Ecology, 79(4), 1998, pp. 1425–1434
© 1998 by the Ecological Society of America

OPTIMIZATION OF VARIABLE AREA TRANSECT SAMPLING USING MONTE CARLO SIMULATION

RICHARD M. ENGEMAN[1] AND ROBERT T. SUGIHARA[2]

[1]*National Wildlife Research Center, USDA/Animal and Plant Health Inspection Service/Wildlife Services, 1716 Heath Parkway, Fort Collins, Colorado 80524-2719 USA*
[2]*National Wildlife Research Center, USDA/Animal and Plant Health Inspection Service/Wildlife Services, Hawaii Field Station, P.O. Box 10880, Hilo, Hawaii 96721 USA*

Abstract. An extensive simulation study was conducted to optimize the number, r, of population members to be encountered from each random starting point in variable area transect (VAT) sampling. The quality of estimation provided by the original calculation formula presented by K. R. Parker in 1979 was compared to another formula that was a Morisita analog intended to reduce bias when sampling aggregated populations. Monte Carlo simulations covered 64 combinations of four spatial patterns, four sample sizes, and four densities. Values of r from 3 through 10 were considered in each case. Relative root mean squared error was used as the primary assessment criterion.

Superior estimation properties were found for $r > 3$, but diminishing returns, relative to the potential for increased effort in the field, were found for $r > 6$. The original estimation formula consistently provided results that were superior to the Morisita analog, with the difference most pronounced in the aggregate patterns for which the Morisita analog was intended. As long as the sampled populations displayed randomness in location of individuals, rather than systematic patterns that are uncommon in nature, the variance formula associated with the original estimation formula performed well.

Additional simulations were conducted to examine four confidence interval methods for potential use in association with the Parker original estimation method. These simulations considered only the sample sizes for which the best estimation was achieved in the earlier simulations. The confidence interval method developed by Parker worked well for populations with random spatial patterns, but it rarely achieved 80% (generally much less) of target coverage for populations displaying aggregation. A nonparametric confidence interval method presented here, or a combination of it with the Parker method, is recommended for general use.

Key words: confidence interval coverage; density estimation; distance methods; plotless methods; spatial pattern.

INTRODUCTION

Many attempts have been made to develop improved methods for addressing a problem common in many fields of biology, estimating the density of immobile objects such as points of animal damage, plant communities, and bird nests. Although the quadrat or plot method is well known to produce unbiased estimates, it can be labor intensive, especially when objects are sparse, unevenly distributed, or otherwise difficult to locate. Many distance or plotless methods have been developed or proposed as more efficient approaches for acquiring a sample from which density can be estimated. In general, plotless density estimators (PDEs) were developed assuming that the sampled population followed a random spatial distribution (e.g., Pollard 1971). A variety of PDEs have been proposed to offer robust estimation over different spatial patterns (e.g., Morisita 1957, Persson 1971, Batcheler 1975, Diggle 1975, Lewis 1975, Patil et al. 1979, Clayton and Cox 1986, Delince 1986). Unfortunately, some PDEs intended to overcome problems with robustness are nearly as difficult to apply in many field situations as is quadrat sampling (e.g., Pollard 1971, Engeman et al. 1994).

Recently Engeman et al. (1994) conducted a simulation study that compared a large number of PDEs on populations from a variety of spatial patterns and a number of densities. Recommendations were made as to which PDEs provided the best estimation properties while remaining feasible to apply in difficult field situations. Here, we follow up on that work by optimizing one of the best performing, but simplest to apply, of the methods from that work, the variable area transect (VAT) developed by Parker (1979).

VARIABLE AREA TRANSECT

The variable area transect estimator can be considered a combination of distance and quadrat methods. At each of n randomly located starting points, a fixed-width (strip) transect is searched until the rth individual is encountered in the strip. Thus, the search for pop-

Manuscript received 11 June 1997; accepted 2 July 1997; final version received 25 July 1997.

ulation members and the measurement of distances to them from the random point are accomplished at the same time, making this an attractive method to apply in the field. If we let w be the width of the strip transect and L_i be the length searched from the ith random starting point to the rth individual, then the formula for the VAT estimator, which we call the original formula, is:

$$\hat{D}_{\text{orig}} = (nr - 1)/(w \sum L_i).$$

Because this estimator was developed under the assumption that the spatial pattern for the sampled population was random, Parker (1979) also proposed an analog to Morisita's (1957) method for producing an unbiased sample with closest individual sampling in an aggregated spatial pattern. The formula for this Morisita analog (not included in the simulations by Engeman et al. 1994) is

$$\hat{D}_{\text{mor}} = [(r - 1) \sum (1/L_i)]/(nw).$$

Following Parker's (1979) example, Engeman et al. (1994) considered only the original form using $r = 3$. Here, we investigate the estimation properties of both formulae when sampling populations that represent a diversity of spatial patterns, and at a variety of densities. In each circumstance we consider estimation based on locating $r = 3, 4, \ldots, 10$ population members from each random starting point. From there we attempt to determine optimal values of r and n. Concurrently, we also examine the properties for the corresponding variance formula for each estimator. The formulae (Parker 1979) used were:

$$\widehat{\text{var}}(\hat{D}_{\text{orig}}) = (\hat{D}_{\text{orig}})^2/(nr - 2)$$

$$\widehat{\text{var}}(\hat{D}_{\text{mor}}) = (\hat{D}_{\text{mor}})^2/[n(r - 2)].$$

Simulation Study Design

A simulation program was written in FORTRAN 77 language (Version 5.0, MS-DOS operating system), each run of which was specified by a combination of population spatial pattern, population density, and sample size (of random VAT starting points). We examined 64 combinations encompassing four spatial patterns, four densities, and four sample sizes. At each random sampling point, the strip transect was extended and searched until 10 population members were encountered. Thus, for each estimation formula, estimates could be calculated using $r = 3, 4, 5, 6, 7, 8, 9$, and 10. This results in a simulation study on five features that potentially could affect estimation by VAT sampling: spatial pattern, sample size, density, estimation method, and number of population members searched for at each random sample point.

