MEASURES OF SPATIAL PATTERN FOR COUNTS

Joe N. Perry

*Department of Entomology and Nematology, Rothamsted Experimental Station, Harpenden,
Hertfordshire, AL5 2JQ, UK*

Abstract. SADIE (Spatial Analysis by Distance IndicEs) is a new methodology to detect and measure the degree of nonrandomness in the two-dimensional spatial patterns of populations. It applies the same principles to data in the form of maps as to data in the form of counts at specified locations, but with different techniques. This paper considers data in the form of counts such as occur commonly in ecology. For such data the method has an advantage over traditional approaches that measure only statistical variance heterogeneity, because all the spatial information in the sample is used. Two indices and associated tests are reviewed, one based on the total distance of the sample from a completely regular arrangement, the other from a completely crowded arrangement. A new diagnostic plot is presented to aid interpretation. Results from some artificial data are studied to survey the properties of both indices for defined patterns of clustering. Indices based on the distance to regularity are powerful at detecting aggregation when several clusters are present; those based on the distance to crowding have the power to detect aggregation only when a single cluster is present. Methods are presented to estimate the typical cluster size and intercluster distance, suitable for data from sample units in the form of a contiguous grid. Examples are given for cyst-nematode field data and plant virus disease.

Key words: aggregation; clusters; heterogeneity; patchiness; randomness; regularity; SADIE; spatial pattern; uniformity.

INTRODUCTION

In ecology it is often difficult to study the movement of individual animals directly, especially for small, numerous animals such as insects. Ecologists have therefore studied the spatial pattern of individuals of a particular species to infer the underlying behavioral rules that govern their movement (Greig-Smith 1952, Lloyd 1967, Kennedy 1972, Taylor 1986). The spatial heterogeneity that results from nonrandom interactions between individuals, from both inter- and intraspecific behavior, tends to stabilize ecological systems (Hassell and May 1973). Spatiotemporal dynamic ecological models (Czárán and Bartha 1992), such as cellular automata (Hassell et al. 1991) and metapopulation models (Hanski and Gilpin 1991, Perry and Gonzalez-Andujar 1993, Perry 1994), increasingly use space explicitly, to locate and move individuals within a two-dimensional coordinate system. Wiens (1989) stressed the importance of spatial scale and emphasized that an ecological process that operates in a certain way at one scale may not operate in the same way, or at all, at a different scale (Heads and Lawton 1983). This paper is about the quantification of spatial pattern for data in the form of counts of animals or plants at specified locations. These counts might, for example, be of beetles sampled by pitfall traps, or of the number of diseased plants in a quadrat. This paper reviews and extends previous work (Perry and Hewitt 1991, Perry

1995*a*, *b*, 1996) on SADIE (Spatial Analysis by Distance IndicEs).

The basis of SADIE is to quantify the spatial pattern in a sampled population by measuring the total effort (in terms of distance moved) that the individuals in the observed sample must expend to move to extreme arrangements, in which the individuals in the samples are either spaced as uniformly (regularly) or are as aggregated (crowded) as possible. The degree of nonrandomness within a set of data is quantified by comparing the observed spatial pattern with rearrangements in which the sampled counts are randomly redistributed among the units. Perry and Hewitt (1991) introduced an index based on crowding for data in the form of a grid of counts, and noted two advantages: a biologically more intuitive index than traditional mathematically based ratios involving sample variance and mean, and increased power due to the greater use of the spatial information in the sample. Alston (1996) argued that distance to regularity provided a better basis for an index, because of the failure of the crowding index to detect multiple clusters, clumps, or patches. Perry (1995*a*) demonstrated the use of distance to regularity for data in the form of counts at specified locations, not necessarily on a grid, and exposed some subtle effects of scale. He showed how to distinguish nonrandomness in the form of statistical heterogeneity from spatial nonrandomness. The former arises from skewness in the frequency distribution of counts and subsequent departures from a Poisson distribution; the latter arises from aggregation of those counts into clus-

ters or from regularity of the counts. Perry (1995*b*) extended the use of the distance-to-regularity index to mapped data, where the two-dimensional coordinates of each individual in the sample are known explicitly. He introduced two diagnostic plots as aids to interpretation and a new method to estimate the initial focus of a cluster. Once again, scale was shown to be an important determinant of overall spatial pattern. Perry (1996) presented an algorithm that would allow a set of given counts to be distributed amongst a set of given sample units to achieve any given degree of spatial pattern, where possible. He noted the importance of constraining the simulated arrangements to allow for the degree to which the observed counts occur towards an edge of the sample area.

This paper studies the behavior of two indices, first for artificial and then for field data: the index based on distance to regularity, I_a, studied by Perry (1995*a*); and an index, J_a, based on distance to crowding, different from that proposed by Perry and Hewitt (1991). In addition, a means to identify and quantify the scale and characteristics of the dominant clusters is offered for the situation when the units form a contiguous rectangular grid. The artificial data are used to provide a baseline library of several simple known patterns, against which to study the behavior of the indices. The field data are then used to test the ability of the indices to expose the main facets of the observed patterns. A new diagnostic plot is introduced to aid interpretation.

Definitions and Notation

The notation follows that of Perry (1995*a*). The data are assumed to be a set of counts of individuals, with one count in each of $i = 1, \ldots, n$ sample units. The two-dimensional position (X_i, Y_i) of the ith sample unit and its associated count, N_i, is assumed to be known. The distance to regularity, D, is the minimum value of the total distance that the individuals in the sample would have to move, from unit to unit, so that all units contained an identical number of individuals. The solution, which may involve a fractional final number in each unit, concerns the optimal way in which individuals would move from each unit with an initial count larger than the mean, to other units with initial counts smaller than the mean. It is provided easily by the transportation algorithm (see especially Kennington and Helgason (1980) for a comprehensive treatment and for the FORTRAN algorithm NETFLO) from the operational research literature. If the observed counts are then randomly permuted between sample units, so that the resulting sample is a simple rearrangement of the original, then P_a represents the proportion of randomized samples with distance to regularity as large as, or larger than, the observed value, D. Intuitively, a large value of D would be expected to imply an aggregated or clustered, i.e., spatially heterogeneous, pattern and, conversely, a small value of D to imply a regular, i.e., spatially uniform, pattern. A value of P_a

