0.75\ d^{-0.5} \qquad (6A)$$

and $$\rho(3) = 0.9375\ d^{-0.5} \qquad (6B)$$

The ratio between observed and expected values then becomes:

$$R(K) = \bar{r}(K) / \rho(K) \qquad (7)$$

The standard deviations at second and third neighbour levels are:

$$\sigma[R(2)] = 0.3630\ N^{-0.5} \qquad (8A)$$

and $$\sigma[R(3)] = 0.2941\ N^{-0.5} \qquad (8B)$$

Critical values at the second and third neighbour levels are listed in Table 1. R(K) = 0 indicates the extreme form of clustering for all values of K, whilst R(2) = 1.4333 and R(3) = 1.1467 indicate the extreme form of dispersion.

Higher-neighbour distances allow consideration of inter-point relationships and patterning at different scales. Figure 3 illustrates just one of many possible examples of patterns which are similar at one level, but very different at others. Whilst the interpretation of the results of higher level analyses may not be as straight forward as for the first neighbour level, their use does provide additional information, and hence allows a

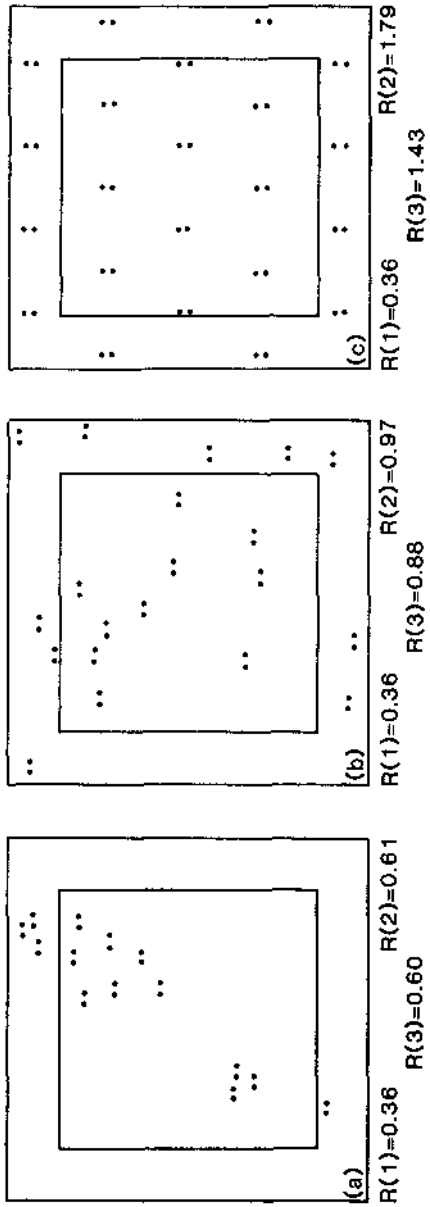


Figure 3. Three patterns comprised of pairs of points. The pairs are (a) clustered; (b) randomly located; and (c) regularly spaced.

fuller description of each pattern. Some authors have provided tables of parameter values for K > 10. Results cannot, however, be interpreted unambiguously when K > 3. In particular, R(K) values for the extreme case of regularity fall within the ranges theoretically indicating randomness. In fact, there are R(4) values (and R(K) values where K > 4) that could arise from moderately highly clustered, random, or dispersed patterns. There may, though, be particular circumstances in which higher neighbour analyses will be particularly valuable, but such applications are beyond the scope of this outline.

Second and third level analyses can, however, be profitably applied to situations like those of Figures 2 and 3. The pattern in Figure 2, constructed to be random, has values of R(2) and R(3) that fall within the relevant a = 0.05 critical values. This pattern is clearly random at the first three levels. One point to note, however, is that the boundary effect becomes much greater at higher levels. The columns headed 'B' in Table 2 give the distances to first, second, and third nearest neighbours that have to be substituted for values in the corresponding column A if buffer-zone measurements are prohibited. For R(1), only one value is affected; for R(2), seven; and for R(3) ten! Differences in the calculated values of R(K) with and without across-boundary measurements also increase as K increases.

In Figure 3, all three patterns consist of pairs of points spaced a constant distance apart. Not surprisingly, R(1) values seemingly indicate three equally highly clustered patterns. But both R(2) and R(3) values show marked differences between patterns and indicate the type of higher level pattern present in each case. At both of these higher levels, Pattern A and Pattern C clearly cannot be accepted as random, as their R(K) values lie outside the a = 0.01 critical values (Table 1). We could conclude, then, that, apart from the presence of the first-neighbour pairs, Pattern A is clustered, Pattern B is random, and Pattern C is dispersed.

(vii) Conclusion

The preceding sections have introduced the fundamentals of order-neighbour analysis. The validity of the results obtained is dependent on the applicability of the assumptions and conditions outlined in Section II(ii) above. In addition, Section II(iv) has suggested that the R statistic, in itself, is of somewhat limited usefulness in describing and classifying non-random patterns. These problems and shortcomings, and some suggestions to overcome them, form the basis for the sections that follow.

III THE BOUNDARY EFFECT

(i) The boundary effect and approaches to it

Theoretical derivations of the expected range of values for R(K) from random patterns depend on the existence of an unbounded area that does not limit distance measurements in any direction. For the use of such a range in hypothesis testing, the empirical $\bar{r}(K)$ value must also be derived from such an unbounded area. If this is not the case, some 'across-boundary' measurements must be discarded in favour of longer ones in other directions. If across-boundary measurements are not possible or not desirable, for whatever reason, the calculated values of R(K) will be inflated.

One solution to this problem was introduced above (Section II(v)). A 'buffer zone' is constructed around the study area and any necessary measurements to points within the extended area are allowed. No points in the 'buffer zone' are used as the origins of measurements, however. In practice, this normally means reducing the effective study area. The inner area is taken to contain a population of points which is an entirely representative sub-section of a larger point population: we are assuming that the pattern in the inner area is part of one extending into the 'buffer zone'. Whether or not such an assumption is valid is often hard to determine before an analysis is carried out; if it is known that the conditions associated with the development of the pattern are identical throughout, the assumption will be soundly based. In addition, however, the appropriateness of this approach depends on the manner in which hypotheses are stated, a point developed further below.

Two other considerations may make the buffer zone approach undesirable in particular circumstances. First, such an approach necessarily leads to a reduction in the number of points available as origins for distance measurements. Secondly, there are any number of situations where the buffer zone approach may be unsuitable because of the way in which it is desirable to frame the hypotheses. Geographers are often confronted with situations in which the value of N is already small. They are also frequently concerned with well-defined study areas best treated as units. Geographical boundaries, whether natural or political, usually cannot be ignored. Measurements cannot be made across a coastline, or across a mountain or other barrier that effectively separates two distinct populations. Nor can they be made to points in a distinctively different cultural or political milieu if the points represent features of the cultural landscape. The alternative is, again, to delimit a smaller area within the whole to create a 'buffer zone'. Great care must then be exercised if the results are to be related to the total unit; it may or may not be possible to infer that the results obtained for the smaller area apply to the whole. Geographers do, like it or not, often wish to deal with bounded patterns with a relatively small number of points - exactly the situation in which the boundary effect is greatest.

Another approach is, however, possible. A set of parameter values, frequency distributions, and critical values from just such bounded patterns can be derived. If this is done, R statistic values from bounded patterns can, with validity, be compared with them in hypothesis testing. Like is again being compared to like. Hsu and Mason (1972) studied the problem in just this manner, and found that the critical values for randomness were significantly larger for bounded patterns. Others have also generated sets of bounded patterns to establish critical ranges, using a simplified Monte Carlo significance test procedure (Hope, 1968). Both Ebdon (1976) and the present author have generated large numbers of patterns with various values of N to empirically establish the ranges of values of R expected from bounded random patterns.

(ii) Critical values of R(K) in bounded random patterns

A computer program was used to generate 200 random patterns in bounded square study areas for N = 25, 50 and 100; and 50 such patterns for N = 20, 40, 60 and 75. The actual degree of divergence of both the mean and standard deviation of R(K) depends on the values of both N and K. The strength of the boundary effect decreases as the value of N increases, but increases as higher

neighbour levels are reached. Curves were fitted to the distributions of values to give the following predictive equations:

$$\mu[R(1)] = 1.0 + 0.4979\ N^{-0.5} \quad (9A)$$

$$\mu[R(2)] = 1.0 + 0.6351\ N^{-0.5} \quad (9B)$$

$$\mu[R(3)] = 1.0 + 0.7706\ N^{-0.5} \quad (9C)$$

$$\sigma[R(1)] = 0.6331\ N^{-0.5} \quad (10A)$$

$$\sigma[R(2)] = 0.5038\ N^{-0.5} \quad (10B)$$

$$\sigma[R(3)] = 0.4687\ N^{-0.5} \quad (10C)$$

It has been established by empirical tests that the R(K) values from bounded patterns follow a normal distribution, and it is thus straightforward to calculate critical values: for certain values of N these are listed in Table 3. Results should not be extrapolated beyond the range of N, $20 \leq N \leq 120$, given.

