



**Food and Agriculture Organization
of the United Nations**

Identifying the Most Appropriate Sampling Frame for Specific Landscape Types

**Publication prepared in the framework of the
Global Strategy to improve Agricultural and Rural Statistics**

July 2014

Contents

Acronyms	3
Preface	4
Acknowledgments	6
1. Purpose and scope	7
2. Master frame of agriculture	9
3. Master frame of household surveys	10
4. Appropriate master frames: country examples	13
4.1 Country examples: frame based on an agricultural census	15
4.2 Country examples: frame based on land use maps or satellite images	16
4.3 Summary and recommendations	19
5. Area sampling and data collection	21
6. Modeling survey costs	22
7. Modeling survey errors: relative efficiency of some area sampling strategies	31
7.1 Simple random sampling	33
7.2 Replicated samples by zones	34
7.3 Systematic replicated samples. Point sampling	35
7.4 Point sampling	36
7.5 Twi-stage sampling	37
8. Choosing an area sampling strategy and optimizing the sample design	39
9. Multipurpose sample design	43
10. Concluding remarks	44
Bibliography	45

Acronyms

BAE	Bureau of Agricultural Economics
EA	Enumeration Area
EO	Elementary Observation
FAO	Food and Agriculture Organization of the United Nations
GPS	Geographical Positioning System
PHC	Population and Housing Census
PSU	Primary Sampling Unit
SAC	Scientific Advisory Committee
SSU	Secondary Sampling Unit
UNSD	United Nations Statistics Division
USDA	United States Department of Agriculture
UTM	Universal Transverse Mercator

Preface

This technical paper on **identifying the most appropriate sampling frame for specific landscape types** has been prepared in the framework of the Global Strategy to Improve Agricultural and Rural Statistics. The Global Strategy is an initiative endorsed by the United Nations Statistical Commission in 2010. It provides a framework and blueprint to meet the current and emerging data requirements and demands of policy makers and other data users. Its goal is to contribute to greater food security, reduced food price volatility, higher incomes, environment sustainability, and greater well-being for rural populations through evidence-based policies. The Action Plan of the Global Strategy is centred on three pillars: (1) establishing a minimum set of core data; (2) integrating agriculture into the National Statistical System; and (3) fostering sustainability of the statistical system through governance and statistical capacity building.

The Action Plan of the Global Strategy to Improve Agricultural and Rural Statistics includes an important research programme to address methodological issues for improving the quality of agricultural and rural statistics. The outcome of the research programme is to produce scientifically sound and cost-effective methods that will be used as inputs to prepare practical guidelines for use by country statisticians, training institutions, consultants, etc.

In order to have countries and partners benefit at an early stage from the results of the research activities already available, it has been decided to establish a **Technical Reports Series** to widely disseminate available technical reports and advanced draft guidelines and handbooks. This will also provide an opportunity to receive feedback from countries on the papers.

Technical reports and draft guidelines and handbooks published in this Technical Report Series have been prepared by senior consultants and experts and reviewed by the Scientific Advisory Committee (SAC)¹ of the Global Strategy, the Research Coordinator at the Global Office and other independent senior experts. For some of the research topics, field tests will be organized before final results are included in guidelines and handbooks.

This technical paper on **identifying the most appropriate sampling frame for specific landscape types** is the result of a comprehensive literature review on the subject, followed by a gap analysis and development of innovative methodological proposals for addressing any issues that emerged.

This paper begins by reviewing the main features of master frames identified in a thorough literature review on the topics of Master Sampling Frames of Agriculture and Master Sampling Frames of Households. Accordingly, this paper goes on to define what constitutes an appropriate master sampling frame by providing several examples from countries that utilize a master frame. It also addresses the characteristics and caveats of area sampling and data collection procedures and proposes a methodology for modelling survey costs. Next, this paper proposes a framework for modelling survey errors and derives a method for measuring the relative efficiency of a set of sampling strategies. Strategies for selecting the optimal sample design are defined and a method to identify the most cost-efficient strategy is proposed. Finally, there is a brief overview of optimizing in the context of multipurpose sample design. This paper

¹ The SAC is composed of ten renowned senior experts in various fields relevant to the Research Programme of the Global Strategy. They are selected for a two-year term. The current membership is composed of: *Vijay Bhatia, Seghir Bouzaffour, Ray Chambers, Jacques Delincé, Cristiano Ferraz, Miguel Galmes, Ben Kiregyera, Sarah Nusser, Fred Vogel, Anders Walgreen.*

will serve as a major input into the *Guidelines on developing and maintaining a master sampling frame for integrated agricultural surveys* which is currently under preparation.

The technical papers will be updated as required with the results of in-country field tests and feedback and experiences of the countries. Any additional comments and/or feedback are greatly appreciated and can be forwarded to ESS-Global-Strategy@fao.org.

Acknowledgments

The Technical Paper on identifying the most appropriate sampling frame for specific landscape types was prepared by Luis Ambrosio, Professor of statistics, econometrics, and operational research and Luis Iglesias, Associate Professor of Geomatics both at the Universidad Politécnica de Madrid, with guidance and supervision by Elisabetta Carfagna, FAO, Javier Galego, JRC and Naman Keita, FAO.

Valuable inputs and comments were provided at different stages by the SAC members and by Loredana Di Consiglio, ISTAT. The editing was carried out by Brett Shapiro.

The preparation of this publication was supported by the Trust Fund of the Global Strategy, funded by the UK's Department for International Development (DFID) and the Bill and Melinda Gates Foundation (BMGF).

Purpose and scope

In this report, we focus on identifying a cost-efficient sampling strategy that is integrated with the national statistical system and allows the connection to be made between the economic, social and environmental dimensions of sustainable development. This focus is part of the Global Strategy of the Food and Agriculture Organization of the United Nations (FAO) to improve agricultural and rural statistics [FAO (2011, 2012a)]. In probability sampling, we first need a sampling frame to design a sample. Furthermore, a sampling strategy is needed which requires a mechanism to randomly select observations from the frame to obtain estimates of the survey variables.

The term “integrated” refers to the use of the same sampling frame and related materials in multiple surveys, as well as the same concepts, survey personnel and facilities. For national statistical systems, the importance of planning a programme of integrated surveys, as opposed to ad hoc design of single surveys, has been highlighted by the United Nations Statistics Division (UNSD): the development of a high quality master sampling frame is expensive and the costs could not be justified if the frame were to be used in only one survey [UNSD (1986)].

We begin by reviewing the Master Frame of Agriculture developed at the Statistical Laboratory of Iowa State College in collaboration with the Bureau of Agricultural Economics (BAE) in the United States Department of Agriculture (USDA) and the Bureau of the Census. Next we reviewed the Master Frame of Household Surveys developed by UNSD. FAO (1996, 1998) and UNSD (1986, 2008) developed guidelines based on these master frames to assist countries in planning and implementing agricultural and household surveys, respectively.

The focus of these guidelines is the development and maintenance of master sampling frames. The guidelines give the same overall recommendations: the use of dual sampling frames and the selection of replicated samples. In fact, they focus on the area sampling frames, since an area frame is complete, accurate and up to date. The list frame is a list of farms or households, sampled at a rate of 100 percent, and used to improve the accuracy of the area estimator.

These guidelines warn about the risk of duplication if the overlap between the area frame and the list frame is not taken into account. [Hartley (1962, 1974) was the first to propose dual frames. Fuller and Burmeister (1972) suggested some improvements. Lohr and Rao (2000, 2006) proposed maximum likelihood estimators. Bankier (1986) proposed a different approach, based on stratified multiple samples.]

We define a master sampling frame for integrated surveys by combining the features found in the literature on Master Sampling Frames of Agriculture and on Master Sampling Frames of Households. The frame construction process begins with a map (or satellite image) of the country that is stratified according to the intensity of land use. The last sampling unit is a segment (a small area, called block (of houses) in household surveys). The size of the segment is defined by looking for a suitable number of reporting units within its boundaries: FAO recommends a target segment size between 10 and 15 holdings; UNSD recommends a target block size of 10 households. Two main area frames are considered in the FAO guidelines: in

one case, segments with identifiable physical boundaries are used; and in the other case, the segment boundaries are geometrics. A third area frame uses points as sampling units. [Nusser and Goebel (1997) describe an area frame based on points and used by the USDA for estimating land use and environmental monitoring.]

The recommended number of sampling stages is one in the FAO guidelines and two in the UNSD guidelines. To avoid having to delimit all segments of the surveyed area, FAO (1996) recommends: (i) dividing the territory into "count units" (4-10 segments); (ii) assigning a number to each "count unit" in a serpentine shape to gain a better geographic distribution of the sample; and (iii) assigning to each "count unit" a theoretical number of segments, taking into account the size of the target segment. It is then recommended that a sample of segments be selected with equal probabilities and that they be delimited only in those "count units" having segments in this sample. UNSD (1986, 2008) recommends the use of Enumeration Areas (EAs) of Population and Housing Census (PHC) (target size equal to ten blocks) as Primary Sampling Units (PSUs) and the use of blocks (of houses) as Secondary Sampling Units (SSUs). Often, the SSU is the house instead of the block, and the sample is selected from a list of households elaborated within each PSU included in the sample.

We focus on finding a cost-efficient sampling strategy, within the class of sampling strategies recommended by FAO and UNSD guidelines. An area frame is costly and uses up a significant part of the available budget (lower using geometric boundaries than physical boundaries). We look to optimize the sample design by finding the segment size, the replicate size and the sample size that minimizes the sampling error, given the available budget and taking into account the cost of building an area frame. The sampling error depends on the correlation structure of the survey variable, and we use correlogram models to derive the expected value of the sampling error. The strategy that achieves the minimum sampling error within the set of alternative sampling strategies under consideration is the cost-efficient sampling strategy.

The report is structured as follows. In sections 2 and 3 we describe the features of master frames found in the literature review. In section 4 we define an appropriate master sampling frame by combining these features and describe several country examples that use this master frame. In section 5 we review some particularities of area sampling and data collection procedures. In section 6 we model the survey costs. In section 7 we model the survey errors and derive the relative efficiency of a set of sampling strategies. In section 8 we optimize the sample design and explain how to identify a cost-efficient sampling strategy. In section 9 we address multipurpose sampling strategies. Finally, in section 10 we offer concluding remarks.

Master frame of agriculture

King (1945) credited the idea of a Master Sample to Rensis Likert, from the BAE in the USDA. The idea was to have a large sample from which subsamples of farmers could be selected. In later usage, the term "Master Sample" has come to be applied to the materials used in the creation of the first sample. That is, the term is often applied to the frame rather than to the sample itself. While large-scale area samples had been used in India and Europe in the 1920s and 1930s [King (1945); Stephan (1948)], the Master Sample of Agriculture developed by the BAE represented a major step forward in sampling techniques.

The master frame should be as complete, accurate and current as practicable, and an area frame does match up with these three requirements. Hence, the use of cartographic materials is required to build the area frame and to obtain the measure of size for the sampling units. In this regard, an extensive use of cartographic materials distinguished the BAE approach from previous approaches. Three primary strata were defined on the basis of density of population [King (1945)]: urban areas, peri urban areas and agricultural areas. The sampling units of the BAE Master Sample were relatively small and the units were designed by the statisticians, specifically for the purpose of the sample. An attempt was made to optimize sampling units with respect to travel time and interviewing costs [Jessen (1945)].

Fecso et al. (1986) and Vogel (1995) provide an overview of how the BAE frame was replaced by the USDA in the 1970s by an area sampling frame stratified by land use, using satellite imagery, and the adoption of multiple frame sampling. Nusser and Goebel (1997) describe a second area frame used by the USDA which is designed for land use estimation and environmental monitoring. These two frames differ in that one is based on segments, the other on points. These innovations set standards for area sampling that continue to this day [Fuller (1984)]. The general features of current master sampling frames of agriculture can be found in FAO (1996, 1998). [Details about the properties of estimators using area frames can be found in Faulkenberry and Garoui (1991).]

In UNSD (1986) one can find examples of countries where surveys were conducted comprising agricultural and household surveys, using list frames (Ethiopia 1980-1983). The problem with lists is that frames are usually incomplete, contain duplications and are not updated. Sampling areas is an improvement over other designs, and therefore an area frame should be a basic component of the master frame [Fuller (1984), UNSD (1986, 2005, 2008, 2009)].

Master frame of household surveys

The United Nations Statistics Division has carried out a series of publications designed to assist countries in planning and implementing household surveys in the context of the National Household Survey Capability Program [(UNSD (1986)]. The central topic of this Program is the development and maintenance of Master Sampling Frames and Master Samples of Household Surveys. In many countries a master sample of households is used to conduct surveys on employment, poverty or malnutrition.

