AREAL SAMPLING STRATEGIES FOR ESTIMATING TOTALS AND AVERAGES ON A GRID OF QUADRATS: APPLICATIONS TO FOREST SURVEYS

Fattorini Lorenzo[1] & Pagliarella Maria Chiara[2]

[1] *Università di Siena (Italy), lorenzo.fattorini@unisi.it*
[2] *Università di Siena (Italy), mariachiara.pagliarella@unisi.it*

Résumé. Le projet REDD (Reduction of Emissions from Deforestation and forest Degradation) a été mis en place en 2005 sous l'égide des Nations-Unies. Le suivi de la couverture forestière, notamment par des méthodes statistiques, est un des objectifs de ce projet. La couverture forestière est le plus souvent estimée à grande échelle par des stratégies d'échantillonnage spatial, dans lesquelles la région étudiée est partitionnée en *N* polygones - ou quadrats - de tailles égales. L'estimation s'appuie alors sur l'interprétation de photographies aériennes prises sur un échantillon de *n* unités. Afin de mesurer l'intérêt d'intégrer des contraintes spatiales dans le plan d'échantillonnage et de tenir compte d'une auto-corrélation spatiale entre les unités échantillonnées, nous proposons de comparer des schémas d'échantillonnage classiques tels que l'échantillonnage aléatoire simple sans remplacement (SRSWR), l'échantillonnage stratifié one-per-stratum (STR) et l'échantillonnage systématique (SIS) à des méthodes ad-hoc d'échantillonnage spatial comme la méthode du pivot local de type I (LPM1, voir Grafström et al., 2012), l'échantillonnage stratifié Generalized Random-Tesselation (GRTSS, voir Steven et Olsen, 2004), l'échantillonnage de Poisson corrélé spatialement (SCPS, voir Grafström, 2012), le schéma drawn-by-drawn qui évite la sélection d'unités contigües (voir Fattorini, 2006) et l'échantillonnage spatial doublement équilibré (voir Grafström et Tillé, 2013). Une étude par simulation permet de comparer les performances de chacune de ces stratégies.

Mots-clés. Echantillonnage équilibré spatialement, estimateur de la différence, estimateur de Horvitz-Thompson, information auxiliaire, schémas ΠPS, schémas PPS, simulation, suivi forestier

Abstract. The Reduction of Emissions from Deforestation and forest Degradation (REDD) project was proposed and initiated in 2005. Monitoring of forest cover by statistical methodologies is a key pre-requisite. Forest cover is usually estimated at large scale by spatial sampling strategies, in which the study region is partitioned into *N* polygons of equal size (e.g. quadrats). Then, a sample of *n* units is selected, aerial photos are provided and visually interpreted to determine the forest cover. In order to incorporate spatial aspect into the design and to account for the presence of spatial autocorrelation among the units, in this paper are compared familiar sampling schemes such as simple random sampling without replacement (SRSWR), one-per-stratum stratified sampling (STR) and systematic sampling (SIS) against ad hoc spatial schemes such as Local Pivotal Method of the first type (LPM1) by Grafström et al. (2012), Generalized Random-Tessellation Stratified Sampling (GRTSS) by Steven and Olsen (2004), Spatially Correlated Poisson Sampling (SCPS) by Grafström (2012), the drawn-by-drawn scheme avoiding the selection of contiguous units by Fattorini (2006) and the Doubly Balanced Spatial Sampling by Grafström and Tillé (2013). A simulation study is performed in order to compare and check the validity of each strategy.

Keywords. spatially balanced sampling, Horvitz-Thompson estimator, auxiliary information, difference estimator, ΠPS schemes, PPS schemes, simulation study, forest monitoring

1 Introduction

The role of forests as the largest reservoir of biodiversity and as a major carbon sink is well recognized. Deforestation and forest degradation are estimated to contribute to about 20% of the total greenhouse gas emissions. Therefore, reducing these phenomena will have a direct impact on the reduction of gas emissions. To this purpose, a post-Kyoto protocol, the Reduction of Emissions from Deforestation and forest Degradation (REDD) was proposed and initiated in 2005. In this framework the monitoring of forest cover at large scale by statistically sound methodologies is a key pre-requisite for the REDD project. The purpose of this paper is to investigate and compare the use of traditional sampling and ad hoc spatial schemes.

2 Statement of the problem

In order to estimate the proportion forest cover in tropical countries, where ground inventories cannot be performed owing to forest inaccessibility, remote-sensed inventories are performed (Sannier et al, 2013). The study area is partitioned into a grid of N quadrats of pre-fixed size (e.g 1 ha) and the proportion of forest cover y_j is recorded within quadrat j using a combination of satellite data (e.g. Landasat imagery) and manual enhancements achieved from online image and map archives (including Bing Maps, Google Earth, Google Maps and Open Street Map) to ensure the greatest possible accuracy. The proportion of forest cover on the whole study area is given by the population mean

$$\overline{Y} = \frac{1}{N}\sum_{j=1}^{N} y_j$$

Owing to the manual operations, the recording of the y_js is a time expensive procedure. Thus, the variable can be recorded only for a sample of n quadrats selected from the population. On the other hand, the proportion of forest cover from satellite data can be obtained automatically from the satellite classification provided at pixel level. Thus, if P_j denotes the set of the M pixels constituting the j-quadrat, the proportion of forest cover from satellite information is given by

$$x_j = \frac{1}{M}\sum_{i\in \mathrm{P}_j} I_{ij}$$

where I_{ij} is an indicator variable equal to 1 if pixel i of quadrat j is classified as forest and 0 otherwise. Because the x_js are readily available for all the quadrats, they can be used as auxiliary information to improve estimation.