Only the lower critical values are given for K = 3 in Table 3, as there is considerable overlap between the upper part of the range and the reference value for triangular lattice patterns, R = 1.1467. R(3) does not offer an unambiguous means of differentiating between random and regular patterns (Section II(vi)). The test for randomness should be applied in a one-tailed form at the K = 3 level, and only when $R(3) < \mu[R(3)]$.

All of the patterns used in establishing these equations were in square study areas. Changes in the shape of the area may affect the strength of the boundary effect as the relationship between the perimeter and the area changes. In fact, experience has shown that the tabled critical values (Table 3) can be safely used for study areas without concave or re-entrant boundaries and with a length-to-breadth ratio not exceeding 4:1. These restrictions may exclude some natural or political units, unless their shape is modified, as such units are notoriously irregular.

(iii) Correction equations

Alternatively, it is possible to counteract the boundary effect by applying correction equations. Such equations allow for the trends in the means and standard deviations implicit in Equations 9 and 10. The corrected, bounded R statistic values can then be compared directly to the theoretical critical values of Tables 1, 3 and 4. The relevant correction equations are:

$$\hat{R}(1) = 0.1744 + 0.8256\ R(1) - 0.4111\ N^{-0.5} \quad (11A)$$

$$\hat{R}(2) = 0.2794 + 0.7206\ R(2) - 0.4576\ N^{-0.5} \quad (11B)$$

$$\hat{R}(3) = 0.3726 + 0.6274\ R(3) - 0.4835\ N^{-0.5}. \quad (11C)$$

These equations allow for the increased value of the means and for the differences between theoretical and bounded standard deviation values. Again, these have been derived specifically for square study areas and the comments in the previous section apply equally here.

Table 3. Critical values and means of R(K) for bounded random patterns

K	N	Lower critical values		Mean	Upper critical values	
		$\alpha = 0.01$	$\alpha = 0.05$		$\alpha = 0.05$	$\alpha = 0.01$
1	20	0.7461	0.8339	1.1113	1.3888	1.4766
	30	0.7927	0.8643	1.0909	1.3175	1.3891
	40	0.8205	0.8825	1.0787	1.2749	1.3370
	50	0.8394	0.8949	1.0704	1.2459	1.3014
	60	0.8534	0.9041	1.0643	1.2245	1.2752
	70	0.8643	0.9112	1.0595	1.2078	1.2547
	80	0.8730	0.9169	1.0557	1.1944	1.2383
	90	0.8803	0.9217	1.0525	1.1833	1.2247
	100	0.8864	0.9257	1.0498	1.1739	1.2131
	110	0.8917	0.9292	1.0475	1.1658	1.2032
	120	0.8963	0.9322	1.0455	1.1587	1.1946
2	20	0.8513	0.9212	1.1420	1.3628	1.4327
	30	0.8786	0.9357	1.1159	1.2962	1.3533
	40	0.8949	0.9443	1.1004	1.2566	1.3059
	50	0.9060	0.9502	1.0898	1.2295	1.2736
	60	0.9142	0.9545	1.0820	1.2095	1.2498
	70	0.9205	0.9579	1.0759	1.1939	1.2313
	80	0.9257	0.9606	1.0710	1.1814	1.2163
	90	0.9299	0.9628	1.0669	1.1710	1.2040
	100	0.9335	0.9648	1.0635	1.1623	1.1935
	110	0.9366	0.9664	1.0606	1.1547	1.1845
	120	0.9393	0.9678	1.0580	1.1481	1.1766
3	20	0.9019	0.9669	1.1723		
	30	0.9199	0.9730	1.1407		
	40	0.9306	0.9766	1.1218		
	50	0.9380	0.9791	1.1090		
	60	0.9434	0.9809	1.0995		
	70	0.9476	0.9823	1.0921		
	80	0.9509	0.9834	1.0862		
	90	0.9538	0.9844	1.0812		
	100	0.9561	0.9852	1.0771		
	110	0.9582	0.9859	1.0735		
	120	0.9599	0.9865	1.0703		

(iv) Conclusion and example

The two approaches of the previous sections are applicable when testing the hypothesis of randomness. They are not generally as relevant when describing and analyzing the particular forms of non-randomness present. The more extreme the non-randomness, the less likely it is that the boundary effect will be of the same type and strength. In fact, the extreme form of regularity will show no boundary effect for $K \leq 3$, as each point is equidistant from six others and, unless N is very small, at least three of these will occur within the bounded area. Highly clustered patterns will also

show a negligible boundary effect, if any, unless the boundary runs through clusters containing few points.

Figure 2 was generated as a random pattern. If the boundary is now taken to limit the directions of measurements, we obtain the results given in each Column B of Table 2. The boundary effect on R(1) is very slight. The bounded versions of R(2) and R(3), however, fall above the theoretical upper critical values for randomness ($\alpha = 0.01$). If these results were interpreted without in any way allowing for the boundary effect, it would be concluded that the pattern is most unlikely to be random at the K = 2 and K = 3 levels and has a tendency to dispersion. But we know the pattern that was constructed is random. If we compare the empirical R(K) values to the critical values of Table 3, however, both R(2) and R(3) are within the relevant critical values for $\alpha = 0.05$. The pattern would thus, correctly, be interpreted as being statistically acceptable as random at all three neighbour levels.

If the correction equations are applied, the corrected R(K) values are $\hat{R}(1) = 0.9576$, $\hat{R}(2) = 1.1067$, and $\hat{R}(3) = 1.0676$. One feature of this approach is that the equations allow for the 'average' boundary effect. As the particular strengths for given patterns vary due to the chance or random elements involved, one cannot expect each R(K) value to be identical to the corresponding R(K) value calculated with across-boundary measurements included. In this example, the boundary effect at the K = 1 level was very small, and the R(1) value has been over-corrected. The R(2) and R(3) values do, however, approximate the unbounded R(2) and R(3) values quite closely. The corrected value at each level falls within the appropriate theoretical critical values ($\alpha = 0.05$), and the hypothesis of randomness cannot be rejected.

IV CUMULATIVE FREQUENCY DISTRIBUTIONS

(i) Introduction

Order-neighbour analysis, as discussed so far, involves the interpretation of a single R(K) value at each neighbour level. The calculation of these values necessarily involves a very considerable loss of information. An alternative approach that would allow some of this to be retained and used in the descriptive and analytical processes is thus intuitively appealing. Cumulative frequency distribution analysis constitutes one such approach. Cowie (1969) strongly advocates this approach, claiming two particular advantages for this method: it circumvents the boundary problem; and it allows a more detailed analysis and description of patterns. The first of these claims is quite inaccurate, but the second is definitely justified.

It is possible to use the gamma distribution as the basis for a test of randomness (Dacey, 1964, 46), but the relevant tables are unlikely to be readily available and are not easy to use. A normal distribution, on the other hand, is readily applied and may be used as a simple, but less accurate, alternative. In either case, the expected mean and standard deviation for the distribution of R(K) values is known (Section II), and hence the complete expected frequency distribution can be established. A standard goodness-of-fit test can then be used to test the null-hypothesis of no significant

difference between observed and expected distributions. This is equivalent to testing the hypothesis of randomness.

Aplin (1979) advocates a third approach. A square root transformation is often used to reduce positive skewness in a distribution. There is also a theoretical justification for its use in this context: if a random variable x follows a normal distribution, a second random variable $v = x^2$ follows a gamma distribution. Hence, the square roots of distance measurements, $r(K)^{0.5}$, from a random pattern will be normally distributed. This approach thus combines the ease of finding expected frequencies from a table of areas under the normal curve with accuracy. Arithmetic probability paper can be used to give a relatively quick and easy visual test for randomness, or a standard goodness-of-fit test can be applied. Unfortunately, theoretical values for the parameters have not, to the author's knowledge, been established. Any test based on $r(K)^{0.5}$ values would thus necessarily be based on empirically derived parameter values.

