Sampling Whiteflies in Cotton: Validation and Analysis of Enumerative and Binomial Plans

STEVEN E. NARANJO,[1] JONATHAN W. DIEHL,[2] AND PETER C. ELLSWORTH[2]

ABSTRACT We tested enumerative and binomial sampling plans developed for *Bemisia tabaci* (Gennadius) in 3,240 ha of commercial cotton as part of the implementation of a community-wide integrated pest management (IPM) program in Laveen and Tolleson, AZ, in 1994. We compared new field observations to sampling distribution models developed previously for all lifestages, and validated and analyzed the performance of 5 sampling plans based on these models by resampling field data from 129 to 284 sites. Mean–variance relationships for the new data differed statistically from mean–variance models previously developed for adults, but not for eggs or nymphs. Resampling analyses indicated that desired precision (SE to mean ratio) was rarely achieved, on average, by fixed-precision sequential sampling plans. These enumerative sampling plans provided better precision than desired at moderate to high densities of eggs and adults and worse precision than desired at most densities of nymphs. An empirical model relating mean density to the proportion of leaves infested with 3 or more adult *B. tabaci* was accurate at mean densities < 2 adults per leaf but over-predicted mean density at higher densities. Resampling analysis revealed that a sequential sampling plan based on this empirical model was accurate at classifying population density relative to an action threshold of 5 adults per leaf. At nominal α and β error rates of 0.10, population density was correctly classified ≈87% of the time. Accuracy was not improved by reducing nominal error rates to 0.05. Resampling analysis of a fixed-sample size plan based on $n = 30$ gave similar results and increasing sample size to 50 increased accuracy only 3%. Further resampling analyses that more closely approximated scouting protocols (15 sample units drawn from each of 2 quadrants in the field) resulted in an average accuracy of ≈70%. Accuracy declined when populations densities differed greatly among quadrants in a field. Most of this error was associated with making a decision to control when pest density was below the action threshold. Based on a robust validation technique using field observations representing a wide range of environmental and agronomic conditions, our sampling plans performed well and should be useful for estimating and classifying population densities of *B. tabaci* in cotton over a wide area.

KEY WORDS *Bemisia tabaci*, *Bemisia argentifolii*, sequential sampling, binomial sampling, resampling, sampling plan validation

SAMPLING IS A fundamental component of any pest management program based on the prescriptive control of pest populations and is an essential element of many field research programs. Regardless of the intended application, any sampling plan should ideally be developed from observations that encompass the geographic area and range of environmental and agronomic conditions that users of the sampling plan will likely encounter. In reality, however, sampling plans are often developed from a fairly restricted range of observations but are then used in situations representing a new array of environmental and agronomic conditions. Thus, it is important that the performance of a sampling plan in novel field situations be examined so that its limitations and strengths can be better defined. These performance qualities are of crucial importance for sampling plans developed for pest management application where an incorrect decision regarding pest suppression could result in costly yield loss or unnecessary environmental disruption.

The B-strain of the sweetpotato whitefly, *Bemisia tabaci* (Gennadius) (=*B. argentifolii* Bellows & Perring), has emerged as one of the most significant pests of agricultural production in the southwestern United States (USDA 1995). Since the early 1990s, this insect has assumed key pest status in several crops in Arizona, California, and Texas, including cotton, *Gossypium hirsutum* L., and cucurbits, *Cururbita* spp. (Butler and Henneberry 1994, Brown et al. 1995). The multivoltine life cy-

cle, high reproductive potential, broad host range, and intercrop movement of *B. tabaci* contributes to its pest potential and complicates the development of pest management strategies for any one commodity. Because *B. tabaci* may disperse readily among adjacent fields (Watson et al. 1992, Byrne et al. 1996), it is likely that sustainable and environmentally sound pest management systems will only be realized through coordinated efforts among agricultural producers over a wide area (Ellsworth et al. 1996).

Since 1993 we have been involved in the development of a community-wide IPM program with cotton producers in the Laveen-Tolleson agricultural community west of Phoenix, AZ. One of the primary goals of this program is to establish avenues for the cooperative management of *B. tabaci* on a community-wide scale. A critical component of this effort is the development and implementation of an efficient sampling protocol that can be used for monitoring pest presence and for aiding decision makers in determining the need for pest suppression.

We have developed several enumerative and binomial sequential sampling plans for estimating the abundance of all life stages of *B. tabaci* in cotton (Naranjo and Flint 1994, 1995) and for classifying the density of *B. tabaci* adults relative to an action threshold (Naranjo et al. 1996a). All of these sampling plans have undergone limited validation with research plot data to establish rough boundaries of utility, but more extensive validation is needed to gauge their performance over the wider set of conditions under which they may be used.

In 1994 we initiated a multifaceted effort to demonstrate and implement a binomial sampling method for adult whitefly within the Laveen-Tolleson agricultural community to allow growers to make more rational pest control decisions (Ellsworth et al. 1996). A critical component of this effort was the rigorous validation and analysis of a binomial sampling plan, which was based on rather limited data gathered in research plots in Phoenix and Maricopa, AZ, during 1993 (Naranjo et al. 1996a). It was essential to examine the accuracy and efficiency of the sampling plan and define its limitations as a tool for pest management decision making. This large-scale effort also afforded us an ideal opportunity to test and evaluate our more research-oriented, fixed-precision sampling plans.

To evaluate the performance of these sampling plans, we used a resampling approach in which random samples are repeatedly drawn from actual field data sets (Hutchison et al. 1988, Hutchison 1994, Naranjo and Hutchison 1997). The major advantage of this technique is that field data and not a theoretical model determines the underlying distribution of counts. This approach permits simultaneous evaluation of sampling error and deviations from the assumed sampling distribution. Given the large geographical area and wide range of pest densities encompassed in the implemen-

tation project, we were confident that analyses would provide rigorous tests of all sampling plans.

Materials and Methods

Sampling Plans Tested. We tested a total of 5 sampling plans for *B. tabaci*, including 3 based on enumerative counts and 2 based on binomial count data. Fixed-precision sequential sampling plans based on enumerative counts have been developed for eggs, nymphs, and adults (Naranjo and Flint 1994, 1995), using Taylor's (1961) power law and Green's (1970) algorithm. Adults are counted on the underside of leaves from the 5th mainstem node from the terminal; eggs and nymphs are counted on 3.88-cm^2 disks from leaves at this same nodal position. The general sequential stop line is given as

$$T_n \geq (an^{1-b}/D^2)^{1/(2-b)}, \quad (1)$$

where T_n is the cumulative count from n sample units, D is precision (SE to sample mean ratio), and a and b are parameters from the Taylor power law, $s^2 = am^b$, which models the sample variance (s^2) as a function of the sample mean (m). The parameters a and b were estimated to be 2.986 and 1.766 for eggs, 2.537 and 1.688 for nymphs and 2.079 and 1.675 for adults, respectively (Naranjo and Flint 1994, 1995).

