STATISTICS TO FIND SPATIAL AND TEMPORAL STRUCTURES IN POPULATIONS

D. Debouzie and J. Thioulouse
Laboratoire de Biométrie
Université Claude Bernard
69622 Villeurbanne. France

Spatial distribution is one of the most characteristic properties of insect populations ; in most cases it allows us to define them (Legay and Debouzie, 1985). Spatial distributions of insect populations are not fixed but dynamic. Thus variations in space and in time of the pattern of individuals occur not only at the scale of the species, but also at the population one. For example, differences between the dynamics of two populations of the butterfly *Euphydryas editha* may be as great as those between *E. editha* and *E. chalcedona* (Ehrlich et al., 1975).

It is important to find spatial and temporal structures in populations for applied and fundamental reasons. No field sampling can be efficient without understanding the underlying spatial distribution (Taylor, 1984). In pest control, we need to know how the insects are dispersed in the region under study ; we have to define the "hot spots" of the infestation so as to modulate the dose of pesticides, or at the orchard scale to take profit of variations in the infestation rate with host varieties ; at last we have to study adult flow from backyard orchards to commercial ones.

Spatial structures **also give** information on theoretical population biology (Legay and Debouzie, 1985). Within a population, groups of individuals may have their own dynamics and contribute differentially to the global population ; then the effective population size will be greatly reduced. The mechanisms that influence or regulate the population dynamics usually differ if, for a given density, individuals are dispersed at random or in clumps. Moreover key-factors implied in the dynamics of a local population, like that in an orchard, may have no influence at a regional scale ; climatic factors or competition for food illustrate such variations. So we need to separate the different levels of heterogeneity when searching the factors that regulate a population.

NATO ASI Series, Vol. G11
Pest Control: Operations and Systems Analysis
in Fruit Fly Management. Edited by M. Mangel et al.

In this paper we propose methods to collect and analyze basic data so as to find the spatial distribution of the individuals inside the universe sampled. Our proposals do not apply to the case where only the mean and the variance of the data are measured with confidence, as did Taylor (1984). We propose to replace random sampling by systematic one. In each sample the spatial positions of the points collected give a relevant information to the structures of the population ; so we must find adequate statistics which take into account such data and give a biologist readable picture of the structures within a sample or between samples when data are collected more than once.

Three main questions will be tackled :

1. What is a population of insects ?
2. How can we sample such a population ?
3. Which statistics can we use to analyze data obtained by systematic sampling ?

1. THE NOTION OF POPULATION

There is an intermediate level between the species and the individuals : the population. We do not consider it as a theoretical object like the mendelian population. We propose to study the populations such as they really live in the field and their characteristics must be found. A population can be defined as one or several ensembles of individuals of the same species that can live together at one time of their life history and then can interact on the mechanisms of their reproduction.

This definition does not imply that populations are fixed in space and in time ; a peculiar population may disappear at one time from a given place or many populations, well-defined in space at one time, may mix up altogether in other environmental conditions.

With insects, we postulate that populations also exist, that they are structured in space and in time and also genetically. So we must find their spatial limits, how these limits vary during the life history of the insects, what are the phenotypic and in some favourable cases the genotypic compositions of the population.

The population approach may have several practical applications, for instance, for fruit fly control. In an heterogeneous area, we must sample all the different types of habitats : large-scale orchards, small orchards and backyard gardens. But we must also sample the classical larval hosts and the alternate host plants which may be attacked by fruit flies under unfavourable conditions. The flow of adults between these different habitats must also be studied and **modelled** , specially for multivoltine fruit flies. Recently this approach has been successfully developed in the agronomic study of the cabbage stem flea beetle infesting rapeseed fields (Thioulouse, 1985).

In a similar way a fruit fly population cannot be correctly described by counting only one stage of develoment, like adults alone or larvae alone. Indeed demographic studies, such as Carey's one (1982), clearly show that adults represent only 4 % of the total population size. So all the stages of development must be studied and their dynamics quantified.

When population structures are known in a location, the sampling intensity can be reduced and pest control improved. Successful practical results have been obtained with rapeseed insects (Kühne, 1977) and cockchafer larvae (Chessel et al., 1984).

To find spatial and temporal structures we need adequate sampling designs and correlatively statistical methods fitted to the data obtained. The concept of population, the sampling design and the statistics used or available are three closely tied questions. An experimental design cannot be conceived if the universe to be studied is not correctly defined, if adequate statistics do not exist, but also if the existence of structures is denied or unknown.

2. THE SAMPLING DESIGN

Four classical sampling designs will be briefly discussed : random, sequential, stratified and systematic. I want to show the weakness or the advantage of each method when spatial structures are searched for.