(ii) Bounded patterns

Although parameter values have not been established for unbounded patterns, the author has used a process analogous to that outlined in Section III to find equivalent values for bounded, random patterns. The comments on the shape of study areas (Section II(ii)) again apply. A process of curve-fitting gave the following predictive equations:

$$\mu[r(1)^{0.5}] = 0.6734 + 0.1916\ N^{-0.5} \quad (12A)$$

$$\mu[r(2)^{0.5}] = 0.8471 + 0.2639\ N^{-0.5} \quad (12B)$$

$$\mu[r(3)^{0.5}] = 0.9512 + 0.3716\ N^{-0.5} \quad (12C)$$

$$\sigma[r(1)^{0.5}] = 0.1948 + 0.0702\ N^{-0.5} \quad (13A)$$

$$\sigma[r(2)^{0.5}] = 0.1727 + 0.0870\ N^{-0.5} \quad (13B)$$

$$\sigma[r(3)^{0.5}] = 0.1548 + 0.1523\ N^{-0.5} \quad (13C)$$

where μ is the mean and σ the standard deviation of the $r(K)^{0.5}$ distribution in question. All of the above equations are standardized to a point density of d = 1.0: for other densities, multiply the values given by $d^{-0.25}$. Values of the parameters for selected N are given in Table 4.

The expected cumulative frequency distribution for random patterns is a cumulative normal distribution with mean and standard deviation values calculated from Equations 12 and 13. Any such distribution plots as a straight line on arithmetic probability paper (Figure 4). Random patterns are indicated by curves that approximate this line. Non-random patterns with a tendency to dispersion are indicated by curves which are concave upwards and to the right of the line; and those with a tendency to clustering by curves which are convex upwards and to the left of the line. Neither extreme pattern can be plotted, as probability paper cannot show either 0 percent or 100 percent. In practical terms, a random pattern is indicated by a distribution which does not differ significantly from the expected normal distribution. Such significance can be tested in a number of ways (Gardiner and Gardiner, 1978, Section IV), including goodness-of-fit tests such as the chi-square and two-sample Kolmogorov-Smirnov tests, or by using the critical values of the latter to plot confidence limits on the probability graph (see Section IV(iv)).

Table 4. Parameters of distributions of $r(K)^{0.5}$ values for bounded, random patterns

K	N	Predicted value of: Mean*	Standard deviation*
1	20	0.7613	0.2141
	25	0.7118	0.2124
	40	0.7037	0.2095
	50	0.7005	0.2083
	60	0.6982	0.2074
	75	0.6956	0.2065
	100	0.6926	0.2054
2	20	0.9061	0.1922
	25	0.8998	0.1901
	40	0.8888	0.1865
	50	0.8844	0.1851
	60	0.8811	0.1840
	75	0.8775	0.1828
	100	0.8735	0.1814
3	20	1.0342	0.1888
	25	1.0255	0.1852
	40	1.0099	0.1789
	50	1.0037	0.1763
	60	0.9991	0.1745
	75	0.9941	0.1724
	100	0.9883	0.1700

*All values are standardized to d = 1.0: for other densities multiply tabled values by $d^{-0.25}$.

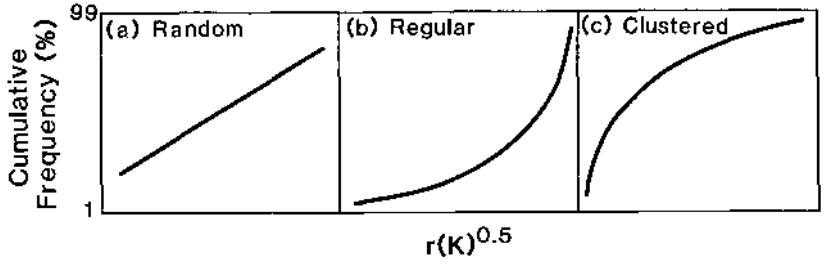


Figure 4. General forms of cumulative frequency plots on arithmetic probability paper.

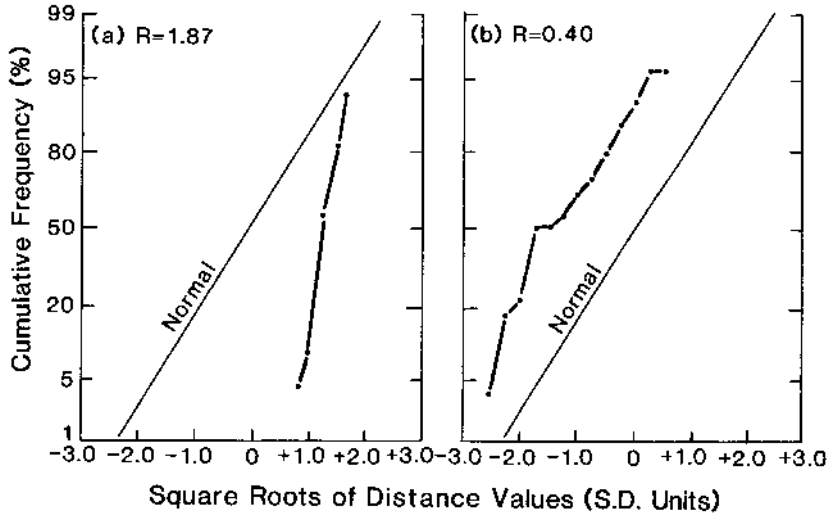


Figure 5. Cumulative frequency plots of $r(1)^{0.5}$ values for (a) regular and (b) clustered patterns: see Section IV(iv).

(iii) Examples

Figure 5 shows two K = 1 cumulative frequency curves plotted on probability paper. Curve A is derived from a regular lattice pattern with small, random displacements of points from their original lattice positions. It indicates an absence of small distance measurements, and the occurrence of the bulk of all measurements in a small range of values at about one standard deviation unit above the mean. It should be noted that raw distance values could have been used on the horizontal scale, but comparability between graphs, and hence between patterns, would then be reduced. Curve B is derived from a clustered (Thomas Double Poisson) distribution, and clearly indicates the prevalence of small distance values.

Only one example, the K = 1 case for the pattern of Figure 2, is worked in detail. The corresponding cumulative frequency distribution is graphed as Figure 6. The numerical values involved are given in Table 5. An analysis of this type involves the following steps:

1) Find the square roots of relevant distance measurements; i.e. calculate $r(K)^{0.5}$ values.

2) Calculate the mean and standard deviation for the expected distribution from Equations 12 and 13: for N = 20 and d = 0.0056, $\mu[r(1)^{0.5}] = 2.6237$, and $\sigma[r(1)^{0.5}] = 0.7842$.

3) Use a table ofareas under the normal curve to establish the expected cumulative frequency at each of a number of z-score or standard

Table 5. Cumulative frequency distribution analysis of Figure 2 (K = 1): Kolmogorov-Smirnov test

Upper class boundary		Cumulative frequencies		D
Standard deviation units	$r(K)^{0.5}$	Observed (%)	Expected (%)	
1	2	3	4	5
-2.5	0.6632	0	0.62	0.62
-2.0	1.0553	0	2.28	2.28
-1.5	1.4474	10	6.68	3.32
-1.0	1.8395	10	15.87	5.87
-0.5	2.2316	35	30.85	4.15
0	2.6237	45	50.00	5.00
+0.5	3.0158	85	69.15	15.85*
+1.0	3.4079	95	84.13	10.87
+1.5	3.8000	95	93.32	1.68
+2.0	4.1921	100	97.72	2.28
+2.5	4.5842	100	99.38	0.62

* D(max) = 15.85

deviation unit values. The values used as class boundaries are shown in Column 1 as z-scores, and in Column 2 as $r(1)^{0.5}$ values; the expected frequencies are given in Column 4.

4) Tabulate the corresponding observed cumulative frequency values for all class boundaries: Column 3 gives these observed values.

5) Find the difference between the observed and expected frequencies in each pair, ignoring the plus and minus signs.

6) Find the maximum value of D, D_{max} = 15.85.

7) Use a table of the critical values of the Kolmogorov-Smirnov statistic, $D(N,\alpha)$. For N = 20, the critical values are $D(20, 0.05) = 0.45$, and $D(20, 0.01) = 0.55$. In other words, for N = 20 and $\alpha = 0.05$, the two distributions must, at some stage, differ by more than 0.45 (if a 0 to 1 proportional scale is used) or by more than 45 percent (if the cumulative frequencies are expressed as percentages) before the difference is significant. Critical values for $D(N,\alpha)$ may be found from Table L of Siegel (1956, 278-9) for N≤40; or from the formulae:

$$D(N, 0.05) = 1.36\sqrt{\frac{2}{N}} \qquad (14A)$$

and

$$D(N, 0.01) = 1.63\sqrt{\frac{2}{N}} \qquad (14B)$$

It is necessary to use the two-sample version of the Kolmogorov-Smirnov test as the mean and standard deviation values have been calculated from samples and the test does not involve a theoretically defined distribution.

