

STATISTICAL ANALYSIS OF OCCUPANCY RATES FOR  
OVERDISPERSED POPULATIONS BY REDISTRIBUTION  
PROCEDURES: APPLICATION TO THE EUROPEAN  
CORN BORER EGG MASSES DISTRIBUTION

J. Vaillant and N. HAWLITZKY\*

Laboratory of Biometry INRA-Versailles; 78000 Versailles, France

\*Laboratory of Zoology, INRA-Versailles; 78000 Versailles, France

Introduction

Many investigations in population ecology lead to data consisting of counts of the numbers of individuals in spatial or temporal sampling units (Lloyd, 1967; Iwao and Kuno, 1971; Taylor, 1984). Various statistical methods are available in accordance with the purpose of the study (Bliss, 1971; Iwao, 1968; Bissel, 1972a, 1972b; Pielou, 1977; Southwood, 1978). On the other hand, there exist situations for which the observed counts correspond to categories in which the population individuals have been classified. It is the case, for example, of the within-plant distribution of insects when we consider the number of individuals per leaf stratum. The problem often put forward is the comparison of the different category occupancy rates. Generally, the well known chi-squared test is applied (see for example Plackett, 1974). Unfortunately, this test is based on the assumption that individuals are distributed independently of each other whereas practitioners may obviously know it is not the case of the population of interest. Thus, for populations with aggregative behavior, chi-squared tests are no longer valid (Denis and Vaillant, 1985). Let us illustrate this by means of the following example.

Consider the simple case of two insect traps. We may want to test that the trap occupancy rates  $p_1$  and  $p_2$  are equal i.e.  $p_1 = p_2 = 1/2$ . Let  $N$  be the total number of insects in both traps and  $N_1$  the number of insects in trap 1. The chi-squared test statistic is:

$$\chi^2 = \frac{(N_1 - Np_1)^2}{Np_1} + \frac{((N - N_1) - N(1 - p_1))^2}{N(1 - p_1)} = 4 \frac{(N_1 - N/2)^2}{N}$$

Now, imagine the unrealistic but meaningful situation for which the insects are distributed by group of size  $k$ , the groups being distributed independently of each other. Therefore, in the above expression of  $\chi^2$ , we have to replace respectively  $N$  and  $N_1$  by  $N/k$  and  $N_1/k$  to obtain the theoretically correct chi-squared test statistic, say  $\chi_*^2$ .

Hence, we obtain easily:

$$\chi_*^2 = \frac{X^2}{k}$$

The nominal level of the test applied is

$$\alpha = P(\chi^2 > t_\alpha) \geq \alpha' = P(\chi^2 > k \times t_\alpha)$$

where  $t_\alpha$  is the value exceeded with probability  $\alpha$  by a chi-squared distributed variate with one degree of freedom.  $\alpha'$  is the exact asymptotic level of this test, i.e. the probability (when  $N$  is very large) of incorrectly rejecting the hypothesis  $p_1 = p_2 = \frac{1}{2}$ . We have  $\alpha = \alpha'$  if and only if  $k = 1$ . In the case  $k > 1$ , the use of the chi-squared test based on  $\chi^2$  will often lead to declare the occupancy rates to be different whereas they are not.

The above example clearly shows that applying test procedures based on the assumption of no deviation from randomness for analysing categorical data from ecological population may lead to erroneous conclusions if the individuals are distributed by groups rather than lonely. Consequently, in the following sections, our aim will be to present valid methods of testing hypotheses about occupancy rates in the case of individuals not necessarily distributed independently of each other. The so called Monte Carlo tests are defined in next section. They were introduced by Barnard (1963) in a quite general context. Besag and Diggle (1977) first applied them to spatial pattern statistical analyses. More recently, Vaillant and Badenhauer (1989) used them for studying transect and grid of counts from ecological populations. The main feature of the Monte Carlo tests is that they do not involve complicated distributional theory and the investigator is free to use a variety of informative statistics of his own choice. Further investigations on the properties and efficiency of Monte Carlo tests may be seen in Jöckel (1986).

