Towards a Plot Size for Canada's National Forest Inventory

Steen Magnussen, P. Boudewyn, and M. Gillis

Abstract.—A proposed national forest inventory for Canada is to report on the state and trends of resource attributes gathered mainly from aerial photos of sample plots located on a national grid. A pilot project in New Brunswick indicates it takes about 2,800 square 400-ha plots (10 percent inventoried) to achieve a relative standard error of 10 percent or less on 14 out of 17 estimates of forest strata (cover type) proportions; it takes about 3,300 plots of similar size to achieve the same precision target on 68 of 85 strata type x age proportions. Minimum cost solutions for fixed targets of relative precision and power include plot sizes from 71 to 192 ha and sampling intensities from 0.2 to 3.5 percent. Multivariate solutions exploiting the covariance in the data called for less sampling than univariate solution.

Plot size and number of sample plots are pivotal design parameters of any forest inventory. Procedures for optimizing these design parameters are well documented for single quantitative characteristics but less so for multilevel categorical variables such as forest strata (Rochon 1989). The effect of plot size on sample variances depends on the extent and spatial distribution of the sampled attribute, and is perhaps best described in terms of the spatial correlation function (Mahalanobis 1944) or the intraplot correlation δ. Briefly, the variance among plots equals the average within-plot covariance. The statistical efficiency of a plot size is governed by the decay rate of δ as plot size increases.

Canada is contemplating a new design for the national forest inventory. Data on area and extent of forest cover types (based on species composition and structure) are pivotal to the success of the inventory. Cover type data will come from interpretation of aerial photos (scale around 1:15000) of inventory plots located on a national grid. Photo interpretation is suited for the estimation of forest strata (Acka 1993) and can accommodate substantial sampling efforts in a cost-effective way. Ideally the photos would be taken and interpreted in a manner compatible to the reporting dates of the national inventory. Issues such as the size and number of plots needed to achieve a precision target on forest cover estimates have not been settled yet. To firm considerations, three pilot projects mimicking various design alternatives were established in New Brunswick, Ontario, and British Columbia to investigate the impact of plot size and number of plots on precision of estimates. In this study, we focus on the precision of estimates of areas of forest strata and strata x age classes (we henceforth use the more generic word strata in place of cover types).

Research Scientist, Research Officer, National Forest Inventory Manager, respectively, Department of Natural Resources Canada, 506 West Burnside Road, Victoria, BC, Canada V87 1M5.

MATERIAL AND METHODS

Data

Classified GIS polygons from the New Brunswick forest inventory (1981-1996) were used as data for this study. Delineation and classification of polygons were done from aerial photos (Leckie and Gillis 1995). Simulated sampling from this GIS coverage with 723 square plots of 6.25, 25, 100, 225, and 400 ha located on a square 10- x 10-km grid furnished the data. The 723 plot locations constitute all the 10- x 10-km nodes on the national forest inventory grid of sample locations in New Brunswick (7,265,912 ha). Figure 1 illustrates the grid and the GIS coverage extracted for each sample plot. Data for a single plot consist of the area and classification of each polygon inside the plot boundaries. The strata classes used in this study are listed in tables 1 and 2. All other classes were lumped into "others." Collectively we call the classes strata when no age distinction is made and we call them strata x age when an age group is associated with the class. The extracted data emulate, in part, the data types that would be generated during a national forest inventory.

Methods

Photo plots in the national inventory are the primary sampling units (PSU). Each PSU is considered an assembly of equally sized square secondary sampling units (SSU). The size of SSU was chosen as the minimum area that a photo interpreter would try to accommodate in the delineation of the photo plot into strata. Although the size of SSU is not explicit, we chose 0.01ha as the size of one SSU (Magnussen 1994). Note this 0.01-ha threshold is not to be confused with the minimum