derived from a sufficiently large number of randomizations provides a formal test of randomness (Perry 1995*a*); the null hypothesis of spatial randomness may be rejected, for example, if $P_a < 0.025$ (in favor of the alternative hypothesis of aggregation), or if $P_a > 0.975$ (in favor of the alternative of regularity) giving the usual 5% probability of rejecting the null hypothesis when it is true. If the arithmetic mean distance to regularity for the randomized samples is denoted as E_a, then the index of aggregation, I_a, is defined as $I_a = D/E_a$. Usually, an aggregated sample is indicated if $I_a > 1$, a spatially random sample if $I_a = 1$, and a regular sample if $I_a < 1$. A total of 2000 randomizations are used for most of the tests in this paper, which should prove sufficient for the derivation of index values; estimation is anyway more important than hypothesis testing (Perry 1986). The quantity I_r (Perry 1995*a*) is no longer thought useful, so is not considered in this paper.

Let C denote the distance to crowding, the minimum value of the total distance that individuals in the sample must move so that all are congregated in one unit. This value is found more readily than D, by using a simple direct search over all the units; the unit with the minimum value being termed the ''focus'' for crowding. Random permutations of the observed counts, as above, lead to a proportion, denoted Q_a, of those permutations with distance to crowding as small as, or smaller than, the observed value, C. Intuitively, for data that comprise a single cluster, a small value of C would be expected to imply a spatially aggregated pattern; conversely a large value of C might imply a spatially regular pattern. Analogously to the above, the null hypothesis of randomness may be rejected if $Q_a < 0.025$ (in favor of the alternative of aggregation) or if $Q_a > 0.975$ (in favor of the alternative of regularity), and if the average distance to crowding for the randomized samples is denoted as F_a, then the index of aggregation, J_a, is defined as $J_a = F_a/C$. As for the index I_a, described above, values of $J_a > 1$ usually indicate an aggregated sample, $J_a = 1$ is expected for spatially random data and $J_a < 1$ for a regular sample. Earlier simulations indicated that values of I_a and J_a for randomized counts were uncorrelated, so the software employed the same randomizations to compute both these indices. Before exemplifying the use of the two indices for field data, which exhibit a complex, multi-scaled pattern, in the next section I present a study of the behavior of these indices for simple patterns using artificial count data. Although the sample units in all the examples given fall on a square grid, this is not essential for the SADIE techniques; units may occupy any location.

Index Behavior for Artificial Data

In ecology, invertebrate or plant data in the form of counts often show clustering, as with the single almost circular cluster with ''diameter'' ~13 units, depicted in the 15 × 15 grid of 225 sample units ($f = 15$) in
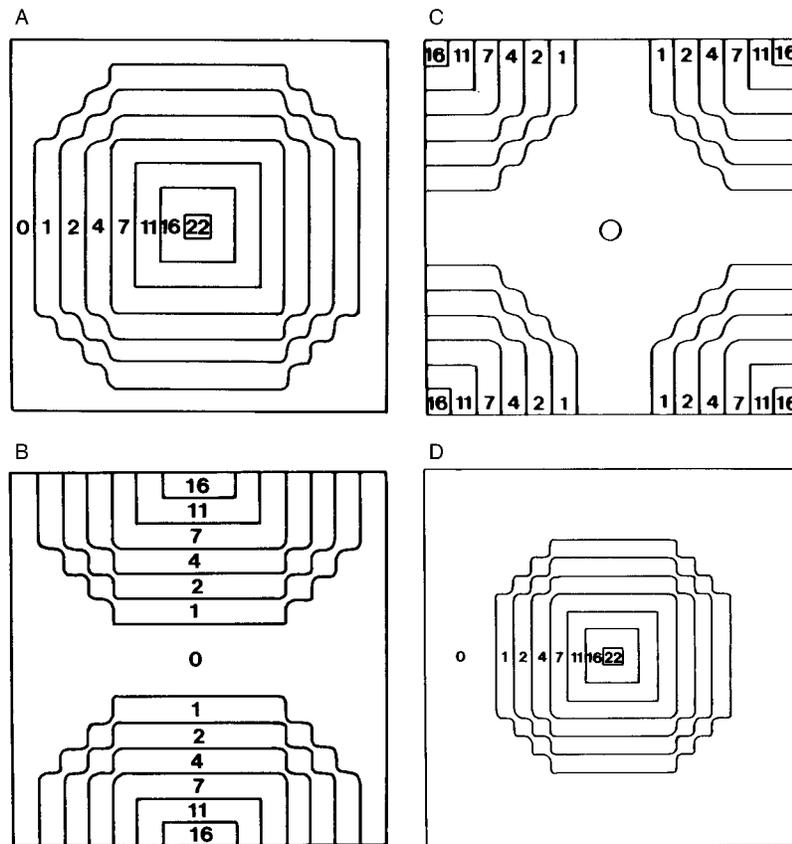
FIG. 1. Artificial data from a square mosaic of clusters of diameter roughly 13 units, where cluster density declines monotonically from cluster centers; the count in each sample unit is 22, 16, 11, 7, 4, 2, 1, or 0 individuals; clusters are separated by gaps of g units and sampled by an $f \times f$ square contiguous grid comprising f^2 sample units. (A–C) $g = 3$, $f = 15$; position of the sample area within the mosaic determines whether one, two, or four clusters are sampled, respectively; (D) $g = 9$, $f = 21$; the cluster is more isolated, but otherwise identical to that in part A.