The principles that govern the establishment of a Master Sampling Frame of Households, according to UNSD's guidelines are somewhat different from those for sampling frames in general, and particularly from that of agriculture. Some of these principles are that the master frame should: (i) be as complete, accurate and current as practicable; (ii) define PSUs in the frame from area units such as census EAs with mapped, well-delineated boundaries and for which population figures are available; (iii) define Master Sample PSUs that are large enough or numerous enough to sustain many surveys, or repeat survey rounds, during inter-censal period; (iv) use census list of households as the frame at the last stage only if very recent (usually no more than one-year old); (v) use dual or multiple frames with caution by ensuring procedures are in place to address duplications; (vi) employ a system of sample rotation – either households or PSUs – in repeat surveys that use master samples.

To follow these principles, particularly (i) and (ii), extensive use of cartographic material such as maps, aerial photos and satellite images is required, since only area frames assure completeness, accuracy and up-to-datedness of a master frame. However, the usual starting point is not an area frame but a country's population census. Labels differs from one country to another but typically include such terms, in descending order, as: province or county; district; tract; ward; and village. For census purposes, administrative sub-divisions are further classified into EAs [(UNSD (1986)].

Frame units

The census EA is the most appropriate area frame unit because it is often the smallest geographical unit defined. During the preparatory phase of the most recent PHC, Geographical Positioning Systems (GPSs) are used to geo-reference EAs, and digitized EA databases are available in most countries. In most countries EAs are intentionally constructed to contain roughly equal numbers of households – often about 100 – in order to provide comparable workloads for census-takers.

Sampling units: country examples

Depending on the surveys, EA frame units turn out to be smaller or larger than the appropriate sampling unit. The size of the PSUs must be sufficiently large to accommodate multiple surveys without the need to interview the same respondents repeatedly. This is why EAs are often redefined by merging PHC EAs, and these new AEs are used as PSUs.

Country examples can be found in UNSD (2005, 2008), where PSUs were defined by aggregating EAs. In Vietnam, for instance, the Master Sample is based on the 1999 PHC. PSUs were defined as communes, in rural areas, and wards, in urban areas. They were defined in this way because it was decided that a minimum of 300 households would be necessary in each PSU to serve the Master Sample. Alternatively, EAs were considered as PSUs but they were too small and would have had to be combined with adjacent EAs in order to qualify satisfactorily as PSUs.

Each sample PSU contained, on average, 25 EAs in urban areas and 14 in rural areas. For the second stage of selection three EAs were selected in each sample PSU, using probabilities proportional to size. The SSUs were the EAs, which contain an average of approximately 100 households according to the 1999 census – 105 in urban areas and 99 in rural areas. For survey applications, a third stage of selection is administered in which a fixed number of households is selected from each sample EA. The number may vary by survey and by urban-rural. For example, 20 households per EA might be chosen for rural EAs and 10 per EA for urban ones.

Mozambique is another country example. Master Sample PSUs were constructed from the 1997 PHC. They consist of geographical groupings of generally 3-7 census EAs, the latter of which contain about 100 households. The second stage of selection was a sub-sample of the PSUs. At the third stage, a sample of one EA was selected in each of the PSUs. The EAs were selected with equal probability because their sizes are roughly the same. The final stage of selection occurred following field work in which a fresh list of households was compiled to bring the 1997 sample frame up to date. From the list, a systematic sample of 20 households in rural areas and 25 in urban was selected for interviews. In new Agricultural Census methodology, the number of sampling stages is reduced to two, in which PSUs are the new PHC EAs redefined and the last sampling unit is the household.

Finally, Cambodia's National Institute of Statistics developed a Master Sample in 1999 from its 1997 PHC. It was decided to use villages as the PSUs because they are large enough (on average, 245 households in urban areas and 155 in rural) to have enough households to accommodate several surveys during the inter-censal period. The alternative of using census EAs was considered but discarded because they are only half the size of villages, on average.

Within each selected Master Sample PSU, blocks of size 10 households (on average) were formed. The number of blocks created within each PSU was computed as the number of census households divided by ten and rounded to the nearest integer. The typical PSU contains about 18-30 blocks, providing an ample number of blocks in each PSU to sustain all surveys. A limitation of the Master Sample of Households design is the use of compact clusters (all the households in the sample block are adjacent to each other): this increases the design effect; to reduce the design effect, EAs are portioned into segments of only size 10 households. Alternatively, a list of all households in the EA is first made and ten are randomly selected with equal probability (sometimes systematic selection is used).

Appropriate master frames: country examples

In this section we define an appropriate master sampling frame by combining the features found in the literature on Master Sampling Frames of Agriculture (reviewed in section 2) and on Master Sampling Frames of Households (reviewed in section 3). The most appropriate master frame is the most cost-efficient, and in section 8 we develop a quantitative approach to identify the most cost-efficient sampling strategy.

The appropriate master sampling frame to design integrated surveys of a target population is a multiple frame, made up of area frames and list frames. The required basic materials are: cartography for the construction of the area frames, including aerial photography and satellite imagery; and censuses for the construction of the list frames. Additional lists are required to improve the estimation of specific characteristics. To support the appropriateness of this frame we describe several country examples that use this master frame.

Primary stratification

A primary stratification consists in dividing the country's territory into three strata: urban areas; peri-urban areas; and rural areas. The limits between these three strata have to appear on the cartographical material.

Frame unit

A sampling frame is made up of frame units. If an agricultural or population census is available, then the frame unit is the EA, as defined in the censuses, where the objective in establishing EAs size is to limit and more or less equalize the workloads of individual census enumerators: this size is, on average, 100 households for urban areas and 100 holdings for rural areas. If neither a population census nor an agricultural census is available, then an area of land having the features of EAs is defined. The EA limits have to appear on the cartographic material. In this way, they are the link between area frames (based on cartographic material) and list frames (based on censuses).

Secondary stratification

The frame units are stratified in a number of strata, using land use intensity as the criterion. According to this criterion, frame units from urban areas (EAs) are classified in the first stratum and frame units from rural areas are stratified in a number of strata, according to the percentage of cultivated land. If census data are not available, then satellite imagery is usually available as a data source on land cover and land use, and is a good basis for integrating agricultural surveys and household surveys with environmental issues.

A Geographical Information System is a useful tool for land use stratification using satellite data, as well as for geo-referencing the administrative units of a country and the EAs [Iglesias (2013)]. Once the EAs have been geo-referenced, both the holdings registered in the agricultural census and the agricultural households and non-agricultural households registered in the population census can be linked with each other and with land use data. [A detailed discussion on linking Households with Agricultural Households and Agricultural Holdings can be found in FAO (2012b).]

Sampling unit size

In the conventional Master Sample of Households, two stages are considered: in the first stage, the PSU is an EA or a cluster of EAs, and the last stage sampling unit is either a small area called “block” (of houses) or directly a household. The recommended number of sampling stages in FAO (1996, 1998) guidelines is one.

The size of the PSU (either an EA or a counting area) and the last sampling unit (either a segment or a block) is the result of an attempt to strike a suitable balance between the smaller area, which is usually statistically more efficient, and the larger area, which is less costly to enumerate. The last stage sampling unit size is fixed at 10 households per block (in urban areas) or 10 holdings per segment (in rural areas), on average. In household surveys, sometimes the last stage sampling unit is not the block but the household.

Sample selection

Usually, a replicated sample is selected [FAO (1996), UNSD (1986, 2005 and 2008)]. Either a zone sampling or a systematic sampling scheme is used to select the sample within each stratum. To maximize geographic distribution of the sampling units, PSUs are numbered in serpentine fashion. PSUs included in the sample are subdivided into the target number of segments or blocks, and a sample of them is selected.

Points sampling

Random selection of a sample of points on a map, via an aerial photography or a satellite image, is especially easy since it does not require segments. Usually the points are on a grid that is overlaid on the map (aerial photography or satellite image) so that it is a systematic sample [Gallego et al (1994); FAO (1998), Chap. 8]. A sample of points is free of coverage bias. The area frame construction requires only maps, in contrast to segment sampling, which also requires satellite images or aerial photographs.

Integrating agricultural surveys and household surveys

A dual frame composed of the master frame of the agriculture, *A* (described in section 2), and the master frame of household surveys, *B* (described in section 3), is an integrated master frame: *A* and *B* use the same materials and the same concepts, and the multiple surveys that can be designed from this frame can be carried out by the same survey personnel, using the same facilities.

To design a master sample that integrates agricultural surveys and household surveys, we consider two domains: (a) agricultural households (which are the households in *B* related with holdings in *A*); and (b) non-agricultural households (which are households in *B* without any

relationship with holdings in A). A third domain with holdings in A without any relationship with households in B could be considered additionally if there is not enough information to link the two sources.

Let N_B be the number of households in B and let N_{Ba} and N_{Bb} be the number of households in the domains (a) and (b), respectively, $N_B = N_{Ba} + N_{Bb}$. Let Y_{Ba} and Y_{Bb} be the total values of the survey variable in the domains (a) and (b), respectively, and let $Y_B = Y_{Ba} + Y_{Bb}$ be the total value of the survey variable in B . We want to estimate Y_{Ba} , Y_{Bb} and Y_B .

The master sample integrates a sample selected from A and a sample selected from B . Let m_A and m_B be the number of sampling units in these two samples. Let n_A be the number of holdings in m_A and let n_{ABa} be the number of agricultural households in B related with at least one holding in n_A . Let n_B be the number of households in m_B and let n_{Bb} be the number of households in n_B without any relationship with holdings in A .

We estimate Y_{Ba} using data in n_{ABa} and using the expansion factors corresponding to m_A . We estimate Y_{Bb} using data in n_{Bb} and using the expansion factors corresponding to m_B . We estimate Y_B as the sum of Y_{Ba} and Y_{Bb} estimates and we estimate the variance of this estimator as the sum of the variances of the estimators of Y_{Ba} and of Y_{Bb} .

4.1 Country examples: master frame based on an agricultural census

a) Chile

In Chile, an Agricultural Census from 2007 was available and we used it to build an area frame [Ambrosio (2012)]. The total area of Chile was classified into two primary strata: (i) urban areas (such as city, town or village); and (ii) non-urban areas (the remaining areas). The areas in stratum (i) were delimited using the digital limits available from the 2011 PHC. The non-urban areas were stratified into three secondary strata, according to land use intensity: agriculture (cultivated land); livestock (meadow and extensive grasslands); and forest and bushes.

Stratification

Satellite imagery was considered as a basis for land use stratification but was discarded because the Chilean landscape is a mosaic of agricultural fields dispersed between non-agricultural areas. Therefore finding sufficient identifiable boundaries for small areas (segments) on the images is a problem. Instead, the agricultural and livestock areas were stratified using the 2007 Agricultural Census [INE (2007)]. In this census the frame units are the EAs and their limits are available in digital form. Land use data aggregated at the EA level are available and were used to classify EAs into strata, using multivariate statistical methods. A similar approach can be found in Abaye (2010).

Sampling units

Giving consideration to the problem of finding sufficient identifiable boundaries for sampling units on the maps and satellite images, it was decided to divide the EAs into segments of geometrical limits, instead of identifiable boundaries. A square grid of sides of 500 metres was superimposed on the stratified EAs to define segments of size 25 hectares. Each sampling unit (segment) is assigned to the strata where the highest part of its area lies, and the strata limits were adjusted to those of square segments.

Multiple frames

This is an adequate Master Sample to estimate areas under the main crops, using the closed segment estimator. It is also a good basis for selecting a sample of holdings, and particularly livestock holdings to estimate livestock farming production using weighted segment estimators [FAO (1996)]. To improve the precision of the area estimator of very specific crops (industrial crops and vegetable crops) and livestock farming, we used a list of outlier holdings sampled at a 100 percent rate.

b) Uruguay

In Uruguay, the EAs of the Census of Agriculture are used as PSUs and the list of holdings inside the selected PSUs are updated by sweeping the PSU in the first-stage sample before selecting a new second-stage sample [MGAP (2011)]. The cartographic material used in the PHC was used in the last (2011) Agricultural Census in an attempt to integrate both censuses.

c) Fiji

In Fiji (2009), EAs from the 2007 PHC were used for primary stratification into urban, peri-urban and agricultural strata. A secondary stratification of the agricultural areas was carried out using the intensity of cultivated land in the EAs as the criterion. The last sampling unit is the segment of physical boundaries, and the segment target size was 100 hectares (one square kilometre grid).

d) A master frame based on a recent population census with several questions on agriculture

FAO (2012b) proposed adding several questions on agriculture in the questionnaire for the population census. A list frame is the list of holdings identified on the basis of these specific agricultural questions included in the census questionnaire. This approach is promising, particularly because the cost of frame construction is low and it could be relatively more cost-efficient than other frames from the same class. However, more work is needed to assess the coverage problem [Carfagna et al (2013)].