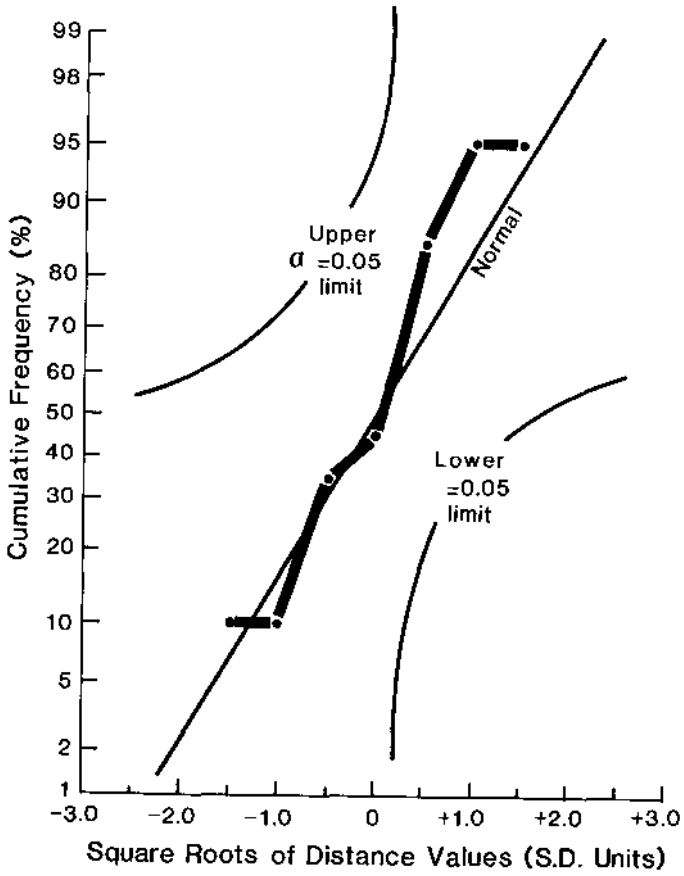


Figure 6. Cumulative frequency plot of $r(1)^{0.5}$ values for the random point pattern of Figure 3.

8) If the observed $D_{max} > D(N,\alpha)$, then the null-hypothesis of no significant difference is accepted at that level of significance. In the example, $D_{max} < D(20, 0.05)$ (i.e. 15.85<45.00) and the pattern is accepted as random.

It is also possible to test the hypothesis graphically, as shown in Figure 6. The additional steps are:

9) Plot the expected cumulative frequencies for random patterns on arithmetic probability paper; such plots are always straight lines.

10) Plot the critical limits for the chosen α value as curves that lie at distances equivalent to the tabled critical value of $D(N,\alpha)$ above and below the normal distribution line. These plots are curved, rather than parallel to the normal line, because of the way in which the scale of the graph paper changes as either 0 percent or 100 percent is approached.

11) Plot the observed distribution: values of 0 percent or 100 percent cannot be plotted.

12) If the observed distribution lies outside the critical limits at any point, the hypothesis of randomness must be rejected at that confidence level. In this example, the observed distribution plot lies entirely within the critical limits defined by D(20, 0.05), and hence the pattern is accepted as random at the α = 0.05 level of significance.

A third alternative is to use the chi-square test. Because of the limitations on the permitted number of cells with small expected frequencies, this test is inappropriate when N is small. Table 6 illustrates its use with a hypothetical case for which N = 120.

Table 6. Hypothetical cumulative frequency distribution analysis using chi-square test

Distance class (standard deviation units)	Frequencies Observed	Expected	$\frac{(O-E)^2}{E}$
<-2.0	2	2.70	0.18
-2.0 to -1.5	8	5.28	1.40
-1.5 to -1.0	12	11.04	0.08
-1.0 to -0.5	26	18.00	3.56
-0.5 to 0	35	22.98	6.29
0 to 0.5	15	22.98	2.77
0.5 to 1.0	10	18.00	3.56
1.0 to 1.5	6	11.04	2.30
1.5 to 2.0	4	5.28	0.31
>2.0	2	2.70	0.18
		$\Sigma \frac{(O-E)^2}{E} = \chi^2 =$	20.63

Critical values for nine degrees of freedom are $\chi^2_{0.05} = 16.92$ and $\chi^2_{0.01} = 21.67$. The hypothesis of no difference between the observed and expected frequencies is thus rejected at the α = 0.05 level, but accepted at the α = 0.01 level. A two-sample Kolmogorov-Smirnov test on this data leads to the same conclusions as D_{max} = 19.17, whilst Equations 14A and 14B give values of D(120, 0.05) = 17.56 and D(120, 0.01) = 21.04.

(iv) Lowering the ambiguity level

It has been suggested that the cumulative frequency approach allows greater use to be made of the available information in reaching conclusions and providing descriptions. Figure 7 serves to reinforce this claim. A

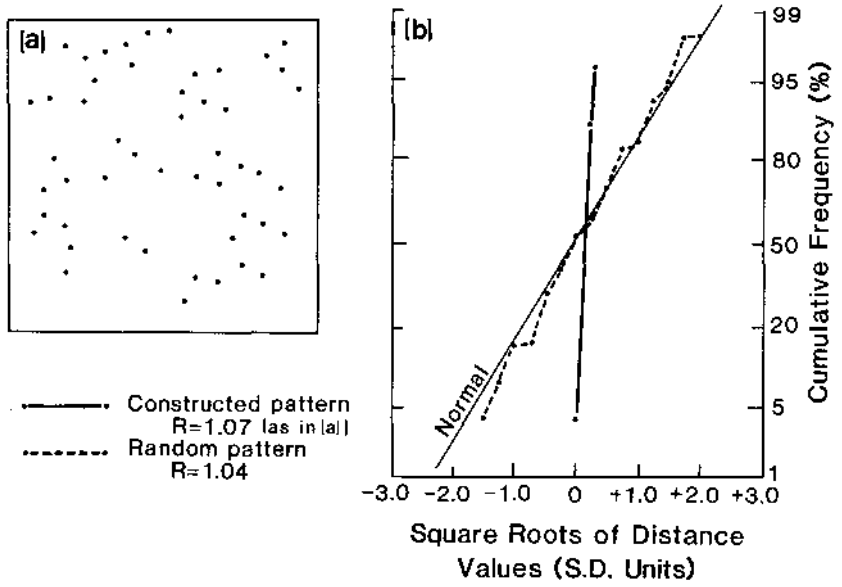


Figure 7. (a) Point pattern generated to give R(1) ≈ 1.0 and all r(1) values approximately equal to ρ (1); (b) cumulative frequency plots of $r(1)^{0.5}$ values for the pattern in (a) above, and for a random pattern.

pattern was purposely constructed to give a value of R(1) similar to that for random patterns, whilst having a cumulative frequency distribution of distances that is quite different. Although the R(1) values for the two patterns are almost identical, the two plotted distributions clearly indicate the differences between them. Most situations will not be as dramatic as this, but it is very often possible to go much further in terms of description and analysis if the cumulative frequency approach is used instead of, or in addition to, the R statistic.

V APPLICATIONS OF ORDER-NEIGHBOUR METHODS

(i) Introduction

Point pattern analysis has been used in many geographical research contexts. Phenomena that have been abstracted to points include such varied objects as drumlins (Trenhaile, 1971) and Romano-British walled towns (Hodder and Hassall, 1971). The majority of such studies have, however, dealt with either more modern human settlements (e.g. Dacey, 1962; Hamilton, 1971; Jensen-Butler, 1972; and King, 1962), or with the locational patterns

of given activities within urban areas (e.g. Sherwood, 1970 (grocery shops, and solicitors and accountants); Yeates, 1974 (various retail activities); Rolfe, 1965 (neighbourhood parks); and Pinder and Witherick, 1972 (schools and hat merchants)). Critiques of these varied uses are given below (Section V(iv)).

(ii) Selected urban activities, Melbourne

The following examples arise from an unpublished study by the author of the locational patterns of a number of activities in Sydney and Melbourne, Australia. The Melbourne study area included a large part of the Melbourne metropolitan area, whilst excluding surrounding non-urban land as much as possible. Whilst a cross-section of urban landscapes, from the central business district to newly developed outer suburbs, is included, a large urban district to the east and south-east is not. Any attempt to relate conclusions to the entire metropolitan area should, then, proceed cautiously. A simple, rectangular study area is used, although this restriction may not have been necessary (see Section III(ii)). Data relate to the situation in 1973.