The uniform random number generator used for placing population individuals and locating sampling points was the UNIF routine (Bratley et al. 1983). Where required, the VNORM routine (Bratley et al. 1983) was used to convert the uniform random numbers to normal random numbers. UNIF has been extensively tested for uniformity, independence, and nonperiodicity of the numbers generated and VNORM tested for accuracy (Brody and Morais 1987).

The density used in a particular run of the program was specified by inputting the size of a rectangular area (the length of each dimension) and the number of individuals to reside in that area. We examined target population densities of 2, 5, 10, and 20 individuals per unit area. The area used for each density was large enough to insure that the target population was several orders of magnitude larger than the number of sampling points.

We considered four spatial patterns for the populations simulated in this study: random, regular, triangular, aggregate, and double clumped. The *random* pattern (also called Poisson in recognition that the points are distributed as a two-dimensional Poisson process) was simulated by generating the appropriate number of random coordinates from an uniform distribution in the designated area. The *regular* spatial pattern was generated by dividing the area into a grid of rectangles, the same number as individuals in the population. The population members were then situated by randomly locating one individual in each rectangle. For the *aggregate* pattern, the centers of user-specified number of clumps were randomly located in the designated area. In addition to the clump center point, user-specified number of "offspring" for the clumps were located within user-specified radius of the center (parent) point. These offspring were located within the clump about the parent point using coordinates randomly generated from the standard bivariate normal distribution. This tended to concentrate the members of the clump near the center point. Aggregate spatial patterns approximate many of the naturally occurring biological population patterns. We considered only clumps of five individuals (the center point or "parent" and four "offspring") that were severely clumped, with the offspring located within a clump radius of 15 distance units. The pattern we label as *double-clumped* is a second-order aggregation that was generated in a similar fashion to the aggregate pattern. The difference is that for the double-clumped pattern, the individuals in the clumps of the aggregate pattern are used for center points (parents) for subclumps of two individuals. The two individuals of the subclumps include the parent plus one other point (offspring) randomly generated from the standard bivariate normal distribution. The radius for the subclump is restricted to one-half that for the clump (7.5 units). This spatial pattern approximates some of the field patterns that we have observed for rodent burrows and animal damage locations. It also provides one of the severest tests of the estimation methods.

Randomly located starting points for the VATs were required to initiate the sampling procedures. The sample sizes considered in this study refer to the number of such random starting points placed in the population.

Sample sizes of random starting points that we examined were 5, 10, 20, and 40.

There were two runs of the simulation program for each spatial pattern-by-density-by-sample size combination. At each replication of each run of the program, a new population was generated and a new set of random sampling points applied. Each simulation run comprised 5000 such replications. The observed statistics accumulated over the 5000 replications for each estimation method at each r included the mean density estimate, variance, relative bias, mean squared error (MSE), and the relative root mean squared error (RRMSE). RRMSE was calculated as

$$\text{RRMSE} = \{[\sum(\hat{D} - D)^2/D^2)]/I\}^{1/2}$$

where $\hat{D}$ was the estimated density, D was the true density, and $I = 5000$ was the number of replications in the simulation run. We used the RRMSE as the primary criteria for comparing the performance of the estimators (see, for example, Patil et al. 1979, Engeman and Bromaghin 1990, Engeman et al. 1994), because it encompasses variance and bias, and it is unitless. The same statistics were used to evaluate the variance estimators for each method, where the density estimate in the above equation is replaced by the appropriate variance estimate and D is replaced by the observed variance of the estimate based on the 5000 simulation replications. Also calculated was the relative bias (RBIAS), the mean observed bias divided by the true parameter value. We present our statistics from the simulations as "relative" statistics (divided by the true density) to standardize the scale across the density parameter being estimated.

The trade-off between n and r was examined by developing a predictive model of RRMSE for the density estimate using the better performing of the two estimation methods (providing one was consistently superior). Only variables under the investigator's control (n and r) were used to develop predictive equations. Density, therefore, was not included. Also, we presumed that: (1) different relationships between RRMSE and n and r might come into play for different spatial patterns of population members, and (2) a field investigator might have information on the population pattern, based on similar populations elsewhere or on other a priori knowledge. For each population pattern, all possible subsets of n, r, and their product were considered as potential predictive variables of RRMSE. We also examined the situation of no prior knowledge of spatial patterns and analyzed all RRMSE results together.

Confidence Interval Simulations

Using the simulation and response surface results, we designed additional simulations to examine confidence interval coverage for the better performing (based on RRMSE) density estimation method. Four calculation methods for producing $1 - \alpha$ confidence intervals were considered, the first of which was given by Parker (1979):

$$C_1/K \leq D \leq C_2/K$$

where C_1 and C_2 are, respectively, the lower and upper $\alpha/2$ percentiles from a chi-square distribution with $2nr$ degrees of freedom, and $K = 2w\Sigma L_i$. The second formula used a normal approximation:

$$\hat{D} - N_{1-\alpha/2}\widehat{\text{var}}(\hat{D}) \leq D \leq \hat{D} + N_{1-\alpha/2}\widehat{\text{var}}(\hat{D}).$$

The third confidence interval formula was a nonparametric method based on applying a confidence interval to the median (e.g., Hollander and Wolfe 1973) from a sample of n L_i's. The calculation steps are given in the Appendix. The final confidence interval method was a combination of the first and third methods, using an analog of Hopkins (1954) index of aggregation to provide a decision criterion for which of the two confidence interval methods to apply. We considered this method because increasing deviations from a random spatial pattern (e.g., greater aggregation) may lessen the quality of confidence interval coverage for the method based on a random pattern assumption developed by Parker (1979). If so, we presumed that there would be an increased likelihood that the nonparametric method would provide superior coverage. The decision criterion that we applied was a ratio of search areas, also obtained from VAT sampling, that was analogous to the ratio given in Hopkins (1954). Let L_{1i} be the distance searched from the ith random point to the first population member and L_{2i} be the distance searched from the ith random point to the second population member. Then

$$U = \sum L_{1i}/\sum(L_{2i} - L_{1i}).$$

VAT sampling of a random spatial pattern should result in the average search distance (area for fixed-width transects) to the first individual encountered approximately equalling the average search distance from the first individual to the second individual ($U = 1$). In aggregated populations the average search distance to the first individual should exceed the average search distance from the first individual to the second individual ($U > 1$). We found through trial simulations on random populations that U rarely exceeded 1.1. We ran test simulations using $U = 1.05$ and found little difference from 1.1, $U < 1.05$ produced diminished accuracy in coverage, as did values of $U > 1.1$. Therefore, we used $U = 1.1$ as the decision point between applying Parker's formula vs. the nonparametric method presented here.