4.2 Country examples: frame based on land use maps or satellite images

If EA maps are available from PHC, they could be used to build an area frame. Otherwise, land use maps and satellite images could be used [Gallego (1995)]. Using satellite images, the starting point is to classify the image by land use categories [Iglesias (2013)]. Then, the sampling units need to be referred to satellite imagery. In this section we describe a country

example (Guatemala) where the frame is based on a map of land use, and another example (Perú) where the frame is based on satellite images.

a) Guatemala

In Guatemala, we decided to build a new master sampling frame of agriculture based on a map of land use [Ambrosio (2013a)], since the last agricultural census is dated 2003 and there is no plan to update soon.

Stratification

We stratify the country using a map of land cover and land use dated 2005. More recent satellite imagery was considered but it was discarded because the Directorate General of Strategic Geographical Information and Risk Management of the Ministry of Agriculture, Livestock and Food is updating the map of land use. The total area of Guatemala was classified into two primary strata: (i) non-agricultural areas (including cities, towns or villages, as well as permanent water areas, forest areas protected by the law and other non-agricultural land); and (ii) agricultural areas (the remaining areas in the country, including peri-urban areas or *traspatio*).

Agricultural areas were stratified into four strata, according to land use intensity [FAO (1996)]. Areas where the percentage of cultivated land is more than 60 percent were divided into two strata according to field size: stratum A is big fields and stratum B is small fields. Stratum C is defined by areas where the percentage of cultivated land is between 20 and 60. The areas where there is less than 20 percent of cultivated land are classified as stratum D. A specific stratum for vegetable crops, coded as E, was initially considered but was integrated in stratum B.

Sampling units

A target segment size was defined for each stratum, as a function of the average field size in the stratum. In order to keep non-sampling errors within tolerable limits, it was considered that the average number of fields per segment should be between 15 and 25 [Taylor et al (1997)]. In stratum A, the average field size is estimated to be 1.25 hectares; as a result, the target segment size is 25 hectares $[((15+25)/2) \times 1.25=25]$. In the same way, target segment sizes for the remaining strata were defined as: 6.25 hectares for stratum B; 50 hectares for stratum C; and 100 hectares for stratum D.

Initially, it was decided to use segments with physical boundaries in the specific stratum for vegetable crops and segments with geometrical boundaries in the remaining strata. Finally, segments with geometrical boundaries were used in all the strata, due to budget and calendar reasons. The target segment size in the specific stratum for vegetable crops was 6.25 hectares.

Land stratification is carried out using the digitized land cover and land use map. The starting point is to compute the percentage of cultivated land in a grid of 1000 metres per side: 100 hectares, which is the target segment size in stratum, D. Then, each square of the grid map is classified in a stratum according to its percentage of cultivated land. The squares of the grid in the stratum of more than 60 percent of cultivated land are sub-stratified according to the average field size in the stratum. The squares of the grid in stratum A are divided into four squares of size 25 hectares each (which is the target segment size in this stratum). The squares of the grid

in the stratum B are divided into 16 squares, each of size 6.25 hectares. The squares of side 1000 meters that is classified in stratum C are divided into two equal parts of 50 hectares (which is the target segment size in this stratum).

The total vegetable crops area in the land use map is partitioned into segments of 6.25 hectares.

Multiple frames

The area frame is used to estimate areas under the main crops, using the closed segment estimator [FAO (1996)]. It is also used to select a sample of holdings, and particularly livestock holdings to estimate livestock farming production using weighted segment estimators. To improve the precision of the area estimator of very specific crops (those where the coefficient of variation of the area estimator is higher than 15 percent) and livestock farming, we use a list of outlier holdings sampled at a 100 percent rate.

b) Costa Region (Perú)

In 2010, when the last agricultural census [CENAGRO (2013)] was not yet available, an area frame was built in the Costa Region of Perú. The country is divided into four Natural Regions (Costa, Sierra, Selva Alta and Selva Baja), according to their orientation with respect to the Andes mountain chain and the altitude. Google Earth satellite imagery was used to stratify the Costa Region into an urban areas stratum, a peri-urban stratum, and three agricultural strata, using the intensity of cultivated land as the criterion.

The last sampling unit is the segment with physical boundaries and the segment target size is 10 parcels. The surface of the segment is a function of the average field size in the stratum. A systematic replicated sample of segments is selected. Where it is not possible to directly delimit segments of the target size (due to lack of physical boundaries) “counting areas” are used [OEEE (2010)].

c) Nicaragua

In Nicaragua, the current survey sample for agricultural statistics is based on a sample of points [FAO (1998), Chapter 8] designed during 1994-1997, using topographic maps at 1:50,000. Using satellite images, an area frame with identifiable physical boundaries was built in 2002-2003. This area frame is used to select a sample of segments, but only in a strip of land parallel to the Pacific coast, for which ortophotomaps are available. Agricultural authorities had a plan to improve agricultural statistics, which was designed in collaboration with FAO [FAO-FIDA (2013)]. In the framework of this plan, we built an area frame with segments of geometric boundaries, using the recent agricultural census [INIDE (2011 a,b)], the non-recent topographic map, and the 2002-2003 area frame with identifiable physical boundaries [Ambrosio (2013b)].

Stratification

Using land use data aggregated at the EA level from the last census, the agricultural and livestock areas were stratified according to land use intensity: areas where the percentage of cultivated land is more than 60 percent were classified in stratum A. Stratum B is defined by areas where the percentage of cultivated land is between 20 and 60 percent. The areas with less than 20 percent of cultivated land and more than 60 percent of extensive grasslands are

classified as stratum C. The areas with less than 20 percent of cultivated land and less than 60 percent of extensive grasslands are classified as stratum D.

Sampling units

Segments with geometrical boundaries were used in all the strata, since neither ortophoto maps nor high-resolution satellite imagery are available. A target segment size was defined for each stratum as a function of the average field size in the stratum, which is known from the last census. In order to keep non-sampling errors within tolerable limits, it was considered that the average number of fields per segment should be between 15 and 25 [Taylor et al (1997)]. In stratum A, the target segment size is 6 hectares, 8 hectares for stratum B, 25 hectares for stratum C and 50 hectares for stratum D.

Multiple frames

This area frame was used to select a sample of segments in order to estimate areas under the main crops, using the closed segment estimator. It was also used to select a sample of livestock holdings in order to estimate livestock farming production using weighted segment estimators. To improve the precision of the area estimator of very specific crops (industrial crops and vegetable crops) and livestock farming, we used a list of outlier holdings sampled at a 100 percent rate.

Data collection using satellite images

Ortophotomaps were not available and it was decided to use satellite images for data collection on the ground. Google Earth images were accessible on line, and the resolution level was adequate for the segment sizes of 25 hectares and 50 hectares. For the segments of 6 and 8 hectares, there were some difficulties but after a test carried out on the ground it was possible to delineate field boundaries on a paper copy of the Google Earth image, as well as to obtain an objective measure of area fields, for all segment sizes, including the small ones.

4.3 Recommendations

We do not recommend a sampling strategy of low quality (low coverage), even if it is a low-cost strategy. Accordingly, we advise against strategies based on a list frame of low coverage or a non-recent census: they are not appropriated because they induce a high bias that will be difficult to estimate. Moreover, we agree with the authors of FAO (1996), p. 18 that a sample from a frame having severe coverage problems cannot be assumed to be a probability sample and should be considered as a subjective sample. We recommend strategies based on list frames built from recent agricultural censuses with good coverage, and those based on an area frame as well as those based on dual or multiple frames that integrate an area frame, including strategies based on and EA Agricultural Census or PHC.

The class of appropriate sampling strategies is large, and choosing one among them is a difficult task, since there is a conflict between the cost and the efficiency of a given sampling strategy. An area frame of segments is costly and consumes a significant part of the available budget: it is less costly to use geometric boundaries than physical boundaries. The fixed cost of sampling points is lower than sampling segments, since using points it is not required to delimit segments. On the contrary, the variable costs are higher using points than segments, since a segment is a

cluster of elementary observation (EO) units (holdings, tracts or households) and the travel costs are lower than sampling points, where the sample of EOs is geographically more dispersed. Using segments (clusters of EOs) is usually less efficient than using points (EOs). In sampling points it is usual to select a systematic sample with only one random start (a grid), and this is more efficient than a systematic replicated sample (more than one random start) [(Ambrosio et al (2003)], but it is less useful in repeated surveys.

To resolve the conflict between cost and efficiency when choosing a sampling strategy, in section 8 we develop a cost-efficiency approach, which allows the sample design to be optimized. To choose a sampling strategy, estimates of the fixed and variable survey costs are required: in section 6 we give some figures regarding the fixed costs, but additional data are required, including data on variable costs. In addition, it is necessary to assess the correlation structure of the survey variables in order to calculate the expected value of the sampling variance. In section 7 we take up this issue.

Area sampling and data collection

In area sampling, the sampling unit and the reporting unit are different. The sampling unit is generally a small land area (it could be a point) called a segment. In agricultural surveys, the land area of a holding inside a segment is called a tract. The reporting unit can be a holding or a tract in agricultural surveys, and a household in household surveys. Several procedures for data collection can be considered. Some survey variables, such as crop acreage, crop yields, soil degradation or water pollution, can be observed directly on the ground using appropriate instruments (for instance, hand-held GPS or computer digitized instruments together with maps, ortophotomaps or satellite images, can be used to measure crop acreage in the tracts). Most of the household data and many agricultural data, such as inputs, cannot be observed directly on the ground and need to be collected by interview.

Different data collection procedures contribute to the total survey errors and survey costs. Measurement instruments are expected to be unbiased and contribute only to the variable error component of the total survey error. In an interview, there is the interviewer, the respondent (the holder for an agricultural survey and the householder for a household survey), the questionnaire and the mode of data collection (face to face, telephone or mail): there is a risk of respondent bias [Carfagna et al. (2013)] and a risk of interviewer effect over the sampling error.

Estimation methods require referring the collected data to sampling units. Three methods have been proposed in the literature to associate segments with reporting units: closed, weighted and open segment methods. In agricultural surveys, the former is used when the reporting unit is the tract and the latter when the reporting unit is the holding. For the closed segment method, the value of the survey variable in the segment is the sum of the values in each of the tracts of the segment. For the weighted method, the survey variable value in a tract is the survey variable in the holding multiplied by the area of the holding that is within the segment and divided by the total area of the holding [FAO (1996)]. The open segment method, associates a segment to all holdings, with headquarters included in the segment [FAO (1996)]. This last method implies special difficulties in area sampling, and the aforementioned guidelines do not recommend its use [see Ford et al (1986) and Nealon (1984)].

Hereon we use the term Elementary Observation (EO) to refer the survey variable value in a tract, for agricultural surveys; and in a household within a block, for the household surveys. We measure the segment size and the block size by the number of EOs inside. In addition, we assume that the sampling units are stratified in a number L of strata, and we use h to identify the generic stratum ($h=1,2,\dots,L$). Y_{hij} denotes the j^{th} EO within the i^{th} segment from the stratum h .

Modelling survey costs

Survey costs depend on a large number of factors that may vary across countries and across surveys within countries [Yansaneh (2005)]. However, a careful analysis of cost components can reveal common cost structures across groups of countries or surveys [Hansen et al (1953), chapter 6, sects 10-15, Groves (1989, chapter 2), Cochran (1977, sects 5.5 and 11.13-11.14)]. To begin, we classify the total costs of conducting a survey in two main categories: fixed and variable.

Fixed costs are those independent of the sample size, including costs for: frame construction; survey planning; development of the survey design; preparatory work; survey management; coding and tabulating; data processing; and editing results [Yansaneh (2005)]. Variable costs are those depending on the sample size and consist of the cost of: selecting the sample; obtaining maps and other material for locating sampling units; locating sampling units; travel by interviewers and supervisors; interviews, including training of interviewers; and supervisors.

Fixed costs: frame-building

Frame-building costs differ according to the kind of boundaries between strata and between sampling units: physical or geometric boundaries are the two main types of boundaries to consider [FAO (1996)]. To guide the search for survey cost estimation, we consider the estimates found in the literature.

Physical boundaries

The estimated time and cost figures to build an area frame can vary widely from one US state to the next, as pointed out by Cotter and Nealon (1989). Factors to be considered include the size of the state, the number of counties, the availability of good boundaries, the type of agricultural activity involved, and the type of materials to be used. Two main procedures for building an area frame with physical boundaries can be considered: those using cartographic material (maps and satellite imagery) on paper and those with electronic support.

When the frame is constructed and maintained on paper, the cost of labour far outweighs the cost of the materials, since many hours are required for the delineation of strata and PSUs on several different media and additional hours are required for reviews. When the frame is constructed using a computer procedure (CASS: Computed Aided Stratification and Sampling Systems), labour costs can be reduced notably [Cotter and Tomczak (1994)].

