

Spatio-Temporal Associations in Beetle and Virus Count Data

S. KORIE, J. N. PERRY, M. A. MUGGLESTONE, S. J. CLARK, C. F. G. THOMAS,
and M. N. MOHAMAD ROFF

This paper analyses two insect-related sets of agricultural field data. Both comprise spatially referenced count data sampled on a series of occasions. One concerns carabids (ground beetles) in cereals, the other the incidence and spread of an aphid-vectored virus disease of lupins. For both sets, the major objective was to describe and quantify the stability through time of the spatial patterns found for each occasion; this was measured by the spatial association between successive samples. Traditional methods for analyzing count data focus on properties of the frequency distribution of the counts and use little or none of the spatial information in the sample. We used methods that utilized all the spatial information, and which, by conditioning on the observed data, provided complementary inferences to the other methods. Our analyses are based on a class of methods termed Spatial Analysis by Distance Indices (SADIE). These methods provide indices and formal randomization tests, both for the spatial pattern in a single population and for the spatial association when the patterns of two populations are compared. Our analyses showed considerable aggregation for both the beetles and the infected lupin plants. Furthermore, both populations displayed positive association between successive samples, that declined as the temporal lag increased. The beetles were affected greatly by the harvest of the cereal crop. The lupin infections showed maximal association for a one-week lag, despite the

fact that the latent period of the virus was a fortnight; it was inferred that the observed pattern of new infections was probably tracking the pattern of the aphid vectors two weeks previously.

Key Words: Aggregation; Bean yellow mosaic potyvirus; *Lupinus albus*; Principal coordinate analysis; *Pterostichus melanarius*; SADIE; Spatial pattern; Spatially referenced count data; Stability.

S. Korie is a statistician at the International Institute of Tropical Agriculture, Ibadan, Nigeria. Professor J. N. Perry is a biometrician in the Entomology and Nematology Department of IACR-Rothamsted, Harpenden, Hertfordshire AL5 2JQ, UK. Dr M. A. Mugglestone is a statistician/epidemiologist at the MRC Institute for Environment & Health, University of Leicester, 94 Regent Road, Leicester LE1 7DD, UK. Ms S. J. Clark (to whom all correspondence should be sent) is in the Statistics Department, of IACR-Rothamsted, Harpenden, Hertfordshire AL5 2JQ, UK. Dr C. F. G. Thomas is an entomologist in the Crop and Environmental Sciences Department of IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS18 9AF, UK. Dr. M.N. Mohamad Roff is a plant pathologist at the Vegetable Research Centre, MARDI Jalan Kebun, P.O. Box 186, 41720 Klang, Selangor, Malaysia. All work described was done while authors held permanent posts at IACR (JNP, MAM, SJC, CFGT), or were visiting workers (SK, MNMR).

1. INTRODUCTION

To measure spatio-temporal change in a field population of organisms, the population may often be sampled in two-dimensional space on a series of occasions. For efficient integrated pest and disease management it is necessary to understand the reasons for aggregation in the spatial pattern of plant diseases, and of pests or their natural enemy species (Parker, Turner, Perry, Blood-Smyth, Ellis and McKinlay 1997). Sometimes such spatial patterns will be the result of variables directly measurable in space, like prey availability or some crop characteristic. Alternatively, the driving variables (such as the density at different locations of a population of insect vectors responsible for spreading a plant virus disease) may not be easy to measure in the field. However, whether or not the spatial patterns on a certain date can be explained, there is often a need to study the change in such patterns over time. This may be measured by the degree to which the spatial patterns on successive occasions are alike, i.e. their spatial association or dissociation. Populations that are very stable spatially, reflecting perhaps constant spatial patterns in plant nutrients driven by soil variables, may be highly associated. Conversely, populations that move in response to disturbances of habitats, such as crops subject to harvesting, may be much less associated, and successive samples may appear to be distributed at random with respect to one another, whether or not each has a strong spatial pattern individually. Pests or diseases that do excessive and rapid harm to their habitat may need to move continually to optimise their environment, and patterns from successive samples may appear negatively associated, i.e. dissociated.

Sophisticated techniques have existed for twenty years to study the spatial pattern of data in the form of maps where the two-dimensional coordinates of every individual are

recorded accurately (e.g. Diggle 1983); botanical data of this form are common. However, the mobility of animals usually precludes their precise location in two dimensions. Instead, in animal ecology, plant pathology and entomology, spatial data are often recorded as counts of the number of individuals occurring in each of several sample units, where the location of each unit is known. For example, in entomology, the sampling of generalist (polyphagous) predatory beetles is often done using pitfall traps inserted into the soil, and the resulting data are in the form of a count and two-dimensional location for each trap. A review of some of the modern statistical methods to study spatial data was given by Cressie (1991), who gave particular attention to methods that allow for spatial dependence, such as those popularized in geostatistics.

In this paper we study spatio-temporal change in two populations through measurement of the spatial association between successive samples. In both populations the change is insect-mediated; in the first, insects are studied directly, the second measures the dispersal of a non-persistently transmitted plant virus spread by an aphid vector. The outline of this paper is as follows. In Section 2 the data are introduced, and more detail is given considering the important questions to be answered. In Section 3 we review traditional approaches to such problems, explain the need to apply methods that have been developed recently, and give brief details of these new methods. In Section 4 the data are analyzed, and in Section 5 the results are discussed. A more formal description of the methods used is given in an Appendix.

2. DATA

The data come from two field experiments done at the Bristol and Harpenden sites of

the Institute of Arable Crops Research (IACR), UK. Both cover the summer field season from June until September. The first set, collected during 1995 at IACR-Long Ashton, Bristol, concerns distributions of carabids (ground beetles) counted in winter oats; the second set, collected on lupins during 1994 at IACR-Rothamsted, Harpenden, pertains to the incidence and spread of the non-persistently transmitted bean yellow mosaic potyvirus (BYMV), vectored between lupins by the lupin aphid *Macrosiphum albifrons* Essig, and the peach-potato aphid *Myzus persicae* (Sulzer). We present total counts of carabid beetles and infected lupin plants on each sample occasion, and the counts at each location for selected occasions; the full data may be obtained from S.J. Clark (Suzanne.Clark@bbsrc.ac.uk).

2.1 CARABID BEETLE DISTRIBUTIONS

The spatio-temporal distribution of the carabid beetles was surveyed within a 145 m x 85 m field, using a 14 x 14 grid of pitfall traps, spaced 10 m x 6 m apart. At each sample location, a single dry pitfall trap, comprising a removable 6 cm diameter plastic pot set within a permanently sited 6.5 cm diameter plastic drain-pipe, was dug into the ground to a depth of 8 cm and kept flush with the soil surface during the experiment. The pitfall traps were used for one day on 15 irregularly spaced occasions (Table 1). The crop was harvested between occasions nine and ten. Live-trapped carabids were identified to species by eye in the field. Once recorded, each individual was released in the same position, but with the pitfall trap closed to allow redistribution and eliminate bias from immediate recapture in the same trap. About twenty species of carabids were identified in total. *Pterostichus melanarius* (Illig.) (Coleoptera: Carabidae) was the most abundant species, with 1298 individuals caught over the season, accounting for 20.6% of the total captures; in this study we analyze data for this species only. Total counts and sample means and variances are given in Table 1.

For the beetle data the problems were: (i) to assess whether there was appreciable spatial pattern, and, if so, (ii) how stable was this pattern over time, i.e. to measure to what degree the successive patterns were associated; (iii) to characterize large-scale movements of the population; and (iv) to assess the effect of harvest on the spatial pattern of the beetles and on its stability through time.

2.2 VIRUS INCIDENCE DISTRIBUTIONS

The incidence and spread of BYMV on lupins was studied within a field plot measuring 60.2 m x 19.2 m (Mugglestone, Mohamad Roff, Plumb, Jones, Deadman, and Lee in prep.). The plot comprised two parts, an inner rectangle measuring 51.2 m x 12.0 m within which a separate experiment was conducted, and an outer edge of guard rows sown with c. 32,000 white lupins, *Lupinus albus* cv. Lucyane, on 25 March 1994. The inner part of the plot is not considered in this study. The outer part was divided into four regions within which the seeds were sown in different directions (Figure 1). In the west (W) and east (E) regions, seeds were sown in 10 rows, each 60.2 m long and 0.36 m apart. Regions W and E were separated by north (N) and south (S) regions in which seeds were sown in 10 and 15 rows, respectively, each 12.0 m long and 0.36 m apart. From June 1994, just over nine weeks after sowing, symptomatic lupin plants were identified and tagged weekly, for 15 weeks. Leaves from these plants were taken back to the laboratory where presence of the virus was confirmed using enzyme-linked immunosorbent assay (ELISA). At the end of the study period, the positions of infected plants in each row were recorded and maps of new infections each week were compiled. BYMV has a latent period of approximately two weeks; i.e. symptoms of infection are the result of virus transfer from an infected plant two weeks earlier.

For the lupin data the problems were: (i) to assess whether there was appreciable spatial

pattern in the location of new BYMV infections, and, if so, (ii) how stable was this pattern over time, i.e. to measure to what degree the successive new infections were associated; (iii) to relate both spatial pattern and spatial association to the growth-stage of the lupin plants; (iv) to characterize large-scale movements of new infections; (v) to determine whether association between new infections declined with time, and if so, how; (vi) to infer the likely incidence and mobility of the vector population over the susceptible plants; (vii) to assess the relative advantages and disadvantages in aggregating the mapped data to form counts within prescribed areas, and their relationships with the size of area used.

3. METHODS

3.1 SPATIAL PATTERN FOR A SINGLE POPULATION

Recently, increasing use has been made in ecology of geostatistical techniques, such as kriging or variograms, to study spatial problems (Liebhold, Rossi and Kemp 1993). These were developed originally for physical variables studied commonly in petroleum geology and soil science, such as chemical content and fertility, that are measured on continuous scales and usually display a stationary, stable covariance structure over a wide area. Usage has increased particularly for environmental remotely-sensed data, typically collected at larger scales than traditional ecological measurements (Sanderson, Zhang, Ustin and Rejmankova 1998); Liebhold, Rossi and Kemp (1993) warn that at least 30-50 pairs of sample units within half the sample area are required for minimum credibility. When sets of insect data are large, such as those collected for forest pests such as gypsy moth, these methods have proved very successful in spatial studies, particularly for prediction (Gribko, Liebhold and Hohn 1995). However, in this paper we are concerned with different aims, and focus less on prediction than on the

detection and measurement of non-randomness, in the spatial pattern of counts and principally in the degree of spatial association between two such patterns. Insect counts are discrete. They are often distributed exceedingly patchily, and frequently comprise a majority of zero values as is the case in much of the data presented here. They may represent populations that are highly dynamic, that have usually evolved to shift ceaselessly in space and time for ecological reasons, and may be characterised by isolated, ephemeral clusters. Data with such discrete clusters and many zeros may be unsuitable for the application of geostatistical techniques that assume a stationary, stable covariance structure over a wide area. Liebhold, Rossi and Kemp (1993) also caution that additional geostatistical techniques that can measure anisotropy are required for most insect data. Hence, while not denying the power of geostatistical techniques for suitable data, we take a different approach in this paper.

