Remote Sensing of Environment 184 (2016) 410–417

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

From one- to two-phase sampling to reduce costs of remote sensing-based estimation of land-cover and land-use proportions and their changes

M.C. Pagliarella [a], L. Sallustio [a,*], G. Capobianco [a], E. Conte [a], P. Corona [b], L. Fattorini [c], M. Marchetti [a]

[a] Dipartimento di Bioscienze e Territorio, Università degli Studi del Molise, Contrada Fonte Lappone snc, 86090 Pesche, IS, Italy
[b] Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Forestry Research Centre (CREA-SEL), Viale S. Margherita 80, 52100 Arezzo, Italy
[c] Department of Economic and Statistics, University of Siena, Piazza San Francesco 8, 53100 Siena, Italy

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 2 July 2016
Accepted 18 July 2016
Available online 30 July 2016

Keywords:
Design-based inference
Proportion and change estimation
Pure panel
Point sampling
Tessellation stratified sampling
One-per-stratum stratified sampling
Italy

ABSTRACT

The estimation of the proportion of land-cover and land-use classes is considered by exploiting remote sensing-based imagery. A pure-panel survey based on point sampling is adopted. An initial sample of points is selected by means of tessellation stratified sampling (TSS). The sample points are classified based on the imagery available for the years of interest to estimate land-cover or land-use proportions and their changes. To reduce the sampling effort, the initial selection of points is viewed as the first phase of sampling and a subsample of these points is selected in a second phase by means of one-per-stratum stratified sampling (OPSS). Land-use estimation at any subsequent year is then based on the classifications performed on the points of the second-phase sample. Two-phase estimators of proportions and of their changes are suggested, and their theoretical properties are derived. Presumably conservative estimators of their variances are proposed. A check of the precision lost involved when changing from one- to two-phase sampling is determined from the assessment of land use in Italy as a case study and from an artificial population generated to resemble the current land-use situation in Italy.

1. Introduction

Land-use and land-use change (LULUC) are among the most important drivers shaping ecosystems at global scale (e.g., Ellis and Ramankutty, 2008). The changes affect both ecosystem structure, e.g., habitat fragmentation, and the ability of ecosystems to provide goods and services for human wellbeing, such as biodiversity (e.g., Corona et al., 2011; Newbold et al., 2015), carbon and nitrogen retention, productivity and pollination (e.g., Haddad et al., 2015), and ecosystem services in general (e.g., Foley et al., 2005). Moreover, LULUC has been shown to reduce ecosystem resilience (Walker et al., 2004). This LULUC effect on ecosystem services provisioning has been recognized for the Mediterranean Basin, including Italy (Corona et al., 2012; Sallustio et al., 2015).

Under such a context and in the framework of decision-making processes, land-cover and land-use monitoring plays a key role as a support for land-cover and land-use policies and planning, especially those related to environmental issues (see, e.g., Marchetti et al., 2014), and are useful for orienting the current exploitation of natural resources towards new concepts such as resilience-thinking (Corona, 2016; Ellis et al., 2013). With specific regard to the forestry sector, for example, the periodic assessment of forest cover is crucial to provide benchmarks for monitoring the performance of various policies in influencing the increase or decrease in forest area over time (McRoberts et al., 2014).

Superior to land-cover and land-use mapping, the use of probabilistic sampling allows identification and quantification, quickly and at relatively small cost, of the key dynamics characterizing the landscape changes, as well as the monitoring of their impact in ecological and functional terms (Corona et al., 2012). Furthermore, the possibility to assess the standard errors, the possibility of frequent updates, and the substantial reduction of commission and omission errors suggest that the sampling approach is a valid and reliable choice for LULUC assessment over time.

