Advances in the use of remote sensing in land use/land cover surveys

Roberto Benedetti*
“G. d’Annunzio” University, Pescara, Italy – benedetti@unich.it

M. Simona Andreano
Universitas Mercatorum, Rome, Italy – s.andreano@unimercatorum.it

Federica Piersimoni
Istat, Rome, Italy – piersimo@istat.it

Paolo Postiglione
“G. d'Annunzio” University, Pescara, Italy – postigli@unich.it

Abstract

Remote sensing is a very important data source in the improvement of the present systems of acquiring and producing agricultural statistics: the remote sensing constitutes a valuable source of information in many phases of the survey process. This paper will analyze the influence of remote sensing technologies on the different phases of sampling spatial statistical units. It is possible to define an *ex-ante* use (i.e., at design level) and/or an *ex-post* use (i.e., at estimation level) of remote sensing information. The remote sensed data might be considered as a covariate that may be used in the definition of the different design and estimation methods. In particular, we will focus on extensions of the classical regression estimator by considering more complex models, and on the use of small area estimation techniques, since this class of methods is more flexible both in the introduction of statistical models and in allowing the use of the covariates at aggregated level and not at unit level (in this case the remote sensing can be viewed as a tool for the areal disaggregation of the estimates). The use of predictive approach to finite population inference (i.e., pure model-based approach) is finally investigated.

**Keywords:** Spatial surveys; auxiliary information; calibration methods; small area estimation.

1. Introduction

Remote sensing is an important tool in the study of natural resources and environment. The possible applications of remotely sensed data are manifold: the identification of potential archaeological sites, the assessment of drought and flood damage, the monitoring and management of land use, the compilation of crop inventories and forecasts, among others. Today, remotely sensed images constitute a basic instrument for monitoring natural resources, the environment, and agriculture. Remote sensing has also become crucial for protecting the global environment, reducing disaster losses, and achieving sustainable development. Furthermore, they provide invaluable information on the state of art of the agricultural sector for both developed and developing countries. As well-stated by Carfagna and Gallego (2005), the spectral response and the identification of crops are not in one-to-one correspondence. Indeed, the radiometric response of the same crop in different conditions can vary across the pixels of an image. A more appropriate approach is to consider the spectral response of a crop as a function of the probability distribution of its spectral reflectance. To solve this problem, the literature has explored some methods for the correct identification of the crops. We refer to this group of techniques as *classification methods*.

There are two broad classes of classification procedure. The first is referred to as unsupervised classification, while the second is defined as supervised classification. In unsupervised classification, an image is segmented into unknown classes. The researcher’s aim is to label these classes at a subsequent stage. Unsupervised classifications seek to group pixels having similar spectral reflective characteristics into distinct clusters. These spectral clusters are then labeled with a certain class name.
Supervised classification uses a set of user-defined spectral signatures to classify an image. The spectral signatures are derived from training areas (or sets) that are created by depicting features of interest on an image. The main difference between the two approaches is that in unsupervised classification, the classes need not be defined a priori. Richards and Jia (2006) provide further detail on this topic. In recent decades, satellite remote sensing technology, in combination with in situ observations, has become an important factor in the enhancement of the current systems for acquiring and generating agricultural data. To gain the benefits of remotely sensed data, managers, consultants, and technicians must be able to understand and interpret the images. Remote sensing techniques are widely used in agriculture and agronomy (Dorigo et al. 2007). Indeed, remotely sensed images provide spatial coverage of a field, and can be used as proxies to measure crop and soil attributes (Fitzgerald et al. 2006). In several developing countries, and over much of the oceans, satellite data is the only source of quantitative information on the state of the atmosphere and of the earth’s surface, and it is an invaluable source of real-time information on severe weather, which is critical for safety in these areas. It is necessary to use remote sensing, since the monitoring of agriculture raises special problems, which are not common to other economic sectors. Indeed, agricultural production depends heavily upon seasonal patterns related to the life cycle of crops.

