Second-Order Neighborhood Analysis of Mapped Point Patterns

Chapter *in* Ecology · August 2010

DOI: 10.1007/978-3-642-01976-0_7

2 authors:

Arthur Getis
San Diego State University

Janet Franklin
University of California, Riverside

Ecology, 68(3), 1987, pp. 473–477

SECOND-ORDER NEIGHBORHOOD ANALYSIS OF MAPPED POINT PATTERNS[1]

ARTHUR GETIS
Department of Geography, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801 USA

AND

JANET FRANKLIN
Department of Geography, University of California, Santa Barbara, California 93106 USA

Abstract. A technique based on second-order methods, called second-order neighborhood analysis, is used to quantify clustering at various spatial scales. The theoretical model represents the degree of clustering in a Poisson process from the perspective of each individual point. The method is applied to point location data for a sample of ponderosa pine (*Pinus ponderosa*) trees, and shows that heterogeneity within the forest is clearly a function of the scale of analysis.

Key words: clustering; heterogeneity; Pinus ponderosa; *point patterns; Poisson process; scale; second-order neighborhood analysis; spatial pattern.*

INTRODUCTION

In any study where spatial data or pattern analyses are required, the appropriate scale for analysis must be chosen. The choice is often arbitrary. Scale is usually defined as the ratio of map distance to the real world distance it represents (Robinson et al. 1984). As scale changes, so does the level of resolution, and new spatial patterns emerge. Theory or subject matter should guide the selection of an appropriate scale, but often researchers need to look at pattern at a number of scales. Spatial pattern has both intensity, the extent to which density varies in space, and grain, the distance over which density is perceived to vary (Pielou 1977:155–156).

Workers in a number of disciplines have attempted to find methods for identifying parameter changes that take place when scale is made to vary. Perhaps *blocking*, or contiguous quadrat analysis, is the most common method used for examining grain of pattern. The study area is covered by an array of N quadrats. These quadrats are combined into larger quadrats in a systematic way. Various guidelines have been proposed for selecting the optimum quadrat size, the most common of which is an analysis of variance (Moellering and Tobler 1972, Grieg-Smith 1983).

Spectral analysis, as a method for selecting scales, has been satisfactorily applied to studies where blocking or hierarchically defined units are used. In addition, data transects have been studied as continuous spectra for scale effects (for a review of the literature see Ripley 1981). Rayner (1971), following the lead of Bartlett (1950), demonstrates how pattern may be studied using two-dimensional spectra. Although more complicated than the technique discussed here, an important advantage is that it allows for an assessment of the effect of orientation on pattern and scale.

While one of the authors was engaged in a remote sensing study of canopy reflectance for a ponderosa pine (*Pinus ponderosa*) forest (Franklin et al. 1985), we developed a technique for describing both intensity and grain of tree spatial patterns simultaneously at a number of scales. The method, second-order neighborhood analysis, is a variation on second-order analysis of point patterns (Ripley 1977, 1981, Diggle 1983, Getis 1984). Second-order analysis is designed to test randomness hypotheses, often based on the Poisson distribution, by examining the proportion of total possible pairs of points in Euclidean space whose pair members are within a specified distance of each other. The analysis is second order because it is the variation rather than the mean of distances that is being studied. The technique discussed below, while similar to second-order analysis, differs in that consideration is given only to those pairs of points having as one of its members a given point i. This method depends on relatively large amounts of digitized point data, from aerial photographs or maps, where coordinates can be accurately recorded.

THE MODEL

Getis' model (1984) has the form

$$\hat{L}_i(d) = \left[A \sum_{j=1}^{n} k_{ij} / \pi(n-1) \right]^{1/2} \qquad (1)$$

where Σk_{ij} is the summation over all points that are within distance d of point i, and it includes a boundary correction where required. If for a given neighborhood point j the specified distance d is more than the distance

[1] Manuscript received 15 May 1986; revised 12 October 1986; accepted 20 October 1986.

