

Efficiency of Adaptive Cluster Sampling for Estimating Plant Disease Incidence

P. S. Ojiambo and H. Scherm

First author: Department of Plant Pathology, North Carolina State University, Raleigh 27695; and second author: Department of Plant Pathology, University of Georgia, Athens 30602.
Accepted for publication 10 March 2010.

ABSTRACT

Ojiambo, P. S., and Scherm, H. 2010. Efficiency of adaptive cluster sampling for estimating plant disease incidence. *Phytopathology* 100:663-670.

Conventional sampling designs such as simple random sampling (SRS) tend to be inefficient when assessing rare and highly clustered populations because most of the time is spent evaluating empty quadrats, leading to high error variances and high cost. In previous studies with rare plant and animal populations, adaptive cluster sampling, where sampling occurs preferentially in the neighborhood of quadrats in which the species of interest is detected during the sampling bout, has been shown to estimate population parameters with greater precision at an effort comparable to SRS. Here, we use computer simulations to evaluate the efficiency of adaptive cluster sampling for estimating low levels of disease incidence (0.1, 0.5, 1.0, and 5.0%) at various levels of aggregation of infected plants having variance-to-mean ratios (V/M) of $\approx 1, 3, 5$, and 10. For each simulation, an initial sample size of 50, 100, and 150 quadrats was evaluated, and the condition to adapt neighborhood sampling (CA), i.e., the minimum number of infected plants per quadrat that triggers a switch from random sampling to sampling in neighboring quadrats, was varied from 1 to 4 (corresponding to 7.7 to 30.8% incidence of infected plants per quadrat). The simulations showed that

adaptive cluster sampling was consistently more precise than SRS at a field-level disease incidence of 0.1 and 0.5%, especially when diseased plants were highly aggregated ($V/M = 5$ or 10) and when the most liberal condition to adapt ($CA = 1$) was used. One drawback of adaptive cluster sampling is that the final sample size is unknown at the beginning of the sampling bout because it depends on how often neighborhood sampling is triggered. In our simulations, the final sample size was close to the initial sample size for disease incidence up to 1.0%, especially when a more conservative condition to adapt ($CA > 1$) was used. For these conditions, the effect of disease aggregation was minor. In summary, both precision and the sample size required with adaptive cluster sampling responded similarly to disease incidence and aggregation, i.e., both were most favorable at the lowest disease incidence with the highest levels of clustering. However, whereas relative precision was optimized with the most liberal condition to adapt, the ratio of final to initial sample size was best for more conservative CA values, indicating a tradeoff. In our simulations, precision and final sample size were both simultaneously favorable for disease incidence of up to 1.0%, but only when infected plants were most aggregated ($V/M = 10$).

Additional keywords: disease survey, survey efficiency.

Many plant disease management decisions must be made at low levels of disease incidence, often below 1% (10,31) to maximize treatment efficacy. At such very low levels of disease, the distribution of infected plants in the field is often spatially clustered, e.g., as a result of aggregation of the initial inoculum. Conventional sampling designs such as simple random sampling (SRS) or stratified random sampling (7,14) may not be very efficient for assessing disease at low incidence levels when disease is aggregated. This is because encountering clusters of infected plants becomes a chance event, resulting in a high error variance (i.e., low precision). In theory, sampling efficiency for rare and spatially clustered populations may be improved by allocating lower and higher sampling efforts where the population density is lower and higher, respectively (4).

A sampling approach that incorporates such information about clustering is adaptive cluster sampling (26). The basic idea is to begin sampling for the species of interest following some conventional (random or systematic) pattern, and if the species is found, to intensify searching nearby. The approach is intuitive for a clustered population in that the likelihood of finding another individual is greater in the vicinity of a previously encountered individual than in some other, randomly selected location. Com-

puter simulations for animal populations (21–23) and forest inventories (1,25) have shown that adaptive cluster sampling can be highly efficient, i.e., produce estimates of population density with smaller variance at equal sampling effort compared with conventional sampling designs. Inevitably, the sample obtained through adaptive cluster sampling is biased (i.e., locations containing the species of interest are represented disproportionately), but unbiased statistical estimators may be obtained based on the Rao-Blackwell theorem. The mathematical and statistical details of the method are beyond the scope of this paper but are explained in Thompson (26) and Dryver and Thompson (9).

Briefly, adaptive cluster sampling is implemented as follows: (i) start selecting the initial sample using a conventional design such as SRS; (ii) once the species of interest (e.g., a rare animal, a particular tree species, or in our context, a diseased plant) is detected in a sampling unit (e.g., quadrat) and observed to exceed a predetermined numerical threshold (referred to as the condition to adapt), then switch to systematic sampling of the neighboring quadrats, whereby the number of nearby quadrats to be included is the neighborhood size; (iii) if the condition to adapt is triggered in one of the neighboring quadrats, then continue with step ii, i.e., proceed sampling additional quadrats in the neighborhood; (iv) if none of the neighboring quadrats meet the condition to adapt, then resume random sampling; (v) switch back to neighborhood sampling if the condition to adapt is met in one of the next quadrats encountered. The approach is illustrated schematically in Figure 1, where the task is to estimate the mean number of dis-

Corresponding author: P. S. Ojiambo; E-mail address: peter_ojiambo@ncsu.edu

doi:10.1094/PHYTO-100-7-0663

© 2010 The American Phytopathological Society

eased plants, scattered unevenly in a field consisting of 400 quadrats. An initial random sample of 10 quadrats is shown in Figure 1A. In this example, whenever two or more diseased plants are observed in one of the randomly selected quadrats (condition to adapt = 2), the adjacent neighboring quadrats to the left, right, top, and bottom are added to the sample (neighborhood size = 4). When the process of adding neighboring quadrats is completed, the final sample consists of the 48 quadrats shown in Figure 1B. Neighborhoods of quadrats may be defined in many ways other than the spatial cross-pattern used in Figure 1 (28).

A collection of neighboring quadrats that results from the condition to adapt being satisfied is called a cluster (Fig. 1C). The cluster consists of edge units, i.e., quadrats at its boundary where the condition to adapt is not met; and of network units, i.e., interior units where selection of any one quadrat within the network leads to inclusion in the sample of every other quadrat in the network (Fig. 1C). Similar to one-way analysis of variance, the total population variance can be partitioned into between-network and within-network components (27). As discussed in

more detail below, the within-network variance (σ_{wn}^2), i.e., the variance among quadrats in the network, is an important property that determines the efficiency of adaptive cluster sampling.