Fig. 1A. This artificial cluster was formed from an imaginary sample in which 22 individuals were counted in the central sample unit of this grid, 16 individuals in each of the eight units immediately surrounding it, a further "ring" of 11 individuals in each of the sixteen units surrounding those, and so on, with population density declining monotonically from the center so that successive rings contained units with 7, 4, 2, and 1 individuals, until beyond the edge of the cluster there was a roughly ring-shaped area comprising eighty units with a count of zero. Further, suppose that outside this grid there is an unsampled area in which the occurrence of this cluster is repeated at regular intervals; that such clusters are separated by a "gap" of length g, exactly three units both horizontally and vertically in space, to give an overall pattern comprising an infinite square mosaic of clusters and gaps (Pielou 1964); and that, purely for convenience, the sampling grid has the same orientation as this mosaic. The example in Fig. 1A represents one of the 11% of cases when a sample grid with $f = 15$, placed randomly on this mosaic, contains units from one and only one cluster. Equally occasion-

ally, the random placement of such a grid on such a mosaic will sample units from each of two clusters, as in Fig. 1B, but in over three-quarters of cases there will be four clusters involved, as in Fig. 1C. The behavior of the indices will be compared for these three different sampling positions of the grid, where the underlying pattern of the mosaic remains the same. Suppose now that the characteristics of the clusters themselves were identical, but that the intercluster gap within the mosaic was increased in length from $g = 3$ to $g = 9$ units. When the sampling area was increased from $f = 15$ to $f = 21$, i.e., to a 21×21 contiguous grid of 441 units, the analogous situation to that in Fig. 1A, where a single cluster was sampled, is now shown in Fig. 1D. In Fig. 1D the relative size of the cluster to the sample area is smaller, but otherwise directly comparable to that in Fig. 1a. (The effect of a similar cluster, displaced towards an edge, will be considered in a separate paper.)

For the single clusters (a and d) both SADIE indices indicated the presence of a significantly aggregated pattern (Table 1), with values well above unity and, as

TABLE 1. Indices and their associated probabilities of aggregation for artificial data in Fig. 1A–D, comprising a square mosaic of clusters of diameter roughly 13 units, separated by gaps of g units, sampled by a square contiguous grid of length f units.

Set	f	g	I_a	P_a	J_a	Q_a
a	15	3	2.69	<0.0005	1.81	<0.0005
b	15	3	2.67	<0.0005	0.92	0.9935
c	15	3	2.65	<0.0005	0.71	>0.9995
d	21	9	3.17	<0.0005	2.54	<0.0005

expected, greater aggregation for the relatively more ''isolated'' cluster in Fig. 1D. Other simulations, not described in detail here, demonstrated that the indices indicate still greater aggregation for yet larger values of f and g, and less aggregation when f was reduced below 13 so that the sample was restricted to the area within the cluster.

For two (b) and four (c) clusters, whilst there was virtually no change in I_a, the value of J_a fell sharply to 0.92 and 0.71, respectively (Table 1), as noted by Alston (1996), and the probability Q_a falsely implied a regular rather than an aggregated pattern. The reason for this is that when data comprise a single cluster only, the individuals within the sample would have a relatively short distance to move to the focus of crowding, which is invariably within the cluster itself. However, when two or more clusters are present, either the focus occurs somewhere between them (as for sets b and c, here) and all individuals would have to move a relatively large distance, or it occurs in one of the clusters, in which case all individuals from the other clusters would have to move a relatively large distance to it; either way the value of the distance to crowding, C, is inflated.

Both indices, except J_a for multiple clusters, were very powerful at detecting the degrees of aggregation simulated. In other simulations, not detailed here, J_a proved more powerful than I_a for single clusters where the cluster was less well-defined.

The different behavior of the indices for several clusters may be used to advantage in the analysis of field data, particularly when there is a need to characterize the typical size of clusters within a set of data in which pattern may occur at several spatial scales. For example, consider the effect on the indices of randomly sampling an idealized mosaic such as that described above, with a steadily increasing sample area f^2. Whatever the precise values of cluster size, or intercluster gap, g, which determines the exact nature of the mosaic, the relationships found above may be qualitatively approximated by typical curves (Fig. 2). When the sample area is very much less than the cluster size, the sample is most likely to be either entirely within a cluster or entirely outside it, and both crowding and regularity indices are fairly small. As the sample area increases, at the point labeled ''a'' in Fig. 2; there is a reasonably large probability that a randomly placed sample area will contain only part of a cluster, and overlap one cluster boundary. Beyond point ''a,'' as this probability increases, both crowding and regularity indices increase, albeit at possibly different rates. Eventually the sample area achieves a size where, at point ''b,'' a randomly placed area will be likely to include parts of two clusters; here the value of I_a reaches a maximum, whereas the value of J_a declines very sharply, to near unity or even below it. With a further increase in size, and a correspondingly greater expected number of clusters per sample area, at point ''c,'' J_a decreases yet further, although I_a is unaffected. By inspecting graphs
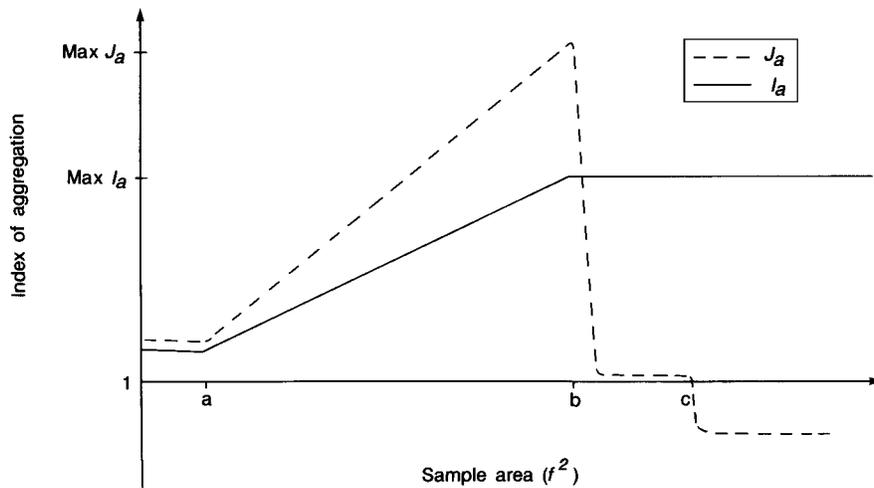
FIG. 2. Approximate qualitative relationship between the indices of aggregation, I_a and J_a, for the idealized patch–gap mosaic (see Fig. 1) and f^2, the size of the square contiguous grid of sample units: up to point ''a'' most samples are either entirely within or entirely outside a cluster, between points ''a'' and ''b'' most samples contain one cluster, between points ''b'' and ''c'' two clusters, and beyond ''c'' more than two.

of I_a and J_a, plotted against sample area for a set of data resampled by various grid sizes, it is possible to infer typical cluster sizes and intercluster distances from the positions of observed points ''a,'' ''b,'' and ''c.'' Examples will be given for two sets of nematode data, described below.

