

## **Sequential Hypothesis Testing with Spatially Correlated Presence-Absence Data**

Author(s): Elijah DePalma, Daniel R. Jeske, Jesus R. Lara, and Mark Hoddle

Source: Journal of Economic Entomology, 105(3):1077-1087. 2012.

Published By: Entomological Society of America

DOI: <http://dx.doi.org/10.1603/EC11199>

URL: <http://www.bioone.org/doi/full/10.1603/EC11199>

---

BioOne ([www.bioone.org](http://www.bioone.org)) is a nonprofit, online aggregation of core research in the biological, ecological, and environmental sciences. BioOne provides a sustainable online platform for over 170 journals and books published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Web site, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/page/terms\\_of\\_use](http://www.bioone.org/page/terms_of_use).

Usage of BioOne content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

# Sequential Hypothesis Testing With Spatially Correlated Presence-Absence Data

ELIJAH DePALMA,<sup>1</sup> DANIEL R. JESKE,<sup>1,2</sup> JESUS R. LARA,<sup>3</sup> AND MARK HODDLE<sup>3</sup>

J. Econ. Entomol. 105(3): 1077–1087 (2012); DOI: <http://dx.doi.org/10.1603/EC11199>

**ABSTRACT** A pest management decision to initiate a control treatment depends upon an accurate estimate of mean pest density. Presence-absence sampling plans significantly reduce sampling efforts to make treatment decisions by using the proportion of infested leaves to estimate mean pest density in lieu of counting individual pests. The use of sequential hypothesis testing procedures can significantly reduce the number of samples required to make a treatment decision. Here we construct a mean-proportion relationship for *Oligonychus perseae* Tuttle, Baker, and Abatiello, a mite pest of avocados, from empirical data, and develop a sequential presence-absence sampling plan using Bartlett's sequential test procedure. Bartlett's test can accommodate pest population models that contain nuisance parameters that are not of primary interest. However, it requires that population measurements be independent, which may not be realistic because of spatial correlation of pest densities across trees within an orchard. We propose to mitigate the effect of spatial correlation in a sequential sampling procedure by using a tree-selection rule (i.e., maximin) that sequentially selects each newly sampled tree to be maximally spaced from all other previously sampled trees. Our proposed presence-absence sampling methodology applies Bartlett's test to a hypothesis test developed using an empirical mean-proportion relationship coupled with a spatial, statistical model of pest populations, with spatial correlation mitigated via the aforementioned tree-selection rule. We demonstrate the effectiveness of our proposed methodology over a range of parameter estimates appropriate for densities of *O. perseae* that would be observed in avocado orchards in California.

**KEY WORDS** Bartlett's sequential test, binomial sampling, generalized linear mixed model

Neglecting the spatial structure of pest populations can result in an inaccurate estimation of pest densities. Spatial analyses have been previously used in studies of diverse groups of pests of agricultural crops such as lentils (Schotzko and O'Keefe 1989), cotton (Gozé et al. 2003), and grapes (Ifoulis and Savopoulou-Soultani 2006, Ramírez-Dávila and Porcayo-Camargo 2008). In all these studies, spatial analyses were conducted by first transforming count data so as to resemble continuous, normally distributed data. Generalized linear mixed models (GLMM), however, are statistical models that are particularly useful for modeling discrete response variables that may be correlated (Breslow and Clayton 1993), such as spatially correlated count data or presence-absence data. GLMMs have been used across multiple scientific disciplines, including ecological studies of pest populations (Candy 2000, Bianchi et al. 2008, Takakura 2009). In this article, we propose a spatial GLMM for a sequential presence-absence sampling program for *Oligonychus perseae*, Tuttle, Baker, and Abatiello (Acari: Tetranychidae), a pest mite of avocados (*Persea americana* Miller [Lau-

raceae]) in California as an example for developing this modeling approach.

The perseae mite, *O. perseae*, is native to Mexico and is an invasive pest in California, Costa Rica, Spain, and Israel. It is a foliar pest of avocados and is most damaging to the popular 'Hass' variety that accounts for 94% of the total production acreage in California (California Avocado Commission [CAC] 2009), it is worth ≈\$300 million each year, and ≈6,000 growers farm ≈27,000 ha of this cultivar (CAC 2010). Feeding by high-density populations of *O. perseae* can cause extensive defoliation to avocados (Hoddle et al. 2000), and in California this pest is typically controlled with pesticides (Humeres and Morse 2005). A scientifically based action threshold and economic injury level (EIL) has not been calculated for *O. perseae* in California. However, work from Israel suggests that the EIL lies between 100 and 250 mites per leaf and the recommended action threshold is in the range of 50–100 mites per leaf (Moaz et al. 2011).

Counting *O. perseae* mites with a hand lens in the field is tedious, time consuming, and an inaccurate approach to monitor population densities for making control decisions. An alternative approach is presence-absence or binomial sampling, which estimates pest population density using the proportion of leaves infested with at least one mite versus the proportion

<sup>1</sup> Department of Statistics, University of California, Riverside, Riverside, CA 92521.

<sup>2</sup> Corresponding author, e-mail: [daniel.jeske@ucr.edu](mailto:daniel.jeske@ucr.edu).

<sup>3</sup> Department of Entomology, University of California, Riverside, Riverside, CA 92521.

**Table 1.** Summary information for the avocado orchards in California from which count data were collected to construct a mean-proportion relationship for *O. perseae*

Orchard	County	Year sampled	Trees sampled	No. leaves	No. sets
1	Ventura	1997	42	6,469	16
2	Orange	1999	66	5,280	8
2	Orange	2000–2001	42	17,220	41
3	San Diego	2009	31	247	1
4	Santa Barbara	2009	30	240	1
5	Santa Barbara	2009	30	240	1
6	Santa Barbara	2009	30	240	1
7	Santa Barbara	2009	30	240	1
8	Ventura	2010	30	240	1
9	Ventura	2010	30	240	1

In total, 72 data sets, each measuring the proportion of infested leaves and the mean leaf mite density, were used to fit the empirical equation 1. The 72 data points and the resulting fitted curve are graphed in Fig. 1.

of clean leaves with no mites. Presence-absence sampling is fast, simple, and allows large areas to be surveyed quickly to quantify pest damage. Presence-absence sampling programs have been developed for a variety of agricultural pests including other spider mite species, eriophyid mites, aphids, flea beetles, leaf hoppers, whiteflies, mealybugs, and leaf miners (Alatawi et al. 2005, Binns et al. 2000, Galvan et al. 2007, Hall et al. 2007, Hyung Lee et al. 2007, Kabaluk et al. 2006, Martinez-Ferrer et al. 2006, Robson et al. 2006).

Sequential sampling procedures are considered a cost effective approach to assessing pest densities (Mulekar et al. 1993, Young and Young 1998, Binns et al. 2000). Cost savings accrue in comparison to fixed sample size procedures, because sequential procedures often require a significantly reduced number of sampled observations to reach a treatment decision, which can result in appreciable savings in the cost of sampling. In applications of sequential sampling, Wald's (1947) sequential probability ratio test (SPRT) is the most often used approach. Wald's SPRT is useful for sampling programs when it can be assumed that, aside from the primary parameter of interest, there are no additional unknown parameters (i.e., nuisance parameters) in the model.