For many years, the techniques applied to measure spatial patterns in counts were limited to descriptions of frequency distributions (e.g. Bliss 1941) and analyses of variance-mean relationships (Greig-Smith 1952; Taylor 1961; Clark, Perry & Marshall 1996). However, these techniques use little or none of the spatial information available from knowledge of the locations of the sample units. Approaches such as that of Mead (1974) overcome this criticism, but usually require the sample units to lie on a regular grid.

New methodology, termed Spatial Analysis by Distance IndicEs (SADIE), now exists to provide indices and tests of non-randomness for spatial patterns of counts using all the available spatial information (Perry 1997a, b, 1998a, b). These techniques may be used for any arrangement of sample units and do not require a grid; they are complementary to those analyses of properties of frequency distributions listed above, because all the inferences are made conditional on the set of counts observed. They provide an improved basis for analysis because

they employ biologically more intuitive measures than the traditional mathematically based ratios involving sample variance and mean, and benefit from increased power due to the greater use of the spatial information in the sample. Consider, as a simple example, the artificial data from a single population on the 5 x 5 square grid shown in Tables 2(a) and (b). In both arrangements only four of the 25 counts are non-zero, the other counts being 1, 2, 3 and 4, occurring once each. Since the list of counts in each arrangement is identical, all statistics pertaining to that list alone, such as mean and variance, are identical for the two arrangements. However, their spatial pattern clearly differs: in (a) the non-zero counts are deliberately crowded together, or aggregated, in a single 'patch' so that none is further than 2.24 units from any other; in (b) although the counts themselves are more heterogeneous than would be expected from a Poisson distribution, they have been arranged randomly.

The basic principle used in SADIE (cf. Perry & Hewitt 1991) is that an observed spatial pattern may be quantified by measuring the minimum total effort which individuals in the observed sample would have to expend were they to move to extreme arrangements, usually in which the individuals in the sample are spaced as uniformly, or regularly, as possible. This effort is quantified by the distance to regularity, D (Perry 1995b), described fully in the Appendix, and calculated by the transportation algorithm (Kennington & Helgason 1980). For example, for the arrangements in Tables 2(a) and (b), $D = 21.6$ and 12.5 , respectively. The degree of non-randomness within a set of count data is quantified by comparing the observed spatial pattern with rearrangements in which the sampled counts are randomly permuted amongst the sampled locations. For the arrangements in Tables 2(a) and (b), the average distance to regularity amongst several hundred randomizations was 14.9. Division of the observed value, D , by this average from the randomization distribution (Perry 1995b, and see

the Appendix) allows the computation of an index, I_a ; for the arrangements in Table 2(a) and (b) the value of I_a was 1.43 and 0.84, respectively. Values of $I_a > 1$ indicate aggregation and $I_a = 1$ indicates random placement of the counts. A test of randomness may be derived from comparison of D with the tails of the randomization distribution (see the Appendix). For the example in Table 2(a) the index value, $I_a=1.43$, indicated significant aggregation ($P_a = 0.02$), whilst the value for Table 2(b), $I_a=0.84$, failed to indicate a non-random arrangement ($P_a = 0.88$).

The coordinates, (x_c, y_c) , of the centroid of the counts provide a simple measure of central location, and are the spatial analogue of an arithmetic mean. Another useful statistic which characterises a particular pattern is δ , the distance between the centroid of the counts and the centroid of the sample units. Whilst necessarily limited in interpretation, this provides a simple measure of overall displacement of the population from the centre of the sample area towards an edge. For example, for the arrangement in Table 2(a), the centroid of the units is (3,3), that of the counts is ($x_c = 4.6, y_c = 4.2$), and $\delta = 2.0$. This form of aggregation due to overall displacement, that occurs at the largest spatial scales that can be resolved within the sample area is indicated by a relatively large value of δ , and for such patterns the value of I_a is correspondingly large; Perry and Klukowski (1997) discuss this in more detail.

3.2 SPATIAL ASSOCIATION FOR TWO POPULATIONS

Techniques for analysing spatial association in mapped populations are available (e.g. Diggle 1983) but usually require a method of simulating independence between component patterns whilst preserving the internal structure of each pattern. The established method of performing the simulations is to hold one pattern fixed and to translate the other pattern in its

entirety by a random horizontal distance and a random vertical distance; the (rectangular) study area is mapped onto a torus by connecting the top and bottom edges and the left and right edges, so that any points moved outside the boundary of the study area reappear at the opposite edge (Lotwick and Silverman 1982). This toroidal shift method was used, for example, by Harkness and Isham (1983) to analyse association between spatial locations of nests of two ant species, *Cataglyphis bicolor* (Fab.) and *Messor wasmanni* (Krausse), in a 0.5 ha field in northern Greece. Such a method depends for practical purposes on having a rectangular sample area, which is not always the case. In this paper, the locations of the virus-infected lupin plants were recorded in mapped form but the toroidal shift method of simulating independence is invalidated by constraints on where plants could occur in the unusually shaped sample area.

For data in the form of counts, there appears to be no traditional methodology for comparing two samples that uses all the spatial information in the samples. The simple correlation coefficient operates on the list of paired counts of two populations, and uses none of the spatial information. For this reason, it may be seriously misleading. For example, consider the two separate arrangements of two populations, shown respectively in bold and italic type, in Tables 2(c) and (d). The arrangements of the italic populations in Tables 2(c) and (d) are identical to each other and to that of Table 2(a), except that, for visual clarity, the zero values are not now shown. The numerical values of the bold population are identical to those of the italic population, in each arrangement; indeed, between Tables 2(c) and (d) it is only the arrangement of the bold population that differs, and that only through a reflection and rotation, so that its spatial pattern remains the same (Perry, Bell, Smith and Woiwod 1996). Visually, the impression in Table 2(c) is one of strong association, and in Table 2(d) of equally strong negative association, or dissociation. However, since the only coincidence of non-zero values

for both arrangements occurs in the central cell, the correlation coefficient for both arrangements is the same small value (-0.083, after transforming the counts to logarithms), that gives no indication of the perceived difference between the two. Clearly, methods are required that can utilise the information from nearby but non-coincident units, for it is just this information that feeds our visual impression.

Two such methods for measuring and testing for spatial association using data in the form of counts are available from within the SADIE class of techniques (Perry 1997b, 1998b), although for brevity, only one will be used here. These utilise all the spatial information in the sample; they require that the location and number of sample units used are the same for both populations. In the approach used here, if, as is usually the case, the total count for each population differs, the counts of each population are multiplied by the total count for the other. These scaled counts are then summed to give a single scaled count for each unit, from which the distance to regularity, denoted T for the two population case, is computed (see Appendix). The idea is that under association, that is, relatively large counts from population 1 being coincident with or close to similarly large values from population 2, aggregations of such counts will be formed that will yield a relatively large value of T ; if there is dissociation, the totals will be more evenly spread, leading to a smaller value. For example, Tables 2(e) and (f) show this process for the arrangements of the bold and italic populations in Tables 2(c) and (d), for which the values of T are 36.6 and 20.8, respectively.

As is the case with mapped populations, a method of simulating independence between component patterns is required to test for association. The method of simulation employed in the SADIE approach is similar in two respects to the toroidal shift method, in that it conditions on the internal structure of the component patterns and it holds one arrangement (say, the first)

fixed whilst randomizing the second (see the Appendix). However, here the randomized version of an arrangement is obtained by applying an algorithm (Perry 1996) that randomly permutes pairs of counts according to certain constraints, until a new pattern is achieved with similar values of I_a and of δ to those observed for population 2 (Perry 1997b, 1998b). This has two advantages over the methods above for mapped data: the sample area need not be rectangular and the sample units may be arranged in any pattern, contiguous or otherwise; also, the mathematically convenient but biologically questionable assumption of periodic boundary conditions is avoided. The aggregation of mapped data to form counts might therefore be worth the inevitable loss of information that results. Manly (199?) discusses alternative forms of randomization to assess association.

Division of the observed value of T by the average from the randomization distributions (combining the values for the randomizations of populations 1 and 2, as long as these are not dissimilar, see the Appendix) allows the computation of an index, I_t ; for the arrangements in Tables 2(e) and (f) the values of I_t were 1.31 and 0.74, respectively. Values of $I_t > 1$ indicate association, $I_t < 1$ indicates dissociation and $I_t = 1$ indicates random placement of one population with regard to the other. Again, a test of randomness is possible, this time in the arrangement of one population with respect to the other, given the spatial pattern of each. Once again this is done by comparing the observed value of T with the tails of the randomization distribution (see the Appendix). For the example in Table 2(d) the index value, $I_t = 1.31$, indicated a tendency towards, but non-significant, association ($P_t = 0.11$), whilst the value for Table 2(b), $I_t = 0.74$, indicated a non-significant tendency towards dissociation ($P_t = 0.87$). The feasibility of this approach was demonstrated by Perry (1998b) using a subset of Harkness and Isham's (1983) ant data and data on bramble, *Rubus fruticosus* L..

To describe the simplest aspect of the difference between two spatial patterns, we define Δ to be the distance between the centroids of their respective counts, by analogy with the use of δ for single populations. A relatively small value of Δ is necessary but not sufficient for association.

4. DATA ANALYSIS

4.1 CARABID BEETLE DISTRIBUTIONS

To study the spatial patterns of beetles on each individual sampling occasion, the SADIE index of aggregation, I_a , was calculated for the counts in each sample (Table 1), based on 200 randomizations, although it is recommended that for general use more be used. All but one of the index values were greater than unity, indicating consistent aggregation, and two-fifths of them were statistically significant. The counts for samples one, four, eleven and 15 are given as examples in Figure 2. Aggregation was very large on three of the first four occasions, declined suddenly on occasion five and remained fairly stable at this moderate level until occasion eleven. After this it appeared to decline further to values that indicated a nearly random distribution of the observed counts. In sample one (Figure 2(a)), there was obvious large-scale clustering ($I_a = 2.50$) demonstrated by the fact that δ was large and 80% of the beetles were in the left-hand side of the plot. Small-scale clustering was also apparent; the large counts occurred close to one another. By sample four (Figure 2(b)), abundance was greater, there were fewer and smaller areas with no beetles, and the centroid had moved slightly to the right (Figure 3). However, although the pattern had evolved, it was very similar to that of sample one, 15 days earlier; aggregation remained substantial ($I_a = 2.14$), the greatest abundance was still in the lower-left quadrant ($20 \leq x \leq 40$, $18 \leq y \leq 36$), and a region of

relatively large density abided near the middle-top of the sample area ($20 \leq x \leq 80$, $66 \leq y \leq 84$). The changes continued consistently; by sample eleven (Figure 2(c)) there were about the same number of zero counts, the centroid had moved still further to the right, and both I_a and δ had decreased, although 60% of the beetles were still caught within the left-hand side of the sample area. By sample 15 (Figure 2(d)), the four quarters of the grid were almost equally represented and the value of δ was very small; aggregation was greater than unity but no longer significant. Also, there were fewer zeros than for the earlier samples shown in Figures 2(a)-(c). Overall, therefore, the spatial pattern was one of increasing colonization by beetles of new regions of the sample area. There appeared to be continual small-scale clustering, overlaid with large-scale clustering that declined over time, reflecting a gradual movement of the centroid of the counts from the left-hand side of the sample area towards the middle; this movement was steady in the long term but with frequent reversals (Figure 3). The informal impression was one of considerable association between successive samples.

