

Methods for efficient estimation of rust incidence in coffee plantations^{1,2}

Raúl E. Macchiavelli³ and Rocío del P. Rodríguez⁴

J. Agric. Univ. P.R. 84(1-2):65-78 (2000)

ABSTRACT

Two systematic patterns were evaluated for efficiency in estimating coffee rust incidence in lots of Caturra and Bourbon. True disease incidence was determined in the middle third of each tree in the selected lot. The different sampling patterns were compared on the basis of the bias and standard error of the estimated incidences obtained with computer simulations of different combinations of sample sizes, average incidence and patterns of spatial dispersion for the disease. The parallel row pattern was at least as efficient as the traditionally recommended W pattern. In order to attain better estimates of coffee rust incidence, it seems more important to sample more leaves per tree than more trees per field. The recommended protocol for estimating rust incidence in coffee plantations is described.

Key words: *Coffea arabica*, *Hemileia vastatrix*, spatial dispersion, systematic sampling

RESUMEN

Método para estimar con eficiencia la incidencia de roya en plantaciones de cafetos

En lotes de las variedades Caturra y Bourbon se evaluaron dos patrones sistemáticos de muestreo para determinar cuál es el más eficiente para estimar la incidencia de la roya del café. Se determinó la incidencia actual de la enfermedad en el tercio medio de cada árbol del lote seleccionado. La comparación de cada patrón se basó en el sesgo y el error estándar de las incidencias estimadas obtenidas mediante simulación bajo diferentes combinaciones de tamaño de muestra, incidencia promedio y distribución espacial de incidencia. El patrón sistemático de hileras paralelas fue tanto o más eficiente que el plan tradicionalmente recomendado en forma de W. Para obtener mejores estimadores parece más importante el número de hojas por árbol que el número de árboles por lote. Se describe el protocolo recomendado para estimar la incidencia de roya en los cafetales.

INTRODUCTION

Coffee is the major traditional crop in Puerto Rico. During 1997-98, the contribution of traditional crops to agricultural gross income was

¹Manuscript submitted to Editorial Board 17 August 1999.

²The authors thank Agronomists Wigmar González and Osvaldo Bosques for their assistance in collecting the data, and Dr. James Beaver for his helpful comments.

³Associate Professor, Agronomy and Soils Department, University of Puerto Rico-Mayagüez, P.O. Box 9030, Mayagüez, P.R. 00681-9030.

⁴Researcher, Crop Protection Department.

\$57,446,000 (Departamento de Agricultura, 1998). Of this overall income, the coffee crop contributed 93.69%. The value of coffee to the economy of the island lies not only in its direct economic contribution but also in that by employing 9,000 to 11,000 laborers in the central rural areas (Monroig-Inglés, 1988), it sustains 9% of the population.

Rust caused by *Hemileia vastatrix* is an important disease of coffee that was detected in Puerto Rico in 1989. The activities of this fungus in the plant reduce coffee berry production and yield by affecting photosynthesis, inducing defoliation and reducing plant vigor (Kushalappa, 1989). Chemical control is one of the strategies for managing the disease and several fungicide formulations have been identified as effective in reducing disease incidence. The regime for fungicide application is based primarily on the epidemic onset regardless of the true disease incidence at the moment. In addition, rust is distributed in foci and not all areas in the coffee plantation are, or will be, affected with the same intensity. Because of this uneven distribution and the fact that disease incidence varies with the years, the decision to use fungicides in an integrated disease management program should be based on disease levels in specific months and at specific sites within the plantation.

According to the procedure described by Hashizume et al. (1975) (cited by Kushalappa, 1989), disease incidence is determined by sampling leaves in the middle third of the coffee tree. This method implies detaching the leaves and determining which of those are rusted or healthy. Since it is not feasible to evaluate all leaves and all trees in the field, it is necessary to know the number of leaves and trees to be sampled to obtain a reliable estimate of the disease incidence at specific sites of the plantation. Another issue is to decide which sampling pattern to use, given the spatial distribution of this disease in the field. The systematic pattern in a W form covering the field is traditionally recommended for pathogens with aggregation tendencies (Campbell and Madden, 1990). For Puerto Rico this methodology is highly impractical because coffee is grown in areas with pronounced slopes. Thus, for the person doing the monitoring, walking diagonally along the slopes, it is almost impossible to keep track of the systematic W pattern. The objective of this study was to determine the most feasible pattern for collecting samples and how many trees and leaves per tree to sample to obtain efficient estimates of rust incidence in coffee fields.

MATERIALS AND METHODS

Two coffee lots, approximately seven years old, at the Adjuntas Experiment Station were selected for this study. Both fields were typical of the coffee growing areas: pronounced slopes with rows of trees par-

allel to the slope and wide variability in disease incidence. Planting distance in the Caturra lot was 1.83×1.22 m and in the Bourbon lot 3.05×1.22 m. Given the complex structure of the spatial dispersion of the rust disease, the problem was approached through simulations using the SAS system (SAS, 1990).

The response of interest in the monitoring is the "disease index", δ , for each tree:

$$\delta = \frac{\text{number of infected leaves in the middle third of the tree}}{\text{total number of leaves in the middle third of the tree}} \quad (1:1)$$

Rust incidence of the lot will be based on the disease index averaged over all trees in the lot, $\bar{\delta}$. In the two-step sampling procedure (n trees sampled out of N trees in the field, and m leaves sampled in each tree out of M leaves in the i -th tree), the estimator of $\bar{\delta}$ is:

$$\bar{d} = \frac{\sum_i d_i}{n}, \quad (1:2)$$

where d_i is the sample disease index for the i -th tree.

In order to estimate the variance of \bar{d} , the formula presented in Steel and Torrie (1980) for the two-stage finite population sampling with binomial-type response can be modified to account for the different number of leaves per tree:

$$\text{Var}(\bar{d}) = \frac{1}{Nn(m-1)} \sum_{i=1}^n \frac{M_i - m}{M_i - 1} d_i(1 - d_i) + \frac{N-n}{Nn} \sum_{i=1}^n \frac{(d_i - \bar{d})^2}{n-1}. \quad (1:3)$$

Since the first stage in the sampling designs to be used in this study is systematic, this formula is only an approximation.

