Sampling design to estimate land use under a budget constraint

Mario E. Niklitschek
and Guillermo Trincado

Facultad de Ciencias Forestales y Recursos Naturales
Universidad Austral de Chile

Based on the published paper:

Niklitschek M. and G. Trincado 2011. A cost effective stratified two-stage sampling design to estimate the forest land area of southern Chile.

Canadian Journal of Forest Research. 41: 1509-1521.

Determining land use: mapping v.s sampling

Advantages of sampling

- Lower cost allowing higher coverage in time and space
- The precision lost associated with the sampling error is compensated by greater precision in cartographic classification.
- However, the estimation of land use changes might require high sampling intensities

Determining land use: mapping v.s sampling

Modelling (prediction)

- Greater precision in the measurement of the response variable and co-variables
- Easier to integrate geo-referenced data from different sources
- Feasible to collect complementary primary data on biophysical and socioeconomic variables.
- Enable to predict scenarios of interest at the parcel or point level.

Stratified design in two stages

- **Stratification**
  - Greater estimators precision
  - Ensure appropriate spatial coverage
- **Clustering**
  - Reduce field costs and the costs of image acquisition and processing.
  - Larger clusters however causes lower statistical efficiency.

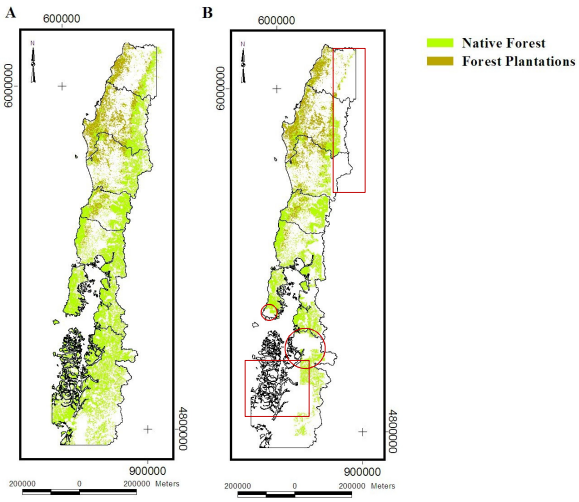
Sampling frame for regions seventh thru eleventh

Study area

The interest area is the area susceptible to change which is obtained subtracting:

- National parks
- Private main parks (Pumalin and Tantauco)
- Fresh water bodies (rivers and lakes)
- High lands above the tree line

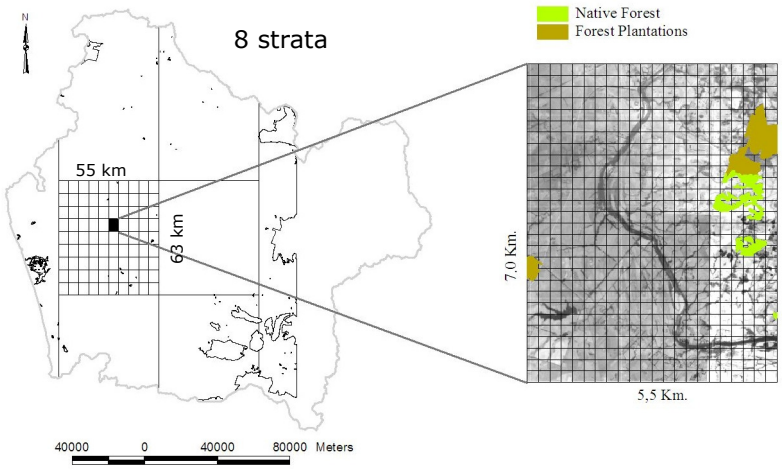
Sampling frame for regions seventh thru eleventh



Strata, primary and secondary units

- The basic sampling unit or secondary unit is the minimum resolution area of the national cadastre (6,25 ha.).
- The strata were obtained overlaying a grid (55 per 63 kilometers) on administrative map with regional borders.
- After integrating small neighboring strata the sampling frame was divided in 46 strata.
- Three grid dimensions to generate primary units were evaluated: 3850 ha, 1925 ha y 962,5 ha.
- Both primary and secondary units have unequal sizes.