To measure the stability of the spatial pattern over time, the SADIE association index, I_t , was computed for each pair of consecutive samples. The index I_t (Table 3) suggested consistent association between the patterns in successive samples, almost half of the values being significant. Association declined steadily from a maximum at samples three and four, increased briefly in samples nine to eleven and subsequently declined to relatively small values, although the association between samples 14 and 15 was statistically significant. Generally the relationship between I_t and Δ was weak. However, the smallest association, occurring between samples ten and eleven, coincided with the largest movement (19.39 m) of the centroid of the distribution of counts (Figure 3, Table 3). The time series of I_t values appeared to vary smoothly, and had a significantly large lag one autocorrelation of 0.677. Overall, the results

confirmed the earlier impression that aggregation was quite stable from sample to sample. The fact that association, as measured by I_t , and aggregation, as measured by I_a , both declined with time, appeared to be a coincidence, particularly since the measurement of association is computed conditionally upon observed aggregation. The association at lag two between a single pair of samples, weeks eleven and 13, when the spatial pattern appeared to change the most and when the length of the intervening period, 29 days, was maximal, was also computed. This yielded a value of $I_t = 0.84$, indicating mild dissociation. Results for another SADIE index of association, I_m (Perry 1997b, 1998b), were also calculated but are not reported here as they were very similar to those of I_t .

4.2 VIRUS INCIDENCE DISTRIBUTIONS

The weekly spatial point patterns of new infections were converted to counts of infected plants in 50 rectangles superimposed on the guard rows. Each rectangle contained five rows of plants approximately 6 m long; there were 20 such rectangles in each of regions W and E, and four and six in regions N and S, respectively (Figure 1). Counts per rectangle ranged from zero to ten; weekly totals of newly infected plants are given in Table 4. Counts for weeks one, six, nine and 12 are given as examples in Figure 4. As the latent period for the disease is two weeks, there were no new infections in week two. Numbers of new infections in weeks four, five and 15 were so small that any formal analysis involving these data is of dubious validity. The coordinates of the centre of each rectangle were computed to locate the total count for that rectangle. Indices of aggregation, I_a , for each week separately (Table 4), based on 400 randomizations, showed a reasonably smooth increasing trend, from week one to week three, that continued (ignoring results for weeks four and five) to a maximum in week eight.

Thereafter, aggregation declined sharply but remained large and often significant, averaging about 1.5. Because the shape of the sample area was a narrow hollow ring, the centroid of the counts generally lay within the inner part of the plot, that is, outside the sample area (Figure 5); however, this presents no problems for interpretation. In week one (Figure 4(a)), the centroid of the counts was located fairly close to the centroid of the sample units; the slight bias towards the left of the sample area was caused by notably low incidence towards the lower part of region E ($16 \leq x \leq 19, 9 \leq y \leq 28$) that persisted up to and including week seven. There was little change in week three; numbers were again probably too small to achieve significance of the index I_a . There was a decline in incidence during weeks four and five, but then the centroid of the counts moved up the plot; this movement continued (Figure 5) through the vegetative/early flowering period of the lupins until weeks six (Figure 4(b)) and seven. By week six the value of δ was large, and there was substantial aggregation; over half of the newly infected plants occupied rectangles in or near region N, for which $y \geq 57$. Week seven displayed a similar pattern, but now the infection appeared to spread from region N down to the top of region W. There was a fourfold increase in infection during week eight, by which time the infection appeared to have spread further down region W and also down the top of region E, resulting in a slight reduction in the value of δ . However, small-scale clustering appeared intense; in week eight, 14 of the 17 largest counts occurred no more than 1.8 m from another member of that set. A substantial further apparent spread of the virus towards region S occurred by week nine (Figure 4(b)) and continued right up until the end of sampling at week 15. Indeed, until week eight there had only been 12 infected plants in the ten rectangles in or near region S, for which $y \leq 5$; in week eight alone there were eight further infections, and in week nine another 16. This process was, of course, accompanied by further movement of the centroid of the counts; from a maximum in

week seven, y_c declined almost monotonically to 3.01 in week 15 (Figure 5). For weeks nine, ten and eleven, the late flowering period, the centroid of the counts was located centrally and consequently δ was small. Between weeks 12 (Figure 4(d)) and 15, representing the pod development stage of the lupins, the value of δ increased, and the spatial pattern became dominated once more by larger-scale aggregation. After a slow start in the spread of the virus (Table 4), there was a steady decrease in new infections, from a peak of over 150 in week nine to a single plant in week 15; as was the case for the beetle data, I_a was uncorrelated with the number of infected plants. Combined counts for the accumulated period, weeks one to 15, also showed significant aggregation (Table 4).

Overall, therefore, the dynamics were dominated by a single cycle of new infections that increased steadily after week four to a maximum in week nine and declined thereafter. The spatial pattern was complex. Small-scale clustering seemed moderate until week five, very intense between weeks six and eight, to decline from weeks nine to eleven, and be absent from week 12 onwards. Large-scale clustering was driven by the steady movement of the centroid, first in the direction of y_c increasing until week seven, then in the direction of y_c decreasing thereafter (Figure 5). The scale of this movement, relative to the size of the study area, was larger than for the beetle data. The initial impression is one of considerable association, although the small total counts for weeks one to seven and 12 to 15 may hamper detection. Furthermore, since the spatial pattern appears to group naturally into weeks one to five, weeks six to eight, weeks nine to eleven, and weeks 12 to 15, spatial association might be expected to be larger within these groups than between them.

Mugglestone et al. (in prep.) previously used K -functions (Diggle 1983) to compare the initial spatial point pattern of infected plant positions with the cumulative patterns of new

infection arising during the vegetative growth stage (weeks two to six), the flowering stage (weeks seven to 12) and the pod development stage (weeks 13 to 15). Their analyses showed that the initial pattern was random, the other patterns were all significantly non-random, and that there was little evidence of spatial association between the spatial patterns of new infections in the initial and later growth stages. The grouping of new infections according to the plant growth stages was partly for convenience but also because the only practicable way of measuring and testing for independence in different samples was via comparison of any given pattern with the pattern in week one. The reason for this is that the sowing pattern of the lupin seed invalidates the use of the toroidal shift method for simulating independence. Had the seed rows all been oriented in the same direction, random multiples of the row spacing could have been used to implement toroidal shifts. Comparisons of the pattern in week one with patterns in subsequent weeks were made possible by the fact that the pattern in week one was consistent with complete spatial randomness; simulations of independence therefore proceeded by simulating new, entirely random initial patterns, each containing 24 events to match the observed data, whilst holding all subsequent patterns fixed.

For the present analysis, after converting the mapped data to counts, SADIE techniques were used to measure association from week to week. The association index, I_t , was computed between pairs of new infections recorded two weeks apart, to match the latent period, i.e. comparing week one and week three, weeks three and five, weeks four and six, and so on, up to weeks ten and 12. After this flowering had ended and new infections were rare. All fortnightly comparisons, except those for weeks three and five, and for weeks ten and 12, showed positive association (Table 5), although only one was significant. After week three, there was a steady increase in association to a maximum, followed by a sharp decline for comparison of weeks

nine and eleven, and weeks ten and 12; there was no significant autocorrelation of I_t . As was the case with the beetle data, association was strongly inversely related to Δ , the distance between the centroids of the units; the smallest associations (weeks three and five; ten and 12) corresponded to the relatively large values of $\Delta = 10.82$ and 13.98 , respectively.

Since all previous infections could potentially influence the distribution of infected plants at a later stage, the association index I_t was computed between counts of infected plants for all pairs of weeks between weeks six and 12 (lags one to six) to determine the cumulative pattern of disease. These weeks had reasonably large numbers of newly infected plants (Table 4) and covered the late vegetative and flowering periods. The resulting matrix of indices is given in Table 6. The expected negative relationship between these values of I_t and Δ was now displayed even more strongly (Figure 6). The matrix in Table 6 appears to confirm the groupings of weeks suggested earlier (i.e. weeks six, seven and eight; nine, ten and eleven; and 12). In a more formal examination of how the patterns in different weeks related to one another we first transformed I_t to a similarity measure, $s = 0.5 + \log_5 I_t$, the logarithmic base being chosen to give a wide interval for s , subject to $0 < s < 1$. Principal coordinates analysis (PCO) was used on these s values to locate the weeks in a conceptual multidimensional space (e.g. Clark, Perry, Tatchell, and Woiwod 1992). The first three principal axes accounted for 46%, 18% and 14% of the variation, respectively (78% in total). The first axis appeared to separate week 12 from the others (Figure 7); the second axis separated weeks six, seven and eight from weeks nine, ten and eleven; the third axis separated weeks nine and eleven. The groupings predicted by comparing the spatial pattern of individual weeks and noted from the individual elements of the matrix in Table 6 were thus confirmed by the PCO. This verifies the impression of a large degree of consistency in the overall results. In general, the association between weeks

declined with increasing lag (Table 6); the geometric mean of the index values for lags one to six were, respectively, 1.35, 1.14, 1.02, 0.85, 0.82 and 0.61. These values suggest some dissociation for lags greater than three weeks; however, they may have been overly influenced by the results for week 12. Excluding this week, the geometric means for lags one to five were 1.47, 1.32, 1.20, 1.16 and 1.13, respectively. Whether week 12 was included or not, association decreased monotonically with lag from a large positive value when the lag was a single week; the decline was initially substantial.

We also calculated the index of association I_m but the results are not reported here as they were very similar to those for I_t .

5. DISCUSSION

The major difference between the two datasets analysed here is that the beetle counts probably represent samples from a single beetle population over time, whilst the new BYMV infections in each week represent increases in the population of infected plants that resulted from the spreading of earlier infections. Some of the beetles recorded in successive samples may well have been the same individuals; estimates of the daily survival rate of this species in arable habitats in the south of England during summer months exceed 0.98 (P.J. Kennedy, pers. comm.).

