L. Daniel Maxim
L. Harrington
Everest Consulting Associates Inc.
Princeton Jct, NJ 08550

On Optimal Two-Stage Cluster Sampling for Aerial Surveys with Detection Errors

Numerical procedures to construct improved survey designs are presented.

Abstract: *This paper extends the statistical theory of two-stage cluster sampling to include the effects of omission and commission errors and variable primary and subcell size. Equations are derived and computational techniques presented and illustrated to enable optimal sampling plans to be developed in cases where two-stage sampling is a feasible design alternative. Depending on the population characteristics and the costs of sampling, these plans can be significantly more efficient than single-stage cluster sampling or simple random sampling alternatives, and for this reason merit careful consideration by aerial survey designers.*

Introduction

In an earlier paper, the authors (1983) discuss, in broad terms, the trade-offs required to optimize an aerial survey. This paper addresses one specific aspect of efficient survey design, the development of optimal two-stage cluster sampling plans. In particular, following Smith (1938), a power function is proposed for modeling the variance as a function of cell size. Associated formulas are developed to compute the sampling error associated with two-stage cluster sampling. Later requisite adjustments are presented to incorporate the effects of omission and commission errors. These results are useful as a planning tool for optimizing aerial surveys using two-stage cluster sampling designs. Broadly, spatial correlation between neighboring cells has the tendency of making large subcells inefficient. Measurement errors will reduce the observed spatial correlation, but the reduction may be small.

Multi-stage cluster sampling plans have, of course, been discussed extensively in the statistical and remote sensing literature: see, for example, Cochran (1977), Aldred (1971), Langley (1975), Bonner (1975), Bonner (1980). Bonner considers the problem of optimizing the primary cell size for two-stage designs assuming a fixed subcell size, while Perry and Hallum (1979) use a power function model to optimize the cell size for single-stage designs. Neither, however, develop the methodology needed to consider the more general problem of optimizing both the primary and secondary cell size simultaneously. Moreover, there has been no explicit treatment of the effects of detection errors on two-stage cluster sampling in an aerial survey context. This paper addresses these topics.

The Problem Setting

The size (and shape) of an area of land contained in one photographic frame of an aerial survey is an important design parameter. For a fixed camera format, this size is inversely related to scale, and scale is often directly related to ease of imagery interpretation. For any task, the minimum exploitable scale is often known from past experience, or it can be determined with a preliminary experiment. Thus, the maximum size of a frame associated with a particular camera format is usually fixed by known scale constraints (in some surveys, such as those which use Landsat imagery, the areal extent of the frame is fixed rather than a decision variable). See Maxim *et al.* (1981b), however, for an alternative

Photogrammetric Engineering and Remote Sensing,
Vol. 50, No. 11, November 1984, pp. 1613-1627.
0099-1112/84/5011-1613$02.25/0

perspective. For some surveys, this maximum size can be quite large, especially if the photointerpreter need only count detected items, rather than make detailed identifications (e.g., agricultural fields, wetlands, etc.). Moreover, the cost of acquiring a frame of film is sometimes substantially smaller than the cost of interpreting it. For small scales, a survey designer may choose to exploit only a few such frames. Alternatively, the frame size could be reduced—either by flying the survey at a lower altitude, by increasing the focal length of the lens, or by reducing the size of the camera format and exploiting more of these "smaller" frames. A third option is to selectively "sample" one or more portions of a larger frame. This last choice is otherwise known as two-stage cluster sampling, and requires decisions regarding the frame and subcell sizes, and the number of subcells to be sampled from each frame. The following sections first develop two-stage cluster sampling in the context of variable cell size absent measurement error and then extend this treatment to include omission and commission errors.

Two-Stage Cluster Sampling

Though the basic concept of two-stage sampling is simple, the associated notation can be intimidating. Table 1 presents a summary of this notation as used by Cochran (1977). In particular, a region of size R (hectares, square miles, etc.) is established and divided into N primary cells, or quadrats, assumed to be of equal size and shape. (The case where all primary cells are not of equal size is not addressed here.) Frequently these quadrats are square or nearly square, depending upon the camera format. However, by combining several contiguous frames together, almost any primary cell shape is possible. For simplicity, multi-frame primary cells are not considered here, although this generalization can be incorporated into the formulas that follow. Each primary cell is further partitioned into M secondary units of equal size and shape, termed subcells. The subscript i indexes the primary cells ($i = 1,N$) in the region, while the subscript j indexes the subcells ($j = 1,M$) within each primary cell.

Table 1. Some Definitions, Notation, and Illustrative Values for Numerical Examples; Two-Stage Cluster Sampling

Symbol	Formula if Applicable	Brief Description	Numerical Value in Example
R		Size of region or strata	4800
N	Integer	Number of primary cells in population.	400
M	Integer	Number of secondary cells in primary cell.	12
i	$i = 1, N$	Index of primary cells.	$i = 1,400$
j	$j = 1, M$	Index of Secondary cells/elements.	$j = 1,12$
y_{ij}	—	Number of objects in j^{th} subcell of i^{th} primary cell.	—
$\overline{Y}_i$	$\overline{Y}_i = \frac{1}{M}\,{}_j\Sigma Y_{ij}$	Mean number of objects per secondary cell in i^{th} primary cell	—
$\overline{\overline{Y}}$	$\overline{\overline{Y}} = \frac{1}{N}\,\Sigma_i\overline{Y}_i$	Overall mean number of objects per secondary cell.	20
S^2	$S^2 = \frac{\Sigma\Sigma(Y_{ij} - \overline{\overline{Y}})^2}{NM - 1}$	Overall variance among secondary cells	820
S_1^2	$S_1^2 = \frac{\Sigma(\overline{Y}_i - \overline{\overline{Y}})^2}{N - 1}$	Variance among primary cell means.	500
S_2^2	$S_2^2 = \frac{\Sigma\Sigma(Y_{ij} - \overline{\overline{Y}}_i)^2}{N(M - 1)}$	Variance among subcells within primary cells.	350
n	Integer	Number of primary cells sampled	A decision variable
m	Integer	Number of secondary cells sampled per primary cell	A decision variable
ρ	$\rho = 1 - \left(\frac{NM}{NM - 1}\right)\frac{S_2^2}{S^2}$	Correlation coefficient between subcells in the same primary cell.	0.573

The subcells contain "objects of survey interest"—fields, ponds (Gilmer *et al.*, 1980; Work and Gilmer, 1976), ant mounds (Green *et al.*, 1977), landslides (Logan 1981), dwellings, infested or diseased trees (Lillesand *et al.*, 1981), etc. The number of objects in the j^{th} subcell of the i^{th} primary cell is denoted y_{ij}. These objects are presumed to be sufficiently small and discrete relative to subcell dimensions that each object of interest can be assigned to one and only one subcell, or alternatively, that a convention is employed to make the assignment unique. It is also assumed that the imagery scale and resolution are such that errors of omission or commission can be ignored—this assumption is relaxed later in the discussion.

