Analysing maps of dispersal around a single focus

S. KORIE[1,6], S.J. CLARK[1], J.N. PERRY[2], M.A. MUGGLESTONE[1], P.W. BARTLETT[4], E.J.P. MARSHALL[5] and J.A. MANN[3]

[1] Statistics Department, [2] Entomology & Nematology Department & [3] ICER, IACR-Rothamsted, Harpenden, Herts, AL5 2JQ, UK, [4] Plant Health Group, MAFF Central Science Laboratory, Sand Hutton, York YO4 1LZ, UK, [5] Crop & Environmental Sciences Department, IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS18 9AF, UK, [6] Present address: International Institute of Tropical Agriculture, Ibadan, Nigeria.

The spatial pattern of organisms may be used to characterize their dispersal, quantify spread or estimate the point of introduction of an alien species. Their distribution may be represented by maps of individuals, or by counts, or presence/absence at known positions within a sampled area. The problems and relative merits of these different forms of data for spatial inference are discussed.

Three datasets concerning dispersal from a single focus are analysed: counts of aphids, *Rhopalosiphum padi* and *Sitobion avenae*, on barley plants, *Hordeum vulgare*, grown in experimental trays; mapped locations of couch grass, *Elymus repens*, tillers within plots of a field experiment; locations of sightings of the lupin aphid, *Macrosiphum albifrons*, as it invaded Great Britain between 1981 and 1984.

A method for generating maps from counts is proposed to overcome problems caused by recording imprecision.

Several statistics are used to quantify dispersal and spatial pattern in the experimental data and together provide a clear picture of the spatial pattern observed; they enabled several effects of the experimental treatments to be identified. The value of the statistics are compared .

Estimates of the source of the lupin aphid invasion are obtained using the backtracking methods of Perry (1995b) and do not contradict previous suggestions.

---

1. Introduction

For many years ecologists have been concerned with analysis of spatial patterns of animals (Bliss, 1941) and plants (Greig-Smith, 1952), especially those arising from their dispersal (Dobzhansky and Wright, 1943). Two-dimensional maps of the locations of individuals can often be compiled, especially for relatively immobile animals (Fleming and Baker, 1936) or plants (Diggle, 1983). These individuals may be distributed in random, regular or aggregated patterns, and a description and quantification of their pattern can contribute to an understanding of the ecology of a species. The map obtained for a single species might represent the early stages of invasion into the sample area (Skellam, 1952; Boag *et al*., 1994); the number of foci that comprise the source may affect invasion rate significantly (Mack, 1985). Alternatively, it may represent a well-established population spread (Taylor, 1986) or continuing to spread (Sharov *et al.*, 1997) throughout the sample area. Previous work has quantified the spread, and estimated the point of introduction, of invading alien species (Tower, 1906; Trewhella and Harris, 1988; Soderstrom and Jonsson, 1989; Bartlett, 1993; Stewart and Blackshaw, 1993; Agassiz *et al.*, 1994; Boag *et al*., 1994; Killion and Grant, 1995; Nash *et al*., 1995; Sharov *et al.*, 1995, 1996).

Maps of dispersal can be generated in several ways. Controlled experiments, usually at the field or glasshouse scale, can be used to study the effects of different experimental factors, or treatments, on dispersal while other sources of variation are held constant. Although it is an advantage to control variation by allowing for known trends within the environment, such experiments may restrict dispersal to a relatively small experimental arena, thereby affecting the natural behaviour of the individuals involved. By contrast, surveys may often be more appropriate for following large-scale dispersal, such as that of an invading

species (Trewhella and Harris, 1988; Bartlett, 1993; Lonsdale, 1993; Boag *et al*., 1994; Killion and Grant, 1995). Such data may be collected according to structured protocols by experienced ecologists, but may also be supplemented by *ad hoc* sightings made by the general public, often in response to media reports. Problems associated with these kinds of data include multiple reporting, inaccuracy of coordinates and variations in sampling intensity over the area sampled. The coverage within a country of persons with accurate knowledge of a particular species may itself be highly aggregated, leading to further potential bias. However, Agassiz *et al*. (1994) showed that this method of data collection, when compared with a systematic survey, led to underestimates of the total range but more accurate estimates of the area in which the invading moth, *Phyllonorycter leucographella*, existed at high densities. For a general introduction to the mathematical modelling of invading species see Shigesada and Kawasaki (1997) and Williamson (1996).

Maps are more difficult to obtain for organisms which are highly mobile and disperse widely, such as aphids, and for sample areas which are relatively large. In these cases, data often take the form of counts, at specified locations (Taylor, 1986) or in grid squares (Augustin, Mugglestone and Buckland, 1996), or of presence/absence in grid squares (Nash *et al*., 1995). Even when the construction of a map is practicable, it is common for individuals to be counted in quadrats arranged in a regular grid or at specific positions within the sample area (Marshall, 1988; Wilson and Brain, 1991). The loss of spatial information in count data generally precludes the use of methods for maps (Diggle, 1983; Ripley, 1988) which might otherwise have provided more accurate or detailed quantifications of spatial pattern. Methods based on relationships between sample variance and mean, such as Taylor's power law (Taylor, 1961), are commonly used for analysing animal (Taylor, Woiwod and Perry, 1978) and plant (Clark, Perry and Marshall, 1996) counts, although they do not utilize even the

limited spatial information in such data (Perry, 1995a).

Dispersal from a single, central point source, if unrestricted and isotropic (the same in all directions), will result in a circular distribution of dispersed individuals. This suggests a natural null hypothesis that may be tested by a Rayleigh statistic (Fisher, Lewis and Embleton, 1987) for uniformity in the circular spread of individuals. Non-uniformity in dispersal may be caused by environmental factors, including host availability, experimental treatments and human activities, especially agricultural and horticultural practices. The spread of plant pests, for example the lupin aphid that is studied here, may be hastened by the transfer of plants from nurseries to gardens centres and on to domestic gardens, creating new foci of dispersal at each relocation.

The analysis of dispersal data takes many different forms, depending on the issues of interest. In an individual-based approach, Silander and Pacala (1985) used maps of seedlings of *Arabidopsis thaliana* to develop an index of neighbourhood interference based on the number, distance and angular dispersal of neighbouring plants. Nelson (1996) derived a comparable proximity index from the density of infected plants in, and the area of, a focus in order to measure the compactness of disease clusters on *Cucurbita pepo* planted on a regular grid. By contrast, considering the population as a whole, Banks, Kareiva and Zia (1988) developed diffusion models of population movement based on two-dimensional transport equations which incorporated density, time, distance and direction; these were used to estimate random and directed components of movement. Dispersal from a focus along a radial axis may be measured at one point in time (Taylor, 1978); changes in distribution over time might be related to distance and more specifically to time after release (Freeman, 1977).

Perry (1995b) extended the SADIE (Spatial Analysis by Distance IndicEs) class of measures (Perry and Hewitt, 1991; Alston, 1994; Perry, 1995a; Perry *et al.*, 1996) for

spatial pattern in animal counts to two-dimensional maps, where the coordinates of each individual and the rectangular area in which they were sampled are known. Briefly, an algorithm is used to move individuals from their initial positions to positions in which they fill the defined rectangular sample area in a regular manner. The distance to regularity, *D*, is the sum of the distances that individuals are displaced from their initial to these final positions; it is computed and compared with a distribution of values obtained from completely random initial arrangements of the points in the sample area. The probability of observing a value as extreme as *D* is computed, along with an index of aggregation, I_p, defined as the ratio of *D* to the mean of the simulated distribution. Values of I_p greater than unity indicate aggregation, values close to unity indicate randomness, and those less than unity, regularity. Once dispersal from a point source has begun, as time progresses the degree of aggregation in the resulting spatial pattern decreases.

Also, it may be of interest to estimate the position of that source, or focus, of the spread. Perry (1995b) proposed a method, termed backtracking, which estimates the focus of a given cluster of individuals. The method assumes that the population moves outwards to colonize available space and would eventually maximize habitat availability by spacing itself in a regular manner, as implemented in the SADIE algorithm described above. The observed arrangement of individuals is assumed to occupy a spatio-temporal position somewhere between the original, unknown and highly aggregated focus of the spread, and the final, eventually regular positions; further details are given below. Perry (1995b) did not realise that the backtracking method must, by definition, estimate the focal point of the invasion to be further from the centre of the sample area than the centroid of the observed data, which will be positioned between the two. The backtracking method is, therefore, only useful for analysing invasions from the edges of the sample area rather than those from a known central

source, for which the estimate is bound to be biased. How seriously this constraint prevents the valid use of this proposed method is examined below.

The purpose of the present paper is to investigate problems related to the analysis of dispersal from a single focus; these we attempt to exemplify with three diverse datasets. Two of these arose from controlled experiments with a known focus in the centre of the sampled area. In the first, dispersing *Rhopalosiphum padi* L. (bird cherry aphid) and *Sitobion avenae* (Fabricius) (grain aphid) were counted on individual *Hordeum vulgare* cv. Puffin (barley) plants in a glasshouse. We propose a method for generating a map of possible aphid positions on these plants, and compare results for the observed counts and generated positions. Here, attention focuses on whether the potential improvement in precision from counts to a map is outweighed by inaccuracies induced by the stochastic map generation process. In the second, tiller positions of the grass, *Elymus repens* (L.) Gould (*Agropyron repens* (L.) Beauv.), were mapped in the field. We attempt to isolate spatial effects of the treatments imposed, aware of possible problems caused by limitations on the size of the sample area. In addition to existing methods, such as the SADIE indices and Rayleigh statistic mentioned above, we suggest some other very simple statistics to aid the description and comparison of the spatial patterns in these two datasets; these might prove useful in the analysis of other similar datasets. Where possible, we compare and contrast the value of all the statistics used to describe the patterns. The third dataset consists of locations of sightings of the lupin aphid, *Macrosiphum albifrons* Essig, as it invaded Great Britain from an unknown focus in the early 1980s. We use backtracking to estimate the position of this focus.

2. Materials and methods

2.1 *Statistical methods*

For two datasets, described in detail below, for which spread was known to have originated from a central source, we computed an appropriate selection of the following statistics to describe movement and spatial pattern. First, I_p, the SADIE index of aggregation (Perry, 1995b). Second, to quantify the displacement of the entire population, we defined δ as the distance between the centroid, *C*, of the observed individuals' positions and the known focus, *P*. Third, to quantify the spread of the population about its displaced position, we defined φ as the average of the squared distances of the observed individuals from *C*. Fourth, to quantify the total distance moved from the focus *P*, we defined Δ to be the average distance of observed individuals from *P*; note that this statistic is an amalgam of the information in δ and φ. Fifth, for the aphid data in set one only (see below), to quantify the relative degree of movement between and within rows of barley plants, we calculated, for each individual, $\cos^2\theta$, where θ denotes the angle subtending the line joining the position of that individual to *P* and the line from *P* along the row in which *P* is located. We then defined γ to be the average of these values of $\cos^2\theta$, over the individuals sampled. Note that by using the squared term (rather than just $\cos\theta$), if movement was entirely along a row (in either direction) then the value of γ would be unity; if it were entirely perpendicular to the rows (in either direction) then γ would be zero; and if there was completely random movement then the expected value of γ would be 0.5. Finally, to quantify the uniformity or otherwise of the circular distribution about *P*, especially useful when there appears to be little overall displacement of the population (i.e. when δ is small), we computed ω, a modification of Fisher, Lewis and Embleton's (1987) Rayleigh statistic, defined as follows. Given coordinates, (x_i, y_i), $i = 1,...,n$, for the *n* observed individuals relative to the original focus, *P*, then the direction cosines, (x_i', y_i'), are defined as:

$$x_i' = \cos\beta_i, \quad y_i' = \sin\beta_i \; ; \; 0 \le \beta_i = \arctan(y_i/x_i) \le 2\pi.$$

Here β_i is the angle, measured anticlockwise, from the *x*-axis to the line joining the position of the individual at (x_i, y_i) and the focus, *P*, at (0, 0). The Rayleigh statistic is the resultant of the direction cosines:

$$R^2 = (\Sigma_i\, x_i')^2 + (\Sigma_i\, y_i')^2 .$$

For $n \geq 10$, $\omega = 3R^2/n$ approximates well to a chi-square distribution under the null hypothesis of uniformity, with three degrees of freedom. We use this test to detect asymmetrical movement away from a point source or, equivalently, large parts of the sample area in which individuals are rare or absent.

The values of φ and ω were transformed to natural logarithms to achieve normality; we then used analysis of variance with these statistics as the response variable to compare the experimental treatments (see below).