Number of leaves per tree

To select a number of leaves per tree which could yield good estimates of $\bar{\delta}$, fifty trees from each coffee variety were randomly chosen. Since the disease is usually concentrated in the middle third of the coffee tree, all leaves in this area were counted and diagnosed for symptoms of rust. Hence the true disease incidence was known for each of the 50 trees of each variety. To prevent unnecessary damage and stress to the trees, the evaluation was done without detaching the leaves. Using a SAS macro, we simulated sampling between 10 and 100 leaves per tree, and 20, 30, 40 and 50 trees. For each combination of number of trees and number of leaves per tree the simulations were repeated 500 times, computing the mean and the standard error of \bar{d} .

Sampling plan

All trees in the Caturra lot ($N = 1269$) were sampled. From the middle third of each tree, 40 leaves were cut and evaluated for disease incidence (this number of leaves per tree was determined from the previous simulations to yield a good approximation of the actual disease index, known for the 50 trees sampled before). Hence, the “true” disease incidence of each tree in the lot was known. Two systematic patterns were selected for the simulation: the traditional W and the more feasible pattern of parallel rows. In the W pattern, trees were selected along a row (every c trees, with c varying from 2 to 6) up to one extreme, then diagonally every c trees up to the other extreme, then diagonally and along a row every c trees in the opposite direction. In the pattern of parallel rows, trees were selected along four equally spaced parallel rows, every c trees, with c varying from 2 to 7. Simulations were carried out in SAS for sampling 2, 5, 10, 30, and 40 leaves per tree every c trees. For each of the 66 cases we repeated the simulations 500 times, and computed the mean and the standard error of \bar{d} .

Spatial distribution

The spatial distribution pattern of the disease index in the Caturra lot was studied. Campbell and Madden (1990) recommend the use of spatial autocorrelation models to describe the spatial dispersion. We applied different correlation models to our data and used Akaike’s information criterion (AIC) (Akaike, 1974) to select the best simple model explaining the correlation structure of the data. The MIXED procedure in SAS (1997) with the following correlation functions was used:

1. Exponential:

$$\text{Cov}(d_i, d_j) = \sigma^2 \exp\left(\frac{-\text{dist}(i, j)}{\rho}\right), \quad (1:4)$$

where $\text{dist}(i, j)$ is the distance, in meters, between trees i and j .

2. Anisotropic Exponential:

$$\text{Cov}(d_i, d_j) = \sigma^2 \exp(-\theta_x(\text{dist}_x(i, j))^{\rho_x}) \exp(-\theta_y(\text{dist}_y(i, j))^{\rho_y}), \quad (1:5)$$

where $\text{dist}_x(i, j)$ and $\text{dist}_y(i, j)$ are the distances (m) between trees i and j along the row and across rows, respectively.

3. Power:

$$\text{Cov}(d_i, d_j) = \sigma^2 \rho^{\text{dist}(i, j)}, \quad (1:6)$$

4. Anisotropic Power:

$$\text{Cov}(d_i, d_j) = \sigma^2 \rho_x^{\text{dist}_x(i,j)} \rho_y^{\text{dist}_y(i,j)}. \quad (1:7)$$

5. Spherical:

$$\text{Cov}(d_i, d_j) = \sigma^2 \left(1 - \frac{3\text{dist}(i,j)}{2\rho} + \frac{\text{dist}^3(i,j)}{2\rho^3} \right) I(\text{dist}(i,j) \leq \rho). \quad (1:8)$$

These same five structures were fitted with a “nugget” effect. This parameter represents additional variation between observations at the same place (i.e., small-scale variability). It is an additive parameter included in each model when the distance is 0. Thus, it increases the variance of the observations, $\text{Cov}(d_i, d_i)$, but not the covariance.

Robustness of the method

Using the best spatial dispersion structure (chosen according to AIC in the previous section) to simulate different scenarios under which the performance of the proposed plan can be validated, we considered four different cases representing changes in the average incidence and/or the spatial correlation. Case 1 had the same spatial correlation but larger incidence ($\bar{\delta} = 0.25$). Case 2 had the same spatial correlation and smaller incidence ($\bar{\delta} = 0.15$). Case 3 had the same incidence ($\bar{\delta} = 0.19$) and larger spatial dependence (nugget effect = 0.0085, half of the value estimated in the last section). Case 4 had the same incidence ($\bar{\delta} = 0.19$) and smaller spatial dependence ($\theta_x = 0.18, \theta_y = 0.27$).

Once the data sets were generated under the corresponding model, simulations were done using the same approach described before. The two systematic patterns under study (W and parallel rows) were used and sampling was simulated for each case every c trees (with c varying from 2 to 7). For each situation 2, 5, 10, 20, 30 and 40 leaves per tree were taken, and each of the 144 cases was iterated independently 500 times. In each case the mean and the standard error of \bar{d} were computed.