In the case of independent and identically distributed (IID) samples, a modification to Wald's SPRT results in Bartlett's (1946) SPRT, which can be applied to pest count models containing nuisance parameters (Shah et al. 2009). However, spatial correlation of pest populations violates the independence assumption required for Bartlett's SPRT. In related work on spatially correlated pest count data, Li et al. (2012) proposed a first-stage initial sample used to assess the effective range of spatial correlation, followed by a second-stage sampling procedure in which each sampled observation is outside of the effective range of all previously sampled observations. Sampling outside of the effective range eliminates any spatial correlation so that Bartlett's SPRT may be applied. In this article, we propose to sequentially sample observations for *O. perseae* so that each sampled observation is maximally spaced from all other previously sampled observations, thereby eliminating spatial correlation. This sampling strategy eliminates the necessity of an initial, first-stage sample as proposed by Li et al. (2012), and

we demonstrate its effectiveness for mitigating spatial correlation sufficiently to allow the application of Bartlett's SPRT for a range of parameter estimates appropriate to *O. perseae* in California avocado orchards. To our knowledge, this article is the first to combine sequential hypothesis testing techniques with presence-absence sampling strategies that account for spatial correlation of pest densities.

## Materials and Methods

**Mean-Proportion Relationship.** The essential component of a presence-absence sampling plan is an accurate relationship between the mean pest density,  $M$ , and the proportion of leaves infested with at least one pest individual,  $P$ . The mean-proportion relationship can be modeled using an empirical equation (Kono and Sugino 1958, Gerrard and Chaing 1970), which has been used to develop binomial sampling plans for pests (Hall et al. 2007, Martinez-Ferrer et al. 2006),

$$\ln(-\ln(1 - P)) = a + b \times \ln(M) \quad [1]$$

The parameters  $a$  and  $b$  can be fit using linear regression.

To construct a mean-proportion relationship for *O. perseae*, Hass avocado leaves were collected randomly from nine avocado orchards in Southern California across various years (Table 1), and counts of all *O. perseae* stages (except eggs) were performed using stereomicroscopes. Seventy-two mite count data sets (incorporating 30,656 leaves with a density range of 0–342 mites per leaf) were used to fit equation 1, with resulting parameter estimates  $a = -1.72762$  and  $b = -0.66527$ . This relationship is shown in Fig. 1 where we plotted the 72 data pairs of mean pest density per leaf and proportion of infested leaves, along with the fitted empirical equation 1.

**Presence-Absence Sampling Hypothesis Test.** The mean-proportion relationship allows a pest control adviser to estimate the mean density of mites per leaf without counting individual mites. This is achieved by sampling a number of leaves and determining the proportion of leaves for which at least one mite is present. In our context, we use the mean-proportion relationship to convert an action threshold for mite

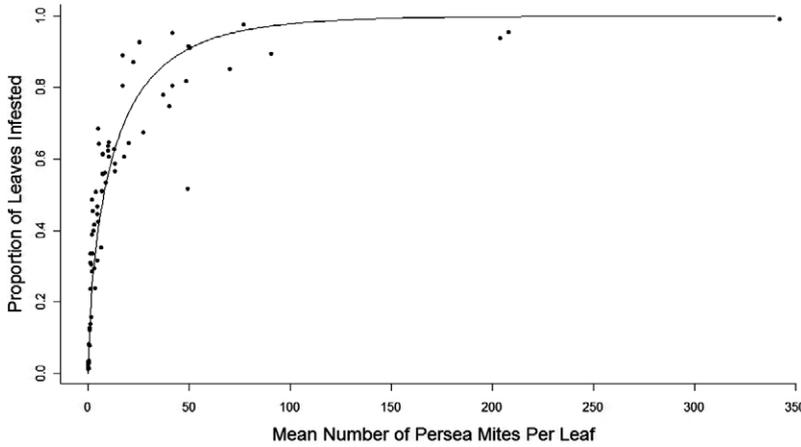


Fig. 1. Plotted values of data sets for *O. perseae*, each measuring the proportion of infested leaves and the mean mite density per leaf, and a graph of the fitted empirical equation 1.

densities per leaf into an action threshold for proportion of infested leaves.

Moaz et al. (2011) determined an action threshold range for *O. perseae* mite densities to be 50–100 mites per leaf. Using the lower bound of this range we construct a statistical decision problem for intervention treatment by proposing the following hypothesis test for mean mite densities,  $M$ :

$$H_0: M = 25 \frac{\text{mites}}{\text{leaf}} \text{ vs. } H_1: M = 75 \frac{\text{mites}}{\text{leaf}}.$$

Because the midpoint of this range is 50 mites/leaf, if the null hypothesis,  $H_0$ , is rejected in favor of  $H_1$ , then mite densities are above 50 mites/leaf and treatment is recommended. Failure to reject  $H_0$  in favor of  $H_1$  implies that mite densities are below 50 mites/leaf and treatment is unnecessary.

Using equation 1 we may explicitly write  $P$  as a function of  $M$ ,

$$P = 1 - \exp[-\exp(a + b \cdot \ln M)], \quad [2]$$

and using the above parameter estimates for  $a$  and  $b$  we convert the hypothesis test for  $M$  into the following hypothesis test for  $P$ :

$$H_0: P = 0.78 \text{ vs. } H_1: P = 0.95.$$

In a binomial sampling plan, the number of infested leaves (i.e., leaves with at least one mite present) in a randomly selected sample of leaves follows a binomial distribution, with the only unknown model parameter being the proportion of infested leaves,  $P$ . The hypothesis test for  $P$  may be evaluated using Wald’s SPRT, which is the most efficient procedure for evaluating a hypothesis test of simple hypotheses when the underlying model contains only a single, unknown parameter (Wald 1947). However, from our field observation and analysis of *O. perseae* count data (Li et al. 2012), mite populations were shown to cluster on individual avocado trees, with neighboring trees having similar population densities. This spatial correla-

tion did not exist once sampled trees were 3–4 trees distant from the last sampled tree (Li et al. 2012). Thus, to more accurately evaluate the above hypothesis test and make a decision regarding treatment, a model must be developed that accounts for the aggregation of mites on individual trees and the spatial correlation of mite densities across trees.

**Spatial GLMM.** To account for the aggregation of mites on individual trees we constructed a Bernoulli response GLMM in which the proportion of infested leaves varies by tree, as determined by a fixed effect common to all trees and a random effect that varies from tree to tree. To account for the spatial correlation of mite densities among trees we allow the random tree effects in the GLMM to be spatially correlated. Specifically, suppose that we select  $n$  trees to be sampled, and that on each tree we randomly sample  $m$  leaves. For  $i = 1, \dots, n$ , on the  $i^{\text{th}}$  tree let  $p_i$  be proportion of infested leaves,  $0 \leq p_i \leq 1$ , and let  $Y_{ij}$  be the corresponding Bernoulli ( $p_i$ ) response for the  $j^{\text{th}}$  leaf sampled,  $j = 1, \dots, m$ , where  $Y_{ij} = 1$  if at least one mite is present and  $Y_{ij} = 0$  otherwise. Let  $\gamma$  denote the fixed effect common to all trees, let  $\xi = (S_1, \dots, S_n)'$  denote the spatially correlated random tree effects for the  $n$  trees sampled, and let  $Y_i$  equal the sum of the  $m$  Bernoulli responses for the  $i^{\text{th}}$  tree,  $Y_i = \sum_{j=1}^m Y_{ij}$ . Therefore, our proposed spatial GLMM is defined as:

$$Y_i | \xi \sim \text{Binomial}(m, p_i)$$

$$\text{logit}(p_i) \equiv \log\left(\frac{p_i}{1 - p_i}\right) = \gamma + S_i \quad [3]$$

$$\xi \sim \text{MVN}(\mathbf{0}, \Sigma) \text{ (multivariate normal),}$$

where  $\Sigma$  is the  $n \times n$  covariance matrix for the random tree effects whose off-diagonal elements determine the correlation structure. We propose allowing for a spatially symmetric correlation structure in which the correlation between the random effects of two trees decreases exponentially with the distance between

the trees, known as a spatial exponential correlation structure (Schabenberger and Gotway 2005). With this correlation structure, the  $(i, i')$  element of  $\Sigma$  is  $\sigma^2 \exp(-d_{i,i'} / \rho)$ , where  $d_{i,i'}$  is the Euclidean distance between the  $i$ -th and  $i'$ -th trees,  $\rho$  is a scale parameter that dictates the strength of the spatial correlation, and  $\sigma^2$  is a scale parameter that determines the variability of the random tree effect on an individual tree. Under this parameterization it can easily be shown that the effective range of the spatial correlation is  $3\rho$ , and that for tree-separation distances beyond this range the spatial correlation is essentially diminished (Schabenberger and Gotway 2005).