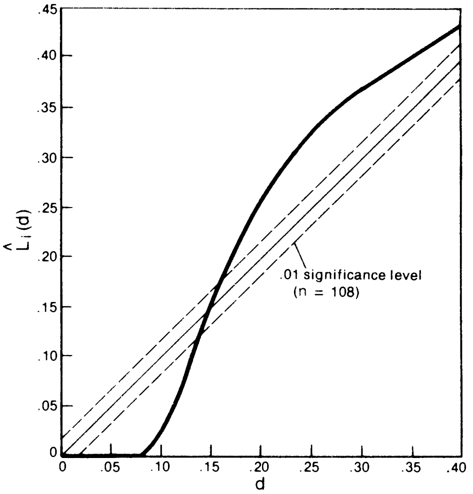

FIG. 1. Cumulative distribution curve (heavy line) of $\hat{L}_i(d)$ for hypothetical tree in a square of area 1. $L_i(d)$ is the number of points within distance d of point i corrected for the boundary effect, and scaled such that $L_i(d) = d$ when $L_i(d)$ represents a pattern produced by a Poisson process in the plane. Dashed lines represent .01 significance levels around the line representing Poisson process.

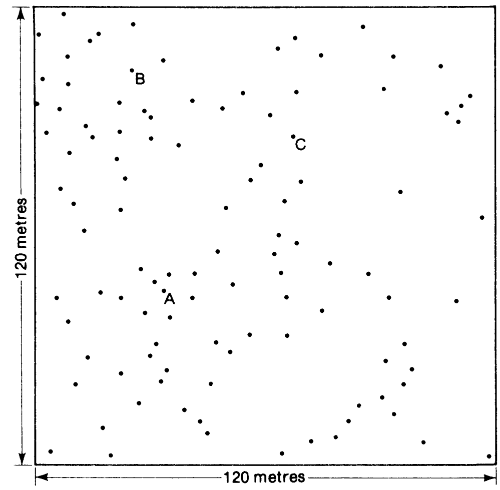

FIG. 2. Point pattern representation of tree locations in the study area. The letters A, B, and C mark particular individual trees, which are referred to in Fig. 4. North is up.

between i and j, then the pair (k_{ij}) counts as 1 (unless the boundary correction is required); otherwise k_{ij} counts 0. The value A is the area of a rectangular region, and $n - 1$ represents all possible pairs of points having i as a pair member. Taking into consideration the circular area centered on point i, for convenience, π and the square root are included in order to make $L_i(d)$ linear with respect to d and to have $L_i(d) = d$ when $L_i(d)$ represents a pattern produced by a Poisson process in the plane.

The boundary correction is as follows: if the distance between i and j is greater than the distance between i and the nearest boundary (e_1), instead of the value 1 for k_{ij}, substitute

$$k_{ij} = [1 - \cos^{-1}(e_1/d)/\pi]^{-1}. \quad (2)$$

If the distance between i and j is greater than the distance to both of two boundaries (e_1, e_2), use

$$k_{ij} = \{1 - [\cos^{-1}(e_1/d) + \cos^{-1}(e_2/d) + \pi/2]/2\pi\}^{-1}. \quad (3)$$

The boundary correction is based on the assumption that the region outside of the boundary in the vicinity of the distance measurement has a spatial pattern similar to the nearby areas within the boundary. If this assumption cannot be accepted, then results must be exclusively for the areas within A greater than d from all boundaries (see Getis 1984 for further discussion of the boundary problem).

The form of the analysis can best be depicted by a diagram. Fig. 1 shows a curve describing the typical values of $\hat{L}_i(d)$ for a given i in a somewhat clustered forest in a square of area 1. The horizontal axis represents d; that is, at any distance from a tree designated as i we can identify an $\hat{L}_i(d)$ value. The diagonal represents $L_i(d)$ values for a pattern that is created by a Poisson process. The initial part of the curve for $\hat{L}_i(d)$ displays a value of 0 as far as distance 0.08. This means that no other tree is within 0.08 of tree i, and so this is the nearest neighbor distance. Up to distance 0.14 from i, the curve remains below the expectation. The fact that the curve for the observations is advancing upward at a faster rate than the theoretical curve implies a tendency for clustering or heterogeneity. It is not until 0.16 from i that one can say that the spatial distribution of pairs displays a statistically significant level (0.01) of clustering (see below). At 0.28, the curve reaches its maximum above $L_i(d)$, implying maximum clustering. In summary, the parameters representing i's relationship with all j are (1) the nearest neighbor distance, (2) the distance at which heterogeneity begins, (3) the distance at which clustering becomes statistically significant, and (4) the distance at which maximum clustering can be observed.