Random sampling, but it is also true for sequential sampling, is based implicitely or explicitly on the lack of spatial structures ; the universe sampled is assumed to be homogeneous and in most examples the population from which samples are drawn, is not referred to. For our point of view that insect populations are structured in space, random sampling must be abandoned, unless previous experiments have shown that individuals disperse randomly in their habitat.

The major difficulty with random sampling comes from the statistical method used. In most papers, dispersion is measured by several indices computed from the average value of the number of individuals per unit of sampling and from the variance of this variable. For example, when the index of dispersion is greater than one, we classically conclude that the distribution is aggregative, that individuals are clumped at the scale of the elementary unit of sampling and biological hypotheses are then put forward.

However, this conclusion is only true when no variation in density occurs in the universe sampled. But we know that natural populations are rarely homogeneous ; several works on fruit flies (Fletcher, 1974 for the Queensland fruit fly ; Neuenschwander and Michelakis, 1979 for the olive fruit fly) lead to the same conclusion.

A distribution of frequency, which is the main statistical product of random sampling, may be generated by many patterns of distributions. For instance, the negative binomial distribution can be generated by more than fifteen different models. We must separate what Feller called in 1943 the true aggregation from the false one ; in true aggregation, the fact that a sample already contains one individual, increases the probability of choice of this sample by the individuals. False aggregation is obtained by numerous ways such as all the kinds of variation in density. When we adjust an observed distribution of frequency to a theoretical law, the true pattern of dispersion within the sample cannot be found. An illustration is given in Figure 1 which represents two distributions corresponding to the same mean and variance of X, the number of individuals per unit. The same value of the index of dispersion (variance/mean) masks two different patterns of dispersion within the area sampled.

Simple graphical representations immediately give a good picture of the true distribution. But they cannot be used alone, without statistics, because graphical tools are never neutral (Auda, 1983).

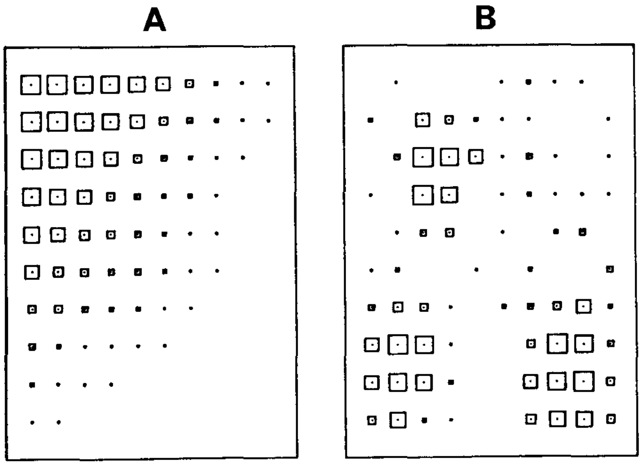


Figure 1 : Two theoretical distributions corresponding to the same mean and variance of individuals per unit of sampling. The index of dispersion equals 3.52. The size of each square is proportional to the number of individuals.

A : Individuals are distributed along a gradient.

B : Three clumps of mean size averaging about 9 units are visible ; elsewhere the distribution is random.

What is important for the population biologist is to find how the individuals disperse in the field, how they occupy their habitat. In 1985 we must go further than the ancient trilogy based on the three classical types of distribution : random, contagious, and regular. Statistical methods should be able to describe the true dispersion, that is to find the size of the aggregates or the presence of a gradient and of any variation in density.

Stratified sampling has no ability to describe spatial structures since on the contrary it assumes that the universe studied is already stratified. Then it is very efficient to estimate the population size and the accuracy of this size.

For many years systematic sampling is well-known to be able to detect spatial structures, specially in phyto-sociology. It offers several practical and theoretical advantages. Usually it is rather easy to be set up in the field ; the starting point of the sampling may be a difficulty but periodic distributions of individuals are rather rare in insect populations. Systematic sampling allows more statistical computations than random sampling. Several scales of heterogeneity can be found in an habitat. Accuracy of the population size estimation can be computed from the theory of regionalized variables (Matheron, 1965). **Finally**, the sampling intensity may be highly reduced when the structures of the population are known.

Till 1970 systematic sampling was difficult to use because of the lack of efficient statistical tests. But in the seventies progresses were made independently by several workers : Chessel, Cliff, Hill, Ludwig, and Ord mainly. We shall present the most powerful tests available, specially those of Chessel who gives a general theoretical framework for this problem.