Adaptive cluster sampling differs from conventional sampling designs in that the process for selecting an adaptive sample is dependent on the population values observed in the field, i.e., the exact number and spatial location of sampling units to be sampled is not known beforehand. Another, commonly used sampling approach for which the final sample size is also not known in advance is sequential sampling (6,16). With sequential sampling, the decision whether to continue sampling is made after each sample is taken based on values observed in previous samples, whereas with adaptive cluster sampling, both the decision on whether and where to sample is based on previously observed values. In practice, therefore, adaptive cluster sampling requires a high degree of record-keeping as the sampling proceeds, which can be accomplished using a hand-held computer with a global positioning system. To alleviate the problem of the unknown final sample size, several stopping rules have been developed and

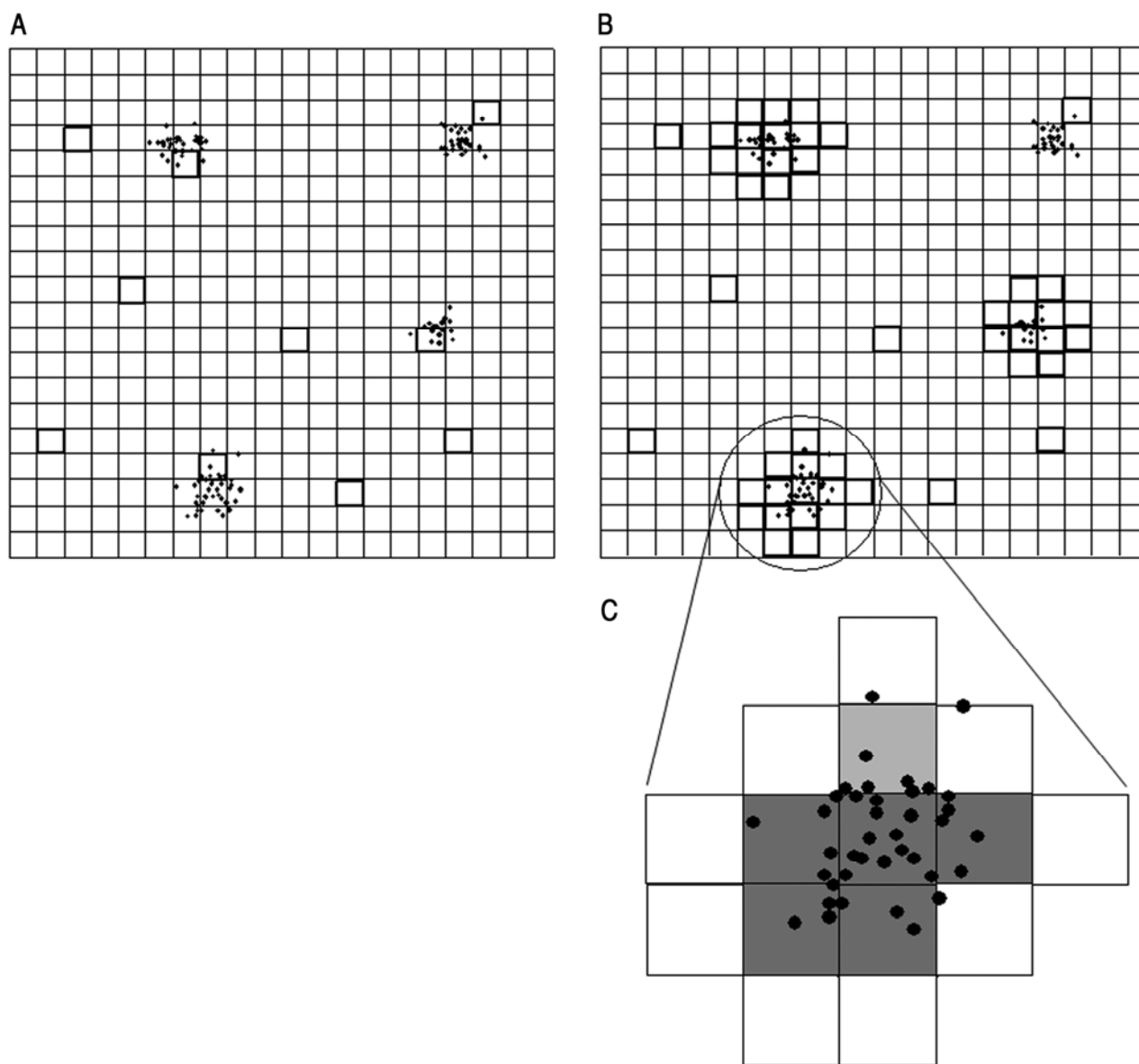


Fig. 1. Illustration of adaptive cluster sampling to estimate the number of infected plants (dots) in a square field consisting of 400 sampling units (quadrats). **A**, First, an initial random sample of 10 quadrats, indicated by the heavier borders, is selected. **B**, Then, four adjacent quadrats are added to the initial sample whenever two or more infected plants are observed in a selected quadrat (condition to adapt = 2), resulting in a final sample of 48 quadrats. **C**, The group of adjacent quadrats generated when neighborhood sampling is triggered is a cluster consisting of edge units (white) and network units (gray). The network unit shown in light gray is the initial sample, a randomly selected quadrat which triggered neighborhood sampling.

evaluated (3,5,24). Conceptually, these stopping rules are similar to those used in sequential sampling in that sampling stops when a predetermined maximum number of sampling units has been evaluated.

In recent years, adaptive cluster sampling has been evaluated for use in several plant and animal ecology studies (1,17,21, 23,25) and in one preliminary report for assessing insect pest density (30), but to our knowledge there are no reports that document the performance of this sampling design for estimating plant disease incidence. In previous research, the efficiency of adaptive cluster sampling was found to depend on the rareness and spatial distribution of the species of interest, quadrat size, method of selecting and the size of the initial sample, and condition to adapt neighborhood sampling. Here, we use computer simulations to assess these design parameters in a plant disease sampling context. Simulations, in general, allows for greater insight into specific biological phenomena in a variety of *ceteris paribus* conditions and at scales not practical in the field (18,20,29), and is used commonly for evaluating disease survey designs (7,13,14,16). Our specific objectives were to (i) demonstrate the application of adaptive cluster sampling for estimating mean plant disease incidence; (ii) determine the relative efficiency of adaptive cluster sampling compared with SRS; and (iii) evaluate the influence of several design factors on the efficiency of adaptive cluster sampling for plant disease incidence. Efficiency is defined broadly to include both precision and the required sample size.