The smallest association between beetle counts coincided with the largest movement of the centroid of the distribution. These samples (weeks ten and eleven) were taken just after the crop was harvested. The reasons for aggregation of the beetles in particular locations is unknown but is likely to be related to either food availability or microclimate. The latter is most

affected by soil conditions, particularly moisture content, and by crop characteristics, particularly crop density. Harvesting the crop destroys the vegetation cover and exposes the soil surface. This is likely to make the microclimate at the soil surface much more uniform and conditions generally less favourable for the beetles. This in turn is likely to lead to a general increase in the movement of the beetles as they search for more favourable sites within the field or leave the cropped area for more suitable habitats at the field edges. Such movement will lead to a disruption of any previously stable pattern and to a more uniform distribution across the entire field.

The detailed analysis for the BYMV infections allows us to study the stability of the spatial patterns over time. At first sight, the fact that association between new infections is greatest for a one-week lag is surprising, given that the latent period is two weeks. Indeed, if the incidence of the viruliferous vector aphids was small and the aphids were randomly distributed over the sample area, this result would be surprising. However, suppose instead that there was a relatively large incidence of virus within the vector population and that the population was distributed patchily, with a spatial pattern that altered systematically through time according to the nutritional quality of the plants. Then, the spatial pattern of new infections might be nothing more than a reflection of the spatial pattern of the aphid population two weeks previously. The association expressed between patterns of virus incidence in successive weeks would then indicate the stability of the vector population. This stability would be expected to be greater between occasions one week apart, rather than two weeks apart. Evidence for relatively large vector mobility at these time scales is supported by the fact that the virus was not introduced deliberately; the initial infections were seed-borne but the virus would also have been brought into the sample area from outside by colonizing aphids. Indeed, the secondary infection process

of transfer of disease from plant to plant during the experimental period may have been supplemented continuously by new primary infection processes caused by newly immigrant viruliferous aphids. If the plants were indeed differentially attractive, and if there was some clustering in this attractiveness, these immigrants would be likely to supplement colonies already in existence in such clusters. A dominant effect of immigrating aphids is more than likely with a virus like BYMV. The explanation of the high degree of association between patterns of virus incidence is but one of many possible explanations; the real situation may be more complex. For example, the large-scale clustering detected might be due to primary infections caused by immigrants, while the small-scale clustering might be due to the secondary infection process driven by transfers of virus between neighbouring plants. Such transfers might be the result of short-range flights by viruliferous aphids that had already fed and were seeking alternative plants.

The use of counts within rectangles rather than the exact locations of virus-infected plants results in a loss of information. The boundaries of such arbitrarily defined areas may well cut through natural clusters of infections and so lead to false inferences. This would certainly have caused problems if the scale at which clustering occurred was larger than the area of the rectangles used, although there was no evidence that this was the case here. One alternative would be to do away with the rectangles and restrict the counts to zero-one data composed entirely of presence-absence of infections. This might have been practicable, although computationally expensive, had the data comprised the locations of all plants, including those that remained uninfected. However, only the locations of plants that became infected during the 15 weeks of the experiment were recorded. Also, the analysis of the zero-one data would have required a further restriction in the randomization procedure to ensure that previously infected

plants would remain infected in simulations of later infections. Failure to do so would induce spurious dissociation. In any case, this approach might not provide sufficient power to detect association if the driving processes occurred at larger scales. For mapped data, some of the simpler aspects of the SADIE techniques have been implemented for data from a single population through an iterative procedure involving Voronoi polygons (Perry 1995a).

In the case studied here the SADIE analysis of counts gave more detailed information than did the K-function analyses of Mugglestone et al. (in prep.); the ability to analyze week to week changes revealed significant association that their analyses based on growth stages did not disclose.

The consistency of the results over time in both examples confirmed the ability of the SADIE index to provide a sensible measure of spatial association, as suggested by Perry (1997a,b, 1998b), although the degree to which this technique is computer intensive precludes the use of the number of randomizations that would be desirable. The putative relationship between large-scale clustering and δ proposed in previous papers (Perry and Klukowski 1997) was extended in this paper to a putative relationship between association and Δ , that was confirmed by example. The analyses presented here have given a broad overview of the analysis of spatial association and its variation through time. More detailed study of particular pairs of occasions could have been pursued using visual diagnostic tools as described in Perry (1998b), but this was not thought necessary here.

Improved SADIE methods that provide a local measure of the degree of small-scale clustering for each individual sample unit have been developed recently (Perry, Winder, Holland and Alston 1999). These allow both patch and gap clusters to be identified and measured; they provide a better environment for detailed interpretation of aggregation patterns

than could the single index I_a . Work in progress has extended this concept to define measures of local association.

ACKNOWLEDGEMENTS

S. Korie was funded by a Rothamsted International Staff Fellowship. M. Bresan conducted the pitfall trapping experiment under an AIR Mobility Grant with the University of Bristol. M.N. Mohamad Roff was supported by the Government of Malaysia. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. The authors thank D.E. Johnson and B.F.J. Manly, for suggestions regarding presentation, and three referees for very useful comments. Free Fortran software to perform SADIE analyses under Windows95 and WindowsNT is available from J.N. Perry (joe.perry@bbsrc.ac.uk) on request and a downloadable tutorial is available from <http://www.res.bbsrc.ac.uk/entnem/about/projects/joeperry/joeperry.htm>

REFERENCES

- Bliss, C.I. (1941), "Statistical Problems in Estimating Populations of Japanese Beetle Larvae," *Journal of Economic Entomology*, 34, 221-232.
- Clark, S.J., Perry, J.N., and Marshall, E.J.P. (1996), "Estimating Taylor's Power Law for Weed Species and the Effect of Spatial Scale," *Weed Research*, 36, 405-417.
- Clark, S.J., Perry, J.N., Tatchell, G.M., and Woiwod, I.P. (1992), "Comparative Phenologies of Two Migrant Cereal Aphid Species," *Journal of Applied Ecology*, 29, 571-580.
- Cressie, N.A.C. (1991), *Statistics for Spatial Data*, Wiley, New York.
- Diggle, P.J. (1983), *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.
- Greig-Smith, P. (1952), "The Use of Random and Contiguous Quadrats in the Study of Plant Communities," *Annals of Botany*, 16, 293-316.
- Gribko, L.S., Liebhold, A.M., and Hohn, M.E. (1995), "A model to predict gypsy moth (Lepidoptera: Lymantriidae) defoliation with kriging and logistic regression," *Environmental Entomology*, 24, 529-537.
- Harkness, R.D., and Isham, V. (1983), "A Bivariate Spatial Point Pattern of Ants' Nests," *Applied Statistics*, 32, 293-303.
- Kennington, J.L., and Helgason R.V. (1980), *Algorithms for Network Programming*, Wiley, New York.
- Liebhold, A.M., Rossi, R.E., and Kemp, W.P. (1993), "Geostatistics and Geographic Information Systems in Applied Insect Ecology," *Annual Review of Entomology*, 38, 303-327.
- Lotwick, H.W., and Silverman, B.W. (1982), "Methods for Analysing Spatial Processes of

Several Types of Points,” *Journal of the Royal Statistical Society, Series B*, 44, 406-413.

Manly, B.F.J. (1998), “Sampling and Modelling of Insect Populations,” In: *Population and Community Ecology for Insect Management and Conservation*, pp. 3-19 (eds, J. Baumgartner, P. Brandmayr & B.F.J. Manly). Balkema, Rotterdam. *Proceedings of the 20th International Congress of Entomology, Florence, 25-31 August 1996*.

Mead, R. (1974), “A Test for Spatial Pattern at Several Scales using Data from a Grid of Contiguous Quadrats,” *Biometrics*, 30, 295-307.

Mugglestone, M.A., Mohamad Roff, M.N., Plumb, R.T., Jones, P., Deadman, M.L., and Lee, B.Y.Y. (1999), “Spatio-Temporal Distribution of Bean Yellow Mosaic Potyvirus on White Lupin cv. Lucyane,” (in prep.).

Parker, W.E., Turner, S.T.D., Perry, J.N., Blood-Smyth, J.A., Ellis, S.A., and McKinlay, R.G. (1997), “Development of a GIS-Based Tool for Testing Field Sampling Plans by Modelling the Within-Field Distribution of Mealy Cabbage Aphid in Brussels Sprouts,” *Precision Agriculture 1997 (Proceedings of the 1st European Conference)* pp. 811-819.

Perry, J.N. (1995a), “Spatial Analysis by Distance Indices,” *Journal of Animal Ecology*, 64, 303-314.

----- (1995b), “Spatial aspects of animal and plant distribution in patchy farmland habitats,” In: *Ecology and Integrated Arable Farming Systems*, pp. 221-242 (eds, D. Glen, M. Greaves & H.M. Anderson). Wiley, Chichester

----- (1996), “Simulating Spatial Patterns of Counts in Agriculture and Ecology,” *Computers & Electronics in Agriculture*, 15, 93-109.

- (1997a), ““Inchworm, Inchworm, Measuring the Marigolds”: Biometry in Action,”
Inaugural Lecture Series of the University of Greenwich. 30pp. ISBN: 1 86166 0499.
- (1997b), “Spatial Association for Counts of Two Species,” *Acta Jutlandica*, 72, 149-
169.
- (1998a), “Measures of Spatial Pattern for Counts,” *Ecology*, 79, 1008-1017.
- (1998b), “Measures of Spatial Pattern and Spatial Association for Counts of Insects,”
In: *Population and Community Ecology for Insect Management and Conservation*, pp.
21-33 (eds, J. Baumgartner, P. Brandmayr & B.F.J. Manly). Balkema, Rotterdam.
*Proceedings of the 20th International Congress of Entomology, Florence, 25-31 August
1996.*
- Perry, J.N., Bell, E.D., Smith, R.H., and Woiwod, I.P. (1996), “SADIE: Software to Measure
and Model Spatial Pattern,” *Aspects of Applied Biology*, 46, 95-102.
- Perry, J.N., and Hewitt, M. (1991), “A New Index of Aggregation for Animal
Counts,” *Biometrics*, 47, 1505-1518.
- Perry, J.N., and Klukowski, Z. (1997), “Spatial Distributions of Counts at the Edges of Sample
Areas,” *Proceedings of the VIth Conference of the Biometric Society (Spanish Region),
Cordoba, 21-24 September 1997*, pp. 103-108.
- Perry, J.N., Winder, L., Holland, J.M., and Alston, R.D. (1999), “Red-Blue Plots for Detecting
Clusters in Count Data,” *Ecology Letters*, 2, 106-113.
- Sanderson, E.W., Zhang, M., Ustin, S.L., and Rejmankova, E. (1998), “Geostatistical Scaling of
Canopy Water Content in a California Salt Marsh,” *Landscape Ecology*, 13, 79-92.
- Taylor, L.R. (1961), “Aggregation, Variance and the Mean,” *Nature*, 189, 732-735.