The third dataset, described below, represents an invasion from the edge of a sample area, with unknown focus. We used the backtracking algorithm (Perry, 1995b) mentioned earlier to estimate the focus of the invasion. The algorithm extrapolates back along the lines joining the predicted final, regular positions, (x_f, y_f), obtained from the SADIE algorithm, to the initial, observed positions, (x_i, y_i), of each individual. Note that the term *backtracking* refers to estimating locations at a time earlier than that observed, and that the final, regular positions are used merely as convenient reference points to aid the process of inferring past movement. Every observed point is backtracked by an identical proportion, *k*, of the distance between its observed and final positions. The backtracked position, (x_b, y_b), of any point is given by:

$$x_b = (k+1)x_i - kx_f \quad \text{and} \quad y_b = (k+1)y_i - ky_f.$$

The positions of the backtracked points approach each other, forming a denser and denser focal cluster as *k* is increased, up to a maximum density at some critical value, k^*; as *k* is increased further, they disperse, forming a cluster that is ever more diffuse. The diffuseness of

the backtracked cluster is measured by the sum of squared distances about its centroid; its relationship with k is therefore quadratic, with a minimum value at $k = k^*$. The unknown focus of the original cluster is estimated as the centroid of the backtracked cluster corresponding to $k = k^*$.

2.2 *Dataset 1*

The spatial and temporal aspects of aphid movement and spread of barley yellow dwarf virus were investigated in an experiment at IACR-Rothamsted. One hundred and twenty-five barley (*H. vulgare* cv. Puffin) plants were grown in each of eight 54 cm x 54 cm trays, in five rows (12 cm spacing) of 25 plants (2 cm spacing), in a glasshouse, under prevailing conditions, to growth stage 12 (GS12, one week old) . Sixteen infected, wingless aphids, either the bird cherry aphid, *R. padi*, or the grain aphid, *S. avenae*, were released in clip cages onto the central plant within each tray. The aphids were allowed to settle and then, after removal of the clip cages, to disperse. Insect trapping glue was applied to the edges of the trays to prevent aphids leaving. The number of aphids on each plant was recorded five times per day, every three hours between 8 am and 8 pm inclusive, for three days, with an extra recording at 8 am on the fourth day. This gave a total of 16 observations per tray. Four replicate trays were set up for each aphid species. Immediately after the last observation the plants were sprayed with insecticide and then left for three weeks to allow virus development; in this study we analyse only the aphid counts. Problems with pregermination meant that only one growth stage could be tested in this run of the experiment; the experiment was repeated with plants at growth stage 22 (GS22, two weeks old), again under prevailing conditions, so that growth stage and other factors, especially temperature, differed between the two runs.

For each observation time, the locations of the observed counts form a map of the

approximate positions of the aphids, assuming that each aphid was at the centre of the plant it occupied. As discussed earlier, it is common for insects to be counted, even though more precise measurements of location may occasionally be possible. The counts of the number of aphids per plant took no account of the overlapping of leaves from plant to plant, either along or between rows. In reality, the aphids would have been distributed throughout the plants and an insect located on an outer leaf of one plant may have been located relatively far from the stem of the plant on which it was found. However, by assuming that the foliage of each plant covered a circle centred on the stem with a particular radius related to its growth stage, and that the two-dimensional position of each aphid was equally likely to have been anywhere within that circle, we were able to simulate possible positions of individual aphids. Appropriate radii were taken to be 1.5 cm and 7.5 cm for plants at GS12 and GS22, respectively (Fig. 1). The smaller radius represents an overlap between adjacent plants within but not between rows, whilst the larger represents plants which also overlap between adjacent rows. As a partial verification of the robustness of analyses based on this method of generating aphid positions, we produced 20 different simulations for one of the trays for observation 16, that had 13 aphids on nine plants, and calculated I_p for each.

We analysed two subsets of the original data. First, aphids from observation 16, by which time the aphids would have been expected to have dispersed the furthest. As an example, the original counts and the simulated aphid positions for two trays at observation 16 are shown in Fig. 2. Second, since many of the released aphids were unrecorded at observation 16, being either dead, on the soil or hidden, observations 13, 14 and 15 were combined, i.e. accumulated, for two replicate trays of each species and growth stage. We thereby obtained larger sample sizes, representing the totality of occupied aphid positions over the entire nine-hour period of the third afternoon. Note that some aphids will be

observed on more than one occasion and so there may be some dependence in the observed locations both within the accumulated dataset and between the accumulated dataset and observation 16; caution is required in interpretation. Also, note that only one simulation was produced for the accumulated dataset. For each subset of simulated aphid positions (occasion 16 or accumulated), I_p, δ and φ were computed for each tray. For comparison, Δ and γ were also computed for the original count data and the generated aphid positions on occasion 16.

2.3 *Dataset 2*

The effects of fertiliser and growth-regulating chemicals on the spread of common couch grass, *E. repens*, in the presence or absence of competing weed species were investigated in an experiment at IACR-Long Ashton, between 1986 and 1988 (Marshall, 1990). Eighteen pairs of 3 m x 3 m plots were laid out in a three-block randomised plaid design. Each pair of plots was either left bare or sown with a mosaic of six grass species. Fertiliser was applied to one plot in each pair and chemical treatments were applied across pairs of plots. Twenty single-node segments of couch rhizome were planted at the centre of each plot in April 1986. The positions of emerging couch tiller complexes were recorded (by eye) for half the plots in autumn 1987 and for the remainder in autumn 1988. In this paper, only the initially bare plots assessed in autumn 1988 are analysed, as growth was often minimal or too dense for accurate recording on the other plots. The treatments therefore form a 4 x 2 factorial with four growth-regulating chemicals (A, "Mowchem", Rhône-Poulenc; B, "Holdfast", ICI; C, "Commando", Shell; D, unsprayed) and two levels of fertiliser (none (-) or 120 kg ha^{-1} (+)).

The hand-drawn maps of tiller positions were digitised using the R$DIGITISE program and a Summasketch II Professional (Summagraphics, Seymour, CT, USA) digitising tablet. Of the 24 plots studied, three had too sparse records ($\leq$ 10 tillers per plot) for analysis.

A further four had couch growth recorded outside the defined plots; this external growth was ignored. A subset of the data is shown in Fig. 3. I_p, δ, φ and ω were computed for each plot.

2.4 *Dataset 3*

The lupin aphid, *M. albifrons*, a native of North America, was first observed in the UK at the Royal Botanic Gardens, Kew, south-west London, in September 1981 (Stroyan, 1981). However, as the infested plants at Kew were grown from seed, rather than imported, and destroyed after the infestation was discovered, it is likely that the invasion began elsewhere within the south-west London area. This area was an important horticultural centre at the time with many nurseries importing plants, but other possible sources within this area include Heathrow airport and the Royal Horticultural Society gardens at Wisley (Bartlett, 1993). The aphid later spread throughout Great Britain and into continental Europe (Piron, 1987; Bartlett, 1993). Between 1982 and 1984, a record of sightings of this devastating pest was made by one of us (PWB) at the Central Science Laboratory (CSL), Harpenden, UK, by means of samples received or through reports in the public media. Information on locations where this aphid was sighted differed in precision, but each location was assigned an Ordnance Survey (OS) grid reference. For example, sightings at university campuses or institutions could be located quite accurately, whilst sightings at private addresses were recorded as the centre of the city, town or village in which they were located. The resulting dataset was divided into aphid sightings for the cumulative periods September 1981 to 7 June 1982 (period A, 11 points, Fig. 4a), September 1981 to 30 June 1983 (period B, 26 points, Fig. 4b) and September 1981 to 30 June 1984 (period C, 61 points, Fig. 4c). A further set of 52 sightings, known only to have occurred sometime during 1982 and 1983, was collected by

E.A. Ellis after publication of articles in local and national newspapers (Fig. 4d). Ellis' dataset was combined with that for period C to investigate the effect of uneven sampling intensity (period D, 113 points, Fig. 4d). In period D, the sightings were roughly equally distributed within (44% of sightings) and outside East Anglia. For period C, the distribution of sightings was more uneven with only 38% of sightings occurring in East Anglia. However, within East Anglia the sightings were visibly aggregated to the north-east. Data for each period were analysed using the backtracking methods described earlier; the SADIE algorithm requires that the sample area be rectangular, so here it was defined as a rectangle which just enclosed mainland Great Britain. The fact that this included a small area of sea was not thought likely to affect precision greatly.

3. Results

3.1 *Dataset 1*

The number of aphids, n_a, recovered per tray ranged from 5 to 15 for observation 16, and the number of aphid observations, n_o, ranged from 18 to 45 for the accumulated period. More aphids were recovered from the experiment with plants at GS22 than at GS12 ($P<0.001$, Table 1a, observation 16; $P=0.019$, Table 2, accumulated period). This was probably the result of the considerably higher ambient temperatures present throughout the GS12 replicate of the experiment. Where necessary, the difference in sample size was accounted for by weighting in the analyses. Losses did not differ significantly between the species.

The values of I_p obtained for the test of robustness with 20 simulations ranged between 1.178 and 1.766, although most were close to the mean of 1.515 (s.e.m.= 0.030). The level of consistency in the results, as reflected in the small s.e.m., led us to conclude that it was reasonable to use a single simulation as a representative map in subsequent analyses.

I_p ranged from 0.972 to 1.625 for observation 16, and 11 out of the 16 trays showed significant aggregation (P<0.05). For the accumulated subset, I_p ranged from 2.058 to 2.666, and all eight trays showed significant levels of aggregation (P<0.02). The greater degree of aggregation demonstrated for the accumulated dataset than for observation 16 illustrates the lesser degree of dispersal in the former dataset. However, there was no significant difference in the degree of aggregation between the two species or growth stages for either time period (Tables 1a and 2).

The displacement distance, δ, did not differ significantly between the two species for observation 16, although aphids were displaced further from the point of release (P=0.029) for the younger (GS12) plants (Table 1a). This trend was also seen for the accumulated subset but the difference was not significant (P=0.155, Table 2).

There was no significant difference between the mean values of the variable representing spread, $\ln(\varphi)$, for the two species or growth stages for either time period (Tables 1a and 2).

At observation 16, the average distance moved, Δ, based on simulated aphid positions, did not differ significantly between growth stages or species (Table 1a), but for the count data Δ appeared larger (P=0.029) for GS12 than for GS22 (Table 1b). Similarly, at observation 16, γ did not differ significantly between species or growth stages for the simulated data (Table 1a), but for the observed counts γ appeared larger (P=0.033) for GS22 than for GS12 (Table 1b). For both GS12 and GS22 with count data, and GS12 with simulated data, movement appeared to be predominantly along the central row (t_{12}=2.42, P=0.032, t_{12}=5.83, P<0.001 and t_{12}=2.44, P=0.032, respectively, against the null hypothesis that γ=0.5). For GS22 with simulated data there was insufficient evidence to reject the hypothesis of isotropic movement (t_{12}=1.48, P=0.164).

3.2 *Dataset 2*

The number of tillers, n_t, recorded per plot ranged from 20 to 680, although the average number, after transformation to natural logarithms, did not differ significantly between chemical treatments or fertiliser applications. None of the other statistics given below, are in any case sensitive to the number of tillers per plot.

I_p differed between the chemicals (P=0.045) and the fertiliser rates (P=0.040). Chemicals A and B produced more highly aggregated growth than chemical C and the control (D) (Table 3, Fig. 3). Couch growth on fertilised plots was more aggregated than on unfertilised plots (Table 3, Fig. 3).

The displacement distance, δ, was smaller (P=0.048) on control plots (D) than on chemically treated plots (Table 3, Fig. 3), but there was no significant difference between fertilised and unfertilised plots. However, there was a significant interaction between chemical treatment and fertiliser application (P=0.019), due mainly to treatments C and D: mean displacements for chemicals A and B were similar in both fertilised and unfertilised plots (Figs 3a and 3b), but the displacement for the untreated and unfertilised plots (D-, Fig. 3e) was particularly small, and that for chemical C on fertilised plots (C+, Fig. 3d) was relatively small, compared with other treatment combinations.

There were no significant differences in the degree of spread as measured by $\ln(\varphi)$, either for the chemical treatments or the fertiliser applications. However, the couch tillers showed large differences in the directional pattern of spread, as measured by $\ln(\omega)$, between chemicals (P=0.008) and also some difference between fertiliser rates (P=0.032) (Table 3). Overall, chemicals A and B tended to produce similar patterns of spread in which invasion appeared to be severely restricted within certain parts of the study area (Figs 3a and 3b) after

allowing for the total spread. By contrast, for chemicals C and D invasion was more uniform radially, again allowing for total spread (Figs 3c to 3f). The untreated (D) plots (Figs 3e and 3f) showed the most uniformity. However, there was a significant interaction between the chemical and fertiliser treatments (P=0.024), mainly due to the very uniform growth observed on untreated, unfertilised (D-) plots (Fig. 3e).