THEORY AND APPROACHES

Simulated fields that varied in disease incidence and clustering of infected plants were generated stochastically for use in this study. Each field measured $50 \times 100 \text{ m}^2$ units (0.5 ha) in size and contained 65,000 plants at a spacing of $0.1 \text{ m} \times 0.76 \text{ m}$ (13 plants/ m^2 representing soybean, for example). Thus, each field was divided into 5,000 equally sized $1 \times 1 \text{ m}^2$ (quadrats), each of which could contain between 0 and 13 infected plants. Field-level disease incidence values of 0.1, 0.5, 1.0, and 5.0% with variance-to-mean (V/M) ratios (based on the $1 \times 1 \text{ m}^2$ quadrats) of approximately 1, 3, 5, and 10 were generated using a variant of a Poisson cluster process (2,8). At low disease incidence, as simulated here, V/M is a very good approximation of a more exact index of dispersion, the D -statistic, which is sometimes used to quantify spatial heterogeneity (15). The cluster process was implemented as follows. First, the number of clusters was selected from a Poisson distribution with mean cluster distance ranging from 1 to 30 m. Cluster centers were positioned randomly within the field grid. In the second stage, diseased plants associated with each cluster were created as a random number from a Poisson distribution with a mean varying from 1 to 500. Infected plants were distributed around the cluster center at a random distance drawn from an exponential distribution and at a random angle between 0 and 360° (Fig. 2). Populations were generated using SAMPLE software version 1.5.4 (U.S. Geological Survey, Leetown Service Center, Kearneysville, WV). Each of the 16 disease incidence-aggregation combinations was simulated stochastically 100 times, i.e., there were 100 disease maps for each combination.

To simulate sampling, a rectangular grid with 1-m^2 quadrats was superimposed on each field, with each quadrat thus containing 13 plants. Using SAMPLE software, sampling was simulated for two designs: (i) SRS, the 'null' sampling model, with the only design factor being initial sampling fraction (f), i.e., the number of quadrats sampled relative to the total number of quadrats in the field; and (ii) adaptive cluster sampling, in which the design factors evaluated were initial sampling fraction (f) and the condition to adapt neighborhood sampling (CA). An initial sample of size $n_i = 50$ ($f = 1.0\%$), 100 ($f = 2.0\%$), or 150 ($f = 3.0\%$) quadrats was selected randomly (without replacement), and the num-

ber of symptomatic plants per quadrat was counted. If the observed value satisfied the condition to adapt (of which values of $CA = 1, 2$, and 4 infected plants per quadrat were evaluated), then four additional quadrats adjacent to that quadrat were sampled in a cross-pattern; in other words, the neighborhood size was 4, which is typical for the first-order neighborhood (28). If the condition to adapt was met in any quadrat of that neighborhood, then their neighborhoods were also sampled similarly. This process was repeated until a boundary of edge units was encountered which no longer satisfied the condition to adapt. At that point, random sampling was resumed if the number of quadrats sampled randomly was $< n_i$. The final number of samples (n_f), which included n_i as well as the quadrats added by neighborhood sampling, was recorded for each field. The ratio of the final to the initial sample size (n_f/n_i) was then calculated. For adaptive cluster sampling to be efficient, this ratio should be close to 1 (2). In the example shown in Figure 1, $n_f/n_i = 4.8$, indicating that the scenario is not very efficient as far as sample size is concerned.

Two estimators, the Horvitz-Thompson, HT (12) and the Hansen-Hurwitz, HH (11) estimators, are available to estimate the population mean in adaptive sampling designs. Thompson (26) presented modified, unbiased HT and HH-type estimators, of which we used the modified HT estimator since it results in smaller estimates of the variance in single-stage adaptive cluster sampling with initial samples selected randomly (19). Based on 500 simulations, the variance of the estimated mean obtained from SRS, $\text{var}(\bar{y})$, was calculated for each disease map. Similarly, the variance of the estimated mean from adaptive cluster sampling, $\text{var}(\hat{\mu})$, was computed for the same map. Values from the 100 maps per disease incidence-aggregation combination were averaged, and the ratio $\text{var}(\bar{y})/\text{var}(\hat{\mu})$ was used as a measure of relative precision (RP). Adaptive cluster sampling is more precise than SRS, i.e., results in a smaller variance, when $RP > 1$. Details on computing the modified HT-estimator and its variance are given in the Appendix.

RESULTS AND DISCUSSION

Precision. The relative precision (RP) achieved with adaptive cluster sampling relative to SRS was highest at the lowest disease incidence and the highest level of clustering of diseased plants (Fig. 3). At $CA = 1$ (i.e., the most liberal condition to adapt where presence of a single infected plant in a 13-plant quadrat triggers neighborhood sampling), RP was consistently greater than 1.0 for disease incidence levels of 0.1 and 0.5% (Fig. 3A), except when diseased plants were distributed randomly ($V/M = 1$). As disease incidence increased to 1.0%, adaptive cluster sampling was more precise than SRS only for the two highest levels of aggregation ($V/M = 5$ and 10). At a disease incidence of 5%, adaptive cluster sampling was consistently less precise than SRS as indicated by $RP < 1$.

The effect of aggregation on precision can be explained in terms of networks (i.e., interconnected quadrats containing diseased plants) formed during adaptive cluster sampling (2). Populations with highly aggregated clusters have networks with large variances relative to the population variance, and adaptive sampling is efficient when the ratio of within-network variance (σ_{wn}^2) to population variance (σ^2) is large (23). For adaptive cluster sampling to be efficient, the size of a network should be large enough such that σ_{wn}^2 is similar to σ^2 (2,26). In our simulations, we observed a direct, positive relationship between σ_{wn}^2/σ^2 and V/M (Fig. 4).

As the condition to adapt became more conservative (i.e., larger numbers of diseased plants per quadrat are required to trigger neighborhood sampling), the gain in precision of adaptive cluster sampling relative to SRS diminished (Fig. 3B and C). For $CA = 2$, RP was greater than 1 for disease incidence levels of up to 1.0% for the two highest level of clustering; whereas at $CA = 4$, a gain

in precision relative to SRS was realized only when disease was most highly aggregated ($V/M = 10$). This result is intuitive in that low levels of clustering combined with low levels of disease incidence will lead to most quadrats having only few infected plants; if the condition to adapt is conservative (i.e., CA is large), neighborhood sampling will be triggered infrequently and adaptive cluster sampling will approximate SRS, leading to RP values around 1. This is also reflected in the effect of spatial aggregation on the within-network to population variance (Fig. 4), where a more conservative condition to adapt neighborhood sampling resulted in a lower ratio of σ_{wn}^2 / σ^2 .

The effect of the initial sampling fraction (f) on RP was relatively small and was most pronounced for high disease incidence

(Fig. 5). Thus, in summary, adaptive cluster sampling was most precise for low levels of disease incidence (0.1 or 0.5%), especially when diseased plants were highly aggregated ($V/M = 5$ or 10) and when the most liberal condition to adapt ($CA = 1$, corresponding to 7.7% of the 13 plants per quadrat in these simulations) was used.