3. STATISTICAL TESTS FOR DATA ORGANIZED IN A GRID

Data collected by systematic sampling are organized along a line when one dimension is sampled, or in a rectangular grid when two dimensions are explored. The case of grids will only be presented since transects are just a peculiar case.

Field distributions of individuals in their habitat are complex and several scales of heterogeneity are often mixed. Intuitively we can reasonably assume that one statistical test cannot detect these different scales ; a set of tests is needed. Chessel (1978) proposed non parametric statistics, the efficiency of which varies according to the hypotheses tested. A test must be preferentially used against only one precise alternative hypothesis ; a test can be very powerful against a hypothesis but weak or bad fitted to another.

The principle of the statistical analyses can be summarized as follows. The elementary units are summed up into blocks ; the size of the blocks varies from one unit to the whole grid. For each block, three measures of heterogeneity are proposed : global heterogeneity where variation is calculated between the blocks ; intra-block local heterogeneity, and heterogeneity at one scale calculated between two contiguous blocks.

For all the tests used, the value of the statistics is plotted against that of k, the size of the block. The shape of the different curves is used to detect the model of distribution ; reference curves have been obtained after analyzing more than a hundred of grids either in vegetation problems or in animal ones (Chessel, *op. cit.*).

We have selected the more efficient tests available for analyzing data that are counts, which are the most frequent measures in fruit fly studies, but other tests exist for binary data.

The systematic sampling of cockchafer larvae in the soil will be treated as example ; one sample of one sixth meter square was collected per hectare in a region of meadows in the North Jura, in France. The goal is to find how larvae disperse in this region ; is there some meadows more infested than others ? Is there a general trend in the infestation rate ?

Basic data are presented in Figure 2.

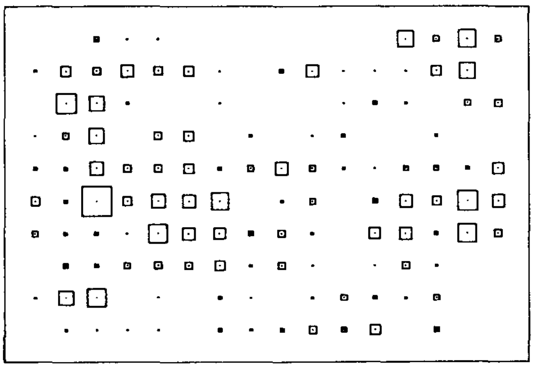
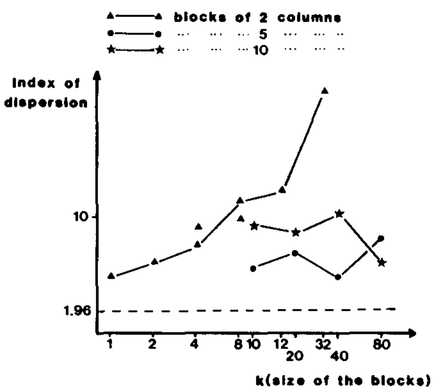


Figure 2 : Distribution of cockchafer larvae (stage L2) in the soil in Nancray (France, 1976). A sample unit 1/6 m2 large is collected per hectare ; about 160 hectares are sampled. The size of the largest square represents 24 larvae.

3.1 Global heterogeneity

The index of dispersion

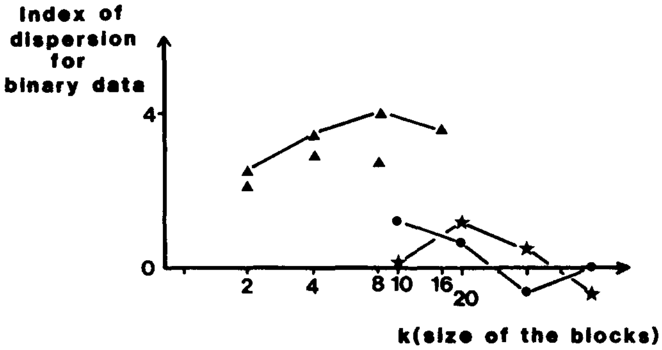
The different curves giving the variations of the index according to the size of the block k are rather classical (Figure 3). We observe an increase when blocks are made of two columns but with blocks of five or ten columns variations are cahotic. In all the cases the values of the index are highly significant which means that at least one scale of heterogeneity exists in the data.



<u>Figure 3</u> : Variations of the index of dispersion (variance/ mean) <u>against</u> the size of the blocks.
The initial grid has 16 lines and 10 columns (see Fig. 2). The abscissa scale is logarithmic.

<u>The index of dispersion for binary data</u>

To eliminate the effect of true aggregation, due for instance with insects to the laying of a mass of eggs, we propose to converse the raw data into binary ones. When a count is greater or equal than the median value, it is converted into 1, and into 0 if lower. The distribution of the ones in the grid allows us to calculate the index of dispersion (see Appendix).