Statistical significance can be ascertained either by simulation or by accepting the values $\pm 1.42\sqrt{A}/(n - 1)$ and $\pm 1.68\sqrt{A}/(n - 1)$ as reasonable approximations of the 5% and 1% significance points, respectively. These are a modification of Ripley's (1978, 1979) estimates for the second-order case.

In addition to the above indicators of the relationship of i to all j are the scale parameters. If we identify the $\hat{L}_i(d)$ value at certain specified distances, say 0.05, 0.10, 0.15, 0.20, for each i, we are then able to compare the spatial *situation* of each tree. One tree may display a high $\hat{L}_i(d)$ value at 0.05, implying that a number of neighbors are close by, while a second tree may have a low $\hat{L}_i(d)$ value at 0.05 but a high value at 0.20. The second tree is much less crowded by near neighbors,

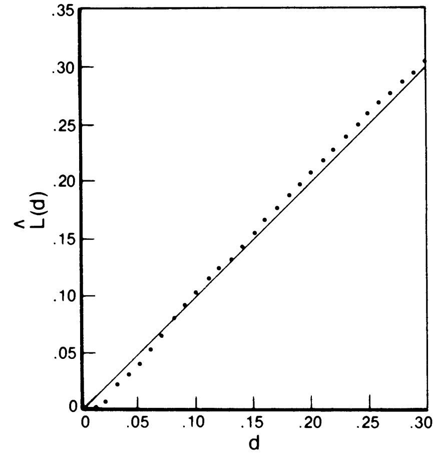

FIG. 3. Values for $\hat{L}(d)$ over the range $.01 \le d \le .30$. $L(d)$ is the number of points within distance d of all points i corrected for the boundary effect, and scaled such that $L(d) = d$ when $L(d)$ represents a pattern produced by a Poisson process in the plane. $\hat{L}(d)$ may be interpreted as the average for all 108 points (from Fig. 2) taken together. Solid line shows expected values given a Poisson distribution. Solid dots show observed values.

but is within a cluster of trees at a distance of 0.20 from it. If the chosen scale of analysis were 0.05, the first tree would be considered a member of a cluster, but the second tree would not. The distance chosen represents the scale at which one can view pattern.

To demonstrate the method, ponderosa pine tree distribution was analyzed. The locations of ≈5000 ponderosa pine trees in the Klamath National Forest in Northern California were determined from United States Forest Service aerial photographs (nominal scale 1:24 000); trees <2.5 m apart were not resolvable, nor were small trees within the canopy of another tree. These points were digitized for automatic analysis (Franklin et al. 1985). A subarea selected for study, 120 × 120 m, included 108 trees that visually display nonrandom characteristics: clumps and clusters of trees appear to dominate the pattern (Fig. 2).

RESULTS

Fig. 3 shows the observed and expected $L(d)$ values; $\hat{L}(d)$ represents the average distance relationships for the 108 trees in the subarea shown in Fig. 2. For convenience the study area was made equal to 1; thus a distance of 0.01 is equivalent to 1.2 m. All data points on Fig. 3 are within the 95% confidence region of the Poisson expectation. This implies that although there are clusters of points and an apparent inhibition effect, the overall pattern cannot be differentiated from one created by a Poisson spatial process.