The average number of objects per subcell in the i^{th} primary cell is denoted $\overline{Y}_i$, while the overall average number of objects per subcell in the entire population is denoted by $\overline{\overline{Y}}$. It is assumed that the principal quantity of interest is the total number of objects in the region, given by the equation

$$T = NM\overline{\overline{Y}}. \qquad (1)$$

The overall variance among all secondary cells is denoted S^2. The variance of the subcell means between primary cells is denoted S_1^2, while S_2^2 is the variance of the subcells within a primary cell. These are defined in Table 1 explicitly. The within S_2^2 and between S_1^2 subcell variances are related by the equation

$$S^2 = \frac{M(N-1)\,S_1^2 + N(M-1)\,S_2^2}{NM-1}. \qquad (2)$$

Finally, the intracluster correlation coefficient, ρ, is defined by the relationship

$$S_2^2 = \frac{NM-1}{NM}(1-\rho)S^2.$$

As ρ approaches 1.0, the within subcell variance, S_2^2, approaches zero and the $y_{ij}s$ within a primary cell vary little when compared to the variability between primary cells. If ρ approaches zero, then the population becomes more homogeneous until ultimately there is no spatial correlation. Negative intracluster correlations are possible (in purely statistical terms) but not frequently encountered in aerial surveys.

To estimate the survey quantity T (defined in Equation 1) using two-stage cluster sampling, a sample of n of the N primary cells is randomly selected and then, from within each of these primary cells, m of the M subcells are selected. This is illustrated in Figure 1. The total is estimated from the formula

$$\hat{T} = \frac{NM}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}, \qquad (3)$$

while the variance of $\hat{T}$ is given by

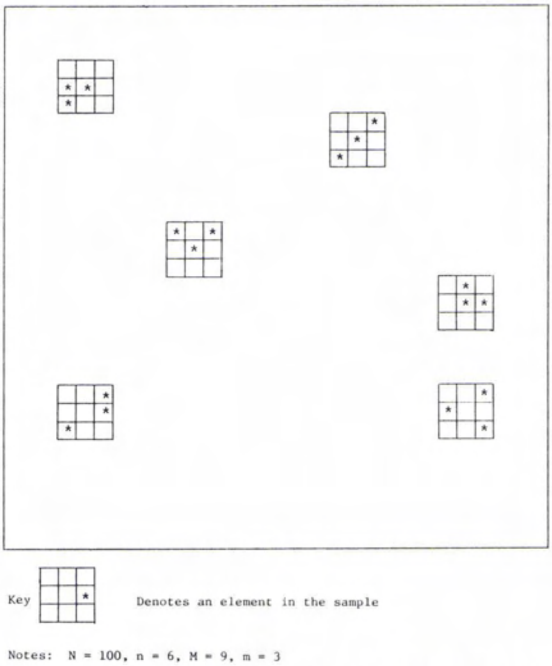

FIG. 1. Schematic representation of two-stage sampling (after Cochran, 1977).

$$\sigma_{\hat{T}}^2 = N^2M^2\left[\left(\frac{1-f_1}{n}\right)S_1^2 + \left(\frac{1-f_2}{nm}\right)S_2^2\right], \qquad (4)$$

where f_1 and f_2 are the sampling fractions, n/N and m/M. These formulas are found in Cochran (1977) together with procedures for estimating $\sigma_{\hat{T}}^2$ (e.g., estimates of S_1^2 and S_2^2 from either a pilot study or from the survey itself).

VARYING CELL SIZE

In order to model the effects of changing cell size, let Var(q) denote the variance of y_{ij} for a subcell that is q units in area. Then, in particular,

$$S^2 = \mathrm{Var}(q). \qquad (5)$$

This notation assumes the shape of the subcell has no effect on the variance. This is often approximately true, but in cases where the cell shapes become extreme, such as long narrow transects, the cell shape can no longer be disregarded. Shape effects are developed further in Harrington (1983) using linear-by-linear models. Also, see Jessen (1978, p. 106) for a discussion on shape.

The between cluster variance, S_1^2, for primary cells of size Q (units of area) is

$$S_1^2 = \frac{\mathrm{Var}(Q)}{M^2}. \qquad (6)$$

The number, N, of primary cells is simply the quotient of the size of the region divided by the primary cell size, i.e.,

$$N = R/Q. \tag{7}$$

Likewise the number of subcells in a primary cell is given by

$$M = Q/q \tag{8}$$

The values of Q and q are constrained by the requirement that N and M above be integers. Equation 4, using these definitions, is

$$\sigma_{\hat{T}}^2 = \frac{\mathrm{Var}(Q)}{n}\left(\frac{R}{Q^2}\right)\left[(R - nQ) - \frac{(Q - mq)(R - Q)}{m(Q - q)}\right] + \frac{\mathrm{Var}(q)}{nm}\left[\frac{R(Q - mq)(R - q)}{q^2(Q - q)}\right] \tag{9}$$

and the intracluster correlation coefficient is

$$\rho = \frac{q}{(Q - q)}\left[\frac{q(R - Q)\mathrm{Var}(Q)}{Q(R - q)\mathrm{Var}(q)} - 1\right]. \tag{10}$$

The next section presents one model for $\mathrm{Var}(q)$ that has been found useful.

The Power Function Model

Models that relate the variance of y_{ij} with the size of a subcell have been examined by several investigators. Jessen (1942), Mahalanobis (1944), and others (see Cochran) proposed a general law that the within variance S_2^2 is related to the size of the primary cell by the empirical formula

$$S_2^2 = \beta_0 Q^{\beta_1} \tag{11}$$

where $0 \leqslant \beta \leqslant 1$. In addition, it was assumed in this formulation that the subcell size was held constant. The above relationship has often fitted experimental data well, but as Cochran (1977) points out, S_2^2 can grow arbitrarily large as Q increases—at variance with the reasonable principle that S_2^2 should approach some upper bound.

An alternative model was suggested by H. Fairfield Smith (1938). In this model the between primary cell variance, S_1^2, is related to the size of the primary cell by the formula

$$S_1^2 = q^2 \beta_0^1 Q^{\beta_1^1}$$

where, for positive ρ, $-1 \leqslant \beta_1^1 \leqslant 0$. Combining the above with Equations 6 and 8,

$$\mathrm{Var}(Q) = \beta_0^1 Q^{(2+\beta_1^1)}, \text{ or}$$

$$\mathrm{Var}(Q) = \beta_0 Q^{\beta_1} \tag{12}$$

where, for positive ρ, $1 \leqslant \beta_1 \leqslant 2$. Note that if subcells are statistically independent, then $\mathrm{Var}(Q)$ should increase linearly with Q (i.e., $\beta_1 = 1$) while if perfectly correlated, $\mathrm{Var}(Q)$ should increase with Q^2 (i.e., $\beta_1 = 2$). Values of β_1 less than one are possible if plots within the primary cell are negatively correlated ($\rho < 0$), but such circumstances are rare. The two models, Equations 11 and 12, are logically inconsistent, but apparently both can fit spatially distributed data well, often over a substantial range of Q. For example, Perry and Hallum (1979) employed Smith's model (Equation 12) to U.S. wheat yields and obtained excellent fits for values of Q ranging five orders of magnitude from 0.27 to 25,463 acres, with values of β_1 between 1.58 and 1.82, depending upon the location.

Harrington (1983) presents an alternative model based upon the double geometric model as developed by Martin (1979). See also Whittle (1956) for a discussion of other properties of the power function model.

Smith's model is analytically convenient in that the same model can be used to describe both S_1^2 and S^2 (since these differ only in the argument, Q or q, respectively) and hence (by Equation 2) S_2^2. Thus, the specification of the two parameters β_0 and β_1 in Equation 12 enables computation of S^2, S_1^2, S_2^2, ρ, and $\sigma_{\hat{T}}^2$. This parsimony, together with the observed adequacy of fit in numerous empirical studies, motivated the choice of the Smith model in the work described here.