3 Familiar sampling schemes

Usually, the schemes adopted to select units are familiar schemes such as stratified or systematic sampling (Sannier et al, 2013). Simple random sampling without replacement (SRSWOR) will be considered as benchmark. In both stratified and systematic sampling, the grid of quadrats is partitioned into n strata of adjacent quadrats. Then, in stratified sampling a quadrat is randomly selected within each stratum, while in systematic sampling a quadrat is randomly selected in one stratum (e.g. the first) and then systematically repeated in the remaining strata.

4 Ad hoc spatial sampling schemes

In order to incorporate spatial aspect into the design, spatially balanced sampling are considered. These are the local pivotal method of the first type (LPM1) by Grafström et al. (2012), the generalized random-tessellation stratified sampling (GRTSS) by Steven and Olsen (2004), the spatially correlated Poisson sampling (SCPS) by Grafström (2012), the drawn-by-drawn scheme to avoid the selection of contiguous units by Fattorini (2006) and the Doubly Balanced Spatial Sampling combining pivotal and cube methods by Grafström and Tillé (2013). Another reference is the inclusion probability proportional to size sampling plans excluding adjacent units (IPPSEA) by Mandal et al. (2008). This scheme makes use of combinatorial properties of incomplete block designs as modification of the approach given by Nigam et al. (1984) and linear programming approach initiated by Rao and Nigam (1990, 1992). The scheme is not treated here because it can be applied only for small population and sample sizes.

5 Simulation study

A simulation study was performed in order to compare and check the validity of each strategy. We were concerned with three populations of 400 quadrats of size 100 ha located in Central Italy in which the value y_j of quadrat *j* represented the forest proportion in the quadrat. The averages of the three populations represented the forest coverage on the whole study areas and were equal to 57%, 18,9% and 6,3% corresponding to high, medium and low coverage, respectively. *Figure 1* plots forest (white) and non-forest (black) areas for the three populations considered in the study.

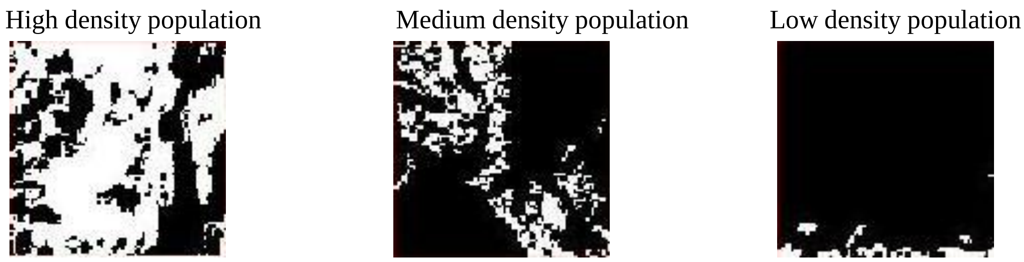

Figure 1

The auxiliary information by satellite map was generated presuming that $x_j = y_j - e_j$ where e_j were normally distributed residuals with 0 mean and variance such that to determine a squared correlation coefficient with the survey variable of 0.3, 0.6, 0.9.

References

[1] Fattorini, L. (2006), Applying the Horvitz-Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities, *Biometrika*, 93, 269-278.
[2] Grafström, A. (2012), Spatial correlated Poisson sampling, *Journal of Statistical Planning and Inference*, 142, 139–147.
[3] Grafström, A. and Tillé, Y. (2013), Doubly balanced spatial sampling with spreading and restitution of auxiliary totals, *Environmetrics*, 24, 120–131.
[4] Grafström, A., Lundström, N.L.P. and Schelin, L. (2012), Spatially Balanced Sampling through the Pivotal Method, *Biometrics*, 68, 514–520.
[5] Mandal, B.N., Parsad, R. and Gupta, V.K. (2008), IPPS sampling plans excluding adjacent units, *Communications in Statistics – Theory Methods*, 37 (16), 2532-2550.
[6] Nigam, A. K., Kumar, P. and Gupta, V. K. (1984), Some methods of inclusion probability proportional to size sampling, *J. Roy. Statist. Soc.*, B 46, 564–571.
[7] Rao, J. N. K. and Nigam, A. K. (1990), Optimal controlled sampling: a unified approach, *Int.*

Statist. Rev., 60(1), 89–98.
[8] Rao, J. N. K. and Nigam, A. K. (1992), Optimal controlled sampling designs, *Biometrika,* 77(4), 807–814.
[8] Sannier, C., Mc Roberts, R.E., Fichet, L.V. and Makaga, E.M.K. (2013), Using regression estimator with Landsat data to estimate proportion forest cover and net proportion deforestation in Gabon, *Remote Sensing of Environment*, online.
[10] Steven, D.J. and Olsen, A.R. (2004), Spatially Balanced Sampling of Natural Resources, *Journal of American Statistical Association*, 99 (465), 262-278.