

## ESTIMATION OF EUROPEAN CORN BORER EGG MASSES DENSITY BY SAMPLING OF RUNS

J. VAILLANT and S. DERRIDJ\*

Laboratory of Biometry INRA-Versailles; 78000 Versailles, France

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

### INTRODUCTION

In the study of maize field infestation by the European Corn Borer (ECB) *Ostrinia nubilalis* HBN., a reliable estimate of the average number  $\lambda$  of egg masses per plant (also called infestation intensity) is often required. Furthermore, an observation method limiting sampling cost is needed because egg mass detection (egg masses are hidden on the underside of the maize leaves) and walking in large fields are tedious. Usually, entomologists apply the well known cluster sampling method which generally provides an unbiased estimate of  $\lambda$ . In this paper, we present a more attractive procedure which facilitates moving and observing in the field. The precision of the estimates obtained by this procedure is investigated under different hypotheses.

Generally, for a fixed observation date, no more than one out of three plants is infested by ECB egg masses in the Paris area (DENECHERE et al., 1982; VAILLANT, 1985). The pest infestation threshold is 0.12 for most maize varieties at the fresh silking stage (STENGEL, 1969) but can be weaker. On the other hand, when the infestation intensity  $\lambda$  increases slightly, the index of dispersion at the plant's scale also increases, rather than the proportion of infested plants. By analogy with the distance-based methods (DIGGLE, 1975; Byth, 1982; Diggle and Cox, 1983), the idea is that the number of sound plants (i.e., not infested) between a plant chosen at random and the nearest infested plant in the same row (in a given direction of observation of this row) enables us to build good estimators of  $\lambda$ . This seems to be reasonable whenever there is not too much information lost when the infestation state of plants is observed instead of the total number of egg masses.

In the sequel, our intention is to show that the estimation of the ECB infestation intensity can be approached in a quite different way from techniques involving full counts on sampled units, and that similar procedures can be considered for other kinds of low intensity spatial distribution of individuals provided that the sampling units are partially aligned as for plants in maize fields.

COMPLETE SPATIAL RANDOMNESS CASE

The hypothesis  $H_0$  of complete spatial randomness for the distribution of ECB egg masses implies that the number of egg masses layed on any plant follows a POISSON law  $P(\lambda)$  (see for example TAYLOR, 1984). Under this hypothesis, all fixed size sampling procedures are equivalent because the numbers of egg masses on different plants are mutually independent and equidistributed (COCHRAN, 1977). Meanwhile, informative sampling procedures (i.e procedures integrating information about the distribution of individuals and taking this into account during the application of the sampling plan) like sequential sampling (DENECHERE et al., 1982; BADENHAUSSER and VAILANT, 1987) can provide a better compromise between sampling cost and estimation precision. Moreover, complete spatial randomness is an idealized standard rarely attainable in practice but tenable as a convenient first approximation (DIGGLE, 1983). In the sequel, we will describe a run-based sampling method and its main features under this hypothesis  $H_0$ . Alternative hypotheses will be investigated in the next paragraph.

Let  $A_1, \dots, A_n$  be completely randomly selected plants (in the following, we shall call them Original Sampling Points (OSP)). For each  $A_i$ , we can measure the number  $u_i^*$  of plants observed before finding an infested plant in a given observation direction within the row containing  $A_i$ . Figure 1 gives an illustration of such a procedure applied within a single row. We can see that, sometimes, only some OSP among  $A_1, \dots, A_n$  provide information about infestation runs. In the following, we shall call them Instructive OSP (IOSP). Of course, when we work in a large field with a low sampling intensity, we can randomly select OSP until we obtain  $n$  IOSP  $B_1, \dots, B_n$  with  $n$  corresponding run lengths  $u_1, \dots, u_n$ .

Under  $H_0$ , the  $u_i, i=1, \dots, n$ , are independent and identically distributed in accordance with a geometric law whose parameter is  $e^{-\lambda}$ . Then, the samplig likilihod is:

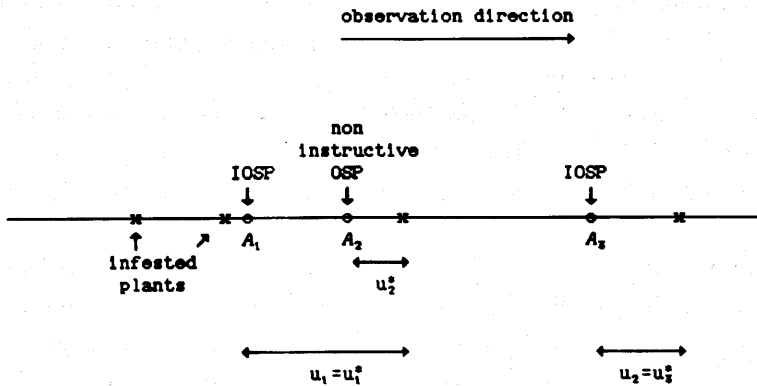


Fig. 1. Illustration of an infestation run-based sampling procedure along a row.  
(I)OSP=(Instructive) Original sampling point.

$$L(u, \lambda, n) = \prod_{i=1}^{i=n} (1 - e^{-\lambda}) e^{-\lambda(u_i - 1)} = (1 - e^{-\lambda})^n e^{-\lambda(\bar{u} - 1)} \quad (2.1)$$

where  $\bar{u} = \sum_{i=1}^{i=n} u_i / n$  is the mean value of the  $u_i$ .

Under  $H_0$ , the maximum likelihood estimate of  $\lambda$  is

$$\hat{\lambda} = \log\left(\frac{\bar{u}}{\bar{u} - 1}\right) \quad (2.2)$$

and it can be shown that  $\tilde{\lambda} = \hat{\lambda} / (1 + n^{-1})$  is almost non-biased. Its asymptotic variance is:

$$\text{Var}(\tilde{\lambda}) = \frac{n(e^\lambda - 1)(1 - e^{-\lambda})}{(n + 1)^2} \quad (2.3)$$

Under  $H_0$ , let us compare this sampling plan noted  $R(n)$  with the simple random sampling of  $m$  plants, say  $S(m)$ , associated to the sample mean  $\bar{\lambda}$  of the number of egg masses. As we said above, under complete spatial randomness, this is equivalent to comparing  $R(n)$  with any fixed size  $m$  sampling procedure associated to  $\bar{\lambda}$ . Another formulation of this choice problem is:

Is it better to count the egg masses on  $m$  plants rather than examining infestation runs from  $n$  sampling points,  $m > n$ ?