<u>Figure 4</u> : Variations of the index of dispersion when data are <u>conversed</u> into binary ones.
For a complete legend, see Fig. 3.

The index has a maximum and significant value for blocks made of 4 lines and 2 columns (Figure 4). But, for blocks of 5 and 10 columns, the dispersion of the local aggregates (corresponding to the ones) is random.

The spatial auto-correlation matrix

The degree of correlation between values recorded at two neighbouring points can be tested by a statistic due to Geary in 1954 and later to Cliff and Ord in 1973. Chessel (1981) applies this general test to two contiguous blocks in a grid : if individuals disperse randomly over the grid, the values measured at two any points are not correlated. The principle of the statistics is to test the variability between two contiguous blocks against the total variability of the measures.

The test applies to quantitative or qualitative data and also to non-rectangular or incomplete grids ; moreover it can be used between two points separated by a distance greater than one.

The test applied to the cockchafer larvae data shows a maximal variability at two sizes of blocks : 8 and 32 (Table 1).

Columns		1	2	5	10
Lines	1	-	2.4	1.0	2.7
	2	2.1	3.0	0.5	1.5
	4	4.3	4.4	0.9	1.4
	8	3.3	3.4	0.0	0.2
	16	4.5	4.8	0.5	-

Table 1 : Values of the spatial auto-correlation test according to the size of the blocks.
Values higher than 1.96 are significant. The test detects two scales of heterogeneity : 4 lin. x 2 col. and 16 lin. x 2 col.

No structure can be detected when all the columns or half the columns are summed up ; so the different lines seem to be homogeneous.

The three tests used to study global heterogeneity do not give contradictory conclusions ; of course they are linked by the sole fact that they use the same data. Nevertheless they are more or less expressive and difficult to interpret according to the examples.

3.2 Heterogeneity at fixed scale

Specific tests were developed to compare the number of individuals in two contiguous blocks (Chessel and DeBelair, 1973). Two blocks are defined as contiguous if their union constitutes a block at a higher scale. So blocks can be compared horizontally on the same line of the grid, or vertically on the same column.

The principle of the test lies on the fact that the difference between two blocks is compared to that calculated under the hypothesis of a random distribution of all the individuals present in the two blocks. Under this hypothesis, the statistic follows a gaussian distribution (see Appendix).

Two major points emerge from Figure 5 which shows the variations of the index of true aggregation.

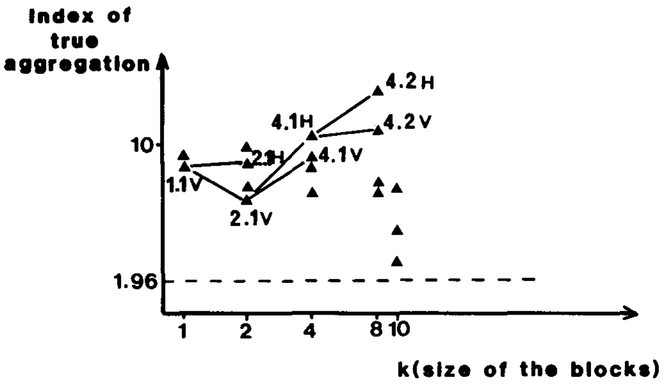


Figure 5 : Variations of the index of true aggregation. Blocks can be compared horizontally (H) or vertically (V). A block formed by 2 lines and 1 column is noted : 2.1 ; only the most interesting combinations of lines and columns are noted.

The counts of two elementary samples are not randomly distributed ; the very high value of the index, between 9 and 10, higher than the threshold which is 1.96, means that many larvae are found in an elementary sample while in the other contiguous sample they are few. This is a true aggregation which is usually observed with insects when many eggs are laid together and when just-born larvae do not migrate very far. In other studies, we observed that the intensity of the true aggregation decreases when cockchafer larvae become older, owing to density-dependent mortality and dispersion ; larvae in the third stage are only slightly contagiously dispersed.

The maximum value of the index corresponds to a block of 4 lines and 2 columns. This scale of heterogeneity has already been found by other statistics and thus is confirmed by the index of true aggregation.

3.3 Local heterogeneity

The local heterogeneity measures the within-block variation calculated between the units of a block. Two statistics are commonly used : the index of local dispersion and the index of the number of empty cases (see Appendix). These statistics are well-fitted to detect locally regular dispersion without any interaction with other heterogeneity structures. The index of the number of empty cases is powerful against the presence of empty units which have a great biological interest, specially in pest control.