Estimating β_0 and β_1

To estimate β_0 and β_1 in the model (Equation 12), it is necessary to acquire sample values of $V(\cdot)$ for at least two values of Q. For example, estimates of S^2 and S_1^2 can be obtained as given in Cochran (1977). Then, by Equations 5 and 6, $\mathrm{Var}(Q)$ and $\mathrm{Var}(q)$ can be estimated. Finally, by the method of moments,

$$\hat{\beta}_1 = \frac{\mathrm{n}(\hat{\mathrm{V}}\mathrm{ar}(Q)) - \mathrm{n}(\hat{\mathrm{V}}\mathrm{ar}(q))}{\mathrm{n}(Q) - \mathrm{n}(q)} \tag{13}$$

and

$$\hat{\beta}_0 = \frac{\hat{\mathrm{V}}\mathrm{ar}(Q)}{Q^{\hat{\beta}_1}}. \tag{14}$$

If more than two sample values of $\mathrm{Var}(\cdot)$ are available, β_0 and β_1 can be estimated by weighted regression on the logarithms as proposed by Proctor (1980) and Hatheway and Williams (1958).

Two-Stage Cluster Sampling (continued)

Substituting the power function model, (Equation 12), into the variance formula, (Equation 9), gives the equation

$$\sigma_{\hat{T}}^2 = Q^{\beta_1 - 2}\left(\frac{\beta_0 R}{n}\right)\left[(R - nQ) - \frac{(Q - mq)(R - Q)}{m(Q - q)}\right] + q^{\beta_1 - 2}\left[\frac{\beta_0 R(Q - mq)(R - q)}{nm(Q - q)}\right] \tag{15}$$

Note that Equation 15 is appropriate for any size primary and secondary cell which satisfies the integer constraints given in Equations 7 and 8. Table 2 gives alternative formulas for Equation 15 that are applicable in special cases. The equation for the within variance is

$$S_2^2 = \frac{Q(R - q)}{R(Q - q)}\left[1 - \left(\frac{R - Q}{R - q}\right)\frac{Q^{\beta_1 - 2}}{q^{\beta_1 - 2}}\right]\beta_0 q^{\beta_1}.$$

If q is small in comparison to Q, the equation above reduces to the approximate relationship

$$S_2^2 \cong S^2(1 - KQ^{\beta_1 - 2}),$$

where K is a constant of proportionality. This should be compared to Equation 11 hypothesized earlier.

The intracluster correlation coefficient reduces to

$$\rho = \left(\frac{q}{Q - q}\right)\left[\frac{(R - Q)Q^{\beta_1 - 1}}{(R - q)q^{\beta_1 - 1}} - 1\right]. \tag{16}$$

For large R, this equation simplifies even further to

$$\rho \cong \frac{q^{2 - \beta_1}(Q^{\beta_1 - 1} - q^{\beta_1 - 1})}{Q - q}$$

For comparison, Nichols (1980) and Bonner (1975) model this relationship as a linear decreasing function of Q. Equation 16 above implies that the intracluster correlation is approximately proportional to $Q^{\beta_1 - 2}$, at least for small (fixed) q. This power function form was also suggested by Hansen *et al.* (1953). For $1 \leq \beta_1 \leq 2$, ρ as given by Equation 16 is nonnegative, and decreasing in Q and increasing in q as would be expected.

Enter Cost Constraints

Equation 15 can be used to evaluate the precision of the survey for any single or two-stage cluster sampling plan over the range of cell sizes and shapes for which the power function model fits the data. A survey planner's goal is often to maximize this precision (minimize $\sigma_{\hat{T}}^2$) by careful choice of the design variables n, m, q, and Q within a specified budget constraint. In particular, one common cost equation linking n and m (see Cochran, 1977) is

$$C = C_1 n + C_2 nm, \tag{17}$$

and if a budget, B, is available for the survey, the budget constraint is given by

$$C_1 n + C_2 nm \leq B,\ n,\ m \text{ integer},$$

where C is the total cost, C_1 is the cost of acquiring imagery of a primary cell, and C_2 is the cost associated with exploiting each subcell. C_1 and C_2, of course, depend upon the decision variables q and Q, but sometimes in a complex manner. These costs might be discontinuous in q and Q if it is necessary to change to different aircraft types or camera formats. Nonetheless, optimization is possible in principle and, considering the relatively small number of decision variables involved (i.e., Q, q, n, m), numerical search techniques can be employed to advantage.

Table 2. Alternative Formulas for $\sigma_{\hat{T}}^2$: Special Cases: Power Function Model

Conditions	$\sigma_{\hat{T}}^2$
None	$\frac{\beta_0 R}{nm(Q - q)}[Q^{\beta_1 - 1}((R(m - 1) + mq(n - 1) - Q(nm - 1)) + q^{\beta_1 - 2}(Q - mq)(R - q)]$
$m = 1$	$\frac{\beta_0 R}{n}[q^{\beta_1 - 2}(R - q) - Q^{\beta_1 - 1}(n - 1)]$
$m = \frac{Q}{q}$ (e.g., single stage cluster sampling)	$\frac{\beta_0 R}{n} Q^{\beta_1 - 2}(R - nQ)$
$n = N$ (e.g., $R = nQ$)	$\frac{\beta_0(R - nmq)}{m(R - nq)}\left[q^{\beta_1 - 2}\ \frac{R(R - q)}{n} - \left(\frac{R}{n}\right)^{\beta_1}(n - 1)\right]$
$m = 1$ and $n = N$	$\beta_0\left[q^{\beta_1 - 2}\ \frac{R(R - q)}{n} - \left(\frac{R}{n}\right)^{\beta_1}(n - 1)\right]$

Cochran (1977) shows that the optimal number of subcells to sample can be calculated as

$$m^* = \frac{S_2}{(S_1^2 - S_2/M)^{1/2}} \left(\frac{C_1}{C_2}\right)^{1/2}$$

$$= \left[\frac{Q\left(q^{\beta_1 - 2} - \dfrac{R - Q}{R - q} Q^{\beta_1 - 2}\right) C_1}{(Q^{\beta_1 - 1} - q^{\beta_1 - 1})\, C_2} \right]^{1/2}$$

$$\cong \left[\frac{\left(\dfrac{q}{Q}\right)^{\beta_1 - 2} - 1}{1 - \left(\dfrac{q}{Q}\right)^{\beta_1 - 1}} \right]^{1/2} \left(\frac{C_1}{C_2}\right)^{1/2} \quad (18)$$

This value should be rounded using the rule: if m^* lies between the integer m and $m + 1$, choose $m^* = m + 1$ if $m^* > m(m + 1)$; otherwise rounded down. Note, however, if the budget is sufficient to enable all the primary cells to be sampled, Equation 18 may no longer be valid. In this case, m^* must be established by numerical search.

As a practical matter, a survey planner can proceed by developing several potential survey plans that satisfy the budget and other constraints, and from these feasible choices, select the best one, as calculated using Equation 15. This procedure is illustrated in the following example.

An Example

Table 1 also contains numerical values (estimated from a subsample) from an aerial survey designed to estimate the number of agricultural fields in a survey region. For this particular survey, the entire region had been photographed previously in response to other requirements—hence, the operative concern was to design an exploitation plan rather than a collection and exploitation plan. Each frame was fixed at 12 area units, (i.e., $Q = 12$). For this survey, time studies established that it required 1/2 hour of effort to register and catalog a frame, and approximately 1.0 hour per unit area to exploit this film. Thus, the cost parameters, assuming subcells of size q are exploited, are

$$C_1 = 1/2$$

and

$$C_2 = q.$$

Finally, an exploitation budget of 500 hours of effort was established.