**Spatial GLMM Hypothesis Test.** It follows from equation 3 that for each tree the proportion of infested leaves,  $p_i$ , is a logit-normal random variable with parameters  $\gamma$  and  $\sigma^2$ . Although the mean of a logit-normal random variable cannot be analytically related to its parameters, a simple analytic relation exists between  $\gamma$  and the median of  $p_i$ ,

$$\gamma = \log\left(\frac{\text{median}(p_i)}{1 - \text{median}(p_i)}\right). \quad [4]$$

In the spatial GLMM model the proportion of infested leaves varies from tree to tree, and a pest manager seeking to make a treatment decision for an entire orchard may use median ( $p_i$ ) as a measure of the proportion of infested leaves over the entire orchard. Thus, using the spatial GLMM the hypothesis test we previously derived for *O. perseae* in terms of  $P$  may be converted into a hypothesis test for  $\gamma$  as follows:

$$\begin{aligned} H_0: \gamma &= \log\left(\frac{0.78}{1 - 0.78}\right) = 1.27 \text{ vs. } H_1: \gamma \\ &= \log\left(\frac{0.95}{1 - 0.95}\right) = 2.94. \end{aligned} \quad [5]$$

Hence, the median pest density over an entire orchard is determined by the spatial GLMM parameter  $\gamma$ , whereas  $\sigma^2$ , and  $\rho$  are nuisance parameters.

**Bartlett's SPRT.** In a model without nuisance parameters, Wald's SPRT is the most efficient test of simple hypotheses, requiring the minimum number of expected samples among all hypothesis tests with the same Type-1 (falsely reject  $H_0$ ) and Type-2 (falsely fail to reject  $H_0$ ) error rates. In a model that contains nuisance parameters, Bartlett (1946) proved that, if the samples are independent and identically distributed (IID), then the Type-1,-2 error rates are asymptotically preserved if the nuisance parameters are replaced with their conditional maximum likelihood estimates at each stage of the sequential testing procedure.

In the context of this study, the IID assumption of Bartlett's SPRT is achieved if spatial correlation is not present, and in a subsequent section we propose a tree-selection rule that effectively diminishes any spatial correlation. Hence, throughout this section we presume that our proposed spatial GLMM has been reduced to a GLMM with no spatial correlation ( $\rho = 0$ ) to which Bartlett's SPRT may be applied.

We apply Bartlett's SPRT to the observations  $\{Y_1, Y_2, \dots\}$ , where  $Y_i$  is the number of mite-infested leaves on the  $i^{\text{th}}$  tree among the  $m$  leaves sampled, with  $m$  determined in the next section. The sequential test for subsequent sampling occasions is based on the log-likelihood ratio,

$$\lambda_n = \log\{f(\underline{Y}_n; \gamma_1, \hat{\sigma}_n^2(\gamma_1)) / f(\underline{Y}_n; \gamma_0, \hat{\sigma}_n^2(\gamma_0))\}. \quad [6]$$

Here,  $n$  denotes the current number of trees sampled in the sequential procedure,  $\underline{Y}_n = (Y_1, \dots, Y_n)'$  are the current observed responses, and the likelihoods are obtained from equation 3 by integrating out the random effects,  $\underline{\xi}_n = (S_1, \dots, S_n)'$ , assuming  $\rho = 0$ ,

$$\begin{aligned} f(\underline{Y}_n; \gamma, \sigma^2) &= \int \dots \int_n f(\underline{Y}_n | \underline{\xi}_n) \cdot f(\underline{\xi}_n) d\underline{\xi}_n \\ &= \int \dots \int_n \left[ \prod_{i=1}^n \binom{m}{Y_i} p_i^{Y_i} (1 - p_i)^{m - Y_i} \right] \\ &\quad \cdot \left[ \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} \underline{\xi}_n' \Sigma^{-1} \underline{\xi}_n\right) \right] d\underline{\xi}_n \\ &= \prod_{i=1}^n \left\{ \int \binom{m}{Y_i} \left( \frac{\exp(\gamma + S_i)}{1 + \exp(\gamma + S_i)} \right)^{Y_i} \right. \\ &\quad \left. \left( 1 - \frac{\exp(\gamma + S_i)}{1 + \exp(\gamma + S_i)} \right)^{m - Y_i} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} S_i^2\right) dS_i \right\}. \end{aligned} \quad [7]$$

Equation 7 consists of a product of  $n$  one-dimensional integrals, each of which is easily numerically evaluated using Gauss-Hermite quadrature. For  $\gamma \in \{\gamma_0, \gamma_1\}$ ,  $\hat{\sigma}_n^2(\gamma)$  denotes the conditional MLE of the unknown nuisance parameter  $\sigma^2$  obtained by setting  $\gamma$  in equation 7 to  $\gamma_0$  or  $\gamma_1$ , respectively, and then maximizing the right-hand side with respect to  $\sigma^2$ . For the hypothesis test in equation 6 used to make a treatment decision for *O. perseae*,  $\gamma_0 = 1.27$  and  $\gamma_1 = 2.94$ .

The upper and lower stopping boundaries of Bartlett's SPRT are

$$A = \ln\left(\frac{\beta}{1 - \alpha}\right) \text{ and } B = \ln\left(\frac{1 - \beta}{\alpha}\right), \quad [8]$$

respectively, so that Bartlett's SPRT rejects  $H_0$  in favor of  $H_1$  at the first  $n$  for which  $\lambda_n \geq B$ , fails to reject  $H_0$  in favor of  $H_1$  at the first  $n$  for which  $\lambda_n \geq A$  and continues by sampling another tree if  $A < \lambda_n < B$ . The resulting Type-1 and Type-2 error rates asymptotically satisfy  $P(\text{Reject } H_0 | H_0) \leq \alpha$  and  $P(\text{Fail to reject } H_0 | H_1) \leq \beta$ , respectively, so that  $\alpha$  and  $\beta$  are Type-1,2 error rate upper bounds, respectively.