National and international agricultural policies, and global agricultural organizations dealing with food security issues, depend to a great extent on reliable and timely crop production information. Carfagna and Gallego (2005) provide a first exhaustive description of the different possible uses of remote sensing for agricultural statistics. In particular, remote sensing techniques may represent a suitable tool for particular problems in agricultural survey, such as: the reliability of data, incomplete sample frames and sample sizes, the methods to select units, the measurement of areas, the non-sampling errors, the gaps in geographical coverage, and the unavailability of statistics at a disaggregated level. Remote sensing can be properly used at the design level. Remotely sensed images provide a synopsis of the area under investigation, and are useful for the construction of the spatial reference frame. Furthermore, classified satellite images can be used as auxiliary variables to improve the precision of ground survey estimates, generally with a regression or a calibration estimator. The remotely sensed information could also represent an auxiliary variable in the process of small area estimation. Finally, remote sensing data have been exploited to estimate the production of crops, using their link with the yield. The most common indicators are based on the Normalized Difference Vegetation Index (NDVI, Benedetti and Rossini, 1993; Benedetti et al. 1994) that can be computed through a remotely sensed image. However, as highlighted by Carfagna and Gallego (2005), the link between the NDVI and crop yield is high only for some crops, under certain conditions.

2. Extension of the regression or calibration estimators

Survey statisticians place considerable efforts in the design of their surveys, to be able to use the auxiliary information for producing precise and reliable estimates. The class of calibration estimators is an example of a very general and practical approach to incorporating auxiliary information into the estimation. These estimators are used in most surveys performed by the major National Statistical Institutes (NSIs).

Agricultural surveys are highly specialized with respect to other surveys. These surveys are conducted to gather information on the crop area, crop yield, livestock, and other agricultural resources. Apart from the difficulties typical of business data, such as the quantitative nature of several variables and their high concentration, agricultural surveys are indeed characterized by certain additional peculiarities. In the case of auxiliary information, two specific issues must be discussed (Benedetti et al. 2015). First, the definition of the statistical units is not unique. The list of possible statistical units is extensive, and its choice depends not only on the phenomenon for which the data is being collecting, but also on the availability of a frame of units (unless Indirect Sampling is used; see Lavallée, 2007). Second, a rich set of auxiliary variables, other than dimensional variables, is available: consider, for example, the information provided by airplane or satellite remote sensing. Concerning the first issue, agricultural surveys can be conducted using a list frame or a spatial reference frame. Generally, a list frame is based on an agricultural census, a farm register or administrative data. A spatial reference
frame is defined by a cartographic representation of the territory and by a rule that defines how it is divided into units. According to the available frame, different statistical units are available. Agricultural holdings are the statistical units of a list frame. Surveys based on agricultural holdings are generally cheaper, since it is possible to collect a significant amount of information in a single interview. However, these presume that the list is recent and is of good quality: conditions that are not always satisfied. Points are an example of statistical units of a spatial reference frame. Surveys based on points are often called point frame surveys. In theory, points are dimensionless, but they can be defined as having a certain size, for the sake of coherence with the observation rules, or the location accuracy that can be achieved. Segments are a second type of statistical units of a spatial reference frame. The choice of segment size depends on the landscape. Also, segments can be delimited by physical elements. Two main differences among the statistical units must be highlighted. Information on the positioning (i.e. geo-referencing) of agricultural holdings is not always available, while it is instead always obtainable for points and segments. Geo-referencing is seen as an important source of data, to be complemented with spatial agricultural information such as satellite images, land cover maps or other geo-referenced information layers. As it is usual with business surveys, the population of agricultural holdings is markedly asymmetrical. Usually, asymmetry is positive, as small family-owned holdings coexist with large industrial companies.

The rich set of auxiliary variables in agricultural surveys is mainly available by means of remote sensing data. Remote sensing can significantly contribute to provide a timely and accurate picture of the agricultural sector, because it is extremely suitable for gathering information over large areas with high revisit frequency. Indeed, a large range of satellite sensors regularly provides us with data that covers a broad spectral range. To derive the information sought, a large number of spectral analysis tools have been developed.