Table 1: Frame-building cost structure (physical boundaries)

State	Procedure	Labour (total hours)	Cost structure (%)	
			Labour	Material
Alabama	Paper	11460	85	15
Arkansas	Paper	10193	79	21
Georgia	Paper	14927	81	19
Louisiana	Paper	10050	81	19
Michigan	Paper	10459	76	24
Oklahoma	CASS	4459	42	58

A more detailed costs structure is shown in Table 2 for an average US state [Cotter and Nealon (1989)].

Table 2: Costs (*) of frame building for an average US state (physical boundaries)

Concept	Description	Cost structure (%)
Labour	Cartographic:	
	Stratification	37.3
	Digitization	16.0
	Statisticians	16.0
	Systems Analyst	2.7
	Administrative and Secretarial	3.3
	Total labour	75.3
Materials	Satellite imagery	10.0
	Aerial photography	6.7
	Transfer maps	0.7
	Scaled overlays	0.6
	Data processing	1.3
	Total materials	19.3
Miscellaneous	SSO support and travel	1.4
	Computer and equipment maintenance	4.0
	Total miscellaneous	5.4
Total costs		100.0

(*) Using maps and satellite imagery on paper

It is clear from Table 2 that labour (75%) and materials (19%) are the two main cost components, and that satellite imagery (10%) is the main material cost. Using data from Cotter and Tomczak (1994), Table 3 shows the labour required per square kilometre to build an area frame with physical boundaries.

Table 3 Number of hours of labour required per square kilometre to build an area frame in several US states (physical boundaries)

State	Hours	
	Total	Per km ²
Alabama	11460	0.08
Arkansas	10193	0.07
Georgia	14927	0.08
Oklahoma(*)	4459	0.025
Louisiana	10050	0.074
Michigan	10459	0.078

(*) Using CASS

Table 4 shows the current prices of satellite imagery.

Table 4a: Satellite imagery prices: COSMO-SkyMed (standard products)

Product	Scene Size (km ²)	Resolution (m ²)	Archive ^(*) (€)	€/Km ²
Spotlight-2	7X7	1X1	3075	62.76
Spotlight-2	10X10	1X1	4725	47.25
Stripmap Himage	40X40	5X5	1800	1.13
Stripmap PingPong	30X30	20X20	960	1.07

(*)New price=2 times archive price Source: e-geos. (list prices, June 2013)

Table 4b: Satellite imagery prices: SPOT-Scene

Product	Scene Size (km ²)	Resolution (m ²)	New (€)	€/Km ²
Spot level 1A,1B, 2A	60X60	2.5X2.5	5400	1.50
SpotView Ortho-Level 3	30X30	2.5X2.5	6000	1.95
Spot level 1A,1B, 2A	60X60	5X5	2700	0.75
SpotView Ortho-Level 3	30X30	5X5	3300	1.07
Spot level 1A,1B, 2A	60X60	10X10	1900	0.53
SpotView Ortho-Level 3	30X30	10X10	2500	0.81

Source: e-geos. (list prices, June 2013)

Table 4c: Satellite imagery for free: LANDSAT-Scene

Product	Scene Size (km ²)	Resolution (m ²)	€/Km ²
L4-5 TM multi-spectral	185X185	30X30	Free
L7 ETM+ multi-spectral		30X30	Free
L1-4 MSS multi-spectral		60X60	Free
L7 ETM+ thermal		60X60	Free
L4-5 TM thermal		120X120	Free

Source: www.landcover.org

The Maps prices are SPOTMaps: UTMWGS 84: 2.0€/Km² (Source: e-geos. List prices. June 2013).

Table 5 shows the average of data on Tables 3 and 4.

Table 5: Average labour and satellite imagery costs per square kilometre (physical boundaries)

Labour	0.076 h/km ² (*)
Materials: Satellite imagery. Resolution and unitary cost	
1X1	55.00 €/km ²
2.5X2.5	1.73 €/km ²
5X5	0.98 €/km ²
10X10	0.67 €/km ²
20X20	1.07 €/km ²
30X30	Free

60X60	Free
120X120	Free

(*) Using maps and satellite imagery on paper

Table 6 shows the total labour and cartography costs per square kilometre, as a function of hourly wage in the country.

Table 6: Average labour costs (as a function of hourly wage) and satellite imagery costs per square kilometre to build an area frame (physical boundaries)

Labour	$0.076 \text{ h/km}^2 (*) w^{(**)} \text{ €/h}$	$0.076 w^{(**)} \text{ €/km}^2$
Materials: Satellite imagery. Resolution. Unitary cost		
30X30	Free	Free
60X60	Free	Free
120X120	Free	Free
Total labour and cartography costs per km^2		$0.076 w^{(**)} \text{ €/km}^2$

(*) Using maps and satellite imagery on paper. (**) w is the hourly wage

Since satellite imagery is free of charge, material costs are negligible and, as a result, only miscellaneous costs need to be added to the labour costs to get an estimate of total costs of building an area frame.

Geometric boundaries

When the frame is constructed using geometric boundaries, the cost of labour is reduced with respect to area frames using physical boundaries, since not many hours are required to delineate strata and PSUs on several different media. The material costs using geometrical boundaries are similar or even lower than using physical boundaries, since high-resolution cartography is not required.

We estimate that the labour costs of building an area frame using geometric boundaries could be reduced to the costs of stratification, which is 37 percent of the total costs of building an area frame with physical boundaries, according to Table 2. These costs can be estimated as a function of hourly wage, using Table 5: $0.37 \times 0.076 \times w \text{ €/km}^2 = 0.028 w \text{ €/km}^2$, where “w” is the hourly wage.

Material costs are negligible since satellite imagery is free of charge and, as a result, only miscellaneous costs have to be added to the labour costs to get an estimate of total costs of building an area frame with geometric boundaries.

Other fixed costs

Fixed costs of conducting the survey include (in addition to building frame costs) those for: survey planning; development of the survey design; preparatory work; survey management; and data processing, editing results and preparing reports [Yansaneh (2005)]. Table 7 shows these costs in Spain.

Table 7: Other fixed costs

Concept	Description	Costs (€/km ²)
Labour and materials	Preparatory work	0.02
	Data base elaboration and estimate computations	0.09
	Editing results and preparing reports	0.08
	TOTAL	0.19

Source: Spanish Ministry of Agriculture, Food, and Environment. 2011

Variable costs: selecting and preparing sampling unit costs

Physical boundaries

Cotter and Nealon (1989) provide estimates of the cost structure for selecting and preparing a segment, as shown in Table 8.

Table 8: Cost structure for selecting and preparing a segment (physical boundaries)

Concept	Description	Cost structure (%)
Labour	Location of selected PSU on frame map, transfer of PSU boundaries to aerial photograph, delineation of segments, selection of sample segment, transfer of segment to county map, identification of photography needed, ordering photography, and quality assurance reviews.	31.6
	Transfer of boundaries to photo enlargement, labeling the photo, make photocopy, digitizing the segment, tape periphery of photo, quality assurance reviews, and mailing to state office.	12.7
	Supervision: Cartographic Statisticians	10.4 7.8
	Technical Support: Statisticians	2.6
	Systems Analyst	2.2
	Office Support: Administrative/Secretarial	2.0
	Annual & Sick Leave (about 15 percent of labour)	10.4
	TOTAL LABOR	80.7
	Materials	17"x17" aerial photography

	12"x12" copy of photograph	3.9
	Office supplies, computer expenses, mailing costs, microfiche, etc.	2.4
	TOTAL MATERIALS	19.3
	TOTAL LABOUR AND MATERIALS	100.0

It is clear from Table 8 that labour (81%) is the main cost component: the authors estimate in 5h/segment the time required for the first task included in Table 8 (location of selected PSU...) and in 2h/segment the time required for the second task (transfer of boundaries...). These two tasks account for more than 44 percent of the total costs. Aerial photography is the main material costs (13%).

Table 9 shows the total labour and material costs per segment, as a function of hourly wage in the country.

Table 9: Average labour costs and satellite imagery costs for selecting and preparing a segment (physical boundaries)

Labour	7 (h/segment)w ^(*) €/h
Materials: Satellite imagery. Resolution. Unitary costs	
1X1	55.00 €/km ²
2.5X2.5	1.73 €/km ²
5X5	0.98 €/km ²
10X10	0.67 €/km ²
20X20	1.07 €/km ²
30X30	Free

(*) w is the hourly wage

Geometric boundaries

Using segments with geometric boundaries, labour costs are reduced with respect to segments with physical boundaries, since tasks such as transfer of boundaries or delineation of segments, which represent more than 40 percent of total costs, are not required.

Table 10 shows the cost for selecting and preparing a segment with geometric boundaries in Spain.

Table 10: Cost for selecting and preparing a segment with geometrical boundaries

Concept	Description	Cost (€/segment)
Labour	Location of selected segment on frame map, make a copy of frame map and quality assurance review	1.50
	Photo enlargement, labeling the photo, make photocopy and quality assurance reviews	1.70
	TOTAL LABOUR	3.20
Materials	Aerial ortophotomaps	Free
	Copy of aerial ortophotomaps	0.30
	Fiches for data collection	0.50
	Office supplies, computer expenses, mailing costs, microfiche, etc.	0.70
	TOTAL MATERIALS	1.50
TOTAL LABOUR AND MATERIALS		4.70

Source: Spanish Ministry of Agriculture, Food, and Environment. 2011

It is clear from Table 10 that labour is the main cost component, mainly due to the fact that aerial ortophotomaps are available for free. Satellite images of resolution 5X5 or lower are generally enough to support data collection on the ground, so that the cost per segment using satellite imagery should be increased in about 1€/segment (see Table 9), reaching a total cost of 5.70 €/segment.

Modelling survey costs

Travel costs are usually modelled as a disaggregated item. For a one-stage sample of size n , we consider the cost model, $C = C_0 + C_1n + C_T\sqrt{n}$, where C_0 is fixed cost, C_1 is the cost of adding a segment to the sample, excluding travel costs, and $C_T\sqrt{n}$ is the cost of travel between segments in the sample. The expected distance that is necessary to cover to visit n sampling points chosen at random with equal probabilities is $d = \sqrt{nA}$, where A is the survey area [Jessen (1978), Sect 4.7]. Travel cost per kilometre is modelled as $(c_T + t_Tw)$, where c_T is the cost of transport by kilometre, t_T is the time (in hours) required to travel a kilometre, and w is wage per hour of interviewers and supervisors. We model the total travel cost, $(c_T + t_Tw)d$ as a function of the sample size, $(c_T + t_Tw)d = C_T\sqrt{n}$, where $C_T = (c_T + t_Tw)\sqrt{A}$.

For a two-stage sample of n_1 PSUs and n_2 segments per PSU, Humphreys (1974) considers the cost model $C = C_0 + C_1n_1 + C_2n_1n_2 + C_{T_1}\sqrt{n_1} + C_{T_2}n_1\sqrt{n_2}$, where C_1 is the average costs for adding a PSU to the sample, excluding the between-PSUs travel cost but including positioning travel cost (travel to the first PSU visited from the interviewer home base and then back to the home base from the last PSU visited during the data-collection trip), and C_2 is the cost of including an extra segment per PSU in the sample, including the costs for locating, contacting and interviewing a segment (interviewer and supervisor salaries, including time spent in training). C_{T_1} is the between-PSUs travel cost, and C_{T_2} is the between-segments-within-PSUs travel costs. In practice the distance to cover between sampling points is greater than the minimum expected, due to limitations of the local road network. Taking this factor into account, this author suggests multiplying C_{T_1} and C_{T_2} by a factor, whose value depends on the characteristics of the road network. (He treats in the same way the cost of revisiting due to absenteeism.)

Modelling survey errors: relative efficiency of some area sampling strategies

Survey errors are of two categories: sampling errors and non-sampling errors. The former are variable and depend on the sample design. The latter are fixed: they remain constant over all possible samples (given an unbiased design) and their effect is a biased estimate. Non-sampling errors arise from three main sources: non-coverage, non-response and measurement [Groves (1989), Biemer et al (1991), Biemer and Lyberg (2003)]. Area frames are complete and protect against non-coverage bias. The risks of non-response and of measurement (respondent) bias are lower when data are observed directly on the ground, using measurement instruments, than when they must be collected by interview [Carfagna et al (2013a)]. These risks can be reduced by an adequate training of interviewers and supervisors.

Sampling variance due to random sample selection is the main source of variation in survey results. However, the interviewer and the measurement instrument are additional sources of variable error that must be controlled [Fuller (2009)]. From now on we focus on modelling the sampling variance, assuming that non-response, the measurement error, and the interviewer effect are under control and that they are small enough to be ignored.

