June 2005

Evaluation of optimized variable area transect sampling using totally enumerated field data sets

Richard M. Engeman
USDA-APHIS-Wildlife Services, s_r100@yahoo.com

Ryan M. Nielson
USDA-APHIS-Wildlife Services

Robert T. Sugihara
USDA/Wildlife Services, National Wildlife Research Center, Hawaii Field Station

Evaluation of optimized variable area transect sampling using totally enumerated field data sets[‡]

Richard M. Engeman[1]*,[†], Ryan M. Nielson[1,2] and Robert T. Sugihara[3]

[1]*USDA/Wildlife Services, National Wildlife Research Center, 4101 LaPorte Ave, Fort Collins, CO 80521-2154, U.S.A.*
[2]*(Current address:) Western Ecosystems Technology, Inc., 2003 Central Ave., Cheyenne, WY 82001, U.S.A.*
[3]*USDA/Wildlife Services, National Wildlife Research Center, Hawaii Field Station, P.O. Box 10880, Hilo, Hawaii 96721, U.S.A.*

SUMMARY

Plotless density estimators (PDEs) can be efficient alternatives to quadrat sampling for estimating the density of stationary objects. Variable area transect (VAT) sampling had been identified, and optimized relative to effort, in previous Monte Carlo simulated population studies as a low-labor field method that demonstrated superior estimation properties among many PDEs considered. However, natural populations tend to be much more complex and less predictable in spatial distribution than computer generated populations. Therefore, we carried out a Monte Carlo simulation study that used 17 fully enumerated field populations rather than simulated populations. These natural populations represented a variety of population densities and spatial patterns. We focused on assessing the effect of the number of population members searched for along each transect (r), sample size, and transect width on estimation quality. Using relative root mean-squared error and relative bias as criteria, the optimal range for r was 5 to 7 population members encountered from each start point. Sample size was best if $n > 20$, but returns in estimation quality diminished by $n = 40$. Transect width was previously uninvestigated and found to be the most important design factor affecting estimation quality. Field studies should strive for transects as wide as logistically reasonable. Published in 2005 by John Wiley & Sons, Ltd.

KEY WORDS: density estimation; distance sampling; spatial pattern

1. INTRODUCTION

The estimation of density of such stationary objects as plant communities, points of crop damage, bird nests, or animal burrows is essential to many areas of ecology. The ideal density sampling method would produce an unbiased estimate, be robust to different population spatial patterns and densities, and be easily applied in various field situations. Sampling with randomly placed quadrats and thoroughly searching for every population member in each quadrat produces unbiased density estimates and is robust over population spatial patterns (given an appropriate quadrat size) (e.g. Engeman *et al.*, 1994; Krebs, 1999) However, it can be labor intensive and difficult to apply in some

---

field situations. Cottom (1947) introduced distance sampling as a method for more easily obtaining density estimates, and many distance methods, or plotless density estimators (PDEs), have since been developed. Besides offering more labor-efficient alternatives to quadrat sampling, PDEs often can be characterized as 'low impact', especially important when studying species listed as 'at risk' under the Endangered Species Act.

Variable area transect (VAT) sampling and estimation (Parker, 1979) was identified in a large simulation study as one of the most promising PDEs, based on estimation quality and in-field practicality (Engeman *et al.*, 1994). The VAT 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 population members and the measurement of distances to them are accomplished simultaneously, making this an attractive method to apply in the field. A Monte Carlo simulation using computer-simulated populations found superior estimation properties for $r > 3$, but diminishing returns, relative to field effort for $r > 6$ (Engeman and Sugihara, 1998).

The results from simulated populations were highly informative, but natural populations are more complex, occurring as mixtures of spatial patterns at varying levels of intensity and grain (intensity is the variability in pattern from place to place and grain expresses the amount of spacing between them [Pielou, 1959]). Statistical performance information derived from simulated populations obtains greater value if verified using naturally occurring populations. The most practical approach for achieving this is to fully enumerate a variety of field populations and repeatedly sample them in Monte Carlo simulations to calculate density estimates and their properties. We took this approach to further evaluate optimization of VAT sampling.