Final sample size. Sampling efficiency is a function not only of precision (as expressed by RP), but also of sample size. Although the initial sample size, n_i , is fixed in adaptive cluster sampling, as defined by the sampling fraction (f), the final sample size, n_f , is variable in that it depends on how often neighborhood sampling is triggered, which itself is a function of disease incidence, V/M , and CA . For adaptive cluster sampling to be

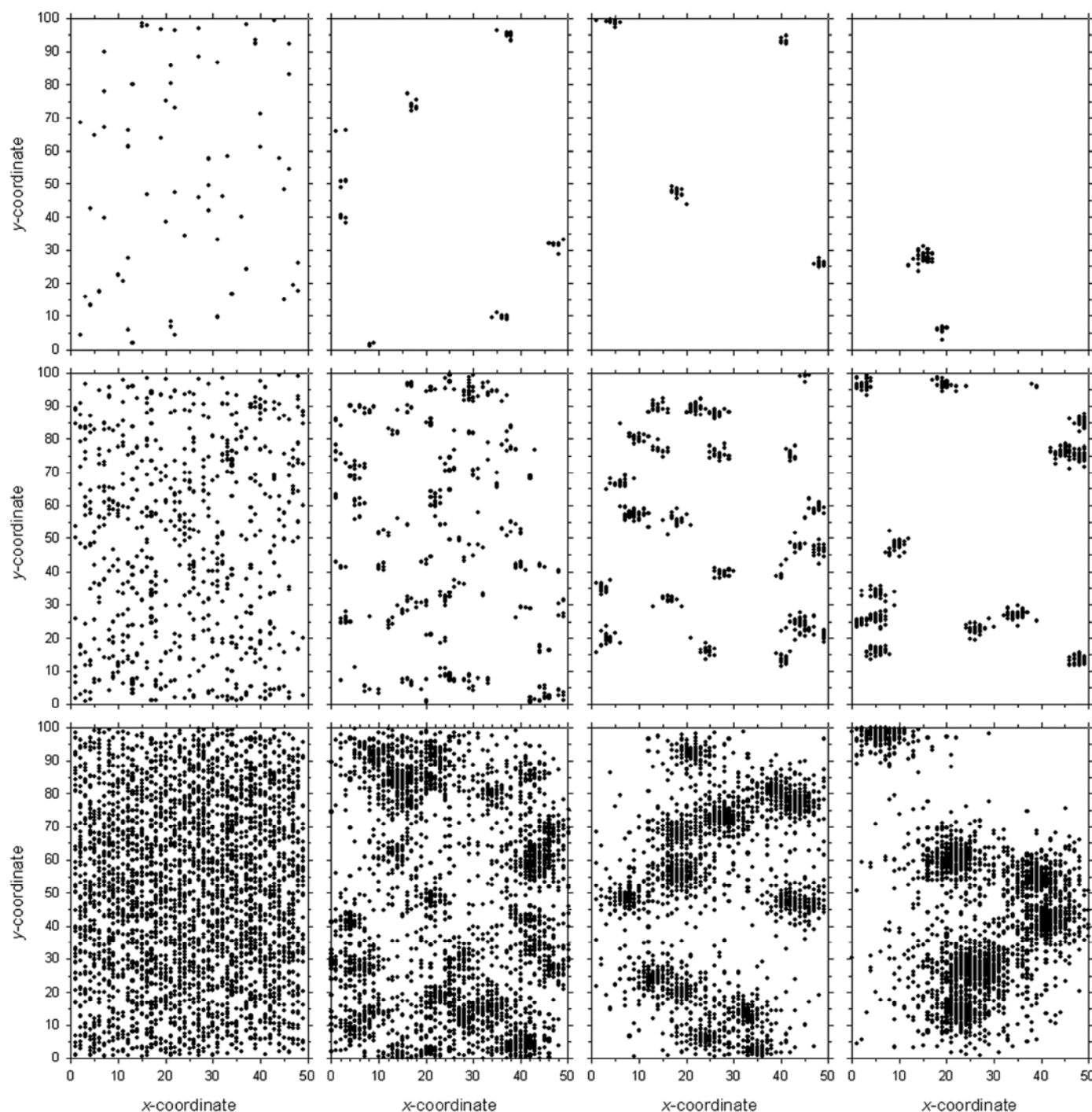


Fig. 2. Illustration of various degrees of spatial aggregation of infected plants (variance-to-mean ratio $[V/M] = 1, 3, 5$, and 10 from left to right) for disease incidence = 0.1% (top), 1.0% (middle), and 5.0% (bottom). Each simulated field contains a total of 65,000 plants.

efficient, the final sample size should not be excessively larger than the initial sample size. In our simulations, final sample size was close to the initial sample size when disease incidence was $\leq 0.5\%$, irrespective of V/M and CA (Fig. 6). For incidence levels $> 0.5\%$, the ratio of final to initial sample size was dependent on V/M and CA . For example, at disease incidence = 1.0% , the final sample size was very close to the initial sample size only for $CA > 1$; whereas for $CA = 1$, the final sample size was approximately

twice the initial sample size. At incidence levels $> 1.0\%$, adaptive cluster sampling became inefficient as expressed by more or less rapidly (depending on V/M) increasing ratios of n_f/n_i . In the worst-case scenario in our simulations (disease incidence = 5.0% , $V/M = 1$, $CA = 1$), the final sample size was ≈ 20 times greater than the initial sample size (Fig. 6A).

The initial sampling fraction (f) had a very limited effect on the ratio of final to initial sample size, except for disease incidence $> 1.0\%$, for which the ratio became less favorable for smaller initial sampling fractions (Fig. 7). Thus, in summary, the final sample size was closest to the initial sample size for disease incidence levels up to 1.0% , especially when a more conservative condition to adapt ($CA > 1$) was used. Under these conditions, the effect of disease aggregation was minor.

Practical implications for sampling. As documented above, both precision and the required sample size responded similarly to disease incidence and aggregation, i.e., both were most favorable at low disease incidence and high levels of clustering. However, the two parameters responded differently to changes in CA ; whereas relative precision was optimized with the most liberal condition to adapt ($CA = 1$), the ratio of final to initial sample size was best for more conservative CA values ($CA = 4$), indicating a tradeoff. The difficulty in optimizing both parameters simultaneously is illustrated in Figure 8 where all 48 tested combinations of disease incidence, V/M , and CA are plotted on a precision-sample size plane. For the purpose of illustration, suppose we demand that (on average) adaptive cluster sampling be at least 1.5 times as precise as SRS ($RP \geq 1.5$), but the required sample size be no more than 1.25 times that of SRS ($n_f/n_i \leq 1.25$). This condition, indicated by gray shading in Figure 8, is satisfied for only 4 of the 48 tested combinations, all with very low disease incidence (0.1 or 0.5%), high levels of aggregation ($V/M = 5$ or 10), and liberal to intermediate conditions to adapt neighborhood sampling ($CA = 1$ or 2). Whether an investigator would be interested in improving precision at such a low level of disease (where the focus may be more on detection than on estimating the mean) is debatable. However, two additional data points (incidence = 0.1% , $V/M = 3$; incidence = 1.0% , $V/M = 10$) are close to meeting the precision-sample size condition defined above. Of these, the latter would appear to be a more realistic scenario for the practical application of adaptive sampling.

