Reprinted from
ENVIRONMENTAL ENTOMOLOGY, VOL. 12, NO. 1, FEBRUARY 1983

Clumping Patterns of Fruit and Arthropods in Cotton, with Implications for Binomial Sampling[1]

L. T. WILSON[2] AND P. M. ROOM[3]

ABSTRACT

Environ. Entomol. 12: 50–54 (1983)

A binomial model is presented which enables the clumping patterns of different species or categories of cotton arthropods and plant parts to be compared, accounting for the effect of their densities. Estimates of the proportion of infested sample units derived with this model are compared with those derived with three other binomial models. Statistical comparison is made. using as a criterion the degree to which each model fit field values of the proportion of infested sample units collected by three sampling methods (visual whole-plant examination. a bag method, and sweep-net). Those models which fail to incorporate the effect of density on clumping behavior fit the data less well. Estimates of sample sizes derived by a binomial sample size equation and a numerical sample size equation both of which incorporate species clumping behavior are also compared. The sample size estimates from the two equations are niost similar at low densities and for species whose distributions appear closest to random; and although binomial sampling requires a larger sample size at higher densities. less time is required to sample each unit (leaf, plant. etc.).

The majority of sampling plans for cotton management evaluate the need to control a pest by recording the number of individuals found per sample unit (enumerative sampling). An alternative is presence-absence (binomial) sampling, which saves time when organisms are abundant.

Ingram and Green (1972) used the 0th term of the negative binomial with a common K to describe the relationship between the proportion of infested plants [P(I)] and the mean number per plant of *Heliorhis armigera* (Hubner) and *Diparopsis castanea* Hmps. Sterling (1975) used a polynomial equation to relate P(I) to mean for 16 insect and plant part categories. Although both of these methods have proven useful, they have some noticeable limitations. Since the K value of the negative binomial is sensitive to population density and since the population densities of different species or categories differ in time, use of a common K value which is only optimal for a relatively low range of densities is not always sufficiently sensitive. The polynomial has limitations which include the lack of biological significance of its coefficients.

This study presents a binomial model which incorporates clumping pattern as a function of population density, and evaluates its fit to field data as compared with three other binomial models. The major purpose of this comparison was to determine whether incorporating the clumping behavior of a species significantly improves the estimation of P(I) values. In addition. binomial and enumerative sample size equations which incorporate clumping pattern are compared.

Materials and Methods

Data

The data were collected during the 1975176 through 1977/78 cotton growing seasons in the Naomi Valley, New South Wales, as part of research aimed at reducing unwarranted pesticide use while maintaining acceptable econoniic returns (Hearn and Room 1979, Room 1979). The sample methods used were: whole-plant visual sampling with 96 individual plants examined in the field every week day; bag sampling, which entailed removing 24 to 48 individual plants weekly (in plastic bags); and sweep sampling, which consisted of 10 to 12 sweep units sampled weekly. each unit being 30 sweeps down one side of a row of cotton and 30 sweeps back up the other side of the same row with a 45-cm-diameter net. During visual sampling, the surface of each plant was examined without opening fruit and leaf buds, in an attempt to mimic the approach normally taken by commercial scouts. Both the bag and sweep samples were exposed to ethyl acetate and returned to the laboratory, where the bagged plants were dissected and up to three cotton fruit and 54 arthropod categories were recorded. This approach provided data for comparing relative efficiencies of the different methods (Wilson and Room, 1983).

Binomial Models

***Model* 1.**—Derivation of model 1 is achieved starting with the negative binomial equation relating K to the estimated proportion of uninfested sample units [$\hat{P}(O)$] (Anscombe 1950).

$$\hat{P}(O) = [K\cdot(K + \bar{x})^{-1}]^K \quad (1)$$

where K is the negative binomial parameter and x is the sample mean. Let c = Kix and

$$\hat{P}(O) = [C\cdot(C + 1)^{-1}]^K \quad (2)$$

Now using the variance-mean relationship from the negative binomial

$$S^2 = \bar{x} + \bar{x}^2\cdot K^{-1} \quad (3)$$

Rearranging and substituting C

$$\bar{x}/S^2 = C\cdot(C + 1)^{-1} \quad (4)$$

Rearranging equation 3 and solving for K

$$K = \bar{x}\cdot(S^2\cdot x - 1)^{-1} \quad (5)$$

[1] Received for publication 3 September 1981
[2] Dept. of Entomology, University of California, Davis. CA 95616.
[3] CSIRO Division of Entomology, P.B. No. 3, Indooroopilly, Qld. Australia 4068