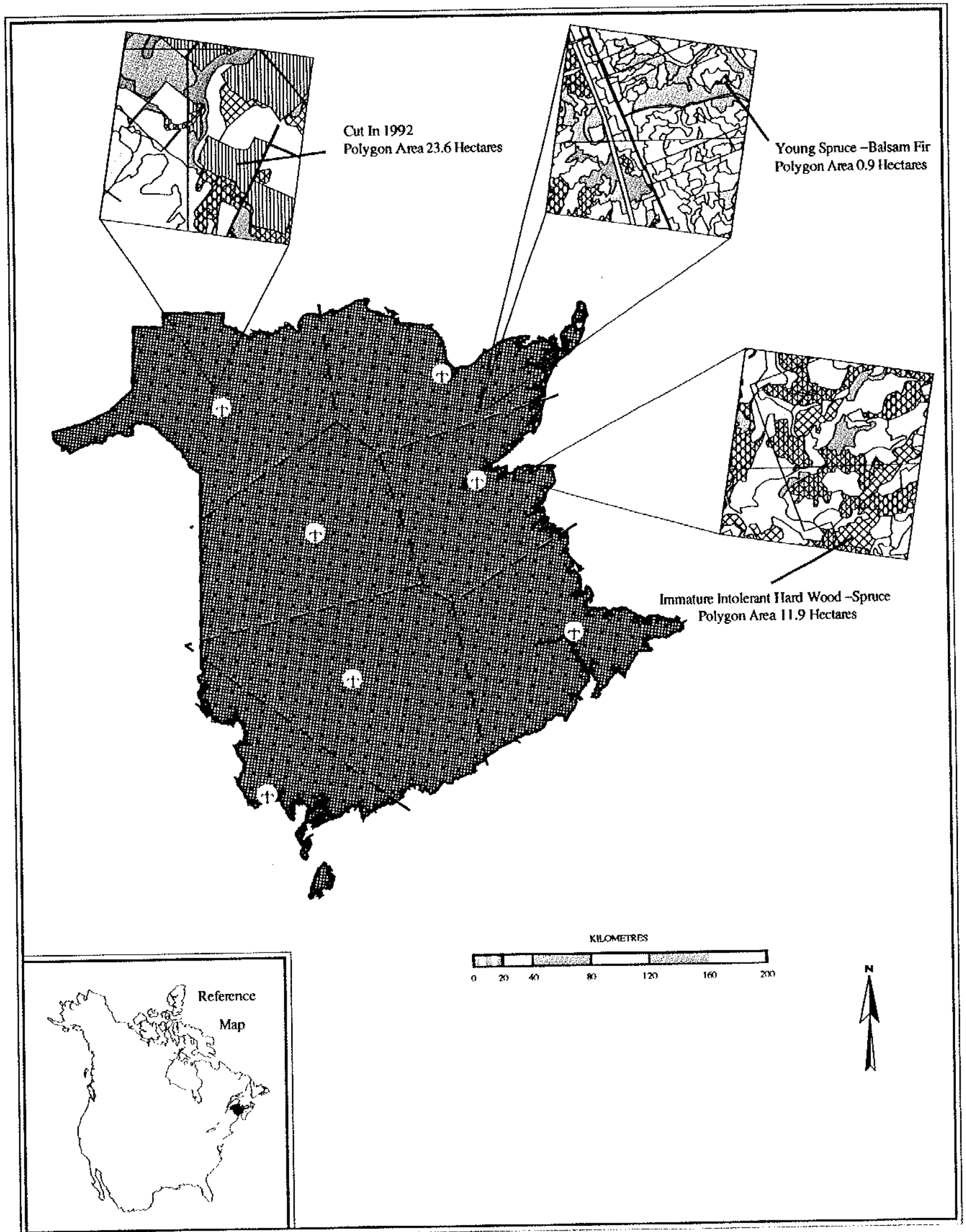


Figure 1.—*Inventory domain (New Brunswick) with a 10- x 10-km sample grid. Blow-ups provide details of three i2 x2 km primary sample units (PSU). Locations of seven airports (hubs) and the area they serve for the acquisition of air-photos are marked with circles and dashed lines.*

Table 1. *Sample estimates of strata proportions* ($\hat{p}$) *in New Brunswick, standard deviation of propensity of finding the strata in any given location* ($\hat{\sigma}_{prop.}$). *Results are based on interpretation of 723 plots of size 400 ha on a regular 10 x 10 km grid. Fit statistics of sample proportions to beta model* ($\hat{D}_{\max}$, $P(\hat{D} \mid H_0)$) *describes the fit of the observed proportions to a beta distribution.*

Strata	$\hat{p}$	$\hat{\sigma}_{prop.}$	$\hat{D}_{\max}$	$P(\hat{D})$
Balsam Fir - Intolerant Hardwood	0.014	0.246	0.037	0.790
Balsam Fir - Spruce	0.060	0.269	0.068	0.290
Balsam Fir - Tolerant Hardwood	0.014	0.245	0.066	0.430
Intolerant Hardwood - Balsam Fir	0.018	0.247	0.061	0.440
Intolerant Hardwood - Spruce	0.032	0.251	0.072	0.300
Intolerant Hardwood - Tolerant Hardwood	0.020	0.249	0.062	0.440
Intolerant - Hardwood	0.046	0.259	0.058	0.350
Other - Softwood	0.028	0.250	0.066	0.310
Pine	0.010	0.246	0.026	0.800
Poor Site - Spruce	0.009	0.244	0.036	0.760
Spruce - Balsam Fir	0.163	0.304	0.052	0.180
Spruce - Intolerant Hardwood	0.023	0.247	0.090	0.300
Spruce - Tolerant Hardwood	0.033	0.251	0.099	0.240
Tolerant Hardwood - Balsam Fir	0.014	0.246	0.062	0.440
Tolerant Hardwood - Intolerant Hardwood	0.021	0.250	0.052	0.490
Tolerant Hardwood - Spruce	0.030	0.252	0.085	0.250
Tolerant - Hardwood	0.084	0.288	0.068	0.220

Table 2. *Sample estimates of age group proportions in New Brunswick, standard deviation of deviation of probability of finding an area of one of the five groups in any given location* ($\hat{\sigma}_{prop.}$). *Estimates are based on interpretation of 723 plots of size 400 ha on a regular 10 x 10 km grid. Fit statistics of sample proportions (area-weighted average across Strata x Age groups) describes fit of observed proportions to a beta model* ($\hat{D}_{\max}$, $P(\hat{D} \mid H_0)$).

Age group	$\hat{p}$	$\hat{\sigma}_{prop.}$	$\hat{D}_{\max}$	$P(\hat{D})$
Regeneration	0.036	0.013	0.011	0.800
Immature-young	0.053	0.018	0.013	0.800
Immature-old	0.205	0.071	0.041	0.562
Mature	0.287	0.101	0.044	0.537
Old	0.039	0.014	0.015	0.799

area assigned to an individual forest stand. A SSU should contain only a single stratum (stratum x age) category. Larger SSU's will inevitably be mixed. Conceptually, an observation is made for each SSU in a PSU. In other words, we are dealing with a single-stage cluster design (Thompson 1992). All areas were rounded to nearest 0.01 ha before the analyses.

Estimation of strata proportions and their sample variances followed standard procedures for single-stage cluster sampling. The variance estimator of a forest type proportion (p) estimated from a sample of n_M primary units each with M SSU's taken from a finite number of N_M primary units of the same size is simply the variance of the primary sample unit corrected for the sampled fraction of the finite population. Another way of expressing this variance, and one that is better suited to explicate the impact of plot size, is (from Thompson 1992, page 122 after division by the square of the total size of New Brunswick , $N^2_M \bullet M^2$):

$$\hat{\sigma}_p^2(M) = \frac{(N_M - n_M)}{n_M \cdot N_M} \cdot \frac{(N_M \cdot M - 1)}{N_M - 1} \cdot \frac{\hat{\sigma}^2}{M^2} \cdot (1 + (M-1) \cdot \hat{\delta}_M) \quad (1)$$

where $\hat{\sigma}^2$ is the estimated total population variance, $\hat{\delta}_M$ the average intraplot correlation in plots with M SSU's, and N_M is the number of primary units of size M in New Brunswick (a unit is counted as "in" if the majority of SSU's in the plot were part of the land base of New Brunswick). Estimation of δ_M is thus central to quantify the impact of plot size on sample variances. Equation 1 can be extended to the covariance between two strata (strata x age) in a straightforward manner.