Fundamentally, the tradeoff between maximizing precision and minimizing final sample size is a function of network size and the

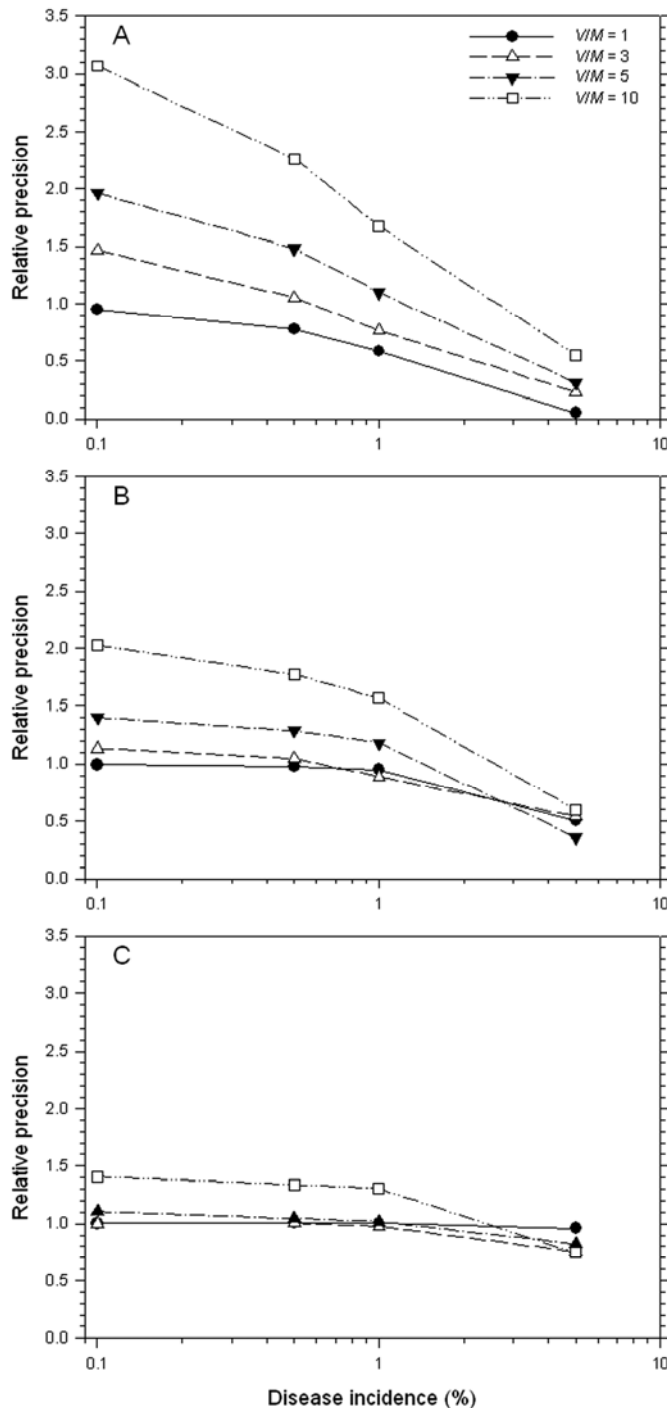


Fig. 3. Relative precision of adaptive cluster sampling as affected by disease incidence and spatial aggregation (variance-to-mean ratio, V/M) when the condition to adapt neighborhood sampling (CA) is **A**, 1, **B**, 2, or **C**, 4 infected plants per 13-plant quadrat, for an initial sampling fraction (f) = 2.0% (100 quadrats). Precision is expressed as the ratio of the variance around the mean incidence from simple random sampling to the variance from adaptive cluster sampling; values > 1 indicate that adaptive cluster sampling is more precise than simple random sampling. Simulations are based on 100 realizations of each incidence-aggregation combination, each sampled 500 times.

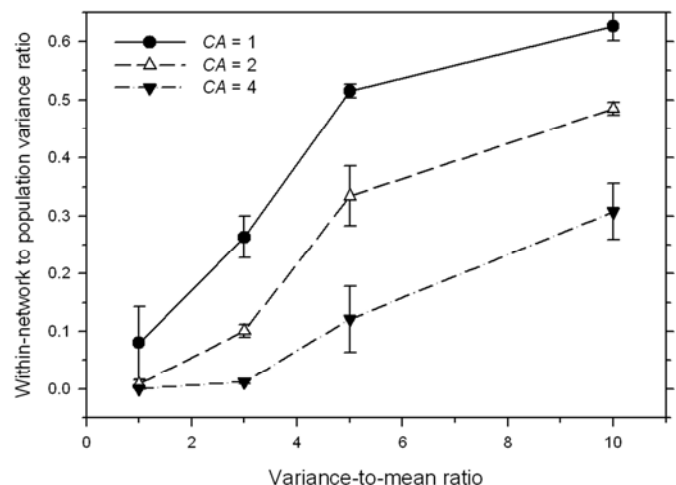


Fig. 4. Ratio of within-network variance to population variance in adaptive cluster sampling for disease incidence as influenced by spatial aggregation (variance-to-mean ratio) of infected plants, for an initial sampling fraction (f) = 2.0% (100 quadrats) when the condition to adapt neighborhood sampling (CA) is 1, 2, or 4 infected plants per quadrat. Means and standard errors are calculated across disease incidence levels of 0.1 , 0.5 , 1.0 , and 5.0% .

associated within-network variance (σ_{wn}^2), but the relationship between σ_{wn}^2 and sample size is not straightforward. As the size of the networks generated by adaptive cluster sampling increases, both σ_{wn}^2/σ^2 and n_f/n_i will increase. Thompson (26) presents