RESULTS AND DISCUSSION

For both cultivars a large variability in disease incidence was detected. Rust incidence varied from 0 to 80% infected leaves. The first step for the identification of a viable methodology to estimate incidence of coffee rust efficiently is to determine the number of leaves to be sampled. The results were consistent for both cultivars as well as for all values of n studied (Figures 1 and 2). When only a few leaves per tree

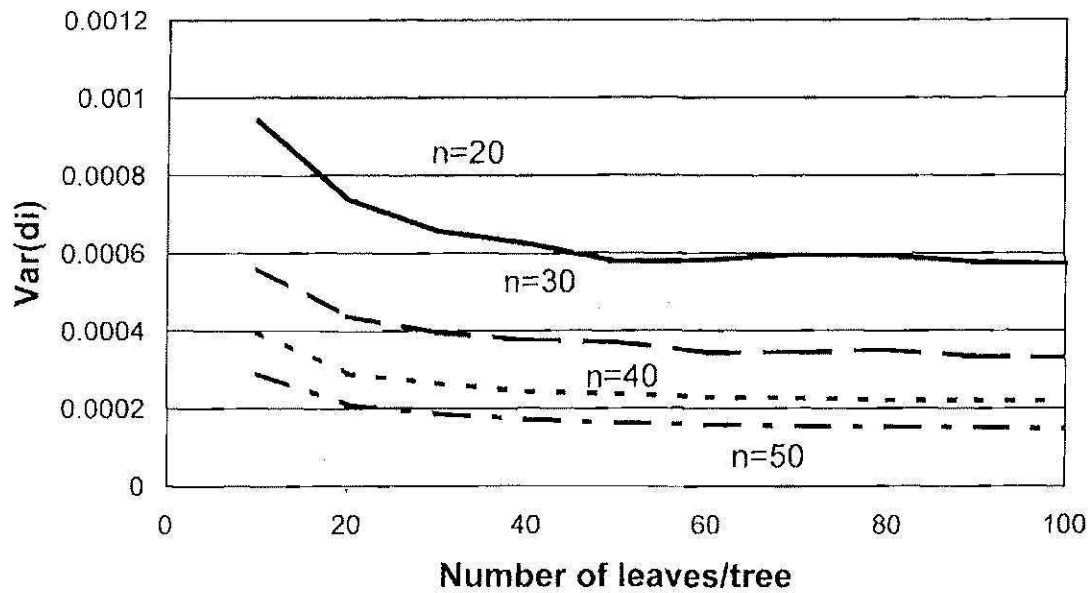


FIGURE 1. Variance of the estimated disease index in cultivar Caturra (n = number of trees).

were sampled the variance was large; then it decreased as m increased and finally it stabilized (the rate of decrease was very small). Forty leaves per tree yielded a reliable approximation to the true δ , and additional leaves increased the precision of the estimate very little. Using

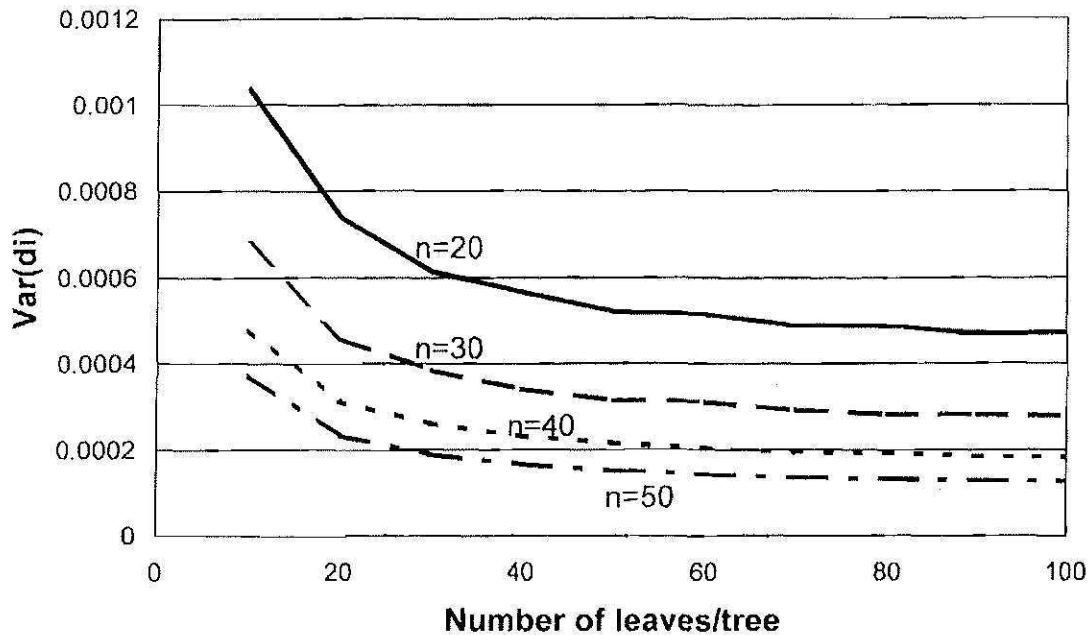


FIGURE 2. Variance of the estimated disease index in cultivar Bourbon (n = number of trees).

40 leaves as an approximation of sampling all leaves in the middle third, we proceeded to study the patterns for sampling.

The average disease index in the Caturra lot was 0.1939, with a standard deviation of 0.1742. The means of \bar{d} indicated that the W pattern overestimates the true average disease index (0.1939), since 90% of the means was greater than the true value. The means observed in the simulation ranged from 0.185 to 0.244. With the parallel row pattern, the means ranged from 0.177 to 0.210 with essentially half of the values (53%) above the true disease index. The standard errors obtained for sampling in parallel rows or in a W pattern are comparable (Figures 3 and 4). It was found that sampling 10 or 20 leaves per tree for a total of approximately 20 trees (equivalent to sampling every sixth tree in either pattern) attained acceptable standard errors (<0.1).