3.3 *Dataset 3*

Aggregation within Great Britain was extreme, since only a small part of Great Britain was invaded (Fig. 4); the values of I_p obtained for the periods A, B, C and D, respectively, were 2.555, 3.537, 4.513 and 6.072 (all P<0.02). The backtracking estimates of the focus of dispersal of the aphid were located at OS grid references TQ 174 576, TQ 111 335, TQ 546 612 and TQ 608 807, for periods A, B, C and D, respectively. These estimates are approximately 19 km, 44 km, 39 km and 43 km from Kew, about 11 km, 25 km, 48 km and 59 km from Wisley, and about 21 km, 42 km, 49 km and 54 km from Heathrow airport, respectively (Fig. 5). The centroids of the observed data were located at grid references TQ 115 720, TQ 924 875, TL 151 442 and TL 243 470, respectively, i.e. about 8 km, 28 km, 68 km and 71 km, from Kew, about 15 km, 32 km, 86 km and 91 km from Wisley, and about 6 km, 19 km, 69 km and 73 km from Heathrow (Fig. 5).

4. Discussion

The analysis of dataset 1 illustrates clearly the importance of recording spatial information accurately. Using data in the form of counts on plants, aphids appear to rarely leave the central row of two-week-old (GS22) plants. However, using our simulation procedure, which we believe locates aphids more realistically within the plant canopy, aphids on plants at GS22

have the potential to be located further from the centre of their host plant than those on plants at GS12 (Fig. 1). The count data do not take into account the known differences in plant size between the two growth stages. This explains the discrepancy between the results for simulated and count data in GS22 for both γ, the statistic measuring the relative degree of movement across and along rows, and Δ, the average gross distance moved by observed individuals from the release point. We believe that the count data give a less accurate picture; indeed, Fig. 2b emphasises that, for such data, Δ will almost certainly be underestimated and γ overestimated. This suggests that when movement on larger, overlapping plants is of interest, recording of counts alone is not sufficient, although a simulation technique such as that adopted here may prove useful in attempting to recover lost information.

The major differences in dispersal in the aphid experiment occurred between the growth stages, for which the whole experiment was replicated on a different occasion. Differences between these replicates were undoubtedly real, but were probably due mainly to differences in the ambient temperature, which was much higher for the GS12 replicate. Because the growth stage effects were investigated sequentially, by separate experiments, rather than simultaneously, they were confounded with other gross effects, such as temperature, and cannot be reliably extracted. An alternative, improved design might have been to allocate the eight trays as two replicates of each of the four combinations of growth stage and species, followed by another replicate of this entire experiment; this would have allowed the same inferences regarding growth stage as regarding species, independently of ambient temperature and other changes between experimental replicates.

There were fewer problems with the analysis of the couch grass data, but even here caution is required in interpretation and lessons can be learned for application to other situations. The restriction of analysis to the bare plots, because growth on other plots was

minimal or too dense for accurate recording, implies that much useful information may have been lost. Also, had it been the case that some of the plots with a particular treatment had been omitted from recording because of their large density, then a possible bias might well have been induced. If accurate counting is impossible, it is still imperative to take a record that preserves the maximum amount of information; this might well entail the use of counts or estimated counts correct to a certain order of magnitude. Some method for the generation of a surrogate map, similar to that used here, could then be used to allow a fully-inclusive analysis of all plots.

Some of the maps in Fig. 3 showed that the invasion had most likely spread beyond, although not far beyond, the edge of the sampling area. Clearly, if the spread of a studied organism does take it far beyond the study area, all kinds of biases may enter into any or all of the six variables we have measured and any others that might be derived. To guard against this, the likely maximum degree of movement must be assessed as accurately as possible prior to dispersal experiments. Distributions of dispersal distances achieved by plants or animals will often be extremely skewed, with a long tail (Taylor, 1978) caused by exceptional individuals who disperse many times the average distance of the population. If possible, a greater area than that likely to be covered should be assigned to dispersal experiments, the outer parts of which may be deleted from analysis if they are not reached. Fortunately, these problems seemed not to be excessive for the data analysed here; results for both experimental aphids and couch showed that the observed patterns following dispersal from a single, central focus remained highly aggregated, despite some dispersal to the edges of experimental areas.

The analysis of variance of I_p, δ, etc., is itself an *ad hoc* procedure. These statistics may not satisfy the usual assumptions of analysis of variance - they are not necessarily normally distributed random variables with constant variance. The results of such an analysis

should, therefore, be interpreted with caution. For the examples presented here, extreme variations from the standard assumptions did not occur; we therefore believe our conclusions are valid. General solutions to the problem of how to analyse replicated (point) patterns or patterns arising from application of different treatment combinations have yet to be established, although some progress has been made in this area (Diggle, Lange and Beneš, 1991; Baddeley *et al*., 1993; Mugglestone, 1996).

Both δ and φ were found to be simple statistics that could describe well some of the gross properties of dispersal of a population. They have obvious analogues to statistics that estimate, respectively, convection and diffusion parameters in diffusion models. These statistics were concerned with distance rather than direction. The parameter Δ was less effective, and might well be thought somewhat redundant for studies where both δ and φ are measured. By contrast, the statistics γ and ω concerned direction rather than distance. Both proved useful. The parameter γ would only be appropriate for studies within a two-dimensional framework where two perpendicular dimensions clearly required different treatments. By contrast, the parameter ω could help to identify such dimensional differences if they were not initially clear from the experimental set-up. By further contrast, the measurement of spatial pattern, through I_p or through any other single statistic, is mostly concerned with the relationship of the individuals that make up the population to each other, rather than to aspects of the population as a whole. Any noticable non-uniformity or non-randomness in the measures of distance or direction adopted would warn of the existence of spatial pattern, but pattern is more than just these aspects and the absence of such indications by no means precludes its existence. For example, a distribution in which the individuals of a population had radiated outwards uniformly from the source to spread themselves evenly over an entire study region, so that their inter-individual spacing was constant, would yield

unremarkable values of δ, φ, γ and ω, and yet display strong regularity in their spatial pattern.

Marshall (1990) established differences between fertilised and unfertilised plots in rhizome growth variables (weight, length and bud number). It is apparent here that the degree of aggregation and the uniformity of growth are also affected by fertiliser application and growth-regulating chemicals. Here, if growth was symmetric, then aggregation would be related directly to rate of spread within the plot. However, only in the unfertilised control plots was there insignificant evidence of asymmetrical growth. Addition of nitrogen can affect rhizome growth and bud dormancy (Leakey, Chancellor and Vince-Prue, 1978), which might lead to variable patterns. Chemical treatments A (mefluidide) and B (paclobutrazol), both growth retardants that affect cell extension or division, resulted in distributions that were significantly more aggregated (less uniform) than other treatments, probably reflecting an increase in rhizome bud numbers and a decrease in rhizome internode length. Chemical C (flamprop-M-isopropyl) had only minor effects on aggregation/uniformity, possibly reflecting only temporary growth inhibition (Richardson and Parker, 1976). The effects of the underlying environment and habitat heterogeneity on dispersal have been noted by several authors. Mortimer and McMahon (1982) proposed that rhizome apices of couch grass were able to respond to resources within the soil, thus explaining the pattern of shoot development. Bergelson, Newman and Floresroux (1993) demonstrated the impact of heterogeneity of gaps within vegetation on patterns of weed spread. Camann *et al*. (1995) found that spatial heterogeneity in the pattern of disease incidence in *Arachis hypogaea* may be caused by cultivar differences. Nelson and Campbell (1993) showed that spatial pattern in viral, fungal and bacterial foliar epidemics in *Trifolium repens* depended on host resistance, pathogen ecology and dispersal methods, defoliation and environmental factors, including year, weather and cultural practices. They concluded that an understanding of all these factors was

necessary for management of plant disease epidemics.

The lupin aphid data were divided into several cumulative periods, rather than into discrete datasets representing new sightings in consecutive periods, under the assumption that sightings up to early June of a particular year were the result of dispersal during the previous summer and not more recently. Records of an invading species often do not list individual sightings from areas in which the organism is already known to be established. This is particularly true of abundant and highly mobile insects, such as aphids; it is unlikely, for example, that the lupin aphid was absent from the London area in the years between June 1982 and June 1984, as suggested by the new CSL records during that time (Figs 4b and 4c) - this is confirmed by Ellis' independent data (Fig. 4d). The aggregation measured was a function both of the spread and number of aphid sightings. Both the backtracked estimates and the centroids of the original data appear to be stable estimates of the source of the lupin aphid invasion for periods A and B, which cover the early stages of the invasion. However, for period C, during which time the invasion had consolidated, the backtracked estimate was notably less affected by changes in sampling occasion than the centroid. No method is likely to be able to pinpoint the source of an invasion very accurately, and that is certainly the case for these data. The four different periods analysed here represent information gained at different stages of the invasion; other studies might not gather such detailed information. It is therefore important that any method is reasonably robust to changes in sample period. The backtracked estimates were not ideal, but were more robust than the centroids of the data, and give possibly the best current estimate possible. Possibly, the accuracy of the estimates falls as dispersal increases, and period A was probably the most reliable, despite the smaller number of sightings involved. The distances between successive estimates give some indication of the variability of the estimates. The results presented here do not contradict

greatly the previous conclusions of Bartlett (1993), that the initial introduction was within the London area. Our results are not inconsistent with a lupin aphid invasion initiated within the smallest area that contains Kew, Wisley and Heathrow airport. These results illustrate the potential for using the backtracking method when estimating dispersal from a point source near the edge of the sample area. Caution is advised, however, in the use of this and any other method, because of sensitivity to unequal and non-uniform sampling efforts. The results presented here suggest that the backtracking method is not seriously affected by unequal sampling intensity, as the estimates for periods C and D do not differ greatly, whilst the estimates from the centroid are strongly affected by the preponderance of points in East Anglia.

Our results have shown that the analysis of dispersal data from a single focus has certain problems, some of which may originate in the design of the experiment or sample survey. Neither of the two experiments were conducted with the objective of testing the methodology used to analyse the data, and both were somewhat flawed in design. The aphid experiment suffered badly from the confounding of host plant growth stage with experimental conditions such as temperature. In the couch grass experiment a proportion of the individuals dispersing left the study area; these might even be the most interesting individuals in the population, possessing an important, though unknown selective advantage. Despite this, we have brought out important aspects of the data, using several different techniques and measures, each of which describes a different aspect of the spatial characteristics of dispersal patterns. None of these alone can provide an adequate description of all aspects of the observed spatial pattern, but in combination they help to provide a clear picture of dispersal from a single, aggregated focus and augment simple observations on plant growth and insect movement. Similar approaches would probably prove useful for most experimental studies of

dispersal from a single focus, with the construction of further *ad hoc* statistics to answer specific questions of interest.

Acknowledgements

S. Korie was funded by a Rothamsted International Staff Fellowship. I.N. Smyrnioudis conducted the aphid movement experiments (dataset 1) as part of an MSc project for the University of Newcastle upon Tyne. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

References

Agassiz, J.L., Nash, D.R., Godfray, H.C.J. and Lawton, J.H. (1994) Estimating the spread of small invading organisms using public information. *Global Ecology and Biogeography Letters* **4**, 1-8.

Alston, R. (1994) *Statistical analysis of animal populations*. Ph.D. thesis, University of Kent.

Augustin, N.H., Mugglestone, M.A. and Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* **33**, 339-47.

Baddeley, A.J., Moyeed, R.A., Howard, C.V. and Boyde, A. (1993) Analysis of a three-dimensional point pattern with replication. *Applied Statistics* **42**, 641-68.

Banks, P.M., Kareiva, P.M. and Zia, L. (1988) Analyzing field studies of insect dispersal using two-dimensional transport equations. *Environmental Entomology* **17**, 815-20.

Bartlett, P.W. (1993) Essig's lupin aphid, (*Macrosiphum albifrons* Homoptera: Aphididae), introduction and spread in Europe. *Plant Health and the European Single Market*, (ed D.L. Ebbels), pp. 361-6. British Crop Protection Monograph no. 54, Farnham, Surrey.

Bergelson, J., Newman, J.A. and Floresroux, E.M. (1993) Rates of weed spread in spatially heterogeneous environments. *Ecology* **74**, 999-1011.

Bliss, C.I. (1941) Statistical problems in estimating populations of Japanese beetle larvae. *Journal of Economic Entomology* **34**, 221-32.