The intraclass correlation within a plot arises because the joint probability of two separate SSU's having the same strata value deviates from their expected value. Formally, for a categorical class T occurring with probability p in the entire population (of SSU's), the average intraclass correlation in a plot with M SSU's (enumerated by U_1, $U_2, \ldots, U_M$, with U_i = {T, not T} for i=1,2,...,M) is :

$$\bar{\delta}_M = \underset{\forall i,j \in M \wedge j \neq i}{E}\left[\frac{\Pr(U_i = T, U_j = T) - p^2}{p(1-p)}\right] \quad (2)$$

The joint probability for a single pair of SSU's will depend on the size of SSU's and the size of the plot. The average δ for a plot, however, depends only on the actual plot size. Equation 2 suggests a direct estimation of δ based on values assigned to the M SSU's in a PSU. Estimation of δ through analysis of variance (Landis and Koch 1977), estimation of a Dirichlet multinomial (Brier 1980), or estimation from the distribution of sampled proportions of strata areas (Prentice 1986) is possible.

We chose to estimate δ from the observed distributions of the 723 values of the proportion of a plot classified to a given stratum (stratum x age). Although not pursued further in this study, this approach is ideally suited for a hierarchical analysis of the effect of plot size (James 1975). When the probability of finding a given stratum (stratum x age) in an area of M SSU's follows a beta distribution with parameters α_M and β_M (the subscripts refer to the fact that the distribution depends on the number of SSU's in a plot), it can be shown that:

$$\delta_M = \frac{1}{(\alpha_M + \beta_M + 1)} \quad (3)$$

Maximum likelihood estimates of α_M and β_M (Prentice 1986) were obtained for each stratum and all strata x age combinations from the observed proportions of a plot occupied by a stratum or strata x age. The cumulative distribution functions of the ensuing beta models were compared to the empirical distribution. A test of no difference was sustained [test: Kolmogorov-Smirnov, Conover paired with Holm's sequential rejection procedure] throughout (tables 1 and 2). When averaged across strata, the intraplot correlations were: 0.47, 0.29, 0.17, 0.12, and 0.11 for the 6.25-, 25-, 100-, 225-, and 400-ha plots, respectively. Relative standard errors, i.e., the standard error in percent of the estimated mean proportion, were 1.0, 0.9, 0.7, 0.3, and 0.3 percent. Corresponding correlations averaged across strata x age combinations were about 0.03 lower and with relative errors of 0.9 to 1.3 percent. In 11 strata x age categories, we failed to estimate the intraclass correlations for the 6.25-ha plot. Another 11 estimates failed to show a monotonic decrease with plot size. We excluded these 22 (of 85) cases from the analyses of trends in the strata x age results (4). A MANOVA approach to estimation of δ gave estimates of δ that were identical to within 0.002. Estimated cross-correlation δ_{ij} between two strata (i,j, i≠j) were exploited for predicting all possible cross-correlation for any plot size M in the same way (4) as was done for the univariate correlation. Estimation of δ by way of estimating a Dirichlet multinomial (Brier 1980) produced frequency weighted results comparable (differences less than 0.01) to the mean intraclass correlations of strata or strata x age observations.

For each stratum and stratum x age, the δ for a plot with M SSU's was predicted from:

$$\hat{\delta}(M) = \frac{1}{2} - \frac{ArcTan[-(\hat{\lambda} + Log[M-1])/\hat{\omega}}{\pi} \quad (4)$$

where λ and ω are regression parameters to be determined by nonlinear least squares (Gallant 1987). This model (a modified cumulative Cauchy distribution) was found suited for predicting the average correlation in plots to within 0.03 (90 percent confidence interval) whenever the spatial correlation (ρ) between two separate units of land centered at (x_i, y_i) and (x_j, y_j) is a power function of distance $\rho^{|x_i - x_j| + |y_i - y_j|}$. We surmise this to be the spatial correlation function behind the estimated average within plot correlation. At the strata level $\hat{\rho}$ was, on average, 0.968.

Total variance ($\hat{\sigma}^2$ in Eq 1) of a strata (strata x age) proportion was determined as the sum of within- (binomial) and between-plot (beta) variances. The estimate of the total variance obtained from the 723 plots of size M was:

$$\hat{\sigma}^2 = \frac{(M-1)\,\Gamma(\hat{\alpha}_M + 1)\,\Gamma(\hat{\beta}_M + 1)}{M\,B(\hat{\alpha}_M, \hat{\beta}_M)\,\Gamma(\hat{\alpha}_M + \hat{\beta}_M + 2)} + \frac{\hat{\alpha}_M \hat{\beta}_M}{(\hat{\alpha}_M + \hat{\beta}_M)^2(\hat{\alpha}_M + \hat{\beta}_M + 1)} \tag{5}$$

where B is the complete beta integral. Each tested plot size gave rise to a separate estimate of the total variance. Provided that the beta distributions are internally consistent, we should not expect any difference in those estimates. However, the lack of fit may give rise to biased estimates. We tested the equality of the five estimates of total variance that were available from each strata and strata x age category with a Bartlett's test of variance homogeneity. The hypothesis was rejected 4 times (out of 17) at the strata level and 24 times (out of 85) at the strata x age level ($P < 0.01$). However, the marked kurtosis (median 25.5 for strata and 144.1 for strata x age) of the data inflated the rejection rate. A multivariate variant of the same test yielded a chi-square of 130.6 with 459 degrees of freedom ($P > 0.5$). Hence, we surmised that the total variance estimates derived from different plot sizes were indeed equal. Area-weighted averages were used henceforth.