A binomial sampling model was developed to estimate the mean density of adult *B. tabaci* on the underside of 5th mainstem node cotton leaves from the proportion of infested leaves (Naranjo et al. 1996a). We used an empirical model (Kono and Sugino 1958, Gerrard and Chiang 1970) given as

$$\mathrm{Ln}(\hat{m}) = \gamma + \delta\mathrm{Ln}[-\mathrm{Ln}(1-P_3)] \quad (2)$$

to define the relationship between mean density ($\hat{m}$) per leaf and proportion of leaves infested with at least 3 adults (P_3). Here, γ and δ are coefficients estimated by linear least-squares regression. The values of γ and δ were 1.777 and 0.918, respectively (Naranjo et al. 1996a). This mean density proportion-infested model formed the basis of 2 sampling plans for classifying population density relative to an action threshold of 5 adults per leaf (Ellsworth and Meade 1994, Naranjo et al. 1996b).

A sequential sampling plan was developed using Wald's (1947) sequential probability ratio test. For binomial count data the general upper and lower sampling stop lines are defined as

$$T_U(n) = Bn + A$$
$$T_L(n) = Bn - C, \quad (3)$$

where n is the number of sample units examined, $T(n)$ is the cumulative number of sample units infested with at least t insects, and A, B and C are parameters derived as standard functions of specified α and β error rates; and lower and upper boundaries bracketing the threshold density, given in terms of proportion infested (Wald 1947, Fowler and Lynch 1987). For this sampling plan we set

Table 1. Summary of field data sets and sampling plants for *B. tabaci* tested in Laveen-Tolleson, AZ, 1994

Sampling plan	No. data sets	Sample size range	Mean density range	Sampling plan reference
Eggs (per leaf disk)				
Green fixed-precision	129	96–100	0.16–29.19	Naranjo and Flint 1994
Nymphs (per leaf disk)				
Green fixed-precision	129	96–100	0.08–15.83	Naranjo and Flint 1994
Adults (per whole leaf)				
Green fixed-precision	284	50–100[a]	0.01–39.91	Naranjo and Flint 1995
Wald binomial SPRT	284	50–100	0.01–39.91	Naranjo et al. 1996a
Fixed-sample-size binomial	284	50–100	0.01–39.91	Naranjo et al. 1996a

SPRT, sequential probability ratio test.

[a] Only 2 quadrants (25 sample units per quadrant) were sampled on several occasions.

the lower (L) and upper (U) boundaries at proportional infestation values associated with 3 and 7 adults per leaf, respectively, and set $\alpha = \beta = 0.10$.

A fixed-sample size plan also was formulated to perform a similar population classification based on 30 binomial counts (Naranjo et al. 1996a). This latter model is our current recommendation for sampling *B. tabaci* in cotton in Arizona for decision making (Ellsworth et al. 1995)

Field Sites. Sampling data were collected in the Laveen-Tolleson area of Maricopa County, Arizona. This agricultural community is typical of central Arizona and encompasses ≈12,000 ha of crops. Cotton is the dominant crop, occupying nearly half (5,700 ha) of this area, with the remaining area planted to alfalfa, cantaloupes, watermelon, cole crops, lettuce, silage corn, sorghum, and citrus. Ellsworth et al. (1996) provides details on the community-wide pest management program. Whitefly samples were collected in 3,240 ha of cotton from this area. This included 190 fields belonging to a total of 14 individual growers. Field size averaged 17 ha. Growers were provided information on insect densities in their field within 24 h of sampling but were not obligated to treat at the recommended threshold of 5 adults per leaf.

Sampling Protocols. A standard set of protocols was used by all scouts to monitor populations of adult *B. tabaci* (Ellsworth et al. 1995). Briefly, this consisted of counting the number of adults on the underside of cotton leaves at the 5th nodal position below the mainstem terminal. Because one of our primary goals was to test the validity of this method for classifying population density relative to an action threshold, scouts counted all adults rather than using a binomial count which only tracks the number of leaves containing ≥3 adults. After an initial training period, scouts took samples weekly from each field. In most of the fields sampling was terminated after the 1st insecticides were applied for whitefly control. The number of weeks that a given field was sampled depended on when the grower first joined the program and when the first decision to spray insecticides was made. The

1st samples were taken as early as 22 May and as late as 4 July. In approximately half of the fields, adults were counted on 15 leaves from each of 2 quadrants per field as recommended for decision making (Ellsworth et al. 1995). In the remaining fields, adults were counted on 25 leaves from each of 4 quadrants for a total of 100 sample units per field. This larger sample size was taken to generate a robust data set for sampling plan validation and analysis (described below). Additionally, in 30 of these latter fields, scouts also collected 5th main-stem node leaves to estimate the population density of immature *B. tabaci*. Eggs and nymphs (all stages) were counted under a dissecting microscope on 3.88-cm^2 leaf disks, extracted near the petiole (Naranjo and Flint 1994). Samples for immatures were collected every other week for a given field from 1 June through 19 August. The total number of sample dates varied from 3 to 6 per field. Sampling for adults (n = 100/wk) also continued in these 30 fields even after the 1st insecticide sprays were applied.

Validation and Analyses of Sampling Plans. A subset of the sample data sets described above was selected for validation and further analyses (Table 1). In total, 1,219 individual field/date data sets were collected for adult *B. tabaci*. Of these, 601 consisted of fields in which 100 leaves were examined for adults, and of these, 571 had mean densities >0 adults per leaf. From the 571 data sets we selected 284 that had unique mean density values ranging from 0.01 to 39.91 adults per leaf. When means of 2 or more data sets were equal we selected 1 at random without regard to sample variance. Most duplicate densities occurred below 1 adult per leaf. In total, 139 individual field/date data sets were collected for immature *B. tabaci*. Of these, 129 data sets had unique mean density values >0 per leaf disk and contained 96–100 sample units each (Table 1).

We first examined the accuracy with which our Taylor power law models predicted the sample variance from the sample means of the data sets described above for all life stages of *B. tabact*. We constructed 95% CI for individual predicted values

about the regression lines to denote the variability of the original models. Further, we used analysis of covariance (ANCOVA) (SAS Institute1989) to compare regression parameters between the original model and the independent data sets. We also examined the accuracy of the empirical relationship which predicts the mean density of adult *B. tabaci* per leaf from the proportion of leaves infested with ≥3 insects. A 95% CI for individual predicted values was constructed about this relationship using the variance term suggested by Schaalje et al. (1991). This variance estimator attempts to account for both biological and sampling error. Again, ANCOVA was used to compare regression parameters (see equation 2) between the original model and a model developed from independent data sets.

To evaluate the overall performance of our sampling plans in the field we used an approach in which actual field data are resampled numerous times on a computer (Hutchison et al. 1988, Naranjo et al. 1995, Naranjo and Hutchison 1997). Specialized public-domain software has been developed to efficiently perform such analyses for several common sampling plans (Naranjo and Hutchison 1997). Basically, after entering sampling plan parameters, the software randomly selects successive observations from a field data set, without replacement, until the sequential stop line criteria are met or, in the case of the fixed-sample-size plan, the prescribed number of sample units have been drawn.

We tested the performance of fixed-precision sequential plans for all life stages by resampling each of the 129 data sets for eggs and nymphs, and each of the 284 data sets for adults 500 times. We tested our sampling plans using a representative precision (SE to mean ratio) of 0.25. The relative performance of our sampling plans would be the same regardless of the level of precision chosen. The minimum sample size was set at 10. From these resampled data we then calculated the mean precision and required sample size for each data set and the associated variances and extreme values of these measures. We then evaluated how closely sample sizes prescribed by the sequential plans compared with those estimated directly from the sample mean and variance of each independent data set using the formula $n = (s/Dm)^2$.