From the values given in Table 1,

$$\mathrm{Var}(1) = S^2 = 820,$$

while

$$\mathrm{Var}(12) = M^2 S_1^2 = 72{,}000.$$

Parameter values for the power function, from Equation 13 and 14, are, therefore, estimated as

$$\beta_1 = 1.801,$$

and

$$\beta_0 = 820.$$

Table 3 presents a variety of sampling plans, each satisfying the 500 hour budget constraint, for several feasible cell sizes. Other considerations required that the subcell size, q, be at least 1/4 area units. Sampling plans for which the primary cell size, Q, is less than 12 area units are not practical alternatives. They are included in Table 3 simply to give a more comprehensive assessment. Also, in order to observe the effects of different values of β_1, the standard deviation, $\sigma_{\hat{T}}$, has been calculated for three values of β_1, equal to 1.801, 1.60, and 1.40.

The most important point to note from Table 3 is that the subcell size should be 1/4 area unit, the minimum possible. The best plan (plan 10, or plan 11 if $\beta_1 = 1.4$) specifies a primary cell size of 12 area units, and from each primary cell, three 1/4 square subcells are exploited. Note also that plan 3 (which exploits the entire frame) has a standard error that is higher by a factor of five compared to that of the optimal two-stage plan (plan 10). Two-stage plans that exploit only one subcell per frame, $m = 1$, are very similar to single-stage plans having the same cell size (e.g., plan 1 versus plan 4). They have slightly smaller standard deviations, however, because two-stage plans exploit knowledge of S_2^2. If S_2^2 is to be estimated from the survey itself, then, of course, m must be at least 2. Note also that plans 8 and 9 exhaust the available primary cells before exhausting the budgeted hours, and, as a consequence, their standard deviations are larger than might have been anticipated. When this occurs, various anomolies are possible, as is shown later.

Smith (1938) and later again Perry and Hallum (1979) showed that the optimal cell size for one-stage cluster sampling is given approximately by the equation

$$Q^* = \frac{(2 - \beta_1)\, C_1}{C_2\, (\beta_1 - 1)}. \quad (19)$$

For this example, Q^* would be approximately 1/8 area unit when $\beta_1 = 1.80$, 1/3 area unit when $\beta_1 = 1.6$, and 3/4 area unit when $\beta_1 = 1.4$. A similar analysis for two-stage cluster sampling is not possible, due in part to the simplistic specification of the cost function given earlier. Figure 2 shows how the standard deviation varies as a function of subcell size,

TABLE 3. ONE AND TWO STAGE SAMPLING PLANS

Type of Sampling	Plan	Primary Cell Size Q	Subcell Size q	M	N	Subcells Sampled m	Primary Cells Sampled n	Actual Cost (Hours)	$b\,\sigma_{\hat{T}}$ β_1 = 1.801	β_1 = 1.6	β_1 = 1.4
Single-stage cluster sampling	1	1/4	1/4	1	19200	1	666	499.5	6007	6905	7932
	2	1	1	1	4800	1	333	499.5	7266	7566	7266
	3	12	12	1	400	1	40	500.0	16101	12543	9783
Two-stage cluster sampling	4	1	1/4	4	4800	1	666	499.5	5783	6742	7825
	5					2	500	500.0	6152	6572	7085
	6					3	400	500.0	6706	6869	7078
	7					4	333	499.5	7266	7266	7266
	8	12	1/4	48	400	1	400	300.0	5788	8049	9894
	9					2	400	400.0	4049	5631	6921
	10					3	400	500.0	3270	4547	5590
	11					4	333	499.5	3900	4662	5446
	12	12	1	12	400	1	333	499.5	5292	6267	6791
	13					2	200	500.0	6754	6675	6627
	14					3	142	497.0	8157	7392	6888
	15					4	11	499.5	9333	8064	7189

assuming that the primary cell size is fixed at 12 and that m is chosen optimally.

- In broad terms, as subcell size is reduced, so too is $\sigma_{\hat{T}}$.
- On a more detailed level, however, it is clear that the relation between $\sigma_{\hat{T}}$ and q is not monotone-decreasing. Rather, the anomoly identified earlier (in cases where the primary cells exhausts the region) creates local optima.
- As q decreases, the optimal value of m, m^*, increases. In this example, m^* is 1 for values of q between 6 and 0.60, and then becomes 2 or more for q smaller than 0.60. This is related, in part, to the fact that the intracluster correlation decreases with q.

Because there are no fixed costs associated with subcells, the variance, $\sigma^2_{\hat{T}}$, will be minimized by making subcells arbitrarily small. This shows that the analysis given in Figure 2 should not be extrapolated too far without careful consideration of the costs and how they relate to cell size. And, of course, additional constraints on cell size are often relevant.

Table 4 shows the effect of budget for fixed primary and subcell size. Larger budgets permit more acquisition and exploitation, so that the total, T, will be estimated with greater precision. Ultimately, a budget of 5,000 hours is sufficient to do a complete census and the sampling error is reduced to zero. For comparison, the single-stage plan using the entire frame is also presented. It is dominated by the two-stage plans for all budgets considered. This result should not be interpreted to imply that two-stage plans will always be superior. The cost structure of a survey is important to the choice between single and two-stage plans. Given specific alternative designs, however, Equation 15 enables alternative plans to be evaluated.

THE EFFECTS OF DETECTION ERROR

All these results assume that all "items of interest" are detected with certainty. It is often the case, however, that detection errors are present and some items are missed, and so the number actually identified, x_{ij}, will be less than or equal to the true number of items in a subcell, y_{ij}. (Commission errors will be addressed later.) It is further assumed that these items have a common detection probability, P_d. If P_d does vary, it is assumed that strata can be defined by the geography of the primary cell or by some variable such as field size (see Maxim and Harrington, 1982) such that P_d can be considered essentially constant within each stratum. Detections are assumed to be independent events. In this case, the expected number of items detected in the ij^{th} subcell is $P_d y_{ij}$, and an unbiased estimate of the total number of objects in the population is

$$\hat{T}' = \frac{NM\bar{\bar{X}}}{P_d},$$

where

$$\bar{\bar{X}} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}/\text{nm}.$$

It is shown in the appendix that the variance of $\hat{T}'$ is given by the equation

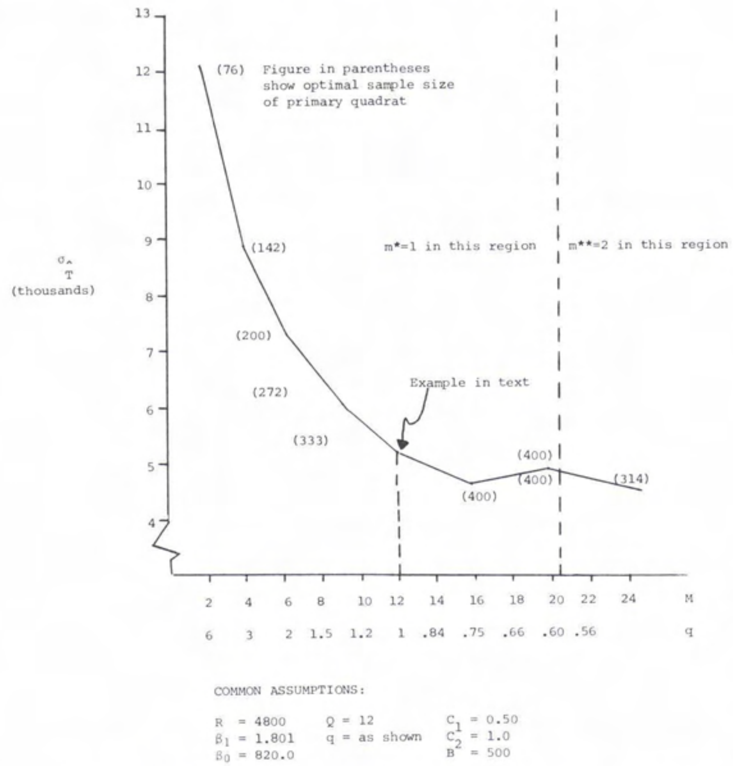

FIG. 2. Standard deviation of sampling plans for varying subcell size, assuming primary cell size is fixed, $Q = 12$, and m is chosen optimally.