First of all, it can be pointed out that we could have compared  $R(n)$  with a fixed-size sampling procedure involving observation of the infestation state of plants. But it is well known that such a sampling strategy is always less precise than the corresponding one involving the full counts on plants. Moreover, under  $H_0$ , when the average sampling size is fixed,  $R(n)$  and any fixed-size sampling strategy based on infestation state observations are equivalent.

Meanwhile,  $R(n)$  also provides a way of testing  $H_0$  as we will see later on, whereas this is not always possible with fixed-size sampling plan based on the infestation state.

Thus, it is more interesting to compare directly  $R(n)$  with a sampling plan involving full counts on the sampled plants and to take into account the gain due to a less heavy examination of plants.

Let us write  $m = \alpha n$  with  $\alpha > 1$ .

Before discussing the sampling cost, let us study the difference  $H(\alpha, n, \lambda)$  between  $\text{Var}(\tilde{\lambda})$  and  $\text{Var}(\bar{\lambda})$ . We have:

$$H(\alpha, n, \lambda) = n e^\lambda \left( \frac{1 - e^{-\lambda}}{n + 1} \right)^2 - \frac{\lambda}{n \alpha} \quad (2.4)$$

An easy study of  $H(\alpha, n, \cdot)$  on  $\mathbf{R}^+$  shows (Fig. 2) that, for  $\alpha$  and  $n$  given,  $R(n)$  is better than  $S(n\alpha)$  when  $\lambda$  is below a certain value  $\lambda_1$ . Intuitively, the reason is that when the mean number of egg masses per plant is not too large, the information given by  $S(n\alpha)$  is not greater than that provided by  $R(n)$ .

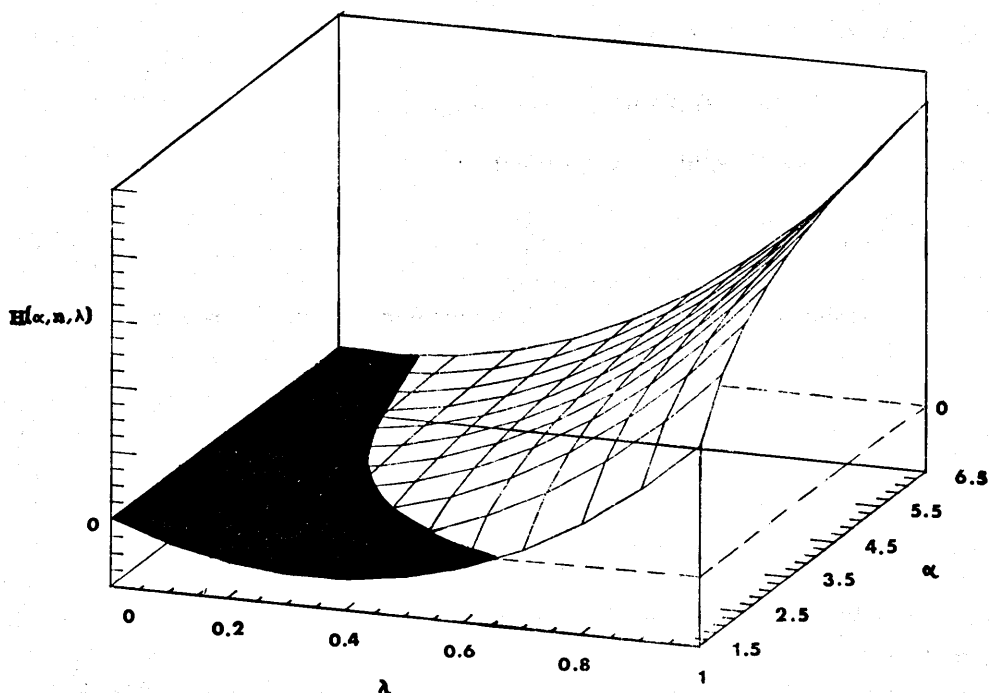


Fig. 2. Comparison of  $R(n)$  and  $S(n\alpha)$  under  $H_0$ , for a given  $n$ . The dark area corresponds to value of  $(\alpha, \lambda)$ , for which  $H(\alpha, n, \lambda) \leq 0$ .

As mentioned previously, an important preoccupation in ecological sampling is the associated cost of a sampling procedure. A simple mathematical approach of observation cost under  $H_0$  (see appendix) shows that  $R(n)$  can be both less expensive and more precise than  $S(n\alpha)$  for a relatively small parameter domain. Time saving due to the stopping procedure plays an important part in these results.

On the other hand, the number of plants observed with  $R(n)$  can theoretically be infinite. In practice, of course, it is necessary to set an upper limit for this number of sampled plants. One way of achieving this is to truncate the run observations  $u_1, \dots, u_n$  in the following way:

Starting from a IOSP, we stop observing the corresponding run if no egg mass has been detected after examining  $n_0$  plants (Fig. 3). Thus, under  $H_0$ , we observe realizations of independent random variables  $v_1, \dots, v_n$  identically distributed in accordance with a modified geometric law. Then, the sample likelihood becomes:

$$L(v, \lambda, n, n_0) = \left( \prod_{v_i \leq n_0} (1 - e^{-\lambda}) e^{-\lambda(v_i - 1)} \right) \prod_{i=1}^{n - N_1} e^{-\lambda n_0} \quad (2.5)$$

with

$$N_1 = \sum_{i=1}^{i=n} I_{[v_i \leq n_0]}$$

( $I_A$  denotes the indicator function for the event A).