For the sequential binomial plan we calculated the mean, variance, and extremes of sample size and estimated the operating characteristic function. Again, each data set was resampled 500 times and the minimum sample size was set to 10. The operating characteristic is a measure of the accuracy with which the population is properly classified relative to the action threshold and is given as the probability of not intervening as a function of true mean pest density. After the sequential decision rule is satisfied (equation 3) this probability was estimated directly from each data set as the proportion of resampling iterations in which the

proportion infested did not exceed the lower sequential stop line. To provide a smooth operating characteristic function from the resampled data we fitted a 4 parameter logistic model to the output:

$$OC(x) = d + (a-d)/(1+[x/c]^b), \qquad (4)$$

where a, b, c, and d are parameters and x is mean density estimated from field data sets ($n = 100$). The Marquardt-Levenberg method was used to fit the model using SigmaPlot (Jandel 1994). We tested the original sampling plan as defined above and then evaluated whether changes in α and β error rates would improve plan performance. For comparison, nominal operating characteristic functions were estimated using the computer program of Nyrop and Binns (1991). Additionally, we estimated the overall probability of making type I (taking action when none is needed) and type II (failing to take action when needed) errors based on all 284 data sets. These error rates were estimated by classifying the outcome of each resampling iteration for each data set according to; mean ≤ 5 and $P_3 \leq 0.57$ (correct decision), mean ≤ 5 and $P_3 > 0.57$ (type I error), mean > 5 and $P_3 > 0.57$ (correct decision), and mean > 5 and $P_3 \leq 0.57$ (type II error), where $P_3 = 0.57$ is the proportion of leaves expected to be infested when the mean density is 5 adults per leaf (equation 2). Because sampling was discontinued in most fields after the 1st insecticide application for *B. tabaci* these estimates of accuracy are based on field observation that are skewed toward low densities. Nonetheless, these densities are representative of those that would be encountered by users of the sampling plans for decision making.

Next, we examined the performance of the fixed-sample-size binomial plan. Here, we estimated the operating characteristic function directly from the proportion of resampling iterations in which the proportion infested did not exceed the action threshold. We examined and compared sampling plan performance using fixed sample sizes of 10, 30, and 50 leaves per field. Type I and type II error rates were estimated as detailed above.

All of the above analyses were based on drawing sample units from the field as a whole and did not account for differences in pest distribution between quadrants. To examine this spatial component in more detail and to mimic the actual scouting protocols more accurately, we conducted a final set of analyses for the fixed-sample-size plan. First, counts from each of the 284 original data sets were sorted by quadrant. Resampling analysis was then conducted by randomly drawing 15 sample units from each of 2 quadrants per field for each possible combination of 2-quadrant sets (6 total). Operating characteristic functions were estimated for the overall average of all 2-quadrant combinations, and for the upper and lower extremes based on individual 2-quadrant sets for each field data set. All these operating characteristics were estimated as a function of mean density averaged over all 4
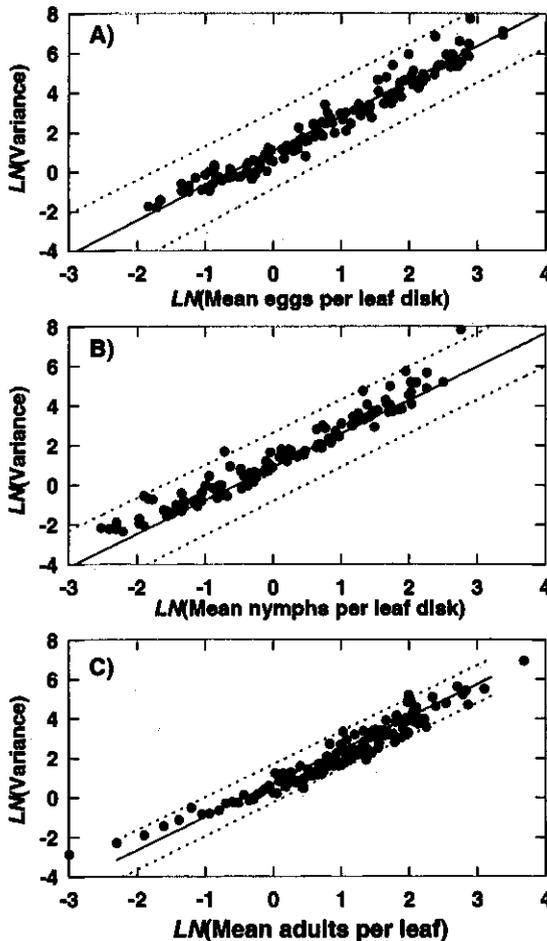
Fig. 1. Comparison of the relationship between sample mean and variance from independent data collected in Laveen-Tolleson, AZ, in 1994 (points) with Taylor power law models (solid line) developed for (A) eggs, (B) nymphs, and (C) adults of *B. tabaci*. Dotted lines denote 95% prediction intervals.

quadrants. Nominal α and β error were set to 0.10. Again, type I and type II error rates were estimated as detailed above.

Results

Performance of Fixed-Precision Sequential Sampling Plans. For *B. tabaci* eggs all of the field observations fell within the 95% prediction interval (Fig. 1A) and ANCOVA indicated no significant differences between either the slope ($P = 0.57$) or intercept ($P = 0.89$) of the original power law model and those estimated from field observations in 1994. Likewise for nymphs, most of the field observations fell within the 95% prediction interval (Fig. 1B) and there were no significant differences in slope ($P = 0.13$) or intercept ($P = 0.23$) parameters between the original power law model and those estimated from field observations. Despite

the lack of statistical differences between independent data sets and original power law models, the consequences of even small deviations from the sampling model can be clearly seen from our resampling analyses (Fig. 2 A and C). The actual sampling precision seldom equaled that specified, and the variance from one sample bout to the next was considerable for egg and nymphs. In general, average precision was better than expected (0.25) for eggs at densities between ≈0.5 and 10 per leaf disk, but worse at higher or lower densities (Fig. 2A). Thus, the sequential plan specified too small of a sample size at high and low densities and to large of a sample size at moderate densities. In comparison, the nymphal sampling plan prescribed too few sample units. The average precision was generally worse than expected when estimating the density of nymphs over all mean densities tested (Fig. 2C; most symbols fall above the horizontal line at $D = 0.25$).

Most field observations for adults fell within the 95% prediction interval (Fig. 1C); however, the Taylor power law generally overpredicted variances, except at very low densities where variance was underpredicted. Here, the slope ($F = 11.82$; df = 1, 293; $P = 0.0007$) and intercept ($F = 12.52$; df = 1, 293; $P = 0.0005$) terms of the power law for the independent field observations significantly deviated from those of the original model. The consequences of this deviation from the underlying sampling model, can again be clearly seen from our resampling analyses (Fig. 2E). For estimating adult density, the sampling plan consistently provided poorer than expected average precision at densities less than ≈0.5 per leaf, but better than expected precision at mean densities > 0.5. The stochastic nature of fixed-precision sequential sampling is evident for all life stages of *B. tabaci*. Even when the average value of precision approached that specified (0.25) there was considerable variability. It was not unusual in these instances for maximum and minimum values of precision when resampling from a data set to be > 0.40 and < 0.15, regardless of the life stage sampled. In contrast, there was little variability in sample size between sampling bouts, and deviations from the expected sample size were slight (Fig. 2 B, D, and F).