$$\sigma^2_{\hat{T}'} = N^2M^2 \left[\left(\frac{1 - f_1}{n}\right) S_1^2 + \left(\frac{1 - f_2}{nm}\right) S_2^2 + \frac{\bar{\bar{Y}}(1 - P_d)}{nmP_d} \right] \tag{20}$$

This expression differs from that obtained earlier only by the addition of the last term. Because $\bar{\bar{Y}}$ is proportional to the subcell size, q, this term can be written as

$$\frac{\mu(1)\ R^2(1 - P_d)}{q\ nm\ P_d},$$

where $\mu(1)$ is the expected number of items when the subcell size is 1. This term is inversely related to subcell size, q, and so can be made smaller by specifying larger subcells. It is independent of the primary cell size, Q, unless a change in Q implies a change in scale (and thus possibly P_d). It is also independent of the primary and secondary finite population correction factors $(1 - f_1)$ and $(1 - f_2)$, and so this additional term will represent a larger percentage of the total variance with increasing sample sizes. Nevertheless, the effects of detection error on $\sigma^2_{\hat{T}'}$ are often small, at least for the cases considered here. Figure 3 plots this variance for various P_d and m. As this shows, the variance is not increased substantially until P_d is less than 0.30, although the assumed value of $\mu(1)$ (=20 here) is equally important.

ANOTHER NUMERICAL EXAMPLE

Table 5 provides another useful numerical example. Specific inputs, e.g., P_d, $\mu(1)$, S^2, S_2^2, etc. are shown at the bottom of this table, but broadly this example assumes that Q is fixed at 12 and q is fixed at 1 (similar to plans 12 and following in Table 3) and that the detection probability is 0.3. Optimization here reduces simply to the specification of m^*, and is readily accomplished numerically. Note that once a value for $m(m = 1, M)$ is assumed, n is immediately determined as the largest integer beneath N that satisfies the budget constraint given in Equation 17. The optimal value of m, m^*, is one in this example. The standard error of the estimated total corresponding to m^*, $\sigma_{\hat{T}}^*$ is equal to 5,588 (a proportional error of ±5.8 percent), somewhat higher than would have been the case with perfect

TABLE 4. EFFECT OF BUDGET ON OPTIMAL SINGLE-STAGE AND TWO-STAGE CLUSTER SAMPLING PLANS

B Assumed Budget (hours)	Optimal Two-Stage Plan			Single-Stage Plan		
	m^*	n^*	$\sigma_{\hat{T}}$	m	n	$\sigma_{\hat{T}}$
500	1	333	5292	12	40	16,101
600	1	400	4300	12	48	14,534
700	1	400	4300	12	56	13,302
800	2	320	4208	12	64	12,298
900	2	360	3541	12	72	11,455
1000	2	400	2899	12	80	10,734
1500	3	400	2246	12	120	8,198
2000	4	400	1833	12	160	6,573
2420	5	400	1534	12	193	5,558
3000	7	400	1096	12	240	4,382
3500	8	400	917	12	280	3,514
4000	9	400	749	12	320	2,684
4500	10	400	580	12	360	1,789
4750	11	400	391	12	380	1,231
5000	12	400	0	12	400	0

Common Assumptions:
$R = 4800$ $q = 1$ $C_1 = 0.5$
$\beta_0 = 820$ $Q = 12$ $C_2 = 1.0$
$\beta_1 = 1.801$ $B =$ as shown

detection ($\sigma_T = 5{,}292$) but appreciably lower than that plan corresponding to simple single-stage sampling, i.e., where $m = M$ or 12. Note also that σ_T is quite sensitive to the choice of m. While this sensitivity is a function of many factors, and the problem can have a "flat optimum" (often noted with sampling problems), it serves as a reminder that "flat optima" do not always result. Among other factors, the sensitivity of σ_T to m is a function of the ratio, C_1/C_2. In particular, when C_1 is very small relative to C_1 (as happens, for example, with Landsat imagery where the user's acquisition cost is small relative to other costs for image processing and exploitation), m tends to be small relative to M, the standard error increases quickly with m, and the two-stage plan is considerably more efficient than the single-stage alternative.

SENSITIVITY ANALYSIS

Table 6 shows the (one-at-a-time) sensitivity of m^* to each of the main inputs to the problem, i.e., μ, P_d, S_1^2, S_2^2, C_1, and C_2 if the budget is fixed at 500. As can be seen, exact *a priori* knowledge of these quantities is *not* required for determination of m^*. For example, S_1^2 was assumed equal to 500 in the base case. Other factors held constant m^* is one for any value of S_1^2 greater than 129—a wide tolerance of error (-74.2 percent to infinity). In this case at least, preliminary estimates of the relevant inputs are likely to be sufficiently accurate for determination of m^*. Subsequent data analysis can refine these inputs for final survey design computation of confidence intervals and survey estimates, etc.

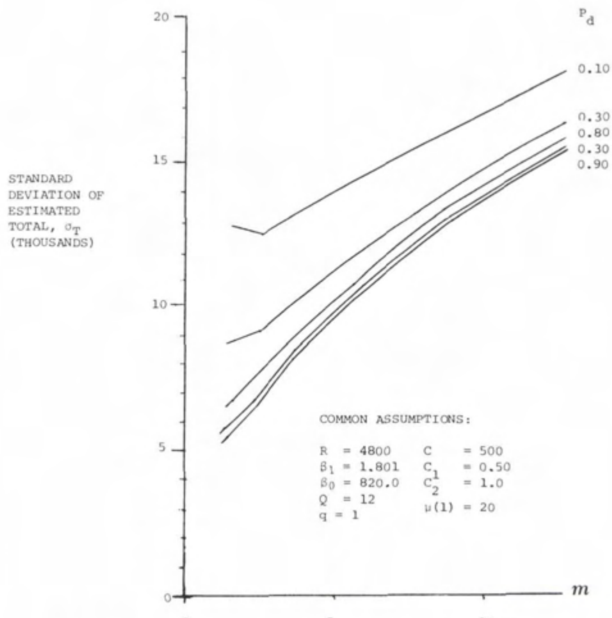

FIG. 3. The effect of detection probability on standard error for fixed primary and secondary cell dimensions and for fixed cost.