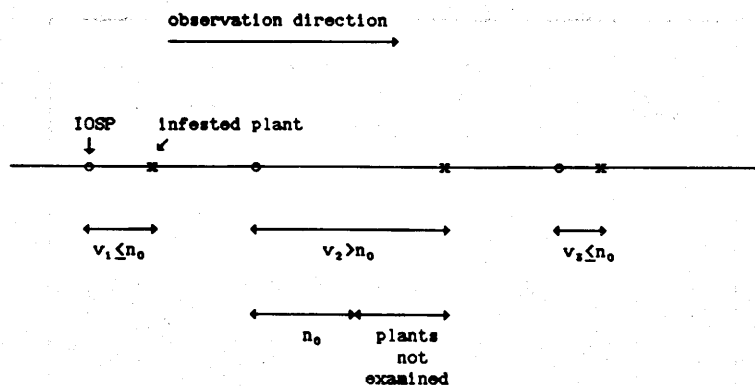


Fig. 3. Illustration of a run-based sampling procedure with truncated observation.

The maximum likelihood estimator  $\lambda^*$  of  $\lambda$  is:

$$\lambda^* = \log \left[ 1 + \frac{1}{\bar{v} - 1 + n_0(n/N_1 - 1)} \right] \quad (2.6)$$

where

$$\bar{v} = N_1^{-1} \sum_{i=1}^{i=n} v_i I_{[v_i \leq n_0]}$$

and its asymptotic variance is

$$\text{Var}(\lambda^*) = \frac{(e^\lambda - 1)(1 - e^{-\lambda})}{n(1 - e^{-\lambda n_0})} \quad (2.7)$$

This sampling procedure, say  $R'(n, n_0)$  is slightly less precise than  $R(n)$  when  $\lambda n_0$  is large enough (Fig. 4), but the number of plants to examine is at most equal to  $nn_0$ , what is an interesting feature in practice.

It is worth noticing that infestation run observations can also provide methods for testing hypotheses about spatial distribution of individuals on aligned discrete units (PIELOU, 1962; VAILLANT, 1985). For example, we can test the goodness-of-fit of such observations with a geometric law derived from the assumption of independence between spatial units' infestations.

#### OVERDISPERSION CASE

Previously, we examined the case which analytically is the most tractable (i.e., complete spatial randomness). Let us now discuss some particular alternative hypotheses.

##### Independent infestations

Suppose again that the plants are infested independently of each other but that the

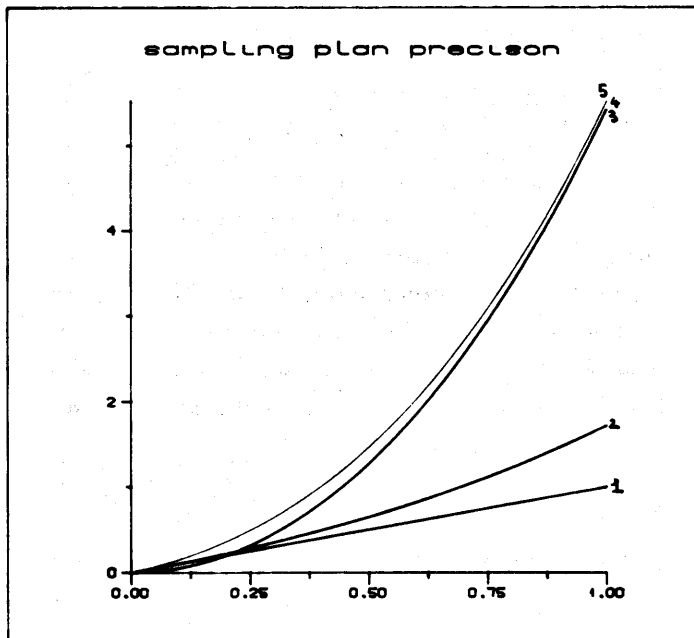
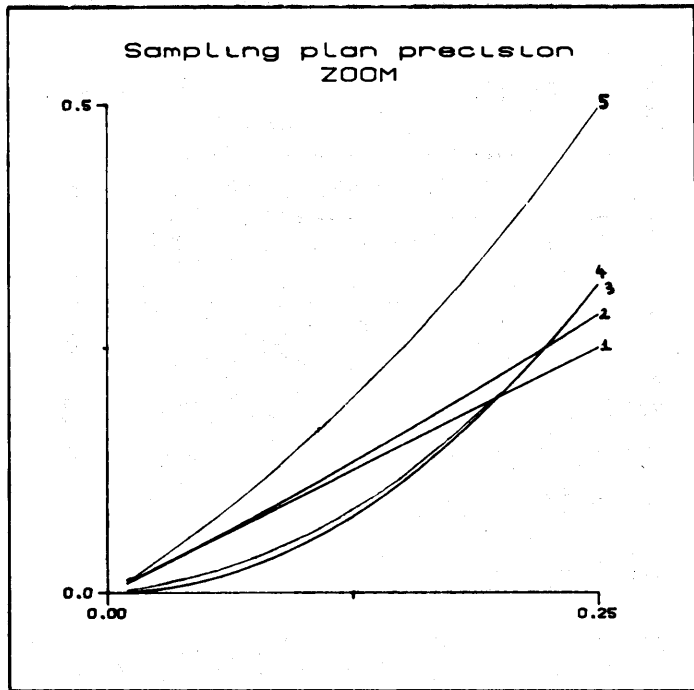


Fig. 4. Effect of truncature and comparison of some sampling procedures under  $H_0$ . In abscissa,  $\lambda$ ; in ordinate, for a given number, say  $n$ , the variance of estimation for:  $^*S(5n)$  when the observations are full counts (curve 1);  $^*S(5n)$  when the observations are presence-absence of individuals (curve 2);  $^*R(n)$  procedure (curve 3);  $^*R'(n, 20)$  procedure (curve 4);  $^*R'(n, 4)$  procedure (curve 5).

numbers of egg masses layed on these plants by the ECB females are equidistributed but not necessarily independent.