From the estimates of total variance, predictions of δ for various plot sizes, and repeated use of equation (1), we computed the expected variance of strata (strata x age) proportions for a large number of designs by varying n_M, and M. Solutions were generated for a total 2,250 combinations of 150 plot sizes and 15 sample sizes. From these results, we then computed the number of plots and plot sizes that would satisfy 12 predetermined relative errors of precision through linear interpolations between available solutions. These calculations were completed in a univariate and a multivariate mode. Finally, we determined, for strata only, the number of plots (n) with M SSU's that would be needed to declare a relative error of x percent significant at the 100(1-0.05) percent level of significance with a probability of 90 percent (test power). Solutions were obtained by inversions from a multivariate F-test and a multivariate chi-square test. Assumptions about the distribution of estimated proportions were deemed satisfied for the entertained sample sizes (Fleiss 1981). For the chi-square test, the solution was the n satisfying (Rochon 1989):

$$F_{\chi^2_{\nu,\hat{\lambda}}}(q_{\chi^2_{\nu,1-\alpha}} \cdot \hat{C}) = 0.90 \tag{6}$$

where F is the cumulative distribution of the non-central chi-square distribution with ν degrees of freedom, non-centrality parameter λ, q is the $1-\alpha$ quantile of the central chi-square distribution with ν degrees of freedom, and $\hat{C}$ is equal to $1 + (M-1) \bullet \hat{\delta}$. For strata analysis $\nu = 16$; α was fixed at 0.05; λ was estimated as the expected value of a chi-square statistic given the subsumed relative error on the estimated strata proportions. Specifically:

$$\hat{\lambda} = \Delta^2 \cdot \hat{\pi}_{strata} \cdot \frac{n \cdot M}{(1 - \frac{n \cdot M}{NB}) \cdot (1 + M(1 - \hat{\delta}_{DM}(M))} \tag{7}$$

where Δ is the predefined subsumed error rate (fraction of estimated proportion), π_{strata} is the proportion of New Brunswick classified as one of the 17 strata (0.62), NB the number of SSU's in New Brunswick (determined by the majority rule), and M the number of SSU's in a plot of a given size (area); $\hat{\delta}_{DM}(M)$ is the predicted intraplot correlation for a plot with M SSU's derived from estimates based on the model in (4) fitted to values of Dirichlet multinomial intraplot correlations as defined by Brier. Solving the same problem but via a multivariate F-test with an observed F-value $\hat{F}$ (11) leads to n satisfying:

$$\Psi_{F_{\hat{\nu},\nu_2}}(\hat{k} \cdot q_{F_{\nu_1,\nu_2,1-\alpha}}) = 0.90 \tag{8}$$

where Ψ is the cumulative distribution function of an F distribution with ν and ν_2 degrees of freedom, k is a constant (9), and q_F is the $100(1-\alpha)$ percent quantile of an F distribution with ν_1 and ν_2 degrees of freedom. For strata-level solutions, $\nu_1 = 17$ and $\nu_2 = n-(17-1)$. The parameters ν and k achieve an approximation of the non-central F-distribution with non-central parameter $\hat{F}$ to a central F-distribution. Direct computation of the probability values of the non-central F-distribution was prohibitive for the involved parameters. ν was estimated as:

$$\hat{\nu} = \frac{(\nu_1 + \hat{F})^2}{(\nu_1 + 2\hat{F})} \tag{9}$$

where $\hat{F}$ is the observed F-ratio (see 11).

$$\hat{k} = \frac{\nu_1 + \hat{F}}{\hat{F}} \tag{10}$$

and finally,

$$\hat{F} = \vec{\Delta}' \cdot (\hat{\Sigma}_{tot}^{-1} \cdot \hat{\Sigma}_{\hat{\delta}}^{-1}) \cdot \vec{\Delta} \tag{11}$$

where $\vec{\Delta}$ is the n_{strata} x 1 column vector of subsumed errors on the strata estimates (relative error times estimated proportion), and $\vec{\Delta}'$ denotes the transposed vector. $\hat{\Sigma}$ is the estimated covariance matrices with subscript *tot* referring to the total variance-covariances (estimated as per Landis and Koch and $\hat{\delta}$ to the predicted intraplot correlations and cross strata correlation predicted from models (4) fitted to estimated correlation (ibid.). Dropping the test-power simplified the above expression to finding an n that satisfied $\Psi_{\hat{F}_{\nu_1, n-(\nu_1-1)}} = 0.95$.

A univariate approach to determine sample size via an F-ratio statistics followed the above procedure except that all covariance was set to zero.

Costs

Under the assumption that new photography and interpretation of photos is acquired for the purpose of the national forest inventory, we seek the inventory design (plot size, number of plots) with the lowest overall costs for a given predefined relative error of estimated proportions. The inventory costs include only design-dependent components: (i) flying to all plots in New Brunswick, (ii) photo acquisition (2 per plot) and processing, (iii) photo interpretation of plots consisting of a delineation of stand boundaries and classification into strata and age group, (iv) checking the photo interpretation by field visits, and (v) data processing. Table 3 lists the unit costs for each of these components. The total cost of a design was computed as the weighted sum where weights were either the number of plots or the area of a plot depending on the unit (table 3). Note while actual costs may vary considerably depending on local circumstances, we trust that their relative magnitudes have been captured. Relative values, not absolutes, determine the optimal design. Flying distances (FD, unit: km) to visit all national forest inventory plots in New Brunswick were predicted from the model on the next page (Eq. 12). The model was derived from measured total travel distance required to