For each of the calculation methods, 90, 95, and 99% confidence intervals were considered. The simulations were conducted essentially as before with 5000 iterations, but only the better performing of the two estimation methods (original vs. Morisita) was considered and sample sizes were restricted to those identified as producing useful results in the initial simulation study.

TABLE 1. Mean RRMSE results for each density estimator using r = 3–10 in each spatial pattern at each sample size.

	Pattern									
	Random					Regular				
Sample size	5	10	20	40	Mean	5	10	20	40	Mean
Original										
r = 3	0.29	0.20	0.14	0.11	0.18	0.23	0.20	0.18	0.17	0.19
4	0.25	0.17	0.12	0.09	0.16	0.17	0.15	0.13	0.12	0.14
5	0.22	0.16	0.11	0.09	0.14	0.14	0.12	0.10	0.10	0.11
6	0.20	0.14	0.10	0.08	0.13	0.12	0.10	0.09	0.08	0.10
7	0.18	0.13	0.10	0.08	0.12	0.11	0.09	0.08	0.07	0.08
8	0.17	0.12	0.09	0.07	0.12	0.09	0.08	0.07	0.06	0.07
9	0.16	0.12	0.09	0.07	0.11	0.09	0.07	0.06	0.05	0.07
10	0.15	0.11	0.08	0.07	0.10	0.08	0.06	0.06	0.05	0.06
Mean	0.20	0.15	0.11	0.08	0.13	0.13	0.11	0.09	0.09	0.10
Morisita										
r = 3	0.45	0.33	0.23	0.17	0.29	0.27	0.20	0.15	0.12	0.19
4	0.32	0.23	0.17	0.12	0.21	0.18	0.14	0.11	0.09	0.13
5	0.27	0.19	0.14	0.10	0.18	0.14	0.11	0.09	0.08	0.10
6	0.23	0.17	0.12	0.09	0.15	0.12	0.09	0.08	0.07	0.09
7	0.21	0.15	0.11	0.08	0.14	0.10	0.08	0.07	0.06	0.08
8	0.19	0.14	0.10	0.08	0.13	0.09	0.07	0.06	0.05	0.07
9	0.18	0.13	0.10	0.07	0.12	0.09	0.07	0.06	0.05	0.07
10	0.17	0.12	0.09	0.07	0.11	0.08	0.06	0.05	0.05	0.06
Mean	0.25	0.18	0.13	0.10	0.17	0.13	0.10	0.08	0.07	0.10
	Aggregate					Double-clump				
Sample size	5	10	20	40	Mean	5	10	20	40	Mean
Original										
r = 3	0.45	0.33	0.28	0.25	0.33	0.52	0.45	0.43	0.43	0.46
4	0.44	0.31	0.24	0.21	0.30	0.49	0.39	0.35	0.34	0.39
5	0.41	0.29	0.23	0.19	0.28	0.49	0.36	0.31	0.29	0.36
6	0.35	0.25	0.20	0.17	0.24	0.50	0.36	0.30	0.27	0.36
7	0.34	0.24	0.19	0.16	0.23	0.51	0.36	0.29	0.26	0.36
8	0.34	0.24	0.19	0.16	0.23	0.51	0.36	0.29	0.26	0.36
9	0.33	0.24	0.19	0.16	0.23	0.48	0.35	0.28	0.25	0.34
10	0.32	0.24	0.19	0.16	0.23	0.40	0.30	0.25	0.23	0.29
Mean	0.37	0.27	0.21	0.18	0.26	0.49	0.36	0.31	0.29	0.36
Morisita										
r = 3	1.05	0.80	0.64	0.53	0.76	1.63	1.21	0.95	0.77	1.14
4	0.92	0.74	0.61	0.54	0.70	1.46	1.15	0.95	0.82	1.10
5	0.82	0.66	0.56	0.50	0.64	1.41	1.14	0.98	0.88	1.10
6	0.69	0.54	0.45	0.30	0.52	1.39	1.14	0.99	0.91	1.11
7	0.67	0.53	0.45	0.40	0.51	1.35	1.13	1.00	0.92	1.10
8	0.65	0.52	0.45	0.40	0.51	1.29	1.08	0.97	0.90	1.06
9	0.62	0.51	0.44	0.40	0.49	1.16	0.98	0.87	0.81	0.95
10	0.58	0.48	0.42	0.38	0.46	0.94	0.78	0.67	0.61	0.75
Mean	0.75	0.60	0.50	0.44	0.57	1.33	1.08	0.92	0.83	1.04

Because it was equally easy in our simulation algorithm to acquire results for all values of r from 3 to 10, we did not restrict ourselves to considering only an optimal range of r. Confidence interval coverage, for a given target coverage, calculation method, n, and r, was calculated as the percentage of the 5000 iterations for which confidence intervals contained the true population density.

RESULTS AND DISCUSSION

The results from the 128 population simulations (two runs of each of the 64 combinations) are summarized as the mean RRMSE for each estimator in each spatial pattern at each sample size in Table 1. We present more condensed results for relative bias (RBIAS) in Table 2. Each simulation run involved 5000 replications of a spatial pattern-by-density-by-sample size combination. Besides the extensive testing involved in developing a simulation program, various aspects of the results served to confirm that the program was functioning properly. Each of these combinations was run twice, with the difference being the seeds for the procedures required by the random number generator. There was very little variability between simulation runs, with some of the resulting statistics often being identical (to four decimal places) between the two runs. In addition, the results for the original estimator at r = 3 were nearly identical to those of Engeman et al. (1994) in those cases where the simulation frameworks were the same.