Fig. 4 contrasts the pattern membership characteristics of three selected trees, labeled A, B, and C in Fig. 2. Note that tree A appears to be a member of a small cluster of trees. Fig. 4a shows a short nearest neighbor distance ($0.03 = 3.6$ m), a rapid rise to clustering status at a distance of 0.03 (3.6 m), and maximum clustering at 0.09 (10.8 m). Visually, point B does not appear to be a member of a cluster, but inspection of Fig. 4b

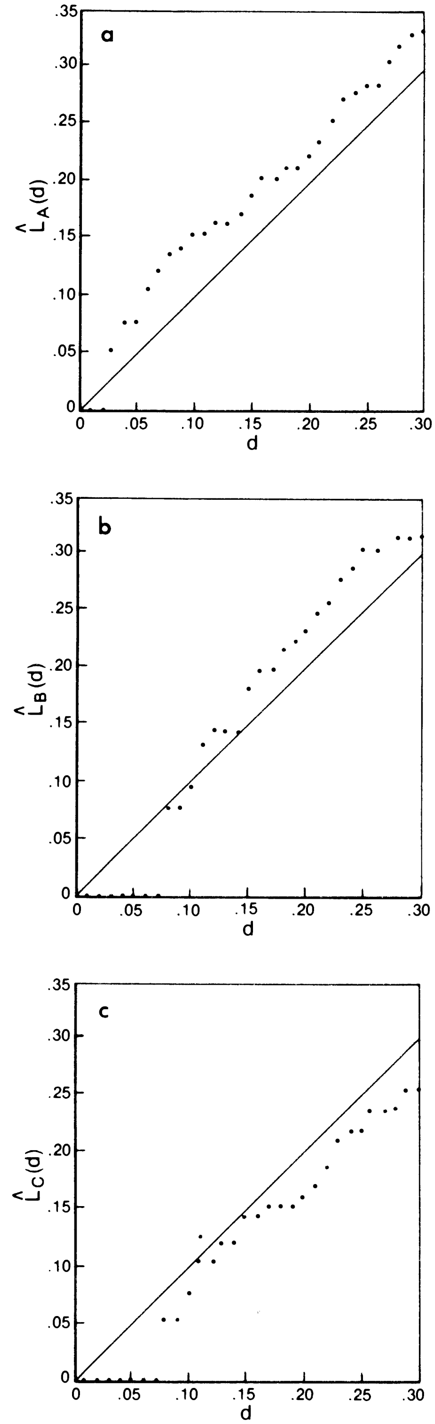

FIG. 4. Values for $\hat{L}_i(d)$ over the range $.01 \le d \le .30$ when i = A (Fig. 4a), B (Fig. 4b), and C (Fig. 4c). The locations of trees A, B, and C are shown in Fig. 2.