In order to make more general recommendations for the monitoring of coffee rust we studied the spatial distribution pattern of the disease index in the Caturra lot. The most appropriate structure is the anisotropic exponential model with nugget, since this model had the highest AIC value (Table 1). The estimated covariance between any two points in this model was:

$$\begin{aligned} \text{Cov}(d_i, d_j) = & \\ 0.0128 \exp(-0.070(\text{dist}_x(i, j))^{0.7668}) \exp(-0.3050(\text{dist}_y(i, j))^{0.5564}) + & \\ 0.0174 I(\text{dist}(i, j) = 0). & \end{aligned}$$

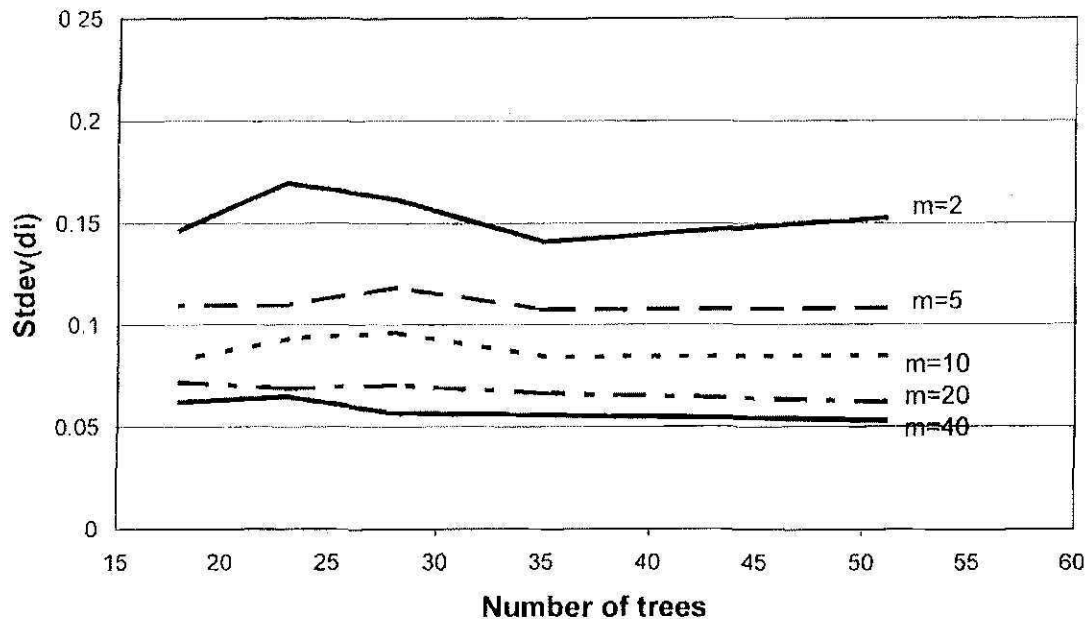


FIGURE 3. Standard errors of estimated disease index. Samples in parallel rows (m = number of leaves per tree).

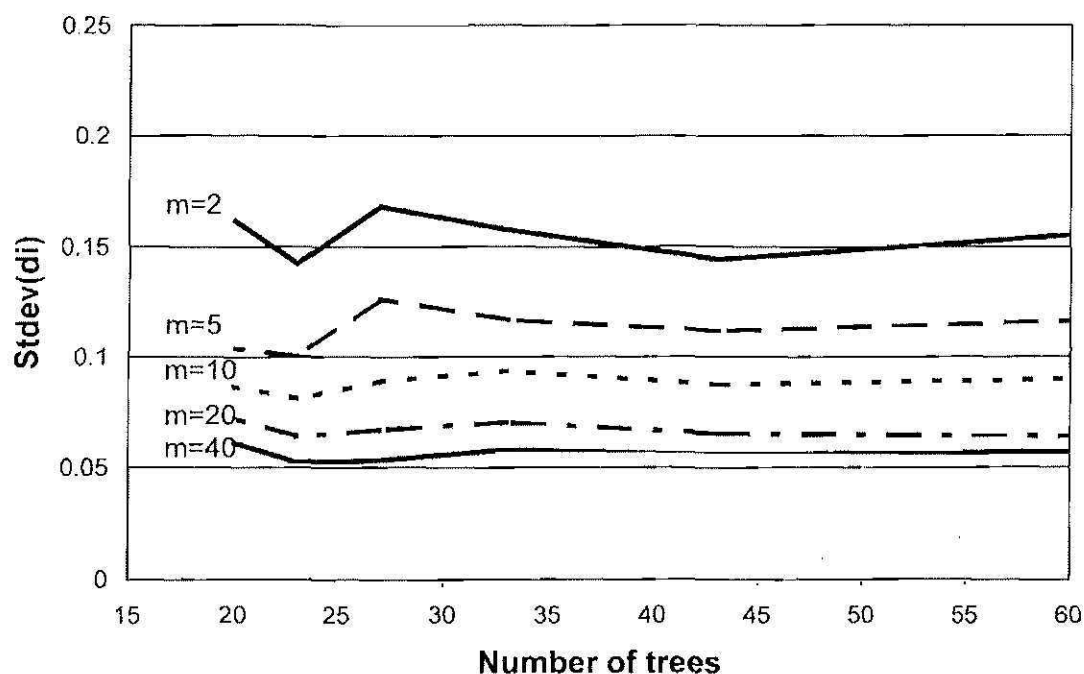


FIGURE 4. Standard errors of estimated disease index. Samples in W pattern (m = number of leaves per tree).

This equation indicates that there is a stronger correlation along the row (i.e., along the x -coordinate) than across rows (i.e., along the y -coordinate). This finding was also evidenced by the fitted correlations between any two observations at a given distance (Figure 5).

An estimate of the variance of \bar{d} can be obtained from the following formula:

$$\text{V}\hat{\text{a}}\text{r}(\bar{d}) = \frac{1}{n^2} \left[\sum_i \text{V}\hat{\text{a}}\text{r}(d_i) + 2 \sum_{i < j} \text{C}\hat{\text{o}}\text{v}(d_i, d_j) \right]. \quad (1.9)$$

Since the proposed sampling plans were based on a particular coffee lot, it is of interest to see how these plans could perform under different

TABLE 1.—Values of Akaike's information criterion (AIC) for fitting different covariance structures.