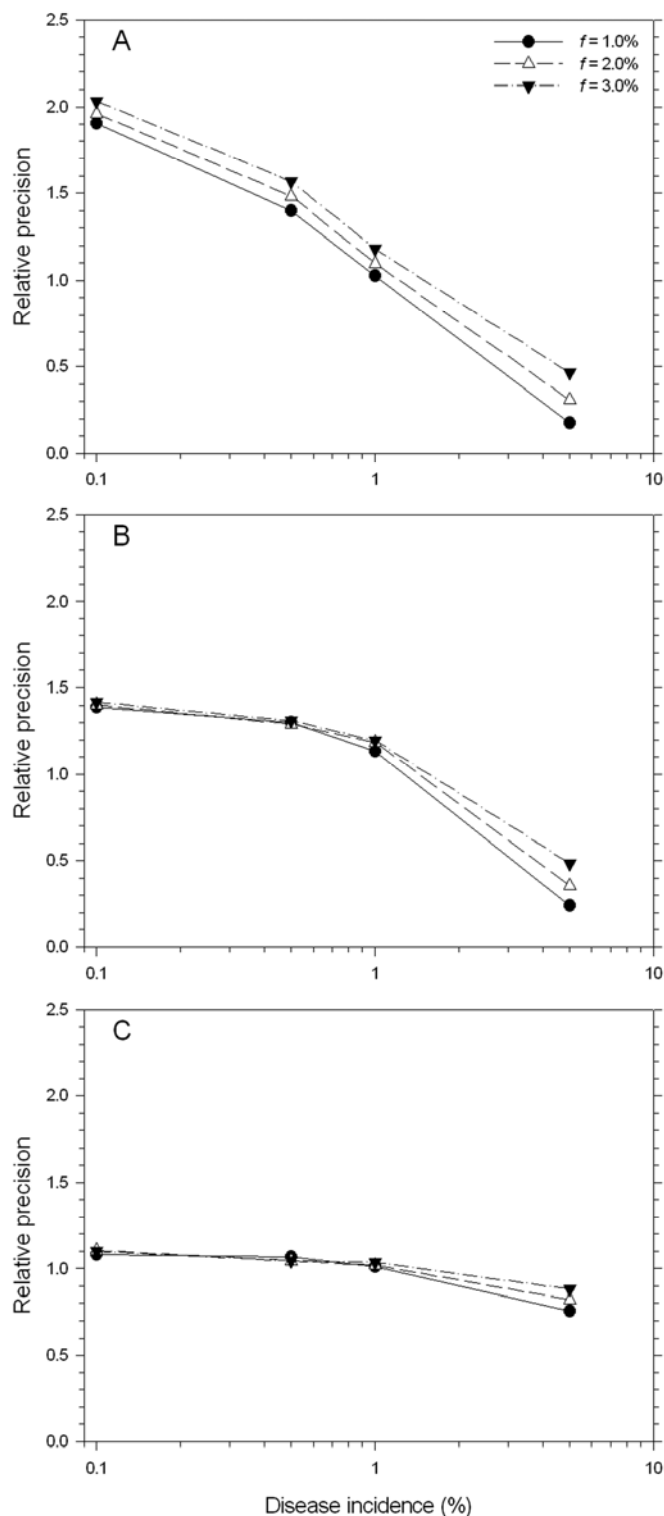


Fig. 5. Relative precision of adaptive cluster sampling as affected by disease incidence and initial sampling fraction (f) when the condition to adapt neighborhood sampling (CA) is **A**, 1, **B**, 2, or **C**, 4 infected plants per 13-plant quadrat, at disease aggregation (variance-to-mean ratio) = 5. Precision is expressed as the ratio of the variance around the mean incidence from simple random sampling to the variance from adaptive cluster sampling; values >1 indicate that adaptive cluster sampling is more precise than simple random sampling. Simulations are based on 100 realizations of each incidence-aggregation combination, each sampled 500 times.

equations that can be used to understand the dynamics of the 'network-effect' via simulation. A rule of thumb for designing an efficient adaptive cluster sample is to aim for small networks so that n_f is not excessively greater than n_i , but networks should not be so small that σ_{wn}^2 is very low (2).

Our simulations showed that both precision and the sample size required with adaptive cluster sampling responded similarly to disease incidence and aggregation, i.e., both were most favorable at the lowest disease incidence with the highest levels of clustering. However, whereas relative precision was optimized with the

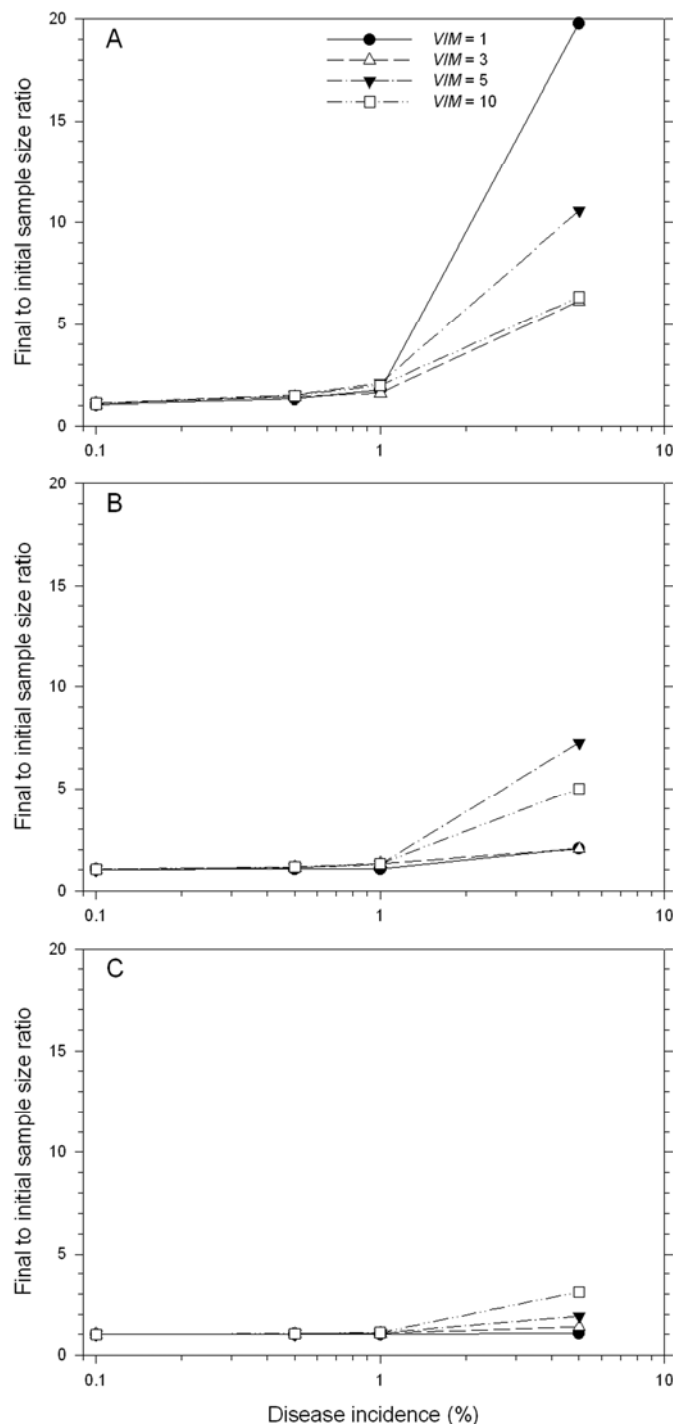


Fig. 6. Ratio of final to initial sample size of adaptive cluster sampling as affected by disease incidence and spatial aggregation (variance-to-mean ratio, V/M) when the condition to adapt neighborhood sampling (CA) is **A**, 1, **B**, 2, or **C**, 4 infected plants per 13-plant quadrat, for an initial sampling fraction (f) = 2.0% (100 quadrats). Simulations are based on 100 realizations of each incidence-aggregation combination, each sampled 500 times.

most liberal condition to adapt, the ratio of final to initial sample size was best for more conservative CA values, indicating a tradeoff. Precision and final sample sized were both simultaneously favorable for disease incidence of up to 1.0%, but only when infected plants were most aggregated ($V/M = 10$).