Our analyses indicated that the desired precision was not being consistently achieved by our sampling plans. To evaluate the consequences on sampling effort we compared sample sizes prescribed by the sequential plans with the sample size that might be estimated directly from a preliminary sample of a given field using the formula $n = (s/Dm)^2$. Based on results above we delineated 5 ranges of density varying from 0.1 to 0.5 to >10 insects per sample unit and averaged prescribed and direct estimates of sample size over these density classes (Table 2). As observed in Fig. 2, all sampling plans produced mean estimates with poorer than desired precision at mean densities
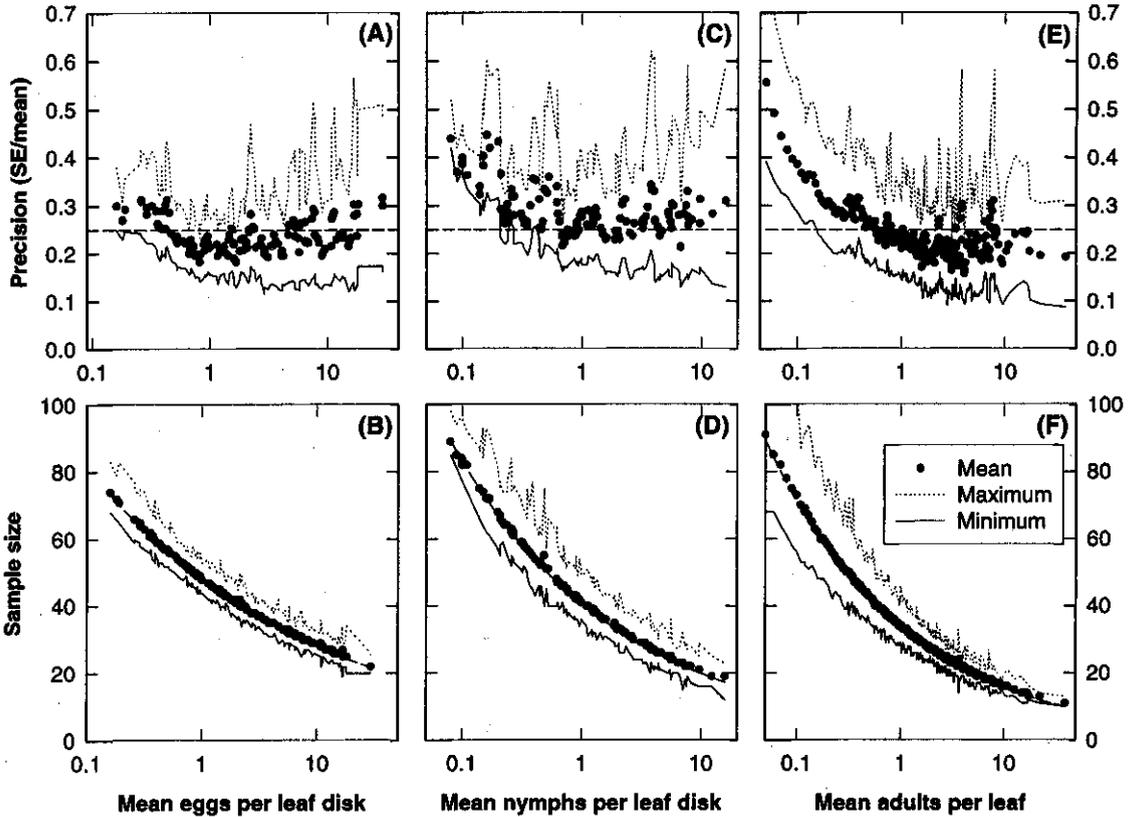
Fig. 2. Actual levels of precision and required sample sizes resulting from using fixed-precision sequential sampling plans developed for (A and B) eggs, (C and D) nymphs, and (E and F) adults of *B. tabaci*. Results were derived from resampling field data sets collected in Laveen-Tolleson, AZ, in 1994. The dashed lines in A, C, and E denote the desired precision specified for the sequential stop lines and those in B, D, and F denote the expected sample size.

<0.5 insects per sample unit. The sequential plan prescribed sample sizes from 33.4 to 73.4% too small for all lifestages at these low densities. At densities >0.5 the sampling plans for eggs and adults generally prescribed sample sizes larger than necessary. However, the largest discrepancy averaged 23.4%, and considering all densities combined the plans achieved precisions close to 0.25 and prescribed sample sizes only 4.1–7.6% larger than required. In contrast, the sequential sampling plan for nymphs consistently prescribed smaller than required sample sizes. Considering all densities combined precision averaged 0.287 and 37.3% fewer samples than required were prescribed. Further resampling analyses indicated that a precision near 0.25 could be achieved for all nymphal densities by specifying a precision D of 0.215 for the sequential stop lines (equation 1).

Performance of Binomial Sampling Plans for Classifying Density. The empirical model relating mean adult density to the proportion of infested leaves was accurate at densities <2 adult per leaf, but generally overpredicted mean densities >2 (Fig. 3). Almost all of the predicted mean densities fell within the 95% prediction interval of the mod-

el. However, because of the consistent bias toward over prediction, ANCOVA indicated that both the slope ($F = 60.00$; df = 1, 279; $P < 0.0001$) and intercept ($F = 34.86$; df = 1, 279; $P < 0.0001$) terms of the ln-ln regression (see equation 2) of the model significantly deviated from those derived from independent field observations.

The consequences of this bias in the model are clearly reflected in the resulting operating characteristic and average sample size functions of the sequential sampling plan (Fig. 4). For $\alpha = \beta$ the operating characteristic should be near 0.5 at the action threshold of 5 adults per leaf. However, because the model overpredicted mean density for a given proportion infested, there was a higher probability of classifying the population as above the threshold (based on binomial counts) at densities below the action threshold. Accordingly, there also was a corresponding lower probability of misclassifying populations at densities above the action threshold. Likewise, the highest sample size requirements should occur at the action threshold, whereas here it occurred at a density slightly lower than the action threshold. Estimated operating characteristic functions deviated from the nominal

Table 2. Comparison of sample sizes from resampling of independent data sets using fixed-precision sequential plans with those estimated directly from the sample mean and variance of the same independent data sets

| Density range	Resampling analysis		Direct estimate[a]	
	Avg precision	Avg sample size	Avg sample size	% difference
Eggs				
0.1–0.5	0.279	62.9	83.9	33.4
0.5–1.0	0.215	51.7	40.2	−22.2
1.0–5.0	0.223	40.3	36.3	−9.9
5.0–10.0	0.237	31.1	27.8	−10.6
>10.0	0.259	26.1	25.9	−0.8
All densities[b]	0.238	—	—	−4.1
Nymphs				
0.1–0.5	0.328	65.5	113.6	73.4
0.5–1.0	0.258	44.8	49.1	9.6
1.0–5.0	0.268	32.6	40.7	24.8
5.0–10.0	0.283	23.1	34.2	48.1
>10.0	0.285	19.0	32.0	68.4
All densities	0.287	—	—	37.3
Adults				
0.1–0.5	0.302	54.2	82.7	52.6
0.5–1.0	0.240	37.5	36.3	−3.2
1.0–5.0	0.211	26.5	20.3	−23.4
5.0–10.0	0.226	18.4	16.4	−10.9
>10.0	0.226	14.0	12.9	−7.9
All densities	0.241	—	—	−7.6

[a] Sample size calculated directly from the sample mean and variance using $n = (s/Dm)^2$ and a precision D of 0.25 where precision is measured as the SE to mean ratio.