Illustration of strata and sampling frame for the ninth region



Selection of units in two stages

- Both primary and secondary units were selected by simple random sampling without replacement
- At both stages we use non-overlapping random sample selection
- This method allows to add subsequently new units to the sample to improve the precision of an estimator without replacing previously selected units

Estimation of the area under forest plantations and native forests

Estimators

- Because the primary units have unequal sizes a ratio estimator might be more appropriate that unbiased estimator.
- With proportional allocation the ratio estimator is the simple average for the selected sample:

$$\hat{P}^{cs} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} a_{hij}^{\,c}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} a_{hij}}$$

- The unbiased estimator has a denominator a fixed area sample size that is $\frac{A}{M}\cdot n\cdot \bar{m}$

Estimation of the area under forest plantations and native forests

Precision

- The mean squared error (MSE) the variability of a biased estimator with respect the population value.
- An approximation of the MSE for the ratio estimator is:

$$MSE(\hat{P}^c) \cong \frac{1}{\hat{A}^2}\left(\frac{M}{n\bar{m}} S\tilde{S}^c + \frac{N}{n}(S\tilde{F}^c - S\tilde{S}^c) - S\tilde{F}^c\right)$$

where $\widetilde{SS}$ and $\widetilde{SF}$ are measures of the variability within primary units and among primary, respectively.

Cost effectiveness analysis

- Smaller primary units reduces the MSE for a given sample area.
- The determination of the number of units in both stages for a given primary unit size z and a budget B is equivalent to solve the following problem:

$$v^z(c_1^z, c_2, B-c_0) = \underset{n,\overline{m}}{Minimize}\left[MSE^z(.): B-c_0 = c_1^z n + c_2 n\overline{m}\right]$$

Post-stratification

- For a given sampling design is possible to improve precision using additional information of the population that is correlated with the variable of interest.
- An estimator with post-stratification uses as weights the proportion of secondary units that fall on each post-strata.
- The estimation of the MSE needs to use linearization procedures (Sha, 2004)

Results: The size of primary units

- The three sizes considered are equivalent to 616, 308 and 154 secondary units.

Cost effectiveness of different PSU (percentage REMC)

Cost variation among PSU sizes	Forest Plantations		
	S	S/2	S/4
Fixed (90 USD)	8.1	8.7	9.4
10 percent reduction*	8.1	8.2	8.5
20 percent reduction	8.1	7.8	7.5
Proportional to S	8.1	6.2	4.8

* PSU cost of S\2 is 0.9 the cost of PSU size S and PSU cost of S\4 is 0.81 the cost of S

Optimum sample size for the ratio and unbiased estimators

Sample size for the entire study area (95 percent confidence level)

Land Use	p	Ratio estimator			Unbiased estimator		
		$\overline{m}*$	n		m^*	n	
			E=10%	E=15%		E=10%	E=15%
Forest Plantations	0.149	30.5	532.2	252.9	28.4	603.3	289.5
Native Forests	0.329	27.3	174.8	79.4	21.8	265.2	121.9

- An average of 31 SUs per PU and 532 PUs are required to predict the forest plantation area with a 10 % error at the 95 % confidence level

Sampling size to predict forest areas at the regional level.

Sample size by region (ratio estimator).

Region	Forest Plantations				Native Forests			
	p		n		p		n	
		$\bar{m}^*$	E=10%	E=15%			E=10%	E=15%
VII	0.212	28.0	241.4	142.0	0.050	39.8	408.8	306.4
VIII	0.312	33.8	178.5	88.4	0.173	30.4	278.7	147.7
IX	0.149	33.8	328.8	193.8	0.204	35.0	219.1	116.4
XIV	0.107	31.8	324.0	223.5	0.468	24.2	118.3	60.6
X	0.016	33.8	1048.5	873.4	0.593	22.8	89.3	41.4
XI	0.005	47.8	544.6	517.2	0.5878	25.0	86.7	42.2

- For the nine region an average of 34 secondary per primary unit and 329 primary units are required with a 10 % error at the 95 % level

Post-stratification using forest growing zones

- We use the growing zones proposed by Schlater and Gerding (1995

Percentage of forest plantations and native forests by post-strata (Schlatter and Gerding 1995).

Poststrata	Geographical location	Forest Plantations (%)	Native Forests (%)
1	West side coastal range	37.56	26.15
2	East side coastal range	37.08	25.02
3	Coastal dryland	12.40	9.03
4	Central valley	13.71	9.54
5	Andes piedmont	16.20	37.55
6	Andes highlands	1.55	79.36
7	Southern region (XI)	0.24	67.59

Post-estratificación utilizando zonas de crecimiento

Sampling precision with the poststratified ratio estimator (n=253, $\overline{m}$ =30).

Estimator	Forest plantations		Native forests	
	RMSE	RMSE %	RMSE	RMSE%
Unbiased estimator	0.01225	8.209	0.01712	5.202
Ratio estimator	0.01142	7.655	0.01380	4.193
Poststratified ratio estimator	0.01132	7.626	0.01214	3.709

Conclusions

- Area sampling can be used to estimate land use categories based on a rigorous statistical sampling design.
- Regional inference requires relatively large samples.
- For area applications is required to determine the optimal primary unit size.
- When variation in the primary unit size is necessary, the ratio estimator provides more precision than the unbiased estimator.
- Area sampling can contribute to identify the drivers of land use changes.