2. METHODS

2.1. Variable area transect estimation

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 is

$$\hat{D} = (nr - 1) \Big/ \Big[w \sum L_i \Big]$$

2.2. Simulation study design

Seventeen naturally occurring populations (Table 1) covering a wide range of densities and spatial patterns each had every population member located and their *X-Y* Cartesian co-ordinates recorded as a data set. The data sets were used in a series of Monte Carlo simulations that examined the effect of sample size (n), the number of population members for which to search along each VAT (r), and the width of the VAT (w). The unit of density of the enumerated populations is unimportant for conceptual purposes as long as the metric for the sampling distances, L_i, is squared to form the density metric. Thus, natural population densities are reported as objects/distance-unit^2 (noted as objects/u^2). Densities ranged from 0.56/u^2 (grasstree distribution) to 1985.02/u^2 (pocket gopher burrows). Pielou's (1959) index of non-randomness, α, was used to describe the direction and degree to which population spatial distribution departs from complete randomness (Table 1), where $\alpha < 1$ for regular or uniform

Table 1. Fully enumerated field populations used in variable area transect sampling simulations. A short name for each data set is given along with a brief description of their origins, its density, and Pielou's (1959) index of non-randomness as a descriptor of their spatial pattern

Population name	Description	Density (objects/u^2)	Pielou's index
Bee-eater	Bee-eater cliffside nests, Africa	581.75	0.75
Koala tree	Trees used by radioed koalas, Queensland	10.49	57.84
Pocket gopher1	Pocket gopher burrows, Washington state	1985.03	0.96
Pocket gopher2	Pocket gopher burrows, Washington state	1867.45	1.01
Rice3	Rat damage in rice study field 3, Philippines	2.17	5.29
Rice4	Rat damage in rice study field 4, Philippines	1.45	1.71
Rice5	Rat damage in rice study field 5, Philippines	8.41	6.94
Rice6	Rat damage in rice study field 6, Philippines	1.74	4.03
Rice7	Rat damage in rice study field 7, Philippines	1.63	1.65
Sugar1	Rat damage in sugarcane field 1, Hawaii	4.35	1.63
Sugar2	Rat damage in sugarcane field 2, Hawaii	14.49	2.88
Sugar3	Rat damage in sugarcane field 3, Hawaii	3.57	2.33
Venus	Craters on planet surface, Venus	355.76	1.19
Waterfowl	Waterfowl nests, Alaska	3.18	7.63
Xanthorhea1	Grass tree plot 1, Bribie Island, Queensland	0.6	1.49
Xanthorhea2	Grass tree plot 2, Bribie Island, Queensland	0.56	1.49
Xanthorhea3	Grass tree plot 3, Bribie Island, Queensland	1.06	1.19

patterns, and $\alpha > 1$ for clustered populations. Spatial patterns of the populations ranged from regular ($\alpha = 0.75$) to highly clustered ($\alpha = 57.84$), with a variety of values in between, including random ($\alpha = 1$).

The number of start points (sample size, n) considered for simulation purposes was 20 or 40, because with more ideal, computer-generated populations, $n < 20$ produced inadequate results, with diminishing improvements for increasing sample size from 20 to 40 (Engeman and Sugihara, 1998). The uniform random number generator, UNIF routine (Bratley *et al.*, 1983), was used to locate random starting points for VAT sampling. Four VAT widths were considered: 10, 20, 30, and 40 distance-units. At each random sampling point, the strip transect was extended and searched until 8 population members were encountered, enabling VAT density estimates to be calculated using $r = 3, 4, 5, 6, 7$, and 8. A single simulation run was defined by the combination of sample size, transect width, and population used. The sampling and estimation was repeated 5000 times (5000 iterations) within each run of the simulation program. Each such run of 5000 iterations was repeated to provide two replicates of each simulation scenario.