Boag, B., Neilson, R., Palmer, L.F. and Jones, H.D. (1994) A further record of the New Zealand flatworm *Artioposthia triangulata* (Dendy) in England. *Plant Pathology* **43**, 220-2.

Camann, M.A., Culbreath, A. K., Pickering, J., Todd, J.W. and Denski, J.W. (1995) Spatial and temporal patterns of spotted wilt epidemics in peanut. *Phytopathology* **85**, 879-85.

Clark, S.J., Perry, J.N. and Marshall, E.J.P. (1996) Estimating Taylor's power law parameters for weeds and the effect of spatial scale. *Weed Research* **36**, 405-17.

Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

Diggle, P. J., Lange, N. and Beneš, F.M. (1991) Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association* **86**, 618-25.

Dobzhansky, T.H. and Wright, S. (1943) Genetics of natural populations. Dispersal rates in *Drosophila pseudobscura*. *Genetics*, Priceton **28**, 304-40.

Fisher, N.I, Lewis, T. and Embleton, B.J.J. (1987) *Statistical Analysis of Spherical Data*. Cambridge University Press, Cambridge.

Fleming, W.E. and Baker, F.E. (1936) A method for estimating populations of larvae of the Japanese beetle in the field. *Journal of Agricultural Research* **53**, 319-31.

Freeman, G.H. (1977) A model relating numbers of dispersing insects to distance and time. *Journal of Applied Ecology* **14**, 477-87.

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

Killion, M.J. and Grant, W.E. (1995) A colony-growth model for the imported fire ant - potential geographic range of an invading species. *Ecological Modelling* **77**, 73-84.

Leakey, R.R.B., Chancellor, R.J. and Vince-Prue, D. (1978) Regeneration from rhizome fragments of *Agropyron repens* (L.) Beauv. III. Effects of nitrogen and temperature on the development of dominance amongst shoots on multi-node fragments. *Annals of Botany* **42**, 197-204.

Lonsdale, W.M. (1993) Rates of spread of an invading species - *Mimosa pigra* in northern Australia. *Journal of Ecology* **81**, 513-21.

Mack, R.N. (1985) *Invading plants: their potential contribution to population biology.* Studies on Plant Demography: A Festschrift for John L. Harper (ed J. White), pp. 127-42. Academic Press, London.

Marshall, E.J.P. (1988) Field-scale estimates of grass weed populations in arable land. *Weed Research* **28**, 191-8.

Marshall, E.J.P. (1990) Interference between sown grasses and the growth of rhizome of *Elymus repens* (couch grass). *Agriculture, Ecosystems and Environment* **33**, 11-22.

Mortimer, A.M. and McMahon, D.J. (1982) Time, space and the growth of couch grass. *Proceedings British Crop Protection Conference - Weeds* (1982), pp. 771-8. British Crop Protection Council.

Mugglestone, M.A. (1996) The role of tessellation methods in the analysis of three-dimensional spatial point patterns. *Journal of Agricultural, Biological and Environmental Statistics* **1**, 141-53.

Nash, D.R., Agassiz, D.J.L., Godfray, H.C.J. and Lawton, J.H. (1995) The pattern of spread of invading species: two leaf-mining moths colonizing Great Britain. *Journal of Applied Ecology* **64**, 225-33.

Nelson, S.C. (1996) A simple analysis of disease foci. Phytopathology **86**, 332-9.

Nelson, S.C. and Campbell, C.L. (1993) Comparative spatial analysis of foliar epidemics in white clover caused by viruses, fungi and a bacterium. *Phytopathology* **83**, 288-301.

Perry, J.N. (1995a) Spatial aspects of animal and plant distribution in patchy farmland habitats. In *Ecology and Integrated Arable Farming Systems* (eds D. Glen, M. Greaves and H.M. Anderson (eds), Wiley, Chichester, pp. 221-42.

Perry, J.N. (1995b) Spatial analysis by distance indices. *Journal of Animal Ecology* **64**, 303-14

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**, 1505-18.

Piron, P.G.M. (1987) The advance of the American lupin aphid (*Macrosiphum albifrons* Essig) (Homoptera: Aphididae) in Europe. *Journal of Applied Entomology* **103**, 111-2.

Richardson, W.G. and Parker, C. (1976) The activity and post-emergence selectivity of some recently developed herbicides: HOE 22870, HOE 23408, flamprop-methyl, metamitron and cyperquat. Technical Report Agricultural Research Council Weed Research Organisation No. 39, 59 pp.

Ripley, B.D. (1988) *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.

Sharov, A.A., Liebhold, A.M. and Roberts, E.A. (1997) Methods for monitoring the spread of gypsy moth (Lepidoptera: Lymantriidae) populations in the Appalachian mountains. *Journal of Economic Entomology* **90**, 1259-1266.

Sharov, A.A., Roberts, E.A., Liebhold, A.M. and Ravlin, F.W. (1995) Gypsy moth (Lepidoptera, Lymantriidae) spread in the central Appalachians: three methods for species boundary estimation. *Environmental Entomology* **24**, 1529-1538.

Sharov, A.A., Roberts, E.A. and Liebhold, A.M. (1996) Spread of gypsy moth (Lepidoptera, Lymantriidae) in the central Appalachians: comparison of population boundaries obtained from male moth capture, egg mass counts, and defoliation records. *Environmental Entomology* **25**, 783-792.

Shigesada, N. and Kawasaki, K. (1997) *Biological Invasions: Theory and Practice*. Oxford University Press, Oxford.

Silander, J.A. Jr and Pacala, S.W. (1985) Neighbourhood predictors of plant performance.

Oecologia **66**, 256-63.

Skellam, J.G. (1952) Studies in statistical ecology I. Spatial pattern. *Biometrika* **39**, 346-62.

Soderstrom, L. and Jonsson, B.G. (1989). Spatial pattern in the leafy hepatic *Ptilidium pulcherrimum*. *Journal of Bryology* **15**, 793-802.

Stewart, V.I. and Blackshaw, R.P. (1993) Genetic variation in populations of the terrestrial planarian *Artioposthia triangulata* (Dendy) and evidence for passive dispersal in Northern Ireland. *Annals of Applied Biology* **123**, 459-68.

Stroyan, H.L.G. (1981) A North American lupin aphid found in Britain. *Plant Pathology* **30**, 253.

Taylor, L.R. (1961) Aggregation, variance and the mean. *Nature* **189**, 732-5.

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

Taylor, L.R., Woiwod, I.P. and Perry, J.N. (1978) The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology* **47**, 383-406.

Taylor, R.A.J. (1978) The relationship between density and distance of dispersing insects. *Ecological Entomology* **3**, 63-70.

Tower, W.L. (1906) *An investigation of evolution in Chrysomelid beetles of the genus* Leptinotarsa. Carnegie Institute, Washington D.C., USA. 320pp.

Trewhella, W.J. and Harris, S. (1988) A simulation-model of the pattern of dispersal in urban fox (*Vulpes vulpes*) populations and its application for rabies control. *Journal of Applied Ecology* **25**, 435-50.

Williamson, M.H. (1996) *Biological Invasions*. Chapman and Hall, London.

Wilson, B.J. and Brain, P. (1991) Long-term stability of distribution of *Alopecurus myosuroides* Huds. within cereal fields. *Weed Research* **31**, 367-73.

Table 1. Results for aphid dataset 1 at observation 16 for each species and growth stage. (a) Mean logged number of recovered aphids, $\ln(n_a)$; mean aphid aggregation, I_p; mean displacement, δ, in cm; mean spread, $\ln(\varphi)$, in cm^2; mean distance moved, Δ, in cm; and mean relative between- to within-row movement, γ; all based on simulated aphid positions. (b) Δ and γ based on original aphid counts. Standard errors of differences between growth stage means in (a) are 0.11, 0.11, 0.70, 0.52, 0.25 and 0.07 for the six variables, respectively, and in (b) are 0.40 and 0.09 for the two variables, respectively; means are over four replicate trays.

	GS12			GS22		
Variable	*R. padi*	*S. avenae*	Mean	*R. padi*	*S. avenae*	Mean
(a)						
$\ln(n_a)$	2.04	2.06	2.05	2.42	2.64	2.53
I_p	1.27	1.36	1.32	1.47	1.47	1.47
δ	3.97	3.71	3.84	2.22	2.02	2.12
$\ln(\varphi)$	4.69	4.37	4.53	3.49	4.11	3.80
Δ	2.05	1.86	1.95	1.63	1.84	1.74
γ	0.61	0.65	0.63	0.55	0.60	0.58
(b)						
Δ	2.03	1.81	1.92	0.57	1.29	0.93
γ	0.65	0.65	0.65	0.90	0.82	0.86

Table 2. Results for aphid dataset 1 for the accumulated observations 13 to 15 for each species and growth stage. Mean logged number of aphid observations, $\ln(n_o)$; mean aphid aggregation, I_p; mean displacement, δ, in cm; and mean spread, $\ln(\varphi)$, in cm^2; all based on reconstructed aphid positions. Standard errors of differences between growth stage means are 0.11, 0.20, 0.85 and 0.92 for the four variables, respectively; means are over two replicate trays.

	GS12			GS22		
Variable	*R. padi*	*S. avenae*	Mean	*R. padi*	*S. avenae*	Mean
$\ln(n_o)$	3.43	3.01	3.22	3.55	3.69	3.62
I_p	2.17	2.21	2.19	2.34	2.37	2.36
δ	3.14	3.80	3.47	2.21	1.75	1.98
$\ln(\varphi)$	4.28	3.48	3.88	3.94	4.00	3.97

Table 3. Results for couch grass dataset 2 for each chemical and fertiliser treatment. Mean couch grass aggregation, I_p; logarithmic mean displacement, ln(δ), in cm; mean spread, ln(φ), in cm^2; and mean uniformity, ln(ω). Standard errors of differences between means in the body of the table are 0.44, 0.31, 0.26 and 0.64 for the four variables, respectively; means are over three replicate plots. Experimental plots were fertilised (+) or unfertilised (-), and treated with growth-regulating chemicals A, B, C (see text) or untreated (D).

		Chemical treatment				
Variable	Fertiliser	A	B	C	D	Mean
I_p	+	3.05	3.37	2.29	2.66	2.84
	-	2.70	2.62	1.89	2.13	2.34
	Mean	2.87	3.00	2.09	2.39	
ln(δ)	+	3.40	3.22	2.80	3.26	3.17
	-	3.30	3.29	3.27	2.12	2.99
	Mean	3.35	3.25	3.04	2.69	
ln(φ)	+	8.71	8.93	8.73	8.97	8.83
	-	8.43	8.92	8.98	8.94	8.82
	Mean	8.57	8.92	8.85	8.95	
ln(ω)	+	3.88	3.91	2.57	3.64	3.50
	-	3.54	3.90	2.64	0.79	2.72
	Mean	3.71	3.90	2.61	2.21	

Figure legends

Figure 1. Plan of a corner of a tray containing barley plants (•) grown 2 cm apart, within rows 12 cm apart (dataset 1). Circles denote the areas assumed covered by the plants and within which aphids were allocated randomly to generate a map. (a) GS 12, assumed radius = 1.5 cm; (b) GS22, assumed radius = 7.5 cm (not all circles drawn). The solid line represents the boundary of the tray.

Figure 2. Examples of data collected for *R. padi* on barley plants (dataset 1) at observation 16. Observed counts of aphids per plant are shown at plant centres along with the map of simulated individual aphid positions (×). (a) 11 aphids on plants as GS12; (b) 13 aphids on plants at GS22. The solid line represents the boundary of the tray.

Figure 3. Examples of couch grass tiller locations (dataset 2), recorded in autumn 1988, from plots infested with couch at a single central source (•) in 1986, and subsequently treated with (a) Holdfast but no fertiliser (B-), (b) Holdfast and fertiliser (B+), (c) Commando but no fertiliser (C-), (d) Commando and fertiliser (C+), (e) unsprayed but no fertiliser (D-), and (f) unsprayed and fertiliser (D+). The solid line represents the boundary of the experimental plot.

Figure 4. Cumulative maps of lupin aphid sightings in Great Britain between September 1981 and (a) 7 June 1982, period A; (b) 30 June 1983, period B; and (c) 30 June 1984, period C; new sightings in the latter year of each cumulative period (•) and previous sightings (❍) are distinguished. In (d), the sightings between September 1981 and June 1984 (❍), are combined with those collected by E.A. Ellis during 1982 and 1983, period D (•). Points appearing off the coastline are on the Isle of Man, Anglesey and the Isle of Wight.