Covariance Structure	With no nugget	With nugget
Exponential	566.1	580.2
Anisotropic Exponential	608.5	620.9
Power	566.1	580.1
Anisotropic Power	544.1	565.4
Spherical	478.5	489.6

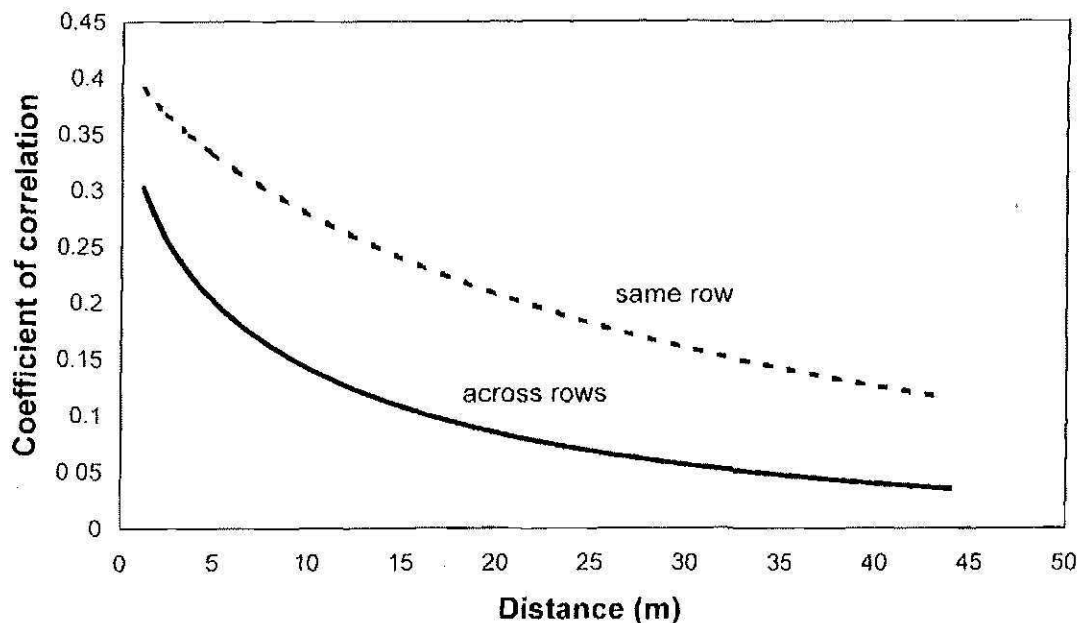


FIGURE 5. Correlations between disease indices of trees at a given distance.

conditions. Computer simulations showed that in all cases the parallel and W patterns gave similar standard errors (Figures 6 to 13). The only situation in which the W pattern performed better than the parallel row pattern was case 3, with strong spatial correlation. From these results it seems more important to sample more leaves per tree than

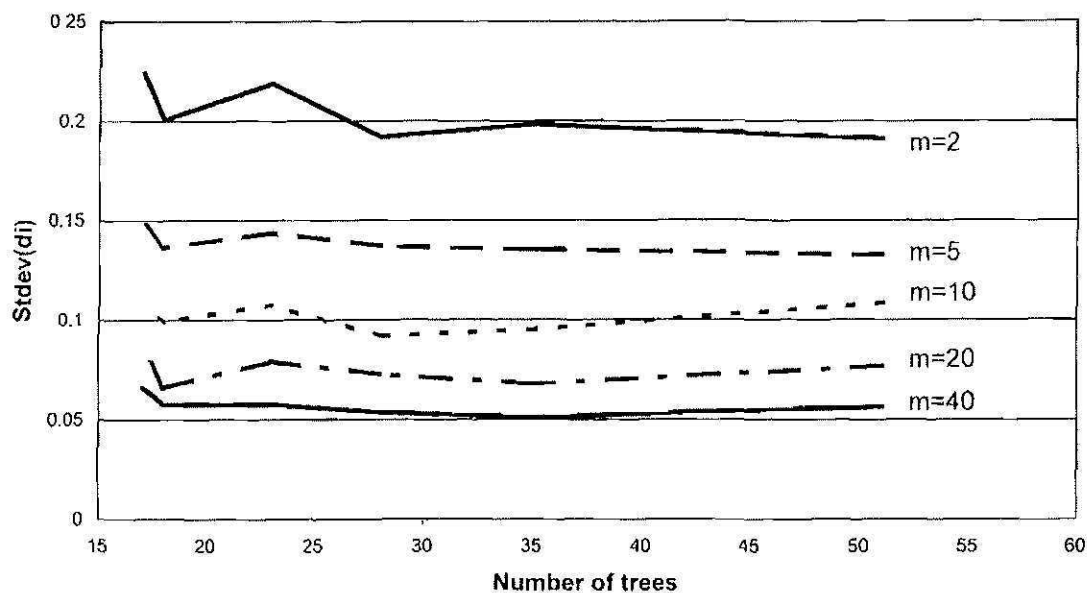


FIGURE 6. Standard errors of estimated disease index for samples in a W pattern. Simulated case 1 (m = number of leaves per tree).

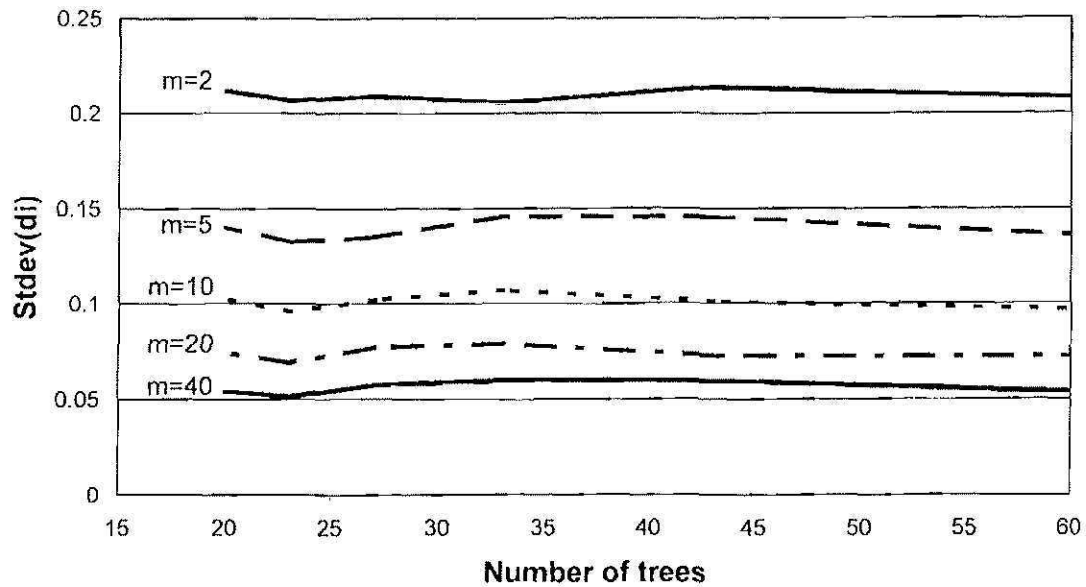


FIGURE 7. Standard errors of estimated disease index for samples in parallel rows. Simulated case 1 (m = number of leaves per tree).