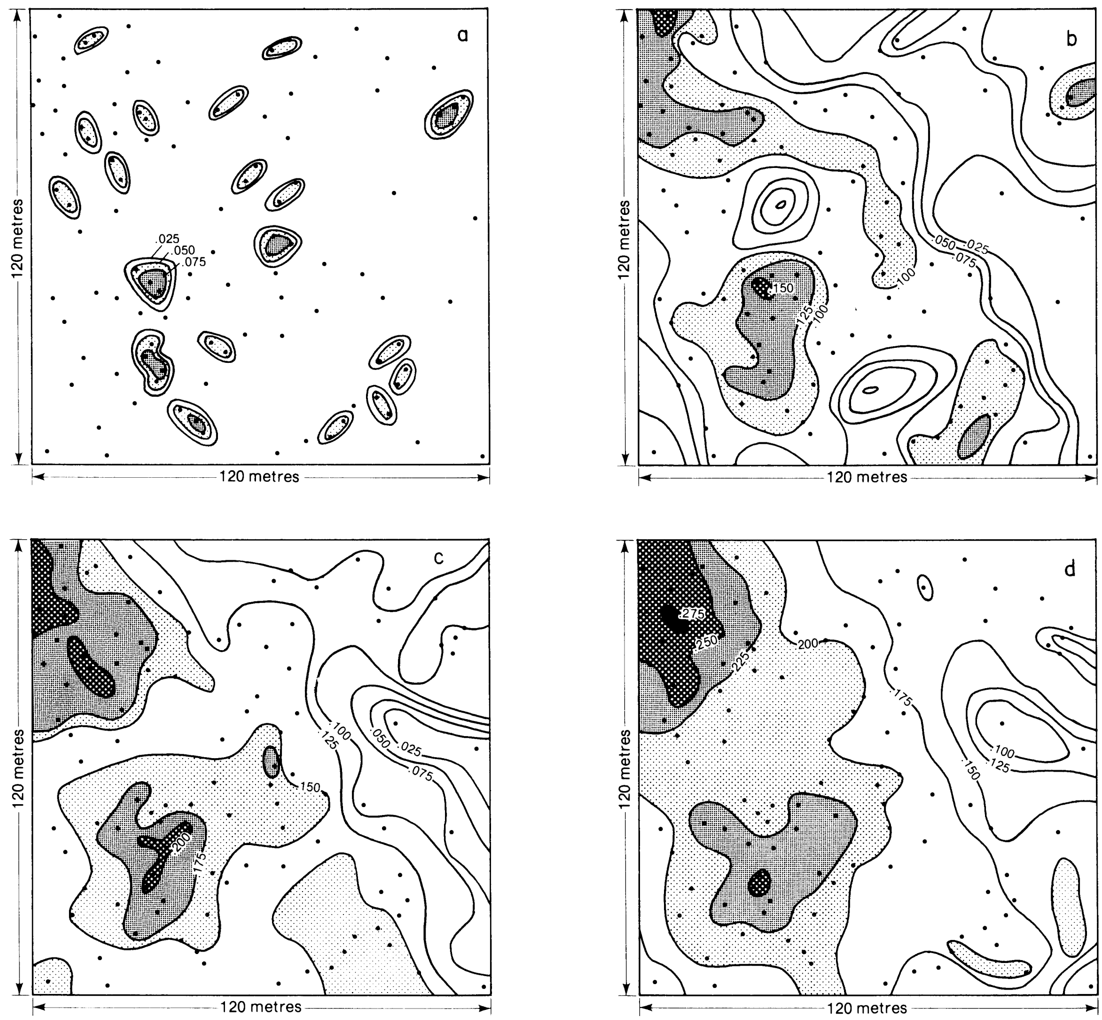

FIG. 5. Pattern created by assigning to each tree its $\hat{L}(d)$ value for the following values of d: (a) 0.05 (6 m), (b) 0.10 (12 m), (c) 0.15 (18 m), (d) 0.20 (24 m). The shaded areas contain trees that have $\hat{L}_i(d)$ values above the Poisson expectation. Intensity of shading corresponds to increase in tree density at a given scale. Isolines are in units of d.

reveals that B is a member of a cluster at distances of 0.11 (13.2 m) and greater. The distance at which maximum clustering takes place (0.25), however, is much greater than for point A. Point C is within an area of the forest where densities are much lower than is true of either A or B. Its $\hat{L}_i(d)$ values, shown in Fig. 4c, reveal that it is not a member of a cluster at most scales.

Fig. 5 shows the pattern created by the trees in our sample for scales (d) of 0.05 (6 m), 0.10 (12 m), 0.15 (18 m), and 0.20 (24 m). Of course, a much finer group of scales could have been selected. The isolines, drawn at intervals of 0.025, indicate areas of greater or lesser tree densities. Accuracy in drawing the isolines was enhanced by the addition of control points to the empty or sparsely vegetated areas; no measurements were taken to a control point. The shaded areas contain all trees that display $\hat{L}_i(d)$ values above the expected, that is above 0.05, 0.10, 0.15, and 0.20, respectively. Intensity of shading corresponds to increases in tree density at the given scale.

Comparisons can be made among all areas of the maps or only among the areas unaffected by the border correction. For the entire area of each map in Fig. 5 and by casual inspection of Fig. 2, it is clear that the higher densities are generally in the west and the lower densities in the east. Fig. 5 reveals, however, a number of further interesting contrasts. For example, note that at a scale of 0.05 (Fig. 5a), only trees within 6 m of one another are considered as members of clusters, so that large areas of Fig. 5a have a relatively low density of trees. When the scale is increased to 0.10 (Fig. 5b), some of the clusters identified at the 0.05 level are now considered part of larger clusters or are not part of any cluster at all. By contrasting the 0.05 (6 m) and the 0.10 (12 m) levels, one can see that geographic interpretations would be greatly different due to the scale chosen. In addition, note that the relatively low-density area in the western half of Fig. 5a becomes part of a clustered region when the scale is increased to 0.20 (Fig. 5d).