Figure 5. Estimates of the focus of the lupin aphid invasion obtained for each of the periods A, B, C and D, using the backtracking method (•) and the centroids of the original datasets (❍). Lines indicate the differences between the estimates obtained from backtracking and the simple centroids of the datasets. Locations of Kew, Wisley and Heathrow Airport are shown for reference.

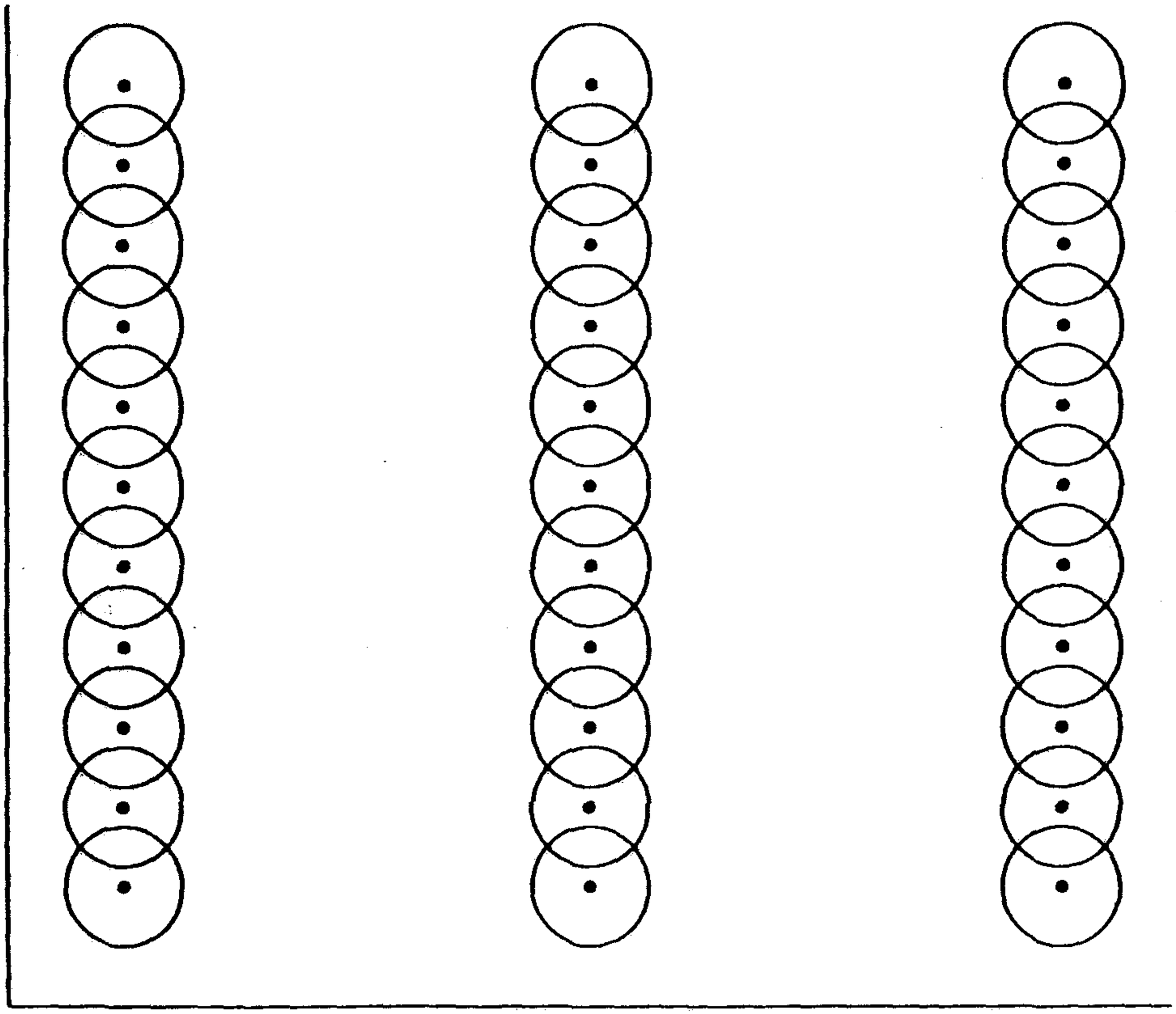

Figure 1a. Plan of a corner of a tray containing barley plants (•) grown 2 cm apart, within rows 12 cm apart (dataset 1). Circles denote the areas assumed covered by the plants and within which aphids were allocated randomly to generate a map. GS 12, assumed radius = 1.5 cm. The solid line represents the boundary of the tray.

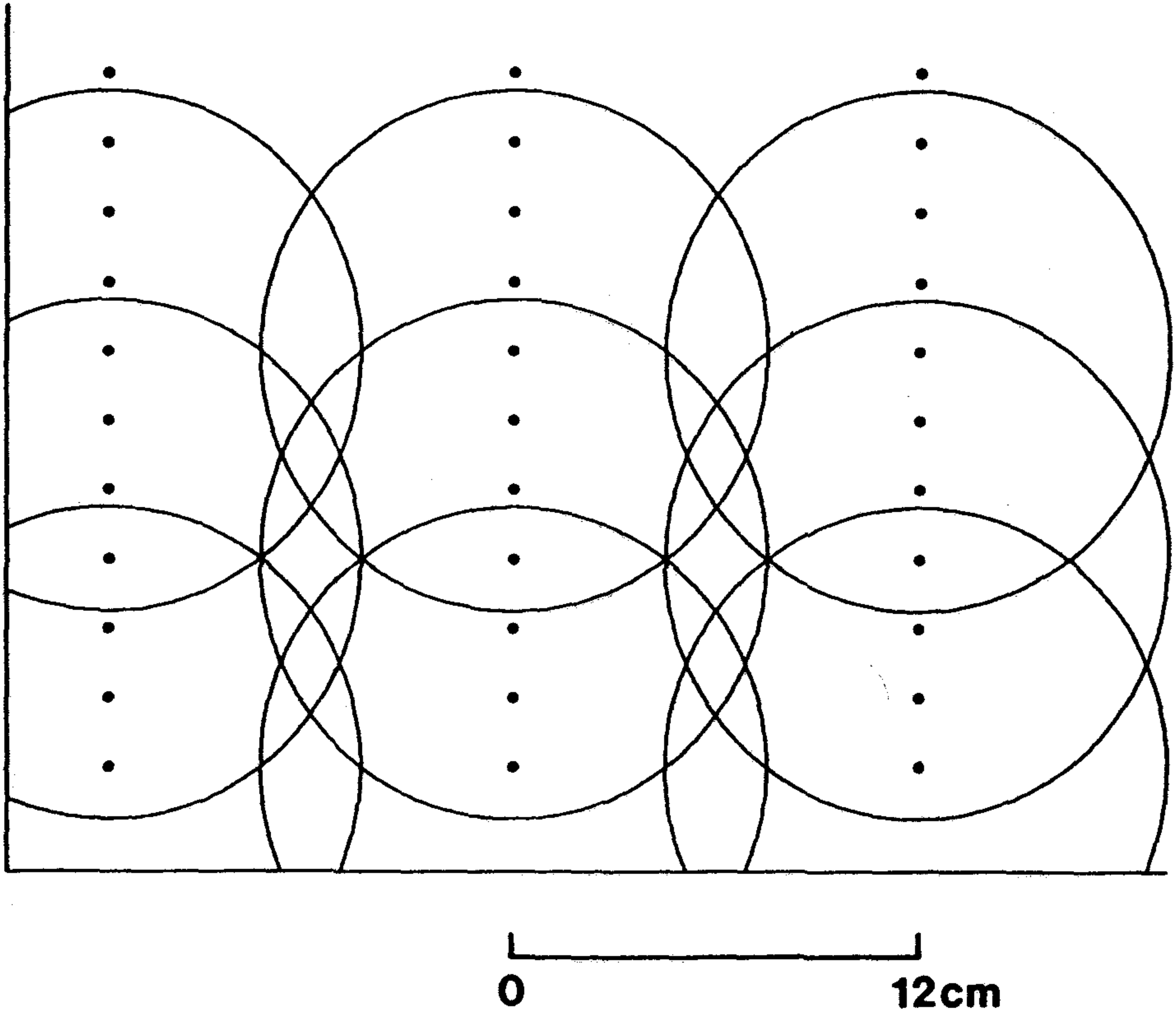

Figure 1b. GS22, assumed radius = 7.5 cm (not all circles drawn).

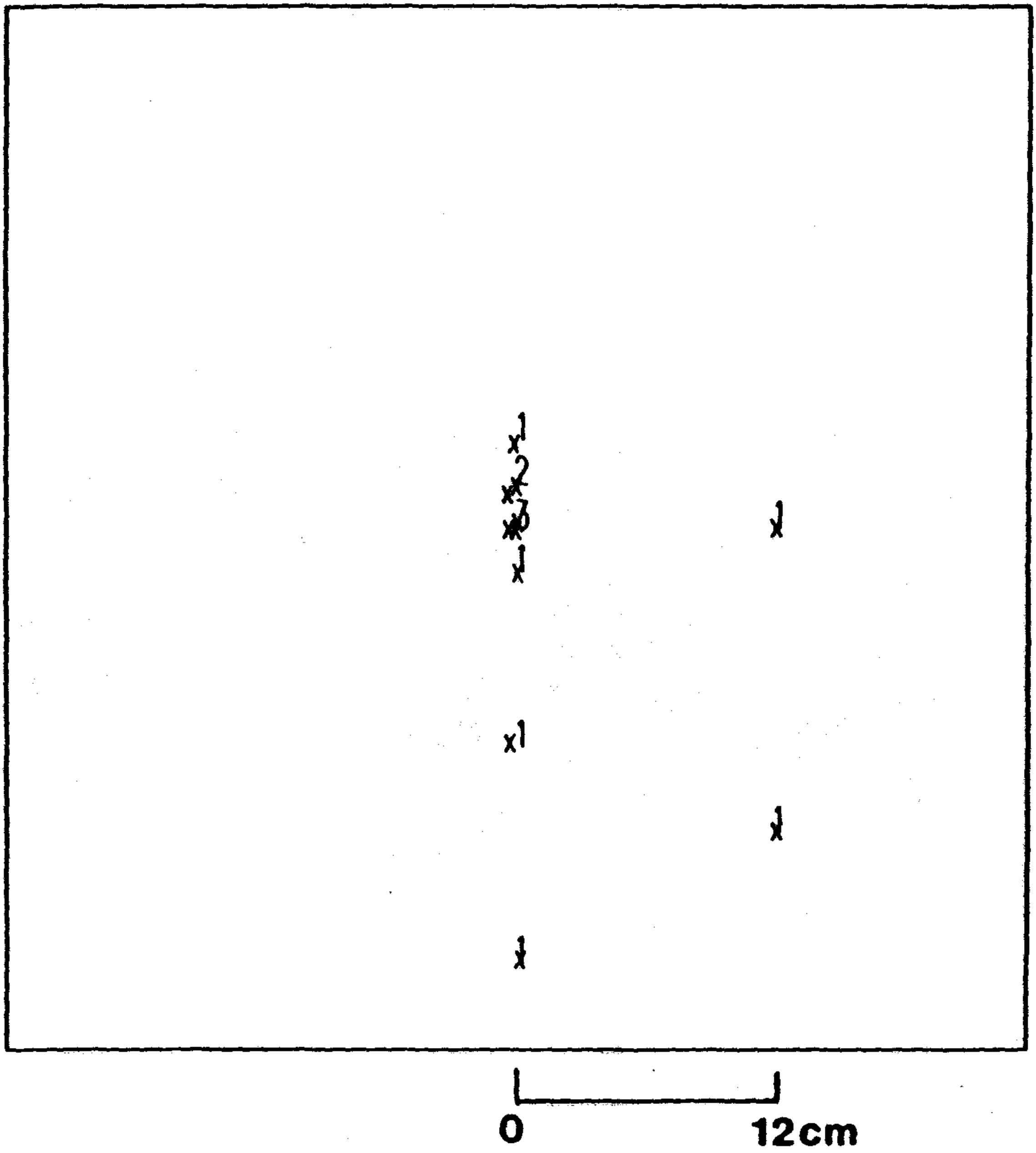

Figure 2a. Examples of data collected for *R. padi* on barley plants (dataset 1) at observation 16. Observed counts of aphids per plant are shown at plant centers along with the map of simulated individual aphid positions (×). 11 aphids on plants at GS12. The solid line represents the boundary of the tray.