#### SIMPLE MONTE CARLO TESTS

Suppose we have to test a simple null hypothesis  $H_0$  by means of a set of data from which the value  $u_1$  of a statistic  $u$  is calculated. Then, by simulating  $s-1$  times the hypothesis  $H_0$ , we obtain a random sample  $\{u_2, \dots, u_s\}$  from the null distribution of  $u$ . Let  $u_{(1)} < u_{(2)} < \dots < u_{(s)}$  be the corresponding order-statistics. Under  $H_0$ , we have

$$P(u_1 = u_{(j)}) = \frac{1}{s} \quad (j = 1, \dots, s).$$

Therefore, the rank of  $u_1$ , say  $r(u_1)$ , may be used to construct exact test of  $H_0$  (see, for example, Diggle, 1983, § 1.6). An exact one-sided test will be based on the following property:

$$P(r(u_1) \leq k) = \frac{k}{s}.$$

The generalization to two-sided tests is obvious.

*Example.* If we have a set of count data  $\{x_1, \dots, x_n\}$  consisting of numbers of individuals in sampling units, we may want to test the null hypothesis of pure random distribution of the  $x_+ = x_1 + \dots + x_n$  individuals in the  $n$  sampling units. This may be done by simulating 199 times the redistribution of  $x_+$  individuals independently of each other in  $n$  units and, then, by calculating a relevant test criterion  $u$  for the set of observed count data, say  $u_1$ , and also for each of the 199 simulated sets of count data, say  $u_2, \dots, u_{200}$ .

$u$  may be the index of dispersion (see for example Taylor, 1984):

$$u = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\bar{x}} \quad \text{with} \quad \bar{x} = \frac{x_+}{n}$$

As pointed out by Vaillant and Badenhauer (1989),  $\bar{x}$  is invariant by redistribution and a simpler criterion, which is furthermore equivalent to the index of dispersion when applying Monte Carlo test, is

$$v = \sum_{i=1}^n x_i^2.$$

The hypothesis  $H_0$  is rejected against underdispersion if  $v_1$  is among the five first indices or against overdispersion if  $v_1$  is among the five last indices. So, we apply a two-sided test of  $H_0$  at level  $\alpha = 10/200 = 0.05$ .

In a similar way, if the  $x_i$ ,  $i = 1, \dots, n$  seen above correspond to count data from  $n$  categories, the same test procedure is valid to test the homogeneity of the occupancy rates (i.e.  $p_1 = p_2 = \dots = p_n$ ) under the assumption of individuals distributed independently of each other.

If ties occur for the order statistics (e.g. when the null distribution of the test statistic  $u$  is discrete), it is suggested to choose the most conservative rank for  $u_1$  (Jöckel, 1986). On the other hand, Hope (1968), Marriot (1979) and Jöckel (1986) have shown that the power of Monte Carlo tests was remarkable high for modest computational effort ( $\alpha \geq 5$ ). Therefore, such a test procedure is particularly useful in many situations for which standard statistical tests are not appropriate.

In the following sections, we will consider a classification with  $r$  categories and assume that an observed individual belongs to a given category  $i$  with probability  $p_i$ . Then, we will deal with testing whether the observed frequencies  $N_1, \dots, N_r$  of individuals in categories 1,  $\dots$ ,  $r$  are in agreement with the probabilities of occupancy  $p_1, \dots, p_r$ .

#### PURE RANDOM DISTRIBUTION CASE

When the individuals are randomly distributed, the chi-squared test is valid for testing hypotheses about  $p_1, \dots, p_r$  if the total number  $N$  of individuals is large

enough. In this case, the nominal level of this test is a very good approximation of the exact level, as seen in the introduction of this paper. Otherwise ( $N$  small), we may obtain spurious significant results. Then, the Monte Carlo test can be applied by numerical redistribution of the  $N$  individuals, independently of each other and in accordance with the probabilities of occupancy  $p_1, \dots, p_r$  and, lastly, by using the simplified test criterion

$$C = \frac{\sum_{i=1}^r N_i^2}{p_i}. \quad (1)$$

This criterion  $C$ , when applying Monte-Carlo tests, is equivalent to the classical chi-squared test statistic

$$\frac{\sum_{i=1}^r (N_i - Np_i)^2}{Np_i} = \frac{\sum_{i=1}^r N_i^2}{Np_i} - N.$$

#### RANDOM DISTRIBUTION OF GROUPS

When individuals are distributed by groups, we can distinguish two cases: firstly, the group sizes are observed. A simple example is the counting of insect eggs when they are distributed by clumps. Secondly, the group sizes are not observed. It is the case when individuals from different groups are mixed inside the sampling units and the belonging to a group is not observable.