Under this hypothesis, say  $H_1$ , for a given plant  $i$ , the number  $N_i$  of egg masses received can be written:

$$N_i = I_i F_i \quad (3.1)$$

where  $F_i$  is the number of egg masses received by  $i$ , ( $F_i > 0$ );

$I_i$  is equal to 1 if plant  $i$  is infested, 0 if not (in the sequel, the probability  $P(I_i = 1)$  of infestation will be noted  $p$ ).

Under  $H_1$ , we have for any couple of distinct plants  $(i, j)$ ,  $Cov(I_i, I_j) = 0$ .

Thus, complete spatial randomness is a special case of  $H_1$ :

- For any couple  $(i, j)$ ,  $Cov(I_i, F_j) = Cov(I_i, F_i) = 0$ ,  $i \neq j$ .
- $F_i$  follows a POISSON law  $P(\lambda)$  truncated at zero.
- $E(I_i) = 1 - e^{-\lambda}$ .

Under hypothesis  $H_1$ , procedure  $R^l(n, n_0)$  provides a sample likelihood based again on realizations of a modified geometric law:

$$L(v, \lambda, n, n_0) = \left( \prod_{v_i \leq n_0} P(1-p)^{v_i-1} \right) \prod_{i=1}^{n-N_1} (1-p)^{n_0} \quad (3.2)$$

and the maximum likelihood estimator  $p^*$  of  $p$  is:

$$p^* = \frac{1}{\bar{v} + n_0(n/N_1 - 1)} \quad (3.3)$$

Therefore, we can estimate  $\lambda$  if the relation between  $p$  and  $\lambda$  is such that:

$\lambda = f(p)$  where  $f$  is an known injective function from  $[0; 1]$  to  $\mathbf{R}^+$ . Then, the maximum likelihood estimator of  $\lambda$  is:

$$\lambda^* = f(p^*). \quad (3.4)$$

When  $g = f^{-1}$  is twice differentiable, the asymptotic variance of  $\lambda^*$  is

$$\begin{aligned} Var(\lambda^*) = & \frac{1}{n} \{ K p^{-2} [g'(\lambda)^2 - p g''(\lambda)] \\ & + ((pK)^{-1} - K + n_0(1-K))(1-p)^{-2} [g'(\lambda)^2 + (1-p)g''(\lambda)] \}^{-1} \end{aligned} \quad (3.5)$$

where  $K = 1 - (1-p)^{n_0}$

Thus, under  $H_1$ , we can also get the maximum likelihood estimate of  $\lambda$  and its precision despite autocorrelation between the  $F_i$ .

For example, overdispersion (see TAYLOR, 1984) at the plants' scale may imply that for any plant  $i$ ,  $N_i$  follows a Negative Binomial law with aggregativity parameter  $k$ . Then, we have the following relation between  $p$  and  $\lambda$ :

$$p = 1 - \left( 1 + \frac{\lambda}{k} \right)^{-k} \quad (3.6)$$

Consequently,

$$\lambda = k((1-p)^{\frac{1}{k}} - 1) \quad (3.7)$$

Therefore, when  $k$  is known, we can calculate  $\lambda$  and estimate its variance.

### Dependent infestations

We will now see two situations for which plants are not infested independently of each other.

The first case arises when individuals are distributed independently of each other but in the presence of random heterogeneity in the natural environment: this can be described by assuming that the parameter  $\lambda$  in hypothesis  $H_0$ , is now a random variable. Therefore, the sample likelihood given by equation 3.2 becomes

$$L(v, \theta, n, n_0) = E\left[\prod_{v_i \leq n_0} (1 - e^{-\lambda}) e^{-\lambda(v_i - 1)} \prod_{i=0}^{n - N_1} e^{-\lambda n_0}\right] \quad (3.8)$$

where the expectation is on the random variable  $\lambda$  whose probability distribution is parametrized by  $\theta$ .

Thus,

$$L(v, \theta, n, n_0) = \sum_{i=0}^{N_1} \binom{N_1}{i} (-1)^i E(e^{-\lambda(i+A)}) \quad (3.9)$$

$$\text{with } A = N_1(\bar{v} - 1) + n_0(n - N_1).$$

For example, if  $\lambda \sim G(k, w/k)$  (where  $G(\dots)$  is the Gamma distribution with exponent parameter  $k$  and expectation  $w$ ), we have:

$$L(v, (k, w), n, n_0) = \sum_{i=0}^{N_1} \binom{N_1}{i} (-1)^i (1 + (i+A)w/k)^{-k} \quad (3.10)$$

$$\text{and } E(\lambda) = w.$$

This likelihood does not provide an analytic expression for the maximum-likelihood estimate of  $w$ . Furthermore, such a model implies that infestations on any pair of plants inside the considered maize line are correlated in the same way, which is not realistic.

On the other hand, neighbourhood correlation can be taken into account by means of nearest neighbour systems (BARLETT, 1978, § 2.22):

Let us assume that infestation on plant  $i$  is only correlated to infestations on plants  $i-1$  and  $i+1$ .

Let us note  $p_{01} = P$  (plant  $i$  is infested/plant  $i-1$  is not infested)

The above hypothesis, say  $H_2$ , includes  $H_1$  which is the case where the non-conditional probability of infestation  $p$  equals  $p_{01}$ .

Procedure  $R^i(n, n_0)$  provides the following sample likelihood:

$$p^{n_1} (1-p)^{n-n_1} p_{01}^{n_2} (1-p_{01})^{(n_0-1)(n-n_1-n_2) + n_2(\bar{v}_2-2)} \quad (3.11)$$



with

$$n_1 = \sum_{i=1}^n I_{[v_i=1]}, \quad n_2 = \sum_{i=1}^n I_{[1 < v_i \leq n_0]}, \quad \bar{v}_2 = \frac{1}{n_2} \sum_{i=1}^n v_i I_{[1 < v_i \leq n_0]}.$$

The maximum-likelihood estimates of  $p$  and  $p_{01}$  are:

$$\hat{p} = \frac{n_1}{n} \quad \hat{p}_{01} = \frac{1}{\bar{v}_2 + (n_0 - 1)(n - n_1)/n_2 - n_0}. \quad (3.12)$$

As for hypothesis  $H_1$ , to make inference about the infestation intensity  $\lambda$ , we need to know what kind of relation exists between  $p$ ,  $p_{01}$  and  $\lambda$ .