Substituting equations **4** and 5 into equation 2 results in

$$\hat{P}(0) = (\bar{x}\cdot S-2)\bar{x}\cdot(S^2\bar{x}^{-1}-1)^{-1} = e^{\bar{x}\cdot(S^2\bar{x}^{-1}-1)^{-1}\ln(\bar{x}\cdot S^{-2})} \tag{6}$$

where ln = log,. Rearranged and converted to proportion infested [P(I)], one has the following relationship:

$$\hat{P}(I) = 1 - e^{-\bar{x}[\ln(S^2\cdot\bar{x}^{-1})\cdot(S^2\cdot\bar{x}^{-1}-1)^{-1}]} \tag{7}$$

Equation 7 is interesting in that. as the ratio of $S^2\cdot\bar{x}^{-1}$ approaches unity,

$$\lim 1 - e^{-\bar{x}[\ln(S^2\cdot\bar{x}^{-1})\cdot(S^2\cdot\bar{x}^{-1}-1)^{-1}]} = 1 - e^{-\bar{x}}$$
$$\text{as } S^2\cdot\bar{x}^{-1} \to 1$$

which is equal to the 0th term of the Poisson distribution.

Model, 2.—Substitution of Taylor's variance-mean model ($S^2 = a\bar{x}^b$) (Taylor 1961, 1965, 1971) for the variance value used in model I (equation 7), increases its flexibility by providing appropriate variance values when the corresponding Taylor coefficients are known, giving the following:

$$\hat{P}(I) = 1 - e^{-\bar{x}[\ln(a\bar{x}^{b-1})\cdot(a\bar{x}^{b-1}-1)^{-1}]} \tag{8}$$

Taylor's coefficients are estimated by regressing $\ln(S^2)$ against $\ln(\bar{x})$.

$$\ln(S^2) = \ln(a) + b\,[\ln(\bar{x})]$$

Model 3.—This model is based on the commonly used Poisson distribution (random), where

$$\hat{P}(I) = 1 - e^{-\bar{x}} \tag{9}$$

Model 4.—This model has the following form:

$$P(I) = 1 - e^{-m'\bar{x}} \tag{10}$$

The coefficient "m'" was determined by using the following transformation:

$$c + m\bar{x} = -\ln[1 - \hat{P}(I)]$$

The "c" intercept was not significantly different from zero for any of the regressions, allowing each regression to be forced through the origin resulting in

$$m'\bar{x} = -\ln[1 - \hat{P}(I)] \tag{11}$$

where m' is the forced regression coefficient. Taking the antilog and then rearranging gives equation 10.

Analyses

The field data used in the analyses were the proportion of infested units, and the mean and variance for each of the three sampling methods for each day's data. Model 1 used all three data types [P(I), $\bar{x}$, and S^2], model 2 used P(I) (field sample value, not equation estimate) and $\bar{x}$, but estimated S^2 by using Taylor coefficients (Taylor 1961, 1965, 1971), model 3 used *P*(I) and $\bar{x}$ and did not estimate or use S^2, and model **4** used P(I) and $\bar{x}$ and again did not estimate or use S^2. All four models then produced $\hat{P}$(I) values which were regressed against the observed field values. The r^2 values resulting from these linear regressions were then compared in an analysis of variance (ANOVA) examining the effects of sample method. analysis method. and species categories, as well as the associated interactions.

Results and Discussion

Figure 1 illustrates the relationship between the proportion of infested plants and the mean per plant for 2 of the 26 species categories presented in this paper. Each point on the figure represents at least 96 visually inspected plants (1 day's data; some points overlap) and attests to the largeness of the data set. The less clumped distribution pattern for jassids is indicated by the higher proportion of infested plants for any density. This diffcrence illustrates that each species or category may have different clumping patterns.

Comparing Binomial Models

Twenty of the 57 species or categories examined during this study were sufficiently abundant (greater than **6** days' data with individuals present) for comparing the fit of the four binomial models. The extremely large set of data used in the analysis (7175 category × days data) prohibits detailed presentation of individual data sets.