##### Known Group Sizes

Let us note  $s_1, \dots, s_T$  the sizes of the  $T$  groups observed. The total number of individuals is  $N = s_1 + \dots + s_T$ .

The principle, here, will be to redistribute the  $N$  individuals by groups of sizes  $s_1, \dots, s_T$ . In other words, we will redistribute the  $T$  groups independently of each other and according to the probabilities of occupancy  $p_1, \dots, p_r$ . The test criterion given by expression (1) will be used for the Monte Carlo test procedures.

##### Unknown Group Sizes

When the  $N$  individuals are known to be distributed by groups but the groups are not identifiable inside statistical units, the redistribution of groups as seen above can not be implemented since the group sizes are not available. Nevertheless, we may have enough information about the probability distribution  $P_{size}$  of the group sizes in given conditions to proceed as follows: draw independently group sizes according to  $P_{size}$  until the cumulated size exceeds  $N$ . Drop the last group in order that the sum of the group sizes be at most equal to  $N$ . Then, distribute the groups independently of each other in the  $r$  categories and apply the Monte Carlo test as in the case of known group sizes.

Thus, when the group sizes are not observed, the simulation based tests are ap-

appropriate only if the probability distribution  $P_{size}$  is well known.

#### DISTRIBUTION WITH PATTERNS AT SEVERAL SCALES

It exists different kinds of departures from a pure random distribution of individuals. We have seen above the one consisting in a random distribution of groups rather than single individuals but many other alternatives have been considered by different authors (Iwao and Kuno, 1971; Vaillant, 1989). For biological populations, the main factors playing an important part in the spatial distribution pattern are the species' mode of life (reproduction, dispersal and interaction among individuals) and the heterogeneity in the environmental conditions. Unfortunately, the study of occupancy rates for population distributions generated by complex mechanisms is difficult with either a classical method or a redistribution-based one. The former demands independence assumptions which are no longer valid and the latter is too much time consuming and, sometimes, theoretically uncertain.

Meanwhile, if the hypothesis to be tested concerns a scale of pattern for which random distribution of groups or single individuals can be accepted, at least as a good approximation of reality, the simulation procedures seen in the previous sections are still of interest, provided that constraints concerning the patterns are taken into account. This will be illustrated in next section by analysing count data of the European Corn Borer (*Ostrinia nubilalis*, Hbn. Lep. Pyralidae) egg masses.

#### ANALYSES OF THE EUROPEAN CORN BORER EGG MASS DATA

In summer 1980, observations were carried out inside a corn field of Paris area (Beauce) during the whole oviposition activity period of the European Corn Borer (ECB). The purpose was to study features of the ECB egg masses spatio-temporal distribution (Hawltitzky, 1986). A quasi-systematic sampling inside a 40 m × 40 m study plot was applied, providing 24 sampling clusters, each of them consisting of 21 adjacent aligned plants. The 504 plants were examined twice a week in order to detect arrivals of new egg masses. Several variables concerning the plant phenology, the egg mass state and the egg mass within-plant location were noted.

Table 1 and Fig. 1 show us some descriptive features of the within-plot distribution of the ECB egg masses at two scales (single plant and cluster). The analysis was performed for the weeks for which the infestation rate was strong enough (6 out of 8 weeks). Overdispersion was detected at both scales which means that the egg masses were not distributed independently of each other. Furthermore, for a given week, the egg masses spatial distribution conditionally to the previous week distribution was studied. Positive time correlation was found at the plant scale for the pairs of weeks (6, 7) and (7, 8) and at the cluster scale for the pairs of weeks (3, 4) and (6, 7) (see table 2). Consequently, if we want to compare the values of the leaf positions occupancy

Table 1a. Descriptive statistics of the ECB egg masses within-plot distribution from a sample of 24 clusters of 21 plants.

week	$m$	$\sigma^2$	$\frac{\sigma^2}{m}$	$m$	$\sigma^2$	$\frac{\sigma^2}{m}$	$P$
	single plant			cluster of size 21			
3	0.024	0.035	1.479*	0.50	1.22	2.43*	92
4	0.075	0.078	1.032	1.58	1.99	1.26	94
5	0.323	0.382	1.182*	6.79	15.65	2.30*	91
6	0.302	0.326	1.082	6.33	17.45	2.76*	88
7	0.103	0.113	1.091	2.17	4.67	2.15*	91
8	0.042	0.040	0.960	0.88	1.33	1.52	93

$m$  and  $\sigma^2$  = mean and variance for the number of egg masses per statistical unit.  $P$  is the percentage of intra-cluster variance for the number of egg masses per plant. \* corresponds to a significant overdispersion at level 0.05 (Index of dispersion-based classical test).