Adaptive cluster sampling is rich in options, but this flexibility complicates survey designs. As such, further research is needed to develop analytical guidelines for designing plant disease surveys that use adaptive cluster sampling. This simulation study utilized realistic parameters for disease incidence, spatial aggregation,

quadrat size, and condition to adapt neighborhood sampling, but it addressed only one field size and host plant density. We evaluated single-stage adaptive sampling, but more complex adaptive sampling designs may be better able to balance the tradeoff between optimizing precision and minimizing sample size. For example, modified adaptive sampling designs such as adaptive grid-based systematic sampling and geographical information system-based adaptive sampling (17) can further increase sampling efficiency. In an effort to limit final sample size, stopping rules have also been developed for use in adaptive cluster sampling. For example, Brown and Manly (3) suggested that sampling could stop after a cumulative sample size meets or exceeds a preset value. We are currently examining these more advanced designs, in conjunction with different field and quadrat sizes, with respect to efficiency and cost of sampling.

APPENDIX

The mean number of individuals per unit (quadrat), \bar{y} , for an initial sample of size n_i selected by simple random sampling without replacement is

$$\bar{y} = \frac{1}{n_i} \sum_{k=1}^{n_i} y_k$$

where y_k is the number of individuals in the k th unit. The variance of the estimator \bar{y} is given by

$$\text{var}(\bar{y}) = \frac{N - n_i}{N n_i} \left[\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{(N - 1)} \right]$$

where N is the total number of units.

Following Thompson (26), the modified Horvitz-Thompson estimator of the mean is

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{n_f} y_k J_k / \pi_k$$

where n_f is the number of units in the final sample, J_k an indicator

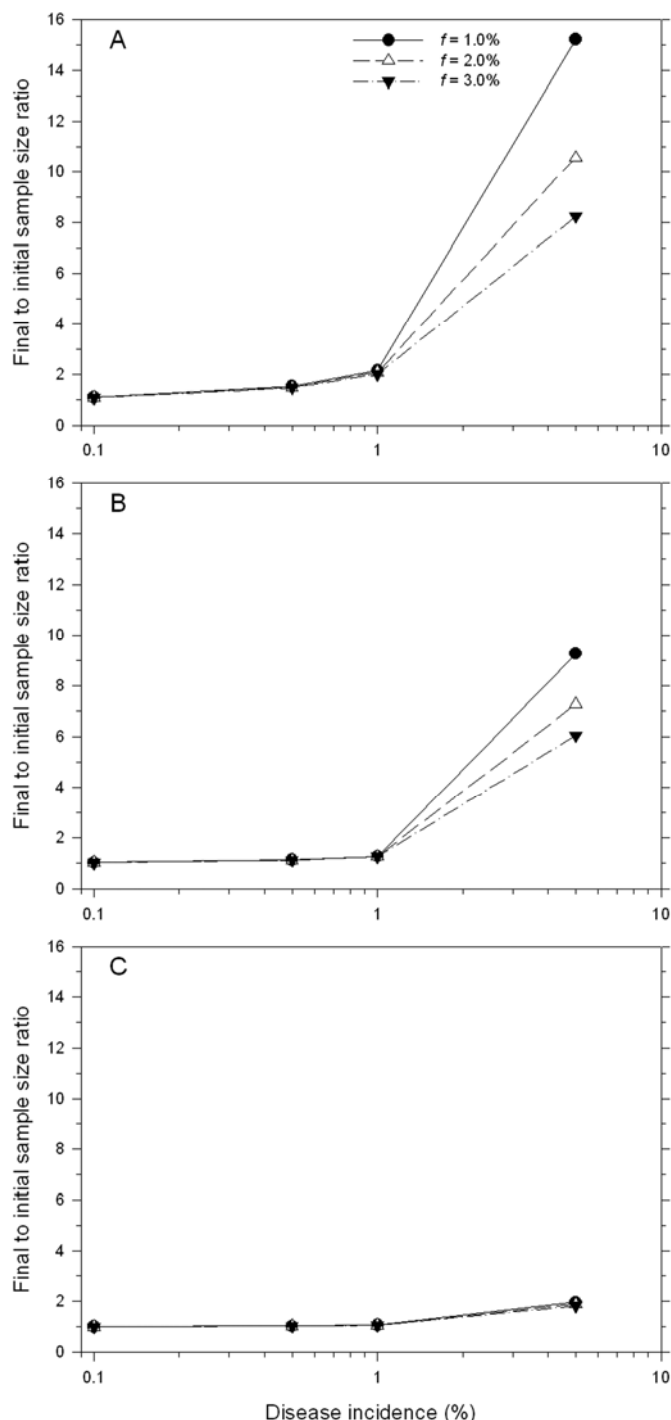


Fig. 7. Ratio of final to initial sample size of adaptive cluster sampling as affected by disease incidence and initial sampling fraction (f) when the condition to adapt neighborhood sampling (CA) is A, 1; B, 2; or C, 4 infected plants per 13-plant quadrat, at disease aggregation (variance-to-mean ratio) = 5. Simulations are based on 100 realizations of each incidence-aggregation combination, each sampled 500 times.

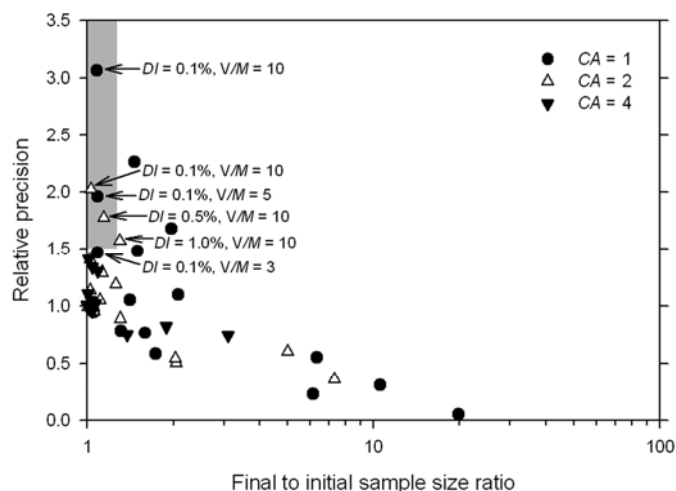


Fig. 8. Relationship between precision and the ratio of final to initial sample size in adaptive cluster sampling for disease incidence. Data points shown are 48 combinations of disease incidence ($DI = 0.1, 0.5, 1.0$, and 5.0%), spatial aggregation (variance-to-mean ratio [V/M] = 1, 3, 5, and 10), and condition to adapt neighborhood sampling ($CA = 1, 2$, and 4), for an initial sampling fraction (f) = 2.0% (100 quadrats). Precision is expressed as the ratio of the variance around the mean incidence from simple random sampling to the variance from adaptive cluster sampling. The gray shading indicates combinations where adaptive cluster sampling is at least 1.5 times as precise as simple random sampling (relative precision ≥ 1.5), but the required sample size is no more than 1.25 times that of simple random sampling (ratio of final to initial sample size ≤ 1.25). Simulations are based on 100 realizations of each incidence-aggregation combination, each sampled 500 times.

variable with values of 0 or 1, and π_k the probability that the k th unit is used in the estimator. $J_k = 1$ if the k th unit is part of the initial (random) sample and/or the condition to adapt neighborhood sampling is met in the unit; otherwise, $J_k = 0$. π_k is given by

$$\pi_k = 1 - \frac{\binom{N-m_k}{n_i}}{\binom{N}{n_i}}$$

where m_k is the number of units in the network of which the k th unit is a member; for units in which the condition to adapt is not met, $m_k = 1$.