more trees per field; hence the previous finding of sampling 20 leaves per tree, selecting every sixth tree in either pattern, seems reasonable.

For patterns of spatial dispersion similar to the ones found in the Caturra lot studied here, the systematic sampling in parallel rows performs better than in W pattern, because it yields estimates of average

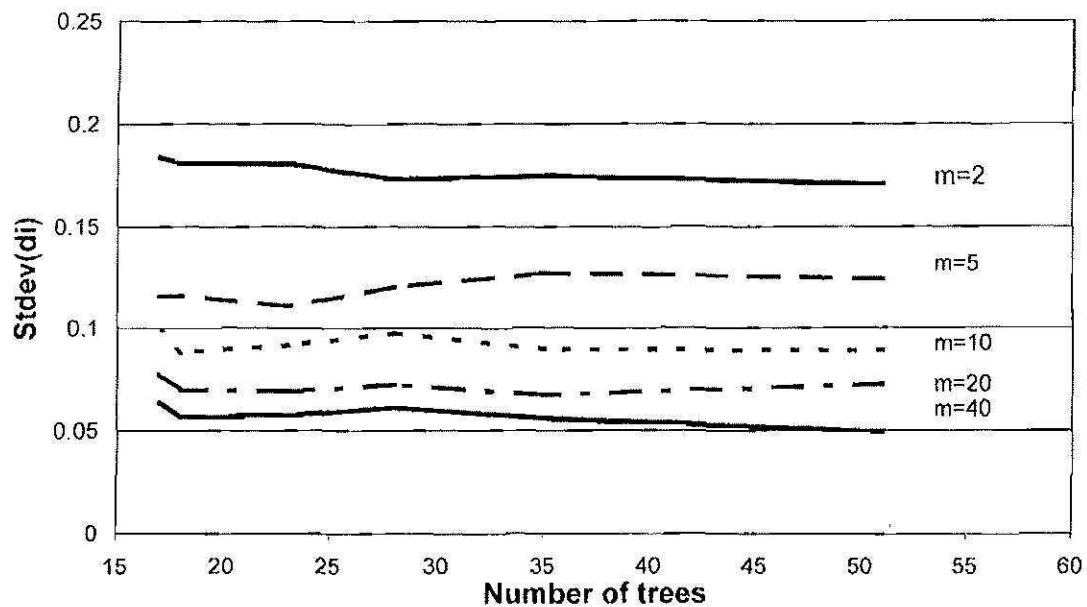


FIGURE 8. Standard errors of estimated disease index for samples in a W pattern. Simulated case 2 (m = number of leaves per tree).

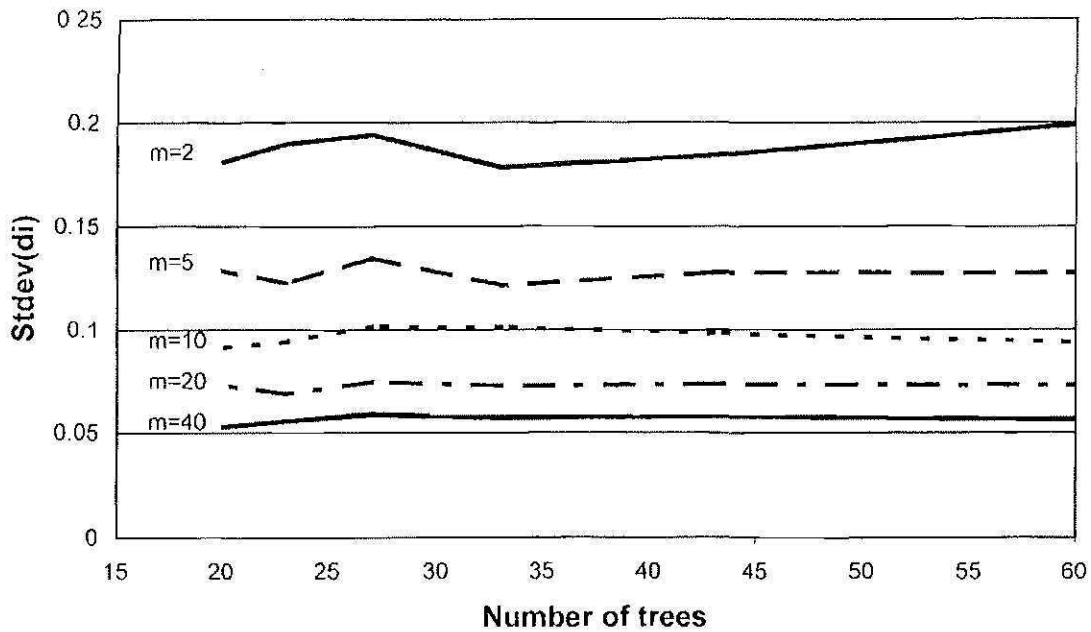


FIGURE 9. Standard errors of estimated disease index for samples in parallel rows. Simulated case 2 (m = number of leaves per tree).

disease index closer to the true disease index of the lot, and the standard errors are similar. It is recommended that trees be sampled along parallel rows, the first row chosen randomly among the first 10 rows in the field lot (these 10 rows represent approximately one-fourth of the