TABLE 2. Mean relative bias (RBIAS) for each density estimator using $r = 3$–10 in each spatial pattern.

Estimator	Random	Regular	Aggregate	Double-clump
Original				
$r = 3$	0.03	0.14	−0.17	−0.38
4	0.03	0.10	−0.08	−0.26
5	0.03	0.07	−0.03	−0.16
6	0.03	0.06	−0.05	−0.07
7	0.03	0.05	−0.02	−0.01
8	0.03	0.05	0.01	0.04
9	0.03	0.04	0.04	0.05
10	0.03	0.04	0.05	0.00
Mean	0.03	0.07	−0.03	−0.10
Morisita				
$r = 3$	0.03	−0.07	0.37	0.52
4	0.03	−0.07	0.44	0.66
5	0.03	−0.06	0.42	0.75
6	0.03	−0.05	0.32	0.81
7	0.03	−0.05	0.33	0.84
8	0.03	−0.04	0.35	0.82
9	0.03	−0.04	0.35	0.74
10	0.03	−0.04	0.33	0.53
Mean	0.03	−0.05	0.37	0.71

We now can make some general observations based on the results, while saving specific recommendations for the *Conclusions* section. The most obvious result was that the original form of the VAT estimator almost invariably outperformed the Morisita analog. The superior performance of the original estimator was most in evidence in the aggregate and the double-clumped spatial patterns, for which the Morisita analog was developed (Parker 1979). For most of the simulation combinations, definite improvement in estimation was made with $r > 3$, although the relative magnitudes of the improvements usually diminished after $r = 6$ (Tables 1 and 2, Fig. 1). As would be expected, estimation improved with increasing sample size. RRMSEs tended to improve (decrease) as densities increased, especially for populations with one of the patterns involving aggregation, although this tendency was not nearly so pronounced as those differences among estimation formulas or population spatial patterns. One might expect performance to improve for aggregated populations as density increases, because populations of randomly placed clumps of individuals begin to appear more like the random populations that have the same density.

The RRMSE results for the variance estimators associated with each estimation formula are given in Table 3. There we can see that quality of estimation was much greater for the populations with a greater degree of randomness in locating individuals (and clumps) than for the regular population where individuals were more systematically placed. Fortunately, population patterns where the members are systematically located

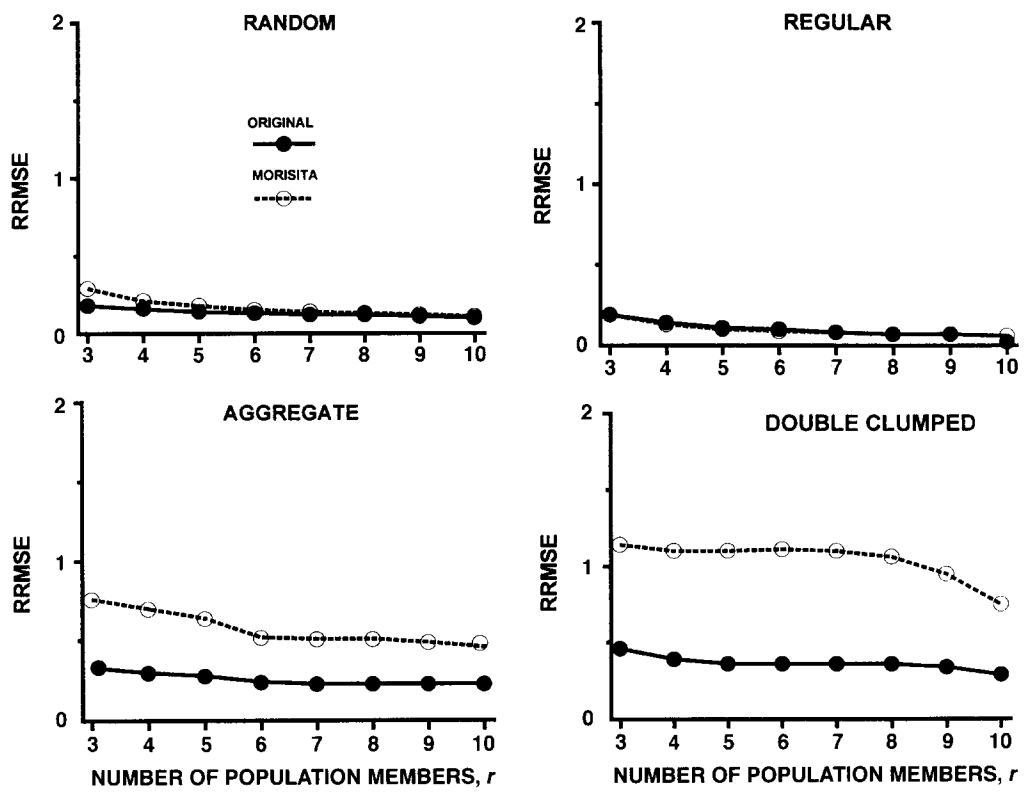

FIG. 1. Mean RRMSE responses in each population spatial pattern for $r = 3$–10 using the original and Morisita analog calculation formulas.

TABLE 3. Mean RRMSE for the variance of each density estimator using r = 3–10 in each spatial pattern at each sample size.

