Assessing bias and efficiency of some area frame sampling strategies.

Francisco Javier Gallego
Joint Research Centre of the EC, Ispra, Italy – javier.gallego@jrc.ec.europa.eu

When comparing list frame surveys with area frame surveys, it is usually considered that the main advantage of area frame is its completeness, that results in a good protection against bias. While in general terms it is true that area frames have less risk of bias than list frames, bias can happen for several reasons, for example excluding presumed non-agricultural strata or defining observation units that do not coincide with sampling units.

We present in this paper some examples of area frame management operations that have been used by developing countries (we do not mention which countries to avoid hitting sensitivities). The assessment is run using a pseudo-truth that is as close as possible to real agricultural landscapes. Simulations have been run in several papers using as pseudo-truth classified satellite images or land cover maps. However both alternatives usually introduce a strong behavior modification to the spatial structure of real agricultural landscapes: land cover maps tend to aggregate small patches and substantially increase spatial autocorrelation at short distances. In exchange classified images often have a speckle effect that decreases the spatial autocorrelation at distances corresponding to one or a few pixels. Additionally classified images often tend to over-classify dominant classes. The data set we have chosen is a GIS-based farmers declaration register of the Netherlands. This type of data set has been operationally produced by many developed countries in the last decades and sporadically by some developing countries. Unfortunately most countries are reluctant to make public such data sets and the Netherlands have been and exception by making plot registers free to download, obviously eliminating personal information. An implicit assumption for the applicability of the conclusion to developing countries is that spatial autocorrelation behaves similarly after zooming out to take into account the different average plot size.

In the framework of a sampling frame of square segments, we simulate the definition of two strata (crop-non crop) by simply converting a GIS polygon layer (farm register) into a raster layer with the resolution corresponding to the sampling frame and the standard option of most GIS software. The result is that a significant part of cropland is excluded from the survey. The underestimation of area ranges between 27% and 56% for major crops. However a remedy is suggested that can eliminate most of the bias, even after the elimination from the sample of a stratum that contains a large crop area.

We have also tested the application of a variant of the so-called "open segment" that we call "extended segment" to highlight that it does not coincide with "open segments" as described by FAO handbooks. has been used in at least one developing country. The principle is entirely including in the segment any crop field that touches the sampled segment. The impact depends on the type of estimator used: with an estimator based on the area proportion, an overestimation is observed for major crops between 6% and 10%.

Keywords: Crop area estimation, Stratification, Area frame sampling, Extended segment.