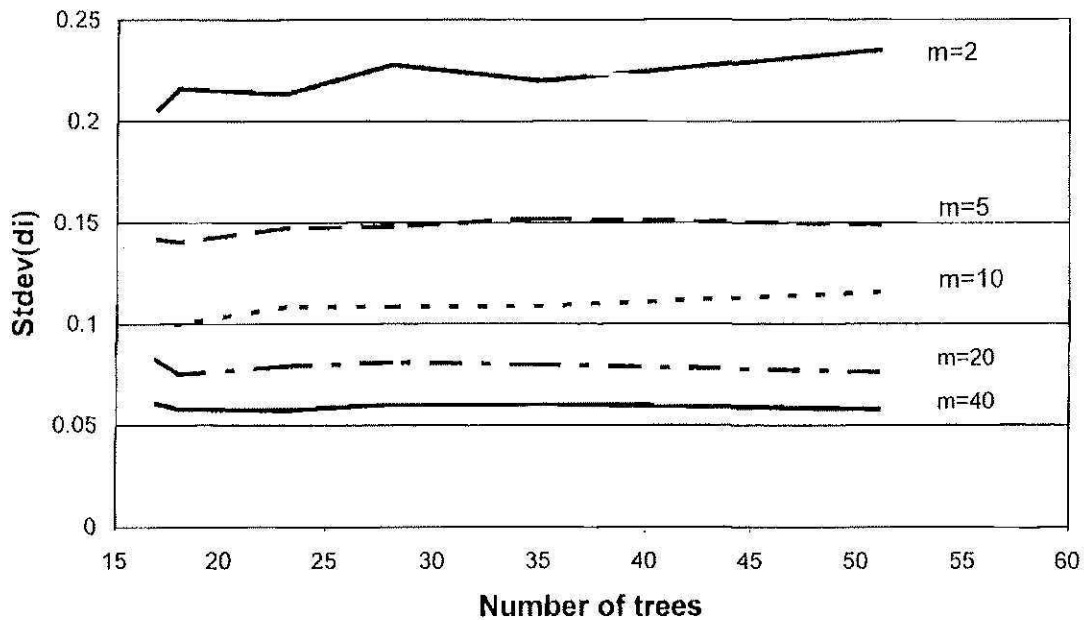


FIGURE 10. Standard errors of estimated disease index for samples in a W pattern. Simulated case 3 (m = number of leaves per tree).

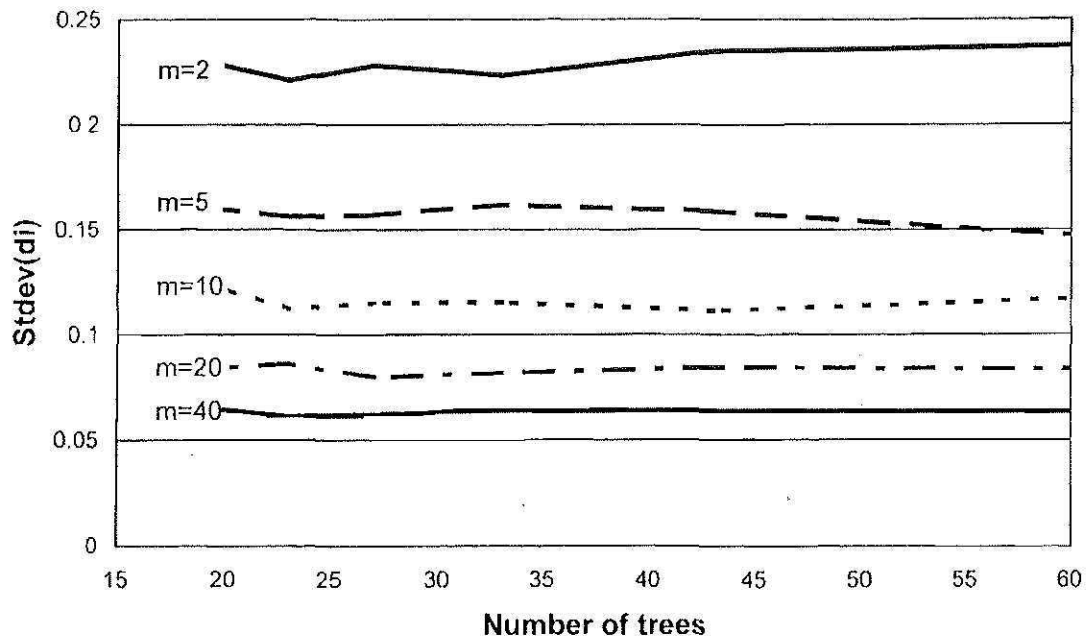


FIGURE 11. Standard errors of estimated disease index for samples in parallel rows. Simulated case 3 (m = number of leaves per tree).

available rows). Take 20 leaves from the middle third of every sixth tree along the row. Each selected row should be separated from the previous one by 10 rows (again approximately one-fourth of the available rows).