The observed statistics accumulated over the 5000 replications for 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

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

where $\hat{D}$ was the estimated density, D was the true density, and $I = 5000$ was the number of replications in the simulation run. RRMSE was the primary criterion for comparing performance among estimates (see, for example, Patil *et al.*, 1979; Engeman *et al.*, 1994; Engeman and Sugihara,

1998; Nielson *et al.*, 2004), because it encompasses variance and bias, and is unitless. Also evaluated was the relative bias (RBIAS)—the mean observed bias divided by the true parameter value. The statistics presented from these simulations were calculated as 'relative' statistics (divided by the true density of the field population used in the simulation) to standardize the scale across the densities for each of the populations. An ideal estimator is one that is robust across many spatial patterns, i.e. RRMSE and RBIAS are low, and where the amount of fieldwork required can be minimized or at least be undertaken efficiently. To better understand how each of the design factors and their combinations influenced RRMSE and RBIAS, we evaluated the simulation results using a four-factor ANOVA that included all of the three investigator-controlled factors r, n, w, and field population as the final factor.

3. RESULTS

There were 48 combinations of the levels of three factors ($r = 3, 4 \ldots 8$; $n = 20, 40$; $w = 10, 20, 30, 40$). The mean RRMSE and RBIAS result for each level of each factor averaged over all simulation runs provided an overview of each factor's impacts on estimation (Table 2). The number of population members (r) for which to search that minimized RRMSE and RBIAS was $r = 7$, although there were diminishing returns beyond $r = 5$ (Table 2). These results are similar to those for computer-generated populations, where r was optimal at $r = 5$ or 6 (Engeman and Sugihara, 1998), but our application to natural populations indicated that $r = 7$ should also be considered. Doubling the sample size from 20 to 40 improved estimation (Table 2), but as was found for computer-generated populations, adequate estimation may be obtained for $n < 40$. The magnitude of improvement in RRMSE and RBIAS between $n = 20$ and $n = 40$ remained virtually constant across all r. The present results indicate that transect width can have a large impact on RRMSE and RBIAS (Table 2), with estimation typically improving as transect width increased.

When considering which factors had the greatest impacts on RRMSE and RBIAS results, the factor in the ANOVA with the largest mean squared error (MSE) by far was transect width, w, followed by field population and its interaction with w (Table 3). The factors controllable by the investigator did not appear to interact with each other. Increasing transect width had the least impact on estimation quality

Table 2. Summary of the mean relative root mean-squared error (RRMSE) and mean relative bias (RBIAS) results for each of the investigator-controlled design factors for variable area transect sampling obtained from a four-factor factorial analysis of variance on Monte Carlo simulation results using 17 fully enumerated field populations

Simulation factor	Level	RRMSE	RBIAS
r	3	0.45	0.29
	4	0.42	0.29
	5	0.39	0.28
	6	0.38	0.27
	7	0.36	0.25
	8	0.37	0.27
Sample size	20	0.41	0.28
	40	0.38	0.27
Strip transect width	10	0.91	0.87
	20	0.31	0.22
	30	0.20	0.05
	40	0.17	−0.04

Table 3. Mean-squared error (MSE) results from four-factor factorial analyses of variance on relative root mean-squared error (RRMSE) and mean relative bias (RBIAS) outputs from Monte Carlo simulations on variable area transect sampling using 17 fully enumerated field populations. The factors in the analysis included three investigator-controlled factors: the number of population members (r) for which to search from each start point, the sample size of start points (n), and the transect width (w). The final factor was the dataset representing each of the fully enumerated populations (dset)