Table 1. Sample Number (S), Sample Occasion (L , Day Number measured from 9 June 1995), Total Count of *Pterostichus melanarius* (N), Sample Mean (m), Sample Variance (s^2), and SADIE Index of Aggregation (I_a ; Significant Values in Bold, $P < 0.05$), for Each of the 15 Sample Occasions

S	L	N	m	s^2	I_a
1	1	66	0.34	0.430	2.50
2	6	26	0.13	0.136	1.21
3	12	185	0.94	2.187	2.72
4	15	147	0.75	1.081	2.14
5	20	73	0.37	0.450	1.34
6	26	29	0.15	0.137	1.60
7	29	35	0.18	0.188	1.41
8	35	36	0.18	0.171	0.89
9	42	79	0.40	0.888	1.56
10	50	49	0.25	0.455	1.28
11	54	122	0.62	1.057	1.39
12	75	35	0.18	0.301	1.27
13	83	123	0.63	1.199	1.40
14	89	91	0.46	0.599	1.09
15	96	202	1.03	2.348	1.08

Table 2. Example Sets of Counts on a 5 x 5 Square Grid to Illustrate: (a) a Spatial Pattern for which the Counts are Aggregated; (b) a Spatial Pattern for which the Counts are Arranged Randomly; (c) Two Sets of Counts showing Strong Association; (d) Two Sets of Counts showing Strong Dissociation; (e) Counts for Sets 1 and 2 in (c) Summed within Each Unit; (f) Counts for Sets 1 and 2 in (d) Summed within Each Unit. Zero Counts are Not Shown in (c), (d), (e) or (f).

	(a)					(b)				
Row 1	0	0	0	2	4	0	2	0	0	0
Row 2	0	0	0	0	0	0	0	0	4	0
Row 3	0	0	1	0	3	0	0	0	0	0
Row 4	0	0	0	0	0	1	0	0	0	0
Row 5	0	0	0	0	0	0	0	0	0	3
	Col. 1	Col.2	Col.3	Col. 4	Col.5	Col. 1	Col.2	Col.3	Col. 4	Col.5
	(c)					(d)				
Row 1			3	2	4				2	4
Row 2				4	2					
Row 3			11		3	3		11		3
Row 4							4			
Row 5							2			
	Col. 1	Col.2	Col.3	Col. 4	Col.5	Col. 1	Col.2	Col.3	Col. 4	Col.5
	(e)					(f)				
Row 1			3	2	4				2	4
Row 2				4	2					
Row 3			2		3	3		2		3
Row 4							4			
Row 5							2			
	Col. 1	Col.2	Col.3	Col. 4	Col.5	Col. 1	Col.2	Col.3	Col. 4	Col.5

Table 3. Association Index I_t (Significant Values in Bold, $P < 0.05$), Distance, Δ m, between the Centroids of the Counts, and Separation (s) in Days for Successive Weekly Comparisons of Counts of *Pterostichus melanarius*

Comparison of Samples	s	I_t	Δ
1 v 2	5	1.40	11.65
2 v 3	6	1.50	7.06
3 v 4	3	1.62	10.60
4 v 5	5	1.50	5.58
5 v 6	6	1.34	11.00
6 v 7	3	1.32	6.57
7 v 8	6	1.14	12.71
8 v 9	7	1.16	15.07
9 v 10	8	1.39	7.55
10 v 11	4	1.37	4.89
11 v 12	21	1.25	8.81
12 v 13	8	1.07	19.39
13 v 14	6	1.10	8.49
14 v 15	7	1.21	1.12

Table 4. Week Number, Total Count of Newly Infected Lupin Plants (N), SADIE Index of Aggregation (I_a ; Significant Values in Bold, $P < 0.05$), Sample Mean (m) and Sample Variance (s^2) for Each of 15 Weeks, and Results for the Accumulated Total Infections

Week	N	m	s^2	I_a
1	24	0.48	0.540	1.12
2	0	-	-	-
3	20	0.40	0.490	1.28
4	11	0.22	0.298	0.72
5	6	0.12	0.108	0.76
6	21	0.42	0.861	1.82
7	33	0.66	1.617	2.32
8	138	2.76	7.207	3.45
9	162	3.24	6.798	1.64
10	123	2.46	4.621	1.38
11	79	1.58	4.698	0.92
12	34	0.68	1.283	1.67
13	10	0.20	0.245	1.84
14	7	0.14	0.164	1.45
15	1	0.02	0.020	-
Total	669	-	-	2.55

Table 5. Association Index I_t (Significant Value in Bold, $P < 0.05$) and Distance, Δ m, between the Centroids of the Counts, for Successive Fortnightly Comparisons of Newly Infected Plants.

There were no New Infections in Week 2

Comparison	I_t	Δ
wk1 v wk3	1.14	8.75
wk3 v wk5	0.90	10.82
wk4 v wk6	1.21	9.43
wk5 v wk7	1.25	11.09
wk6 v wk8	1.64	1.32
wk7 v wk9	1.27	12.25
wk8 v wk10	1.33	9.54
wk9 v wk11	1.11	4.13
wk10 v wk12	0.62	13.98

Table 6. Association Indices, I_i , for All Pairwise Comparisons of Counts of Newly Infected Plants Between Weeks 6 and 12 (Significant Values in Bold, $P < 0.05$)

Sample							
6	-						
7	1.51	-					
8	1.64	1.75	-				
9	1.27	1.27	1.38	-			
10	1.25	1.24	1.33	1.45	-		
11	1.13	1.07	1.11	1.11	1.29	-	
12	0.61	0.60	0.46	0.63	0.62	0.90	-
Sample	6	7	8	9	10	11	12

FIGURE LEGENDS

Figure 1. Plan Showing Location and Orientation of Rows of Lupin Plants in the Guard Rows of a 60.2 x 19.2 m Experimental Plot. Dashed and Solid Lines Represent the Edge of the Experimental Area and Rows of Lupin Plants, Respectively. The Central Area is Excluded from the Analyses. Counts of Infected Plants in Fifty 6 m x 1.8 m Rectangles were Obtained: Two Such Rectangles are Shown in Regions E and S.

Figure 2. Counts of Pterostichus melanarius in 196 Pitfall Traps in (a) Sample 1, $\delta = 24.3$; (b) Sample 4, $\delta = 14.2$; (c) Sample 11, $\delta = 10.7$; and (d) Sample 15, $\delta = 2.4$. ● Locates the Centroid of the Counts.

Figure 3. ● Centroids of Counts of Pterostichus melanarius on Each of the 15 Sample Occasions.

Figure 4. Counts of Newly Infected Lupins in Fifty 6 m x 1.8 m Rectangles in (a) Week 1, $\delta = 4.6$; (b) Week 6, $\delta = 14.5$; (c) Week 9, $\delta = 4.7$; and (d) Week 12, $\delta = 10.1$. ● Locates the Centroid of the Counts. The Shaded Area Contains a Separate Experiment that is Excluded from the Analyses.

Figure 5. ● Centroids of Counts of Newly Infected Lupin Plants in Weeks 1,3...15. No New Infections Occurred in Week Two.

Figure 6. Plot of SADIE Index, I_t , Against Distance Moved by Centroid, δ , Between Pairs of Samples of Newly Infected Lupin Plants. Each Point is Labelled by the Lag in Weeks Between the Pair of Samples.

Figure 7. First Two Principal Coordinate Axes for $s = 0.5 + \log_5 I_t$. The Axes Account for 46% and 18% of the Variation, Respectively. Points are Labelled by Week Number.

APPENDIX

1. SPATIAL PATTERN FOR A SINGLE POPULATION

Suppose there are $k = 1, \dots, n$ sample units located on a two-dimensional plane with coordinates (x_k, y_k) , and a single integer count, c_k , recorded for each unit. The sample mean of the counts, $\sum_k c_k / n$, is denoted by m . Let the set of units with $c_k > m$ be denoted as $\{\mathfrak{R}_i, i = 1, \dots, N_i\}$, the set of units with $c_k < m$ be denoted as $\{\mathfrak{S}_j, j = 1, \dots, N_j\}$, and the set of units with $c_k = m$ be denoted as $\{\mathfrak{N}_d, d = 1, \dots, N_0\}$. Note that $N_i + N_j + N_0 = n$, that $\{\mathfrak{R}\}$, $\{\mathfrak{S}\}$ and $\{\mathfrak{N}\}$ are mutually exclusive sets, and that $\{\mathfrak{N}\}$ is always empty if m is non-integer, and will often be empty even if m is integer. There are $N_i N_j$ possible pairings of sample unit i in $\{\mathfrak{R}\}$ with unit j in $\{\mathfrak{S}\}$. For these pairs of units, consider the set of $N_i N_j$ quantities v_{ij} , where $v_{ij} \geq 0$ and v_{ij} may be non-integer, and are subject to the constraints that, for each unit in $\{\mathfrak{R}\}$, $\sum_j v_{ij} = c_i - m$, and that, for each unit in $\{\mathfrak{S}\}$, $\sum_i v_{ij} = m - c_j$. Associated with each such pair is the Cartesian distance, d_{ij} say, between the two units, where $d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}$. Now denote the function, $\sum_i \sum_j v_{ij} d_{ij}$, by D_{ij} . The value of D_{ij} represents the total distance required to move individuals and fractions of individuals from their observed locations, to achieve a new arrangement over the given sample units in which an identical (usually non-integer) quantity, m , is located in each unit, i.e. to achieve a completely regular spatial distribution. The minimum value, D , of the function D_{ij} is termed the distance to regularity. D may be found as the unique solution of the transportation algorithm from the operational research literature. Kennington and Helgason (1980) described the solution and illustrated how their Fortran subroutine NETFLO could be used to find it. Their method identifies which values of i and j are associated with non-zero values of v_{ij} and calculates these values of v_{ij} . Usually there are $N_i + N_j - 1$ such values, but there may be fewer.

Consider a random permutation of the observed counts, c_k , amongst the given sample units, and let the value of distance to regularity found for the r th such randomization be denoted as D_r . If the sample mean of the values of D_r over R such permutations, $r = 1, \dots, R$, is denoted as E_a , then an index of aggregation, I_a , may be defined as $I_a = D / E_a$. Furthermore, suppose P_a represents the proportion of the permutations with $D_r \geq D$. Then a test of the null hypothesis of a random arrangement of the observed counts amongst the given sample units is made by rejecting if $P_a < 0.05$, against the alternative that the counts are aggregated, i.e. that relatively large counts tend to occur closer together than would be expected by chance. Aggregated arrangements are indicated by $I_a > 1$; random arrangements by $I_a = 1$. The alternative hypothesis of a more regular spatial distribution of counts than would be expected by chance, i.e. arrangements in which relatively large counts are separated by larger distances than would be expected by chance alone, is not usually of interest, since very few such arrangements occur in field data.