The sampling variance depends on the correlation structure of the survey variable. We use correlogram models to assess this structure. Bellhouse (1977) showed that there is no uniformly optimum sampling scheme for a general class of correlograms. This is why we propose to identify a cost-efficient sampling strategy by comparing the expected value of the sampling variance (anticipated variance) of a set of alternative sampling strategies, keeping constant the total survey cost. To derive this expected value, we assume that the universe of elementary observations is generated according to a second-order stationary random process with the following characteristics:

The mean, $EY_{hij} = \mu_h$, and the variance, $VY_{hij} = \sigma_h^2$, within a stratum are constants.

The covariance $Cov(Y_{hij}, Y_{hi'j'}) = \sigma_h^2 \rho_h(\text{dist}(s_{hij}, s_{hi'j'}))$ between two EOs, $(Y_{hij}, Y_{hi'j'})$, referred to the points of coordinates s_{hij} and $s_{hi'j'}$, is positive, $\rho_h(\cdot) \geq 0$, and it decreases when the distance between these points, $\text{dist}(s_{hij}, s_{hi'j'})$, increases (note that the expectation operator, E , and as a result the variances, V , and covariances, $Cov(\cdot)$, are defined over the model-based distribution and not over the design-based distribution).

There are a great many quotations on the empirical evidence supporting this model, including Tobler's first law of Geography: "Everything is related to everything else but near things are more related than distant things." This is why the existence of positive correlation among near observations is a basic principle (the proximity principle) in survey sampling [Jessen (1978,

p.439)]. To assess the correlation structure, theoretical variogram and correlogram models are proposed in the literature [Cressie (1991), chapter 2; Gallego and Carfagna (1995), Gallego et al. (1998), Gallego et al. (1999)].

Two often used correlogram models are the exponential model, $\rho_h(d|a_h, \tau_h) = (1 - \tau_h)e^{-d/a_h}$, and the spherical model, $\rho_h(d|a_h, \tau_h) = (1 - \tau_h) \left[1 - \frac{3d}{2a_h} + \frac{d^3}{2a_h^3} \right]$ if $d \leq a_h$ and $\rho_h(u, v|a_h, \tau_h) = 0$ if $d > a_h$, where $d = \text{dist}(s_{hij}, s_{hij'})$. The model parameters are the range rate, a_h , and the ratio $\tau_h = \tau_{h0} / (\tau_{h0} + \tau_{hd})$, where τ_{h0} is the nugget effect, i.e., the variation at the origin or near the origin (independent of the distance), τ_{hd} is the partial sill (a function of the distance d between sampling points) and $(\tau_{h0} + \tau_{hd})$ is the sill, i.e. the maximum variation far from the origin.

Note that, as $\tau_h \rightarrow 1$, in such a way that the partial sill tends towards zero ($\tau_{hd} \rightarrow 0$), the variogram tends to be a pure nugget effect, independent of distance, and the spatial correlation tends to be zero. When $\tau_{hd} > 0$, the variogram is equal to the sum of two components (both positive), one of them, τ_{h0} , is independent of the distance but the other, τ_{hd} , depends on the distance and the spatial correlation increases when τ_{hd} increases with respect to τ_{h0} in such a way that $\tau_h \rightarrow 0$. As we will show below, $\rho_h(d|a_h, \tau_h)$ makes the difference between sampling strategies: if there is no correlation, $\rho_h(d|a_h, \tau_h) = 0$, all sampling strategies have the same expected sampling variance and are equally efficient.

Cressie (1991) and Journel and Huijbregts (1978) show that the moment estimator of the correlation function is biased, but that of the variogram is unbiased, so that the variogram function will be estimated here. The correlogram function will be derived from the variogram function. The empirical variogram is computed using pilot sample data and the moment estimator. The moment estimator of the variogram is (Cressie, 1991)

$$2\hat{\gamma}_h(d) = \frac{1}{|N_h(d)|} \sum_{N_h(d)} (y_{s_h} - y_{s'_h})^2, \text{ where } (y_{s_h}, y_{s'_h}) \text{ are the survey variable values in the pair}$$

of sampling units (s_h, s'_h) included in the pilot sample and $N_h(d) = \{(s_h, s'_h) | \text{dist}(s_h, s'_h) = d\}$ is the number of pairs of sampling units from the pilot sample a distance a part d . The correlation function will be estimated from the variogram function, using the relationship

$$\hat{\rho}_h(d) = 1 - \frac{\hat{\gamma}_h(d)}{\hat{\sigma}_h^2}. \text{ As the estimator, } \hat{\sigma}_h^2, \text{ the sampling quasi-variance is considered.}$$

We will fit a theoretical semivariogram (exponential or spherical) to the empirical semivariogram in order to estimate the range, a_h , the nugget effect, τ_{h0} , and the partial sill, τ_{hd}

. The ratio $\tau_h = \tau_{h0} / (\tau_{h0} + \tau_{hd})$, is estimated by replacing τ_{h0} and τ_{hd} by their sample estimates.

Note that by replacing these estimates in the formulas of the expected sampling variance, we get model-based consistent estimates of the sampling error in the sense that the sampling variance estimators are design-consistent and the expected values estimator is model-based unbiased.

7.1 Stratified sampling

We consider first a simple strategy where a population of N segments is stratified into L strata: in stratum h , ($h=1,2,\dots,L$), there are N_h segments and $N_h n_0$ EOs. We select a simple random sample of n_h segments, independently from each stratum. Let $\hat{Y} = \sum_{h=1}^L \hat{Y}_h$ be the estimator of the total survey variable, where $\hat{Y}_h = N_h \hat{\bar{Y}}_h$ is the estimator of the total survey variable in the stratum h , where $\hat{\bar{Y}}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi}$, is the mean estimator, $Y_{hi} = \sum_{j=1}^{n_0} \omega_{hij} Y_{hj}$ is the survey variable value in the i^{th} segment, where ω_{hij} is the proportion of the total area of the j^{th} EO in the i^{th} segment and Y_{hj} is the survey variable value in the j^{th} EO.

The sampling variance is $V\hat{Y} = \sum_{h=1}^L V\hat{Y}_h$, where $V\hat{Y}_h = N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$, where $S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$ is the variance between segments and $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$ is the mean per segment. The expected value

of this sampling variance is $EV\hat{Y} = \sum_{h=1}^L EV\hat{Y}_h$ where $EV\hat{Y}_h = N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} ES_h^2$, where [Das (1950)]

$$ES_h^2 = \sigma_h^2 \Psi(N_h, n_0 | a_h, \tau_h) \quad \text{and}$$

$$\Psi(N_h, n_0 | a_h, \tau_h) = \frac{n_0(N_h n_0 - 1)}{(N_h - 1)} \left[1 - \Phi(N_h, n_0 | a_h, \tau_h)\right] - \frac{N_h n_0 (n_0 - 1)}{(N_h - 1)} \left[1 - \Phi(n_0 | a_h, \tau_h)\right]$$

In this last expression, $\Phi(N_h, n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{N_h n_0}^2$ pairs of EOs that can be formed with the $N_h n_0$ EOs from the stratum h . $\Phi(n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{n_0}^2$ pairs of EOs that can be formed with the n_0 EOs of a segment.

An alternative sampling strategy is using EOs instead of segments as last sampling units. We select a simple random sample of $(n_h n_0)$ EOs, instead of n_h segments with n_0 EOs each. The

sampling variance is $V\hat{Y}_0 = \sum_{h=1}^L V\hat{Y}_{h0}$, where $V\hat{Y}_{h0} = (N_h n_0)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{h0}^2}{n_h n_0}$, where $S_{h0}^2 = \frac{1}{N_h n_0 - 1} \sum_{i=1}^{N_h n_0} (Y_{h0i} - \bar{Y}_{h0})^2$ is the variance between Eos, and $\bar{Y}_{h0} = \frac{1}{N_h n_0} \sum_{i=1}^{N_h n_0} Y_{h0i}$ is the mean per EOs.

The expected value of the sampling variance with this alternative strategy is $EV\hat{Y}_0 = \sum_{h=1}^L EV\hat{Y}_{h0}$

where $EV\hat{Y}_{h0} = N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{n_0 \sigma_h^2}{n_h} \left[1 - \Phi(N_h, n_0 | a_h, \tau_h)\right]$. The relative efficiency of sampling

EOs with respect to sampling segments in stratum h is $\frac{EV\hat{Y}_h}{EV\hat{Y}_{h0}} = \frac{\Psi(N_h, n_0 | a_h, \tau_h)}{n_0 \left[1 - \Phi(N_h, n_0 | a_h, \tau_h)\right]}$. If

the correlation is null, $\rho_h(d | a_h, \tau_h) = 0$, then $\Phi(N_h, n_0 | a_h, \tau_h) = 0$, $\Psi(N_h, n_0 | a_h, \tau_h) = n_0$ and

$EV\hat{Y}_h = EV\hat{Y}_{h0}$ and, as a result, these two sampling strategies are equally efficient. Generally, $\rho_h(d|a_h, \tau_h) > 0$ and it increases when $d = dist(s_{hij}, s_{hi'j'})$ decreases and, as a result, $\Phi(n_0|a_h, \tau_h) > \Phi(N_h, n_0|a_h, \tau_h)$, $\Psi(N_h, n_0|a_h, \tau_h) < n_0[1 - \Phi(N_h, n_0|a_h, \tau_h)]$, $EV\hat{Y}_h > EV\hat{Y}_{h0}$ and using EOs as last sampling units is most efficient than using segments.

7.2 Replicated sampling by zones

The FAO and UNSD guidelines recommend selecting a replicated sample, instead of only one sample. Here we focus on the relative efficiency of this sampling strategy, and we refer to these guidelines for details on its advantages for managing surveys. We consider the case where the replicate is defined by partitioning the stratum h in a set of M_h zones (a kind of sub-stratification) and by selecting an independent simple random sample of one segment from each zone. We assume that the zone size is equal to N_{hz_0} segments. To select a sample of n_h segments

we select $r_h = \frac{n_h}{M_h}$ replicated samples of M_h segments.

Let $\hat{Y}_r = \sum_{h=1}^L \hat{Y}_{rh}$ be the estimator of the survey variable total in the population, where $\hat{Y}_{rh} = \sum_{z=1}^{M_h} \hat{Y}_{rhz}$

is the estimator of the survey variable total in stratum h and $\hat{Y}_{rhz} = N_{hz_0} \hat{\bar{Y}}_{rhz}$ is the estimator of the survey variable total in z^{th} zone from this stratum, where $\hat{\bar{Y}}_{rhz} = \frac{1}{r_h} \sum_{i=1}^{r_h} Y_{hzi}$ is the estimator of the mean per segment and Y_{hzi} is the survey variable value in the i^{th} segment.

The sampling variance is $V\hat{Y}_r = \sum_{h=1}^L V\hat{Y}_{rh}$, where $V\hat{Y}_{rh} = \sum_{z=1}^{M_h} V\hat{Y}_{rhz}$ and $V\hat{Y}_{rhz} = N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{S_{hz}^2}{r_h}$,

where $S_{hz}^2 = \frac{1}{N_{hz} - 1} \sum_{i=1}^{N_{hz}} (Y_{hzi} - \bar{Y}_{hz})^2$ is the variance of the survey variable between segments within the z^{th} zone from stratum h . The expected value of this sampling variance is $EV\hat{Y}_r = \sum_{h=1}^L EV\hat{Y}_{rh}$ where $EV\hat{Y}_{rh} = \sum_{z=1}^{M_h} EV\hat{Y}_{rhz}$, where

$$EV\hat{Y}_{rhz} = N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{1}{r_h} \sigma_h^2 \Psi(N_{hz_0}, n_0 | a_h, \tau_h). \quad \text{As a result,}$$

$$EV\hat{Y}_{rh} = M_h N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{1}{r_h} \sigma_h^2 \Psi(N_{hz_0}, n_0 | a_h, \tau_h), \text{ where}$$

$$\Psi(N_{hz_0}, n_0 | a_h, \tau_h) = \frac{n_0(N_{hz_0} n_0 - 1)}{(N_{hz_0} - 1)} \left[1 - \Phi(N_{hz_0}, n_0 | a_h, \tau_h)\right] - \frac{N_{hz_0} n_0 (n_0 - 1)}{(N_{hz_0} - 1)} \left[1 - \Phi(n_0 | a_h, \tau_h)\right]$$

where $\Phi(N_{hz_0}, n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{N_{hz_0} n_0}^2$ pairs in a zone.