For example, if we consider the two-nearest-neighbours spatial model for negative binomially distributed counting observation on aligned units (see VAILLANT, 1988), the count  $x_j$  on plant  $j$  follows the negative binomial distribution  $NB(k, \lambda)$ , (where  $k$  is the exponent parameter and  $\lambda$  the expectation). Moreover, conditionally to counts  $x_1, \dots, x_{j-1}$  on plants  $1, \dots, j-1$ ,  $x_j$  follows the Negative Binomial

Table 1. Descriptive statistics of the maize rows and the ECB egg masses distributions.

row	$N$	$p$	$m$	$\sigma^2$	$\sigma^2/m$	$N_0$	$N_1$	$N_2$	$N_3$	$N_4$	$\bar{l}$	$l_1$	$l_{max}$
1	485	.157	.169	.166	.982	409	70	6	0	0	6.13	12	22
2	513	.150	.166	.174	1.048	436	70	6	1	0	6.54	12	29
3	503	.181	.203	.218	1.074	412	82	8	0	1	5.42	15	26
4	412	.114	.119	.120	1.006	365	46	0	1	0	8.89	7	52
5	634	.142	.151	.148	.975	544	84	6	0	0	6.70	17	42
6	585	.096	.104	.111	1.061	529	51	5	0	0	10.20	10	62
7	314	.248	.283	.293	1.034	236	69	8	0	1	4.01	22	14
8	495	.293	.329	.302	.918	350	129	14	2	0	3.38	50	16
9	449	.283	.334	.330	.988	322	105	21	1	0	3.40	42	15
10	533	.161	.176	.176	.996	447	78	8	0	0	6.19	16	29
11	525	.179	.196	.192	.980	431	85	9	0	0	5.59	17	22
12	548	.201	.219	.212	.966	438	101	8	1	0	4.93	25	17
13	586	.118	.125	.126	1.014	517	66	2	1	0	8.21	8	39
14	529	.102	.106	.102	.967	475	52	2	0	0	9.30	7	35
15	631	.100	.103	.099	.960	568	61	2	0	0	10.02	15	69
16	511	.239	.282	.305	1.081	389	104	14	4	0	4.19	28	13
17	442	.201	.235	.267	1.133	353	77	10	1	1	4.90	18	19
18	455	.213	.237	.230	.968	358	86	11	0	0	4.64	23	22
19	390	.223	.262	.276	1.055	303	73	13	1	0	4.12	25	25
20	390	.259	.321	.383	1.195	289	84	11	5	1	3.79	28	18
21	390	.200	.251	.307	1.221	312	61	14	3	0	4.53	14	21
22	390	.141	.154	.161	1.049	335	51	3	1	0	6.59	12	63
23	390	.215	.251	.266	1.058	306	71	12	1	0	4.66	17	19
24	390	.123	.136	.143	1.055	342	43	5	0	0	8.23	9	28
25	390	.244	.318	.377	1.185	295	68	25	2	0	3.88	34	20
26	390	.267	.333	.377	1.131	286	82	18	4	0	3.55	33	14
27	390	.282	.354	.399	1.127	280	86	21	2	1	3.23	42	23
28	390	.197	.228	.264	1.157	313	69	5	2	1	4.72	16	38
29	390	.213	.262	.307	1.173	307	67	13	3	0	4.55	19	25
30	390	.177	.197	.200	1.013	321	61	8	0	0	5.32	11	20

$N$ =number of plants in the row,  $p$ =proportion of infested plants,  $(m, \sigma^2)$ =(mean and variance for the number of egg masses per plant),  $N_i$ =number of plants having  $i$  egg masses,  $\bar{l}$ =mean length of the infestation runs,  $l_1$ =number of unity length infestation run,  $l_{max}$ =maximum infestation run length.

$$NB\left(k + x_{j-1}, \lambda \frac{k + x_{j-1}}{k + \lambda}\right). \quad (3.13)$$

Therefore, we have

$$p = 1 - \left(\frac{k}{k + \lambda}\right)^k \quad \text{and} \quad p_{01} = 1 - \left(\frac{k + \lambda}{k + 2\lambda}\right)^k. \quad (3.14)$$

Hence,

$$\lambda = \frac{k}{2} \left( \left[ (1 - p_{01})(1 - p) \right]^{\frac{1}{k}} - 1 \right). \quad (3.15)$$

It is possible to estimate both  $\lambda$  and  $k$  from expression 3.14. For a given value of  $k$ , the maximum-likelihood estimate of  $\lambda$  can be obtained analytically or numerically. An estimate of its variance may be obtained using an equation similar to equation 3.5.

Table 2. Ranks of Monte-Carlo test criteria for 199 simulations.  $I_D$  corresponds to the variance-mean ratio of the number of egg masses per plant,  $\rho_1$  corresponds to the first-order autocovariance of the number of egg masses per plant,  $V(l)$  is the variance of the infestation run length,  $l_1$  is defined as for Table 1.

row	Hypothesis $H_0$ simulated		Hypothesis $H_1$ simulated		
	$I_D$	$\rho_1$	$V(l)$	$l_1$	$\rho_1$
1	72	59	39	84	50
2	159	116	82	112	121
3	174	170	149	56	158
4	107	170	196	148	174
5	61	160	187	187	165
6	162	194	197	193	199
7	121	165	66	146	159
8	17	195	155	189	190
9	87	186	60	184	190
10	87	154	150	140	131
11	59	53	81	99	44
12	45	152	48	154	163
13	87	65	114	74	56
14	39	134	105	149	130
15	30	200	196	200	200
16	176	148	40	79	145
17	192	193	112	93	180
18	52	11	142	142	126
19	152	170	191	187	159
20	197	100	133	137	96
21	198	48	19	65	59
22	135	198	154	190	197
23	150	44	107	74	49
24	144	171	137	180	162
25	200	200	199	200	199
26	186	177	57	186	172
27	183	198	197	199	199
28	194	80	188	121	83
29	196	117	175	125	114
30	126	57	3	60	62

## ANALYSES OF DATA

In this paragraph, our purpose is to study the characteristics of the ECB egg masses spatial distribution within maize lines from data carried out in the Paris area on several maize fields.