Table 1b. Frequencies distribution of ECB eggs on the 504 maize plants observed in summer 1980.

Week	Number of plants per infestation class							
	$N_0$	$N_1$	$N_2$	$N_3$	$N_4$	$N_5$	$N_6$	$N_7$
1	503	1	—	—	—	—	—	—
2	503	1	—	—	—	—	—	—
3	494	9	—	1	—	—	—	—
4	468	34	2	—	—	—	—	—
5	376	98	26	3	1	—	—	—
6	377	105	20	1	1	—	—	—
7	456	45	2	1	—	—	—	—
8	483	21	—	—	—	—	—	—
Whole period	241	140	86	25	9	2	—	1

$N_i$  = number of plants having  $i$  egg masses.

rates with some pre-established values, we can not use the classical chi-squared test procedure since it assumes that plant infestations by ECB egg masses are mutually independent events.

Information about the within-plant distribution of the egg masses (respectively the eggs) is given by table 3 (respectively table 4a). The relation between egg mass size and infested leaf position is described in table 4b. Precautions must be taken when interpreting these tables because, for a given date, all the corn plants have not the same number of available leaves. Nevertheless, it seems clear that most egg masses or eggs are laid on leaves 4 to 10. This confirms an intuition of practitioners who often consider that observing all leaves of a plant may be useless when performing counts on sampled corn plants in order to estimate egg masses density (Despins and Roberts, 1986). Examination for egg masses being confined to certain leaves, much sampling time could be saved. The question then is which leaf stratum has to be observed and which correcting factor must be applied to the estimated egg masses

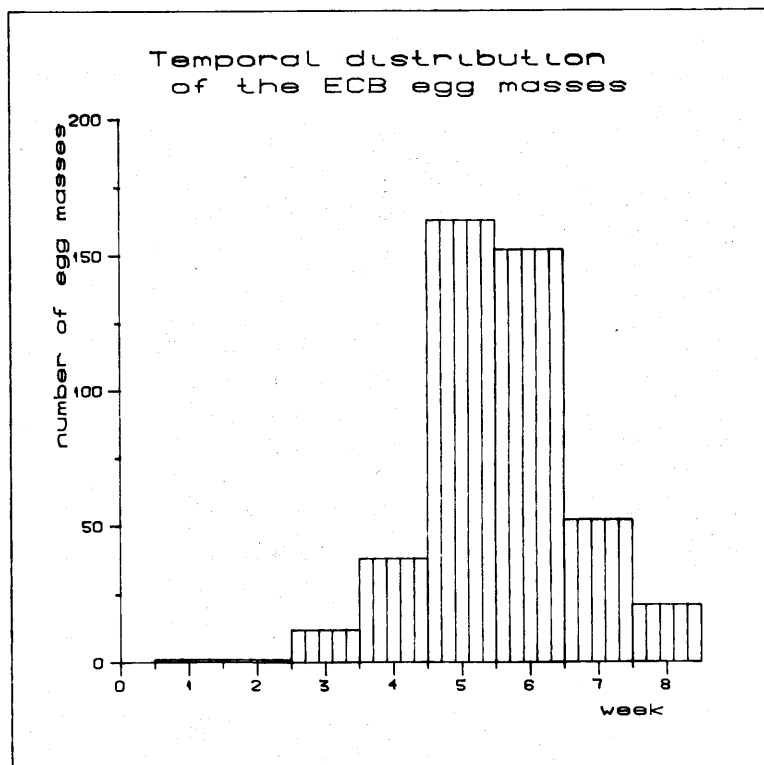


Fig. 1. Observed distribution of the European Corn Borer egg masses over the whole oviposition period for a sample of 504 corn plants.

density from this stratum to obtain a good estimation precision. Table 5 and practical and biological considerations suggest that this stratum must consist of middle plant leaves. The lowest leaves may be withered and the highest too exposed to the weather