Table 3. *Variable and fixed cost items for photo acquisition, photo-interpretation and field inspections.*

Item	Unit Cost ($)	Unit	Remarks
Flying	6.56	km	twin engine ($1000-$2200 per hour)
Interpretation	0.10	ha	strata delineation and age determination
Field visits	80.00	plot	site visits to one plot in 50 (2 people, one hour travel time to and from plot, $20/hr.) to check quality of interpretation
Field checks	0.12	ha	check interpretation (strata, age) by visual comparing interpretation results to field realities. 2 people can check 50 ha in 6 hours at $20/hour.
Fixed costs	20.00	plot	photos, plot registration, GIS work

visit all points on 40 different regular grids covering New Brunswick. The grids contained 100 to 5,000 plot locations. We assumed that the flying would be done from seven hubs serving as refueling stations (fig. 1). New Brunswick was, therefore, tiled into seven Voronoi cells, and travel distances for various grids were computed separately for each cell and then summed to a province-wide prediction model.

$$FD = Exp[5.94741 + 0.4787326 \cdot Log[n_M] + \frac{1}{2} \cdot 0.000176141] \tag{12}$$

where the last term in the exponential is one half the residual variance of the regression. The range of the airplane was set to 1,400 km; excess flying includes travel to and from fuel stations. Initial travel distances to dispatch the airplane to the hubs are ignored. For example, the total travel distance for a design with 500 plots systematically allocated on a grid is predicted to be about 7,500 km and 10,500 km for one with 1,000 plots. An approximate formula is provided by FD ~ 335 x $\sqrt{n_M}$

RESULTS AND DISCUSSION

Intraplot correlation for strata proportions declined at first rapidly and then more slowly as plot size increased (fig. 2

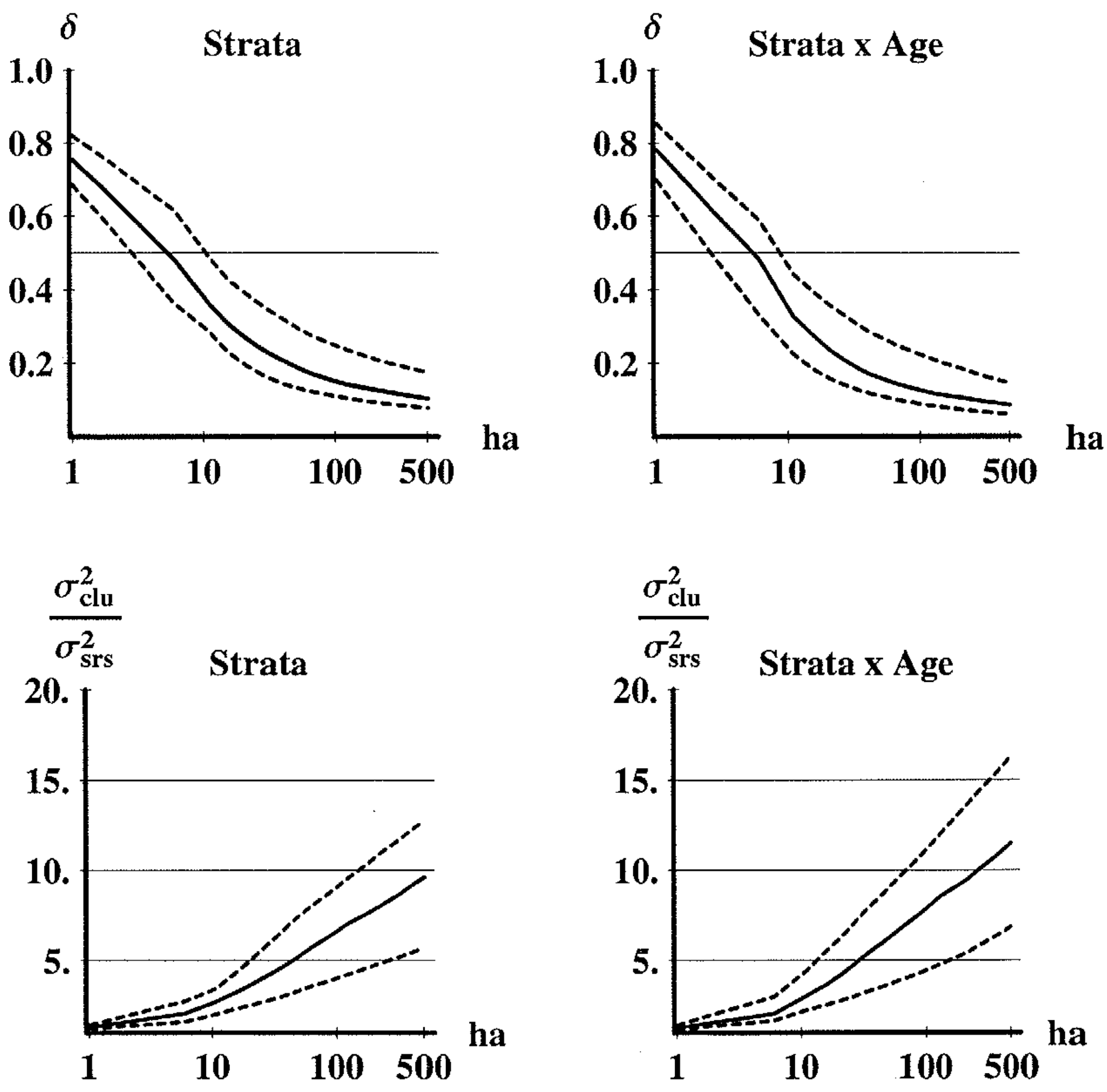


Figure 2.—*Top: Intraplot correlation (δ) versus plot size (ha). Full line: median. Dashed lines: upper and lower 10 percent quantiles. Bottom: Ratio of variance in cluster sampling (σ^2_{clu}) to the variance of simple random sampling (σ^2_{srs}) for various sizes (ha) of primary sampling units. Secondary sample unit = 100m².*