Having obtained baseline data concerning the behavior for artificial data with known patterns, ecological field data will now be analyzed to illustrate interpretation featuring the simultaneous use of both indices. For the first time, presence/absence data, a special case of counts, will be analyzed using SADIE.

MATERIALS AND METHODS

Field data

The first two sets of data concern counts of two pest cyst-nematode species, collected by B. Boag (Webster and Boag 1992). In two fields in Scotland, an area of 1 ha was sampled by taking soil cores on a 15×15 square grid at intervals of 7.1 m; the cysts were separated from each 200-g sample of soil and those with viable eggs counted. *Heterodera avenae* was sampled at Invergowrie (Fig. 3A) and *Globodera rostochiensis* at Drumkilbo (Fig. 3B). Webster and Boag (1992) were concerned to quantify the patchiness, to estimate cluster size, and to determine whether there was more infestation in the centers of clusters than near their margins.

When sample units form large contiguous rectangular grids of $m \times n$ units, such as these sets of nematode data, a hierarchical scheme (Bliss 1941, Greig-Smith 1952, Mead 1974) may be employed to investigate change of pattern with spatial scale and to find the dominant cluster size, as described above. As an example, for both sets of nematode data, the 15×15 grid was sampled firstly as a whole, but then additionally using successively smaller sample areas. For example, it is easy to see that there are nine separate ways in which a 13×13 continuous subgrid may be imposed on the overall data. In general, there are $(n - f + 1)^2$ different ways in which an $f \times f$ contiguous subgrid may be imposed on an $n \times n$ grid. Of course, when each of these possible $f \times f$ subgrids is used once, some sample units, especially those closest to the center of the 15×15 grid, contribute to more than one subgrid. The $(n - f + 1)^2$ separate values of I_a, J_a, P_a, and Q_a obtained are therefore not independent. The median values of these indices and probabilities were selected to be representative of the pattern at each of the $f \times f$ scales examined. All integer values of f from $f = 2$ up to $f = n$ could be studied, but $f = 2$ was deemed too small to give useful information; here $f = 3, 5, 7, 9, 11, 13,$ and 15 were used. Furthermore, the number of possible subgrids for small values of f may be prohibitively large to study comprehensively; for example, for $n = 15$ and $f = 5$ there are 121 possible subgrids, and for $n = 15$ and $f = 3$ there are 169. For this reason,

A) *Heterodera avenae*

0	0	0	1	0	0	1	1	1	0	1	0	2	0	10
0	0	0	0	1	0	2	1	7	2	0	0	1	0	0
0	0	0	3	11	3	1	0	28	16	0	0	1	31	22
0	0	0	2	9	39	24	9	24	11	1	0	0	0	16
0	0	0	0	15	3	15	12	46	7	0	1	0	0	0
0	0	0	0	7	19	18	3	2	2	0	0	0	0	5
0	0	0	0	4	1	0	1	0	0	6	0	2	3	2
0	0	0	0	13	3	3	0	2	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
7	0	1	0	3	0	15	4	4	3	0	0	0	0	0
0	2	0	2	8	16	23	18	5	0	5	0	6	1	0
0	2	0	2	12	0	13	6	0	1	0	32	12	42	0
0	0	0	1	5	21	27	4	25	2	0	12	9	1	0
0	1	0	1	7	22	19	15	13	0	4	0	0	0	18

Position number (vertical axis), *Position number* (horizontal axis)

B) *Globodera rostochiensis*

20	26	35	12	20	9	15	7	5	2	3	11	11	12	8
22	21	24	11	19	9	12	9	4	3	1	5	5	10	7
16	18	9	6	14	18	20	14	11	2	5	3	13	7	10
9	14	19	10	12	6	14	6	8	5	4	1	7	8	12
17	8	7	10	6	15	8	27	10	6	16	16	13	25	41
17	23	16	12	13	23	19	13	10	9	7	7	5	17	14
34	47	28	7	10	36	29	11	9	6	4	6	16	7	5
18	12	22	19	17	17	10	8	12	21	12	7	8	6	10
6	10	14	21	21	11	13	6	8	17	16	10	7	7	11
15	20	26	17	8	13	7	10	4	6	10	8	12	7	6
11	13	5	9	10	7	4	5	4	3	5	5	4	6	3
7	10	8	14	26	16	17	7	6	6	2	2	1	4	7
8	8	7	4	2	3	6	7	2	3	1	5	8	11	4
11	3	18	9	4	9	7	2	13	5	6	4	7	23	6
11	6	18	14	8	7	10	4	1	4	11	15	8	24	20

Position number (vertical axis), *Position number* (horizontal axis)

FIG. 3. Counts of cyst nematodes, collected by B. Boag, from Webster and Boag (1992). (A) *Heterodera avenae* at Invergowrie ($m = 4.15$, $s^2 = 64.9$); (B) *Globodera rostochiensis* at Drumkilbo ($m = 11.0$, $s^2 = 55.4$).

for the *Heterodera avenae* data only nine 5×5 subgrids were employed, placed so as not to overlap with each other, and nine 3×3 subgrids, placed so that their centers coincided with those of the 5×5 subgrids; the information within a single 3×3 grid is relatively small. The median values of I_a and J_a, and of P_a and Q_a, were then plotted against f, as in Fig. 2, to estimate the dominant cluster size and intercluster distance.