		Mean squared error	
Source	df	RRMSE	RBIAS
r	5	0.3475	0.0536
n	1	0.3561	0.0192
w	3	49.5128	70.1421
dset	16	6.1063	9.7170
$r*n$	5	0.0005	<0.0001
$r*w$	15	0.4937	0.0795
$r*$dset	80	0.0122	0.0320
$n*w$	3	0.0042	0.0032
$n*$dset	16	0.0029	0.0027
$w*$dset	48	2.8603	2.4193
$r*n*w$	15	<0.0001	<0.0001
$r*n*$dset	80	<0.0001	<0.0001
$r*w*$dset	240	0.0071	0.0042
$n*w*$dset	48	0.0329	0.0034
$r*n*w*$dset	240	<0.0001	<0.0001
Replicate	816	0.0006	0.0006

for populations with random or nearly random spatial patterns. Clustered populations usually showed improved estimation with increases in transect width, although for mild clustering the bulk of the improvement occurred by going from $w = 10$ to $w = 20$.

4. DISCUSSION

This study provides investigators further, reality- based guidelines for VAT sampling. It offers a tangible comparison of how optimization of VAT sampling on naturally existing populations relates to computer-generated theoretical populations. We found that the number of population members for which to search from each start point should be at least $r = 5$. However, in contrast to computer-generated populations (Engeman and Sugihara, 1998), the results on natural populations suggest that if it is logistically feasible, $r = 7$ might offer the best tradeoff between estimation quality and effort. While we do not consider the improvement from $n = 20$ to $n = 40$ to be substantial, experimental resources probably should dictate the sample size, as long as $n \geq 20$. Of all factors studied, transect width, which had not been previously examined using computer-generated populations, had the greatest impact on estimation quality. Estimation typically improved as transect width increased. This effect also has been observed empirically in an experiment on damage sampling in row crops (Engeman and Sterner, 2002). Nevertheless, there are logistical limits to the transect width for any sampling situation whereby observations can be effectively and accurately made. We recommend that the transect be as wide as can be readily accommodated in a single pass. For example, airplane surveys

might be flown as high as possible without losing the ability to spot every population member in the transect path.

The data sets used in the simulations were from a variety of natural situations and characterized by a spectrum of patterns and densities. How well our particular data sets represent all possible populations is impossible to discern, but the results obtained from repeatedly sampling these populations provide general insights into VAT sampling beyond what we can learn from computer-generated populations. We provide our view for optimizing the tradeoff between in-field labor and the design factors minimizing RRMSE and RBIAS. We do not suggest exact values for *r*, *w*, and *n*, but ranges in which the investigator is most likely to achieve acceptable estimation properties, and which overlap with experimental resources and logistics.

REFERENCES

Bratley P, Fox B, Schrage L. 1983. *A Guide to Simulation*. Springer-Verlag: New York, U.S.A.
Cottam G. 1947. A point method for making rapid surveys of woodlands. *Bulletin of the Ecological Society of America* **28**: 60.
Engeman RM, Sterner RT. 2002. A comparison of potential labor-saving sampling methods for assessing large mammal damage in corn. *Crop Protection* **21**: 101–105.
Engeman RM, Sugihara RT. 1998. Optimization of variable area transect sampling using Monte Carlo simulation. *Ecology* **79**: 1425–1434.
Engeman RM, Sugihara RT, Pank LF, Dusenberry WE. 1994. A comparison of plotless density estimators using Monte Carlo simulation. *Ecology* **75**: 1769–1779.
Krebs CJ. 1999. *Ecological Methodology*. Addison Wesley: Reading, MA, 620.
Nielson R, Sugihara R, Boardman T, Engeman RM. 2004. Optimization of ordered distance sampling. *Environmetrics*. **15**: 119–128.
Parker KR. 1979. Density estimation by variable area transect. *Journal of Wildlife Management* **43**: 484–492.
Patil SA, Burnham KP, Kovner JL. 1979. Nonparametric estimation of plant density by the distance method. *Biometrics* **35**: 597–604.
Pielou EC. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. *Journal of Ecology* **47**: 607–613.