To accomplish these aims, any land-cover and land-use sample survey should be planned to provide reliable estimates of land-cover and land-use proportions at any year of interest as well as of their changes over the years. Once an unbiased and efficient estimator of proportions is adopted, the very natural way to estimate the change between two years is to take the difference of the proportion estimates at the two years. In this case, the variance of the change estimator obviously decreases as the covariance of the two estimators increases. Thus, irrespective of the criterion adopted to estimate proportions, a convenient way to plan land-cover and land-use surveys is to maintain the same sample over the years so that a positive covariance between the

* Corresponding author.
E-mail address: lorenzo.sallustio@unimol.it (L. Sallustio).

http://dx.doi.org/10.1016/j.rse.2016.07.027
0034-4257/

estimators of proportions is induced. Many land-cover and land-use surveys adopt this solution, which is usually referred to as the *pure panel* (e.g., the U.S. Land Cover Trends Project described in Stehman et al., 2003).

Remote sensing-based imagery is essential in estimating land-cover and land-use proportions and their changes when large-scale ground surveys cannot be performed periodically because of high costs but are mandatory for constructing historical records (Corona et al., 2007; Sannier et al., 2014, 2016). For these reasons, several remote sensing-based land-cover and land-use surveys have been implemented at the global (e.g., http://www.fao.org/forestry/fra/remotesensing/en/), continental (e.g., LUCAS, the Land Use/Cover Area frame statistical Survey in Europe: http://ec.europa.eu/eurostat/web/lucas/overview) and country (e.g., Sannier et al., 2014; the AGRIT project, http://www.itacon.it) levels. It is worth noting that all of these surveys constitute pure panels because they maintain the same sample locations over the years.

In Italy, to implement the national greenhouse gas assessment under the Kyoto Protocol framework, the Ministry of Environment and Protection of Land and Sea promoted and carried out in 2008 a land-use pure-panel survey (IUTI, from the Italian acronym of *Inventario dell'Uso delle Terre d'Italia*) based on intensive point sampling. A spatially balanced sample of approximately 1,200,000 points was selected by means of *tessellation stratified sampling* (TSS). The sample points were photo-interpreted on the very accurate resolution imagery available for the years 1990 and 2008 to estimate land-use proportions at both years and their changes over time (Corona et al., 2012). The current need to repeat the survey and reduce survey costs has made the reduction of the sampling effort unavoidable. To this end, adopting the terminology of Särndal et al. (1992, Chapter 9), the initial selection of points is viewed as the first phase of sampling, and a spatially balanced sample of these points has been selected in a second phase by means of a finite-population sampling scheme usually referred to as *one-per-stratum stratified sampling* (OPSS). Therefore, land-use estimation at any future year will be based on remotely sensed imagery interpretations performed on the points of the second-phase sample. In this way, a positive covariance is induced not only between any pair of future estimates but also between the 1990 and 2008 estimates and the future estimates.

The present study pursues two objectives. First, starting from IUTI as the inspiring situation, the present study proposes a two-phase sampling strategy for achieving statistically sound estimators of land-cover and land-use proportions and their changes at a large scale. The basic idea is to use OPSS for sub-sampling from the first-phase sample in a manner that retains the advantageous property of spatial balance ensured in the first phase by TSS. The second objective of the present study is the assessment of the precision sacrificed by the second phase, based on IUTI.

Section 2.1 describes the first phase of IUTI carried out by means of TSS, and the statistical properties of the one-phase estimators are emphasized. Section 2.2 proposes a two-phase strategy in which a second-phase sample is selected from the first-phase sample by means of OPSS. The scheme has a long standing in statistical literature (e.g., Breidt, 1995) and constitutes a very simple method of achieving spatial balance. The statistical properties of the two-phase estimators under OPSS are reported, whereas the derivations are allotted in the Supplementary Material file. The statistical properties of the two-phase estimators can also be extended to any other second-phase probabilistic sampling and are reported in the same file. Section 2.3 is devoted to the assessment of the proposed two-phase strategy in the IUTI framework. Investigations are performed on the basis of the first-phase sample of points classified using on-screen interpretation of the 2008 imagery as well as on the basis of an artificial population roughly resembling the current Italian land-use scenario. The investigation results are given in Section 3. An application to the estimation of land-use proportions at the year 2013 and their changes from the year 1990 is reported in Section 4. Final comments are provided in Section 5.