Sample size	Pattern: Random					Regular				
	5	10	20	40	Mean	5	10	20	40	Mean
Original										
r = 3	0.67	0.41	0.27	0.20	0.39	1.72	1.48	1.38	1.30	1.47
4	0.54	0.34	0.23	0.18	0.32	2.04	1.86	1.77	1.69	1.84
5	0.46	0.30	0.21	0.17	0.28	2.37	2.23	2.14	2.04	2.19
6	0.41	0.27	0.19	0.16	0.26	2.59	2.50	2.40	2.33	2.45
7	0.38	0.25	0.18	0.15	0.24	2.84	2.79	2.67	2.63	2.73
8	0.35	0.23	0.17	0.14	0.22	3.02	2.98	2.87	2.81	2.91
9	0.33	0.22	0.15	0.13	0.21	3.29	3.22	3.11	3.09	3.18
10	0.31	0.21	0.15	0.12	0.20	3.40	3.35	3.27	3.20	3.30
Mean	0.43	0.28	0.19	0.16	0.26	2.66	2.55	2.45	2.39	2.51
Morisita										
r = 3	1.81	1.07	0.65	0.43	0.99	3.05	2.36	2.03	1.81	2.31
4	0.89	0.54	0.35	0.24	0.50	2.72	2.51	2.42	2.26	2.48
5	0.64	0.41	0.27	0.19	0.38	3.02	2.95	2.84	2.74	2.89
6	0.52	0.34	0.23	0.17	0.32	3.25	3.18	3.12	3.05	3.15
7	0.45	0.30	0.21	0.16	0.28	3.50	3.52	3.43	3.41	3.46
8	0.40	0.27	0.19	0.15	0.25	3.67	3.70	3.60	3.56	3.63
9	0.37	0.25	0.17	0.14	0.23	3.93	3.93	3.85	3.84	3.88
10	0.34	0.23	0.16	0.13	0.22	4.02	4.01	3.97	3.90	3.97
Mean	0.68	0.43	0.28	0.20	0.40	3.40	3.27	3.16	3.07	3.22
Sample size	Aggregate					Double-clump				
	5	10	20	40	Mean	5	10	20	40	Mean
Original										
r = 3	0.78	0.67	0.65	0.65	0.69	0.93	0.74	0.72	0.73	0.78
4	0.73	0.68	0.68	0.69	0.69	0.85	0.77	0.76	0.78	0.79
5	0.72	0.69	0.69	0.70	0.70	0.83	0.79	0.79	0.80	0.80
6	0.72	0.67	0.68	0.70	0.69	0.83	0.80	0.80	0.82	0.81
7	0.70	0.69	0.69	0.71	0.70	0.83	0.81	0.81	0.83	0.82
8	0.71	0.70	0.70	0.72	0.71	0.84	0.82	0.82	0.83	0.83
9	0.72	0.71	0.71	0.73	0.71	0.83	0.82	0.82	0.83	0.83
10	0.72	0.71	0.71	0.73	0.72	0.81	0.80	0.81	0.83	0.81
Mean	0.72	0.69	0.69	0.70	0.70	0.84	0.79	0.79	0.81	0.81
Morisita										
r = 3	1.93	1.08	0.70	0.61	1.08	2.51	1.18	0.81	0.77	1.32
4	0.89	0.76	0.67	0.66	0.74	0.93	0.84	0.80	0.80	0.84
5	0.76	0.71	0.69	0.70	0.71	0.85	0.82	0.82	0.82	0.83
6	0.76	0.73	0.74	0.74	0.74	0.85	0.83	0.84	0.84	0.84
7	0.77	0.75	0.75	0.76	0.76	0.85	0.84	0.85	0.85	0.84
8	0.77	0.76	0.76	0.77	0.77	0.86	0.85	0.86	0.86	0.86
9	0.77	0.77	0.77	0.78	0.78	0.86	0.87	0.87	0.87	0.87
10	0.78	0.78	0.78	0.79	0.78	0.88	0.88	0.89	0.89	0.88
Mean	0.93	0.79	0.73	0.73	0.80	1.07	0.89	0.84	0.84	0.91

are probably the least likely to occur in nature. The RRMSE results for the original variance formula were invariably smaller than for the Morisita formula. The Morisita variance estimate showed considerable improvement in RRMSE at the two smaller sample sizes when r was increased from 3 to 4. In general, the effects from increasing n or r were more pronounced for the Morisita variance estimates than for the original formula, even though the RRMSE magnitudes remained consistently higher (poorer) for the Morisita variance estimate. The most substantial improvement in RRMSE for the original variance formula was shown when n or r increased in the random spatial pattern.

The original estimator, by providing consistently superior estimation to the Morisita formula, was used for development of prediction equations for the RRMSE results. Various degrees of quality (Table 4) were achieved across the spatial patterns for predicting RRMSE (using the original estimator). For each pattern the best prediction was produced by an additive model using n and r as predictive variables. The product of n and r, by itself or in combination with n and/or r, did not contribute to improved prediction. If an investigator has information as to which of these spatial patterns a population tends to follow, then the equations (Table 4) may be of use in some situations for developing a sampling plan by examining the trade-offs between sample size and r. Because it is common for an

TABLE 4. Parameters at each spatial pattern of the best predictive models for RRMSE using sample size and r, where the density estimate is calculated by Parker's (1979) original formula.

Pattern	Parameter estimates			
	Intercept	Sample size	r	R^2
Random	0.2625	−0.0031	−0.0109	0.72
Regular	0.2335	−0.0010	−0.0170	0.77
Aggregate	0.4376	−0.0046	−0.0143	0.47
Double-clump	0.5637	−0.0047	−0.0171	0.28
Overall	0.3743	−0.0033	−0.0148	0.16

investigator not to have a clear notion in advance of the spatial pattern of the population to be sampled, we also examined the quality of predictive model produced overall without regard to spatial pattern (Table 4). Again, n and r in an additive model produced the best prediction, but the quality of that prediction was rather low ($R^2 = 0.16$). Thus, as we describe in the Conclusions section, it may be best for the investigator to identify an overall range for n and r where performance is acceptable. Within this range for n and r, the logistics of the field situation can determine the trade-off between n and r.

