

Area Frames for Land Cover Estimation: Improving the European LUCAS Survey.

Javier Gallego.

JRC-IES, Ispra (Italy). javier.gallego@jrc.it

Abstract

LUCAS (Land Use/Cover Area frame Survey) has been launched by the European Union (EU) in 2001 based on an area frame of points with a two-stage systematic sampling. We discuss some possible options to improve LUCAS assuming that it could combine in the future an annual survey focused on arable land with a general purpose survey. A test is being carried in Greece out with a two-phase area frame of unclustered points, following the current approach of the Italian AGRIT survey. The high share of the variance coming from the first sampling stage suggests that unclustered points can be better. The same conclusion is reached considering the "equivalent number of points of a Primary Sampling Unit". A simulation shows that a stratification based on photo-interpretation is much more efficient with unclustered points (single stage sampling) than with the current two-stage design if the photo-interpretation is very accurate. The gain is debatable with a moderate amount of photo-interpretation errors. A remark is made showing that the current LUCAS two-stage sampling scheme can be seen as a single-stage systematic sampling, which simplifies the estimation of variances

Area frame surveys of points: two-stage versus two-phase schemes.

We speak of an area frame survey when the sampling units are defined on a cartographic representation of the surveyed territory. Area frames generally match more precisely the population than list frames. The units of an area frame can be points, transects (straight lines of a certain length) or pieces of territory, often named segments. The LUCAS (Land Use/Cover Area-frame Survey) was launched by the European Union (EU) in 2001. LUCAS is based on a non-stratified, systematic, two-stage sampling scheme, with Primary Sampling Units (PSUs) defined as a rectangle of 1500 x 600 m following a grid of 18 km [Delincé, 2001, Bettio et al., 2002]. In each PSU 10 points (SSUs) are selected arranged on two rows of 5 points with a step of 300 m. The "point" (SSU) is defined as a circle of 3 m to be consistent with ground survey specifications. This two-stage systematic design can be described as a single stage systematic sample if the PSU is defined as containing only the 10 "square points" of 3 x 3 m, so that the 1500 x 600 m rectangle contains $100 \times 100 = 100000$ PSUs. This scheme simplifies the problem of variance computation.

An alternative is the current design of the Italian AGRIT survey [Consortio ITA, 2003, Martino, 2003]: single stage (unclustered points) two-phase sampling: the first phase gives a systematic sample of unclustered points that are photo-interpreted and subsampled (second phase) with higher rates in agricultural strata. A similar approach approach has tested in Greece in 2004: its data are being analysed at the time of writing this paper, that focuses on the analysis of the reasons why an improvement can be obtained with this approach. For the comparison of the two-stage scheme with the single-stage, two-phase scheme, it is useful to estimate the part of variance coming

from each sampling stage. A good idea can be obtained using the formulae for two-stage random sampling [Cochran, 1977]:

$$v(\bar{y}) = \left(\frac{N-n}{N} \right) \frac{s_1^2}{n} + \frac{n}{N} \left(\frac{M-m}{M} \right) \frac{s_2^2}{mn} = (1-f_1) \frac{s_1^2}{n} + (1-f_2) \frac{s_2^2}{mN}$$

Some care is needed when estimating the contribution of each stage. The proportion of

variance coming from the first stage is estimated by
$$rb = \frac{(1-f_1)}{n} \left(\frac{s_1^2 - s_2^2(1-f_2)}{m} \right) \frac{1}{v(\bar{y})}$$

The share of the variance due to the first stage is between 70% and 85% for major crops and even higher for forest, due to its higher spatial auto-correlation. This means that improving the accuracy requires a higher number of PSUs rather than increasing the number of SSUs per PSU.

A different way to compare both approaches is pretending that only 1 of the 10 points in the PSU has been visited. For this approach we have estimated variances with a slightly modified Matern formula [Matern, 1986], better adapted to systematic sampling. Comparing the estimated variances leads to the “equivalent number of points of a PSU”, with values around 4 for main annual crops and around 3 for grassland and forest. This means that 3-4 unclustered points give as much information, in terms of variance, as the current cluster of 10 points.

An additional question is the possible improvement in the efficiency of stratification. Let us assume that LUCAS becomes a set of two co-ordinated surveys: a general land cover survey with a periodicity of 4-5 years and a specific survey for annual crops every year. For the specific survey on annual crops, the efficiency of stratification needs to be assessed. Classifying each point into one stratum may be more efficient than classifying a whole PSU in the same stratum, because a PSU contains points belonging to different categories. A simulation using three types of information for stratification suggests that unclustered points only give a clear advantage from this point of view if the stratification is based on a very accurate photo-interpretation. The question must be studied more in depth with real photo-interpretation data.

References

- Bettio M., Delincé J., Bruyas P., Croi W., Eiden G., (2002), Area frame surveys: aim, principals and operational surveys. *Building Agri-environmental indicators, focussing on the European Area frame Survey LUCAS*. EC report EUR 20521, pp. 12-27. <http://agrienv.jrc.it/publications/ECpubs/agri-ind/>
- Cochran W., (1977), *Sampling Techniques, chapter 10*. New York: John Wiley and Sons,
- Consorzio ITA (2003). *Bollettino finale di statistiche agricole regionali mediante point frame in Italia*.
- Delincé J. (2001), A European approach to area frame survey. *Proceedings of the Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR)*, June 5-7, Vol. 2 pp. 463-472 <http://www.ec-gis.org/>
- Martino L. (2003), The Agrit system for short-term estimates in agriculture: A project for 2004. *Polish Seminar : Information Systems in Agriculture*. Krakow, July 9-11. 2003.
- Matérn B. (1986), *Spatial variation*, Springer Verlag lecture notes in statistics, n. 36, 144 pp