Table 2. Correlation between counts of ECB egg masses on the statistical units for two consecutive weeks.

week	Time correlation coefficient	
	single plant $n=504$	cluster of size 21 $n=24$
3	-0.034	0.670*
4	0.043	0.147
5	0.072	0.018
6	0.107*	0.379*
7	0.114*	-0.131
8	—	—

$n$  is the number of pairs of counts ( $x_{i,t}$ ,  $x_{i,t+1}$ ) used for calculating the correlation coefficient,  $i$  being the statistical unit number and  $t$  the week number. A \* beside the time correlation coefficient for week  $t$  corresponds to a significant positive correlation between week  $t$  and week  $t+1$  at level 0.05 (unilateral test of correlation between two series of observations).

Table 3. Within-plant distribution of ECB egg masses on the 504 maize plants observed in summer 1980.

Week	Maximal number of available leaves	Number of egg masses found per leaf position												
		$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	$F_{11}$	$F_{12}$	$F_{13}$
1	11	0	0	0	0	0	0	0	1	0	0	0	0	0
2	13	0	0	0	0	1	0	0	0	0	0	0	0	0
3	15	0	0	0	0	0	3	0	7	1	1	0	0	0
4	16	0	0	0	1	3	7	10	9	4	3	0	1	0
5	16	0	0	3	12	29	38	23	28	20	6	3	1	0
6	16	0	0	1	6	20	21	23	32	23	18	5	3	0
7	16	0	0	0	1	5	7	14	10	4	4	4	2	1
8	16	0	0	0	0	0	4	8	2	2	1	2	2	0

$F_i$  corresponds to leaf position  $i$ , the leaves of the corn plant being arbitrarily numbered from the bottom up; the lowest leaf on the plant is designated as leaf position 1.

conditions. Furthermore, these leaves are less easy to reach for examination. The more convenient stratum on the various points of view seen above seems to be leaf positions 4 to 11. We have included leaf 11 in this stratum for not losing too much information about the within-plant distribution. During the three first weeks of the oviposi-

Table 4a. Within-plant distribution of ECB eggs on the 504 maize plants observed in summer 1980: Number of eggs found per leaf position.

Week	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	$F_{11}$	$F_{12}$	$F_{13}$
1	0	0	0	0	0	0	0	5	0	0	0	0	0
2	0	0	0	0	12	0	0	0	0	0	0	0	0
3	0	0	0	0	0	36	0	62	26	2	0	0	0
4	0	0	0	9	29	57	102	82	59	22	0	7	0
5	0	0	21	156	316	419	272	257	192	69	34	11	0
6	0	0	14	61	231	245	281	350	242	177	38	30	0
7	0	0	0	11	37	67	147	88	53	40	35	27	15
8	0	0	0	0	0	37	72	16	13	10	19	9	0

Table 4b. Within-plant distribution of ECB eggs on the 504 maize plants observed in summer 1980: Mean size of egg mass at each leaf position.

Week	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	$F_{11}$	$F_{12}$	$F_{13}$
1	—	—	—	—	—	—	—	5.0	—	—	—	—	—
2	—	—	—	—	12.0	—	—	—	—	—	—	—	—
3	—	—	—	—	—	12.0	—	8.9	26.0	2.0	—	—	—
4	—	—	—	9.0	9.7	8.1	10.2	9.1	14.8	7.3	—	7.0	—
5	—	—	7.0	13.0	10.9	11.0	11.8	9.2	9.6	11.5	11.3	11.0	—
6	—	—	14.0	10.2	11.6	11.7	12.2	10.9	10.5	9.8	7.6	10.0	—
7	—	—	—	11.0	7.4	9.6	10.5	8.8	13.3	10.0	8.8	13.5	15.0
8	—	—	—	—	—	9.3	9.0	8.0	6.5	10.0	9.5	4.5	—

$F_i$  corresponds to leaf position  $i$ , the leaves of the corn plant being arbitrarily numbered from the bottom up; the lowest leaf on the plant is designated as leaf position 1.



tional period, all the egg masses were found in this stratum. During the following weeks, only 14 egg masses were found outside this stratum. For the maximal oviposition period (weeks 5 and 6), of the total of 315 found, 97.5% were on leaf positions 4 to 11. For the two last weeks, the within-plant distribution was more spread over the different leaf positions.