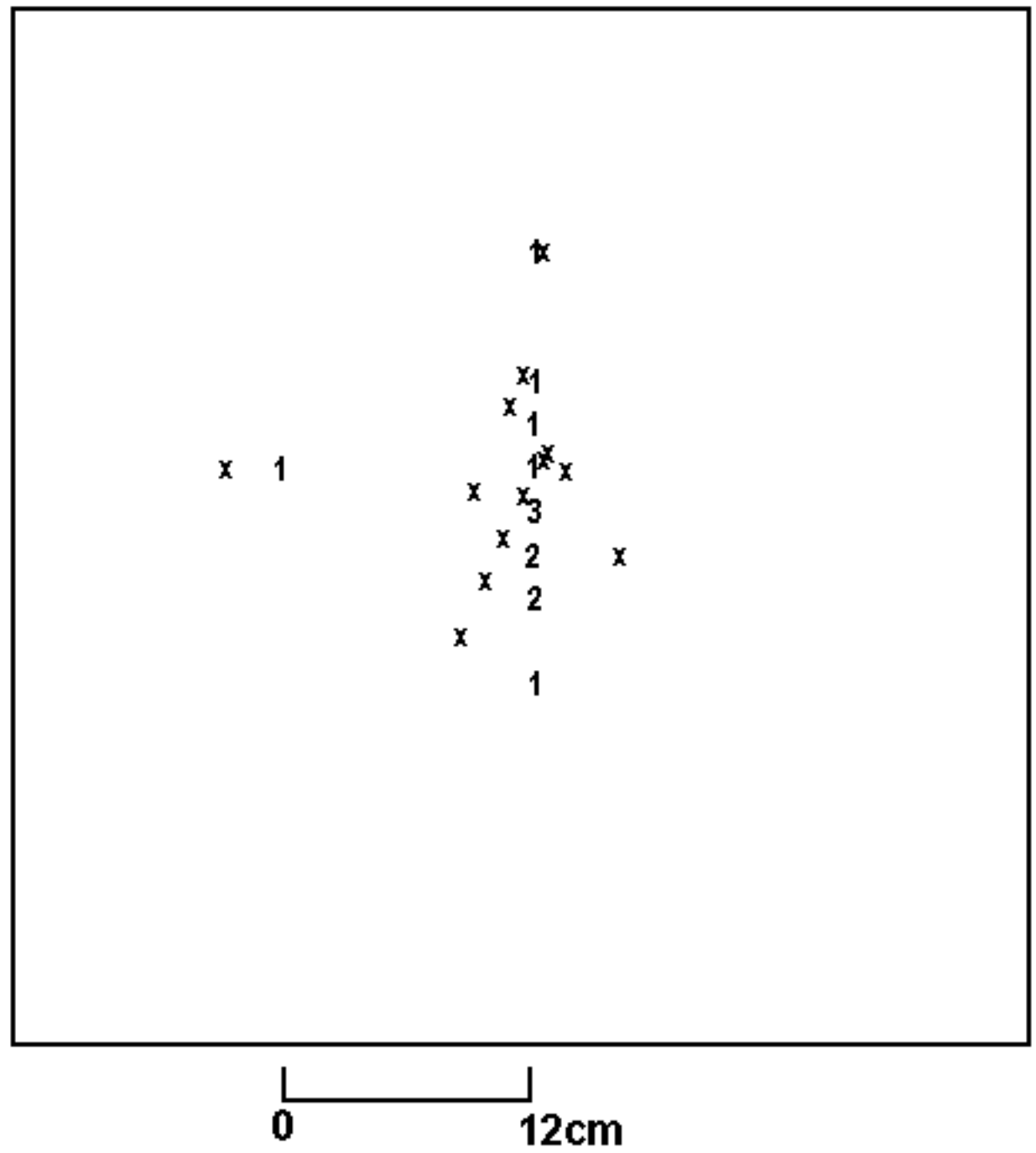

Figure 2b. 13 aphids on plants at GS22

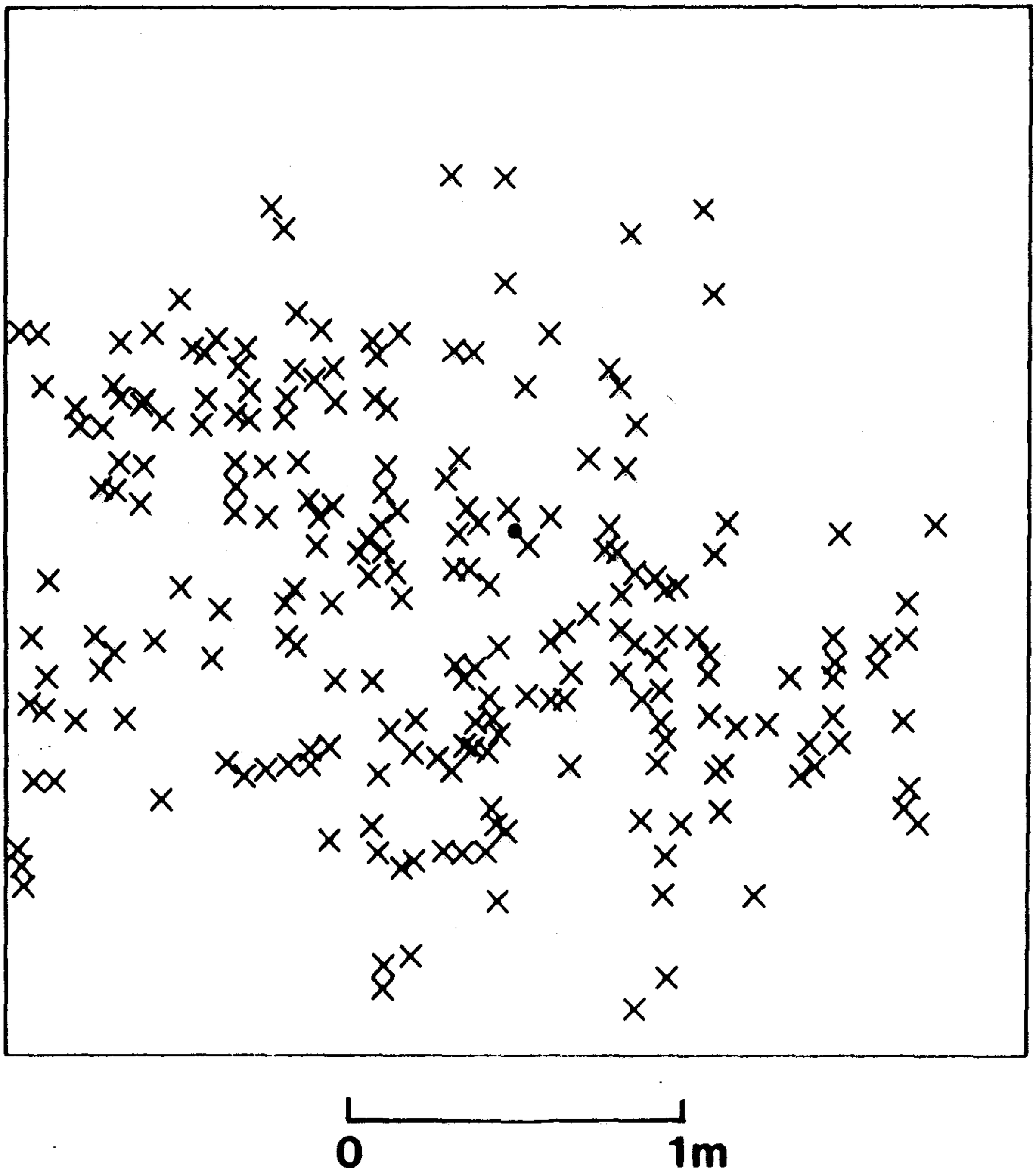

Figure 3a. Examples of couch grass tiller locations (dataset 2), recorded in autumn 1988, from plots infested with couch at a single central source (•) in 1986, and subsequently treated with Holdfast but no fertilizer (B−). The solid line represents the boundary of the experimental plot.

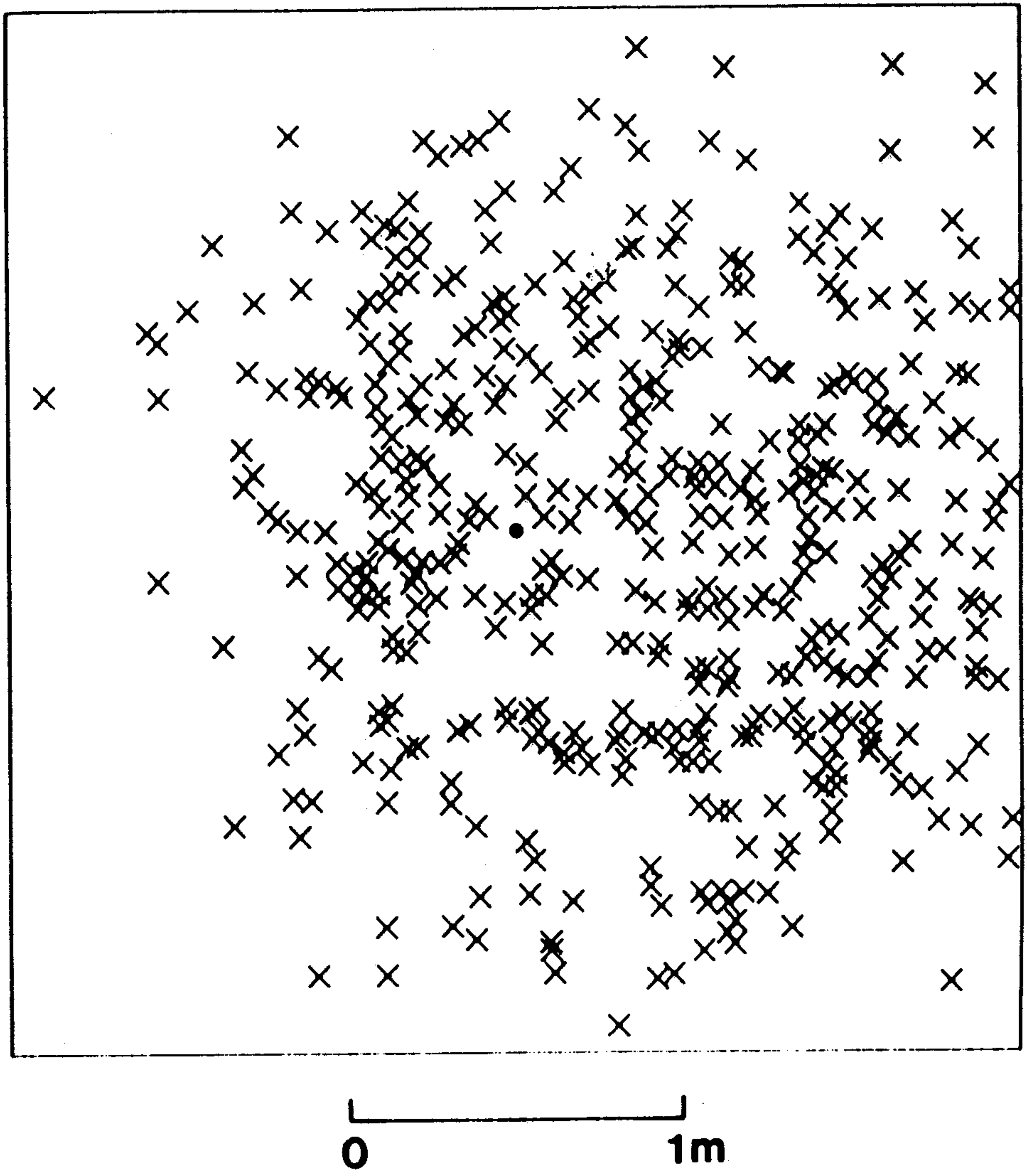

Figure 3b. Holdfast and fertilizer (B+).