All main effects of the **ANOVA** were significant ($P < 0.05$). as were both first-order interactions with sampling method as a factor. Table **1** presents means of the main effects (see Table 2 for complete species names). The bag method gave the best fit, followed by the visual and sweep methods.

Model I gave the best average r^2 value (0.99), followed by model *2* (0.92), model 3 (0.89). and model **4** (0.85) methods, The near-perfect fit of the $S^2/\bar{x}$ method attests to the close relationship between P(I), S^2 and $\bar{x}$ for the different species or categories examined. Model **4** has the disadvantage for some categories at low densities or producing P(I) estimates which are impossible [P(I) > $\bar{x}$]. The fit of model 3 (Poisson) was not suitable for most categories having clumped distributions.

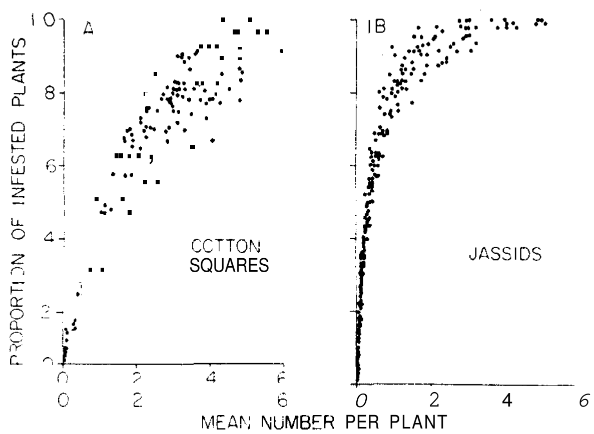

FIG. 1.—Proportion of infested plants [P(I)] as a function of density for (A) cotton squares and (B) jassids, *Austroasca viridigrisea*.

Table 1.—Means of the main effects and tests for significant differences using Student-Newman-Keul's multiple range test[a]

$\bar{r}^2$ Values for main effects			
Sample method		Analysis method	
Visual	0.90b	$S^2/\bar{x}$	0.99a
Bag	0.96a	$\hat{S}^2/\bar{x}$	0.92b
Sweep	0.89b	Poisson	0.89b
		b'	0.85c
Arthropod categories			
Heliothis spp.			
Very small larvae			0.92ab
Small larvae			0.95ab
Medium larvae			0.96ab
Large larvae			0.96ab
Earias heugeli			0.96ab
Austroasca viridigrisea			0.84ab
Oxycarenus luctuosis adults			0.83b
Geocoris lubra adults			0.95ab
Campylomma livida adults			0.90ab
C. livida nymphs			0.95ab
Verania frenata adults			0.91ab
Cocinella repanda adults			0.89ab
Diomus notescens adults			0.94ab
Luius bellulus adults			0.92ab
Oxyopes spp. adults			0.96ab
Salticidae spp. adults			0.86ab
Chiracanthium diversum adults			0.86ab
C. diversum nymphs			0.83b
Achaearanea veraculata adults			0.97a
A. veruculata nymphs			0.84ab

[a] Numbers followed by same letter differ at the 5% level, by Student-Newman-Keuls test statistic.

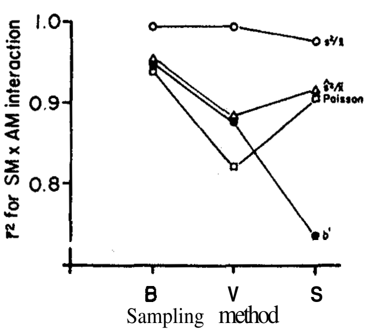

FIG. 2.—Sampling methods—analysis method interactions.

As mentioned, both first-order interactions involving sampling method were significant ($P < 0.05$). The fit of model 1 was little affected by sampling method, whereas at the other extreme model 4 fit progressively less the bag, the visual, and then the sweep data (Fig. 2).

Of the 20 categories examined in the SM × CA interaction, 9 had average best fits in the order V-B-S, 6 in the order B-V-S, 4 in the order B-S-V, and 1 V-S-B. Attempts at correlation with age class and organism type (Araneida, Lepidoptera, etc.) lead to no biological interpretation of this interaction.