The locations of activities in a number of categories were established from commercial and telephone directories, and with the aid of detailed maps. Each location was identified by a six-figure grid reference. Calculations were carried out by a computer, thus enabling large populations and a considerable number of different analyses to be handled with ease. Output from the program included the R(K) statistics and indices given in Table 7, as well as values for a number of other statistics not covered here. It also provided the basic cumulative frequency distribution data needed to draw graphs (Figure 9) and test hypotheses. The locations of three activities are discussed in the remainder of this section: government secondary schools, ladies' clothing manufacturers, and Roman Catholic churches (Figure 8).

The population for the analysis of government secondary schools includes all secondary schools operated by the Victorian Department of Education. There are, in fact, two parallel systems - the academically oriented high schools, and the more technically and vocationally oriented technical schools. This, in large part, explains the occurrence of the occasional closely spaced pairs of schools. Whether or not the two types of schools should be considered together is open to argument, but that issue will not affect the usefulness of this example as a demonstration of order-neighbour analysis.

Table 7 gives the results for the various analyses. It is not possible to give an unambiguous result in the case of R(3), as the value calculated could be from either a random or dispersed pattern (Section III(ii)). Both R(1) and R(2) fall well within the range delimited by the a 0.05 critical values for bounded, random patterns. All observed cumulative frequency distributions are entirely within the critical limits (a = 0.05) and, in fact, very close to the normal line. The K = 1 case is graphed in Figure 9(a): the K = 2 and K = 3 graphs are very similar. This range of analyses can only lead one to conclude that the pattern is a random one.

If the area were homogeneous, an intra-urban version of central place theory would lead to a hypothesis that the schools are regularly spaced. The analysis above leads to that hypothesis being rejected. One of the major reasons is, of course, that the study area is anything but homogeneous

Table 7. Analyses of three patterns of establishments in Melbourne, Australia

	Government secondary schools	Ladies' clothing manufacturers	Roman Catholic churches
N	88	94	97
R(1)	1.0678	0.2065	1.3545
H_0	Accept; α = 0.05	Reject; α = 0.01	Reject; α = 0.01
CFD(1)	Random	Clustered	Regular
H_0	Accept; α = 0.05	Reject; α = 0.01	Reject; α = 0.01
R(2)	1.1313	0.2483	1.1388
H_0	Accept; α = 0.05	Reject; α = 0.01	Accept; α = 0.05
CFD(2)	Random	Clustered	Random
H_0	Accept; α = 0.05	Reject; α = 0.01	Accept; α = 0.05
R(3)	*	0.2556	1.0733
H_0	*	Reject; α = 0.01	Accept; α = 0.05
CFD(3)	Random	Clustered	Random
H_0	Accept; α = 0.05	Reject; α = 0.01	Accept; α = 0.05

*R(3) value greater than mean expected for bounded, random distribution; unambiguous interpretation not possible (see Section III(vi)).

in terms of population density. In other words, the regular spacing model assumes that would-be 'customers' are uniformly distributed. It also assumes the schools are of equal size and that the bodies responsible for planning and locating schools have done their job and have been able to respond to demographic change and urban growth. The seemingly random location of schools is most likely a result of three factors acting together: heterogeneity in terms of the density of 'customers'; varying sizes of schools; and an inability to locate in an ideal, theoretically sound manner. In this particular case, the dual nature of the school system may also be important. The analysis does illustrate the value of a point pattern analysis in leading to further questions and, potentially, further hypotheses, once the original one (in this case, one of dispersion) has been rejected. If the theory is applicable, this result focuses attention on the imperfections in the ideal situation summed in the assumptions necessary for that theory to hold.

In the second example, the particular manufacturing establishments included were those producing ladies' suits and coats. Many geographers have commented on the highly concentrated nature of this industry. The point pattern representing their locations in Melbourne immediately confirms that they are similarly concentrated in that city. In fact, the points cannot be

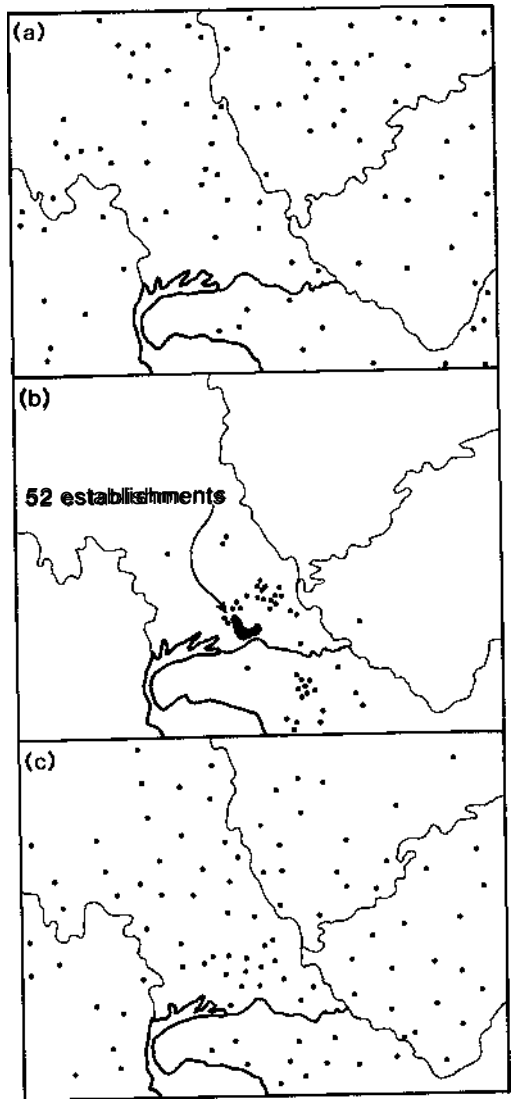


Figure 8. Location of (a) government secondary schools, (b) ladies' clothing manufacturers, and (c) Roman Catholic churches in Melbourne, Australia, 1973.

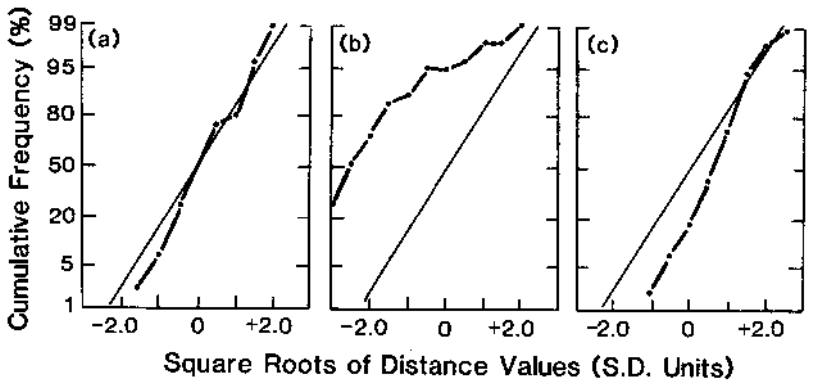


Figure 9. Cumulative frequency plots of $r(1)^{0.5}$ values for the three patterns of Figure 8

plotted as individuals at all within a small area of the inner city: there are 70 establishments in an area of approximately 6 km^2; 46 of those within an area of 1.25 km^2. There is also an identifiable secondary concentration further south. In a number of cases, firms even share premises, and hence, at the scale of any practicable study, have the same location. The hypothesis is thus one of a high degree of clustering.

All R(K) values are well below the lower critical values for randomness, indicating a high degree of clustering. Furthermore, Figure 9(b) shows a cumulative frequency distribution in which the bulk of the distance values are small: again, the K = 2 and K = 3 cases are similar. The obvious conclusion is that the manufacturers of ladies' suits and coats are strongly agglomerated. There is one question, however, that an analysis such as this cannot answer by itself: to what extent is this clustering due to the fact that manufacturing as a whole is concentrated in certain districts of Melbourne? One approach would be to undertake similar studies of other industries to compare them. Both foundries and plastics manufacturing in Melbourne are much less highly clustered, for example.

The locations of Roman Catholic churches are shown in Figure 8(c). Knowing that the population density in general, and that of Roman Catholics in particular, varies considerably from one part of the city to another, one cannot expect the pattern to be completely and regularly spaced. This heterogeneity, and an attempt to cater for the relevant population, leads one to expect clustering or, if the various factors interact in a certain way, randomness. It would again be useful to have some means of allowing for the heterogeneity of 'customer' densities, either by means of a map transformation (Tobler, 1963) or by some form of weighting of distance values. Both are far beyond the scope of this work.