We examined confidence interval coverage by the four methods for sample sizes of 20 and 40 (Table 5). The method given by Parker (1979) produced excellent coverage in the random spatial patterns for all values of r from 3 to 10, for both samples sizes (20 and 40). This method provided inconsistent results among target coverages for the regular spatial pattern, although the results for $n = 20$ and $r = 4$ or 5 were probably adequate for each of the $1 - \alpha$ target coverages. Most notably, the Parker method, derived under the assumption of a Poisson pattern, produced very poor coverage for population patterns having aggregation, never achieving 80% (generally much less) of any target coverage, for any r, at either sample size.

The normal confidence intervals followed a pattern of results similar to the Parker method. Excellent coverage in the random pattern occurred for all r and n. This method, too, exhibited inconsistent coverage in the regular pattern, producing much higher than target coverage in most situations (very conservative). In contrast, this method produced even poorer coverage than the Parker formula for the population patterns having some form of aggregation.

The nonparametric method generally exceeded target coverage in the random and regular patterns for $n = 20$, but generally fell slightly short of target for $n = 40$. (With $r = 6$ or 7, results were excellent for $n = 40$.) The regular spatial pattern was a challenge for this method too, as coverage results were always short of target. This method also tended to be conservative for the aggregated patterns by providing actual coverage in excess of the target coverage. As one might expect, the combination of the nonparametric method with the Parker method based on a decision criterion concerning spatial pattern provided results balanced between the two methods. Using $r = 5$ or 6 usually resulted in coverage reasonably near target, even for the regular pattern with $n = 20$.

CONCLUSIONS

We conclude this paper by giving our recommendations, based on the results presented here, as to how VAT sampling would be best applied in the field. Although Engeman et al. (1994) demonstrated that VAT sampling with $r = 3$ was one of the more practical methods among the spectrum of plotless density estimators displaying good estimation properties, we believe that valuable improvement for the design of field studies would be realized using $r > 3$. The greatest improvement in estimation properties generally occurred when r was increased from 3 to 4, but additional improvements in performance occurred using $r = 5$ or 6. The two estimators that displayed the best overall estimation properties in Engeman et al. (1994) involved angle-order sampling (Morisita 1957) with measurements to the second and third nearest population member in each quadrant. It should be noted that had the VAT estimator in that same study used $r = 5$ or 6, it would have displayed comparable estimation properties as these two angle-order estimators, but it would have required locating fewer population members while employing a much easier (and more practical) search pattern from each random starting point.

We suggest that the largest $r \leq 6$ that can be practically applied in the field should be used. Saying that, we also realize that, depending on the field situation and the cost or effort of locating population members along each VAT relative to locating the starting points for the VATs, there may be trade-offs between the number of VATs (n) that can be applied in the field and the number of population members that can be located along each VAT (r). Beyond $r = 6$, the relative improvement in estimation could well be counterbalanced by additional effort in many field situations. That is, if field conditions are "easy" enough to use $r > 6$, then quadrat sampling should be considered.

For calculation of estimates from VAT sampled data, we recommend the use of the original formula rather than the Morisita analog. Unless the population follows a very regular or systematic spacing pattern (unlikely, but may occur occasionally), the variance formula associated with the original formula appears satisfactory. For confidence interval calculation we recommend using the nonparametric/Parker combined method with a decision criterion of $U = 1.1$. If one wanted to consider only one confidence interval formula, then the nonparametric formula is preferable, due to the very poor results in aggregated populations by the Parker formula. However, in the relatively uncommon situations where the population is systematically placed in a repeated pattern (regular population pattern), and if sam-

TABLE 5. Average percentage confidence interval coverage results using sample sizes of 20 and 40 in four spatial patterns, across four population densities, for four methods of a calculation using r = 3–10.