We will now deal with applying numerical redistribution techniques for statistically testing hypotheses about the proportion of egg masses laid on leaf positions 4 to 11 during the oviposition period. Since we have seen above that the egg masses were not distributed at random among the plants (see tables 1a and 2), we will redistribute them only within the plants. This means that the number of egg masses on each plant for each week will not be changed. The redistribution will only consist in randomly assigning a leaf position to each egg mass within the plant in accordance with the number of available leaves on this plant. Different values for the probability of occupancy  $p$  of the stratum defined by leaf position 4 to 11 (or leaf positions at least equal to 4 for plants having less than 11 leaves) have been tested for the most infested weeks. Table 6 gives us the results. The values 0.95 and 0.96 provide a good fitting between observed frequencies and simulated frequencies for every week. Thus, these values may be used as correcting factor of the sample mean for sampling plans consisting in counting the number of egg masses in the stratum described above.

Another point we wanted to clarify is: are the biggest egg masses laid on particular leaf positions? The mean egg mass size was calculated for each week and each leaf stratum (table 4b). Despite Fig. 2 indicates no significant difference between the

Table 5. Relation between the position in which egg masses are laid and the number of available leaves.

Leaf position	Number of egg masses found per type of plant														T	T*
	$Pl_4$	$Pl_5$	$Pl_6$	$Pl_7$	$Pl_8$	$Pl_9$	$Pl_{10}$	$Pl_{11}$	$Pl_{12}$	$Pl_{13}$	$Pl_{14}$	$Pl_{15}$	$Pl_{16}$			
$F_3$	0	0	0	0	0	0	0	3	1	0	0	0	0	4	4	
$F_4$	1	0	0	0	1	0	5	2	8	2	1	0	0	20	20	
$F_5$	—	0	0	0	0	0	7	20	20	10	1	0	0	58	57	
$F_6$	—	—	0	0	0	1	6	27	23	19	3	1	0	80	77	
$F_7$	—	—	—	0	0	0	6	17	30	19	5	1	0	78	78	
$F_8$	—	—	—	—	0	0	3	20	31	20	10	4	1	89	81	
$F_9$	—	—	—	—	—	0	3	2	21	16	8	2	2	54	53	
$F_{10}$	—	—	—	—	—	—	1	4	7	12	6	3	0	33	32	
$F_{11}$	—	—	—	—	—	—	—	1	3	4	4	1	1	14	14	
$F_{12}$	—	—	—	—	—	—	—	—	2	1	2	3	1	9	9	
$F_{13}$	—	—	—	—	—	—	—	—	—	0	0	1	0	1	1	
$F_{14}$	—	—	—	—	—	—	—	—	—	—	0	0	0	0	0	
$F_{15}$	—	—	—	—	—	—	—	—	—	—	—	0	0	0	0	
$F_{16}$	—	—	—	—	—	—	—	—	—	—	—	—	0	0	0	

$Pl_i$  is the class of plants having  $i$  leaves at the infestation date.  $T$  corresponds to the total of egg masses laid per leaf position in the whole infestation period.  $T^*$  corresponds to the total of egg masses found per leaf position when plants were at an advanced phenology stage (weeks 4, 5, 6, 7, 8).

Table 6. Test about the occupancy rate of the middle plant stratum.

Week	Ranks of the Monte-Carlo test criterion for 199 simulations					
	$p=0.98$	$p=0.97$	$p=0.96$	$p=0.95$	$p=0.94$	$p=0.93$
4	75	75	118	109	141	155
5	91	75	161	180	187	196*
6	84	63	143	177	191*	196*
7	199*	128	90	43	47	85
8	200*	192*	190	116	122	117

$p$  is the tested value of occupancy rate for the middle plant stratum (leaf positions 4 to 11). The test criterion is the chi-squared statistic for two categories. The simulations consist in redistributing the egg masses within the plants taking into account the number of available leaves on each plant and without modifying the field distribution. The test is significant when the rank is greater or equal to 191. Then, the value of  $p$  is rejected.

mean sizes at the different dates (overlapping confidence intervals), the weeks were separately analysed. Table 7 presents these analyses. Without modifying the field and the within-plant distributions, random permutations of the egg mass sizes were simulated and a related Monte Carlo test was performed at level 0.05. The result is that at any date, there is no correlation between the egg mass size and the infested leaf position.