Using the process of partitioning (see next section, *A new diagnostic plot*), it was possible to identify the
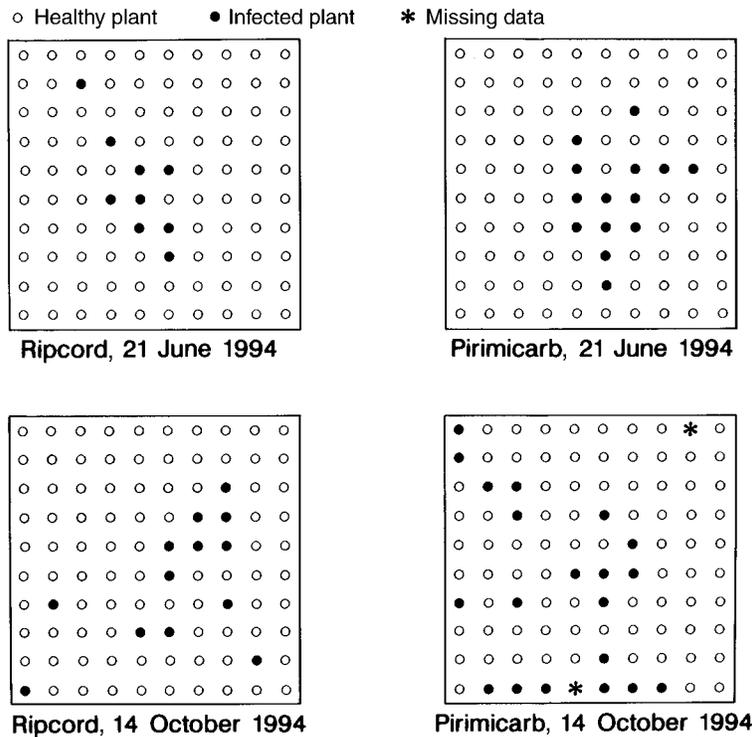
FIG. 4. Positions of virus-infected pepper plants on a 10×10 grid with two insecticide treatments on two replicate dates (*unpublished data* of J.-L. Collar and A. Fereres).

units that comprised each major cluster together with their associated surrounding units, for each set. For each such individual cluster, its ''center'' was defined as that unit that was the focus for crowding, after reduction of the sampled data to that cluster alone. Then, the relationship between the geometric mean of the counts at a given distance, say, s, from the cluster center, and that distance, s, was studied to determine whether the margins of a cluster were less infested than the center, a question posed by Webster and Boag (1992).

The third set of data concerns the presence or absence of the virus disease Potato virus Y (PVY) (pepper strain, 0 pathotype) on pepper plants *Capsicum annuum*, spread by viruliferous *Myzus persicae* aphids, collected by J.-L. Collar and A. Fereres (*personal communication*). Healthy plants were arranged symmetrically in a 10×10 grid in an indoor cage and the aphids were released from the center of the cage. The plants were treated with insecticide prior to aphid release, and assessed for virus 30 d after release. One replicate was done on each of two dates, in each of which one of two insecticides, Pirimicarb and Ripcord (a.i. Cypermethrin) was applied (Fig. 4). Since the virus was either present or not, the counts are either zero or one. The SADIE analysis proceeded exactly as for unconstrained counts. The unknown rate of aphid dispersal from the cage center was assumed to be related to two observable variables. Firstly, to the concentration of disease at the grid center, measured by the degree of patchiness of the virus infection. Secondly, to the relationship between disease incidence and distance from the grid center, measured by the steepness of the regression slope, b, of logit(p), where p is the proportional disease incidence, on distance from the grid center, and by the significance of this regression (F-statistic on 1, 13 degrees of freedom).

A new diagnostic plot

The new diagnostic plot presented here is the version, for counts, of the ''initial and final'' (IAF) plot of moves to regularity described for mapped data by Perry (1995*b*). Briefly, the output from the transportation algorithm gives the optimum number of individuals required to move from each of the cells with initially more individuals than the sample mean, to cells with initially fewer, to achieve regularity. When such flows are plotted, possibly excluding relatively trivial flows, a good visual impression is gained of clusters in the data and, conversely, of less dense areas. As for mapped data (Perry 1995*b*) such flows are shown graphically as lines radiating out from units within clusters to fill initially sparse areas, but now with no relationship between the length of a line and the position of the corresponding unit with respect to its cluster center, and with no overlapping of lines. An example of such a plot for *Heterodera avenae,* where units with inflows are indicated by an asterisk (Fig. 5A), showed
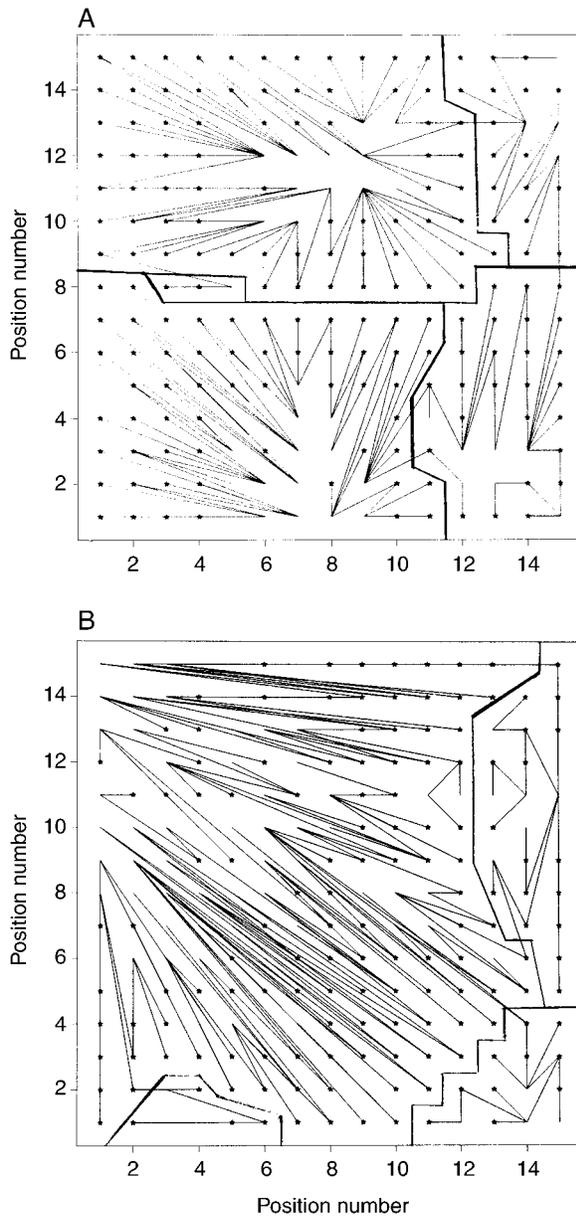
FIG. 5. (A) Initial and final (IAF) plot for Boag's cereal cyst-nematode data of Fig. 4A, for all flows, with destination units indicated by stars. Polygonal lines show the heuristic partitioning of the sample units into five sets, corresponding to the major clusters in the data. (B) As for (A) but for Boag's potato cyst-nematode of Fig. 3B, partitioned into four sets.