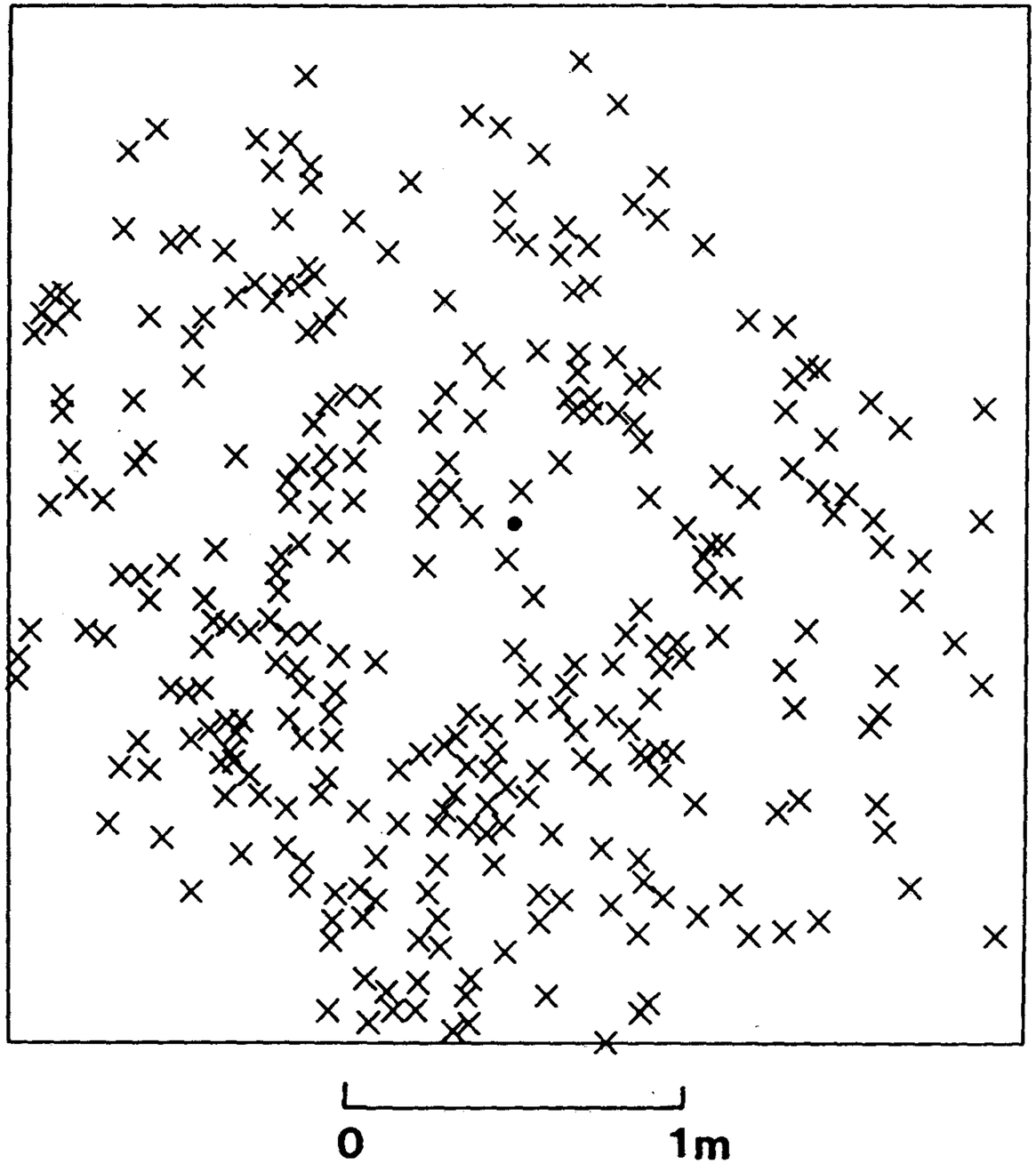

Figure 3c. Commando but no fertilizer (C–).

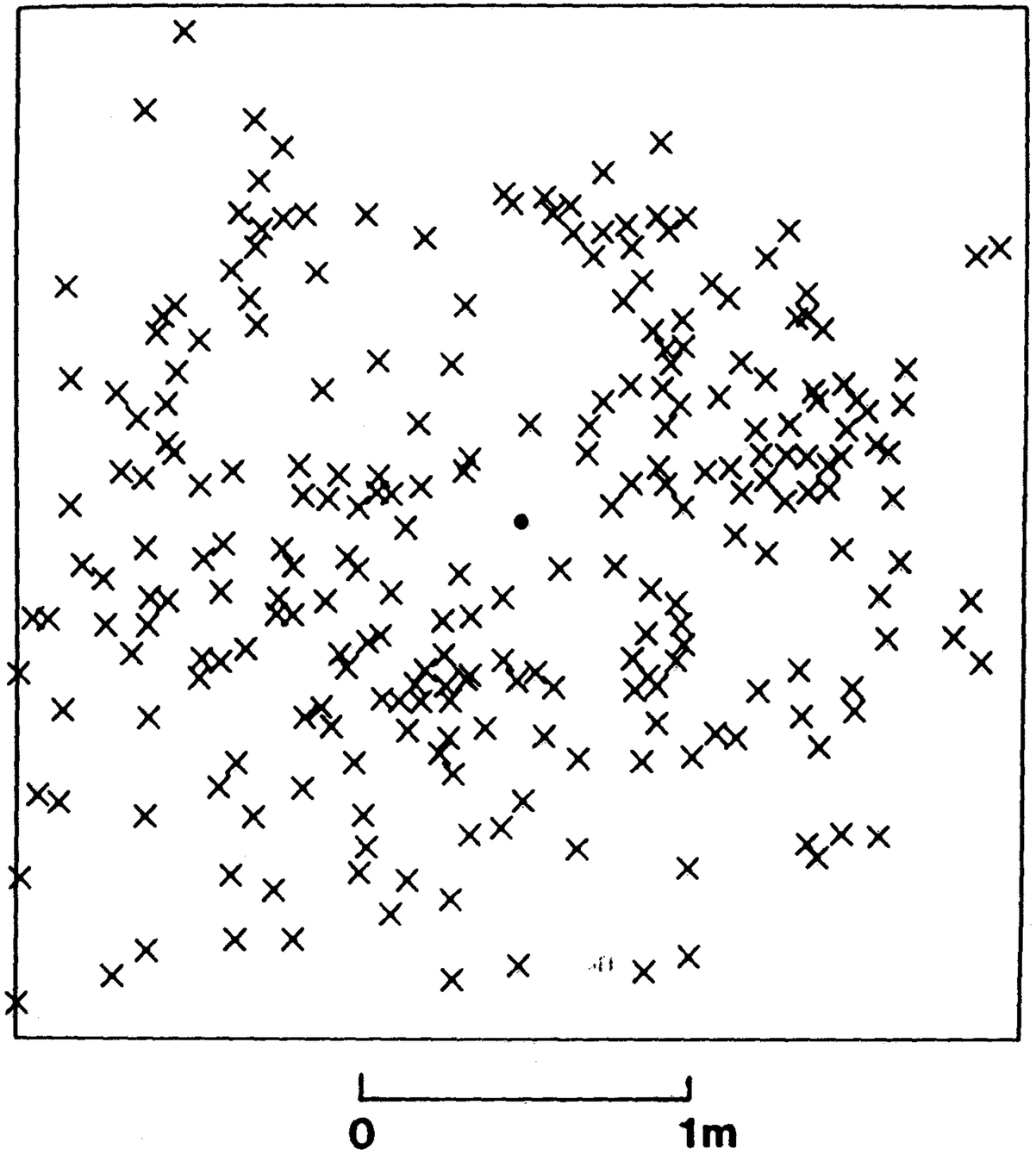

Figure 3d. Commando and fertilizer (C+).

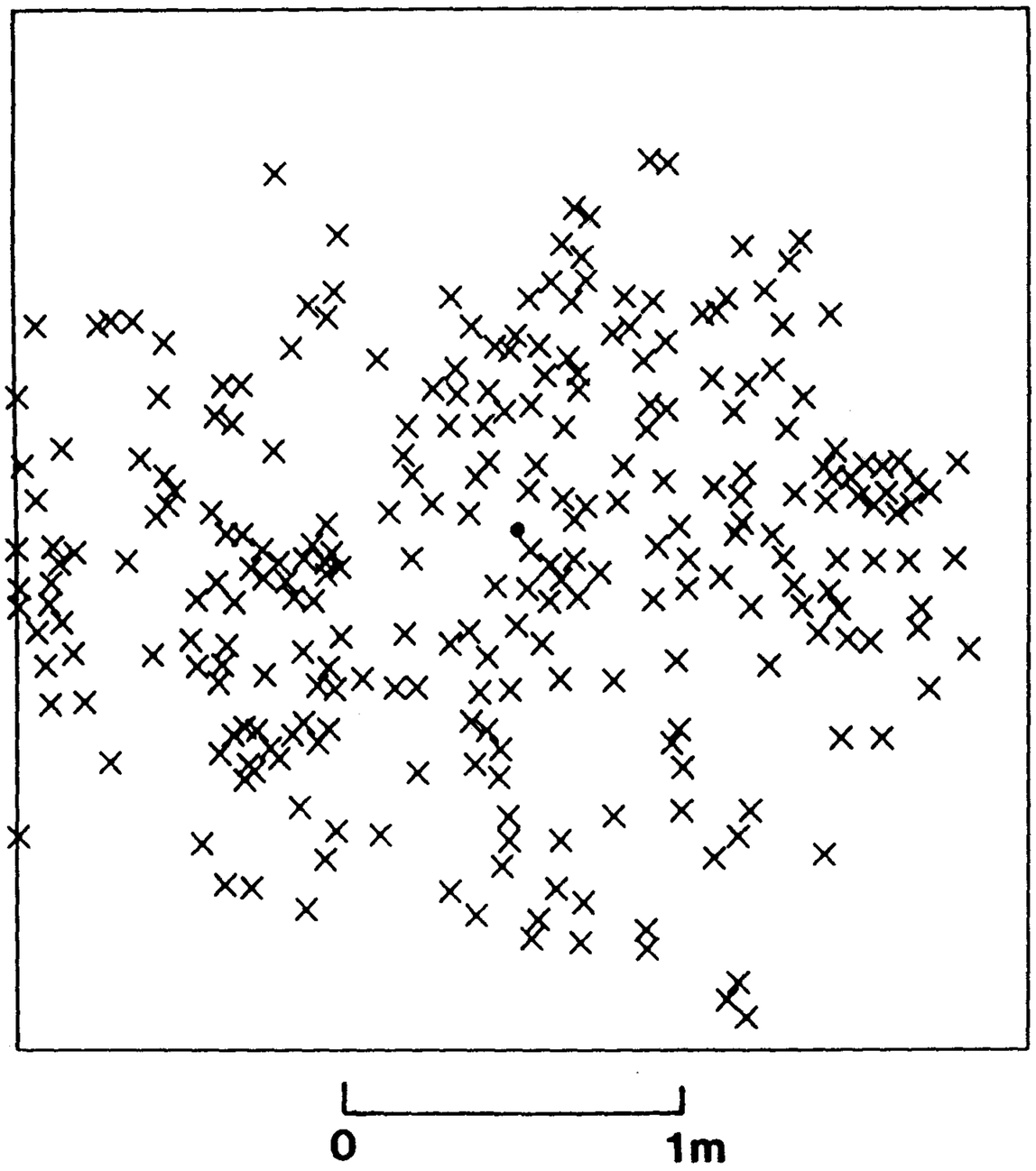

Figure 3e. Unsprayed but no fertilizer (D−).