Clumping Pattern

Sufficient data were available for comparing the effect of age on the clumping pattern for cotton fruit and for *Heliothis* spp. The results confirm previous findings in showing that as most populations age they become progressively less clumped (Salt and Hollick 1946, Guppy and Harcourt 1970). This decreased clumping is reflected in smaller "a" and "b" coefficients for Taylor's equation (Table 2), as well as by an increase in the proportion of plants infested for a given mean. Figure 3, produced by using model 2, illustrates the decrease in clumping with age for *Heliothis* spp. The close similarity of the three larval curves as compared with the egg curve implies that mortality during the later part of the egg stage or early larval stage, or dispersal during the early larval stage is largely responsible for the observed difference. Room (1979) has shown that most *Heliothis* mortality occurs during the later part of egg development and the early part of larval development. Wilson et al. (1980) further showed that 1st-instar *Heliothis* disperse very little, and it is therefore unlikely that dispersion to adjoining plants would be occurring at this stage of development.

Sample Size

The clumping behavior of an organism affects the number of samples required to estimate the population density with a given level of reliability. Karandinos (1976) presented a set of equations for use in estimating sample size when the underlying distribution was of several types. These can be simplified to:

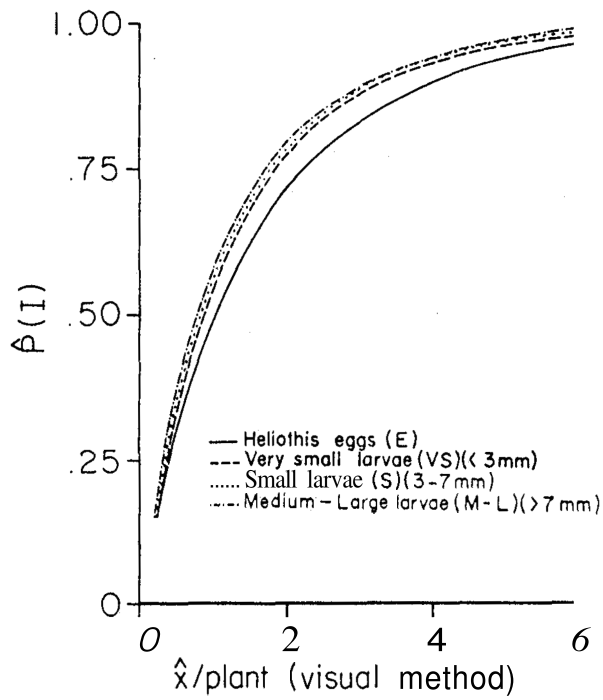

FIG 3.—Proportion of infested plants [P(I)] as a function of density for different *Heliothis* spp. age classes.

Table 2.—Taylor's coefficients for cotton data categories, Naomi Valley, N.S.W.