clearly the presence of two large, equal-sized central clusters, one at the top and one at the foot of the sample area, and gave a good indication of two further smaller clusters, one each to the right of the larger ones. These smaller two clusters appeared to extend beyond the boundary of the sample area. The original data (Fig. 3A) confirmed this visual impression readily. Notice that units with particularly large counts were identified by the emanation of several lines. Also, since the max-

imum flow between any two units must be equal to the sample mean, m, minus the minimum count, and since the minimum count here, as in most animal data sets, was zero, the maximum flow was also $m = 4.15$. Fig. 3A does not distinguish between the strengths of the different flows, which vary both between and within units. However, in principle, flow strength could be represented graphically by using different thicknesses of lines, by labeling, or by plotting only a subset of flows, say those within a certain range.

It is also possible to partition heuristically the sample units into a small number of sets according to the IAF plot, such that almost all the inflow units in a set receive flows from outflow units that are in the same set, and such that the sets correspond in general location with the major clusters and their close vicinity. One such partition into five sets is shown by the polygonal lines imposed upon the IAF plot of Fig. 5A. In only five units was there any ambiguity in this partition of the grid; e.g., the unit at (11, 3) receives flows from units at (9, 2) and (12, 2) that are in different sets. Such inflow units were categorized as belonging to the set from which they received the greatest outflow. In this case, since the flow from the unit at (12, 2) was 8.42 times that from the unit at (9, 2), the unit at (11, 3) was assigned to the set to which the unit at (12, 2) belonged, i.e., that covering the lower-right portion of the grid. This partition produced sets (Fig. 5A) of unequal size: 90, 76, 22 and 34 units in the lower-left, upper-left, upper-right and lower-right sets, respectively, that resembled closely the shapes of the clusters visible in Fig. 3A, and three units in the very small set between the first two of these clusters that could not be ascribed to either with any certainty.

RESULTS

Heterodera avenae

The frequency distribution of counts was very skewed; the value of s^2/m was 15.6, the median count was zero whilst the mean was 4.15. The values of I_a and P_a were respectively 1.46 and 0.007; J_a and Q_a were respectively 1.03 and 0.284. There is considerable overall pattern, indicated by the large value of I_a, mainly caused by more than one cluster, indicated by the small value of J_a. Median values of I_a, J_a, P_a, and Q_a were calculated for various values of f (Table 2). The effects of changes in f were apparent, but were, for these field data, more gradual than those nominal values shown for the idealized mosaic in Fig. 2. The maximum value of J_a occurred between $f = 5$ and $f = 7$, and the value of I_a, except for a large value that exceeded 1.8 for $f = 9$, also seemed to plateau between $f = 5$ and $f = 7$. Because the two major clusters occurred near the top and foot of the full sample area, those values of f (5, 7, 9) for which the majority of subgrids contained part of a single cluster were also those for which the sampled cluster occurred at an edge

TABLE 2. Median values of indices of aggregation and randomization test statistics, for the cereal cyst-nematode *Heterodera avenae* and the potato cyst-nematode *Globodera rostochiensis,* for various sizes of sample area, f^2, using an $f \times f$ contiguous sample grid.

f	Number of subgrids	I_a	J_a	P_a	Q_a
Heterodera avenae					
3	9	1.01	0.99	0.395	0.528
5	9	1.42	1.31	0.015	0.065
7	81	1.66	1.30	<.005	0.005
9	49	1.83	1.21	<.005	0.010
11	25	1.57	1.08	<.005	0.120
13	9	1.42	1.00	0.020	0.485
15	1	1.46	1.03	0.007	0.284
Globodera rostochiensis					
3	169	1.09	0.99	0.270	0.510
5	121	1.25	1.01	0.080	0.445
7	81	1.57	1.02	0.005	0.310
9	49	1.91	1.04	<.005	0.130
11	25	2.26	1.05	<.005	0.010
13	9	2.49	1.05	<.005	0.010
15	1	2.62	1.02	<.0005	0.1035

of the subgrid and for which the value of I_a was therefore slightly inflated (Perry 1996). When $f = 11$ there was a sharp doubling in the proportion of subgrids containing units from more than one cluster, from about one-half to unity, the number of subgrids with larger values of I_a was greatly reduced, and stabilization of I_a for larger values of $f = 11, 13, 15$ occurred around a value of about $I_a = 1.5$. These results, and the IAF plot (Fig. 5A) both indicated that the main contribution to the spatial pattern of the data was made by the two major clusters, and supported an approximately estimated cluster diameter of five to six units (40–45 m) with an intercluster distance of about four units (~30 m). Significant aggregation was detected for all values of $f \geq 5$. For $f = 3$ there are only nine units in the subgrid, so the power to detect nonrandomness was relatively small.

Following the partition of the IAF plot as defined above, the units contributing to the two partitions corresponding to the two major clusters were separately analyzed to find their focus for crowding, as described above. The centers of the lower-left and upper-left clusters were estimated from the reduced data sets to be at (7, 3) and (8, 12), respectively. In neither case were these units those that contained the maximum count in their respective cluster. The decline in nematode density from the center of the former cluster (Fig. 6A) was roughly sigmoidal, with maximum rate at a distance of 1.5 units (~10 m), from the center. Broadly the same result, albeit with a slightly more steady decline, was obtained for the latter cluster, so there was good evidence for nonuniformity of density within each of the two major clusters.