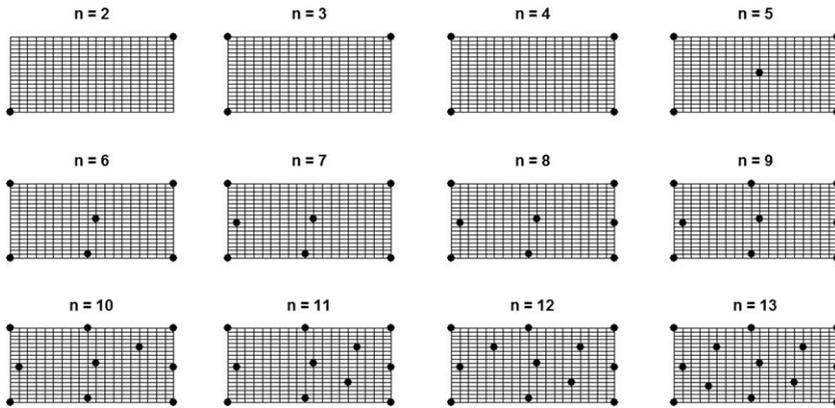


Fig. 2. Visual illustration of the sequential, maximin tree-selection rule applied to a 20 × 20 grid of trees, demonstrating how the first 13 trees are selected. Here we arbitrarily chose the first tree selected to be the lower, left-hand corner tree.

**Leaf-Selection Rules and Sampling Cost.** To determine the optimal number of leaves to sample per tree,  $m$ , and assuming that our tree-selection rule has effectively diminished spatial correlation ( $\rho = 0$ ), we conducted a simulation study to analyze average sample numbers (ASN) of Bartlett’s SPRT applied to the hypothesis test in equation 6, over a range of  $m$  and  $\sigma^2$  parameter values appropriate to *O. perseae*, and Type-1,-2 error rate upper bounds of  $\alpha = 0.10$  and  $\beta = 0.10$ .

Let  $N$  denote the number of sampled trees required to reach the stopping rule in Bartlett’s SPRT. As the number of leaves sampled per tree,  $m$ , increases, the expected number of sampled trees,  $E(N)$  decreases, but the expected total number of sampled leaves,  $m \cdot E(N)$  increases (see the Results section for details). To determine an optimal value for  $m$ , we constructed a simple sampling cost function that includes a sampling cost for each tree and an additional sampling cost for each leaf:

$$\text{Cost} = (\text{cost per tree}) \cdot N + (\text{cost per leaf}) \cdot m \cdot N \tag{9}$$

For a given value of  $m$  the expected cost,  $E(\text{Cost})$ , depends on  $E(N)$  that varies with  $\gamma$ . For each value of  $m$ , we evaluate  $E(N)$  at the value of  $\gamma$ , say  $\gamma_{\max}$ , for which  $E(N)$  is maximized. We choose  $m$  to minimize the expected cost, which up to a constant of proportionality can be written as:

$$E(\text{Cost}) \propto (1 + cm) \cdot E(N)|_{\gamma_{\max}},$$

$$\text{where } c = \frac{\text{cost per leaf}}{\text{cost per tree}}. \tag{10}$$

In practice, the costs associated with selecting an additional leaf should be much less than the costs associated with selecting and locating an additional tree, so that the leaf-to-tree cost ratio,  $c$ , should be much less than one. Given a value of  $c < 1$ ,  $E(\text{Cost})$  versus  $m$  is plotted and in the resulting graph an optimal value of  $c < 1$ ,  $E(\text{Cost})$  versus  $m$  is chosen so as to minimize  $E(\text{Cost})$ .

**Sequential Maximin Tree-Selection Rule.** To mitigate spatial correlation of mite counts between adja-

cent trees we propose to sequentially select each tree to be maximally spaced from all other previously selected trees. We base our notion of ‘maximally spaced’ on a maximin distance criterion, in which each tree is selected so as to maximize the minimum distance it has to all other previously selected trees. A design constructed by this rule has been referred to as a ‘coffee-house’ design for the similar way in which customers select their tables in a coffee-house (Müller 2007).

In a nonsequential, fixed-size spatial sampling setting, maximin designs possess optimality properties that we now briefly describe. In a fixed-size sampling setting, a maximin design simultaneously selects all points so that the minimum distance between all pairs of selected points is maximized. The index of a fixed-size maximin design is the number of pairs of points separated by this maximal, minimum distance. For any statistical model in which the correlation between two points is a decreasing function of the distance between the two points, a fixed-size maximin design of smallest index is asymptotically related to an optimal design that minimizes the variances of parameter estimates (Johnson et al. 1990). This result enables the construction of an asymptotically optimal fixed-size sampling design based on geometric criteria alone.

In the context of this article, we adopt the above notion of ‘index’ to a sequential, maximin tree-selection rule, as follows. At each stage in the sequential procedure, we define a maximin tree to be a tree (not necessarily unique) whose minimum distance to all previously selected trees is maximal. The index of a maximin tree is defined to be the number of previously selected trees separated by this maximal, minimum distance. Our proposed sequential, maximin tree-selection rule is to select a maximin tree of smallest index.

In Fig. 2 we provide a visual illustration of the sequential, maximin tree-selection rule applied to a 20 × 20 grid of equally spaced trees, demonstrating how the first 13 trees are selected.

**Evaluation of Proposed Methodology.** Our proposed methodology for developing a presence-absence sampling plan is to use the mean-proportion

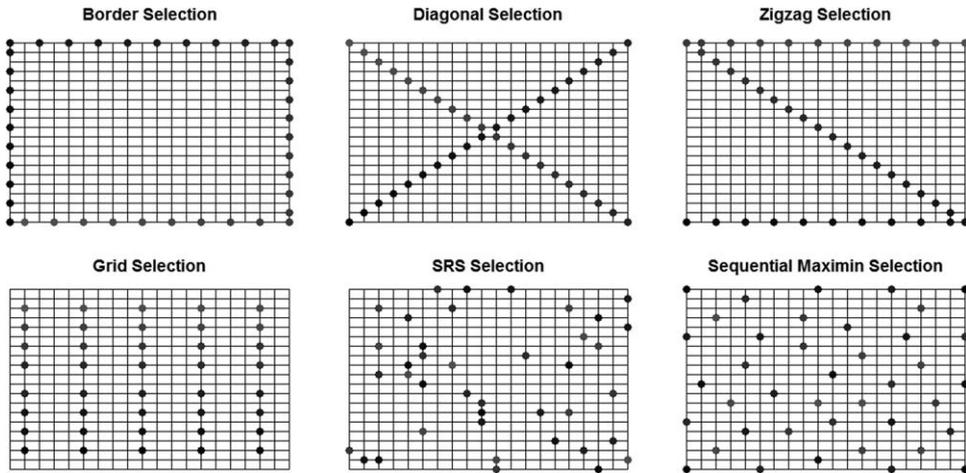


Fig. 3. The six tree-selection rules for which we evaluated the error rates of Bartlett's SPRT. The size indicates the order of tree selection from largest to smallest. We truncated the sequential hypothesis test at an upper bound of 40 trees, all of which are graphed here. However, in all cases the sequential hypothesis test typically terminated after sampling the first 5–10 trees.

relationship coupled with the spatial GLMM to construct the treatment decision hypothesis test in equation 5, to which we apply Bartlett's SPRT coupled with the sequential, maximin tree-selection rule and the leaf-selection rule. We validate the proposed methodology by verifying that the sequential, maximin tree-selection rule successfully diminishes spatial correlation sufficient to preserve the Type-1,2 error rates of Bartlett's SPRT applied to the hypothesis test in equation 5, for a range of  $\sigma^2$  and  $\rho$  parameter values appropriate to *O. perseae*.