The variance about the observed mean, $\bar{L}_i$, for a particular d indicates the extent of the heterogeneity within the pattern. The scale at which the variance is maximized will show the greatest contrast in pattern. This may be a reasonable choice for an investigation when no other information is available to indicate an appropriate scale. In our example, the variance reaches its first peak at 0.15 (18 m), decreases, and then increases to a maximum at 0.35 (42 m) before decreasing again. The border correction contributes greatly to the creation of the second peak.

Conclusions and Discussion

We have shown that second-order neighborhood analysis can identify different dominant patterns at different scales for mapped point data. In the example given, the overall pattern cannot be differentiated from one created by a spatial Poisson process, but close inspection of the spatial relationships of individual trees to nearby trees reveals noteworthy variations. The influence of nearest neighboring trees dominates the pattern at or below a scale of 6 m. At 12 m, clustering is seen, but it is stronger at 18 m. From 24 to 42 m the effect of the border correction appears to play a role in intensifying the clustering (for example, near the northwest border).

Second-order analysis identifies several important scales of pattern: (1) the distance to nearest neighbor, (2) the distance where heterogeneity begins, (3) the distance where clustering becomes significant, and (4) the distance where maximum clustering is observed. The technique presented here, neighborhood analysis, can be applied to selected individuals, and maps of pattern density at a given scale can be produced. A knowledge of the scale-dependent spatial setting of individuals would be useful in testing neighborhood models of population dynamics and competition (Weiner 1984, Pacala and Silander 1985).

This study of scale pertains specifically to points each valued ostensibly as 1. It requires only a slight modification in our model, however, to place interval scale values at each point, such as size of a tree (see Getis 1984). In addition, if data were given for units having areal extent (nonpoint data), such as a lattice of quadrats, the analysis could be carried out if the researcher assigns the data values to points representing each sample area.

Acknowledgments

This study was partly supported by a grant from the National Science Foundation (SES-8219170). The authors appreciate the comments of W. R. Tobler and the referees.

Literature Cited

Bartlett, M. G. 1950. Periodogram analysis and continuous spectra. Biometrika **37**:1–16.

Diggle, P. J. 1983. Statistical analysis of spatial point patterns. Academic Press, London, England.

Franklin, J., J. Michaelsen, and A. H. Strahler. 1985. Spatial analysis of density dependent pattern in coniferous forest stands. Vegetatio **64**:29–36.

Getis, A. 1984. Interaction modeling using second-order analysis. Environment and Planning A **16**:173–183.

Grieg-Smith, P. 1983. Quantitative plant ecology. Third edition. Blackwell Scientific, Oxford, England.

Moellering, H., and W. R. Tobler. 1972. Geographical variances. Geographical Analysis **4**:34–50.

Pacala, S. W., and J. A. Silander, Jr. 1985. Neighborhood models of plant population dynamics. I. Single-species models of annuals. American Naturalist **125**:385–411.

Pielou, E. C. 1977. Mathematical ecology. Wiley-Interscience, New York, New York, USA.

Rayner, J. N. 1971. An introduction to spectral analysis. Pion, London, England.

Ripley, B. D. 1977. Modelling spatial patterns (with discussion). Journal of the Royal Statistical Society, Series B **39**:172–212.

———. 1978. Spectral analysis and the analysis of pattern in plant communities. Journal of Ecology **66**:965–981.

———. 1979. Tests of 'randomness' for spatial point patterns. Journal of the Royal Statistical Society, Series B **41**: 368–374.

———. 1981. Spatial statistics. Wiley, New York, New York, USA.

Robinson, A. H., R. D. Sale, J. L. Morrison, and P. C. Muehrcke. 1984. Elements of cartography. Fifth edition. Wiley, New York, New York, USA.

Weiner, J. 1984. Neighborhood interference amongst *Pinus rigida* individuals. Journal of Ecology **72**:183–197.