top, note the logarithmic x-axis). Increasing plot size from 100 m^2 to 10 ha, for example, caused a decline in average within plot correlation of 0.2, whereas an increase from 10 to 100 ha lowered the correlation by about 0.3, a pattern consistent with a spatial first-order autocorrelation process. Intraplot correlation varied among strata; figure 2 shows the 20 and 80 percent quantiles of the 17 strata and 63 strata x age combinations used in this study. Overall, an 80 percent confidence interval would be about 0.2 for smaller plots (< 10 ha) and shrink to about 0.1 for large plots (> 200 ha). Variance-weighted means for the five plot sizes were 0.46, 0.33, 0.21, 0.16, and 0.14, respectively. Simple means were 0.46, 0.29, 0.15, 0.13, and 0.11, while frequency-weighted estimates of δ for strata were 0.45, 0.29, 0.17, 0.13, and 0.11, respectively. For strata x age combinations, the decline was slightly steeper and the width of the 80 percent confidence band was somewhat wider, as dictated by their smaller proportions and higher degree of spatial scattering. Our choice of SSU (0.01 ha) has no effect on the above results; rather, a different SSU would merely impact on the resolution of results by virtue of the discreteness of SSU.

Intraplot correlation declined marginally at a rate below 1/M for a plot size increase from M to M+1, which means that the variance of a sample proportion based on a fixed sample area increased with sample size (fig. 2 bottom). Median sampling variances of strata proportions arising from sampling a fixed area with square 10-ha plots, for example, can be expected to be two times as high as sampling the same area with 100-m^2 plots. Conversely, 100- and 500-ha plots would inflate the variance sevenfold and tenfold. Results for individual strata varied greatly. Lower 10 percent bounds were roughly 0.5 the median value and the upper bound roughly 1.5 times the median. Going to the strata x age level (fig. 2 bottom) exaggerated parallel results by 10 to 50 percent.

The consequences of the tradeoff between plot size and number of plots are illustrated in figure 3. For a fixed inventoried area (x-axis), the expected precision drops exponentially as plot size goes up. For example, one could achieve a median error of 5 percent on a strata proportion by an inventory of about 1 percent of New Brunswick with square plots of 1 ha. Conversely, one would need to inventory about 4 percent of the province with a plot size of 25 ha to achieve the same precision. Insuring that most (80 percent) of the estimated strata proportions satisfy the target precision would add between 30,000 and 70,000 ha to the area inventoried. For a given inventory design, the median or 80 percent quantile precision at the strata x age is 1 to 5 percentage points lower than at the strata level. Finally, when inventory efforts to achieve a given precision with a fixed plot size are translated into necessary sample sizes, we get results as those in figure 4. We notice a dramatic increase in the minimum sample size needed to achieve, say a relative error of 10 percent, when plot size goes below 80 ha. For

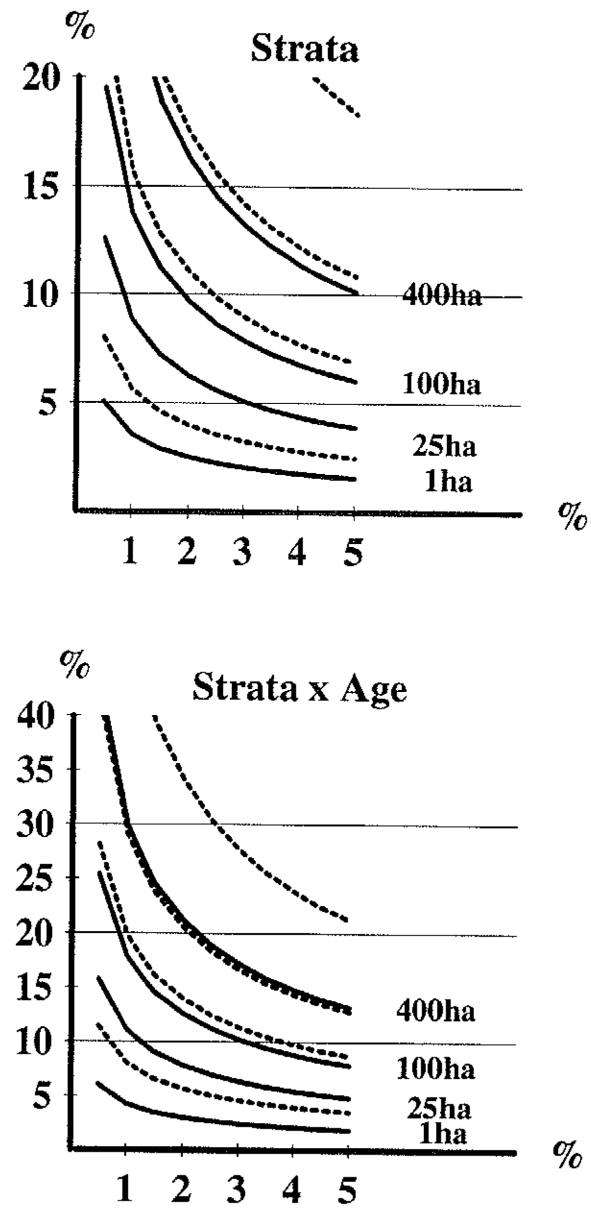
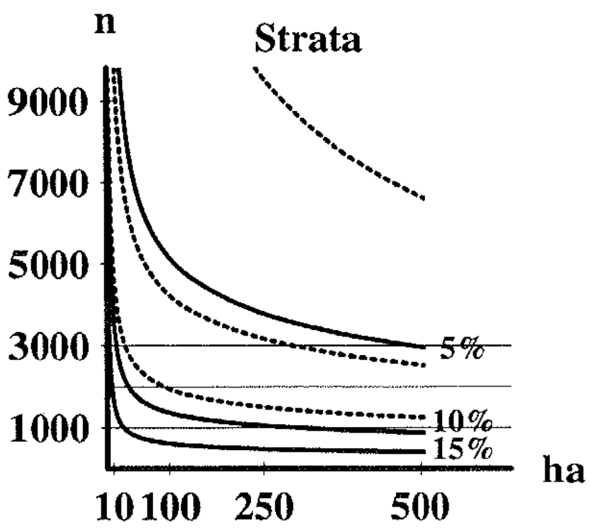