The centroid of the sample units is denoted as location O , with coordinates (x_o, y_o) , where $x_o = \sum_k x_k / n$, and $y_o = \sum_k y_k / n$. The centroid of the observed counts is denoted as location C , with coordinates (x_c, y_c) , where $x_c = \sum_k c_k x_k / \sum_k c_k$, and $y_c = \sum_k c_k y_k / \sum_k c_k$. The Cartesian distance, OC , between O and C , is denoted as δ .

2. SPATIAL ASSOCIATION FOR TWO POPULATIONS

The above notation is generalized as follows. Suppose there are $k = 1, \dots, n$ sample units located on a two-dimensional plane with coordinates (x_k, y_k) , with two integer counts recorded in each unit: c_{k1} , for population 1, and c_{k2} , for population 2. Let the total count over all units for population 1 be denoted $M_1 = \sum_k c_{k1}$, and for population 2 be $M_2 = \sum_k c_{k2}$. Form new scaled

integer counts C_{k1} , for population 1, from $C_{k1} = M_2 c_{k1}$, and C_{k2} , for population 2, from $C_{k2} = M_1 c_{k2}$, $k = 1, \dots, n$. The total count in both scaled populations is now $M_1 M_2$. Denote the distance to regularity for the scaled counts of population 1 as D_1 , and for population 2 as D_2 . Now form the total scaled count, $C_{k1} + C_{k2}$, for each unit, $k = 1, \dots, n$. Denote the distance to regularity, for this set of observed total scaled counts, as T .

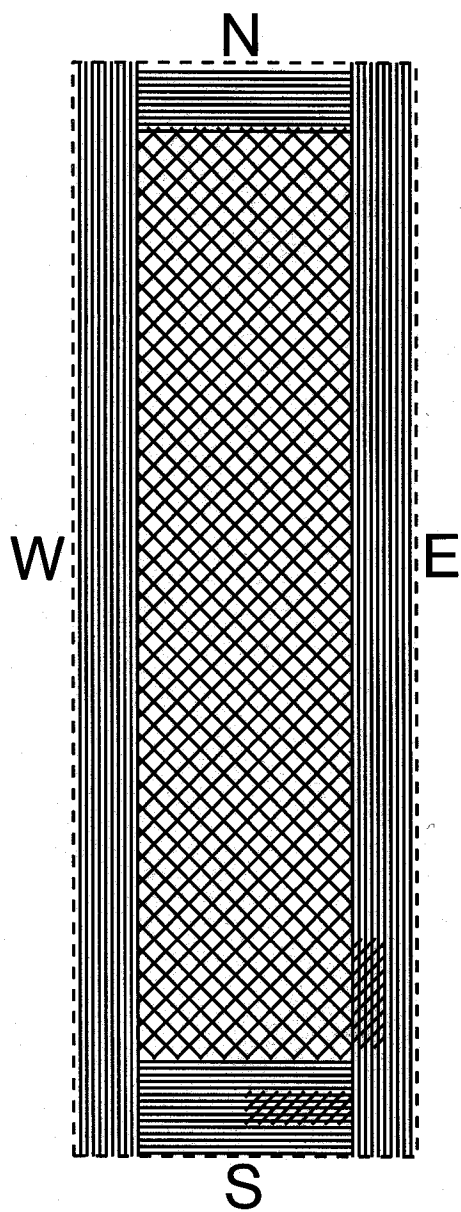
Now consider two separate sets of constrained random permutations, to test the null hypothesis that there is no spatial association of one population with respect to the other, given the spatial pattern of each population individually. To explain the generation of these, the following notation is defined. Denote the distance between the centroid of the sample units, location O , at $(x_o = \sum_k x_k / n, y_o = \sum_k y_k / n)$, and the centroid, C_1 , of the scaled counts for population 1, at $(x_{c1} = \sum_k C_{k1} x_k / \sum_k C_{k1}, y_{c1} = \sum_k C_{k1} y_k / \sum_k C_{k1})$, as δ_1 . Similarly, denote the distance between location O and the centroid, C_2 , of the scaled counts for population two, at $(x_{c2} = \sum_k C_{k2} x_k / \sum_k C_{k2}, y_{c2} = \sum_k C_{k2} y_k / \sum_k C_{k2})$, as δ_2 . The distance between C_1 and C_2 , the counts centroids of the two populations, is denoted as Δ .

The first set of constrained permutations randomly rearranges the scaled counts, C_{k1} , for population 1 amongst the given sample units, whilst holding the scaled counts, C_{k2} , of population 2 in their original locations. The two constraints on the permuted values of C_{k1} , for each randomization r , are: (i) the distance to regularity of the permutation, D_{1r} , satisfies $D_1(1-\epsilon_1) < D_{1r} < D_1(1+\epsilon_1)$ where $0 < \epsilon_1 < 1$ is a small value, typically 0.05; and (ii) the distance $C_{1r}O$ for the permutation satisfies $\delta_1/(1 + \epsilon_2) < C_{1r}O < \delta_1(1 + \epsilon_2)$, where $0 < \epsilon_2$ is a small value, again typically 0.05. This procedure yields a set of $r = 1, \dots, R$ constrained permutations for each of which there is a total scaled count, $C_{k1r} + C_{k2}$, for each unit, $k = 1, \dots, n$; for each such permutation there is a distance to regularity denoted as $T_{r(2)1}$. By analogy with the single

population case above, if the sample mean of the values of $T_{r(2)1}$ over the R constrained permutations is denoted as $E_{r(2)1}$, then an index of association, $I_{t(2)1}$, is formed from $T/E_{r(2)1}$. The proportion of constrained permutations that have values of $T_{r(2)1}$ as large as, or larger than T is denoted as $P_{t(2)1}$.

The second of the constrained set of permutations is obtained by analogy with the above, but now the scaled counts from population 2 are permuted, with scaled counts of population 1 held at their original values. This yields the set of values $T_{r(1)2}$, the proportion $P_{t(1)2}$ and the index $I_{t(1)2}$. In principle, it is possible for distinct inferences to pertain to each set of permutations. However, in practice, the values of $I_{t(2)1}$ and $I_{t(1)2}$, and of $P_{t(2)1}$ and $P_{t(1)2}$ are usually similar, and their arithmetic average values are then calculated, to form a combined index $I_t = (I_{t(2)1} + I_{t(1)2})/2$, and proportion $P_t = (P_{t(2)1} + P_{t(1)2})/2$. Given the spatial pattern of each population, spatial association between them is indicated by $I_t > 1$; dissociation by $I_t < 1$; random placement of one population with regard to the other by $I_t = 1$. A test of the null hypothesis is made by rejecting if $P_t < 0.025$, in favour of the alternative that the counts are associated, or if $0.975 < P_t$, in favour of the alternative that the counts are dissociated.