The relative efficiency of sampling replicated samples of segments with respect to simple random sampling in stratum h is $\frac{EV\hat{Y}_h}{EV\hat{Y}_{rh}} = \frac{\Psi(N_h, n_0 | a_h, \tau_h)}{\Psi(N_{hz_0}, n_0 | a_h, \tau_h)}$ and with respect to simple random

sampling of EOs is $\frac{EV\hat{Y}_{h0}}{EV\hat{Y}_{rh}} = \frac{n_0 [1 - \Phi(N_h, n_0 | a_h, \tau_h)]}{\Psi(N_{hz_0}, n_0 | a_h, \tau_h)}$. If the correlation is null,

$\rho_h(d | a_h, \tau_h) = 0$, then $\Phi(N_h, n_0 | a_h, \tau_h) = 0$, $\Psi(N_h, n_0 | a_h, \tau_h) = \Psi(N_{hz_0}, n_0 | a_h, \tau_h) = n_0$, $EV\hat{Y}_{rh} = EV\hat{Y}_h = EV\hat{Y}_{h0}$ and, as result, these three sampling strategies are equally efficient. Generally, $\rho_h(d | a_h, \tau_h) > 0$ and as a result, $\Psi(N_{hz_0}, n_0 | a_h, \tau_h) < \Psi(N_h, n_0 | a_h, \tau_h)$, $EV\hat{Y}_{rh} < EV\hat{Y}_h$, and replicated sampling is more efficient than simple random sampling.

7.3 Systematic replicated sampling

Now we consider the case where replicates are systematic samples. We number the segments of the stratum h from 1 to N_h and we divide this set into M_h sections of size $N_{hz_0} = N_h / M_h$ segments each. We select a systematic sample by choosing with equal probability a random start between 1 and N_{hz_0} and include in the sample the segment from each section occupying the same position as the chosen random start. There are N_{hz_0} systematic (replicated) samples, and to select a sample of n_h segments we select $r_h = \frac{n_h}{M_h}$ systematic samples of M_h segments each.

Let $\hat{Y}_{sy} = \sum_{h=1}^L \hat{Y}_{syh}$ be the total estimator, where $\hat{Y}_{syh} = N_{hz_0} \hat{Y}_{syh}$ is the total estimator in the stratum

h , and $\hat{Y}_{syh} = \frac{1}{r_h} \sum_{i=1}^{r_h} Y_{hi}$ is the mean per replicated estimator, where $Y_{hi} = \sum_{j=1}^{M_h} Y_{hij}$ is the total in the

replicated $i = 1, 2, \dots, N_{hz_0}$. The sampling variance is $V\hat{Y}_{sy} = \sum_{h=1}^L V\hat{Y}_{syh}$, where

$V\hat{Y}_{syh} = N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{S_{syh}^2}{r_h}$, and where $S_{syh}^2 = \frac{1}{N_{hz_0} - 1} \sum_{i=1}^{N_{hz_0}} (Y_{hi} - \bar{Y}_h)^2$ is the variance between replicated totals.

The expected value of the sampling variance is $EV\hat{Y}_{sy} = \sum_{h=1}^L EV\hat{Y}_{syh}$, where

$$EV\hat{Y}_{syh} = N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{1}{r_h} \sigma_h^2 \Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right), \text{ where}$$

$$\Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right) = \frac{(N_{hz_0} M_h n_0 - 1) M_h n_0}{(N_{hz_0} - 1)} \left[1 - \Phi \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right)\right]$$

$$- \frac{N_{hz_0} (M_h - 1) M_h n_0}{(N_{hz_0} - 1)} \left[1 - \Phi \left(M_h, n_0 | a_h, \tau_h\right)\right],$$

$$- \frac{N_{hz_0} M_h (n_0 - 1) M_h n_0}{(N_{hz_0} - 1)} \left[1 - \Phi \left(n_0 | a_h, \tau_h\right)\right]$$

where $\Phi \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right)$ is the average of the correlation between pairs of observations over the $C_{N_{hz_0} M_h n_0}^2$ pairs that can be formed with the $N_{hz_0} M_h n_0$ EOs of the stratum h and $\Phi \left(M_h, n_0 | a_h, \tau_h\right)$ is the average of the correlation between pairs of observations over the $C_{M_h n_0}^2$ pairs that can be formed with the $M_h n_0$ EOs within a systematic sample.

The relative efficiency of replicated systematic sampling with respect to zone replicated in stratum h is $\frac{EV\hat{Y}_{rh}}{EV\hat{Y}_{syh}} = \frac{M_h \Psi \left(N_{hz_0}, n_0 | a_h, \tau_h\right)}{\Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right)}$, with respect to simple random sampling of

segments is $\frac{EV\hat{Y}_h}{EV\hat{Y}_{syh}} = \frac{M_h \Psi \left(N_h, n_0 | a_h, \tau_h\right)}{\Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right)}$ and with respect to simple random sampling of

EOs is $\frac{EV\hat{Y}_{h0}}{EV\hat{Y}_{syh}} = \frac{M_h n_0 \left[1 - \Phi \left(N_h, n_0 | a_h, \tau_h\right)\right]}{\Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right)}$. If the correlation is null, $\rho_h \left(d | a_h, \tau_h\right) = 0$, then

$\Phi \left(N_h, n_0 | a_h, \tau_h\right) = 0$, $\Psi \left(N_h, n_0 | a_h, \tau_h\right) = \Psi \left(N_{hz_0}, n_0 | a_h, \tau_h\right) = n_0$, $\Psi_{sy} \left(N_{hz_0}, M_h, n_0 | a_h, \tau_h\right) = M_h n_0$

and $EV\hat{Y}_{syh} = EV\hat{Y}_{rh} = EV\hat{Y}_h = EV\hat{Y}_{h0}$. As a result, these four sampling strategies would be equally efficient. Generally, $\rho_h \left(d | a_h, \tau_h\right) > 0$ and systematic sampling is often relatively more efficient than simple random sampling [Das (1950), Bellhouse (1977), Ambrosio et al (2003)].

7.4 Point sampling

Point sampling corresponds to the particular case where we select a systematic sample of size $M_h = n_h$ EOs ($n_0 = 1$) from the number of possible systematic samples, $N_{hz_0} = N_h / n_h$, using only one random start, $r_h = 1$. The expected value of the sampling variance, $EV\hat{Y}_{syh}$, in this particular case is obtained by replacing N_{hz_0} , M_h , r_h and n_0 by the values above indicated in the general formula.

7.5 Two-stage sampling

The UNSD guidelines provide a number of country examples using two-stage sampling strategies to select a sample of households. Often, the PSU is the census EA, and the last sampling unit is a segment whose target size is usually ten households. In the first stage we select a replicated-by-zones sample of PSUs and in the second stage we select a simple random sample of segments within the PSUs included in the first-stage sample. We assume that the PSUs from stratum h are divided into a set of M_h zones of size N_{1h} PSUs. The first-stage sample size is n_{1h} PSUs and we select $r_h = \frac{n_{1h}}{M_h}$ replicated samples of M_h segments. The size of a PSU is N_{2h} segments and in the second stage we select a simple random sample of n_{2h} segments within each one of the n_{1h} PSUs included in the first-stage sample.

Let $\hat{Y}_{2s,r} = \sum_{h=1}^L \hat{Y}_{2s,rh}$ be the total estimator, where $\hat{Y}_{2s,rh} = \sum_{z=1}^{M_h} \hat{Y}_{2s,rhz}$ is the total estimator in the stratum h and $\hat{Y}_{2s,rhz} = N_{hz1} \hat{\bar{Y}}_{2s,rhz}$ is the estimator of the survey variable total in the z^{th} zone from this stratum and $\hat{\bar{Y}}_{2s,rhz} = \frac{1}{r_h} \sum_{i=1}^{r_h} \hat{Y}_{hzi}$ is the mean per PSU estimator in this zone, where $\hat{Y}_{hzi} = N_{2h} \hat{\bar{Y}}_{hzi}$ is the total estimator in the PSU $i = 1, 2, \dots, r_h$ included in the first-stage sample of the z^{th} zone from the stratum h , where $\hat{\bar{Y}}_{hzi} = \frac{1}{n_{2h}} \sum_{j=1}^{n_{2h}} Y_{hzij}$ is the mean per segment estimator and Y_{hzij} is the survey variable in the segment $j = 1, 2, \dots, n_{2h}$ included in the second-stage sample.

The sampling variance is $V\hat{Y}_{2s,r} = \sum_{h=1}^L V\hat{Y}_{2s,rh}$, where $V\hat{Y}_{2s,rh} = \sum_{z=1}^{M_h} V\hat{Y}_{2s,rhz}$, where

$$V\hat{Y}_{2s,rhz} = N_{1h}^2 \left(1 - \frac{r_h}{N_{1h}} \right) \frac{S_{1hz}^2}{r_h} + N_{1h} \frac{1}{r_h} \sum_{j=1}^{N_{1h}} N_{2h}^2 \left(1 - \frac{n_{2h}}{N_{2h}} \right) \frac{S_{2hzi}^2}{n_{2h}}, \text{ where } S_{1hz}^2 = \frac{1}{N_{1h} - 1} \sum_{i=1}^{N_{1h}} (y_{hzi} - \bar{y}_{hz})^2$$

is the variance between PSUs, and $S_{2hzi}^2 = \frac{1}{N_{2h} - 1} \sum_{j=1}^{N_{2h}} (y_{hzij} - \bar{y}_{hzi})^2$ is the variance between

segments within PSUs. The expected value of the sampling variance is $EV\hat{Y}_{2s,r} = \sum_{h=1}^L EV\hat{Y}_{2s,rh}$

where $EV\hat{Y}_{2s,rh} = \sum_{z=1}^{M_h} EV\hat{Y}_{2s,rhz}$, where

$$EV\hat{Y}_{2s,rhz} = N_{1h}^2 \left(1 - \frac{r_h}{N_{1h}} \right) \frac{1}{r_h} \sigma_h^2 \Psi_1(N_{1h}, N_{2h}, n_0 | a_h, \tau_h) + N_{1h} N_{2h}^2 \left(1 - \frac{n_{2h}}{N_{2h}} \right) \frac{1}{n_{2h}} \sigma_h^2 \Psi_2(N_{2h}, n_0 | a_h, \tau_h)$$

. As a result,

$$EV\hat{Y}_{2s,rh} = M_h \left[N_{1h}^2 \left(1 - \frac{r_h}{N_{1h}} \right) \frac{1}{r_h} \sigma_h^2 \Psi_1(N_{1h}, N_{2h}, n_0 | a_h, \tau_h) + N_{1h} N_{2h}^2 \left(1 - \frac{n_{2h}}{N_{2h}} \right) \frac{1}{n_{2h}} \sigma_h^2 \Psi_2(N_{2h}, n_0 | a_h, \tau_h) \right]$$

where

$$\Psi_1(N_{1h}, N_{2h}, n_0 | a_h, \tau_h) = \frac{(N_{1h}N_{2h}n_0 - 1)}{(N_{1h} - 1)N_{2h}n_0} [1 - \Phi(N_{1h}, N_{2h}, n_0 | a_h, \tau_h)]$$

$$- \frac{N_{1h}(N_{2h} - 1)n_0}{(N_{1h} - 1)N_{2h}n_0} \Psi_2(N_{2h}, n_0 | a_h, \tau_h)$$

$$- \frac{N_{1h}N_{2h}(n_0 - 1)}{(N_{1h} - 1)N_{2h}n_0} [1 - \Phi(n_0 | a_h, \tau_h)]$$

$$\Psi_2(N_{2h}, n_0 | a_h, \tau_h) = \frac{(N_{2h}n_0 - 1)}{(N_{2h} - 1)n_0} [1 - \Phi(N_{2h}, n_0 | a_h, \tau_h)] - \frac{N_{2h}(n_0 - 1)}{(N_{2h} - 1)n_0} [1 - \Phi(n_0 | a_h, \tau_h)]$$

where

$\Phi(N_{1h}, N_{2h}, n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{N_{1h}N_{2h}n_0}^2$ pairs that can be formed with the $N_{1h}N_{2h}n_0$ EOs from the stratum h ; $\Phi(N_{2h}, n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{N_{2h}n_0}^2$ pairs that can be formed with the $N_{2h}n_0$ EOs of a PSU from the stratum h ; and $\Phi(n_0 | a_h, \tau_h)$ is the average of the correlation between pairs of observations over the $C_{n_0}^2$ pairs that can be formed with the n_0 EOs of a segment.

Choosing an area sampling strategy and optimizing the sample design

Planning surveys and designing samples to produce information are economic processes, and it is important to optimize the available resources. In the case of a single variable, the cost-efficiency approach to optimizing the design of the sample is well established in the literature. However, surveys are usually multipurpose and a sample that is optimal for one purpose is not necessarily optimal for other purposes. The optimization problem is even more complex when it comes to designing a master sample, which includes many multipurpose surveys.

In spite of these difficulties, we think that the cost-efficiency approach offers a systematic way to evaluate sampling strategies and is better than blindly accepting some standard solution [Groves (1989, p.56)]. We evaluate the cost-efficiency of a set of alternative sampling strategies, using the survey costs model specified in section 6 and the survey errors models specified in section 7. The strategy that achieves the minimum sampling error within the set of alternative sampling strategies under consideration, keeping constant the total cost of the survey, is the cost-efficient sampling strategy.