r	CI	Parker	Normal	Nonpara-metric	Com-bined	r	CI	Parker	Normal	Nonpara-metric	Com-bined
Random spatial pattern (n = 20)						Regular spatial pattern (n = 20)					
3	90	90.0	90.6	91.3	90.5	3	90	82.1	90.1	69.7	80.4
	95	95.0	95.3	96.7	95.6		95	92.1	97.0	84.6	91.2
	99	99.0	99.0	99.0	99.0		99	95.2	99.9	93.7	98.3
4	90	89.7	90.3	92.2	90.6	4	90	87.8	93.4	73.0	85.6
	95	94.9	95.4	97.2	95.7		95	95.5	98.3	86.4	94.1
	99	98.9	99.0	99.2	99.0		99	97.9	>99.9	94.9	99.1
5	90	89.7	90.5	92.7	90.7	5	90	92.7	96.0	75.8	90.1
	95	94.8	95.6	97.3	95.7		95	97.7	99.1	88.6	96.4
	99	98.9	99.1	99.3	99.2		99	99.4	>99.9	95.7	99.3
6	90	89.7	90.7	93.0	91.1	6	90	94.3	96.9	77.3	91.2
	95	94.9	95.4	97.7	95.9		95	98.4	99.4	89.3	97.0
	99	99.0	99.2	99.4	99.1		99	99.6	>99.9	95.9	99.3
7	90	89.7	90.5	93.1	91.1	7	90	95.9	98.0	78.4	92.9
	95	94.9	95.4	97.7	96.0		95	99.1	99.7	90.0	97.9
	99	98.9	99.1	99.4	99.1		99	99.8	>99.9	96.4	99.4
8	90	89.5	90.3	93.4	91.0	8	90	97.0	98.4	80.3	94.3
	95	94.7	95.2	97.8	95.7		95	99.4	99.8	91.2	98.2
	99	98.8	99.1	99.4	99.0		99	99.9	>99.9	96.9	99.5
9	90	89.5	90.3	93.2	91.0	9	90	97.9	98.9	81.2	95.0
	95	94.6	95.2	97.9	95.9		95	99.6	99.8	91.5	98.3
	99	98.9	99.1	99.5	99.1		99	99.9	>99.9	97.2	99.5
10	90	89.3	90.1	93.5	91.0	10	90	98.0	99.0	82.1	95.4
	95	94.7	95.2	97.7	95.7		95	99.7	99.9	92.2	98.4
	99	98.9	99.1	99.5	99.0		99	>99.9	>99.9	97.4	99.6
Random spatial pattern (n = 40)						Regular spatial pattern (n = 40)					
3	90	89.4	90.1	85.4	88.2	3	90	57.8	67.6	41.7	56.8
	95	94.7	95.0	91.6	93.6		95	75.8	85.0	55.2	74.5
	99	98.8	99.0	97.9	98.4		99	99.0	98.4	79.7	94.1
4	90	89.2	89.9	87.3	88.7	4	90	68.2	76.3	47.5	66.5
	95	94.4	94.9	93.0	94.1		95	84.5	90.3	61.1	82.6
	99	98.7	98.8	98.4	98.6		99	99.7	99.3	83.0	96.7
5	90	89.2	89.7	88.3	88.9	5	90	80.2	85.4	53.7	77.9
	95	94.3	94.6	93.7	93.9		95	92.0	95.4	66.5	89.7
	99	98.7	98.9	98.7	98.7		99	99.9	99.8	87.0	98.4
6	90	88.9	89.7	89.7	89.1	6	90	83.5	87.9	56.7	81.0
	95	94.1	94.6	94.4	94.2		95	94.0	96.4	68.8	91.7
	99	98.7	98.9	98.7	98.7		99	>99.9	99.8	88.2	98.5
7	90	88.6	89.4	90.0	89.2	7	90	87.8	91.2	59.3	84.9
	95	93.9	94.6	94.6	94.3		95	96.2	97.8	71.5	93.8
	99	98.7	98.9	98.8	98.8		99	>99.9	>99.9	89.3	98.8
8	90	88.6	89.4	90.3	89.3	8	90	91.0	93.6	62.0	87.9
	95	93.9	94.6	95.0	94.5		95	97.3	98.5	74.0	95.1
	99	98.6	98.9	98.9	98.7		99	>99.9	>99.9	90.9	99.1
9	90	88.3	89.2	90.4	89.0	9	90	93.0	95.1	64.3	89.8
	95	93.8	94.4	94.9	94.1		95	98.2	98.9	76.2	95.9
	99	98.5	98.9	99.0	98.7		99	>99.9	>99.9	91.8	99.1
10	90	88.4	89.2	90.6	89.1	10	90	93.7	95.5	66.4	90.1
	95	93.9	94.5	95.1	94.3		95	98.5	99.1	77.4	96.1
	99	98.6	98.9	98.9	98.7		99	>99.9	>99.9	92.2	99.2

ple sizes are adequate (say $n \geq 20$), then a bootstrap variance and confidence interval might be considered. In obvious regular patterns alternative sampling plans to VAT, such as quadrat sampling, might also be considered.

We sought to optimize the application of VAT sampling through extensive simulations. In natural situations the pattern and density of populations can vary greatly and a simulation using artificial populations can only approximate natural processes. Therefore, the true test of the results presented here would be verification using a wide variety of fully enumerated field data sets. We currently are accumulating a series of such field data sets, which eventually can be used for further evaluation of VAT sampling. At present, it appears that the VAT sampling method using r = 5 or 6 and the original formula for calculating estimates offers the field investigator a method that is relatively easy to apply, confusion-free, and produces high-quality estimates. Future efforts might point toward developing a sequential stopping criterion in which r could differ among the individual VATs according to a measure of the immediate density of objects. This would place VAT sampling in the realm of adaptive sampling

TABLE 5. Extended.

r	CI	Parker	Normal	Nonpara-metric	Com-bined	r	CI	Parker	Normal	Nonpara-metric	Com-bined
Aggregate spatial pattern (n = 20)						Double-clump spatial pattern (n = 20)					
3	90	29.8	27.1	95.5	88.8	3	90	3.3	2.8	88.2	85.3
	95	36.4	32.3	98.8	92.5		95	4.5	3.6	95.2	92.3
	99	50.1	43.5	99.6	94.8		99	7.7	5.9	98.5	95.7
4	90	42.3	39.7	94.9	88.9	4	90	9.2	8.2	93.3	90.6
	95	50.2	46.5	98.4	93.0		95	11.8	10.4	97.6	94.8
	99	63.8	58.8	99.6	95.6		99	17.9	14.9	99.3	96.7
5	90	47.9	45.7	94.6	89.1	5	90	17.4	16.0	95.3	92.7
	95	55.2	52.6	98.2	93.2		95	21.3	19.5	98.5	95.8
	99	68.8	64.7	99.6	96.1		99	29.8	26.5	99.7	97.3
6	90	44.6	42.5	95.5	89.9	6	90	25.3	24.1	95.2	92.8
	95	52.4	49.7	98.7	93.8		95	30.6	28.7	98.7	96.3
	99	66.0	62.1	99.8	96.2		99	40.6	37.8	99.7	97.5
7	90	48.8	47.0	95.3	90.2	7	90	32.0	30.9	94.5	92.3
	95	57.0	54.4	98.6	94.1		95	38.1	36.5	98.2	96.0
	99	70.4	67.0	99.7	96.6		99	49.3	46.7	99.5	97.6
8	90	52.9	51.6	94.7	90.0	8	90	35.7	34.8	93.5	91.3
	95	61.0	59.0	98.3	94.1		95	42.3	41.0	97.8	95.8
	99	74.3	71.5	99.6	96.7		99	54.0	52.0	99.4	97.7
9	90	55.2	54.1	94.4	90.0	9	90	36.6	35.8	94.0	91.9
	95	63.4	62.0	98.2	94.4		95	43.0	42.0	97.9	95.9
	99	76.3	74.7	99.6	97.0		99	55.2	53.3	99.5	97.7
10	90	56.7	56.1	94.1	89.9	10	90	33.1	32.1	95.3	93.2
	95	64.6	63.6	98.0	94.3		95	38.9	37.6	98.6	96.5
	99	77.8	76.1	99.5	97.1		99	50.5	48.2	99.6	97.8
Aggregate spatial pattern (n = 40)						Double-clump spatial pattern (n = 40)					
3	90	13.0	11.6	95.8	90.1	3	90	0.2	0.2	80.8	79.9
	95	17.4	15.2	98.1	92.7		95	0.3	0.3	88.3	87.3
	99	27.4	23.7	99.8	95.1		99	0.7	0.6	96.7	95.7
4	90	27.5	25.7	94.4	89.3	4	90	1.8	1.6	91.4	90.4
	95	33.8	31.3	97.3	92.4		95	2.4	2.1	95.5	94.5
	99	46.8	42.6	99.5	95.3		99	4.1	3.4	99.2	98.1
5	90	34.7	33.3	93.7	88.9	5	90	6.6	6.2	94.9	93.9
	95	41.7	39.3	96.9	92.3		95	8.6	7.9	97.6	96.7
	99	55.3	51.9	99.3	95.6		99	13.3	11.8	99.6	98.7
6	90	30.7	29.0	95.7	91.0	6	90	14.1	13.6	95.0	94.1
	95	37.2	35.1	98.1	93.8		95	17.5	16.5	97.6	96.7
	99	50.9	47.3	99.7	96.2		99	24.6	22.9	99.6	98.7
7	90	37.2	35.6	95.4	90.9	7	90	22.0	21.3	93.1	92.2
	95	42.3	42.2	97.8	93.7		95	26.1	25.2	96.6	95.6
	99	57.8	54.9	99.7	96.5		99	34.9	33.0	99.3	98.4
8	90	43.5	42.1	94.3	90.2	8	90	26.5	26.0	91.1	90.2
	95	51.0	49.8	97.2	93.4		95	31.8	30.9	95.1	94.2
	99	65.0	62.5	99.5	96.7		99	41.4	39.9	98.8	97.9
9	90	47.5	46.5	93.0	89.3	9	90	27.7	27.2	91.8	91.0
	95	55.9	54.6	96.4	93.0		95	32.9	32.2	95.6	94.8
	99	69.1	67.3	99.3	96.9		99	43.1	41.8	98.9	98.1
10	90	49.5	48.7	92.7	89.1	10	90	22.4	21.7	94.8	93.9
	95	57.6	56.4	96.2	93.1		95	26.7	26.0	97.4	96.5
	99	71.3	69.8	99.2	96.9		99	36.1	34.3	99.6	98.8