There follows, in this order:
Figures 1, 2a-d, 3, 4a-d, 5, 6 & 7

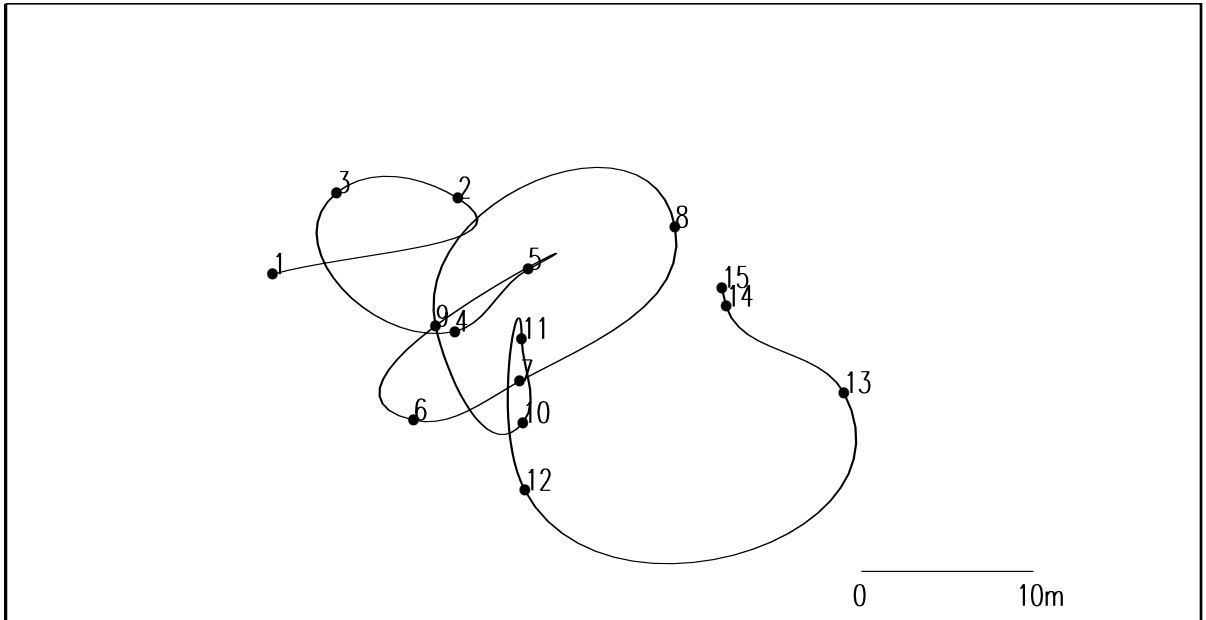


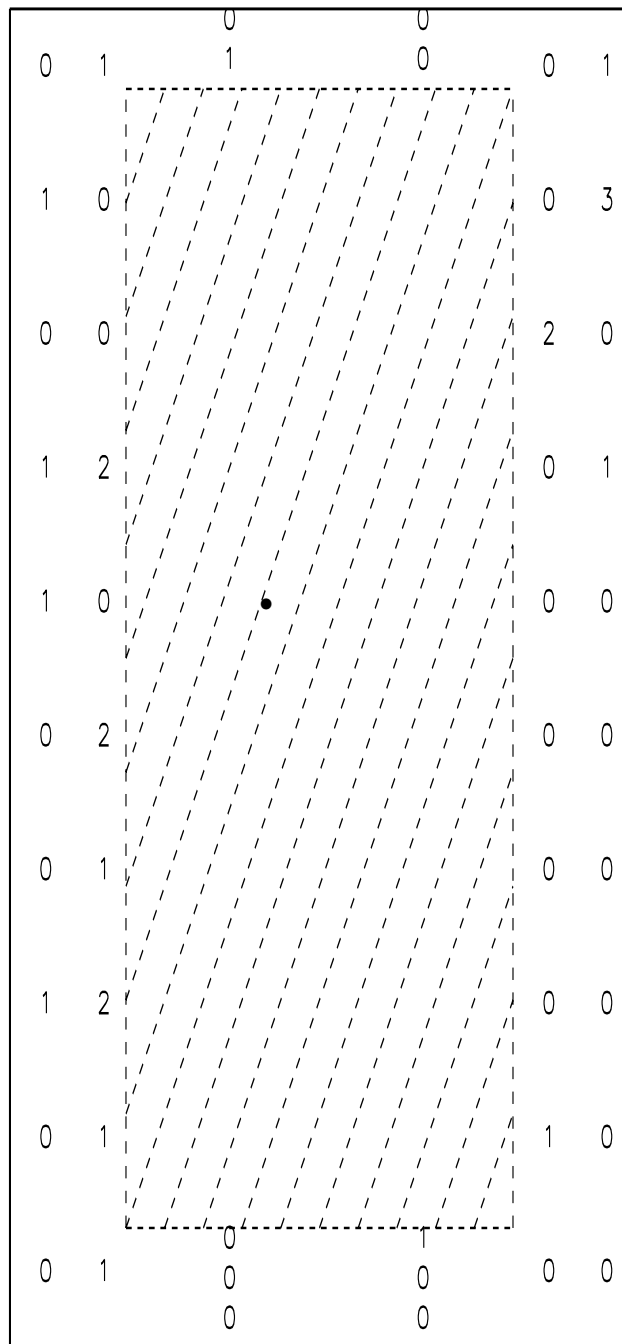
0	0	0	1	1	1	0	0	0	0	0	0	1	0
0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	1	2	1	2	1	0	1	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0
0	1	2	1	0	0	0	0	0	1	0	0	0	0
0	3	1	1	0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	1	0	1	1	1	0	0	0	0
0	0	0	3	0	0	0	1	0	0	0	0	0	0
0	3	1	3	1	0	0	1	0	1	0	0	1	0
0	1	2	0	1	1	2	0	0	0	0	0	0	0
0	1	0	3	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	1	0	0	0	0	0	0

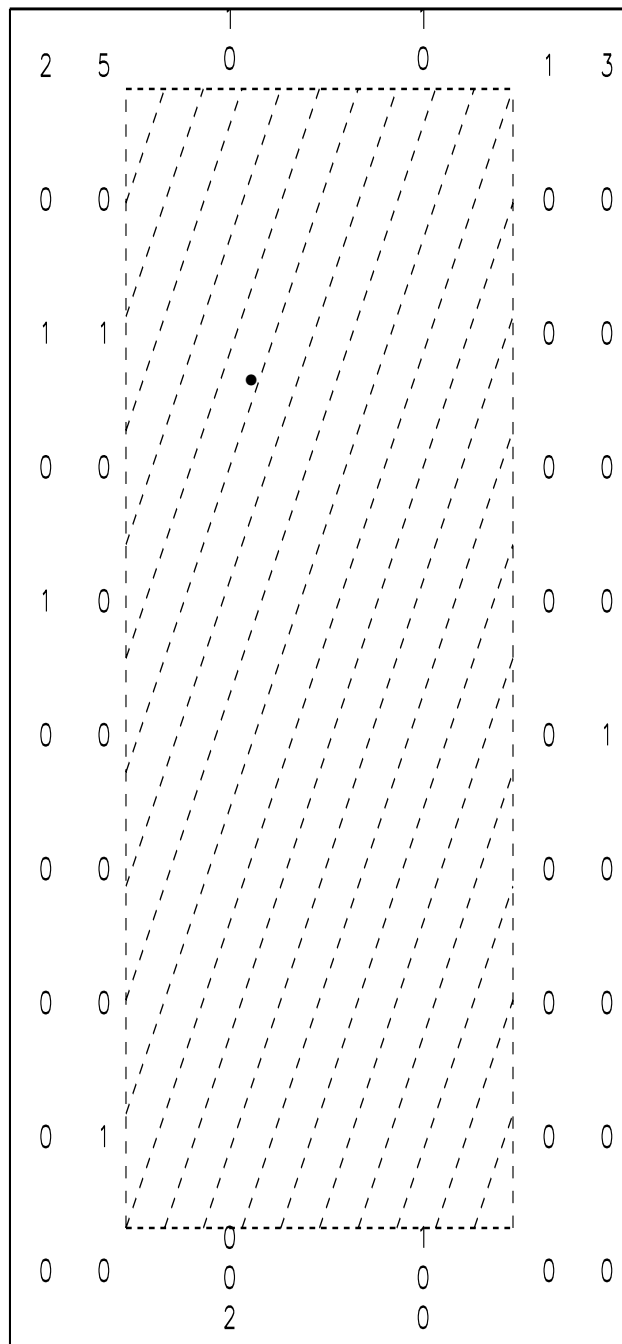
0	3	2	2	0	1	1	0	1	0	0	0	0	0
0	2	1	0	1	1	1	0	1	1	0	0	1	0
0	3	0	0	0	2	3	0	0	0	0	0	3	0
0	0	0	0	3	0	1	2	2	0	0	3	0	0
0	0	1	1	0	1	1	0	0	1	1	1	2	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	2	1	0	1	0	1
0	0	1	0	0	1	1	1	0	1	1	0	0	0
2	2	1	0	2	1	3	2	1	0	0	0	0	0
0	5	0	3	0	2	2	2	2	0	0	0	1	0
0	0	3	0	0	0	1	0	0	0	0	0	1	0
0	6	1	1	2	3	1	2	0	3	1	0	0	0
1	2	1	0	2	3	0	1	1	1	0	0	0	0
2	2	1	2	3	1	1	0	0	0	1	0	1	0

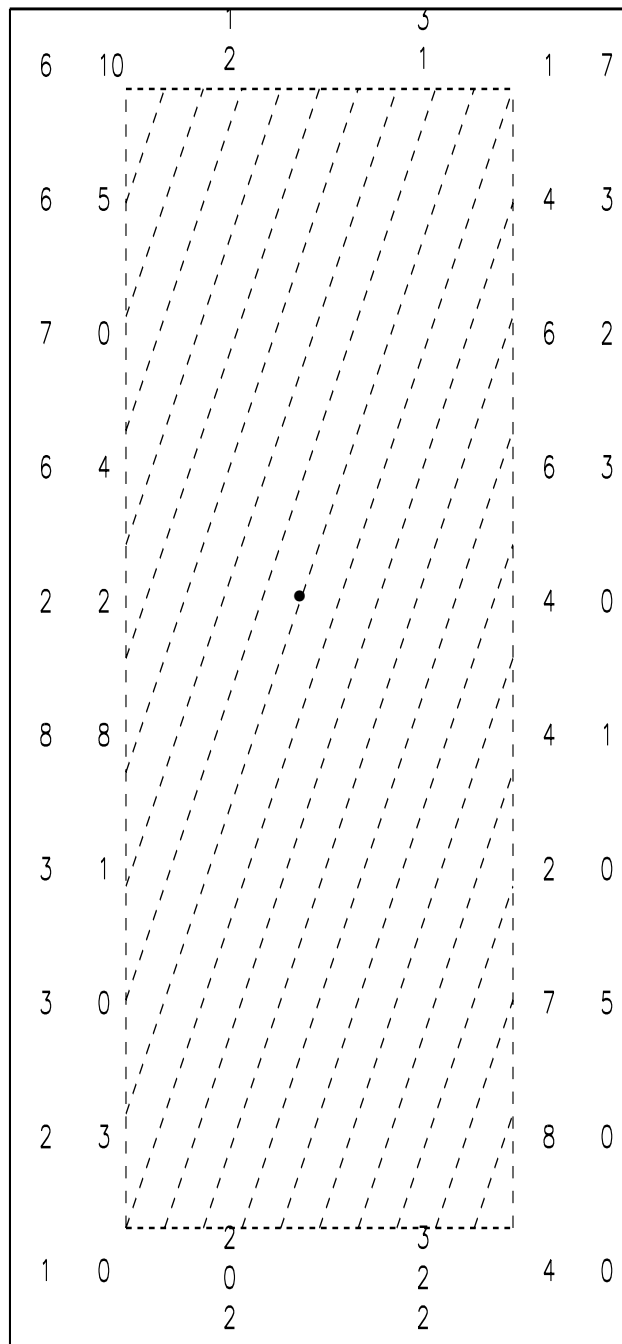
0	1	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	1	0	3	0	0	0	0	1	0	0	0	0	0
0	0	0	3	1	4	1	0	1	0	1	0	0	1
0	2	1	0	0	2	0	0	1	1	0	0	0	0
0	2	1	0	4	0	0	0	1	3	2	1	1	0
1	0	1	0	0	2	3	4	0	0	0	2	0	1
0	1	0	0	0	2	0	0	1	1	1	0	0	0
2	1	0	3	1	0	0	1	3	0	0	0	0	0
0	1	2	1	0	1	1	0	0	0	0	0	0	0
0	2	0	0	1	1	0	0	1	0	4	0	1	2
2	4	0	1	0	0	0	0	2	3	1	0	0	0
1	0	0	0	0	0	1	0	0	0	0	3	0	2
5	0	0	1	0	1	2	0	0	0	0	0	0	0