In a simulation study we simulated presence-absence data from a spatial GLMM for a range of values of the spatial correlation parameter,  $\rho$ , and the nuisance parameter,  $\sigma^2$ , appropriate to *O. perseae*. Assuming the optimal leaf selection rule of  $m = 6$  leaves per tree (see Results section), we simulated data from a  $20 \times 20$  grid of 400 equally spaced trees. For each simulation we evaluated the hypothesis test in equation 5 by applying Bartlett's SPRT with Type-1,2 error rate upper bounds of  $\alpha = 0.10$  and  $\beta = 0.10$ . However, we truncated Bartlett's SPRT so that the maximum possible number of trees sampled is 10% of the orchard, or 40 trees in this example. If a stopping rule had not been reached after 40 trees had been sampled, then the sequential procedure was halted and a decision made based on whether the sequential hypothesis test statistic,  $\lambda_{40}$ , was closer to  $B$  the stopping rule upper boundary (reject  $H_0$ ), or closer to  $A$ , the stopping rule lower boundary (fail to reject  $H_0$ ).

We compared the sequential, maximin tree-selection rule to several other tree-selection rules, all of which are illustrated in Fig. 3: 1) border selection, where trees were sampled along the orchard borders; 2) diagonal selection, where trees were sampled along a diagonal in the orchard; 3) zigzag selection, where the lower orchard border is sampled, followed by the orchard diagonal, followed by the upper orchard bor-

der; 4) grid selection, where trees were sampled on a grid pattern uniformly spaced throughout the orchard; 5) SRS selection, where trees were selected using simple random sampling throughout the orchard. In Fig. 3 we indicate the order in which trees were selected in decreasing size from largest to smallest. Although all 40 trees are designated for each truncated sequential hypothesis test, in practice the average number of trees sampled to reach a stopping rule typically ranged between 5 and 10 trees.

We caution the reader to distinguish between the SRS tree-selection rule, which at each sequential step randomly selects a tree from all remaining trees over the entire orchard, and what might be referred to as a random tree-selection rule in which a pest manager walks through a grove haphazardly, randomly selecting trees to sample. Because this latter type of tree-selection rule does not sequentially select trees to be spaced far apart, our results from patterned tree-selection rules suggest that it will not mitigate spatial correlation sufficiently to apply Bartlett's sequential test.

## Results

**Illustrated Examples: Sample Parameter Estimates.** Various statistical software packages implement model fitting and parameter estimation for GLMMs, such as SAS Proc Glimmix. To provide realistic parameter estimates for *O. perseae* distributions in avocado orchards we fitted the spatial GLMM model, equation 3, to four presence-absence sets of data, with the fitted parameters provided in Table 2. Based on these estimates, in the simulation studies we allowed  $\sigma^2$  to vary from 0.5 to 2.0, and  $\rho$  to vary from 0 to 5.0.

**Leaf-Selection Rules: Outcome.** Fig. 4 shows ASN curves for the expected number of sampled trees, and Fig. 5 shows ASN curves for the expected total number

**Table 2.** Parameter estimates for four sets of *O. perseae* presence-absence data fitted to the spatial GLMM model, equation 3

Orchard	$n$	$m$	$\gamma$	$\sigma^2$	$\rho$
4	30	8	1.53	1.67	1.06
5	60	8	5.48	0.00024	0.073
8	400	4	3.24	1.23	4.37
10	402	4	-0.85	0.87	1.32

For each data set,  $n$  is the no. of trees and  $m$  is the no. of leaves sampled per tree. In each orchard, trees were approximately equally spaced on a grid, and in fitting the spatial GLMM distance is measured in tree-separation units. Note that in the second data set (orchard 5) mites were present on nearly every sampled leaf.

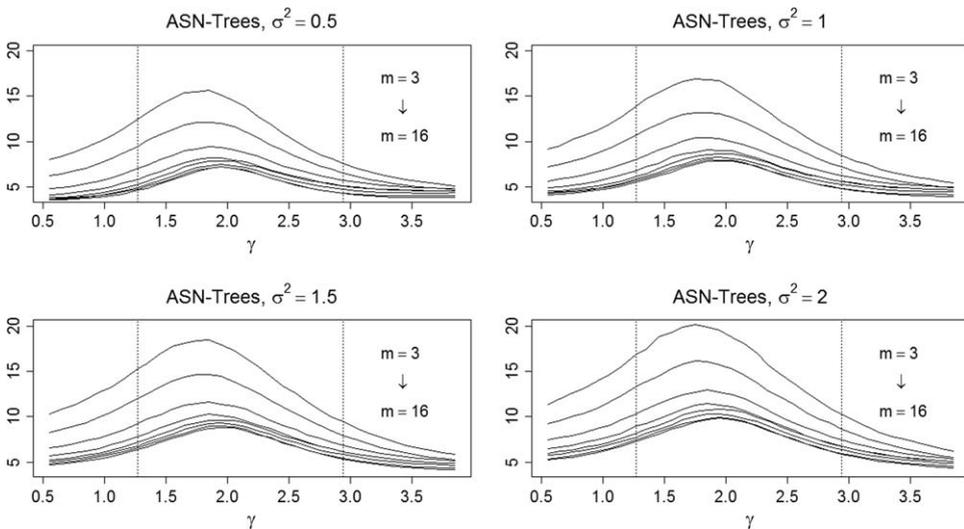
of sampled leaves, where each point was obtained using 20,000 simulations. We observe that as more leaves per tree were sampled (i.e., as  $m$  increases), one expects to sample fewer trees but more total leaves. The ideal choice for  $m$  minimizes the expected cost,  $E(\text{Cost})$ , that depends upon the leaf-to-tree cost ratio,  $c$ . In Fig. 6,  $E(\text{Cost})$  versus  $m$  is plotted for several values of  $c < 1$ ,  $c = 0.01, 0.10, 0.25$ , and  $0.50$ . We observe that as  $m$  increases beyond six leaves per tree the expected cost does not significantly decrease for smaller values of  $c$  and  $\sigma^2$ , and increases for larger values of  $c$  and  $\sigma^2$ . Thus, for *O. perseae* we conclude that, if spatial correlation has been effectively diminished, then an ideal leaf-selection rule for evaluating the hypothesis test in equation 5 that applies to a range of parameter values and leaf-to-tree sampling cost ratio values is to randomly select  $m = 6$  leaves per tree.

**Evaluation of Proposed Methodology: Outcome.**

Figs. 7 and 8 display the results of our simulation study, which show how the observed Type-1 and Type-2 errors vary in the truncated sequential hypothesis test as the strength of spatial correlation increases from 0 to 5.0. Each point was obtained using 20,000 simulations, and the percentage of simulations for which the

stopping rule was not reached after sampling 40 trees was negligibly small, never exceeding 1.5%. All of the patterned tree-selection rules show strong inflations of the observed Type-1, 2 error rates, from which we conclude that patterned tree-selection rules cannot be used in Bartlett’s SPRT if spatial correlation is present. Although the SRS tree-selection rule performs better than the patterned tree-selection rules, the sequential, maximin tree-selection rule outperforms all other tree-selection rules, preserving the 10% Type-2 error rate over the range of parameters tested, and preserving the 10% Type-1 error rate up to a spatial correlation strength of  $\rho = 2.0$ .