An auxiliary variable that is commonly used for crop area estimates is the Land Use/Land Cover (LULC) data. LULC refers to data that is a result of raw satellite data classification into categories based on the return value of the satellite image. LULC data are most commonly presented in a raster or grid data structure, with each cell having a value that corresponds to a certain classification. LULC have been widely applied in estimating crop area. Hung and Fuller (1987) combine data collected by satellite with data collected by means of area survey, to estimate crop areas. Basic survey regression estimation is compared with two methods of transforming the satellite information, prior to regression estimation. González and Cuevas (1993) used thematic maps to estimate crop areas. The estimates were made using regression methods. Pradhan (2001) presents an approach to develop a Geographic Information System (GIS) for crop area estimation that supports a crop forecasting system at a regional level. The overall system combines spatial reference frame sampling and remote sensing. Remote sensing data also provide information on different factors that influence crop yield. The most popular indicator for studying vegetation health and crop production is the NDVI. This is a normalized arithmetic combination of vegetation reflectance in the red and near infrared. Studies have shown that NDVI values present a significant correlation with crop yields. For a comprehensive review of the different ways to use remote sensing for agricultural statistics, see also Gallego (2004). The availability of remote sensing data does not eliminate the need for ground data, since satellite data do not always present the accuracy required. However, this information can be used as auxiliary data to improve the precision of the direct estimates. In this framework, the calibration estimator can improve the efficiency of crop areas and yield estimates for a large geographical area, when classified satellite images and NDVI can, respectively, be used as auxiliary information.

3. Small-area estimators

Regression and calibration estimator are techniques used to improve the precision of a sample estimator. However, these estimators are not sufficiently precise to produce Small Area Estimates (SAE) of surveyed variables, due to the small sample sizes in the small area considered. The literature features several contributions seeking to increase the precision of the SAE.

The term Small Area (SA) generally refers to a small geographical area or a spatial population unit for which reliable statistics of interest cannot be produced, due to certain limitations of the available data.
For example, small areas include small geographical regions such as counties, municipalities or administrative divisions; domains or subpopulations, such as a particular economic activity or a subgroup of individuals obtained by cross-classifying demographic characteristics. SAE is a research topic of great importance, due to the rising demand for reliable small-area statistics even when only very small samples are available for these areas. The problem affecting SAE is twofold. The first issue is how to produce reliable estimates of characteristics of interest for small areas or domains, based on the very small samples that can be taken from these areas. The second issue is how to assess the estimation error of these estimates (Benedetti et al. 2015).

In the context of agriculture, SA usually refers to crop areas and crop yield estimates at the level of the small geographical area. Agricultural statistics are generally obtained through sample surveys, where the sample sizes are chosen to provide reliable estimators for large areas. A limitation of the available data in the target small areas severely affects the precision of estimates obtained from area-specific direct estimators. When auxiliary information is available, the design-based regression estimator is a classical technique used to improve the precision of a direct estimator. This technique has been widely applied to improve the efficiency of crop area estimates, where the auxiliary information used is given by satellite image data. Unfortunately, direct area-specific estimates may not be able to provide adequate precision at the SA level. In other words, they are expected to return undesirable large standard errors due to the small, or even zero, size of the sample in question. Furthermore, when there are no sample observations in some of the relevant small domains, the direct estimators cannot be calculated. To increase the precision of area-specific direct estimators, various types of estimators have been developed that combine both the survey data for the target small areas, and auxiliary information from sources outside the survey, such as data from a recent agricultural census, remote sensing satellite data and administrative records. Such estimators, referred to as indirect estimators, are based on (implicit or explicit) models that provide a link to related small areas by means of auxiliary data, to borrow information from the related small areas and thus increase the effective sample size. Torabi and Rao (2008) derived the model mean squared error of a GREG estimator and two-level model-assisted GREG estimator of a small area mean. They show that, due to the borrowing of strength from related small areas, estimators based on explicit model exhibit significantly better performance compared to the GREG.