[b] Average over all data sets with means ≥0.1: eggs, $n = 129$; nymphs, $n = 127$; adults, $n = 275$.

functions calculated for a binomial sequential probability ratio test; they were slightly steeper and offset toward lower mean densities.

Lowering the nominal α and β error rates in the sequential sampling model had very little influence on the operating characteristic functions (Fig. 4 C and E), reflecting little improvement in the accuracy of classifying population density. However, reducing α and β errors greatly increased required sample sizes (Fig. 4 D and F; Table 3). An average of ≈26 samples were needed when densities ranged from 3 to 7 adults per leaf (lower and upper decision boundaries) with $\alpha = \beta = 0.10$. Setting $\alpha = \beta = 0.05$ increased the average sample size to near 40 with little improvement in accuracy of classification. There was considerable variation in sample size requirement among sampling bouts of the same field data set. Interestingly, even at densities near the action threshold there were certain sampling bouts in which sampling was terminated after as few as 10 sample units. Recall that the minimum sample size was set to 10; thus, actual sample size requirements may have been even less.

To gauge the overall accuracy of the sequential sampling plan for classifying population density we estimated the probabilities of taking action when none was warranted (type I error) and of failing to take action when action was needed (type II error)
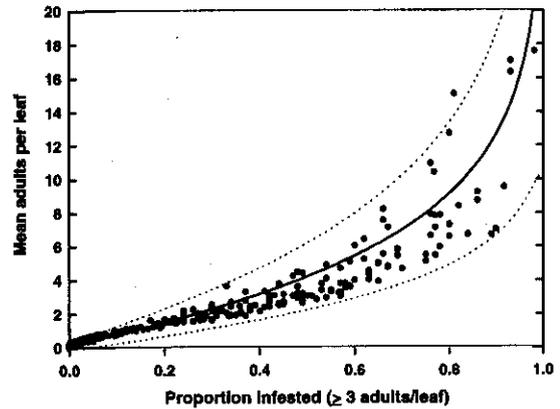
Fig. 3. Comparison of the relationship between the mean density of adults and the proportion of sample units infested with ≥3 adults from field data collected in Laveen-Tolleson, AZ, in 1994 (points) with an empirical model developed for adult *B. tabaci*. Dotted lines denote a 95% prediction interval.

over all the 284 data sets examined. These probabilities were generally not affected by changing nominal α and β error rates in the sequential model (Table 3). The overall accuracy in classifying the population densities observed in our study relative to the action threshold ranged from 87.8 to 89.1%.

The performance of fixed-sample-size plans with $n \geq 30$ was similar to that of the sequential plans. The operating characteristic function of the fixed-sample-size plan with $n = 30$ was essentially identical to that of the sequential model with $\alpha = \beta = 0.10$ (Fig. 5). The operating characteristic function improved considerably when the sample size was increased from 10 to 30, but there was relatively little difference in the operating characteristic between sample sizes of 30 and 50 (Fig. 5). These patterns were reflected in type I and type II error rates (Table 4). The overall accuracy of classifying observed population densities relative to the action threshold ranged from 87.2 to 89.9% for sample sizes of 30 to 50 (Table 4). Still, even with as few as 10 samples the correct classification was made ≈80% of the time.

Up to this point, all analyses were based on the assumption that pest populations had no explicit spatial structure despite the fact that sample units were collected from 4 distinct quadrants in each field. Because our current sampling protocol (Ellsworth et al. 1995) specifies that 15 sample units be collected from each of 2 quadrants in a field, we performed further analyses to more rigorously evaluate the accuracy of decision making based on this scouting method.

The mean operating characteristic, which was estimated by averaging results over all possible 2-quadrant combinations, differed from the operating characteristic derived from resampling of whole field data sets with $n = 30$ (Fig. 6). The quadrant-based function was less steep and offset
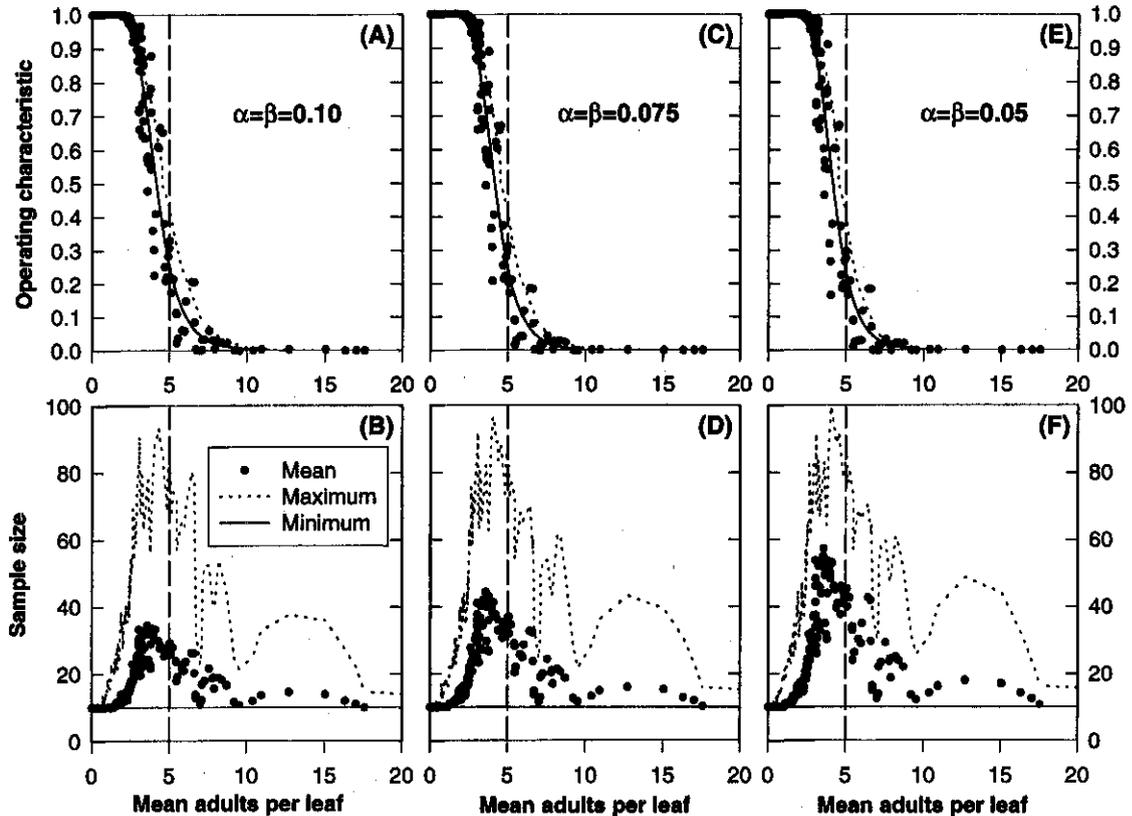
Fig. 4. Operating characteristics and average sample sizes resulting from using the Wald sequential probability ratio test to classify population density of adult *B. tabaci* relative to an action threshold of 5 adults per leaf (dashed vertical lines). Results were derived from resampling of field data sets collected in Laveen-Tolleson, AZ, in 1994 and specifying 3 different levels of nominal α and β errors. Smooth operating characteristic functions were fitted to the data using a logistic model (equation 4). The dotted lines in A, C, and E denote the nominal operating characteristic function for a binomial sequential probability ratio test.