The estimates for the spatial correlation parameter reported in Li et al. (2012), based on count data, ranged from 0.24 to 1.55, so that a reasonable range of study for  $\rho$  was taken to be 0–2.0. Our presence-absence data analyses suggest allowing  $\rho$  to increase up to 5.0. More typically, we do not expect  $\rho$  to achieve values beyond 2.0, but this extended range was used to introduce robustness into our conclusions. In Figs. 7 and 8, as  $\rho$  ranges from 0 to 2.0 we see that our proposed methodology consistently preserves the Type-1 and Type-2 error rates. Even if the spatial correlation is as high as 5.0, the proposed methodology still preserves the Type-2 error rates, although the Type-1 error rates become slightly elevated. In the context of making a treatment decision based on an action threshold, making a Type-2 error corresponds to failing to treat an orchard for *O. perseae* when mite densities are  $>50$  mites/leaf and treatment is necessary, and making a Type-1 error corresponds to treating an orchard when mite densities are below 50 mites/leaf and treatment is unnecessary. Thus, using our proposed methodology, even under high levels of spatial correlation, a pest manager will not fail to treat a grove needing treatment, but may



**Fig. 4.** ASN curves for the expected number of sampled trees in Bartlett’s SPRT. The number of leaves sampled per tree,  $m$ , ranges between three (upper curve), 4, 6, 8, 10, 12, 14, and 16 (lower curve).

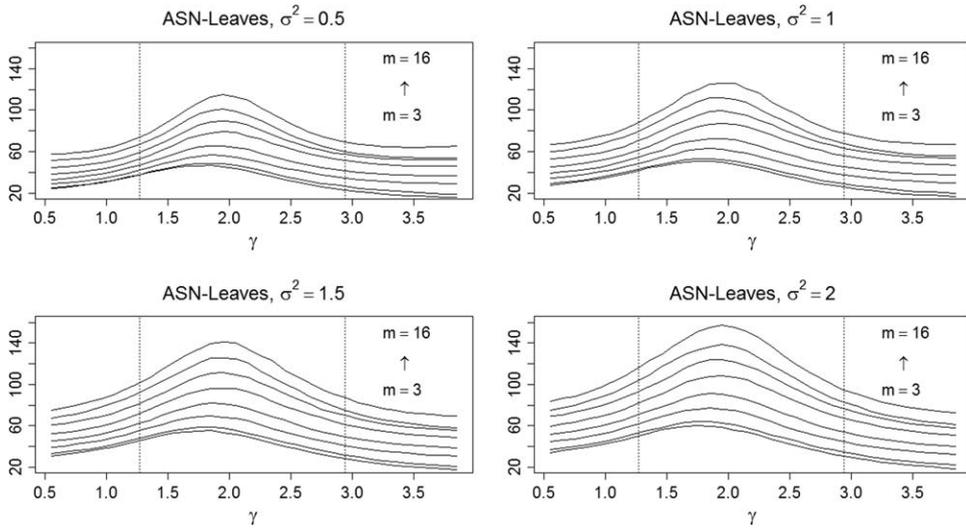


Fig. 5. ASN curves for the expected number of sampled leaves in Bartlett's SPRT. The number of leaves sampled per tree,  $m$ , ranges between three (lower curve), 4, 6, 8, 10, 12, 14, and 16 (upper curve).

conservatively treat a grove for which treatment is not required.

This simulation study confirms the effectiveness of the proposed methodology for the range of parameter values appropriate to *O. perseae*. In particular, the methodology proposed here eliminates the need for an initial pilot sample as suggested by Li et al. (2012).

Discussion

The ultimate purpose of developing a sampling plan is to provide an easy to use tool for pest managers to use to allow them to quickly and accurately reach decisions on whether or not avocado orchards need to

be treated for *O. perseae*, an important foliar mite pest of avocados in California, Mexico, Costa Rica, Spain, and Israel. Because a reliable sampling tool does not exist, integrated pest management (IPM) programs for *O. perseae* in California are relatively nonexistent and it is likely that numerous pesticide applications are applied annually for the control of this pest when they are not needed. Analysis of pesticide use trends in California avocados shows a remarkably rapid increase in pesticide applications after the invasion of *O. perseae* in 1990 (Hoddle 2004), and the adoption of a sampling plan similar to that proposed here may help reverse this trend by reducing the rate of unnecessary applications for this pest.

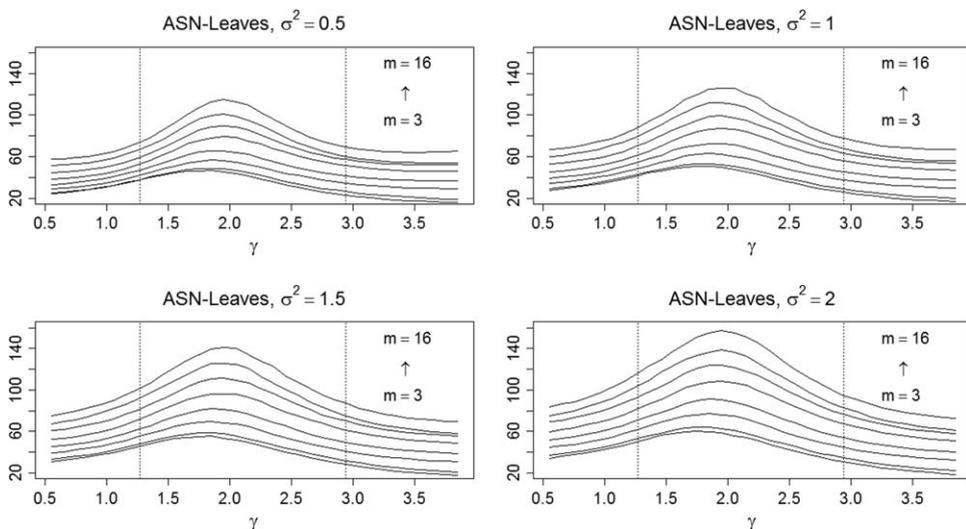


Fig. 6. Expected sampling cost versus the number of leaves selected per tree,  $m$ , where the leaf-tree sampling cost ratio,  $c = \frac{\text{cost per leaf}}{\text{cost per tree}}$ , ranges between 0.01 (lower curve), 0.10, 0.25, and 0.50 (upper curve).

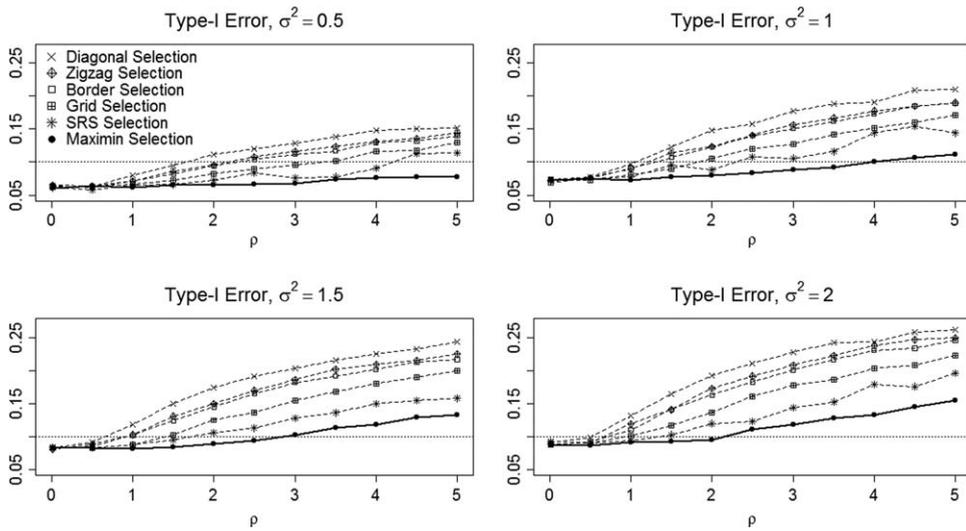


Fig. 7. Observed Type-1 error rates for Bartlett’s SPRT for data simulated with correlation parameter  $\rho$ . The solid curve corresponds to the sequential, maximin tree-selection rule, the dashed curves correspond to the patterned and SRS tree-selection rules, and the horizontal dotted line is the theoretical error rate upper bound of  $\alpha = 0.10$ .