As mentioned above, the literature features many contributions on the topic of SAE. In particular, Ghosh and Rao (1994), Rao (2003), and Pfeffermann (2013) have highlighted the main theories upon which the practical use of small area estimators are based. To compare the performance of small-area estimators with the calibration estimators, we must identify the appropriate SAE, considering the different types of agricultural data. SAEs based on linear mixed models may be inefficient when dealing with agricultural data. Two different models can be considered:

i - In the presence of variables with a high portion of values equal to zero and a continuous skewed distribution for the remaining values, we propose using zero-inflated models, as suggested by Chandra and Chambers (2008);

ii - Small domains are often geographical areas. An adequate use of geographic information and spatial modeling can provide more accurate estimates for small area parameters. Several attempts to generalize the Fay–Herriot model (Fay and Herriot, 1979) considering the correlated random area effects between neighboring areas have been performed using a Simultaneously Autoregressive (SAR) process (Petrucci and Salvati, 2006; Pratesi and Salvati, 2009). However, it is common in agricultural and environmental studies for the population to be divided into latent heterogeneous spatially-dependent subgroups of areas, in which the effect of covariates on the variable being studied is stationary. Here, we suggest using a local stationarity approach to estimate the parameters of the Fay–Herriot model (Benedetti et al. 2013). The problem of outliers and missing data, often present in satellite information, is another issue that can be examined.

4. Benchmarking the estimators

Agricultural and rural statistics are essential to inform policies and decisions regarding a variety of important issues, including economic development, food security and environmental sustainability.
Statistical information on land use and rural development can be derived with reference to different observation units, such as households, agricultural holdings, and parcels of land or points. Furthermore, agricultural and rural statistics can be derived through sampling and non-sampling methods. Non-sampling methods mainly include agricultural censuses and the use of administrative data collected for different purposes. Sampling methods can be based upon a list frame or an area frame, or rely on the combined use of different sampling frames. The use of administrative sources in producing agricultural and rural statistics was discussed by Carfagna and Carfagna (2010). The main advantages and drawbacks deriving from using list and area frames in agricultural sample surveys were analyzed by Carfagna and Carfagna (2010), and Cotter et al. (2010).

Sample surveys are usually designed to provide reliable estimates of finite population parameters for large areas. Design-unbiased, or approximately design-unbiased, direct estimates can be derived for these areas, when the sample size is sufficiently large. Auxiliary information from agricultural censuses, administrative sources, or remotely sensed data, can be used before the sample selection – in designing the survey – as well as after the sample selection, during the estimation procedure. The ex-ante use of auxiliary information mainly concerns the construction of optimal sample stratification, and the definition of balanced sampling designs. The auxiliary information can be introduced into the estimation procedure by means of generalized regression estimators (see e.g. Beaumont and Alavi, 2004) or by following the calibration approach (Deville and Särndal, 1992; Särndal, 2007). The use of calibration and regression estimators to combine information from ground data and remotely sensed data in agricultural surveys has been discussed by Gallego et al. (2010). The use of model-based small area estimators raises the question of the robustness of the inference in light of possible model misspecifications. Furthermore, when a reliable direct estimate is available for an aggregate of small areas, the model-based small area estimates must be consistent with the direct estimate for the larger area. This condition is crucial when the direct estimate for the larger area is official endorsed. A potential difficulty with these model-based estimates is that when they are aggregated, the overall estimate for a larger geographical area may be different from the corresponding direct estimate, which is often assumed to be rather reliable. One way to avoid this problem is the so-called benchmarking approach, which consists in modifying these model-based estimates so that, when aggregated, they match the direct estimate for the larger geographical area. A number of benchmarking procedures, intended to ensure consistency between model-based small-area estimates and direct estimates for large areas, have been developed (see e.g. Pfeffermann, 2013). The benchmarking procedures make robust the inference forcing the model-based small-area predictors to agree with the design-based estimator for an aggregate of the areas (Pfeffermann, 2013).

5. Conclusions
In this paper, we have presented a further step in assessing the possibility of using remotely sensed images in sampling design and estimation. Satellite and/or aerial remote sensing technology, in combination with in-situ observations, are a very important tool in enhancing the monitoring system of the earth and, in particular, of agriculture. Remote sensing provides information that is available for several countries at a certain spatial and temporal resolution. We have sought to identify possible solutions to the gaps outlined. Applicability depends essentially on the availability of images, which may not always be satisfied in developing countries. We consider that new satellite technology will provide images that will solve this problem in the very near future.

References