Statistical techniques based on the infestation runs' length and the number of ECB egg masses per plant were applied on data obtained from 30 maize lines. Table 1 shows us some descriptive features of the spatial distribution like the index of dispersion, the mean number of egg masses per plant and the maximum length of the infestation runs. Monte-Carlo tests (see DIGGLE, 1983) based on different criteria were applied for each line from simulations of hypotheses  $H_0$ ,  $H_1$ ,  $H_2$ . Their results are given in Table 2. We can see that the tests of  $H_0$  and  $H_1$  are significant for some of the maize rows, but not necessarily for the most infested ones. Thus,  $H_1$  is rejected for

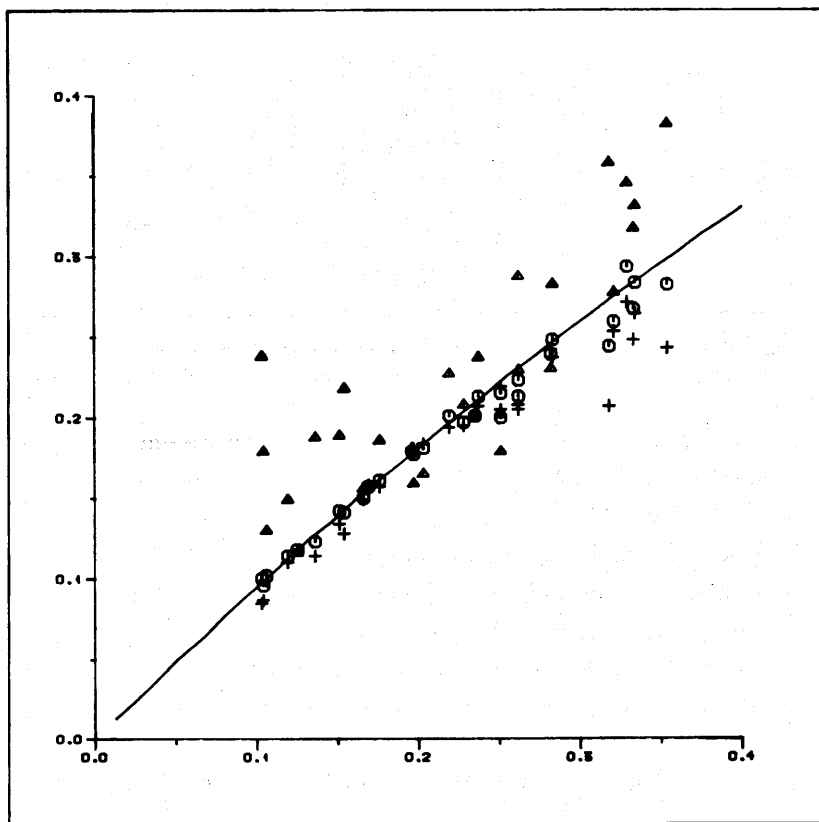


Fig. 5. Relation between the mean number of egg masses per plant (abscissa) and the (conditional or non-conditional) proportion of infested plants for a maize line.  $\Delta$  (respectively +) indicates the probability for a plant in the maize line to be infested, given the preceding plant was (respectively was not) infested.  $\circ$  indicates the non-conditional probability. The solid line represents the theoretical relation under complete spatial randomness.

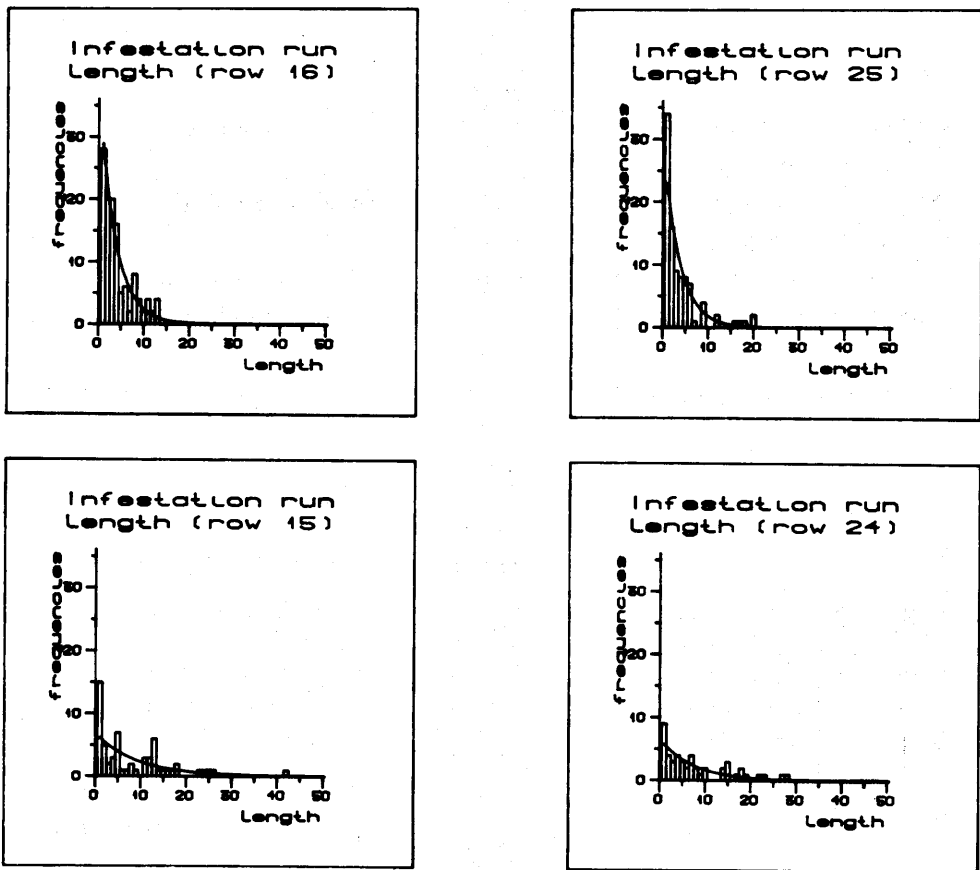


Fig. 6. Distribution of the infestation run length for rows 15, 16, 24, 25. The solid line corresponds to the theoretical distribution under complete spatial randomness.

row 6 and row 15 while the corresponding infestation intensities are 0.104 and 0.103. On the other hand, the tests of  $H_2$  are all significant. Figures 5 and 6 illustrate the tendency detected by the Monte-Carlo tests: The infested plants are rather close to each other compared to their expected locations under hypothesis  $H_1$ .