The unbiased sample variance for adaptive cluster sampling for the Horvitz-Thompson estimator is

$$\text{var}(\hat{\mu}) = \frac{1}{N^2} \sum_{k=1}^K \sum_{m=1}^K y_k^* y_m^* \frac{(\pi_{km} - \pi_k \pi_m)}{\pi_k \pi_m \pi_{km}}$$

where K is the number of networks in the sample; y_k^* and y_m^* the sum of the number of individuals in two networks, k and m ; and π_{km} the joint probability that the initial sample contains at least one unit in networks k and m , that is

$$\pi_{km} = 1 - \frac{\left\{ \binom{N-m_k}{n_i} + \binom{N-m_m}{n_i} - \binom{N-m_k-m_m}{n_i} \right\}}{\binom{N}{n_i}}$$

Application of the Rao-Blackwell method to improve the above estimators is described by Thompson (26).

ACKNOWLEDGMENTS

We thank D. R. Smith, U.S. Geological Survey, for suggestions relative to the use of adaptive cluster sampling tools.

LITERATURE CITED

- Acharya, B., Bhattarai, G., de Gier, A., and Stein, A. 2000. Systematic adaptive cluster sampling for the assessment of rare tree species in Nepal. *For. Ecol. Manage.* 137:65-73.
- Brown, J. A. 2003. Designing an efficient adaptive cluster sample. *Environ. Ecol. Stat.* 10:95-105.
- Brown, J. A., and Manly, B. J. F. 1998. Restricted adaptive cluster sampling. *Environ. Ecol. Stat.* 5:49-63.
- Christman, M. C. 2000. A review of quadrat-based sampling of rare, geographically clustered populations. *J. Agric. Biol. Environ. Stat.* 5:168-201.
- Christman, M. C., and Lan, F. 2001. Inverse adaptive cluster sampling. *Biometrics* 57:1096-1105.
- Copes, W. E., Scherm, H., and Ware, G. O. 2001. Sequential sampling to assess the incidence of infection by *Monilinia vaccinii-corymbosi* in mechanically harvested rabbiteye blueberry fruit. *Phytopathology* 91:348-353.
- Delp, B. R., Stowell, L. J., and Marois, J. J. 1986. Evaluation of field sampling techniques for estimation of disease incidence. *Phytopathology* 76:1299-1305.
- Diggle, P. J. 2003. *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Arnold, London.
- Dryver, A. L., and Thompson, S. K. 2005. Improved unbiased estimators in adaptive cluster sampling. *J. Roy. Stat. Soc. Ser. B* 67:157-166.
- Fry, W. E. 1977. Integrated control of potato late blight—Effects of polygenic resistance and techniques for timing fungicide applications. *Phytopathology* 67:415-420.
- Hansen, M. M., and Hurwitz, W. N. 1943. On the theory of sampling from finite populations. *Ann. Math. Stat.* 14:333-362.
- Horvitz, D. G., and Thompson, D. J. 1952. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* 47:663-685.
- Hughes, G., and Gottwald, T. R. 1998. Survey methods for assessment of citrus tristeza virus incidence. *Phytopathology* 88:715-723.
- Lin, C. S., Poushinsky, G., and Mauer, M. 1979. An examination of five sampling methods under random and clustered disease distributions using simulation. *Can. J. Plant Sci.* 59:121-130.
- Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distributions, heterogeneity, and temporal analysis. *Annu. Rev. Phytopathol.* 33:529-546.
- Madden, L. V., and Hughes, G. 1999. Sampling for plant disease incidence. *Phytopathology* 89:1088-1103.
- Morrison, L. W., Smith, D. R., Young, C. C., and Nichols, D. W. 2008. Evaluating sampling designs by computer simulation: A case study with the Missouri bladderpod. *Pop. Ecol.* 50:417-425.
- Sackett, K. E., and Mundt, C. C. 2005. The effects of dispersal gradient and pathogen life cycle components on epidemic velocity in computer simulations. *Phytopathology* 95:992-1000.
- Salehi, M. M. 2003. Comparison between Hansen-Hurwitz and Horvitz-Thompson estimators for adaptive cluster sampling. *Environ. Ecol. Stat.* 10:115-127.
- Skelsey, P., Rossing, W. A. H., Kessel, G. J. T., Powell, J., and van der Werf, W. 2005. Influence of host diversity on development of epidemics: An evaluation and elaboration of mixture theory. *Phytopathology* 95:328-338.
- Smith, D. R., Brown, J. A., and Lo, N. C. H. 2004. Application of adaptive cluster sampling to biological populations. Pages 93-152 in: *Sampling Rare or Elusive Species*. W. L. Thompson, ed. Island Press, Washington, DC.
- Smith, D. R., Conroy, M. J., and Brakhage, D. H. 1995. Efficiency of adaptive cluster sampling for estimating density of winter waterfowl. *Biometrics* 51:777-788.
- Smith, D. R., Villella, R. F., and Lemarie, D. P. 2003. Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environ. Ecol. Stat.* 10:7-15.
- Su, Z., and Quinn, T. J., II. 2003. Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. *Environ. Ecol. Stat.* 10:17-41.
- Talvitie, M., Leino, O., and Holopainen, M. 2006. Inventory of sparse forest populations using adaptive cluster sampling. *Silva Fenn.* 40:101-108.
- Thompson, S. K. 1990. Adaptive cluster sampling. *J. Am. Stat. Assoc.* 85:1050-1059.
- Thompson, S. K., and Seber, G. A. 1996. *Adaptive Sampling*. John Wiley & Sons, Inc., New York.
- Turk, P., and Borkowski, J. J. 2005. A review of adaptive cluster sampling: 1990-2003. *Environ. Ecol. Stat.* 12:55-94.
- Willocquet, L., and Savary, S. 2004. An epidemiological simulation model with three scales of spatial hierarchy. *Phytopathology* 94:883-891.
- Zhang, N.-S., Zhu, Z.-R., and Hu, B.-M. 2000. On two-stage adaptive cluster sampling to assess pest density. *J. Zhejiang Univ. (Agric. & Life Sci.)* 26:617-620.
- Zink, F. W., Grogan, R. G., and Welch, J. E. 1956. The effect of seed transmission upon subsequent spread of lettuce mosaic virus. *Phytopathology* 46:662-664.