Figure 3.—*Expected relative error (%) of inventory estimates of strata and strata x age areas (y-axis) plotted against the percentage of New Brunswick inventoried (x-axis). Full lines: median result obtained with a plot size of 1, 25, 100, and 400 ha. Dashed lines are the corresponding 80% quantiles.*

plot sizes in excess of 100 ha, the tradeoff between size and number of plots becomes more equitable. While relationships displayed in figures 2-4 are quite general, the actual numbers and their interrelationships are unique to the landscapes of New Brunswick.

As expected, costs for designs meeting a specified target precision varied considerably. Designs with plot sizes in the range of 70 to 130 ha for a target precision of 6 to 14 percent appear attractive (fig. 5). Smaller plots incur more flying costs but require less interpretation and data



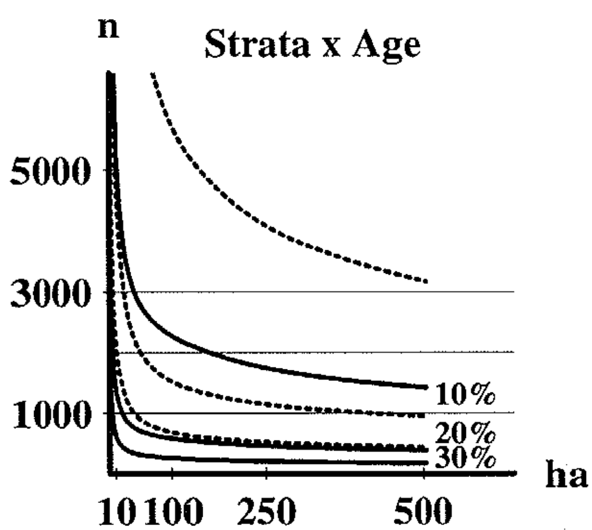


Figure 4.—*Number of plots (n) required to achieve a median relative error of 5, 10, and 15 percent on inventory estimates of strata and strata x age areas as a function of plot size (x-axis). Dashed lines are the combinations of plot size and number of plots needed to ensure that 80 percent of the results have errors less than 5, 10, and 15 percent.*

processing than larger plots. Table 4 lists the individual minimum cost solutions. The size of the most cost-effective plot increases when the target relative precision is lowered. Insuring that the target precision applies to 80 percent of the strata instead of just the 50 percent (table 4) appears to lower the plot size at the strata level, while the reverse is true at the strata x age level. The high number of plots needed to insure that 80 percent of the relative errors meet the target flipped the marginal cost relations between size and numbers of plots.

Precision targets can be further refined by imposing a threshold not to be exceeded with a probability of $100(1-\alpha)$ percent and by requesting a minimum probability of detecting excess. These targets can be expressed as average expectations in a univariate or multivariate context. The latter exploits the covariance between inventoried attributes while the errors are averaged over all strata (strata x age). Estimates of total covariances were those from the MANOVA's, and predictions of cross-correlation were made via equation (4) adapted to a bivariate case. Table 5 presents five alternative minimum cost solutions, each based on a different set of probability statements or test statistics. In general, achieving a probability of 90 percent to declare an error in excess of the targets significant ($P \leq 0.05$) instead of one around 50 percent would entail an increase in the area inventoried of about 0.3 percent. In a multivariate context, the tradeoff between number of plots and size of plots is less clear than in a univariate framework due to the impact of covariance and the distribution of the test statistics. On average, a multivariate approach (Solutions I, II, and V) calls for less area to be inventoried than a univariate approach. The difference is most marked for low error targets and decreases with increasing error tolerance. Also, the choice of test statistics may modify the solution. For example, the II and V solutions in table 5 are accompanied by identical probabilistic statements and produce very similar solutions in terms of the total area to inventory, yet larger plots are clearly favored by solutions based on the F-test, a fact attributed to the trend for the determinant of the inverse of the covariance matrix to increase at a rate below $1/M$ when plot size is increased by one unit from M to $M+1$.

CONCLUSIONS

The tradeoff between the number of plots and the size of plots to use in a national forest inventory poses a classical dilemma for those who design the inventory. The quantitative analyses presented here cover only a narrow range of issues. Ultimately, it should be the value of the inventoried data that drives the design decisions about plot size and number of plots. For strata and strata x age area estimates, it is possible that the value may be inversely proportional to the estimate itself. In the context of conservation and biodiversity, focus will be on rare strata.

Figure 5.—*Effect of plot size (x-axis) on inventory costs (y-axis) when the number of plots is chosen to ascertain a target (median or the 80 percent quantile) relative error on the estimate of strata (strata x age) areas. The minimum cost solutions are indicated with gray squares.*

Sampling designs under the aegis of conservation and biodiversity issues are dictated by a very different set of imperatives than those epitomized by this study. Until the value of forest inventory information has been formulated, the best approach to settle the issue of plot size and number of plots is to minimize costs for a given precision target. Within this albeit narrow scope, a plot size in the order of 100 to 150 ha for stratifying the landbase by strata and age categories appears attractive when the precision targets fall in the 8 to 12 percent range. Note, however, that a precise formulation of these targets and the level (strata or strata x age) to which they apply can have a dramatic impact on the total area to inventory.