Globodera rostochiensis

The frequency distribution of counts was skewed ($s^2/m = 5.0$, median count = 9, mean = 11.0) but less so

than for the other nematode species; no unit was uninfested. The IAF plot (Fig. 5B) showed clearly one large cluster in the top left of the sample area, extending down ten to twelve rows and covering most of its left-hand side, and about three, or possibly more, further smaller clusters, located mostly towards the other edges of the sample area. The values of I_a and P_a were, respectively, 2.62 and <0.0005; J_a and Q_a were, respectively, 1.02 and 0.1035. These results point towards very strong spatial pattern, indicated by the very large value of I_a, again with more than one cluster, but here even the relatively small value of J_a that only just exceeded unity was associated with a fairly small value of Q_a, indicating that the influence of the subdominant clusters was limited. The presence of a single dominant cluster was confirmed by the steady increase in I_a with f (Table 2). The value of J_a also increased steadily with f, achieving significance at $f = 11$ and $f = 13$ despite the relatively small values of $J_a = 1.05$, until, when $f = 15$, the inclusion of all of the five smaller edge clusters reduced J_a to the slightly smaller value found for the full data set.

The units assigned to the large cluster by the partitioning of the IAF plot (Fig. 5B) were analyzed separately and the focus for crowding was found to be at (6, 9), which was taken as the cluster center. The count in this unit, 36, was the third largest recorded of the 225 in the full data set. Because of this there was an almost inevitable sharp decline in density from this unit to those immediately surrounding it. However, in marked contrast with the other nematode species, for *G. rostochiensis* there was virtually no further decline in density beyond a distance of ~1 unit from the cluster center all the way out to the margins of the cluster (Fig. 6B).

Presence of PVY

On 21 June both treatments gave similar results (Table 3); large and significant aggregation and limited
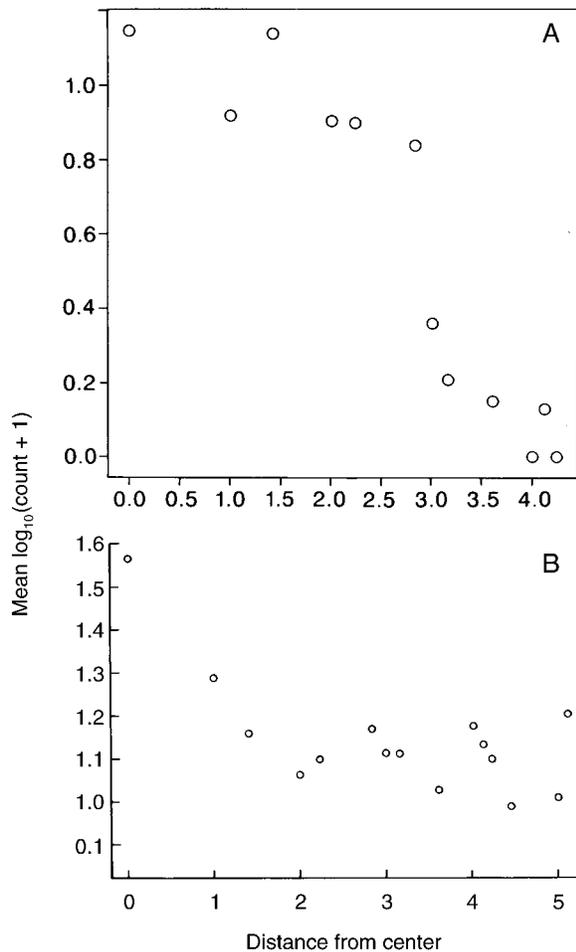
FIG. 6. (A) Decline of geometric mean count (number of nematodes per soil core at that location; note $\log_{10}$ transformation on ordinate scale) with distance from center unit (7, 3) of major cluster, at lower left of sample area, for Boag's cereal cyst-nematode data of Fig. 3A; (B) as for (A) but for Boag's potato cyst-nematode of Fig. 3B and for major cluster at top left of sample area with center unit (6, 9).

dispersal was indicated both by the indices of aggregation and by the regression slope, b, of logit-transformed proportional disease incidence on distance from the grid center. Results for Pirimicarb were slightly more extreme than for Ripcord. Note the greater power to detect a single cluster shown by the index based on crowding than by that based on regularity. However,

on 14 October, the values of I_a, J_a, and b all confirmed the visual impression in Fig. 4 that aggregation was less marked and dispersal greater. For Ripcord, aggregation was demonstrated by J_a and the regression slope, although both were substantially reduced from their values for June. Results for Pirimicarb in October were more extreme. The significant aggregation indicated by I_a was due not to limited dispersal but to the restriction of diseased plants to the lower-left half of the cage, and this was confirmed by the nonsignificant regression; for this treatment the disease had spread right to the cage boundaries. The possible presence of more than one disease cluster (Fig. 4) was suggested by the relatively small value of J_a.

DISCUSSION

Forms of spatial pattern are richer than implied by a one-dimensional continuum from uniformity through randomness to aggregation. The patterns of counts investigated in this paper, not surprisingly, cannot be adequately described by the single index I_a, introduced previously (Perry 1995a). The use of J_a as a supplementary index has been justified by the extra insights afforded in the examples given here. Usually J_a detects aggregation for a single cluster more powerfully than can the indices based on regularity (e.g., dispersal of virus with Ripcord insecticide on 14 October, Table 3). In other cases it was crucial to the quantification of spatial pattern at several scales (Fig. 2, Table 2), and, in conjunction with the partitioned IAF plot, facilitated a coherent measurement of the cluster centers for Boag's nematode data. The use of supplementary visual diagnostics, such as IAF plots (Fig 5) and density-distance plots (Fig. 6), and plots not illustrated here of the frequency distributions from randomizations, contribute important aids to interpretation; these are necessary to help identify which of the diverse list of possible patterns underlies a particular set of data. Further indices may need to be developed to aid the interpretation of different aspects of spatial pattern, for example to study distributions close to an edge, or the presence of spatial association between two populations (Perry et al. 1996, Perry 1997).

There is a current fashion to apply geostatistical techniques such as kriging or variograms (Matheron 1976, Webster and Oliver 1989) to the analysis of spatial pattern in ecology. This is fostered by the frequent need

TABLE 3. Indices of aggregation, randomization test statistics, and regression slopes for PVY virus spread; $\hat{b}$ is the regression slope, $F_{1,13}$ is the F statistic for the regression.