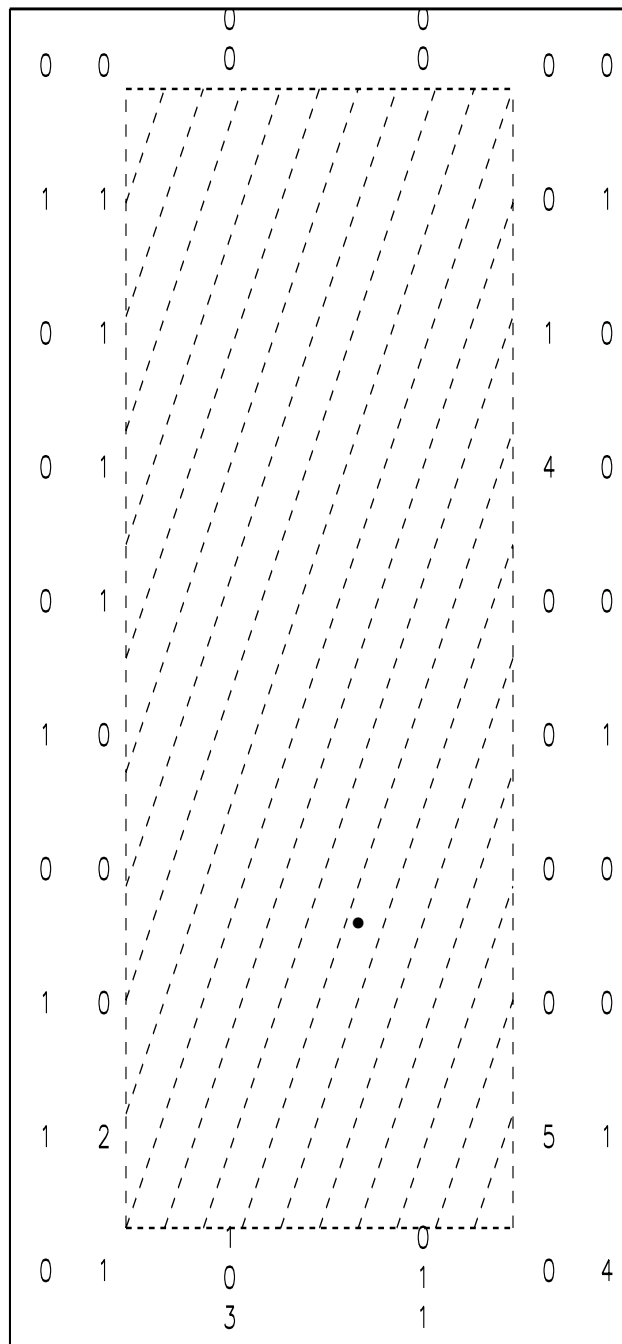
2	0	2	0	0	2	1	0	0	0	0	3	1	0
2	1	2	0	0	0	1	0	3	0	4	0	0	0
2	2	0	2	2	0	0	0	0	2	0	2	0	1
0	0	0	0	0	4	1	0	1	2	6	1	1	2
3	0	1	0	0	0	1	0	0	0	1	4	0	0
2	0	0	0	0	0	2	0	1	1	1	1	1	1
9	0	0	0	1	1	0	0	2	0	3	0	0	0
0	0	1	1	0	0	0	• 0	3	2	3	5	5	0
1	1	5	1	1	0	0	2	0	0	5	0	0	0
1	1	2	0	0	2	2	3	0	1	1	1	0	0
1	1	4	1	1	0	1	1	0	3	1	0	2	1
1	0	0	0	2	1	0	0	8	4	2	0	0	0
0	0	0	0	0	0	0	2	0	0	5	6	1	1
3	3	1	0	0	0	0	0	0	0	3	1	0	0

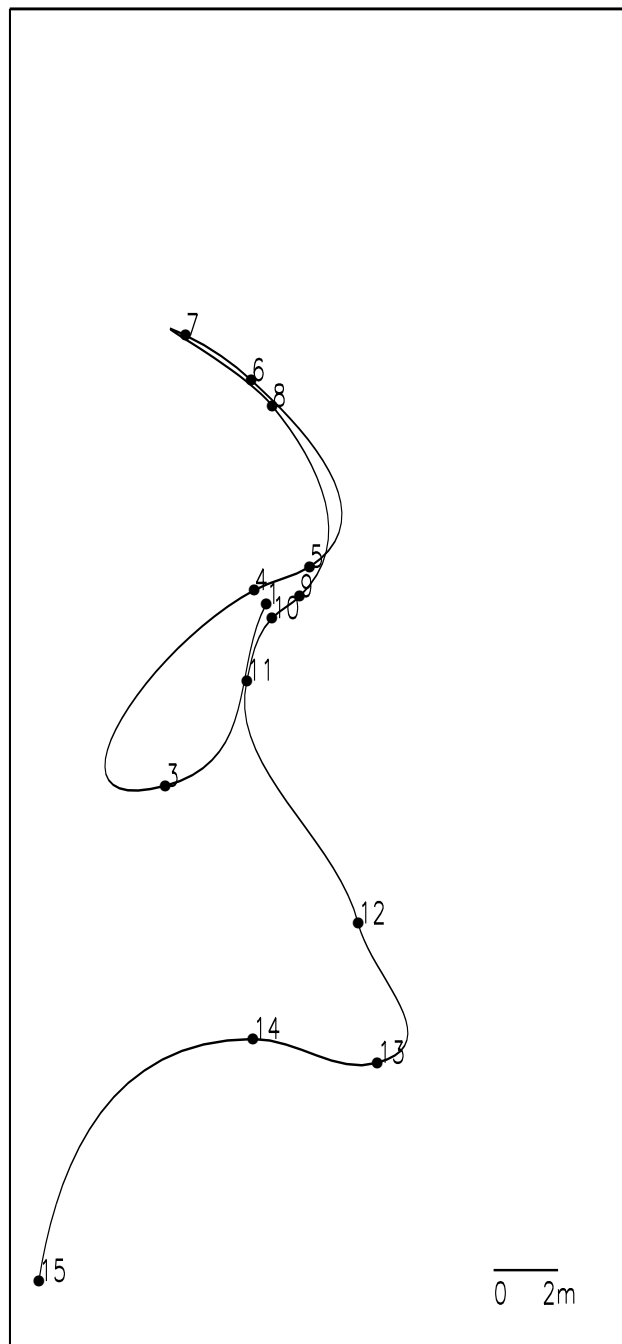


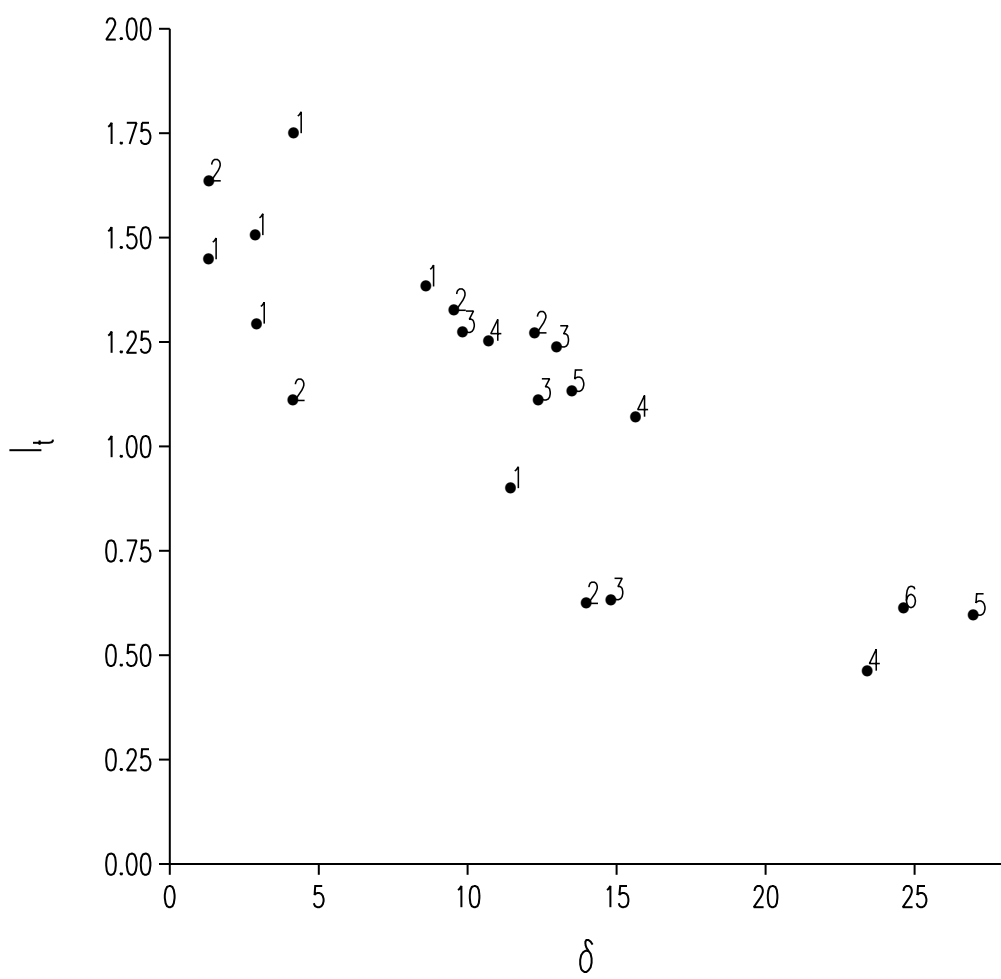




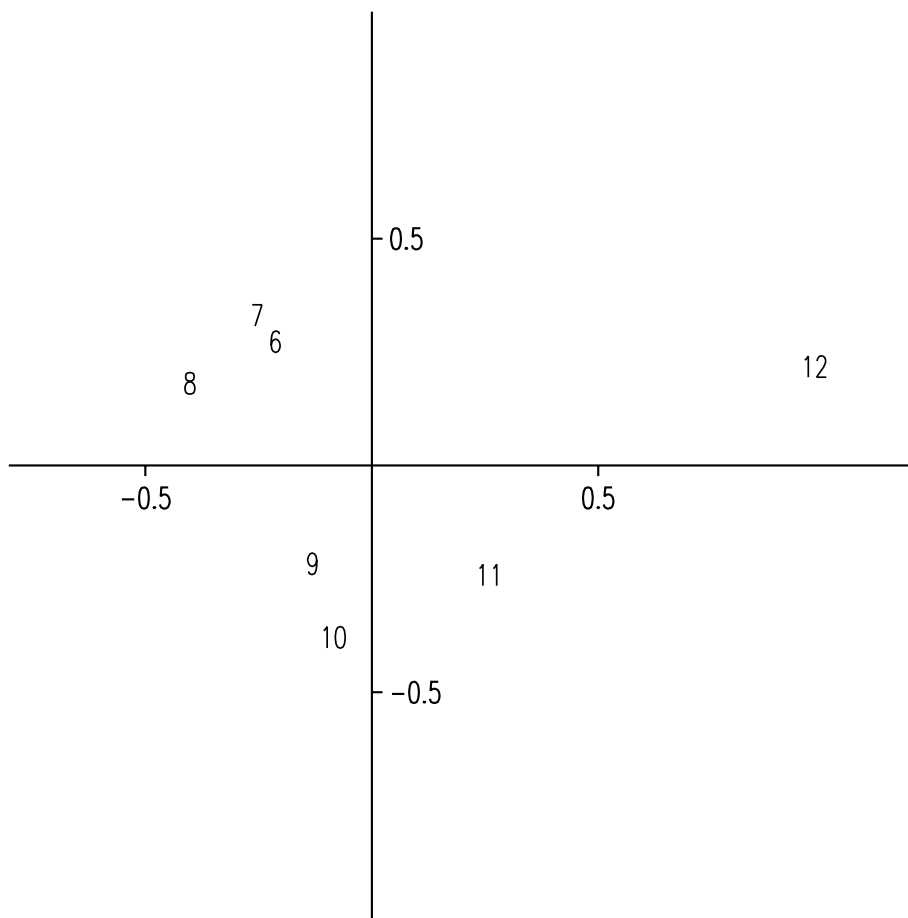








Axis 2



Axis 1