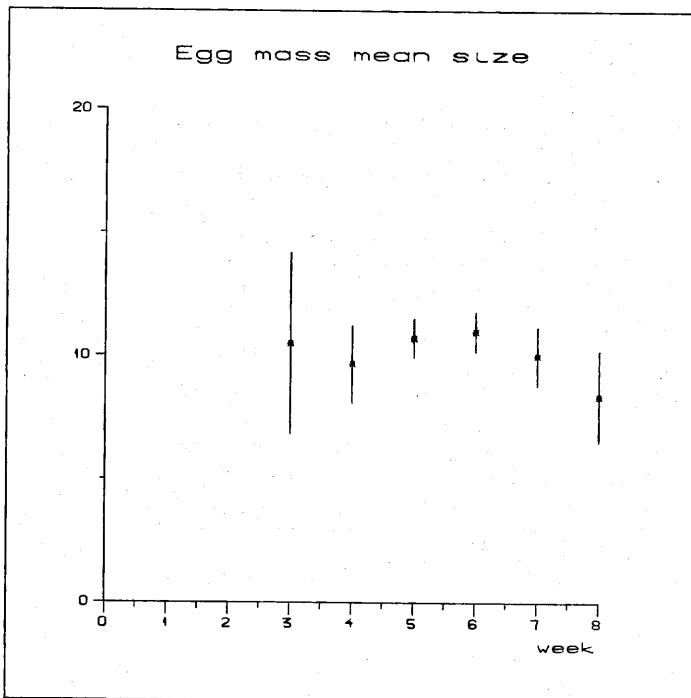


Fig. 2. Mean size of the ECB egg masses and its 95% confidence interval (except for weeks 1 and 2 for which only one egg mass was found).

Table 7. Test of non correlation between the egg mass size and the egg mass position within the plant.

week	Ranks of the Monte-Carlo test criterion for 199 simulations	
	$p=0.95$	$p=0.96$
4	93	93
5	63	62
6	161	160
7	119	131
8	127	115

$p$  is the occupancy rate for the middle plant stratum (leaf positions 4 to 11). The test criterion is  $C = \sum_{i=2}^9 N_i^2/p_i$  where  $N_1$  is the number of eggs found outside this stratum,  $N_i$ , for  $i=2, \dots, 9$  is the number of eggs found on leaf position  $i+2$ . The corresponding occupancy rates are  $p_1=1-p$  and  $p_i=p/8$ , for  $i=2, \dots, 9$ . The simulations consist in assigning at random the egg mass sizes to the egg masses for a given week without modifying their spatial position in the field and within the plants. The test is significant when the rank is greater or equal to 191.

#### DISCUSSION

We have only considered single classifications in this paper but the generalization of the Monte Carlo tests to cross-classified categorical data is straightforward. Indeed, it often happens that individuals from a sampled population are cross-classified according to several categorical variables. In such a case, the redistribution of individuals is to be subject to constraints in order to respect the marginal totals of the multidimensional contingency table. When individuals are distributed independently of each other and their total number is large enough, several statistical techniques exist (relevant references and illustrations by actual data are given in Fienberg (1980)). Otherwise, it is worth applying redistribution based Monte Carlo test.

Techniques involving redistribution of individuals or groups independently of each other have been mainly considered. Other kinds of population distribution involving complex non-random distribution of groups or solitary individuals are met in practice. If their simulations can be implemented without too much time consuming, Monte Carlo tests are still attractive and may be the only way of testing hypotheses about the occupancy rates. Meanwhile, when the parameters of the simulated distribution are estimated from the same observations used for testing the null hypothesis, the properties of Monte-Carlo tests may be doubtful (Besag and Clifford, 1989; Hall and Terrington 1989).

An interesting feature of numerical redistribution based tests is the possibility of taking into account constraints due to the non random distribution of individuals, for example when some individuals can not belong to certain categories or when they are distributed by groups: in the case of the European Corn Borer eggs, hypotheses about leaf position occupancy rates were tested without modifying the field and the within plant distributions and the number of available leaves on each plant was taken into account.

## SUMMARY

Standard statistical analyses of distributions of individuals from contingency tables are generally invalid if the individuals are not distributed independently of each other. In this paper, we discuss a method of testing hypotheses about classification category occupancy rates for overdispersed population or for population whose individuals are distributed by groups rather than lonely. These methods are based on population redistribution simulations and provide valid, exact and powerful tests in situations for which classical methods are not appropriate. Illustrations are given from the European Corn Borer eggs data.