No application of the tests to the cockchafer larvae will be given since no peculiar scale of variation was found inside the blocks.

3.4 Conclusions

Two scales of heterogeneity are found : the elementary sample where true aggregation is observed, and a block of eight elementary units, representing an area of about 8 hectares. It means that some meadows, or groups of meadows, are heavily infested, about 45 millions of second stage larvae per km square. At the opposite, other meadows are slightly infested with a rate near 15 millions per km square. The initial map (Fig. 2) could be drawn again, by representing the number of larvae within each block of 8 hectares.

It is worth noting that the two intermixing scales of heterogeneity imply different biological factors : egg-laying and density-dependent mortality and dispersion for the first, ecological characteristics of the meadows and memory of the adult females for the second (see for instance Chessel et al., 1984).

4. STATISTICS FOR DATA ORGANIZED IN SPACE AND IN TIME

Usually population biologists do not collect grid-organized data only once ; they must take into account the temporal structures which are always observed with temperate insects. Then two major questions are asked to the statistician :

- describe the spatial structures at each date,
- test the stability of these structures during the period of survey.

The first question can be answered by using the statistics described in the previous section. But for the second, till now we do not know any test able to study intra-grid dispersion over several times. The approach of Taylor and Iwao, who modelled the relation between the variance and the mean, either for different locations of samples or for different dates, is not supported by the same philosophy. These authors do not aim to describing the distribution within the basic samples ; they compare samples on the basis on the relations between variance and mean. In ecological words, their approach refers more to the species, while ours concerns more the population.

Always with this population view, we propose to describe temporal variations in spatial structures by multivariate analyses. We choose an example from the agronomic field to present the statistical method.

The data correspond to captures of cabbage stem flea beetle adults in 33 traps regularly spaced in a 1.24 hectare rapeseed field ; the traps are examined 64 times from the beginning of the infestations in September 1983 to the last emergence of the new generation in June and July 1984.

Data can be presented as a collection of 64 grids, each of the same size, or as a collection of 33 curves, each curve corresponding to a sampling point (Figure 6).

We propose to use the factorial analysis of correspondences (F.A.C.) to treat the data. The latter are reorganized in a table with two dimensions, instead of three as in the raw data (two for

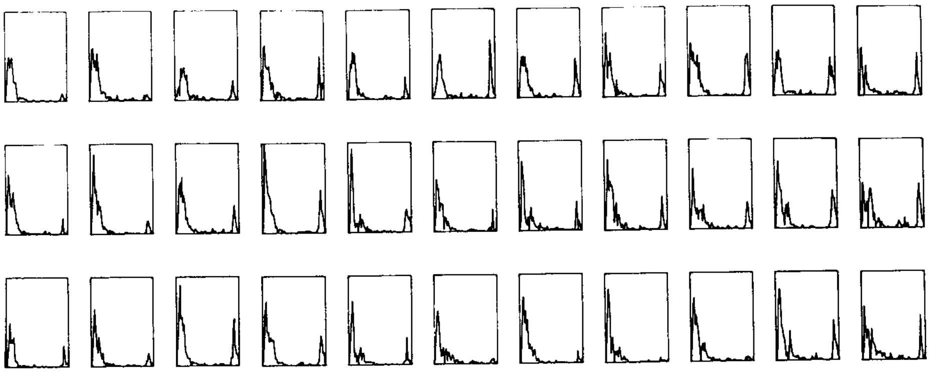


Figure 6 : Curves giving the numbers of cabbage stem flea beetle adults captured in water-traps regularly spaced in a rape field.

Each curve corresponds to a sampling point ; traps are collected 64 times since the invasion of the field in September 1983 to the emergence of the new generation in June and July 1984. For the ordinate axis, the maximum value corresponds to 125 adults captured in a trap (for the curve at the second line and the fourth column).

space, one for time). The lines of the table correspond to the sampling points and the columns to the dates (or inversely since in the F.A.C. model chosen here, lines and columns play a symmetrical role).

The goals are twofold : i) compare the temporal curves of catches per location, that is compare the distribution of each line of the table ; ii) compare the spatial distribution of catches per date, which is the distribution of each column.

The factorial analysis of correspondences is well-aimed at responding to these goals. It gives factors of decreasing magnitude according to the eigen values. Coordinates of the lines and of the columns can be computed for each factor. In the inertia model of the F.A.C. which is used here, contributions of lines and columns to the factors and inversely descriptions of a line or a column by a factor, help the interpretation.

Since in F.A.C. the factor-line maximizes the variance of the conditionnal means of the columns and reciprocally, we can represent graphically the factors-lines and the factors-columns in space or in time when data are organized in space or in time, as it is the case in our problem.