2. Materials and methods

2.1. The IUTI one-phase sample survey

To carry out IUTI implementation, the Italian territory was covered by a network of 25-ha quadrats, each of them containing at least a portion of the national territory. Then, one point was randomly located within each quadrat in accordance with TSS (Fattorini et al., 2004). In contrast with the *uniform random sampling* (URS), which selects N points completely at random within the survey territory and therefore may provide an uneven survey of the study region, with points aggregated in some part of the territory and voids in other parts, TSS selects one point in each quadrat, thereby providing a representative sample of the whole study region.

Sample points, henceforth referred to as the *first-phase points*, were assigned to land-use classes by means of on-screen imagery classification in accordance with a land-use classification based on the greenhouse gas reporting system introduced by the Good Practice Guidance for Land Use, Land Use Change and Forestry (International Panel on Climate Change, 2003). The coarsest classification adopted six land-use classes: *Forest land* (1), *Cropland* (2), *Grassland* (3), *Wetland* (4), *Settlements* (5) and *Other lands* (6). Classification was performed and officially released for the years 1990 and 2008 (ISPRA, 2014) and is currently available on the Geoportale Nazionale (http://www.pcn.minambiente.it/GN/accesso-ai-servizi/servizi-di-visualizzazione-wms).

More precisely, the Italian territory was covered by a network of $N = 1{,}217{,}032$ quadrats for a total area of $Q = 30{,}425{,}800$ ha against the actual area of the Italian territory of $A = 30{,}148{,}676$ ha. The difference of 277,124 ha was the portion of network lying outside the Italian territory.

Because under TSS, points are located within the network rather than within the study area, for any land-use class k, the parameter under estimation, say φ_k, represents the fraction covered by the class with respect to the network area Q. Accordingly, under TSS, the ϕ_ks fail to sum to one. In the case of IUTI, the ϕ_ks summed to 0.9909, i.e., one minus the fraction 0.0091 of the network area outside the Italian territory, which constituted an additional class.

Henceforth, ϕ_k will be referred to as the proportion of the class k. As is customary in equal-probability point sampling (Fattorini et al., 2004), each ϕ_k is estimated by the fraction of the points falling within the class k, i.e.,

$$\bar{Y}_k = \frac{N_k}{N} = \frac{1}{N} \sum_{j \in \mathbf{U}} y_{j,k} \tag{1}$$

where $\mathbf{U}$ denotes the set of the N first-phase points, N_k is the number of points falling within the k class and $y_{j,k}$ is an indicator variable equal to 1 when the point j falls within the class k and 0 otherwise.

Under TSS, $\overline{Y}_k$ is a design-unbiased estimator of the proportion ϕ_k with design-based variance

$$\mathrm{Var}(\bar{Y}_k) = \frac{1}{N^2} \sum_{j \in \mathbf{U}} \phi_{j,k} \left(1 - \phi_{j,k}\right) \tag{2}$$

where $\phi_{j,k}$ is the proportion of the k class within the quadrat j. Because TSS selects one point per quadrat, there is no possibility of estimating the terms $\phi_{j,k}(1-\phi_{j,k})$ in Eq. (2). Thus, $\mathrm{Var}(\overline{Y}_k)$ is routinely estimated as if URS were performed by means of

$$V_k = \frac{\overline{Y}_k (1 - \overline{Y}_k)}{N - 1} \tag{3}$$

Under TSS, V_k is conservative; i.e., V_k is likely to avoid the unsuitable occurrence of overestimating the precision of the sampling strategy (Fattorini et al., 2004). From (3), $V_k^{1/2}$ is the standard error (SE)