TABLE 5. ILLUSTRATIVE OPTIMIZATION OF SUBCELL SAMPLING IN TWO-STAGE CLUSTER SAMPLING WITH FIXED BUDGET

Number of Subcells Sampled Per Primary Cell m	Number of Primary Cells Sampled n	Standard Error of Estimated Total σ_T	Standard Error As % of Estimate	Efficiency Relative; To Optimal Plan (%) $\frac{100\ \sigma_T^2(m^*)}{\sigma_T^2(m)}$
1	333	5,588	5.8	100
2	200	6,949	7.2	64.7
3	142	8,310	8.7	45.2
4	111	9,461	9.9	34.9
5	90	10,585	11.0	27.9
6	76	11,575	12.1	23.3
7	66	12,464	13.0	20.1
8	58	13,339	13.9	17.5
9	52	14,119	14.7	15.7
10	47	14,881	15.5	14.1
11	43	15,582	16.2	12.9
12	40	16,169	16.8	11.9

Common Assumptions:
$N = 400$, $\mu(1) = 20$, $S_1^2 = 500$, $S^2 = 820$
$M = 12$, $P_d = 0.3$, $S_2^2 = 350$
$C_1 = 0.5$, $B = 500$
$C_2 = q$, $q = 1.0$, $Q = 12.0$

SPATIAL CORRELATION WITH DETECTION ERRORS

The variance of the x_{ij} can be shown to be

$$\mathrm{Var}_x(q) = \bar{\bar{Y}}\, P_d(1 - P_d) + P_d^2 S^2. \tag{21}$$

Thus, the intracluster correlation associated with the x_{ij} (as distinct from the y_{ij}) is, by Equation 16,

$$\rho_x = \left(\frac{q}{Q - q}\right)\left[\left(\frac{R - Q}{R - q}\right)\left(\frac{\mu(1)(1 - P_d) + P_d\beta_0 Q^{\beta_1 - 1}}{\mu(1)(1 - P_d) + P_d\beta_0 q^{\beta_1 - 1}}\right) - 1\right]$$
$$\cong \frac{\beta_0 q}{Q - q}\left[\frac{Q^{\beta_1 - 1} - q^{\beta_1 - 1}}{\dfrac{\mu(1)\,(1 - P_d)}{P_d} + \beta_0 q^{\beta_1 - 1}}\right] \tag{22}$$

Figure 4 plots values of ρ_x for various P_d and Q. The effect of detection errors is to reduce the *observed* correlation. But, at least for this example, detection errors again need to be very small before this effect is large.

TABLE 6. ONE-AT-A-TIME SENSITIVITY ANALYSIS FOR EXAMPLE IN TABLE 5 WHERE $m^* = 1$ AND $B = 500$

Parameter	Base Case Value	Value Necessary to Increase m^* to 2	% Difference from Base Case
S_1^2	500	129	−74.2
S_2^2	350	1457	316.3
P_d	0.3	0.01299	−95.7
$\mu(1)$	20	652	3,160.0
C_1	0.5	2.3785	375.7
C_2	1.0	.559	−44.1

For some data sets (see Bonner (1975), for example), the power function model does not appear to be suitable because the observed intracluster correlation decreases less rapidly with Q than is predicted from Equation 16. The presence of detection errors (or other measurement errors) could account for this at least in part. Figure 5 plots the apparent (observed) intracluster correlation, Equation 22, for a variety of cases where β_1 and P_d have been chosen jointly so that $\rho = 0.5$ when $Q = 2$. This shows that the intracluster correlation given by Equation 22 can appear to persist over large areas under suitable values of the parameters.

Appendix B provides an extension of the proof given by Cochran (1977) to show that if detection errors are present, then the variance is minimized when the number of subcells sampled per primary cell is

$$m^{**} = \frac{S_2}{(S_1^2 - S_2^2/M)^{1/2}}\left(\frac{C_1}{C_2}\right)^{1/2}\left[1 + \frac{\bar{\bar{Y}}(1 - P_d)}{P_d\, S_2^2}\right]^{1/2}. \tag{23}$$

This equation reduces to m^* obtained earlier when $P_d = 1$; otherwise, m^{**} above is larger than m^*. For the example considered earlier, this adjustment,

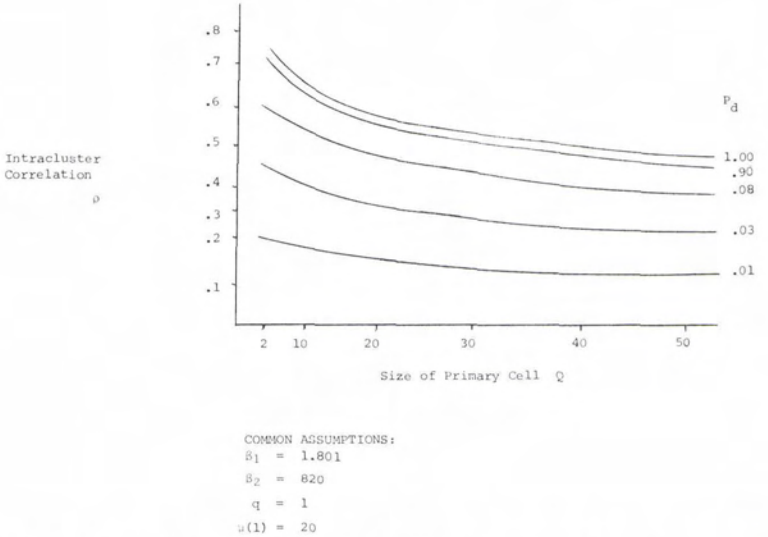

FIG. 4. Effect of the detection probability on the observed intracluster correlation.

$$\left[1 + \frac{\overline{\overline{Y}}(1 - P_d)}{P_d S_2^2}\right]^{1/2}$$

is not large unless P_d is less than about 0.30.

CLASSIFICATION/IDENTIFICATION ERRORS

In many surveys, commission errors as well as omission errors need to be considered. There are several approaches for modeling commission errors. As one example, commission errors might be proportional to the number of "items of interest" in a cell because objects so misclassified are collocated with the objects of interest. In this case, commission errors might be modeled as independent Poisson events with a mean and variance θy_{ij}. That is,

$$P(k) = \frac{(\theta y_{ij})^k \, e^{-\theta y_{ij}}}{k\,!} \quad k = 0, 1, 2 \ldots$$

Given a suitable ground truth data set, this Poisson assumption can be assessed directly. An unbiased estimate of the total would be

$$\hat{T}'' = \frac{NM\overline{\overline{X}}}{(P_d + \theta)},$$

and the variance can be shown (by a procedure similar to that given in Appendix A) to be

$$\sigma^2_{\hat{T}''} = N^2M^2\left(\frac{(1 - f_1)\,S_1^2}{n} + \frac{(1 - f_2)\,S_2^2}{nm} + \frac{(P_d(1 - P_d) + \theta)\,\overline{\overline{Y}}}{nm\,(P_d + \theta)^2}\right)$$

As a second example, commission errors might simply be a function of the area searched. In this case, the number of such errors would be independent of y_{ij} and could be modeled as Poisson with mean and variance θ for all subcells within a stratum. Given this assumption, an unbiased estimate of the total is

$$\hat{T}'' = \frac{NM}{P_d}\,(\overline{\overline{X}} - \theta),$$

and variance of this estimate can be shown to be

$$\sigma^2_{\hat{T}''} = N^2M^2\left(\frac{(1 - f_1)\,S_1^2}{n} + \frac{(1 - f_2)\,S_2^2}{nm} + \frac{1}{nm\,P_d^2}\,\left(P_d(1 - P_d)\,\overline{\overline{Y}} + \theta\right)\right).$$

Of course, distributions other than the Poisson may be appropriate. If so, formulas can be derived similar to Appendix A. See Maxim *et al.* (1981c) for yet other models applicable to commission errors.