In spite of this, the relation between the expected infestation run length and the infestation intensity  $\lambda$  seems close to the one provided under  $H_0$  (Fig. 7). Consequently, expression 2.2 provides an interesting estimator of  $\lambda$  for non truncated observations while it can lead to overestimation when procedure  $R^t(n, n_0)$  is applied. For this reason, the correction factor  $n_0(n/N_1 - 1)$  must be applied in expression 2.6. Therefore, estimator  $\lambda^*$  given by this expression is better since it takes into account the truncature.

#### DISCUSSION

This paper presents a framework for estimation of population density when

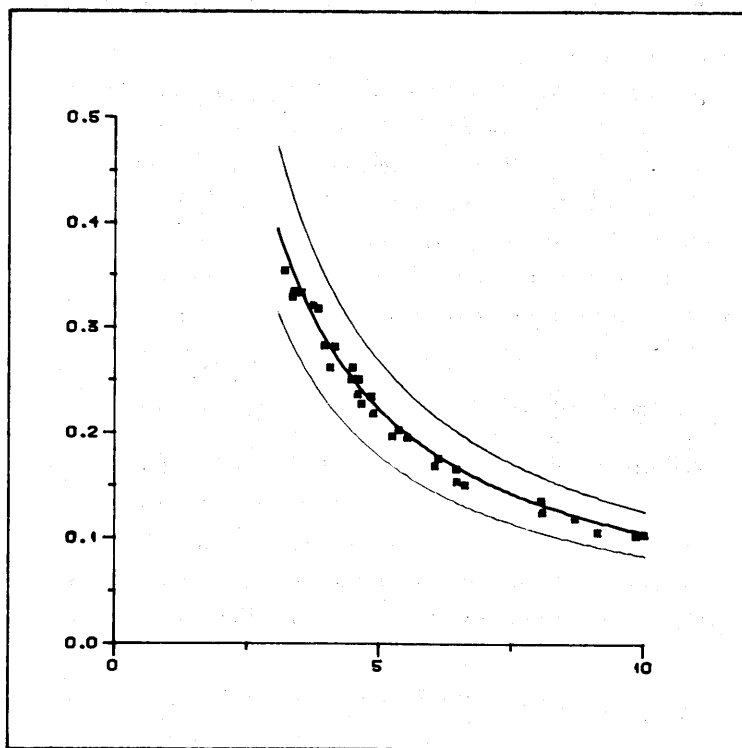


Fig. 7. Relation between the mean length of the infestation runs (abscissa) and the mean number of egg masses per plant (ordinate). The solid lines indicate the theoretical confidence interval under complete spatial randomness.

individuals are distributed on aligned discrete units. Its first motivation is to propose a more attractive sampling procedure to estimate ECB borer egg mass infestations. Run-based strategies do not require full counts of individuals in sample units and provide a way of testing complete spatial randomness.  $R'(n, n_0)$  which limit the sampling cost, still provides a good precision for low infestation rates. In fact, ECB egg masses in the Paris area show features which fully justify the use of  $R'(n, 20)$ :

1) Detection of egg masses is tedious because they are hidden on the underside of the maize leaves (at least, 2/3 of time gained simply by identifying infested plant instead of searching all egg masses on a plant);

2) The length of infestation run rarely exceeds 20;

3) Generally,  $\lambda < 0.5$  (to give a rough idea, the pest infestation threshold from an agronomical point of view is 0.12 in France) and the relation between infestation intensity and expected infestation run length is not significantly different from the complete spatial randomness case one.

On the other hand, our intention was to show that estimation of population density can be approached in a quite different way from methods based on full count on sampled units. Meanwhile, this must not involve excessive loss of information. This

technique is efficient when the relation between infestation intensity and expected infestation run length is well known. Average number of individuals per unit must be relatively small to preserve a certain reliability (Fig. 2, Fig. 4). Furthermore, infestation run observations can provide methods to test randomness from aligned discrete units (we find again the classical duality between estimate of intensity and test of randomness in spatial distribution of individuals (HINES and O'HARA, 1979; DIGGLE and COX, 1983). However, some a priori information is needed to choose value  $n_0$  for truncation. Cost reduction due to observation of infestation states instead of counting individuals in each unit can be an attractive feature of such methods.

#### SUMMARY

Spatial distributions of individuals may be considered in two ways: Firstly, individuals are situated on a continuum. Secondly, individuals are situated on discrete units, for example, insect eggs on plants. The aim of this paper is to show that sometimes, in the latter case, sampling methods based on infestation runs can be of interest in estimating population density. Analytical results are obtained under complete spatial randomness hypothesis and alternative hypotheses. Sampling procedures with limited cost are discussed and the European corn borer (ECB) case is mainly considered.