	Sampling method											
	Bag				Visual				Sweep			
Category	a	b	r^2	n^a	a	b	r^2	n	a	b	r^2	n
Cotton fruit												
Squares					2.79	1.19	0.94	144				
Bolls					2.54	1.22	0.97	116				
Open bolls					2.50	1.18	0.96	36				
Lepidoptera: Noctuidae												
Heliothis armigera (Hübner)												
H. punctigera Wallengren												
White eggs					1.80	1.13	0.94	408				
Brown eggs					1.82	1.13	0.95	385				
Total eggs					1.86	1.14	0.96	438				
Very small larvae (<3 mm)	1.43	1.12	0.90	29	1.22	1.05	0.96	385	1.29	1.12	0.90	38
Small larvae (3-7 mm)	1.05	1.01	0.95	30	1.17	1.04	0.98	369	1.17	1.08	0.91	78
Medium-sized larvae (7-19 mm)	0.84	0.95	0.98	22	1.16	1.03	0.98	260	1.04	1.09	0.93	73
Large larvae (>19 mm)	1.00	0.98	0.92	11	1.14	1.03	0.98	210	1.07	1.03	0.91	61
Earias huegeli Rogenhofer larvae	0.94	0.98	0.99	9	1.30	1.06	0.99	215	1.19	1.10	0.96	32
Hemiptera: Cicadellidae												
Austroasca viridigrisea (Paoli)	1.61	1.17	0.97	42	1.96	1.21	0.97	474	1.65	1.32	0.87	64
Hemiptera: Lygaeidae												
Oxycarenus luctuosis M.&S. adults	4.62	1.43	0.95	37	6.37	1.40	0.93	452	1.57	1.33	0.90	81
Geocoris lubra (Kirkaldy) adults	1.60	1.12	0.96	13	1.21	1.04	0.95	150	1.06	1.02	0.92	73
Hemiptera: Miridae												
Campylomma livida Reuter adults	1.52	1.11	0.93	12	1.61	1.11	0.94	270	1.31	1.14	0.90	69
C. livida nymphs	0.89	0.97	1.00	11	1.33	1.06	0.97	257	1.35	1.12	0.84	17
Coleoptera: Coccinellidae												
Verania frenata Er. adults	1.87	1.18	0.93	7	1.15	1.03	0.98	195	1.31	1.26	0.94	68
Coccinella repanda Er. adults	1.14	1.01	0.81	23	1.14	1.04	0.97	436	1.17	1.20	0.87	79
Diomus notescens Blackburn adults	0.86	0.96	1.00	17	1.49	1.09	0.91	204	1.13	1.11	0.92	70
Coleoptera: Melyridae												
Laius bellulus (Guerin) adults	0.99	1.00	0.98	20	1.33	1.07	0.94	384	1.14	1.32	0.85	80
Araneida: Oxyopidae												
Oxyopes spp. adults	0.93	0.98	1.00	8	1.29	1.06	0.98	94	0.99	1.00	0.90	75
Araneida: Salticidae												
Salticidae spp. adults	0.86	0.93	0.84	11	1.44	1.08	0.93	119	1.27	1.13	0.88	66
Araneida: Clubionidae												
Chiracanthium diversum (Koch) adults	3.30	1.39	0.87	25	1.14	1.03	0.98	263	1.18	1.08	0.88	72
C. diversum immatures	0.87	0.97	1.00	12	2.05	1.18	0.89	282	0.95	0.96	0.88	60
Araneida: Theridiidae												
Achaearanea veruculata (Urquhart) adults	0.80	0.94	0.99	19	1.03	1.01	0.99	210	1.51	1.17	0.96	31
A. veruculata immatures	0.88	0.97	1.00	14	1.65	1.13	0.89	339	1.08	1.04	0.90	48

[a]Regression for categories with at least 7 days' data.

$n = C\,S^2/\bar{x}^2$ general distribution

$n = C\,(1/\bar{x} + 1/K)$ negative binomial

$n = C/\bar{x}$ Poisson

$n = C\,q/p$ binomial

where

$$C = (Z_{\alpha/2}/D)^2$$

with $(1 - \alpha)$ the confidence coefficient, $Z_{\alpha/2}$ the upper $\alpha/2$ part of the standard normal distribution, $p + q = 1$, p in this case equals P(I), and D being a fixed proportion of the mean. Wilson and Room (1983) converted the sample size equation for the general distribution, the negative binomial, and the Poisson into the form:

$$n = C\,a\,\bar{x}^{b-2} \tag{12}$$

This equation incorporates Taylor's equation and has the advantage of allowing estimates of sample sizes to be made for a range of densities for any species whose coefficients are known. The Poisson is a special case of equation 12 and occurs when a = b = 1. The negative binomial, also a special case, fits a range of data for which the variance exceeds the mean which occurs for nearly all species described in this paper (see Table 2 for calculating variances).

Figure 4A presents the P(I) – $\bar{x}$ curves derived by using model 2 for a Poisson distribution variate, a structure which is slightly clumped (cotton squares), and an organism which is very clumped (*Oxycarenus* adults) for the visual sampling method. The more clumped the species. smaller is P(I) for a given mean. Of interest, none of the species examined in this study were Poisson (randomly) distributed: although some sampling methods, such as the sweep-net, may mask the true distribution. These findings are in agreement with Taylor et al. (1978), who found that organisms are rarely distributed randomly.

Figure 4B shows the sample size required to estimate the mean within 10% by using the numerical and the binomial methods. For all three categories at low densities, the sample size estimates from the binomial and numerical equations are quite similar. For *Oxycarenus*, the most clumped of these three, a considerably larger sample size is required at higher densities, due to the relatively slow increase in P(1) as 100% infestation of the plants is approached. Sample size for the numerical method, however, decreases asymptomatically with increase in density. The rather large difference in sample size estimates for *Oxycarenus*, cotton squares, and a Poisson-distributed variate attests to the importance of considering species clumping patterns when developing sampling programs. The difference in the fit of the four binomial models to field data further confirms this point.

Acknowledgment

We thank D. González, P. M. Ives, R. B. Nowierski, and C. G. Summers for reading the manuscript, and unknown reviewers for their helpful criticisms of earlier drafts of the paper.