towards lower densities resulting in higher rates of type I error (Table 4). The operating characteristic functions based on extreme 2-quadrant samples deviated greatly from the average operating char-

Table 3. Summary of the performance of the Wald sequential probability ratio test for binomial count data based on resampling analysis of whole field data from Laveen-Tolleson, AZ, 1994

Nominal error rates[a]		Overall error rates[b]		Avg sample size[c] (avg min./max)
α	β	Type I	Type II	
0.05	0.05	0.068	0.041	39.8 (10/75.2)
0.075	0.075	0.069	0.044	32.2 (10/73.8)
0.10	0.10	0.071	0.051	26.2 (10/72.1)

[a] Error rates sepecified in the sample model.
[b] Overall probability of making type I (taking action without need) and type II (failing to take action when action is needed) errors based on all field data sets.
[c] Average number of samples needed to terminate sampling at mean densities between the lower and upper decision boundaries.

acteristic suggesting that pest populations may vary considerably among quadrants. The vast majority of error, particularly for the lower extreme, was associated with the decision to control populations when they were actually below the action threshold.

Discussion

The fields from which sampling data were collected in this study encompassed a wide range of agronomic characteristics, including different planting dates, different cotton cultivars, and varying cultural practices such as irrigation frequency, cultivation, fertilization schedules, and use of insecticides. A range of environmental variables also was represented, including different soil types, spatial patterning of whitefly hosts other than cotton, and changing weather conditions over the course of the growing season. This agronomic and environmental diversity is representative of the range of conditions under which we expect our sampling plans to be used. Many of these factors
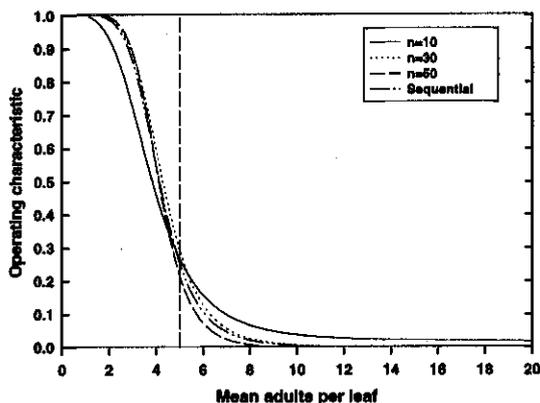
Fig. 5. Operating characteristic functions from using fixed-sample size binomial sampling plans to classify population density of adult *B. tabaci* relative to an action threshold of 5 adults per leaf (dashed vertical line). Results were derived from resampling of field data sets collected in Laveen-Tolleson, AZ, in 1994. Smooth operating characteristic functions were fitted to the data using a logistic model (equation 4). Actual data points are omitted for clarity. The operating characteristic function from the sequential sampling plan with $\alpha = \beta = 0.10$ is shown for comparison.

would be expected to influence the spatial patterning of arthropod populations and, thus, affect sampling distributions and the performance of sampling plans based on these distributions (e.g., Trumble et al. 1989, Jones 1990). We are confident that our studies provided a robust test of actual sampling plan performance in the field.

Our original mean-variance models for eggs and nymphs (Naranjo and Flint 1994) were statistically similar to ones estimated from the sample data collected in this study. This similarity probably resulted from the relatively large variability of the independent sample data and the variability of the original models as evidenced by the wide prediction intervals (see Fig. 1). In contrast, the variability of the mean-variance model for adult *B. tabaci* (Naranjo and Flint 1995) was relatively low and may have contributed to the significant deviation of this model from the independent sample data. Regardless, our resampling analysis indicated that fixed-precision sampling plans based on these models for all lifestages often did not meet the desired precision. This paradoxical result for eggs and nymphs suggests that simply testing parameters of the sampling distribution model alone is insufficient for evaluating the performance of a sampling plan. These results also highlight the stochastic nature of fixed-precision, sequential sampling plans. The desired precision specified in the sampling plan is only the average precision that would be expected over a large number of independent samples even if the mean/variance model was correct (Hutchison et al. 1988, Nyrop and Binns 1991). Sources of error include that associated with the sequential selection of samples from

Table 4. Summary of the performance of fixed-sample-size plans for binomial count data based on resampling analysis of whole field or quadrant data sets from Laveen-Tolleson, AZ, 1994

Fixed sample size	Overall error rates[a]		Quadrants[b] (n = 30)	Overall error rates[a]	
	Type I	Type II		Type I	Type II
10	0.109	0.092	Avg	0.218	0.090
30	0.064	0.064	Lower extreme	0.428	0.006
50	0.068	0.033	Upper extreme	0.095	0.228

Quadrant analysis explicitly accounts for differences in the spatial distribution of insects among the 4 quadrants sampled.

[a] Overall probability of making type I (taking action without need) and type II (failing to take action when action is needed) errors based on all field data sets.

[b] Analyses were based on repeatedly drawing 15 sample units from each of 2 quadrants (total $n = 30$). Average performance was estimated by averaging results over all possible combinations of 2 quadrants (6 combinations total); lower and upper extremes were estimated from the extreme values resulting from the 6 possible combinations of 2-quadrant samples.

the field (Anscombe 1952, Kuno 1972), those arising from the fact that mean density is being used to predict the variance, a randomly distributed variable in the mean/variance model (Nyrop and Binns 1991), and that the mean itself is estimated with error (Taylor 1984). An additional source of
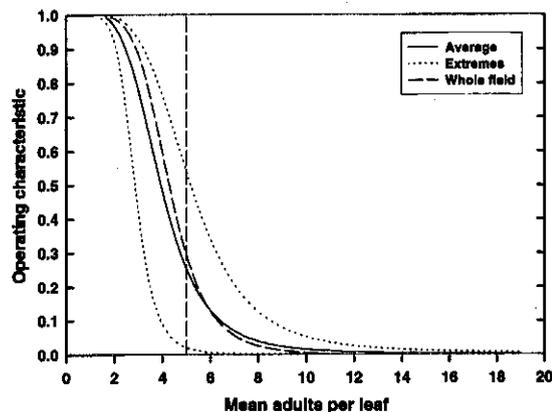

Fig. 6. Operating characteristic functions derived from using a fixed-sample size binomial sampling plans $(n = 30)$ to classify population density of adult *B. tabaci* relative to an action threshold of 5 adults per leaf (dashed vertical line) and explicitly accounting for differences in the spatial distribution of insects among the 4 quadrants sampled in Laveen-Tolleson, AZ, 1994. Analyses were based on repeatedly drawing 15 sample units from each of 2 quadrants. The average curve was estimated by averaging results over all possible combinations of 2 quadrants (6 combinations total); lower and upper extremes were estimated from the extreme values resulting from the 6 possible combinations of 2-quadrant samples. Smooth operating characteristic functions were fitted to the data using a logistic model (equation 4). Actual data points are omitted for clarity. The operating characteristic function for a fixed-sample size of 30 drawn from whole-field data sets is shown for comparison. Mean density on the x-axis is the average over all 4 quadrants.

error involves actual deviations from the mean variance model resulting from differing patterns of spatial distribution. Because we resampled field observations independent of those used to develop the sampling plans, our results reflect the contribution and impact of all these sources of error.