#### REFERENCES

- BADENHAUSSER, I. and J. VAILLANT (1987) Sequential sampling of spatially distributed populations. *C. R. Acad. Agric. Fr.* 73: 83-92.
- BARTLETT, M. S. (1978) *Stochastic Processes*. Cambridge Univ. Press.
- BYTH, K. (1982) On robust distance-based intensity estimators. *Biometrics* 38: 127-135.
- CASSEL, C. M., C. E. SARNDAL and J. H. WRETMAN (1973) *Foundations of inference in survey sampling*. Wiley, New-York.
- COCHRAN, W. G. (1977) *Sampling techniques*. Wiley, New-York.
- DENECHERE, M., S. DERRIDJ and C DUBY (1982) Etude d'une methode d'echantillonnage sequentiel appliquee a l'estimation du nombre de pontes de la pyrale du maïs. *Agronomie* 2: 341-346.
- DIGGLE, P. J. (1975) Robust density estimation using distance methods. *Biometrika* 62: 39-48.
- DIGGLE, P. J. (1977) A note on robust density estimation for spatial point patterns. *Biometrika* 64: 91-95.
- DIGGLE, P. J. (1983) *Statistical analysis of spatial point patterns*. Academic Press, London.
- DIGGLE, P. J. and T. F. COX (1983) Some distance-based tests of independence for sparsely-sampled multivariate spatial point patterns. *International statistical review* 51: 11-23.
- HINES, W. G. S and R. J. O'HARA (1979) The Eberhardt statistic and the detection of non randomness of spatial point distributions. *Biometrika* 66: 73-79.
- MILNE, A. (1959) The centric systematic area-sample treated as a random sample. *Biometrics* 15: 270-297.
- NACHMAN, G. (1981) A mathematical model of the functional relationship between density and sapatial distribution of a population. *J. Anim. Ecol.* 50: 453-460.

- PATIL, S. A., K. P. BURNHAM and J. L. KOVNER (1979) Non parametric estimation of plant density by the distance method. *Biometrics* 35: 597-604.
- PIELOU, E. C. (1962) Runs of one species with respect to another in transects through plant populations. *Biometrics* 18: 579-593.
- STENGEL, M. (1969) Influence de l'attaque de la Pyrale (*Ostrinia nubilalis*) sur le poids et la qualite des grains de Mais. *Rev. de Zool. Agri.* 68: 101-112.
- TAYLOR, L. R. (1984) Assessing and interpreting the spatial distributions of insects populations. *Ann. Rev. Entomol.* 29: 321-357.
- VAILLANT, J. (1985) Etude statistique des répartitiones spatiales et temporelles des pontes de Pyrale (*Ostrinia nubilalis*) dans le Bassin Parisien. Problemes d'échantillonnage. *Thèse 3e cycle. Toulouse.*
- VAILLANT, J. (1988) Spatial-temporal models involving correlated negative binomial distributions. *Tech. report Dep. Math. Sciences. Durham University.*

### 連を用いたサンプリング法によるアワノメイガ卵塊の密度推定

J. VALLANT and S. DERRIDJ

個体の空間分布を考える場合、個体が連続体の上に位置している場合と、個体が点々と散らばっている植物体上の昆虫の卵のように離散した住処に位置している場合のふたつに分けられる。この論文では後者の場合を取り上げ、その個体群の密度推定において、感染している植物の連を用いたサンプリング法が有効に利用できることを示した。個体が完全なランダム分布をしている場合と、そうでない場合について解析し、さらに使える費用が限定されている場合のサンプリング方法を示し、アワノメイガの例を提示した。

## APPENDIX

Let  $c_i$  be the cost due to total observation of plant  $i$ .

Let  $g_i$  be the gain due to partial observation of plant  $i$ .

Assume that  $E(c_i) = c$ ,  $E(g_i) = c\beta(\lambda) \forall i$ , with  $\beta(\cdot)$  (the gain proportion) which is an increasing concave function from  $\mathbb{R}^+$  to  $[0; 1]$ .

If we only consider cost  $C$  due to plant observation time, we obviously have:

$$C(S(n\alpha)) = n\alpha c \quad \text{and} \quad C(R(n)) = nE(cu_1 - g_1) = ncE(u_1) - nE(g_1)$$

then, under  $H_0$ ,

$$C(R(n)) = nc \left( \frac{1}{1 - e^{-\lambda}} - \beta(\lambda) \right)$$

and

$$C(R(n)) - C(S(n\alpha)) = nc \left( \frac{1}{1 - e^{-\lambda}} - \alpha - \beta(\lambda) \right)$$

This cost difference is null for  $\lambda = \lambda_2$  ( $\lambda_2$  depending on  $\alpha$  and  $\beta(\cdot)$ ), and negative for  $\lambda > \lambda_2$ . On the other hand, we have:

$$nH(\lambda_2, n, \alpha) = \left( \frac{n}{n+1} \right)^2 \frac{1}{(\alpha + \beta(\lambda_2) - 1)(\alpha + \beta(\lambda_2))} - \frac{1}{\alpha} \log \left( \frac{\alpha + \beta(\lambda_2)}{\alpha + \beta(\lambda_2) - 1} \right).$$

The sign of this latter expression provides a way of determining which of the two strategies is the more precise at equal expected cost. The closer  $\beta(\cdot)$  will be to unit, the more interesting will be  $R(n)$  in comparison with  $S(n\alpha)$ .

When  $\lambda_2 < \lambda_1$  ( $\lambda_1$  being the unique value of  $\lambda$  for which  $H(\lambda, \alpha, n) = 0$ ),  $R(n)$  is more precise and cheaper than  $S(n\alpha)$  for  $\lambda \in [\lambda_2; \lambda_1]$ . In practice, we can assume that  $\beta(\cdot)$  takes a constant value  $\hat{\beta}$  which belongs to  $[2/3, 5/6]$  and depends on the plants' development stage. Note that  $\lambda_2$  decreases when  $\alpha$  or  $\hat{\beta}$  increases and that  $\lambda_1$  increases when  $\alpha$  or  $n$  decrease.

Numerical examples: Under  $H_0$ ,  $(n, \alpha, \hat{\beta}) = (25, 5, 2/3)$  implies  $[\lambda_2; \lambda_1] = [0.19; 0.22]$

and  $n > 20$ ,  $\alpha \in [3; 10]$ ,  $\hat{\beta} \in [2/3, 5/6]$  implies  $0 < (\lambda_1 - \lambda_2) < 0.10$