(Thompson 1992), whereby some of the associated efficiencies could be realized.

ACKNOWLEDGMENTS

We would like to thank William Dusenberry for review and comments on this work prior to submission. Gary White, Doug Johnson, and an anonymous reviewer provided suggestions that substantially improved the work.

LITERATURE CITED

Batcheler, C. L. 1975. Development of a deer census from pellet groups. Journal of Wildlife Management **39**:641–652.

Bratley, P., B. L. Fox, and L. E. Schrage. 1983. A guide to simulation. Springer-Verlag, New York, New York, USA.

Brody, M., and D. Morais. 1987. UNIF is fast, reliable random number generator for IBM-AT microcomputer. Research Information Bulletin Number 87-84. USDI Fish and Wildlife Service, Washington, D.C., USA.

Clayton, G., and T. F. Cox. 1986. Some robust density estimators for spatial point processes. Biometrics **42**:753–767.

Delince, J. 1986. Robust density estimation through distance measurements. Ecology **67**:1576–1581.

Diggle, P. J. 1975. Robust density estimation using distance methods. Biometrika **62**:39–48.

Engeman, R. M., and J. F. Bromaghin. 1990. An approach

to estimating density from line transect data where animals move in response to the observer. Journal of Statistical Computation and Simulation **34**:145–154.

Engeman, R. M., R. T. Sugihara, L. F. Pank, and W. E. Dusenberry. 1994. A comparison of plotless density estimators using Monte Carlo simulation. Ecology **75**:1769–1779.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric statistical methods. Wiley, New York, New York, USA.

Hopkins, B. 1954. A new method for determining the type of distribution of plant individuals (with appendix by J. G. Skellum). Annals of Botany, New Series **18**:213–227.

Lewis, S. M. 1975. Robust estimation of density for a two-dimensional point process. Biometrika **62**:519–521.

Morisita, M. 1957. A new method for the estimation of density by spacing method applicable to nonrandomly distributed populations. (In Japanese; available as Forest Service translation Number 11116, USDA Forest Service, Washington, D.C., USA.) Physiology and Ecology **7**:134–144.

Parker, K. R. 1979. Density estimation by variable area transect. Journal of Wildlife Management **43**:484–492.

Patil, S. A., K. P. Burnham, and J. L. Kovner. 1979. Nonparametric estimation of plant density by the distance method. Biometrics **35**:597–604.

Persson, O. 1971. The robustness of estimating density by distance measurements. Pages 175–187 *in* G. P. Patil, E. C. Pielou, and W. E. Waters, editors. Statistical ecology. Pennsylvania State University Press, University Park, Pennsylvania, USA.

Pollard, J. H. 1971. On distance estimators of density in randomly distributed forests. Biometrics **27**:991–1002.

Thompson, S. K. 1992. Sampling. Wiley, New York, New York, USA.

APPENDIX

Calculation method for a nonparametric $1 - \alpha$ confidence interval on density from variable area transect sampling, when the transect is followed from each random sampling point until r population members are encountered. At any given random sampling point let w = the width of the variable area transect (VAT) and let L_i = the distance searched along that transect to the rth population member. Let n = the number of such random sampling points from which VATs were applied.

1) Order the n L_i's from smallest to largest such that

$$L^{(1)} < L^{(2)} < \ldots < L^{(n)}.$$

2) Calculate the value C to the nearest integer where

$$C = (n/2) - Z_{\alpha/2}(n/4)^{1/2}.$$

3) The lower (LL) and upper (UL) $1 - \alpha$ confidence limits are, respectively:

$$\text{LL} = (nr - 1)(nwL^{(n+1-C)})$$

$$\text{UL} = (nr - 1)(nwL^{(C)}).$$