For eggs and adults, our sampling plans generally erred on the conservative side; the sequential stop lines achieved better than expected precision, but at the cost of requiring a slightly larger sample size than actually needed. The opposite was true for nymphs where the prescribed sample size was too small to achieve the desired precision. Sampling plans for all life stages performed poorly at low densities. For example, with adults, even the best observed precision out of 500 sampling bouts exceeded the desired level at densities less than ≈ 0.2 adults per leaf (see Fig. 2E). These outcomes are reflected in the mean-variance models, which consistently underpredicted variances associated with low mean values (see Fig. 1). However, this pattern may depend more on a general failing of the Taylor power law at low mean densities (Taylor and Woiwod 1982, Taylor 1984, Perry and Woiwod 1992) rather than deficiencies in our models which predict variance over a wide range of population densities. The overall performance of sequential sampling plans for eggs and adults was good and the sampling plan for nymphs could be improved by adjusting the stop lines by increasing the specified precision.

The binomial model (Naranjo et al. 1996a) was biased generally toward overprediction of mean density from the proportion of infested sample units at mean densities >2 adults per leaf. However, sequential or fixed-sample-size plans based on this binomial model were reasonably accurate in classifying population density relative to a critical threshold. Bias in the binomial model resulted in asymmetrical, but conservative, operating characteristic functions. Thus, there was a greater probability of deciding to treat at mean densities below the action threshold and a corresponding lower probability of deciding against treatment at densities above the action threshold. This pattern was even more pronounced when simulating scouting protocols for the fixed-sample-size plan.

The nominal α and β error rates specified in the binomial sequential probability ratio test are measures of these two types of error when actual densities are equal to those that define the lower and upper sequential stop line boundaries, respectively. Nyrop and Binns (1991) suggest that α and β, as well as the upper and lower boundaries, be thought of as parameters that can be adjusted to achieve desired operating characteristic and average sample size functions. We did not vary the boundary conditions in our analysis, but changes in nominal α and β errors had very little effect on operating characteristic functions and resulting errors in classifying population density. The relatively large increase in average sample size resulting

from reductions in nominal error rates is not offset by the correspondingly small gain in accuracy. Specifying $\alpha = \beta = 0.10$, the likelihood of taking action when the population is actually below the action threshold was $\approx 7\%$, whereas the chance of failing to take action at densities above the threshold was $\approx 5\%$. Reducing nominal error rates to 0.05 increased accuracy <1%, while increasing average sample size by 52%. Very similar levels of accuracy were achieved with fixed-sample-size plans using $n = 30$ (see Table 4) and assuming no spatial structure in the pest population. Considering all population densities observed, correct decisions were made >87% of the time. Increasing sample size by 67% to $n = 50$ only improved accuracy of classification to 90%.

Resampling analyses explicitly based on the spatial patterning of pest populations among quadrants of a field revealed large variation in error probabilities associated with decision making. When 15 sample units are drawn from each of 2 quadrants in a field it was possible for decision accuracy to drop to as low as 57% (see Fig. 6; Table 4). The vast majority of this error was associated with making the decision to control when average pest densities over the whole field were below the action threshold (type I error). From a grower perspective the more critical type II error never exceeded 23% even in the most extreme case. Because neither the 2 quadrants a scout chooses to sample nor the densities of whiteflies in these quadrants can be predicted, the accuracy of decision making would be expected to fall somewhere between the low and high extremes (see Fig. 6).

A 23% probability of treating a field that does not require treatment may represent unacceptable risks from the perspective of increased environmental contamination or higher probabilities of the development of insecticide resistance. Several factors ameliorate such concerns. First, the average density at which growers in this study first applied insecticides for control of *B. tabaci* was 2.1 adults per leaf (Diehl and Ellsworth 1995). Even our extreme operating characteristic function (Fig. 6) indicates a small probability (<0.05) of making the decision to treat at this density. Thus, our sampling and decision-making plan represents a considerably improvement compared with current practices. Further, our estimates of accuracy assume that the goal is to correctly classify overall population density in a field. Given that most insecticides are applied on a whole-field basis, an equally realistic goal may be to treat a field even if only 1 quadrant exceeds the threshold. From this perspective and based on highly variable populations among quadrants in some fields, the accuracy of our sampling plan in identifying a field in need of treatment may improve considerable. Overall, results highlight the importance of accounting for actual spatial distributions in evaluating sampling plans.

The relative merits of using a sequential versus a fixed-sample-size plan depend on the actual population density in the field being monitored. When the density of adults is far below or far above the threshold level the sequential plan is most efficient because most decisions could be made with ≤10 samples. However, when densities are near the threshold, sample size increases dramatically, and there may be individual sample bouts that would require nearly 100 samples to reach a decision (see Fig 4). In contrast, a fixed-sample-size model with a similar operating characteristic function would be less efficient at densities far above and far below the threshold but less demanding at densities near the threshold.

As mentioned, our binomial plan for sampling whiteflies is implemented using a fixed-sample-size of 30, with 15 sample units being collected at 2 different sites per field (Ellsworth et al. 1995). One of our primary concerns was adequate coverage of large commercial fields (>16 ha) which is more difficult to achieve with a sequential approach, particularly at low or high population densities when sampling would be terminated after observing relatively few sample units. Variable-intensity sampling (Hoy 1991), which allocates sequential samples more evenly within a field, may be an alternative approach worth consideration. In general, this technique requires more training to use and may be more difficult to implement consistently, particularly with the inevitable turnover of scouts through the season. Our recommended plan proved to be relatively accurate for classifying population density and on average took <10 min to complete in a given field (Ellsworth et al. 1996).

Many sampling models have been developed for a wide variety of insects of economic importance. However, relatively few of these models have been adequately tested in the field. There are 2 basic approaches for assessing the performance characteristics of sampling plans, both based on computer-intensive methods that draw repeated samples from a distribution to arrive at expected values of performance measures and their associated variances. The Monte Carlo simulation (e.g., Nyrop and Binns 1991) draws random samples from a theoretical distribution (e.g., negative binomial) that is assumed to mimic actual sampling distributions. A 2nd approach is to randomly resample sets of actual field data collected independently of those used to develop the sampling plan (Hutchison et al. 1988). This approach is less tractable analytically, but more robust because the underlying sampling distribution is defined by the data set, and the technique simultaneously evaluates sampling error as well as deviations from the underlying sampling model (Hutchison 1994, Naranjo and Hutchison 1997). Based on this robust validation technique and using field observations representing a wide range of environmental and agronomic conditions, our sampling plans performed well and should be useful for estimating and clas-sifying population densities of *B. tabaci* in cotton over a wide area.