These graphical representations have been done for the first four factors ; a synthesis of the spatial partition in three homogeneous zones is given in Figure 7 (part A). These zones, called A, B, and C, are defined by the sign of the factors calculated for each sampling point and by the contribution of these points to the definition of the factors. The catch curves corresponding to the three zones are drawn in Figure 7 (part B).

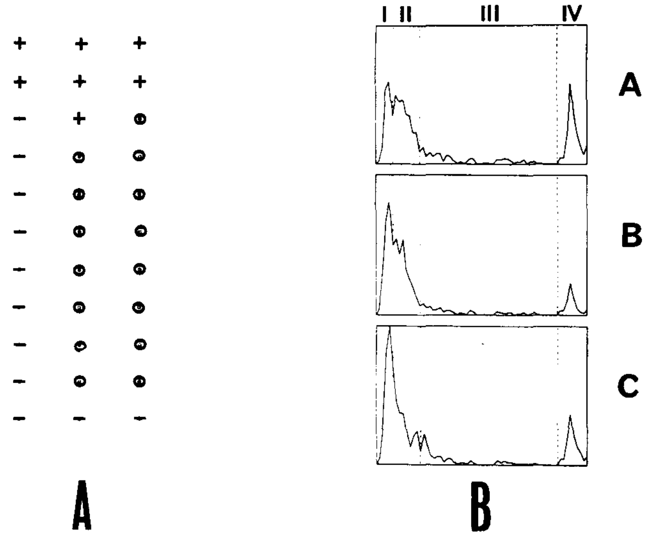


Figure 7 : Part A : Spatial partition of the rape field.
Part B : Curves of captures of adult insects relative to the three zones defined within the field (part A).
Zone A (symbol +) : 7 points. Factors 1 and 2 positive, factor 4 negative.
" B (" -) : 11 " . Factor 1 negative.
" C (" o) : 15 " . Factors 1 and 4 positive.
Period I : 6 dates from 12/09 to 29/09/83.
" II: 8 " " 30/09 to 27/10.
" III: 41 " " 28/10 to 04/06/84.
" IV: 9 " " 05/06 to 11/07/84.
In the ordinate axis, the maximum value represents 93 adults captured per trap.

In a symmetrical way, dates may be partionned in four periods ; of course the spatial partition is linked to the temporal one owing to the reciprocal averaging of the factors-lines and the factors-columns. It is worth noting that the spatial partition and the temporal one have biological and ecological interpretations, since zones are formed by contiguous points, located at the North of the field (zone B) or at the East (zone C), and since periods correspond to different phases in the life history of the insect (main invasion, late invasion and autumn survival, winter survival, and emergence).

At last, data can be reorganized by summing up all the captures for a given period and a given zone (Table 2).

Zone		A	B	C
Period	I	152	218	265
	II	249	232	194
	III	101	67	97
	IV	167	61	112

Table 2 : Mean number of adult insects captured per trap for a given period and zone.
Periods and zones are defined in Figure 7. The numbers correspond to the total captures during a period but periods have variable lengths.

We have shown that spatial and temporal structures may exist within a rape field : in the example, the emergence rate may be three times greater in a zone than in another. The factorial analysis of correspondences describes such structures.

5. CONCLUSIONS

Spatial and temporal structures exist in insect populations even at a fine scale. The existence of local heterogeneities is now well-documented ; for example one fruit tree may be heavily attacked by insects and not its neighbour (Debouzie *et al.*, 1985) ; rape fields may be infested by insects at very variable rates (Thioulouse *et al.*, 1984). These fine structures, which often vary during the life history of the insect, can only be found because we abandon random sampling and adopt systematic survey ; it is also because a set of statistical methods specially fitted to this sampling design is now available. The population concept, the sampling design, and the statistics used remain three closely tied subjects.

Multivariate analyses offer much more possibilities to the population biologist. For example, using another model of the F.A.C., the factorial analysis of multiple correspondences, the interactions between an insect crop insect and its vegetal host can be studied. The canonic model of the F.A.C. contributes to defining the width of the niches.

One last interesting advantage of the systematic sampling is that the precision of the population size estimate can be obtained. Of course the question of the estimation of the population size remains a difficult one but we must go further in some favourable examples ; in some applied works a confidence interval for this estimation is necessary. The theory of the regionalized variables applied to regular spaced data gives such an interval as a function of the intensity and the nature of the spatial structures found within the grid (see an example for the unidimensionnal case in Thioulouse *et al.*, 1985).