As with the schools, a visual inspection of the map tells us little of the type of pattern present. Table 7 suggests mild dispersion at the K = I level, and randomness at the two higher levels. At the K = 2 level,the

Table 8. Rank ordering of urban activities, Sydney and Melbourne, 1973

Activity	Ranking (1 = most clustered)					
	Sydney			Melbourne		
	K = 1	K=2	K=3	K=1	K=2	K=3
Ladies' clothing manuf.	1	1	1	1	1	1
Foundries	2	2	2	4	3	3
Plastics manufacturing	3	4	3	3	4	4
Public hospitals	4	3	3	2	2	2
C.B.C. banks	5	5	5	5	5	5
Govt. secondary schools	6	7	7	7	9	10
Totalizator agencies	7	6	6	6	6	6
Police stations	8	8	8	8	7	7
Presbyterian churches	9	10	10	9	8	8
Roman Catholic churches	10	9	9	10	10	9

R(K) value is not far below the upper a = 0.05 critical value, but the R(3) value is less than the mean. It appears, then, that the tendency to dispersion becomes progressively weaker to the point of disappearing as higher neighbour levels are reached. The first level tendency to dispersion can be explained by the absence of the very small distances found in all random patterns. Both the map and Figure 9(c) show this. The churches have avoided locating too close to each other. In the case of the secondary schools, these small distance values were provided somewhat artificially by the adjacent locations of some pairs of high and technical schools. School locations may not, then, be as random as the pattern analysis seems to indicate. Despite the tendency to regularity in the locational pattern of the churches at the K = 1 level, the results as a whole show that the pattern is, in overall terms, more adequately described as random.

(iii) Comparative studies, Sydney and Melbourne

Raw R(K) values from different populations or different study areas cannot be directly compared as the strength of the boundary effect varies as N and d vary. Corrected values (Section III(iii)) do allow direct comparisons to be made, however, provided that the shapes of the study areas are acceptable (Section III(ii)). Table 8 gives the rank ordering of ten activities at each of three neighbour levels for Sydney and Melbourne, Australia. It can be seen that the activities do fairly consistently rank in the same general order on the clustering-random-dispersed continuum, as defined by values of the R statistic. A Spearman rank order correlation analysis confirms that the rank orders do not differ significantly between the two cities at any neighbour level.

In these two cities, the manufacturing industries studied are more clustered or agglomerated than any of the other activities except the public hospitals, particularly in Melbourne. Industry in general is located almost exclusively in designated industrial areas. The garment trade, in particular, is much more highly concentrated within very limited areas within these general industrial zones. Public hospitals have tended to concentrate in locations near the central business districts and in the older, more densely populated, and generally more accessible, inner suburbs. It is only in recent years that efforts have been made, especially in Sydney, to open large new hospitals in outer areas, even if this has meant closing some beds in the inner suburban hospitals.

Banks and Totalizator Agency Board branches - the commercial activities - form a second, less clustered group. Their locations are fairly evenly spaced throughout the suburban areas, but there are also obvious concentrations in the central business districts. A third group is comprised of service activities, police stations and churches, and these are somewhat more regularly located as they rely less on locations in commercial districts, and more on proximity to residential populations. Secondary schools are the one activity that fit this general categorization of service activities poorly: they show the widest variety in their rankings. It should be added, however, that none of these patterns shows a particularly strong tendency to regularity, no doubt because of the concentration of the urban population and economic activity into certain areas.

(iv) Other published studies

This section is a critique of a number of papers and sections of books that involve the use of order-neighbour analysis. It is by no means exhaustive, due to space criteria and to a desire to limit the treatment to relatively easily accessible material. Particular reference will be made to aspects of these studies that highlight or illustrate points made in previous sections.

Pinder and Witherick (1972) offer possibly the best short introduction to the technique available. As the title of their paper suggests, they deal with both the principles and the pitfalls of this method of analysis. In addition, they include four empirical analyses. Two of these, involving tors on the Bodwin Moor granitic boss and Junior schools in Southampton, illustrate applications in very different contexts. In both cases, it is hypothesized that the patterns will be dispersed. What if that is found to not be the case? 'Should the analyses reveal that the patterns are not uniform (i.e. dispersed), then clearly distorting forces have been at work, and it is these that require identification and evaluation' (p282). That is precisely the approach taken in the studies of functions in Melbourne (Section V(i)). In more general terms, the authors stress the value of order-neighbour analysis in helping pose further questions. The other two analyses concern hat materials merchants in Luton in 1914 and 1965, and is an example of temporal comparisons, even though no rigorous statistical approach is taken. This paper demonstrates the care needed in interpreting the results of order-neighbour analyses.

A considerable number of studies have involved settlement patterns; that of King (1962) was the earliest to become widely known. He sees the need to have a more precise connotation to descriptive terms such as

'dispersed' and 'agglomerated' and suggests order-neighbour analysis as a means of achieving that. He takes twenty areas within the U.S. with values of N ranging from 20 to 177 to give a series of settlement patterns in a cross-section of physical and economic regions. Cross-boundary measurements are allowed (see Section III(i)). Only the K = 1 cases are considered and the results are seen as being mainly descriptive, although he does suggest that one can bring 'powerful and sophisticated statistical techniques to bear' to explain variations in R(1) values. The ability to do this does not appear as great as King implies.

Barr et al (1971) apply the same approach to twenty sample areas in the U.S.S.R., but extend the analyses to the K = 6 level. The fact that some areas have values of N as low as 3, however, throws very considerable doubt on the reliability of some of their results (see Section II(iii)). Whilst admitting that the higher order results are difficult to interpret, they find that those results do give 'greater utility and power of description than statistics derived for first order distances alone' (p.222; c.f. Section II(vi)).

Such studies are extended by Jensen-Butler (1972) in that he deals with a number of levels in the settlement hierarchy. The claim that the nearest neighbour statistic can measure the extent to which a distribution of points approaches a uniform (equilateral triangular) distribution' (p.356) is not justifiable in that no statistically reliable method of doing so exists. It is more accurate, however, to claim that 'nearest neighbour analysis is essentially an exploratory technique' (p.359). This is illustrated in the same paper by the alternative hypotheses suggested as a result of the analyses carried out.

Although Hamilton (1971) is dealing with industrial location patterns in Eastern Europe, his analyses are of industrial towns rather than individual sites. Whilst his analyses of macro- and meso-scale regions involve 80 or more points in each case, the bulk of his micro-regions for 1945 have N < 5: he, in fact, gives an R(1) value where N = 2 and even quotes a result of R(1) = 0.0000 for N = 0. Such a tabulation of results shows a clear lack of understanding, as do two other aspects of the paper: first, no mention is made of the boundary problem; and secondly, his chosen regions are often across national boundaries and, although entirely within the Eastern Bloc, may be affected by important differences in national policy.

Two further studies provide a useful contrast. Hirst (1971), in a study of the administrative centres of Uganda, displays a similar lack of concern for both small N values and the boundary problem. In marked contrast, Hammond and McCullagh (1978; 275-7), in their example of order-neighbour analysis involving the settlements of Norfolk and Suffolk, give a very thorough treatment of the boundary problem and N is large.

A second group of studies involves a range of functions and their locations within urban areas as, indeed, do three of the examples given by Pinder and Witherick (1972: see above). The earliest study in this group was that of Getis (1964). He analyzed the locational patterns of grocery stores in Lansing, Michigan, and the manner in which they changed over time. His use of order-neighbour analysis involved two variations on the method outlined in Section 2: the three nearest neighbours are measured in three sectors (see Dacey and Tung, 1962); and he deals with the boundary problem

by defining circular study areas well within the total urban area and measuring across the boundary where necessary. He also states that, following from central place theory, we would expect the grocery store location pattern to approximate the population density characteristics' (p.392: c.f. Section V(ii)). He had, in an earlier paper (Getis, 1963), attempted to allow for variations in population densities by means of a map transformation.

Sherwood (1970) builds on Getis' work by analyzing patterns of a number of activities in Shrewsbury and by extending the analysis to the K = 3 level. This paper contains an interesting use of comparisons between different neighbour levels in an attempt to analyze the scales of clustering present (c.f. Section II(vi)). He also compares functions with each other and over time by means of a simple ranking system (c.f. Section V(iii)). This technique may not warrant his description of it as a 'rigorous descriptive tool', however, unless one is aware of its many pitfalls.