If the detection probability is to be estimated (see Maxim and Harrington, 1981a), then the uncertainty, Var (P_d), of this estimate needs to be considered. (It also biases the estimate, $\hat{T}'$, although this bias is usually small.) In particular, an estimate of the variance of $\hat{T}'$ is, using a Taylor series approximation,

$$\mathrm{Var}\left(\frac{NM\overline{\overline{X}}}{\hat{P}_d}\right) \cong \sigma^2_{\hat{T}'} + T'^2\,\frac{\mathrm{Var}(\hat{P}_d)}{\hat{P}_d^2} \qquad (24)$$

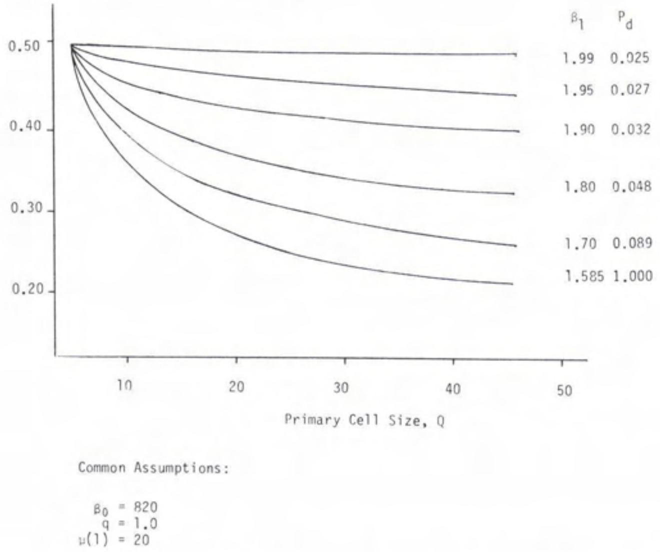

FIG. 5. The persistence of the observed intracluster correlation when detection errors exist.

The first term is calculated as given by Equation 20, substituting in sample values where necessary. The second term in Equation 24 can be significant and should be included if the size of the ground truth sample used to estimate $\hat{P}_d$ is small.

Extensions

An aerial survey often has a variety of errors other than those associated with detection and identification. For example, the accuracy of linear or area measurements will be a function of the mensuration equipment used, uncertainty in aircraft altitude above ground level, optical distortions, the skill of the photointerpreters, etc. In other cases, regression models are used because the variable of interest cannot be determined directly (e.g., the wood volume of a tree (see Bonner, 1975)). In these instances, the variable of x_{ij} will be "contaminated" with error (measurement or otherwise), e_{ij}, and the observed values, x_{ij}, can be represented as

$$x_{ij} = y_{ij} + e_{ij}.$$

If e_{ij} has an expected value of zero and is distributed independent of the other e_{ij}s, then it can be similar to Appendix A that the variance of

$$\hat{T}_x = NM \frac{\Sigma\, \Sigma x_{ij}}{nm}$$

is given by

$$\sigma^2_{\hat{T}_x} = N^2M^2 \left[\left(\frac{1 - f_1}{n}\right) S_1^2 + \left(\frac{1 - f_2}{nm}\right) S_2^2 + \frac{1}{nm}\,\sigma_e^2 \right] \tag{25}$$

where σ_e^2 is the variance of e_{ij}. Because the e_{ij} are assumed independent, one model for σ_e^2, for varying subcell size, would be

$$\sigma_e^2 = \sigma_e^2(1)q,$$

where $\sigma_e^2(1)$ is the measurement error variance for a subcell of size 1. Thus, the term given in Equation 25 for this added error reduces to

$$\frac{R^2\ \sigma^2(1)}{q\ \ nm}$$

The functional form of Equation 25 is similar to those related to detection errors. Thus, this more general case is a simple extension of these earlier results.

Summary

The variance of two-stage cluster sampling plans is calculated for varying primary and subcell sizes. The equation for the intracluster correlation is also derived assuming a power function model, Equation 12. Adjustments for omission and commission errors are then calculated. Generally, when data exhibit large intracluster correlations, efficient sampling plans will utilize small subcells. When this intracluster correlation is small, then the total area exploited is more important than how it is partitioned into primary and secondary subcells. However, as Equation 16 demonstrates, the intracluster correlation varies with cell and subcell size, and so the optimization of a survey design can be complicated. The effect of detection errors is to reduce the apparent intracluster correlation, and so make larger subcells more attractive. Also, the correlation can appear to persist over longer distances. With realistic cost components, these results could be incorporated into a survey planning models similar to those proposed by Aldred (1971), Titus (1979), Wensel and Eriksson (1980), and others.

Acknowledgment

The authors would like to acknowledge the helpful discussion of our colleagues Dr. Nancy David and Mary Kennedy. Ms. Lauren Maxim kindly wrote the necessary computer codes and checked the computations throughout. The referees comments were most appreciated and helped to improve the clarity and focus of this paper.

References

Aldred, A. H., 1971. *A Planning Model for Forest Inventories Based on Combined Photo and Ground Sampling Methods*, Ph.D. Thesis, University of California, Berkeley, 192 pp. University Microfilms No. 73-25339.

Bonner, G. M., 1975. A Test of Cluster Sampling in Forest Inventories, *Canadian Journal of Forest Resources*, Volume 5, pp. 269-272.

Bonner, W. G., Jr., and J. Morgart, 1980. Landsat: A Sampling Frame for Arid Land Inventories, *Proceedings of Arid Land Resource Inventories Workshop: Developing Cost-efficient Methods*. USDA Forest Service General Technical Report No. WO-28, pp. 230-239.

Cochran, W. G., 1977. *Sampling Techniques*, Third Edition, John Wiley and Sons, New York.

Gilmer, D. S., E. A. Work, J. E. Colwell, and D. L. Rebel, 1980. Enumeration of Prairie Wetlands with LANDSAT and Aircraft Data, *Photogrammetric Engineering and Remote Sensing*, Vol. 46, pp. 631-634.

Green, L. R., J. R. Olson, W. G. Hart, and M. R. Davis, 1977. Aerial Photographic Detection of Imported Fire Ant Mounds, *Photogrammetric Engineering and Remote Sensing*, Vol. 48, pp. 1581-1592.

Hanson, M. H., W. H. Hurwitz, and W. C. Madow, 1953. *Sample Survey Methods and Theory*, Vol. 1, John Wiley and Sons, New York.

Harrington, L., 1983. *On Optimizing Agricultural Surveys using "Linear-by-Linear" and Power Function Models*, Everest Consulting Associates, Inc., Technical Report.

Hatheway, W. H., and E. J. Williams, 1958. Efficient estimation of the relationship between plot size and the variability of crop yields, *Biometrics*, Vol. 14, pp. 207-222.

Jessen, R. J., 1942. *Statistical Investigation of a Sample Survey for Obtaining Farm Facts*. Iowa Agr. Exp. Sta Res. Bull. 304.

Jessen, R. J., 1978. *Statistical Survey Techniques*, John Wiley and Sons, New York.

Langley, P. G., 1975. *Multistage Variable Probability Sampling: Theory and Use in Estimating Timber Resources From Space and Aircraft Photography*. Ph.D. Thesis, University of California, Berkeley, 123 pp. University Microfilms No. 75-22538.

Lillesand, T. M., D. E. Meisner, D. W. French, and W. L. Johnson, 1981. Evaluation of Digital Photographic Enhancement for Dutch Elm Disease Detection, *Photogrammetric Engineering and Remote Sensing*, Vol. 48, pp. 1581-1592.

Logan, T. L., 1981. A Data Base Approach for Prediction of Deforestation-Induced Mass Wasting Events, *Technical Papers, The American Society of Photogrammetry*, 47th Annual Meeting, Washington, D.C., pp. 197-211.