6. LITERATURE CITED

AUDA Y., 1983. Rôle des méthodes graphiques en analyse des données : application au dépouillement des enquêtes écologiques. Thèse Doct. 3° cycle, Lyon.
CAREY J.R., 1982. Demography and population dynamics of the Mediterranean fruit fly. Ecol. Model., 16 : 125-150.
CHESSEL D., 1978. Description non paramétrique de la dispersion spatiale des individus d'une espèce. In : Biométrie et Ecologie. LEGAY J.M. and TOMASSONE R. (Eds.). Soc. Fr. Biom., Jouy (France), pp. 45-135.
CHESSEL D., 1981. The spatial autocorrelation matrix. Vegetatio, 46 : 177-180.
CHESSEL D. and DeBELAIR G., 1973. Mesure de la contagion vraie en échantillonnage par carrés dans l'analyse des populations végétales. C. R. Acad. Sc., sér. D, Paris, 277 : 1483-1486.
CHESSEL D., DEBOUZIE D., ROBERT P. and BLAISINGER P., 1984. L'échantillonnage des larves du hanneton commun *Melolontha melolontha* L. Acta Oecol., Oecol. Appl., 5 : 173-189.
CLIFF A.D. and ORD J.K., 1973. Spatial autocorrelation. Pion, London.
DEBOUZIE D., LEBRETON J.D., ALLAINE D. and PALLEN C., 1985. Contribution à la notion de groupes et de populations. Exemples de populations d'oiseaux et d'insectes. Rapport A.T.P. C.N.R.S., Biol. des Popul., 73 p.
EHRLICH P.R., WHITE R.R., SINGER M.C., Mc KECHNIE S.W. and GILBERT L.E., 1975. Checkerspot butterflies : a historical perspective. Science, 188 : 221-228.
FELLER W., 1943. On a general class of contagious distributions. Ann. Math. Stat., 14 : 389-400.
FLETCHER B.S., 1974. The ecology of a natural population of the Queensland fruit fly, *Dacus tryoni*. VI. Seasonal changes in fruit fly numbers in the areas surrounding the orchard. Austral. J.

Zool., 22 : 353-363.

GEARY R.C., 1954. The contiguity ratio and statistical mapping. Incorp. Statist., 5 : 115-145.

KUHNE W., 1977. Untersuchungen zur Befallsverteilung der Rüsselkäfer Ceutorrhynchus napi Gyll., Ceutorrhynchus quadridens Panz. und Ceutorrynchus assimilis Payk. innerhalb grossflächiger Winterrapsbestände. Arch. Phytopatho. Pflanzenschutz, 13 : 109-115.

LEGAY J.M. and DEBOUZIE D., 1985. Introduction à une biologie des populations. Masson, Paris.

MATHERON G., 1965. Les variables régionalisées et leur estimation : une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson, Paris.

NEUENSCHWANDER P. and MICHELAKIS S., 1979. Mc Phail trap captures of Dacus oleae (Gmel.) (Diptera, Tephritidae) in comparison to the fly density and population composition as assessed by sondage technique in Crete, Greece. Bull. Soc. entomol. Suisse, 52 : 343-357.

TAYLOR L.R., 1984. Assessing and interpreting the spatial distributions of insect populations. Ann. Rev. Entomol., 29 : 321-357.

THIOULOUSE J., 1985. Structures spatio-temporelles en biologie des populations d'insectes. Application à l'étude de l'altise du colza (Psylliodes chrysocephala L.) : résultats méthodologiques et biologiques. Thèse Doct. 3° cycle, Lyon.

THIOULOUSE J., DEBOUZIE D. and BALLANGER Y., 1984. Structures spatiales et temporelles des populations d'un ravageur du colza (Psylliodes chrysocephala L., (Col., Chrysomelidae)) dans plusieurs parcelles de culture. Acta Oecol., Oecol. Appl., 5 : 335-353.

THIOULOUSE J., HOUILLER F. and ONILLON J.C., 1985. Variables régionalisées et dénombrement d'insectes : cas unidimensionnel. C. R. Acad. Sc., sér. III, Paris, 301 : 423-428.

APPENDIX

1. The index of dispersion for binary data

We note N = the number of elementary units,
k = the size of a block,
B = the number of blocks (= N/k),
M = the number of values greater (or equal) than the median, that is the number of ones (see text).

Y_i represent the number of ones in the block i.

We calculate :

$$T = \sum_i Y_i^2$$

$$\varepsilon = (T - E(T)) / \sqrt{var(T)}$$

with

$$E(T) = M (1 + (k-1)(M-1)/(N-1))$$

$$var(T) = 2M(M-1)(k-1)(N-k)(N-M)(N-M-1)/(N-1)^2/(N-2)/(N-3)$$

Under the null hypothesis, ε follows a gaussian distribution, slightly skewed at right, since the null hypothesis is rejected at the 5 % level when ε >3.