Uses of order-neighbour analysis in physical geography are harder to find, largely, perhaps, due to the obvious areal extent of most features at the usual scales of investigation. Indeed, the most trenchant criticism of those studies that have been published is that they tend to ignore the areal extent of the individual objects. Whilst settlements may have a negligible areal extent when mapped at the national or regional level, that can hardly be said to be true of the volcanic craters of the Bunyaruguru volcanic field of West Uganda, as studied by Tinkler (1971). His Map A shows craters with considerable areal extents and, in fact, almost covering parts of the study area. They certainly cannot be accepted as capable of legitimate abstraction as points (see Section I). Similar criticisms can be levelled at the studies of drumlins that have employed order-neighbour analysis (Smalley and Unwin, 1968; Trenhaile, 1971). In Smalley and Unwin, Figure 4 (p.384) and Figure 6 (p.386) both show that the drumlins do have considerable areal extent at the scale of the study and in relation to the study area. Assuming that drumlins are not superimposed on other drumlins, much of the total area cannot, then, contain points, thus contravening the basic definitions of randomness (Section II), the form of point pattern on which the technique is ultimately based.

We have thus, in a sense, come full circle. What, then,of the uses made of order-neighbour analysis by its originators, the plant ecologists? In fact, they have made relatively little such use; at least, little that has been reported in the readily available literature. Why might that be so? They do not seem to have positively rejected order-neighbour analysis, but it is rather that quantitative ecological work, as reflected in such works as Greig-Smith (1964) and Kershaw (1964), has concentrated on quadrat analysis. Quadrat surveys have long been a field investigation method of ecologists, whereas concern with distance as such has been relatively minor. They have, furthermore, been frequently concerned with attempting to model empirical distributions with mathematically defined probability generating functions. Despite the shortcomings of quadrat analysis, and despite Dacey's attempts to link probability functions with distance approaches, they have found the former to be more amenable to this type of study.

VI CONCLUSION

(i) Summary of order-neighbour method

Previous sections have dealt with the order-neighbour approach to point pattern analysis in considerable detail. They have highlighted its strengths as well as its weaknesses. They have also contained proposals for extensions to the basic technique that may help to overcome some of the weaknesses and to add new strengths. It is appropriate to bring this material together as a general overview of the method as treated in this work.

Order-neighbour analysis involves two stages: the measurement or calculation of point-to-point distances; and a small number of relatively straightforward arithmetic calculations based on them. Neither stage is at all difficult, but the former can be decidedly tedious and time consuming if N is large and it is not possible to use a computer. Even mini-computers can handle large numbers of such repetitive calculations quickly and easily, as long as their memory capacity is not exceeded. Once the measurements have been made, values for the density, for the expected mean distance, and for the parameters that characterize random patterns are easily found using the equations given in Section II. So are the critical values needed to test the hypothesis of randomness. It is no more difficult to carry out these steps for higher neighbour levels, although interpretation of the results is less straightforward. While it is considered not justified to proceed beyond the third neighbour level in this work, an extension of the order-neighbour method to the second and third neighbour levels is often a most rewarding one in that it provides valuable additional information of patterning at different scales and with different degrees of generalization.

The theoretical derivation of this approach is based on the assumption, amongst others, of an unbounded point population. If a boundary is present, and it does limit distance measurements in some directions, then the effect of that limitation must be taken into account. In some circumstances, it may be feasible to delimit a 'buffer zone' (Section II(v)). If not, allowance for the boundary effect must be made in some other way. Section III(ii) introduces equations and tables of critical values that are derived from a large number of generated, bounded random patterns. If both the observed R(K) value and the expected critical values relate to bounded patterns, then the comparisons used in hypothesis testing are again valid. Alternatively, Section III(iii) gives a series of equations that allow empirical R(K) values from bounded study areas to be corrected for the likely boundary effect. In this case, theoretically derived, unbounded parameter and critical values are used in reaching conclusions.

Whenever an analysis or description is ultimately dependent on a single R(K) value at each neighbour level, a great deal of information has been, in a sense, passed over. We did, after all, measure or calculate many more than one distance value. A given mean distance value, r(K), can arise from any number of different distributions of r(K) values. Section IV raises the possibility of directly comparing the entire distribution of distance values with an expected distribution. Such an approach often gives valuable additional insights into the nature of point patterns.

Perhaps the greatest shortcoming of order-neighbour analysis, given the inherent interest of geographers in non-random patterns, is its inability to

deal readily with questions relating to the type and degree of non-randomness present. Like all distance-based methods, the R statistic approach gives, above all else, a test for randomness. It does not, in anything like the same sense, give a test for clustering or a test for dispersion. Somewhat more detail as to the form of non-randomness present can be obtained from cumulative frequency distributions, however.

(ii) Comparisons between patterns

Most of the work in order-neighbour analysis has concentrated on the description and analysis of individual point patterns. Whilst many of the papers listed in Part C of the Bibliography do consider more than one pattern, comparisons between patterns are almost always on the simplest possible level. In other words, the authors find R statistic values for a number of patterns representing different phenomena, different areas, and/or different times, and then proceed to compare the values without applying any statistical tests. Such comparisons lead to rather subjective interpretations and conclusions. There are, however, at least two more direct, and statistically more rigorous, approaches to comparisons between pairs of patterns.

Work on methods of comparing patterns point by point, i.e. on measuring the point by point correlation between the locations of points of one population with those of points of another, could fill a volume the size of this one; and that is despite the fact that much further work needs to be carried out before such comparisons can be said to be based on established, proven techniques. The two methods introduced below rely on comparisons of the overall point patterns as characterized by either the R(K) values or the cumulative frequency distributions of distance measurements.

First, the 14(K) values may be compared, and the null-hypothesis of no statistically significant difference between the two values can be tested. This test depends on the applicability of the two-sample z-test or t-test outlined in almost all elementary statistical texts. Unfortunately, the circumstances in which it is applicable are limited. The distribution of R(K) values from random patterns for a given value of N is normal. Each R(K) or R(K) value is the mean distance value divided by a constant, p (K). Hence, the common tests for significant differences between sample means can be applied. They cannot, however, be validly applied to patterns exhibiting an extreme degree of non-randomness. Both Clark and Evans (1954) and Thompson (1956) suggest the use of similar z-tests in this context, but only for unbounded patterns. It must be remembered that the strength of the boundary effect changes with variations in population size, and thus more meaningful comparisons can be achieved when such differentials are allowed for. It is for this reason that it is suggested that the corrected R(K) values be used.

A second approach is to apply a two-sample Kolmogorov-Smirnov goodness-of-fit test to a comparison between two r(K) or r(K) ^{5} cumulative frequency distributions. The latter will often be more convenient as relevant values will have already been calculated for the individual pattern analyses. This approach has the decided advantage of being applicable regardless of the type of pattern involved, as the underlying statistical test is independent of the type of distribution taken by the distance values. The two sample version of the Kolmogorov-Smirnov test is not as simple to apply, or as well known, as the one sample version: Siegel (1956) gives a valuable summary of

its use. Valid application of this test presupposes standardization of the distributions of distance values using standard deviation units.

(iii) Comparisons with other methods and conclusions

A perfect method of point pattern analysis does not exist. All methods have their strengths and weaknesses. One might be forgiven for thinking, after reading in the general subject area, that the weaknesses greatly outnumber the strengths. It is certainly true that too much has been claimed for many of the techniques, and that the strengths have often been enthusiastically extolled while the weaknesses have been played down or ignored. Point pattern analysis techniques, like many others, were sometimes used rather recklessly during the heady early days of geography's 'quantitative revolution'. But there is no need to throw them away completely because they are imperfect. As long as their basic assumptions and their definite limitations are understood, they are still very often valuable tools.

Quadrat analysis is one of two major approaches to point pattern analysis: it is covered fully in another CATMOG (Thomas, 1976). One fundamental, and unsolved, shortcoming of that approach is that the size of quadrat used affects the results. For example, the same pattern may be differently classified (using the clustered, dispersed, random classification) if small quadrat squares are used than if large ones are substituted. Unless some objective solution to the problem of choice of quadrat size can be found, the validity and reliability of any method based on cell counts must remain in doubt. Some useful suggestions have been made, but the overall problem is far from solved. A second major problem is that this method is based on a knowledge of how many quadrats have 1, 2, 3, ... n points in them, but does not, in any way, take into account how the cells are distributed with relation to each other. In other words, all the cells with large numbers of points may be grouped together on one side of the study area, or they may be scattered throughout it. Whilst distance measurements may likewise be distributed over the study area in different ways, the problem is less acute than with quadrats, and is very considerably reduced if second and third neighbour measurements are used.