References Cited

Anscome, F. J. 1952. Large-sample theory of sequential estimation. Proc. Camb. Philos. Soc. 48: 600–607.

Brown, J. K., D. R. Frolich, and R. C. Rosell. 1995. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex. Annu. Rev. Entomol. 40: 511–534.

Butler, G. D., Jr., and T. J. Henneberry. 1994. *Bemisia* and *Trialeurodes* (Hemiptera: Aleyrodidae), pp. 325–352. *In* G. A. Matthews and J. P. Turnstall [eds.], Insect pests of cotton. CAB, Wallingford, U.K.

Byrne, D. N., R. R. Rathman, T. V. Orum, and J. C. Palumbo 1996. Localized migration and dispersal by the sweetpotato whitefly, *Bemisia tabaci*. Oecologia (Berl.) 105: 320–328.

Diehl, J. W., and P. C. Ellsworth. 1995. Community-wide implementation of sampling and action thresholds for whiteflies in cotton, pp. 255–258. *In* Cotton, a College of Agriculture Report, Series P-99. University of Arizona, Tucson.

Ellsworth, P. C., and D. Meade. 1994. Action thresholds for the sweetpotato whitefly in Arizona, pp. 878–881. *In* Proceedings of the Beltwide Cotton Conferences. National Cotton Council, Memphis, TN.

Ellsworth, P., J. Diehl, T. Dennehy, and S. Naranjo. 1995. Sampling sweetpotato whiteflies in cotton. Cooperative Extension, IPM Series 2 (revised). University of Arizona, Tucson.

Ellsworth, P. C., J. W. Diehl, and S. H. Husman. 1996. Establishment of integrated pest management infrastructure: a community-based action program for *Bemisia* management, pp. 681–693. *In* D. Gerling and D. Mayer [eds.], *Bemisia* 1995: taxonomy, biology, damage, control and management. Intercept, Hants, U.K.

Fowler, G. W., and A. M. Lynch. 1987. Sampling plans in insect pest management based on Wald's sequential probability ratio test. Environ. Entomol. 16: 345–354.

Gerrard, D. J., and H. C. Chiang. 1970. Density estimation of corn rootworm egg populations based upon frequency of occurrence. Ecology 51: 237–245.

Green, R. H. 1970. On fixed precision sequential sampling. Res. Popul. Ecol. (Kyoto) 12: 249–251.

Hoy, C. W. 1991. Variable-intensity sampling for proportion of plants infested with pests. J. Econ. Entomol. 84: 148–157.

Hutchison, W. D. 1994. Sequential sampling to determine population density, pp. 207–244. *In* L. Pedigo and G. Buntin [eds.], Handbook of sampling methods for arthropods in agriculture. CRC, Boca Raton, FL.

Hutchison, W. D., D. B. Hogg, M. A. Poswal, R. C. Berberet, and G. W. Cuperus. 1988. Implications

of the stochastic nature of Kuno's and Green's fixed-precision stop lines: sampling plans for the pea aphid (Homoptera: Aphididae) in alfalfa as an example. J. Econ. Entomol. 81: 749–758.

Jandel. 1994. SigmaPlot user's manual, version 2.0. Jandel, San Rafael, CA.

Jones, V. P. 1990. Developing sampling plans for spider mites (Acari: Tetranychidae): those who don't remember the past may have to repeat it. J. Econ. Entomol. 83: 1656–1664.

Kono, T., and T. Sugino. 1958. On the estimation of the density of rice stem infested by the rice stem borer. Jpn. J. Appl. Entomol. Zool. 2: 184–188.

Kuno, E. 1972. Some notes on population estimation by sequential sampling. Res. Popul. Ecol. (Kyoto) 14: 58–73.

Naranjo, S. E., and H. M. Flint. 1994. Spatial distribution of preimaginal *Bemisia tabaci* (Homoptera: Alyerodidae) in cotton and development of fixed-precision sequential sampling plans. Environ. Entomol. 23: 254–266.

1995. Spatial distribution of adult *Bemisia tabaci* in cotton and development and validation of fixed-precision sequential sampling plans for estimating population density. Environ. Entomol. 24: 261–270.

Naranjo, S. E., and W. D. Hutchison. 1997. Validation of insect sampling plans using a resampling approach: software and analysis. Am. Entomol. 43(1): 48–57.

Naranjo, S. E., H. M. Flint, and T. J. Henneberry. 1996a. Binomial sampling plans for estimating and classifying population density of adult *Bemisia tabaci* in cotton. Entomol. Exp. Appl. 80: 343–353.

Naranjo, S. E., C. C. Chu, P. C. Ellsworth, T. J. Henneberry, R. L. Nichols, D. G. Riley, N. C. Toscano, and T. F. Watson. 1996b. Results of a two year regional project to determine an action threshold to manage whiteflies in cotton, pp. 132–133. *In* Pro-ceedings of the Beltwide Cotton Conferences. National Cotton Council, Memphis, TN.

Nyrop, J. P., and M. Binns. 1991. Quantitative methods for designing and analyzing sampling programs for use in pest management, pp. 67–132. *In* D. Pimentel [ed.], Handbook of pest management in agriculture, vol. 2. CRC, Boca Raton, FL.

Perry, J. N., and I. P. Woiwod. 1992. Fitting Taylor's power law. Oikos 65: 538–542.

SAS Institute. 1989. SAS/STAT user's guide, version 6, 4th ed. SAS Institute, Cary, NC.

Schaalje, G. B., R. A. Butts, and T. L. Lysyk. 1991. Simulation studies of binomial sampling: a new variance estimator and density predictor, with special reference to the Russian wheat aphid (Homoptera: Aphididae). J. Econ. Entomol. 84: 140–147.

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

1984. Assessing and interpreting the spatial distribution of insect populations. Annu. Rev. Entomol. 29: 321–357.

Taylor, L. R., and I. P. Woiwod. 1982. Comparative synoptic dynamics. I. Relationships between inter and intra-specific spatial and temporal variance/mean population parameters. J. Anim. Ecol. 51: 879–906.

Trumble, J. T., M. J. Brewer, A. M. Shelton, and J. P. Nyrop. 1989. Transportability of fixed-precision sampling plans. Res. Popul. Ecol. (Kyoto) 31: 325–342.

[USDA] U.S. Department of Agriculture. 1995. Silverleaf whitefly: 1995 supplement to the 5-year national research and action plan, USDA–ARS 1995-2, Washington DC.

Wald, A. 1947. Sequential analysis. Wiley, New York.

Watson, T. F., J. C. Silvertooth, A. Tellez, and L. Lastra. 1992. Seasonal dynamics of sweetpotato whitefly in Arizona. Southwest. Entomol. 17: 149–167.