2. The spatial auto-correlation matrix

Let X = the number of individuals per unit,
$\bar{x}$ = the mean value of X,
σ^2 = the estimated variance of X.

V is a symmetrical matrix defined by :

$V_{ij} = V_{ji} = 1$ if i and j are two neighbouring points,
$V_{ij} = V_{ji} = 0$ otherwise.

Note : $L_i = \sum_j V_{ij}$; $A = \sum_i L_i$; $C = \sum_i L_i(L_i - 1)$.

The total variability of the measures is : $H_T = 2\sigma^2$, and the variability between two neighbouring points is :

$$H_V = (\sum_{i,j} V_{ij}(x_i - x_j)^2)/A.$$

We calculate $Z = H_V / H_T$
and $\varepsilon = (E(Z) - Z) / \sqrt{var(Z)}$

with
$$E(Z) = 1$$
$$var(Z) = (U_1+U_2+U_3)/(A^2.N(N-2)(N-3))$$

and
$$U_1 = ((N^2-3)-(N-1)^2.U_4).A^2$$
$$U_2 = 2A(N-1)(N^2-3N+3-(N-1)U_4)$$
$$U_3 = (N-1)(D+A)((N^2-N+2).U_4 -(N^2+3N-6))$$
$$U_4 = N.\sum_i(x_i-\bar{x})^4/(\sum_i(x_i-\bar{x})^2)^2$$

Under the null hypothesis, ε follows a gaussian distribution.

3. The index of true aggregation

We note :
$$D = |L - R|$$
$$P = L + R$$

where L and R are the numbers of individuals in two contiguous blocks (left and right).

We calculate : $\varepsilon = \sum_{\text{couples}} (D - E(D))/ \sqrt{var(D)} / \sqrt{n}$

with : n = number of couples of blocks where $P > 1$
$E(D;P=2) = 1$

Let P be an even number, so :

$$E(D;P) = E(D;P-1)$$
$$E(D;P+1) = E(D;P) + (P!)/(((P/2)!)^2.2^P)$$

For all values of P :
$$var(D;P) = P - (E(D;P))^2$$

Under the null hypothesis, ε follows a gaussian distribution as soon as $n \geqslant 10$.

4. The index of local dispersion

In the block i made of k elementary units, we note s_{ij} the number of individuals in the unit j.

We calculate : $LD_i = \sum_{j=1}^{k} s_{ij}^2$; $S_i = \sum_j s_{ij}$

and $\varepsilon = (1/\sqrt{B})(\sum_{i=1}^{B} (LD_i - E(LD_i))/ \sqrt{var(LD_i)}$

with

$$E(LD_i) = S_i(S_i + k - 1)/k$$

$$var(LD_i) = 2(S_i - 1)S_i(k - 1)/k^2.$$

Under the null hypothesis, ε follows a gaussian distribution as soon as $B \geqslant 10$.

5. The index of the number of empty units

We note ZO_i the number of empty units in the block i ; S_i is defined as in §4.

We calculate : $\varepsilon = (1/\sqrt{B}) (\sum_{i=1}^{B} (ZO_i - E(ZO_i))/ \sqrt{var(ZO_i)})$

with

$$E(ZO_i) = k((1 - 1/k)^{S_i})$$

$$var(ZO_i) = k(k-1)((1-2/k)^{S_i}) + E(ZO_i) - (E(ZO_i))^2$$

Under the null hypothesis, ε follows a gaussian distribution as soon as $B \geqslant 10$.

ABSTRACT

Spatial structures are one of the most characteristic properties defining insect populations ; they are not fixed but vary during the life history of the insect.

Fitted sampling designs and adequate statistical tests are needed to find spatial and temporal structures. We propose to use systematic sampling and to abandon random designs.

Non parametric statistics (Chessel, 1978) allow us to find how the individuals disperse within the habitat studied. The tests are not based on the mean and the variance of the number of individuals per elementary unit, but on the spatial arrangement of the individuals. Several tests, instead of one index as in the classical studies, are needed to find several scales of heterogeneity. For all the tests, the elementary units are summed up

into blocks of growing size and we study the variations of the statistics according to the size of the blocks. We propose statistics adapted to detect three kinds of variation : global (between blocks of units), local (between units within a block), and between two contiguous blocks. An example is given to illustrate the tests.

When grid organized data are collected many times, we propose to use the factorial analysis of correspondences. This method is well aimed to describing temporal variations in spatial structures inside the grid, as seen in an example with insects infesting rape fields.