Stratified sampling

Using segments as last sampling units, we find the minimum expected value of the sampling variance, given the cost, C , by solving with respect to $\{n_0, n_h; h = 1, 2, \dots, L\}$ the optimization problem:

$$\min_{\{(n_h, n_0), h=1, 2, \dots, L\}} \sum_{h=1}^L EV\hat{Y}_h = \min_{\{(n_h, n_0), h=1, 2, \dots, L\}} \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \sigma_h^2 \Psi(N_h, n_0 | a_h, \tau_h)$$

$$\text{Subject to: } C = C_0 + \sum_{h=1}^L C_{1h} n_h + \sum_{h=1}^L C_{2h} n_h n_0 + \sum_{h=1}^L C_{T_1h} \sqrt{n_h} + \sum_{h=1}^L C_{T_2h} n_h \sqrt{n_0}$$

The solution to this problem is the optimum segment size, n_0 , and the optima sample size of segments in each stratum, $\{n_h; h = 1, 2, \dots, L\}$. In addition to the cost, C , this optima solution is conditioned to the correlogram model parameters, (a_h, τ_h) , and to the restriction that $N_h n_0$ must be equal to the number of EOs in the stratum h .

Using EOs as last sampling units, we find the minimum expected value of the sampling variance with respect to $(n_h n_0)$ and given the cost, C , by replacing $EV\hat{Y}_h$ by $EV\hat{Y}_{h0}$ in the first equation

and $C_{T_1h}\sqrt{n_h} + C_{T_2h}n_h\sqrt{n_0}$ by $C_{T_1h}\sqrt{n_h n_0}$ and $C_{1h}n_h + C_{2h}n_h n_0$ by $(C_{1h} + C_{2h})n_h n_0$ in the second equation and solving the optimization problem. Using segments as last sampling units is relatively more efficient than using EOs if
$$\min_{\{(n_h, n_0), h=1, 2, \dots, L\}} \sum_{h=1}^L EV\hat{Y}_h < \min_{\{(n_h n_0), h=1, 2, \dots, L\}} \sum_{h=1}^L EV\hat{Y}_{h0},$$
 given the cost C .

Replicated sampling by zones

We find the minimum expected value of the sampling variance of this sampling strategy, given the cost C , by solving the following optimization problem with respect to $\{n_0, M_h, r_h; h = 1, 2, \dots, L\}$:

$$\min_{\{(n_0, M_h, r_h), h=1, 2, \dots, L\}} \sum_{h=1}^L EV(\hat{Y}_{rh}) = \min_{\{(n_0, M_h, r_h), h=1, 2, \dots, L\}} \sum_{h=1}^L M_h N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{1}{r_h} \sigma_h^2 \Psi(N_{hz_0}, n_0 | a_h, \tau_h)$$

Subject to $C = C_0 + \sum_{h=1}^L C_{1h} M_h r_h + \sum_{h=1}^L C_{2h} M_h r_h n_0 + \sum_{h=1}^L C_{T_1h} M_h \sqrt{r_h} + \sum_{h=1}^L C_{T_2h} M_h r_h \sqrt{n_0}$

The solution to this problem is the optimum segment size, n_0 , the optima replica size, $\{M_h; h = 1, 2, \dots, L\}$, and the optima replicated sample size, $\{r_h; h = 1, 2, \dots, L\}$, in each stratum. In addition to the cost, C , this optima solution is conditioned to the correlogram model parameters, (a_h, τ_h) , and to the restrictions $M_h N_{hz_0} = N_h$ and $N_h n_0$ equal to the number of EOs in the stratum h .

Replicated sampling by zones is the relatively most efficient within the considered set of alternative sampling strategies, if the minimum expected value of its sampling variance,

$$\min_{\{(n_0, M_h, r_h), h=1, 2, \dots, L\}} \sum_{h=1}^L EV(\hat{Y}_{rh}), \text{ given the cost, } C, \text{ is the lowest one.}$$

Systematic replicated sampling

We find the minimum expected value of the sampling variance of this sampling strategy, given the cost C , by solving the following optimization problem with respect to $\{n_0, M_h, r_h; h = 1, 2, \dots, L\}$:

$$\min_{\{(n_0, M_h, r_h), h=1, 2, \dots, L\}} \sum_{h=1}^L EV(\hat{Y}_{syh}) = \min_{\{(n_0, M_h, r_h), h=1, 2, \dots, L\}} \sum_{h=1}^L N_{hz_0}^2 \left(1 - \frac{r_h}{N_{hz_0}}\right) \frac{1}{r_h} \sigma_h^2 \Psi_{sy}(N_{hz_0}, M_h, n_0 | a_h, \tau_h)$$

Subject to $C = C_0 + \sum_{h=1}^L C_{1h} M_h r_h + \sum_{h=1}^L C_{2h} M_h r_h n_0 + \sum_{h=1}^L C_{T_1h} \sqrt{M_h r_h} + \sum_{h=1}^L C_{T_2h} M_h r_h \sqrt{n_0}$

The solution to this problem is the optimum segment size, n_0 , the optima replicated size, $\{M_h; h = 1, 2, \dots, L\}$, and the optima replicated sample size, $\{r_h; h = 1, 2, \dots, L\}$, in each stratum. In addition to the cost, C , this optima solution is conditioned to the correlogram model parameters, (a_h, τ_h) , and to the restrictions $M_h N_{hz_0} = N_h$ and $N_h n_0$ equal to the number of EOs in the stratum h .

Systematic sampling is the relatively most efficient strategy within the considered set of alternative sampling strategies, if the minimum expected value of its sampling variance,

$$\min_{\{(n_0, M_h, r_h); h=1, 2, \dots, L\}} \sum_{h=1}^L EV(\hat{Y}_{syh}), \text{ given the cost, } C, \text{ is the lowest one.}$$

Point sampling

Point sampling corresponds to the particular case where we select a systematic sample of size $M_h = n_h$ EOs ($n_0 = 1$) from the number of possible systematic samples, $N_{hz_0} = N_h / n_h$, using only one random start, $r_h = 1$. We find the minimum expected value of the sampling variance of this sampling strategy, given the cost C , by replacing in $EV\hat{Y}_{syh}$ and C in the above equations N_{hz_0}, M_h, r_h and n_0 by the values above indicated, using $C_{T_2h} = 0$, and solving the optimization problem with respect to $\{n_h; h = 1, 2, \dots, L\}$

Two-stages

We find the minimum expected value of the sampling variance of this sampling strategy, given the cost C , by solving the following optimization problem with respect to $\{n_0, M_h, r_h, n_{2h}; h = 1, 2, \dots, L\}$

$$\begin{aligned} & \min_{\{(n_0, M_h, r_h, n_{2h}); h=1, 2, \dots, L\}} \sum_{h=1}^L EV\hat{Y}_{2s, r_h} = \\ & = \min_{\{(n_0, M_h, r_h, n_{2h}); h=1, 2, \dots, L\}} \sum_{h=1}^L M_h \left[\begin{aligned} & N_{1h}^2 \left(1 - \frac{r_h}{N_{1h}} \right) \frac{1}{r_h} \sigma_h^2 \Psi_1(N_{1h}, N_{2h}, n_0 | a_h, \tau_h) \\ & + N_{1h} N_{2h}^2 \left(1 - \frac{n_{2h}}{N_{2h}} \right) \frac{1}{n_{2h}} \sigma_h^2 \Psi_2(N_{2h}, n_0 | a_h, \tau_h) \end{aligned} \right] \end{aligned}$$

$$\begin{aligned} \text{Subject to } C = & C_0 + \sum_{h=1}^L C_{1h} M_h r_h + \sum_{h=1}^L C_{2h} M_h r_h n_{2h} + \sum_{h=1}^L C_{3h} M_h r_h n_{2h} n_0 + \\ & + \sum_{h=1}^L C_{T_1h} \sqrt{M_h r_h} + \sum_{h=1}^L C_{T_2h} M_h r_h \sqrt{n_{2h}} + \sum_{h=1}^L C_{T_3h} M_h r_h n_{2h} \sqrt{n_0} \end{aligned}$$

The solution to this problem is the optimum segment size, n_0 , the optima first-stage replicated size, $\{M_h; h = 1, 2, \dots, L\}$, the optima replicated first-stage sample size, $\{r_h; h = 1, 2, \dots, L\}$, and the optima second-stage simple random sample, $\{n_{2h}; h = 1, 2, \dots, L\}$, in each stratum. In addition to the cost, C , this optima solution is conditioned to the correlogram model parameters, (a_h, τ_h) , and to these two restrictions: $M_h N_{1h}$ must be equal to the number of PSUs in stratum h , and $M_h N_{1h} N_{2h} n_0$ must be equal to the number of EOs in stratum h .

This two-stage sampling strategy is the relatively most efficient one within the considered set of alternative sampling strategies, if the minimum expected value of its sampling variance,

$$\min_{\{(n_0, M_h, n_h, n_{2h}); h=1, 2, \dots, L\}} \sum_{h=1}^L EV\hat{Y}_{2s, rh}, \text{ given the cost, } C, \text{ is the lowest one.}$$

Abacuses

To assess the correlation structure effect on the sampling error and to facilitate the calculus of the average correlations between pairs of elementary observations, $\Psi(\cdot|a_h, \tau_h)$ and $\Psi_{sy}(\cdot|a_h, \tau_h)$, we suggest the development of abacuses in function of the values of the correlogram model parameters, (a_h, τ_h) . For instance, in Ambrosio et al (2004) we consider a set of a_h values and a set of τ_h values ranging from high to low correlation and we calculate the average correlations for a set of population sizes arranged in a lattice of rows and columns. This kind of arrangement is often the basis for spatial sampling and is found or can be constructed from many cartographic or other representations of space, including remote sensing. In the Universal Transverse Mercator (UTM) system, the territory is divided into square units by lines parallel to the axis of a Cartesian system of reference. The UTM co-ordinates identify the row and the column of the unit. This is the basis to build area frames of segments with geometric boundaries.

Multipurpose sample design

The usual approach to address multipurpose sample design is to seek a compromise between the optima corresponding to the set of single purposes. Following Kish [see Kalton and Heeringa (2003)], we seek some form of *averaging* between all the optima single-purpose sample designs. As Kish pointed out, this involves assigning relative values of importance (weights), I_g , to all purposes; this may be difficult but an “ignorant” decision-maker can assign equal I_g value to all them.

Let $\min EV\hat{Y}_g$ be the minimum expected value of the sampling variance corresponding to the optimum sample design for the g^{th} survey variable. Let $EV\hat{Y}_g$ be the expected value of the sampling variance for the g^{th} survey variable corresponding to compromise sample design. Let

$L_g = \frac{EV\hat{Y}_g - \min EV\hat{Y}_g}{\min EV\hat{Y}_g} = \frac{EV\hat{Y}_g}{\min EV\hat{Y}_g} - 1$ be the loss due to using the compromise sample instead

of the optimum sample. The relative increment of the compromise sampling variance with respect to that of the optimum sample is $C_g^2 = \frac{EV\hat{Y}_g}{\min EV\hat{Y}_g} = 1 + L_g$ and the weighted average over

the set of multipurpose is $\sum_g I_g C_g^2 = \sum_g \frac{I_g}{\min EV\hat{Y}_g} EV\hat{Y}_g$.

Using segments as last sampling units, we find the minimum expected value of the sampling variance, given the cost, C , by solving with respect to $\{n_0, n_h; h = 1, 2, \dots, L\}$ the optimization problem:

$$\min_{\{(n_h, n_0), h=1, 2, \dots, L\}} \sum_g \frac{I_g}{\min EV\hat{Y}_g} \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \sigma_{gh}^2 \Psi_g(N_h, n_0 | a_{gh}, \tau_{gh})$$

$$\text{Subject to: } C = C_0 + \sum_{h=1}^L C_{1h} n_h + \sum_{h=1}^L C_{2h} n_h n_0 + \sum_{h=1}^L C_{T_1h} \sqrt{n_h} + \sum_{h=1}^L C_{T_2h} n_h \sqrt{n_0}$$

The solution to this problem is the multipurpose compromise segment size, n_0 , and the multipurpose compromise sample size of segments in each stratum, $\{n_h; h = 1, 2, \dots, L\}$. In addition to the cost, C , this multipurpose compromise solution is conditioned to the correlogram model parameters for each survey variable, (a_{gh}, τ_{gh}) , and to the restriction that $N_h n_0$ must be equal to the number of EOs in the stratum h .

This sampling strategy is the relatively most efficient one within the considered set of alternative sampling strategies, if the minimum expected value of its weighted sampling

variance, $\min_{\{(n_h, n_0), h=1, 2, \dots, L\}} \sum_g \frac{I_g}{\min EV\hat{Y}_g} EV\hat{Y}_g$, given the cost, C , is the lowest one.