The work presented here is the first statistical application of spatial analyses coupled with sequential sampling for the development of a sampling plan for pest management. Our proposed presence-absence sampling methodology for *O. perseae* evaluates a sequential hypothesis test of pest population densities which, 1) accounts for aggregation of pest populations on individual trees, and 2) mitigates spatial correlation of pest populations on adjacent trees using a tree-selection rule that sequentially selects trees to be maximally spaced from all other previously selected trees (sequential, maximin tree-selection). Based on a simulation study we determined that the expected

sampling cost is essentially minimized with a random selection of  $m = 6$  leaves per tree, and based on a separate simulation study of Bartlett’s SPRT with 10% Type-1, 2 error rates, we demonstrated that the sequential, maximin tree-selection rule preserves the error rates in the presence of spatial correlation, with average sample numbers for the sequential test being 5–10 trees. Although our results demonstrate the effectiveness of our presence-absence sampling methodology for parameter estimates relevant to *O. perseae*, the methodology can easily be applied to other pests, and even other nonpest spatial sampling situations. Furthermore, although it is not the focus of

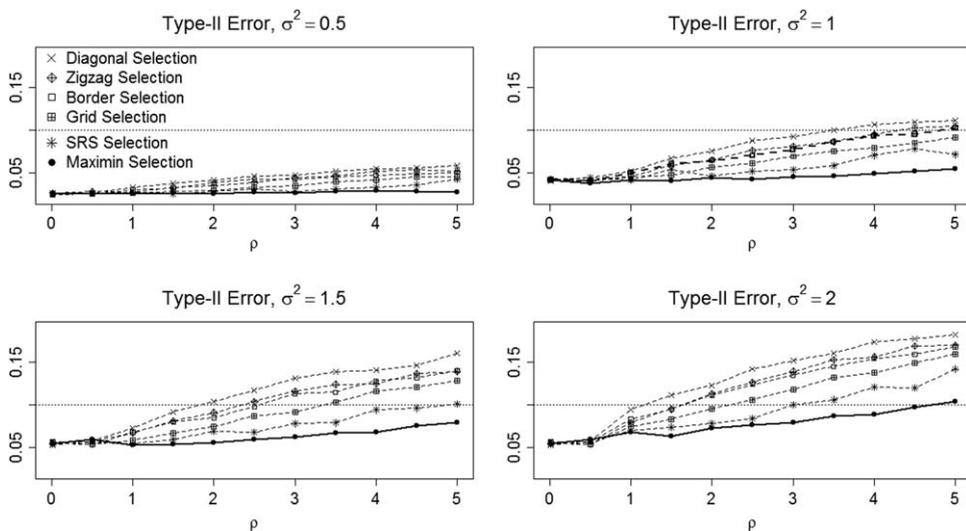


Fig. 8. Observed Type-2 error rates for Bartlett’s SPRT for data simulated with correlation parameter  $\rho$ . The solid curve corresponds to the sequential, maximin tree-selection rule, the dashed curves correspond to the patterned and SRS tree-selection rules, and the horizontal dotted line is the theoretical error rate upper bound of  $\beta = 0.10$ .

this article, we briefly point out that the spatial modeling and sample plan development may be relevant to other aspects of pest management such as sampling for plant diseases (van Maanen and Xu 2003, Kelly and Guo 2007). For example, Perring et al. (2001) and Groves et al. (2005) demonstrated the importance of spatial analyses for understanding the distribution of the disease-causing bacterium, *Xyllela fastidiosa*, which is vectored by cicadellids feeding on grapes and almonds. However, these studies do not directly address the manner in which sampling should be conducted. Consequently, the development of spatial models that provide an unbiased snapshot of incidence levels across sampled blocks, similar to that presented here, may have utility beyond sampling for aggregated populations of pest mites in orchards.

The results of the simulations conducted here demonstrate the effectiveness of our spatial presence-absence sampling methodology for parameter estimates relevant to *O. perseae* in California avocado orchards. With further research involving field validation, our sampling model has the potential to be customized as a reliable decision-making tool for pest control advisers and growers to use for control of this mite in commercial avocado orchards. To meet this goal, software would be needed to help a pest manager with tree selection and with evaluating the treatment decision hypothesis test at each sequential step. A component of any new technology is end-user adoption, especially if underlying concepts appear difficult and application potentially complicated. With the widespread ownership and use of smart phones, sampling programs like the one developed here could be made available as a downloadable "application." This has several major attractions for users: 1) by following simple sampling instructions on a screen (such as GPS directions to the next tree to sample) and punching in sampling data (yes or no for the presence or absence of *O. perseae* for each sampled leaf), user uncertainty about sampling methodology (both tree and leaf selections) and correct calculations and interpretation of outcomes are potentially minimized. 2) Smart phone applications would return management decisions in real time and can be immediately emailed to a supervisor. Photos and GPS coordinates generated by the smart phone could also be included in reports if extra details are useful for decision-making. 3) All sampling events have the potential to be archived electronically eliminating the need for expensive triplicate docket books and storage space for these article records. 4) Because the popularity of smart phone applications is increasing, a well-developed application that is attractive in appearance and easy to use may help greatly with the adoption of sampling plans, like that developed here for *O. perseae*, for IPM programs.

In this article we used a sampling cost function, equation 9, which includes a fixed cost for each tree sampled. Future work might include a more sophisticated per tree sampling cost that varies during the sequential sampling process to account for both the distance and the land topography between subse-

quently sampled trees, which may be of interest to a pest manager seeking to minimize their distance traveled and seeking to avoid sampling from trees that are difficult to reach (e.g., trees on steep hillsides). Additionally, the spatial GLMM model of pest populations that we used assumes that pest individuals are distributed randomly within a tree, and that correlations of pest populations on adjacent trees are spatially symmetric. Future research on sequential sampling with spatial components that extends beyond these model assumptions may address issues pertaining to pest populations that are systematically distributed within trees, and may include anisotropic (i.e., asymmetric) correlation structures of pest populations, allowing for stronger correlation along orchard edges or within orchard rows (see Ifoulis et al., 2006).

### Acknowledgments

This work was supported in part by a USDA-NIFA Regional Integrated Pest Management Competitive Grants Program Western Region Grant 2010-34103-21202 to MSH and DRJ. Financial support to JRL was provided, in part, by the University of California Riverside Cota Robles Fellowship and the Herbert Kraft Scholarship. The UC Hansen Trust and California Avocado Commission helped cover costs associated with field surveys. The owners of the avocado orchards in Carpinteria, CA, kindly provided unlimited access to study sites. Ruth Amrich, Allison Bistline, Mike Lewis, Naseem Saremi, Martin Castillo, John Jones, and Nick Salvato assisted with field collections and counting perseae mites in the lab.