Date	Insecticide treatment	I_a	P_a	J_a	Q_a	$\hat{b}$	$F_{1,13}$
21 June	Ripcord	1.17	0.138	2.15	<.0005	−0.526	16.7
21 June	Pirimicarb	1.41	0.0195	2.21	<.0005	−0.667	34.4
14 October	Ripcord	1.13	0.1765	1.36	0.005	−0.345	4.69
14 October	Pirimicarb	1.34	0.0275	1.08	0.178	−0.035	0.05

to consider many variables, some of which vary continuously, and the ready availability of geographic information system software (e.g., Fry 1995) to process these large volumes of data. Such approaches were developed originally for physical variables studied commonly in soil science, such as fertility and chemical content, that are measured on continuous scales and display a stationary, stable covariance structure over a wide area (but see Kleingeld and Lantuejoul's [1993] application to diamond numbers). However, counts of individuals of a particular animal or plant species are not continuous but discrete, are often, as here, distributed exceedingly patchily, and frequently comprise a majority of zero values. By contrast with physical variables, such population counts are highly dynamic, and have usually evolved to shift ceaselessly in space and time for ecological reasons (Taylor 1986). Such variables might not possess the stable spatial covariance structure assumed by geostatistical methods. They are often characterized instead by isolated clusters, which may be acting as metapopulations with varying degrees of intercluster dispersal (Perry and Gonzalez-Andujar 1993). Webster and Boag's (1992) application was based on the fact that cyst-nematodes move at very slow rates and may be so ubiquitous in the soil that they fulfil the criteria required. Certainly it is true that the main aim of geostatistical analysis, that of local estimation, has some overlap with that of the analysis of spatial pattern. However, I believe that in general the contribution of geostatistical methods to pattern analysis for population count data may prove limited.

Literature Cited

Alston, R. 1996. Statistical analysis of animal populations. Dissertation. University of Kent, Kent, UK.

Bliss, C. I. 1941. Statistical problems in estimating populations of Japanese beetle larvae. Journal of Economic Entomology **34**:221–232.

Czárán, T., and S. Bartha. 1992. Spatiotemporal dynamic models of plant populations and communities. Trends in Ecology and Evolution **7**:38–42.

Fry, G. 1995. Landscape ecology of insect movement in arable ecosystems. Pages 177–202 *in* D. M. Glen, M. P. Greaves, and H. M. Anderson, editors. Ecology and integrated farming systems. Wiley, Chichester, UK.

Greig-Smith, P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. Annals of Botany **16**:293–316.

Hanski, I., and M. Gilpin. 1991. Metapopulation dynamics: brief history and conceptual domain. Biological Journal of the Linnean Society **42**:3–16.

Hassell, M. P., H. Comins, and R. M. May. 1991. Spatial structure and chaos in insect population dynamics. Nature **353**:255–258.

Hassell, M. P., and R. M. May. 1973. Stability in insect host-parasite models. Journal of Animal Ecology **42**:693–726.

Heads, P. A., and J. H. Lawton. 1983. Studies on the natural enemy complex of the holly leaf-miner: the effects of scale on the detection of aggregative responses and the implications for biological control. Oikos **40**:267–276.

Kennedy, J. S. 1972. The emergence of behaviour. Journal of the Australian Entomological Society **11**:168–176.

Kennington, J. L., and R. V. Helgason. 1980. Algorithms for network programming. Wiley, New York, New York, USA.

Kleingeld, W. J., and C. Lantuejoul. 1993. Sampling of orebodies with a highly dispersed mineralization. Pages 953–964 *in* Geostatistics Tróija. 1992. Volume two. Kluwer Academic, Dordrecht, Germany.

Lloyd, M. 1967. 'Mean crowding.' Journal of Animal Ecology **36**:1–30.

Matheron, G. 1976. A simple substitute for the conditional expectation: the disjunctive kriging. Pages 221–236 *in* M. Guarascio, M. David, and C. Huijbregts, editors. Advanced geostatistics for the mining industry. Reidel, Dordrecht, Germany.

Mead, R. 1974. A test for spatial pattern at several scales using data from a grid of contiguous quadrats. Biometrics **30**:2095–307.

Perry, J. N. 1986. Multiple-comparison procedures: a dissenting view. Journal of Economic Entomology **79**:1149–1155.

———. 1994. Chaotic dynamics can generate Taylor's power law. Proceedings of the Royal Society of London series **B** **257**:221–226.

———. 1995*a*. Spatial aspects of animal and plant distribution in patchy farmland habitats. Pages 221–242 *in* D. M. Glen, M. P. Greaves, and H. M. Anderson, editors. Ecology and integrated farming systems. Wiley, Chichester, UK.

———. 1995*b*. Spatial analysis by distance indices. Journal of Animal Ecology **64**:303–314.

———. 1996. Simulating spatial patterns of counts in agriculture and ecology. Computers and Electronics in Agriculture **15**:93–109.

———. 1997. Spatial association for counts of two species. Acta Jutlandica, *in press.*

Perry, J. N., E. D. Bell, R. H. Smith, and I. P. Woiwod. 1996. SADIE: software to analyse and model spatial pattern. Aspects of Applied Biology **46**:95–102.

Perry, J. N., and J. L. González-Andujar. 1993. A metapopulation neighbourhood model of an annual plant with a seedbank. Journal of Ecology **81**:453–463.

Perry, J. N., and M. Hewitt. 1991. A new index of aggregation for animal counts. Biometrics **47**:1505–1518.

Pielou, E. C. 1964. The spatial pattern of two-phase patchworks of vegetation. Biometrics **20**:156–167.

Taylor, L. R. 1986. Synoptic dynamics, migration and the Rothamsted Insect Survey. Journal of Animal Ecology **55**:1–38.

Webster, R., and B. Boag. 1992. Geostatistical analysis of cyst nematodes in the soil. Journal of Soil Science **43**:583–595.

Webster, R., and M. A. Oliver. 1989. Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. Journal of Soil Science **40**:497–512.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology **3**:385–397.