## REFERENCES

- Barnard, G. A. (1963) Discussion of Professor Bartlett's paper. *J. R. Statist. Soc. B* **25**: 435-441.
- Besag, J. and P. Clifford (1989) Generalized Monte Carlo significance tests. *Biometrika* **76**: 633-642.
- Besag, J. E. and P. J. Diggle (1977) Simple Monte Carlo tests for spatial pattern. *J. R. Statist. S. B* **26**: 327-333.
- Bissel, A. F. (1972a) A negative binomial model with varying element sizes. *Biometrika* **59**: 435-441.
- Bissel, A. F. (1972b) Another negative binomial model with varying element sizes. *Biometrika* **59**: 691-693.
- Bliss, C. I. (1971) The aggregation of species within spatial units. *Stat. Ecol.* **1**: 311-335.
- Denis, J.-B. and J. Vaillant (1985) Analyse statistique de données de piégeage d'insectes pour mettre en évidence des comportements agrégatifs. *Revue de Statist. Appl.* **33**: 29-42.
- Despins, J. L. and J. E. Roberts (1986) Within-plant and within-field distribution of first-generation European Corn Borer (Lepidoptera: Pyralidae) egg masses on field corn in Virginia. *Environ. Entomol.* **15**: 106-108.
- Diggle, P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Fienberg, S. E. (1980) *The Analysis of Cross-Classified Categorical Data*. MIT Press, London.
- Hall, P. and D. M. Terrington (1989) The effect of simulation on level accuracy and power of Monte Carlo tests. *J. R. Statist. Soc. B* **51**: 459-467.
- Hawlitzky, N. (1986) Etude de la biologie de la pyrale du maïs, *Ostrinia nubilalis*, Hbn. (Lep. Pyralidae) en région parisienne durant quatre années et recherche d'éléments prévisionnels du début de ponte. *Acta OEcologica, OEcol. Applic.* **7**: 47-68.
- Hope, A. C. A. (1968) A simplified Monte Carlo test procedure. *J. R. Statist. Soc. B* **30**: 582-598.
- Iwao, S. (1968) A new regression method for analyzing the aggregation pattern of animal populations. *Res. Popul. Ecol.* **10**: 1-20.
- Iwao, S. and E. Kuno (1971) An approach to the analysis of aggregation pattern in biological populations. *Stat. Ecol.* **1**: 461-513.
- Jöckel, K.-H. (1986) Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Ann. Statist.* **14**: 336-347.
- Lloyd, M. (1967) Mean Crowding. *J. Anim. Ecol.* **36**: 1-30.
- Marriot, F. H. C. (1979) Monte Carlo tests: how many simulations? *Appl. Statist.* **28**: 75-77.
- Pielou, E. C. (1977) *Mathematical Ecology*. Wiley, New-York.
- Plackett, R. L. (1974) *The analysis of categorical data*. Griffin, London.

- Southwood, M. (1978) *Ecological Methods with Particular Reference to the Study of Insect Populations*. Chapman and Hall, London.
- Taylor, L. R. (1984) Assessing and interpreting the spatial distributions of insects populations. *Ann. Rev. Entomol.* **29**: 321-357.
- Vaillant, J. and I. Badenhauer (1989) Etude de répartition d'individus sur transect ou grille régulière par tests de Monte-Carlo. *Biom. Praxim.* **29**: 153-172.
- Vaillant, J. (1989) Spatial-temporal models for counting processes in ecology. *C. E. R. O. Operat. Res. Statist. and Appl. Math.* **31**: 277-296.

集中分布する個体群の分布の再構成手法による存在確率の統計的解析.  
ヨーロッパ産アワノメイガの卵塊の分布に対する適用

J. VAILLANT · N. HAWLITZKY

個体が互いに独立に分布していない場合、分割表によって行う標準的な統計的解析は一般に正しくない。この論文では、集中分布もしくは単独よりもむしろグループを作って分布する個体群のカテゴリカルな存在確率に関する仮説を検定する方法を論じた。これらの方法は、シミュレーションにより個体群の分布を再構成することに基づいており、標準的な方法が適当でない場合でも正しく、正確で効力のある検定を与える。解説例にヨーロッパ産アワノメイガの卵塊のデータを用いた。