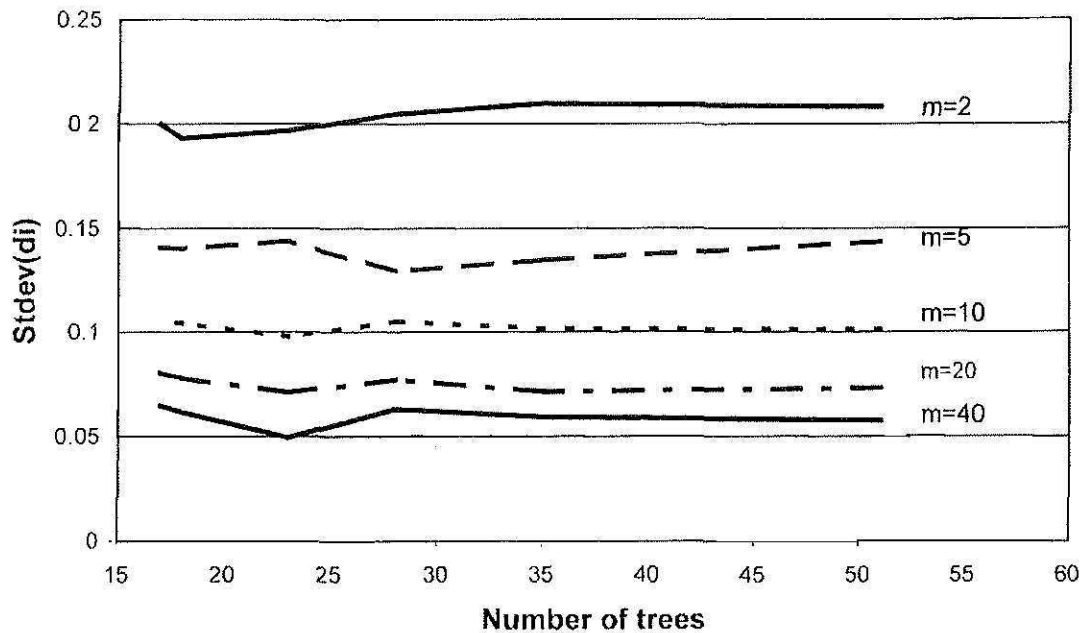


FIGURE 12. Standard errors of estimated disease index for samples in a W pattern. Simulated case 4 (m = number of leaves per tree).

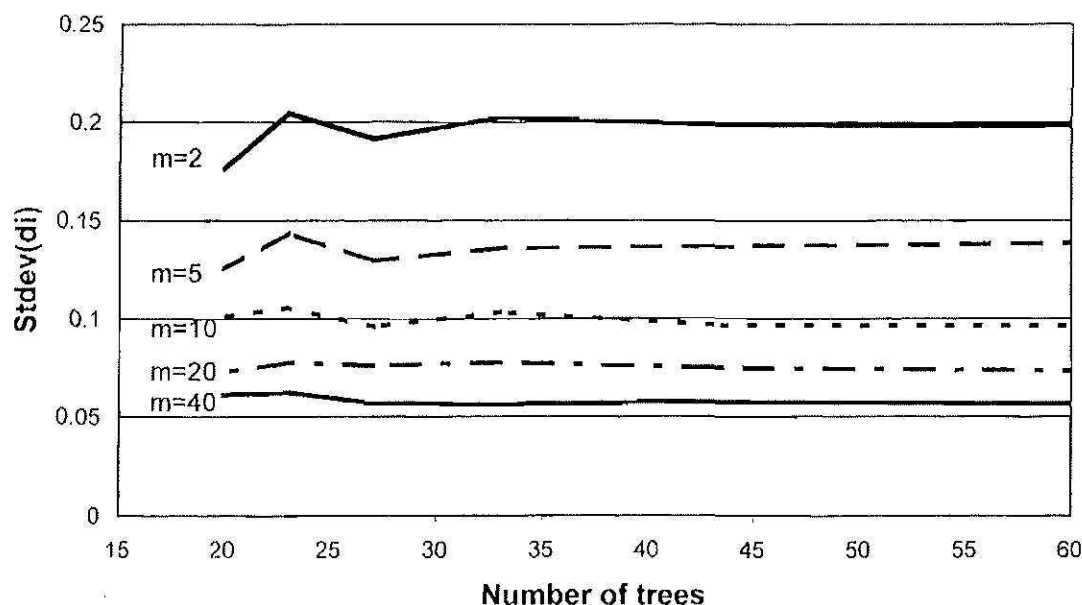


FIGURE 13. Standard errors of estimated disease index for samples in parallel rows. Simulated case 4 (m = number of leaves per tree).

This plan should select approximately 20 trees in a lot of 1,269 trees (about one-sixtieth of the overall population of trees).

The estimated standard errors, calculated with formula 1:3, were often smaller than the actual standard errors of \bar{d} (calculated from the 500 values of \bar{d} obtained in each case). This finding could be due to the fact that this formula was developed for random sampling, whereas our plans are systematic and the dispersion of the disease is non-random. Some formulas developed for systematic sampling could be introduced (Wolter, 1984), but they may need to be modified to account for the non-random dispersion. Standard errors calculated with the covariance model, estimated on the basis of the anisotropic exponential model with nugget, were generally larger than the ones calculated from the 500 realizations of \bar{d} . In order to find better estimates of standard errors, a bootstrap method may be necessary, given that the incidence of this disease almost never presents random dispersion, since it is distributed in foci.

LITERATURE CITED

- Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Transaction on Automatic Control*, AC-19, 716-723.
- Campbell, C. L. and L. V. Madden, 1990. *Introduction to Plant Disease Epidemiology*. John Wiley and Sons, Inc., NY.
- Departamento de Agricultura, 1998. Ingreso Agrícola. Cifras revisadas para 1997-98. Oficina de Estadísticas Agrícolas, Departamento de Agricultura de Puerto Rico.

- Kushalappa, A. C., 1989. Biology and epidemiology. p. 13-80. *In* A.C. Kushalappa and A. B. Eskes (eds.). *Coffee Rust: Epidemiology, Resistance and Management*. CRC Press.
- Monroig-Inglés, M., 1988. *Prácticas Modernas en el Cultivo del Café en Puerto Rico*. SEA, Colegio de Ciencias Agrícolas, Recinto Universitario de Mayagüez, Mayagüez, PR. Banco Santander de Puerto Rico.
- SAS Institute, Inc., 1990. *SAS Guide to Macro Processing*. Version 6, 2nd Edition. Cary, NC.
- SAS Institute, Inc., 1997. *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC.
- Steel, R. and J. Torrie, 1980. *Principles and Procedures of Statistics: A Biometrical Approach*. 2nd Edition. McGraw-Hill, NY.
- Wolter, K., 1984. *Introduction to Variance Estimation*. Springer-Verlag, NY.