The terms 'distance-based methods' and 'nearest-neighbour analysis' both include a range of techniques, a series of variations on a theme. Unfortunately, no authoritative studies have been carried out to compare these various approaches. Some comments have been made, however, although without any supportive evidence. Dacey and Tung (1962, 86), for example, compare the order-neighbour approach of Section II with Dacey's (1962) regional neighbour approach which is based on the distance to the nearest neighbour in each of six sectors around each origin point. They conclude that the latter seems to have a greater power efficiency for discerning random and uniform (dispersed) patterns, whilst the former is more efficient in detecting clustering. Quite apart from the inefficiency of using two different sets of measurements in each study, there would seem to be no sound basis on which to choose the more complicated regional method over, or as well as, the order method. The former also involves one additional, subjective choice on the part of the researcher, that of the orientation of the sectors. It may be informative to note that virtually all of Dacey's own published work at later dates uses the order-neighbour method. Whilst there are many other methods of point pattern analysis, they are all beyond the scope of this publication and comparisons with them are not included.

If this outline has succeeded in its purpose, it should be perfectly clear by now that the order-neighbour method is far from perfect. It is not universally applicable. It cannot answer all the myriad questions that researchers may wish to ask about point patterns. It is not completely objective, nor does it always give unambiguous results. It does require considerable tedious measurement, unless you can use a computer to take the tedium out of the exercise. But, despite all this, it can offer valuable insights into the nature of point patterns.

BIBLIOGRAPHY

A. General

Aplin, G.J. (1973), Distance methods of point pattern analysis. *Monash Publications in Geography,* 7, 1-20.

Curry, L. (1964), The random spatial economy: an exploration in settlement theory. *Annals, Association of American Geographers,* 54, 138-146.

Dacey, M.F. (1973), Some questions about spatial distributions. In: *Directions in geography,* (ed) R.J. Chorley. (Methuen, London).

Eliot Hurst, M.E. (1973), Establishment geography: or how to be irrelevant in three easy lessons. *Antipode,* 5(2), 40-59.

Getis, A. & B. Boots (1975), *Models of spatial processes* (Cambridge Geographical Studies No. 8, (Cambridge U.P., London), 26-33.

Gould, P. (1970), Is statistix inferens the geographical name for a wild goose? *Economic Geography,* 46, 439-448.

Haggett, P., A.D. Cliff & A.E. Frey (1977), *Locational analysis in human geography 2: Locational Methods,* (Arnold, London).

Hammond, P. & P.S. McCullagh (1978), *Quantitative techniques in geography: an introduction.* Second Edition (Clarendon Press, Oxford), 270-278.

King, L.J. (1969), *Statistical analysis in geography.* (Prentice-Hall, Englewood Cliffs, N.J.).

Pinder, D.A. & M.E. Witherick (1972), The principles, practice and pitfalls of nearest neighbour analysis. *Geography,* 57, 277-288.

Silk, J.A. (1979), *Statistical concepts in geography.* (Allen & Unwin, London).

Vos, S. de (1973), The use of nearest neighbour methods. *Tijdschrift voor Economische en Sociale Geografie,* 64, 307-319.

Yeates, M. (1974), *An introduction to quantitative analysis in human geography.* (McGraw-Hill, New York).

B. Theoretical

Aplin, G.J. (1979), Nearest neighbour cumulative frequency distributions. *Area,* 11, 102-106.

Clark, P.J. & F.C. Evans (1954), Distance to nearest-neighbour as a measure of spatial relationships in populations. *Ecology,* 35, 445-453.

Cowie, S.R. (1969), The cumulative frequency nearest neighbour method for the identification of spatial patterns. *Seminar Paper, Series A,* 10, University of Bristol, Department of Geography.

Dacey, M.F. (1960), A note on the derivation of nearest neighbour distances. *Journal of Regional Science,* 2, 81-87.

Dacey, M.F. (1962), Analysis of central place and point patterns by a nearest neighbour method. In: *Proceedings of the I.G.U. Symposium in Urban Geography,* (ed) K. Norberg. (Lund studies in geography, Series B, 24; Gleerup, Lund).

Dacey, M.F. (1964), Two-dimensional random point patterns: a review and an interpretation. *Papers & Proceedings of the Regional Science Association,* 13, 41-55.

Dacey, M.F. (1965), Order distance in an inhomogeneous random point pattern, *Canadian Geographer,* 9, 144-153.

Dacey, M.F. & T.Tung (1962), The identification of randomness in point patterns, I. *Journal of Regional Science,* 4, 83-96.

Diggle, P.J. (1979), On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics,* 35, 87-101.

Donnelly, K.P. (1978), Simulations to determine the variance and edge effect of total nearest-neighbour distance. In: *Simulation studies in archaeology,* (ed) I. Hodder. (Cambridge University Press, London).

Gardiner, V. & G. (1978), *Analysis of frequency distributions.* Concepts and techniques in modern geography, 19. (Geo Abstracts Ltd., Norwich).

Greig-Smith, P. (1964), *Quantitative plant ecology* (2nd edition). (Butterworths, London).

Ebdon, D. (1976), On the underestimation inherent in the commonly-used formulae. *Area,* 8, 165-169.

Harvey, D.W. (1968), Some methodological problems in the use of the Neyman Type A and the negative binomial probability distributions for the analysis of spatial point patterns. *Transactions of the Institute of British Geographers,* 45, 71-78.

Hope, A.C.A. (1968), A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B,* 30,582-598.

Hsu, S. & J.D. Mason (1972), A comparison between the Poisson and the simple random distribution with respect to the nearest neighbour technique in map analysis. In: *International Geography 1972* (22nd International Geographical Congress). (University of Toronto Press, Toronto).

Jones, A. (1971), An order neighbour approach to random disturbances on regular point lattices. *Geographical Analysis,* 3, 361-378.

Kershaw, K.A. (1964), *Quantitative and dynamic ecology.* (Arnold, London).

Pielou, E.C. (1969), *An introduction to mathematical ecology.* (Wiley, New York).

Siegel, S. (1956), *Nonparametric statistics for the behavioral sciences.* (McGraw-Hill, New York).

Thomas, R.W. (1976), *An introduction to quadrat analysis.* Concepts and techniques in modern geography, 12. (Geo Abstracts Ltd., Norwich).

Thompson, H.R. (1956), Distribution of distance to nth neighbour in a population of randomly distributed individuals. *Ecology,* 37, 391-394.

Tobler, W.R. (1963), Geographic area and map projections. *Geographical Review,* 53, 59-78.

C. Applications

Barr, B.M., I.J. Lindsay & E.R. Reinelt (1971), Patterns of urban spacing in the U.S.S.R.: analysis of order neighbour statistics in two-dimensional space. *Journal of Regional Science,* 11, 211-224.

Getis, A. (1963), The determination of the location of retail activities with the use of a map transformation. *Economic Geography,* 39, 14-22.

Getis, A. (1964), Temporal land-use pattern analysis with the use of nearest-neighbour and quadrat methods. *Annals, Association of American Geographers,* 54, 391-399.

Hamilton, F.E.I. (1971), The nearest neighbour statistic: an application to the spatial distribution of industry in Eastern Europe. In: *Festschrift for A.E.F. Moodie,* (ed) F.E.I. Hamilton. (Northwestern University, Studies in Geography, 18, 79-110).

Hirst, M.A. (1971), The changing pattern of district administrative centres in Uganda since 1900. *Geographical Analysis,* 3, 90-98.

Hodder, I. & M. Hassall (1971), The non-random spacing of Romano-British walled towns. *Man,* 6, 391-407.

Jensen-Butler, C. (1972), Nearest neighbour analysis of a central place system. *Tijdschrift voor economische en sociale geografie,* 63, 353-359.

King, L.J. (1962), A quantitative expression of the pattern of urban settlements in the United States. *Tijdschrift voor economische en sociale geografie,* 53, 1-7.

Rolfe, E. (1965), Analysis of the spatial distributions of neighbourhood parks in Lansing: 1920-60. *Papers of the Michigan Academy of Science,* 50.

Sherwood, K.B. (1970), Some applications of nearest neighbour techniques to the study of the movement of intraurban functions. *Tijdschrift voor economische en sociale geografie,* 61, 41-48.

Smalley, I.J. & D.J. Unwin (1968), The formation and shape of drumlins and their distribution and orientation in drumlin fields. *Journal of Glaciology,* 7, 377-390.

Tinkler, K.J. (1971), Statistical analysis of tectonic patterns in areal volcanism: the Bunyaruguru volcanic field of West Uganda. *Mathematical geology,* 3, 335-355.

Trenhaile, A.S. (1971), Drumlins: their distribution, orientation and morphology. *Canadian Geographer,* 15, 113-126.