Concluding remarks

In this document we focus on identifying a cost-efficient sampling strategy that integrates agricultural surveys and household surveys. The starting point is the guidelines developed by FAO and UNSD for the planning and realization of agricultural and household surveys, respectively. These two guidelines basically give the same recommendations: using dual (or multiple) sampling frames, with an area frame component; and selecting replicated samples of segments. Area sampling frames are recommended because they protect against non-coverage bias. The cost of building an area frame is fixed and depends largely on the type of segment boundaries: it is higher when using recognizable physical boundaries on the ground than when using geometric boundaries. The efficiency of the sample design depends on the correlation structure of the survey variable. For a given correlation structure, the sampling error and the cost of the survey depend on the sample design, mainly on segment size, replicate size and sample size.

Since there is no uniformly optimal sampling strategy for a general class of correlation structures, we propose identifying a cost-efficient sampling strategy by comparing the expected value of the sampling error of a set of alternative sampling strategies, keeping constant the total cost of the survey. We derive the expected value of the sampling error using a model of the correlation structure, and the result is a function of the characteristics of the sample: segment size, replicate size and sample size. We optimize the sample by minimizing the expected value of the sampling error, keeping constant the total cost of the survey. The strategy that achieves the minimum sampling error within the set of alternative sampling strategies under consideration is the cost-efficient sampling strategy.

One of the main difficulties in our approach is to assess the correlation structure and calculate the average correlations between pairs of elementary observations. To address this difficulty, we suggest the development of abacuses to determine the average correlations in function of the values of the correlogram model parameters, given the characteristics of the sample. The other main difficulty is to optimize multipurpose samples. To address this difficulty, it is suggested that an importance value be assigned to each individual purpose and that the mean of the sampling errors of individual purposes be minimized, weighted by their importance. We suggest the development of specific software for solving the optimization problems formulated in sections 8 and 9.

We have limited ourselves to the simplest case in dual frame, where the list frame is sampled at 100 percent rate and only the area sample needs to be designed. In future developments of this research for choosing a sampling strategy, we will treat the general case, where an additional sample to be selected from the list frame needs to be designed and the weight of each one of these two samples in the final estimator needs to be estimated.

Bibliography

- Abaye, A.T. 2010. Combining enumeration area maps and satellite images (land cover) for the development of area frame (multiple frames) in an African country: Preliminary lessons from the experience of Ethiopia. National Statistical Data Quality Directorate. Director Central Statistical Agency. Addis Ababa.
- Ambrosio, L., Iglesias, L. & Marin, C. 2003. Systematic sampling design for the estimation of spatial means, *Environmetrics*. 14: 45-61.
- Ambrosio, L., Iglesias, L., Marin, C. & Del Monte, J.P. 2004. Evaluation of sampling methods and assessment of the sample size to estimate the weed seedbank in soil, taking into account spatial variability. *Weed Research*, 44: 224-236.
- Ambrosio, L. 2012. Marcos múltiples de áreas y listas. Diseño de encuestas sobre la estructura y la producción agrícola en Chile. Informe técnico. Instituto Nacional de Estadísticas de Chile. FAO. Universidad Politécnica de Madrid.
- Ambrosio, L. 2013a. Marco de muestreo y diseño de la Encuesta Nacional Agropecuaria de Guatemala 2013. Informe Técnico. Ministerio de Agricultura. Ganadería y Alimentación de Guatemala. Instituto Nacional de Estadística de Guatemala. FAO. Universidad Politécnica de Madrid.
- Ambrosio, L. 2013b. Marco de muestreo y diseño de la Encuesta Agropecuaria de Nicaragua. Informe Técnico. Ministerio Agropecuario y Forestal de Nicaragua. FAO. Universidad Politécnica de Madrid.
- Bankier, M.D. 1986. Estimators Based on Several Stratified Samples with Applications to Multiple Frame Surveys. *Journal of the American Statistical Association*, 81: 1074-1079.
- Bellhouse, D.R. 1977. Some optimum designs for sampling in two dimensions. *Biometrics*, 64: 605-611.
- Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. & Sudman, S. (Ed). 1991. *Measurement errors in surveys*. Wiley. New York.
- Biemer, P.P. & Lyberg, L.E. 2003. *Introduction to survey quality*. Wiley. New York.
- Carfagna, E., Pratesi, M. & Carfagna, A. 2013. Methodological developments for improving the reliability and cost-effectiveness of agricultural statistics in developing countries. The 59th World Statistical Congress, Special Topic Session (STS043) "Using geospatial information in area sampling and estimation for agricultural and environmental surveys", Hong Kong, 25-30 August 2013.
- Censo Nacional Agropecuario (CENAGRO). 2013. IV Censo Nacional Agropecuario. Instituto Nacional de Estadística e Informática. Lima.
- Cochran, W.G. 1977. *Sampling Techniques*. Wiley. New York.
- Cotter, J. & Nealon, J. 1989. *Area frame design for agricultural surveys*. NASS. USDA. Washington DC.
- Cotter, J. & Tomczak, C. 1994. *An image analysis system to develop area sampling frames for agricultural surveys*. USDA-NASS. Washington DC.
- Cressie, N.E. 1991. *Statistics for Spatial Data*. Wiley. New York.

- Das, A.C. 1950. Two-dimensional systematic sampling and associated stratified and random sampling. *Sankhya*, 10: 95-108.
- FAO. 1996. Multiple frame agricultural surveys. Vol.1. Current surveys based on area and list sampling methods. Statistical Development Series. 7. Rome.
- FAO. 1998. Multiple frame agricultural surveys. Vol.2. Agricultural survey programmes based on area frame or dual frame (area and list) sample designs. Statistical Development Series, 10. Rome.
- FAO, World Bank and United Nations Statistical Commission. 2011. Global Strategy to Improve Agricultural and Rural Statistics. The World Bank. Washington DC.
- FAO, World Bank and United Nations Statistical Commission. 2012a. Action Plan of the Global Strategy to Improve Agricultural and Rural Statistics. FAO. Rome.
- FAO. 2012b. Guidelines for Linking Population and Housing Censuses with Agricultural Censuses with selected country practices. Special Issue of the FAO Statistical Development Series. Rome.
- FAO-IFAD. 2013. Misión “Propuesta de Mejoras del Sistema de Información y Estadísticas Agropecuarias”. Informe final. Managua. Rome.
- Faulkenberry, G.D. and Garoui, A. 1991. Estimating a population total using an area frame. *Journal of the American Statistical Association*, Vol. 86, n 414, Theory and Method: 445-449.
- Fecso, R., Tortora, R. D. and Vogel, F. 1986. Sampling Frames for Agriculture in the United States. *Journal of Official Statistics*, 2: 279-292.
- Fiji. 2009. Fiji National Agricultural Census. Department of Agriculture. Economic Planning and Statistics Division.
- Ford, B. L., Nealon, I. and Tortora, R. D. 1986. Area frame estimators in agricultural surveys: Sampling versus non-sampling errors. *Agricultural Economics Research*, Vol. 38, N 2: 1-10.
- Fuller, W. A. and Burmeister, L. F. 1972. Estimators for samples selected from two overlapping frames. *Proceedings of the Social Statistics Section. American Statistical Association*: 245-249.
- Fuller, W.A.1984. The master sample of agriculture. In *Statistics: An Appraisal*, David, H.A. and David, H.T., eds. Ames: Iowa State University Press.
- Fuller,W.A. 2009. *Sampling statistics*. Wiley. New York.
- Gallego, F.J., Delincé J. and Carfagna, E. 1994. Two-Stage Area Frame Sampling on Square Segments for Farm Surveys. *Survey Methodology* , Vol 20, N 2: 107-115.
- Gallego, F.J.1995. Sampling Frames of Square Segments, Report. EUR 16317 EN, Office for Publications of the E.C. Luxembourg. pp. 68.
- Gallego, F.J. and Carfagna, E. 1995. Extrapolating Intracluster Correlation to Optimize the Size of Segments in an Area Frame. *Applied Statistics in Agriculture*. Manhattan. Kansas State University: 261-270.
- Gallego, F.J., Carfagna, E. and Feunette, I. 1998. Geographic Sampling Strategies and Remote Sensing. Report to EUROSTAT, F2, Agricultural Products and Fisheries.

- Gallego, F. J., Feunette, I. and Carfagna, E. 1999. Optimising the size of sampling units in an area frame. Gómez J., Soares A. and Froidevaux R. (editors). *geoENV II- Geostatistics for Environmental Applications*. Kluwer Academic Publisher.
- Groves, R. M. 1989. *Surveys errors and surveys cost*. Wiley. New York.
- Hansen, M.H, Hurwitz, W.N. and Madow, W.G. 1953. *Sample Survey Methods and Theory*. Wiley. New York.
- Hartley, H. O. 1962. Multiple Frame Surveys. *Proceedings of the Social Statistics Section. American Statistical Association*.
- Hartley, H. O. 1974. Multiple Frame Methodology and Selected Applications. *Sankhya*, Ser. C, 36, pp. 99-118. *Proceedings of the Social Statistics Section. American Statistical Association*.
- Humphreys, Ch. P. 1974. ADA Baseline Survey: Technical Appendix. Working Paper No.10. Part II. Addis Abeba: Institute of Development Research. University of Ethiopia.
- Iglesias, L. 2013. Improving the use of GPS, GIS and RS for setting up a master sampling frame. Technical report. FAO. Universidad Politécnica de Madrid.
- Instituto Nacional de Información de Desarrollo (INIDE). 2011a: IV Censo Nacional Agropecuario. Informe Final. Nicaragua
- Instituto Nacional de Información de Desarrollo (INIDE). 2011b: IV Censo Nacional Agropecuario. Aspectos Metodológicos. Nicaragua
- Jessen, R. J. 1945. The Master Sample of Agriculture: II. Design. *Journal of the American Statistical Association*, 40: 46-56.
- Jessen, R. J. 1978. *Statistical Survey Techniques*. Wiley. New York.
- Journel, A.G and Huijbregts, C.H.J. 1978. *Mining Geostatistics*. Academic Press.
- Kalton G. and Heeringa H. 2003. *Leslie Kish: Selected Papers*. Wiley. New York.
- King, A. J. 1945. The Master Sample of Agriculture: I. Development and use. *Journal of the American Statistical Association* 40: 38-45.
- Lohr, S. and Rao, J. N. K. 2000. Inference from Dual Frame Surveys. *Journal of the American Statistical Association*. Vol. 95: 271-280.
- Lohr, S. and Rao, J. N. K.. 2006. Estimation in Multiple-Frame Surveys. *Journal of the American Statistical Association*. Vol. 101: 1019-1030.
- El Ministerio de Ganadería, Agricultura y Pesca (MGAP). 2011. Encuesta Agrícola. Serie Encuestas n° 301. Dirección de Estadísticas Agropecuarias de Uruguay. Ministerio de Ganadería Agricultura y Pesca.
- Nusser, S. and Goebel, J. 1997. The National Resources Inventory: A Long-Term Multi-Resource Monitoring Programme. *Environmental and Ecological Statistics*, 4(3): 181-204.
- Oficina de Estudios Económicos y Estadísticos (OEEE). 2010. Diseño muestral y construcción del marco de muestreo de área en la Región Natural Costa. Oficina de Estudios Económicos y Estadísticos. Ministerio de Agricultura. Lima.
- Stephan, F. F. 1948. History of the uses of modern sampling procedures. *Journal of the American Statistical Association* 43: 12-39.

- Taylor, J., Sannier, C., Delincé, J. and Gallego, F.J. 1997. Regional Crop Inventories in Europe Assisted by Remote Sensing: 1988-1993. Synthesis Report. EUR 17319 EN, Office for Publications of the EC. Luxembourg.
- UNSD. 1986. National Household Survey Capability Programme. Sampling Frames and Sample Designs for Integrated Household Survey Programmes. Department of Technical Co-Operation for Development and Statistical Office. United Nations. New York.
- UNSD. 2005. Household sample surveys in developing and transition countries. ST/ESA/STAT/SER.F/96. United Nations Department of Economic and Social Affairs. Statistics Division. United Nations. New York.
- UNSD.2008. Designing Household Survey Samples: Practical Guides. ST/ESA/STAT/SER.F/98 Department of Economic and Social Affairs. Statistics Division. Studies in Methods. Series F N. 98. United Nations. New York.
- UNSD. 2009. Handbook on Geospatial Infrastructure in Support of Census Activities. Studies in Methods Series F No. 103 ST/ESA/STAT/SER.F/103. Department of Economic and Social Affairs. Statistics Division. United Nations. New York.
- Vogel, F.A.1995. The evolution and development of agricultural statistics at the United States Department of Agriculture. Journal of Official Statistics, 11:161-180.
- Yansaneh, I. 2005. An analysis of cost issues for surveys in developing and transition countries. In UNSD.Chap. XII.