ACKNOWLEDGMENTS

We thank the New Brunswick Department of Natural Resources and Energy for making the GIS data available for our study. A special thank you goes to Director R. Dick, Forest Management Planning Section, for arranging

Table 4. *Lowest cost combinations of plot size and number of plots for a given relative error on estimated strata (strata x age) proportions. Error rates apply to either the median expectation or as an upper limit of 80% of the strata viz. strata x age combinations.*

	Relative error of estimate					
	6%	8%	10%	12%	14%	16%
Strata, Median						
Number of plots	3777	2169	1354	921	663	508
Spacing, approx., km	4	6	7	9	10	12
Size of plots, ha	91	91	101	111	121	121
π_{sample} %	3.9	3.6	1.9	1.5	1.2	1.0
Strata, 80%						
Number of plots	11062	6900	4542	3203	2291	1766
Spacing, approx., km	3	3	4	5	6	6
Size of plots, ha	91	81	81	81	91	91
π_{sample} %	9.3	6.1	4.4	3.3	2.7	2.3
Strata x age, Median						
Number of plots	6672	3863	2414	1690	1248	929
Spacing, approx., km	3	4	5	7	8	9
size of plots, ha	71	71	81	81	81	91
π_{sample} %	6.5	3.8	2.7	1.9	1.4	1.2
Strata x age, 80%						
Number of plots	12798	8550	5717	4069	3035	2270
Spacing, approx., km	2	3	4	4	5	6
Size of plots, ha	121	101	101	101	101	111
π_{sample} %	21.3	11.9	7.9	5.7	4.2	3.5

Table 5. *Strata level multivariate F-test and Chi-square solutions to lowest cost combinations of plot size and number of plots for a given relative error on estimates.*

Solution I: Error is declared significant (P~ 0.5) at the $\alpha = 0.05$ level of significance of a type I error. Test: multivariate F-test.

Solution II: Error is declared significant (P=0.90) at the $\alpha = 0.05$ level of significance of a type I error. Test: multivariate F-test.

Solution III: Error is declared significant (P~ 0.5) at the $\alpha = 0.05$ level of significance of a type I error. Test: multivariate F-test with all covariances equal to zero.

Solution IV: Error is declared significant (P= 0.90) at the $\alpha = 0.05$ level of significance of a type I error. Test: multivariate F-test with all covariances equal to zero.

Solution V: Error is declared significant (P= 0.90) at the $\alpha = 0.05$ level of significance of a type I error. Test: Chi-square.

	Relative error of estimate					
	6%	8%	10%	12%	14%	16%
Solution I						
Number of plots	915	589	394	297	200	193
Size of plots, ha	124	93	88	79	108	85
π_{sample} %	1.6	0.8	0.5	0.3	0.3	0.2
Solution II						
Number of plots	1101	616	395	277	215	160
Size of plots, ha	124	138	152	164	156	192
π_{sample} %	1.9	1.2	0.8	0.6	0.5	0.4
Solution III						
Number of plots	2055	1173	771	580	401	302
Size of plots, ha	105	108	108	113	115	129
π_{sample} %	3.0	1.7	1.1	0.9	0.6	0.5
Solution IV						
Number of plots	2494	1403	899	625	456	347
Size of plots, ha	103	111	117	123	136	147
π_{sample} %	3.5	2.1	1.4	1.1	0.9	0.7
Solution V						
Number of plots	2233	1248	791	539	387	299
Size of plots, ha	71	76	80	87	97	94
π_{sample} %	2.2	1.3	0.9	0.6	0.5	0.4

and facilitating data extraction and transfer, and to David Seemann for GIS support.

The following people reviewed this paper: Dr. J. Flewelling, Consultant Applied Statistics and Biometrics, Kent, WA, USA; Dr. S. Otukol, Ministry of Forest, Resource Inventory Branch, Victoria, BC, Canada; and Dr. M. Penner, Consultant, Huntsville, ON, Canada

LITERATURE CITED

Acka, A. 1993. Technique and importance of continuous forest inventory. Allgemeine Forst und Jagdzeitung. 164: 193-198.

Brier, S.S. 1980. Analysis of contingency tables under cluster sampling. Biometrika. 67: 591-596.

Gallant, A.R. 1987. Nonlinear statistics models. New York, NY: John Wiley and Sons.

Fleiss, J.L. 1981. Statistical methods for rates and proportions. 2d ed. New York, NY: John Wiley and Sons.

James, I.R. 1975. Multivariate distributions which have beta conditional distributions. Journal of the American Statistical Association. 70: 681-684.

Landis, J.R.; Koch, G.G. 1977. A one-way components of variance model for categorical data. Biometrics. 33: 671-679.

Levy, P.S.; Lemeshow, S. 1991. Sampling of populations: methods and applications. New York, NY: John Wiley and Sons.

Loetch, F.; Haller, K.E. 1964. Forest inventory. BLV. - Verlagsgesellschaft. Munich, GE.

Magnussen, S. 1989. Inter-plant interactions and their influence on within and among plot variances. Scandanavian Journal of Forest Research. 4: 369-377.

Magnussen, S. 1994. A coordinate-free area variance estimator for forest stands with a fuzzy outline. Forest Science. 42: 76-85.

Mahalanobis, P.C. 1944. On large-scale sample surveys. Journal of the Royal Statistical Society B. 231: 41-451.

Matérn, B. 1980. Spatial variation. Berlin, GE: Springer.

Okabe, A.; Boots, B.; Sugihara, K. 1992. Spatial tesselations concepts and applications of Voronoi diagrams. Chicester, UK: John Wiley and Sons.

Prentice, R.L. 1986. Binary regression using an extended beta-binomial distribution, with discussion of correlations induced by covariate measurement errors. Journal of American Statistical Association. 81: 321-327.

Rochon, J. 1989. The application of the GSK method to the determination of minimum sample sizes. Biometrics. 45: 193-205.

Thompson, S.K. 1992. Sampling. New York, NY: John Wiley and Sons.