### References Cited

- Alatawi, F. J., G. P. Opit, D. C. Margolies, and J. R. Nechols. 2005. Within-plant distribution of twospotted spider mites (Acari: Tetranychidae) on impatiens: development of a presence-absence sampling plan. *J. Econ. Entomol.* 98: 1040-1047.
- Bartlett, M. S. 1946. The large sample theory of sequential tests. *Proc. Cambridge Philos. Soc.* 42: 239-244.
- Bianchi, E.J.J.A., P. W. Goedhart, and J. M. Baveco. 2008. Enhanced pest control in cabbage crops near forest in The Netherlands. *Landsc. Ecol.* 23: 595-602.
- Breslow, N. E., and D. G. Clayton. 1993. Approximate Inference in Generalized Linear Mixed Models. *J. Am. Statistical Assoc.* 88: 9-25.
- Binns, M. R., J. P. Nyrop, and W. Van Der Werf. 2000. Sampling and monitoring in crop protection: the theoretical basis for developing practical decision guidelines. CABI Publishing, Wallingford, United Kingdom.
- (CAC) California Avocado Commission. 2009. Acreage inventory summary, 2009 update using remote sensing technology. California Avocado Commission. (<http://www.avocado.org/acreage-inventory-summary/>).
- (CAC) California Avocado Commission. 2010. 2009-10 annual report. California Avocado Commission. (<http://www.californiaavocadogrowers.com/annual-report/>).
- Candy, S. G. 2000. The application of generalized linear mixed models to multi-level sampling for insect population monitoring. *Environ. Ecol. Stat.* 7: 217-238.
- Galvan, T. L., E. C. Burkness, and W. D. Hutchison. 2007. Enumerative and binomial sequential sampling plans for the multicolored Asian lady beetle (Coleoptera: Coccinellidae) in wine grapes. *J. Econ. Entomol.* 100: 1000-1010.

- Gerrard, D. J., and H. C. Chaing. 1970. Density estimation of corn rootworm egg populations based upon frequency of occurrence. *Ecology* 51: 237–245.
- Gozé, E., S. Nibouche, and J. P. Deguine. 2003. Spatial and probability distribution of *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae) in cotton: systematic sampling, exact confidence intervals and sequential test. *Environ. Entomol.* 32: 1203–1210.
- Groves, R. L., J. Chen, E. L. Civerolo, M. W. Freeman, and M. A. Viveros. 2005. Spatial analysis of almond leaf scorch disease in the San Joaquin Valley of California: factors affecting pathogen distribution and spread. *Plant Dis.* 89: 581–589.
- Hall, D. G., C. C. Childers, and J. E. Eger. 2007. Binomial sampling to estimate rust mite (Acari: Eriophyidae) densities on orange fruit. *J. Econ. Entomol.* 100: 233–240.
- Hoddle, M. S. 2004. Invasions of leaf feeding arthropods: why are so many new pests attacking California-grown avocados? *Calif. Avocado Soc. Yrbk.* 87: 65–81.
- Hoddle, M. S., L. Robinson, and J. Virzi. 2000. Biological control of *Oligonychus perseae* on avocado: III. Evaluating the efficacy of varying release rates and release frequency of *Neoseiulus californicus* (Acari: Phytoseiidae). *Int. J. Acarol.* 26: 203–214.
- Humeres, E. C., and J. G. Morse. 2005. Baseline susceptibility of persea mite (Acari: Tetranychidae) to abamectin and milbemectin in avocado groves in southern California. *Exp. Appl. Acarol.* 36: 51–59.
- Hyung, D. L., J. J. Park, J.-H. Lee, K.-I. Shin, and K. Cho. 2007. Evaluation of binomial sequential classification sampling plan for leafmines of *Liriomyza trifoli* (Diptera: Agromyzidae) in greenhouse tomatoes. *Int. J. Pest Manag.* 53: 59–67.
- Ifoulis, A. A., and M. Savopoulou-Soultani. 2006. Use of geostatistical analysis to characterize the distribution of *Lobesia botrana* (Lepidoptera: Tortricidae) larvae in northern Greece. *Environ. Entomol.* 35: 497–506.
- Johnson, M. E., L. M. Moore, and D. Ylvisaker. 1990. Minimax and maximin distance designs. *J. Stat. Plann. Inference* 26: 131–148.
- Kabaluk, J. T., M. R. Binns, and R. S. Vernon. 2006. Operating characteristics of full count and binomial sampling plans for green peach aphid (Hemiptera: Aphididae) in potato. *J. Econ. Entomol.* 99: 987–992.
- Kelly, M., and Q. Guo. 2007. Integrated agricultural pest management through remote sensing and spatial analyses, pp. 191–207. *In* A. Ciancio and K. G. Mukerji (eds.), *General concepts in integrated pest and disease management*. Springer, Dordrecht, The Netherlands.
- Kono, T., and T. Sugino. 1958. On the estimation of the density of rice stems infested by the rice stem borer. *Jpn. J. Appl. Entomol. Zool.* 2: 184–188.
- Li, J. X., D. R. Jeske, J. L. Lara, and M. S. Hoddle. 2012. Sequential hypothesis testing with spatially correlated count data. *Integration: Math. Theory Appl.* 2: 269–284.
- Martinez-Ferrer, M. T., J. L. Ripolles, and F. Garcia-Mari. 2006. Enumerative and binomial sampling plans for citrus mealybug (Homoptera: Pseudococcidae) in citrus groves. *J. Econ. Entomol.* 99: 993–1001.
- Moaz, Y., S. Gal, M. Zilberstein, Y. Izhar, V. Alchanatis, M. Coll, and E. Palevsky. 2011. Determining an economic injury level for the persea mite, *Oligonychus perseae*, a new pest of avocado in Israel. *Entomol. Exp. Appl.* 138: 110–116.
- Mulekar, M. S., L. J. Young, and J. H. Young. 1993. Testing insect population density relative to critical densities with 2-SPRT. *Environ. Entomol.* 22: 346–351.
- Müller, W. 2007. *Collecting spatial data: optimum design of experiments for random fields*. Springer, Berlin, Germany.
- Perring, T. M., C. A. Farrar, and M. J. Blua. 2001. Proximity to citrus influences Pierce's disease in Temecula valley vineyards. *Calif. Agric.* 55: 13–18.
- Ramírez-Dávila, J. F., and E. Porcayo-Camargo. 2008. Spatial distribution of the nymphs of *Jacobiasca lybica* (Hemiptera: Cicadellidae) in a vineyard in Andalucía, Spain. *Rev. Colomb. Entomol.* 34: 169–175.
- Robson, J. D., M. G. Wright, and R.P.P. Almeida. 2006. Within-plant distribution and binomial sampling of *Pentalonia nigronervosa* (Hemiptera: Aphididae) on banana. *J. Econ. Entomol.* 99: 2185–2190.
- Schabenberger, O., and C. A. Gotway. 2005. *Statistical methods for spatial data analysis*. Chapman & Hall/CRC, Boca Raton, FL.
- Schotzko, D. J., and L. E. Okeeffe. 1989. Geostatistical description of the spatial-distribution of *Lygus hesperus* (Heteroptera: Miridae) in lentils. *J. Econ. Entomol.* 82: 1277–1288.
- Shah, P., D. R. Jeske, and R. Luck. 2009. Sequential hypothesis testing techniques for pest count models with nuisance parameters. *J. Econ. Entomol.* 102: 1970–1976.
- Takakura, K. 2009. Reconsiderations on evaluating methodology of repellent effects: validation of indices and statistical analyses. *J. Econ. Entomol.* 102: 1977–1984.
- Van Maanen, A., and X. M. Xu. 2003. Modelling plant disease epidemics. *Eur. J. Plant Pathol.* 109: 669–682.
- Wald, A. 1947. *Sequential analysis*. Wiley, New York.
- Young, L. J., and J. H. Young. 1998. *Statistical ecology*. Kluwer Academic Publishers, Boston.

Received 18 June 2011; accepted 11 February 2012.