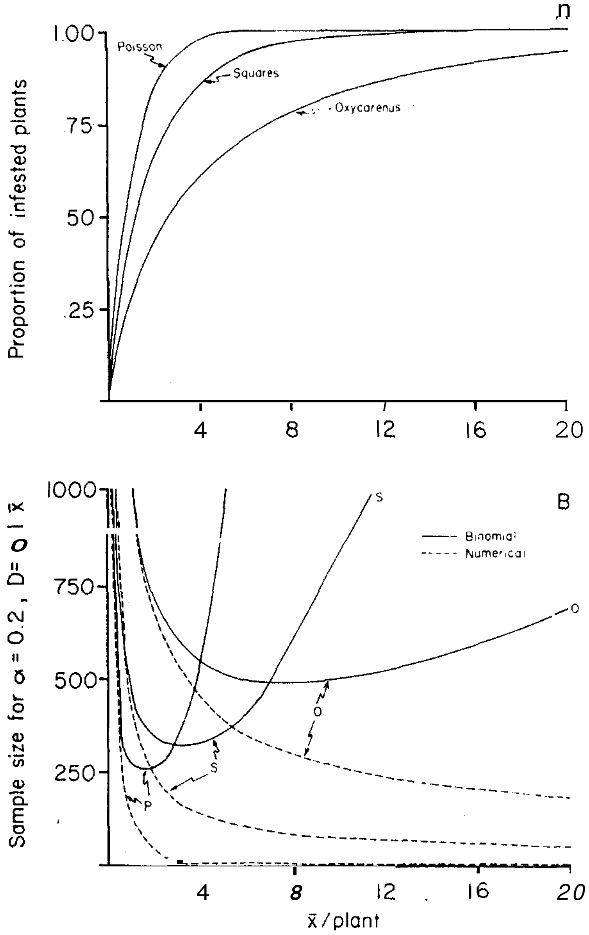

FIG. 4.—(A) Proportion of infested plants [P(1)] and (B) sample size as a function of density for a Poisson-distributed variate, for cotton squares, and for *Oxycarenus* adults.

REFERENCES CITED

Anscombe, F. J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika 37: 358–382.

Guppy, J. C., and D. G. Harcourt. 1970. Spatial pattern of the immature stages and teneral adults of *Phyllophaga* spp. (Coleoptera: Scarabaeidae) in a pernianent meadow. Can. Entomol. 102: 1254–1259.

Hearn, A. B., and P. M. Room. 1979. Sequential analysis of crop development for pest management in cotton. Prot. Ecol. 1: 265–277.

Ingram, W. R., and S. hl. Green. 1972. Sequential sampling for bollworms on rain grown cotton in Botswana. Cotton Grow. Rev. 40: 265–275.

Karandinos, M. G. 1976. Optimal sample size and comments on some published formulae. Bull. Entomol. Soc. Am. 22; 417–421.

Room, P. M. 1979. A prototype on-line pest management system for cotton in the Naomi Valley, New South Wales. Prot. Ecol. 1: 245–264.

Salt, G., and F. S. Hollick. 1946. Studies of wireworm populations. II. Spatial distribution. J. Exp. Biol. 23: 1–46.

Sterling, W. 1975. Sequential sampling of cotton insect populations. Proc. Beltwide Cotton Prod. Conf. 19 pp.

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature (London) 189: 732–735.

1965. A natural law for the spatial disposition of insects. Proc. XIIth Int. Congr. Entoniol., 1964: 396–397.

1971. Aggregation as a species characteristic. *In* G. P. Patil, E. C. Pielou, and W. E. Waters [eds.]. Statistical ecology. Vol. 1. pp. 357–377.

Taylor, L. R., I. P. Woiwod, and J. N. Perry. 1978. The density-dependence of spatial behavior and the rarity of randomness. J. Anim. Ecol. 47: 383–406.

Wilson, L. T., A. P. Gutierrez, and T. F. Leigh. 1980. Within-plant distribution of the immatures of *Heliothis zea* (Boddie) on cotton. Hilgardia 48: 12–23.

Wilson, L. T., and P. hl. Room. 1982. The relative efficiency and reliability of three methods for sampling arthropods in Australian cotton fields. J. Aust. Entoniol. Soc. (in press).