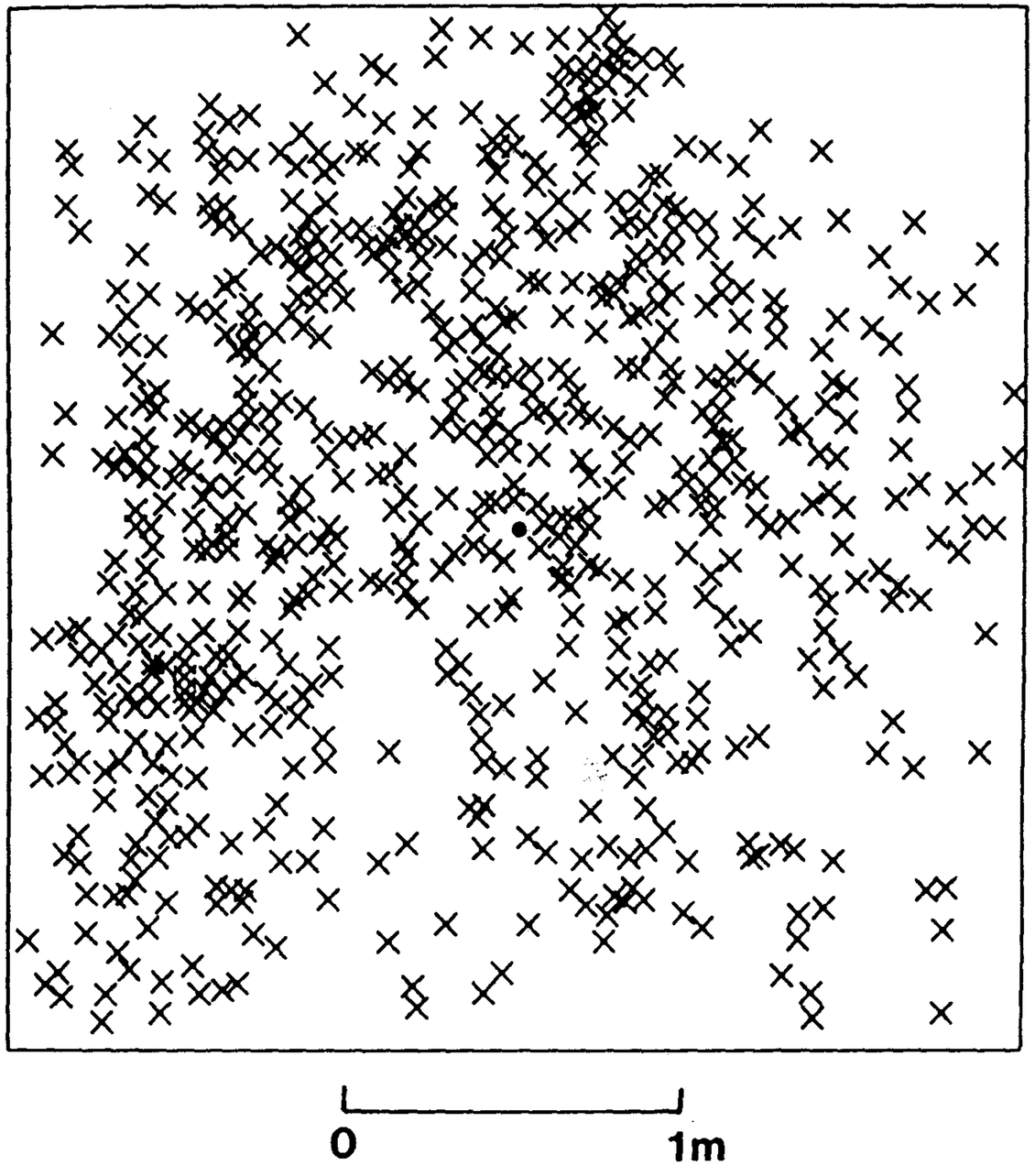

Figure 3f. Unsprayed and fertilizer (D+).

Figure 4. Cumulative maps of lupin aphid sightings in Great Britain between September 1981 and (a) 7 June 1982, period A; (b) 30 June 1983, period B; and (c) 30 June 1984, period C; new sightings in the latter year of each cumulative period (•) and previous sightings (o) are distinguished. In (d), the sightings between September 1981 and June 1984 (o), are combined with those collected by E.A. Ellis during 1982 and 1983, period D (•). Points appearing off the coastline are on the Isle of Man, Anglesey and the Isle of Wight.

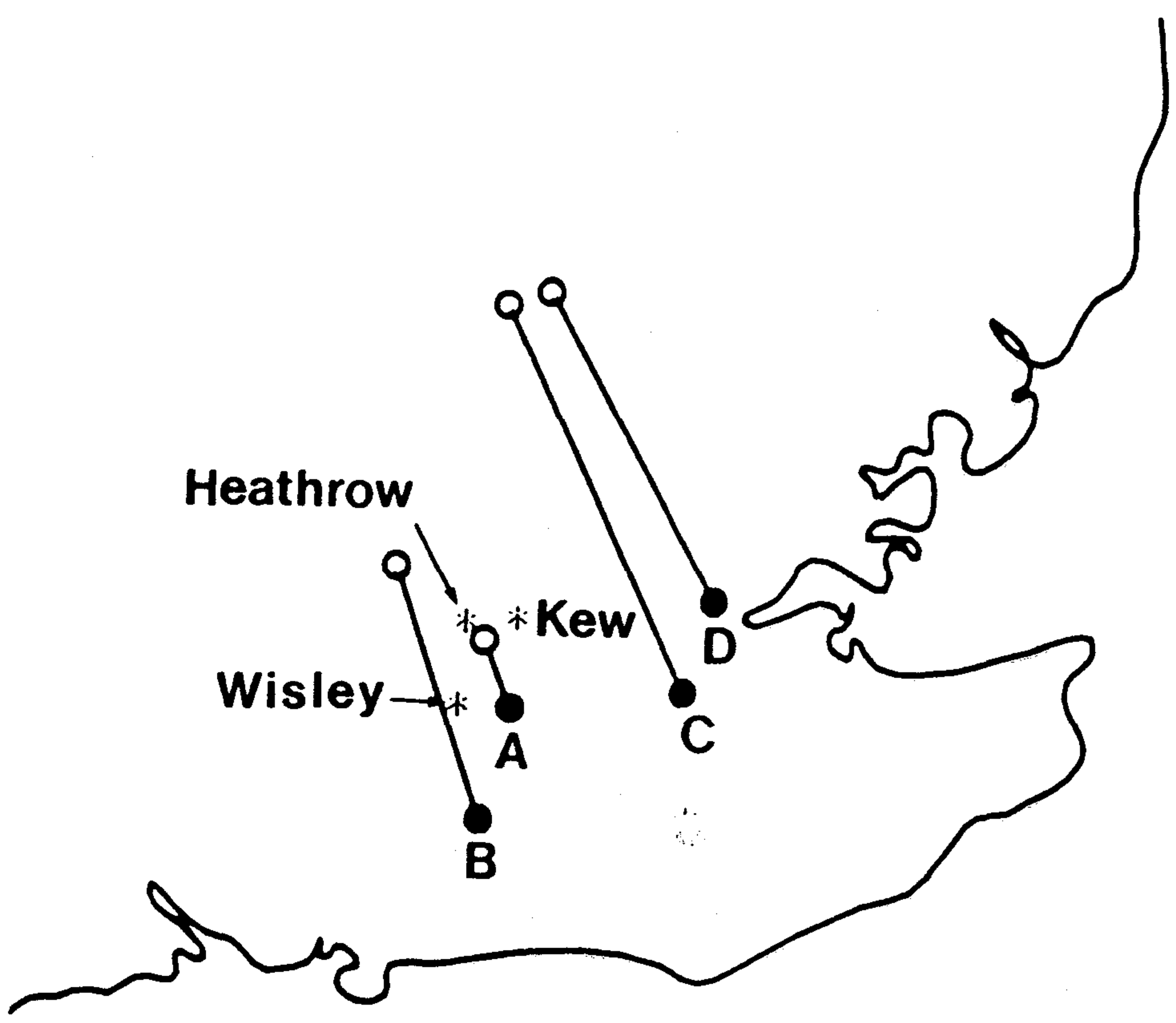

Figure 5. Estimates of the focus of the lupin aphid invasion obtained for each of the periods A, B, C and D, using the backtracking method (•) and the centroids of the original datasets (o). Lines indicate the differences between the estimates obtained from backtracking and the simple centroids of the datasets. Locations of Kew, Wisley and Heathrow Airport are shown for reference.

Rejoinder We are pleased to receive Professor Rathbun's comment on our paper and reply as follows. It seems to us that he has misunderstood the two major motivations for the paper, although we are pleased to note that the culmination of his comment is an acknowledgement of the need for collaboration between biologist and statistician. The multi-authorship of our paper reflects just that, with contributions from the biologists involved in the study, as well as from statisticians. The content, too, reflects what the biologists among us were seeking: not only an appreciation of what new insights spatial analysis could bring to our data, but also a recognition of important considerations to help improve similar future studies. This latter focus, on experimental design rather than analysis, is ignored in Professor Rathbun's comment. For the cereal aphid data, it enabled us to expose the problem of not recording the spatial information accurately and to discuss the danger of confounding treatment effects with differences in ambient environment; for the couch grass data, to warn against the introduction of bias from the non-measurement of variables on some plots and to promote the use of large plots for experimentation; for the lupin aphid data to disclose the dangers of unequal sampling effort in surveys. These traditional design considerations form an important part of the development of Statistics that must not be forgotten; they are embraced keenly by biologists. We remind readers that no increase in the sophistication of analysis may help to address a problem if the design is flawed to begin with.

Turning to the question how well our analyses have addressed the questions posed by the biologists involved, there is a further difference between us. We have used purely descriptive, empirical methods; Professor Rathbun favours mechanistic analyses, based on modelling. In our experience, the choice between these approaches depends on how much previous analysis has been performed and the degree of information known concerning the biological interactions. None of the three sets of data examined had been analysed spatially before; few details were known, in each case, concerning how dispersal mechanisms interacted with environmental variables, either measured or unmeasured. In short, these scientific projects were all at an early stage of development. Our judgement was that there was insufficient information for detailed modelling and we therefore deliberately took a heuristic approach. The initial analyses we have done have gone a long way towards answering the simple biological questions posed: do cereal aphids move further along rows than between rows; is there evidence that treatments alter the spatial heterogeneity of rhizomes as well as their weight; was Wisley or Heathrow as likely a point of introduction for the lupin aphid outbreak as Kew? Note that the couch grass data were from a designed field experiment; the natural choice would be to routinely study the response variables using the analysis of variance, a traditional and empirical technique, but one that is still of immense value to statisticians. We do not reject the modelling approach when it is appropriate. Indeed, we have often used mechanistic population models developed by biologists, some that provide far closer representations of the organism's ecology than those suggested by Professor Rathbun (e.g., Brain & Marshall, 1996; Harrington, Mann *et al*. 1994; Perry & Clark 1983). However, we did not consider that the stage reached by any of our previous studies (Mann unpublished; Marshall 1990; Bartlett 1993) was sufficient to benefit greatly from that form of analysis. This was an example of pragmatic collaboration between biologist and statistician colleagues that Professor Rathbun extols in his final sentence.

Some details in Professor Rathbun's comment require response. It was not our intention to provide a comprehensive review of methods for analysing spatial point patterns, of which we are aware. These can be found elsewhere, for example, in Cressie (1993) and in Stoyan, Kendall and Mecke (1995). Rather, we focused on methods applicable to our data: mobile organisms dispersing from a point source. It is true that the first-order indices that we have used to measure the displacement of the entire population (δ), its spread (φ), the total distance moved by individuals from the focus (Δ), the anisotropy of movement (γ), and the uniformity of distribution (ω), neglect inter-event distances. We do not agree that 'considerable' information is thereby lost, especially since this is a multiple index approach. It is untrue to state that the index I_p does not use inter-event distances, although its use is implicit rather than explicit. It is also nonsensical to suggest that our indices ignore the distribution of events about their focus; ω is designed specifically to provide a direct test of the uniformity of that distribution, and γ is equally concerned to measure this in a more specific context. It is true that the SADIE model is very unlikely

for any organism; complete regularity is merely a convenient baseline against which to measure observed arrangements in much the same way that complete spatial randomness, while seldom encountered in nature, is a convenient null hypothesis for the test based on Ripley's *K*-function. But it is as unreasonable to infer from this that 'any index based on (regularity) is meaningless' as it would be to claim this for the *K*-function based on randomness. Professor Rathbun is correct to point out that compressing spatio-temporal data over time can sacrifice opportunities to distinguish between alternative models by following the evolution of the spatial pattern over time; we have tried to address the issue of spatial association through time elsewhere (Korie *et al*., submitted). However, we believe he will agree with our judgement that, for the cereal aphid data, very little information was lost by our procedure.

Notwithstanding the defence of our approach set out above, we would be enthusiastic to see Professor Rathbun's proposed model fitted to any of our sets of data. If his approach leads to further biological insights, then we would see that as a very useful extension to our initial analyses, complementary to and not competing with them. Professor Rathbun appears not to have obtained copies of our datasets whilst writing his comment, despite instructions for doing so being included in the paper (in accordance with the editorial policy of this journal). However, we reiterate our personal invitation to everyone to have access to our data and we will particularly welcome Professor Rathbun's developments of the spatio-temporal aspects of the analyses. We therefore look forward to productive discussions with him in the near future.

References not given in the main paper

Brain & Marshall (1996) Estimating cultivation effects on weed seed dispersal using Fourier transforms. *Aspects of Applied Biology*, **46**, 173-179.

Cressie, N.A.C. (1993) Statistics for Spatial Data, revised edition. Wiley, New York.

Harrington, R.H., Mann, J.A. and twelve other authors (1994) Monitoring and forecasting for BYDV control - the way forward? *Aspects of Applied Biology*, **37**, 197-206.

Perry, J.N. & Clark, S.J. (1983) Use of population models to estimate the proportion of cyst-nematode eggs killed by fumigation. *Journal of Applied Ecology*, **20**, 857-864.

Stoyan, D., Kendall, W.S. and Mecke, J. (1995) Stochastic Geometry and its Applications, second edition. Wiley, New York.

Korie, S., Perry, J.N., Mugglestone, M.A., Clark, S.J., Thomas, C.F.G. & Mohamad Roff, M.N. Analyzing association between counts. *Journal of Agricultural, Biological & Environmental Statistics*. (submitted).