Mahalanobos, P. C., 1944. On Large-Scale Sample Surveys, *Phil Trans Roy Soc. London*, pp. 231, 329-451.

Martin, R. J., 1979. A Sub-cell of Lattice Processes Applied to a Problem in Planned Sampling, *Biometrika*, Vol. 66, pp. 209-217.

Maxim, L. D., L. Harrington, and M. Kennedy, 1981a. A Capture-Recapture Approach for Estimation of Detection Probabilities in Aerial Surveys, *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 779-788.

Maxim, L. D., H. D. Weed, L. Harrington, and M. Kennedy, 1981b. Intensity Versus Extent of Coverage, *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 789-797.

Maxim, L. D., L. Harrington, and M. Kennedy, 1981c. Alternative Scale-Up Estimators for Aerial Surveys Where Both Detection and Classification Errors Exist, *Photogrammetric Engineering and Remote Sensing*, Vol. 47, pp. 1227-1239.

Maxim, L. D., and L. Harrington, 1982. Scale-Up Estimators for Aerial Surveys With Size-Dependent Detection, *Photogrammetric Engineering and Remote Sensing*, Vol. 48, pp. 1271-1287.

———, 1983. How to Design an Aerial Survey: A Systems-Analytic Perspective, *Photogrammetric Engineering and Remote Sensing*, Vol. 49, pp. 1425-1435.

Nichols, J. D., 1980. Intra-cluster Correlation in Sample Design. *Proceedings of Arid Land Resource Inventories Workshop: Developing Cost-Efficient Methods*, USDA Forest Service General Technical Report No. WO-28, pp. 142-147.

Perry, C. R., and R. S. Chikata, 1980. Stratum Variance Estimation for Sample Allocation in Crop Surveys, *American Statistical Association 1980 Proceedings of the Section on Survey Research Methods*, pp. 782-787.

Perry, C. R., and C. Hallum, 1979. Sampling Unit Size Considerations in Large Area Crop Inventorying Using Satellite-Based Data, *American Statistical Association 1979 Proceedings of the Section on Survey Methods*, pp. 430-433.

Proctor, C. H., 1980. Estimating Smith's b from Sample Data, *American Statistical Association 1980 Proceedings of the Section on Survey Research Methods*, pp. 761-765.

Smith, H. F., 1938. An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops, *Journal of Agricultural Science*, Vol. 28, pp. 1-23.

Titus, S. J., 1979. *Design of Multivariate Surveys for*

Sampling Wildland Ecosystems, Ph.D. Thesis, University of California, Berkeley. 123 pp. University Microfilms No. 80-00546.

Wensel, L. C. and M. Eriksson, 1980. Survey Planning Model: Application for Arid Land Resource Inventories. *Proceedings of Arid Lands Resource Inventories Workshop: Developing Cost-efficient Methods*, USDA Forest Service General Technical Report No. WO-28, pp. 136-141.

Whittle, P., 1956. On the Variation of Yield Variance with Plot Size, *Biometrika*, Vol. 43, pp. 339-343.

Work, E. H., and D. S. Gilmer, 1976. Utilization of Satellite Data for Inventorying Prairie Ponds and Lakes, *Photogrammetric Engineering and Remote Sensing*, Vol. 42, pp. 685-694.

(Received 29 October 1982; revised and accepted 3 June 1984)

Appendix A. The Variance of $\hat{T}'$

This appendix shows the variance of

$$\hat{T}' = \frac{NM}{P_d nm} \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} \tag{A1}$$

is

$$\sigma^2_{\hat{T}'} = N^2M^2 \left[\left(\frac{1-f_1}{n}\right) S_1^2 + \left(\frac{1-f_2}{nm}\right) S_2^2 + \frac{\overline{\overline{Y}}\,(1-P_d)}{nmP_d} \right] \tag{A2}$$

Given the true number of objects in the ij^{th} subcell, the number detected, x_{ij}, will be binomially distributed, assuming such detections are independent of one another. Thus

$$\mathrm{E}(x_{ij}|y_{ij}) = P_d y_{ij}$$

and

$$\mathrm{Var}(x_{ij}|y_{ij}) = y_{ij} P_d (1 - P_d)$$

From Cochran (e.g., 10.2)

$$\mathrm{Var}\left(\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}\right) = V_2\,(E_2\,(\Sigma\,\Sigma\, x_{ij})) + E_1\,(V_2\,(\Sigma\,\Sigma\, x_{ij})) \tag{A3}$$

where $E_2(\cdot)$ and $V_2(\cdot)$ are the mean and variance *given* a particular set of primary and secondary cells, and $E_1(\cdot)$ and $V_1(\cdot)$ denote the mean and variance over all possible selections of primary and secondary cells. Now,

$$E_2\left(\sum_i \sum_j x_{ij}\right) = P_d\,\Sigma\,\Sigma\, y_{ij}$$

$$V_2\left(\sum_i \sum_j x_{ij}\right) = P_d(1 - P_d) \sum_i \sum_j y_{ij}$$

From Equation (4) in the text,

$$V_1\left(P_d \sum_i \sum_j y_{ij}\right) = P_d^2\, n^2 m^2 \left[\left(\frac{1-f_1}{n}\right) S_1^2 + \left(\frac{1-f_2}{nm}\right) S_2^2 \right] \tag{A4}$$

Also

$$E_1\,(P_d(1 - P_d)\,\Sigma\,\Sigma\, y_{ij}) = nm\, P_d(1 - P_d)\,\overline{\overline{Y}}. \tag{A5}$$

Adding Equations A4 and A5 to give Equation A3, the result, Equation A2, follows from Equation A1.

Appendix B. Optimal Subcell Sample Size, m**, Assuming Detection Errors

Equation 23 in the main text can be proven similar to that given by Cochran, that assumes no detection errors. That is, letting

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}}{nm},$$

the variance of $\bar{\bar{X}}$ can be written as

$$\mathrm{Var}(\bar{\bar{X}}) = P_d^2 \left[\frac{1}{n} (S_1^2 - S_2^2/M) + \frac{1}{nm} \left(\frac{\bar{\bar{Y}}(1 - P_d)}{P_d} + S_2^2 \right) - S_1^2/N \right].$$

Minimizing Var $(\bar{\bar{X}})$ for fixed cost $C = nC_1 + nmC_2$ is equivalent to minimizing the product

$$C(\mathrm{Var}(\bar{\bar{X}}) + S_1^2/N) = \frac{P_d^2}{n} \left[(S_1^2 - S_2^2/M) + \frac{1}{m} \left(\frac{\bar{\bar{Y}}(1 - P_d)}{P_d} + S_2^2 \right) \right]$$
$$(nC_1 + nmC_2)$$

By the Cauchy-Schwarz inequality, the above expression is minimized if

$$\frac{C_2 m}{C_1} = \frac{\dfrac{\bar{\bar{Y}}(1 - P_d)}{P_d} + S_2^2}{m(S_1^2 - S_2^2/M)}.$$

Solving for m gives m^{**}.

CALL FOR PAPERS

Envirosat "85"

Sheraton Washington Hotel, Washington, D.C.
10-13 September 1985

The objective of NOAA's 1985 Environmental Satellite Conference, whose theme is "25 Years of Service," is to increase user awareness of current and future applications of remotely sensed environmental data. Topics will include hardware and software, sensor applications, public and private sector interests, and future trends.

Those wishing to present a paper should submit an abstract of no less than 500 words by *15 December 1984* to

Program Committee
2nd Envirosat Conf
NOAA/NESDIS, E/ER2
FB#4, Mail Stop D
Washington, DC 20233