

Chapter 5

Procedures for the Estimation of Forest Inventory Quantities



Procedura di stima delle grandezze inventariali

Maria Michela Dickson and Diego Giuliani

Abstract This chapter aims at illustrating the statistical procedures adopted to estimate the unknown values of the parameters of interest of the forest inventory. In particular, it firstly describes how the data collected during the second phase of the sampling plan have been used to estimate the areal extents of the different land use and cover categories. Secondly, it illustrates the procedures to properly estimate the total and density values of the quantities measured during the third phase of the survey campaign. These procedures were developed for INFC2005 and, as explained in this chapter, are still valid for INFC2015.

Keywords Unbiased estimators · Areal extent estimation · Total and density estimation

5.1 Introduction

In order to estimate the unknown values of the parameters of interest of the forest inventory, using the data collected through the survey campaign, it is required to apply the correct statistical procedures. In other words, it is necessary to identify the proper unbiased estimators. This chapter describes the unbiased estimators that were developed for the survey campaign of INFC2005 and explains why they can still be used for the campaign of INFC2015.

First of all, it is important to clarify which parameters of the forest inventory need to be estimated. It should be considered that the national territory is subdivided into L territorial districts (corresponding to the 21 Italian regions and autonomous provinces, called “regions” in the following text), with areal extents equal to $A_1, A_2, \dots, A_i, \dots, A_L$, respectively. The total areal extent of Italy is therefore equal to $A = \sum_{i=1}^L A_i$. Each region is characterised by $K = NF + F + 1$ land categories, where NF is the number of non-woodland categories (set equal to one in

M. M. Dickson (✉) · D. Giuliani
Department of Economics and Management, University of Trento, Trento, Italy
e-mail: mariamichela.dickson@unitn.it

INFC to simplify the estimation), F indicates the number of woodland categories, while the residual class comprises the categories that are excluded from the inventory.

The main population parameters of interest that can be estimated with the data collected during the second phase of the sampling plan are the $K \times L$ region-level areal extents of the categories $\{a_{kl} : k = 1, \dots, K, l = 1, \dots, L\}$. Aggregations of these quantities provide other interesting information, such as the areal extent of category k in the entire national territory, $A_k = \sum_{l=1}^L a_{kl}$; the areal extent of a subset C of some categories for a given region l , $a_{Cl} = \sum_{k \in C} a_{kl}$, or for the entire country, $a_C = \sum_{l=1}^L a_{Cl}$.

The main population parameters that can be estimated at the end of the third phase of the sampling plan concern the measured quantitative variables, such as the number of trees, the growing stock volume and the biomass. In particular, for any measured variable, it is possible to estimate the $F \times L$ total values t_{kl} for any combination of woodland category and region. Analogously to the a_{kl} parameters, aggregations of t_{kl} may also be of particular interest. Specifically, for a given variable, the quantity $T = \sum_{k=1}^K \sum_{l=1}^L t_{kl}$ corresponds to the overall value in the entire territory; $T_l = \sum_{k=1}^K t_{kl}$ represents the overall value in region l ; and $T_k = \sum_{l=1}^L t_{kl}$ indicates the overall value for the category k in the whole area. Moreover, for a given subset C of categories, it is also interesting to know the overall value of the variable in region l , $t_{Cl} = \sum_{k \in C} t_{kl}$, and the overall value over the entire territory, $t_C = \sum_{l=1}^L t_{Cl}$. Another set of relevant parameters is the $F \times L$ density values $d_{kl} = t_{kl}/a_{kl}$ and their aggregations, that is $d_k = t_k/a_k$, $d_l = t_l/a_l$, $d_{Cl} = t_{Cl}/a_{Cl}$ and $d_C = t_C/a_C$.

Proper estimates of all these parameters and their corresponding variances can be obtained using the estimators developed for the survey campaign of INFC2005 (Fattorini et al., 2011). Such estimators are explained in the following sections of this chapter. In particular, Sect. 5.2 provides an essential description of the estimators of areal extents, while Sect. 5.3 briefly illustrates the estimation procedures of the total and density values of quantitative variables. Section 5.4 concludes explaining why these estimators are still valid for the survey campaign of INFC2015.

5.2 Estimation of Areal Extents

According to the sampling design adopted for INFC2005, during the first phase of the survey campaign (cf. Chap. 2), a point is selected at random in each of the NQ quadrats of the sampling grid. The $N_0 \leq NQ$ selected points are then classified, according to aerial photointerpretation, into $H = NF + V + 1$ land use and cover strata, where $V < F$ strata refer to a less detailed classification of the F categories (cf. Sect. 5.1). The last residual stratum includes all the points that the aerial photointerpretation cannot properly classify.

The N_0 points are also classified according to the region, thus leading to a two-way stratification characterised by $H \times L$ strata, each denoted as U_{hl} , with size $N_{hl} (h = 1, \dots, H, l = 1, \dots, L)$.

For any stratum hl relative to the V categories and non-classifiable land category, that is where $h = NF + 1, \dots, H$, if $N_{hl} > 0$, a sample S_{hl} of n_{hl} points is selected through simple random sampling without replacement. The sampled points are then observed on the field in order to correct possible classification errors arising during the aerial photointerpretation. To the contrary, points belonging to the strata relative to NF categories, those with $h = 1, \dots, NF$, are not sampled (cf. Chap. 2).

Following Fattorini et al. (2006) an unbiased estimator of a_{kl} is

$$\hat{a}_{kl} = R \left\{ w_{kl} + \sum_{h>NF} w_{hl} w_{khl} \right\}, k = 1, \dots, NF, l = 1, \dots, L \quad (5.1)$$

or

$$\hat{a}_{kl} = R \sum_{h>NF} w_{hl} w_{khl}, k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.2)$$

where R is the overall areal extent of the NQ quadrats, $w_{hl} = N_{hl}/NQ$ is the weight of stratum hl , $w_{kl} = N_{kl}/NQ$ is the weight of stratum kl relative to the non-woodland categories that are not sampled during the second phase, and $w_{khl} = n_{khl}/n_{hl}$ is the share of points of S_{hl} belonging to land category k .

Fattorini et al. (2006) also show that a conservative estimator of the variance of \hat{a}_{kl} is

$$\hat{v}(\hat{a}_{kl}) = \frac{R^2}{NQ - 1} \left\{ w_{kl} + \sum_{h>NF} \frac{N_{hl} - 1}{n_{hl} - 1} w_{hl} w_{khl} - \sum_{h>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl}^2 - p_{kl}^2 \right\},$$

$$k = 1, \dots, NF, l = 1, \dots, L \quad (5.3)$$

or

$$\hat{v}(\hat{a}_{kl}) = \frac{R^2}{NQ - 1} \left\{ \sum_{h>NF} \frac{N_{hl} - 1}{n_{hl} - 1} w_{hl} w_{khl} - \sum_{h>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl}^2 - p_{kl}^2 \right\},$$

$$k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.4)$$

where $p_{kl} = \hat{a}_{kl}/R$.

In applying Eqs. 5.3 and 5.4, it is necessary that if $N_{hl} > 1$ then $n_{kl} \geq 2$ while if $N_{hl} = 1$ then $n_{kl} \geq 1$.

In order to estimate the variances of aggregations of \hat{a}_{kl} , such as $\hat{A}_k = \sum_{l=1}^L \hat{a}_{kl}$, the covariances among the estimates involved in the aggregations are also needed. According to Fattorini et al. (2006), the covariance between \hat{a}_{kl} and $\hat{a}_{k'l}$ can be properly estimated with

$$\hat{c}(\hat{a}_{kl}, \hat{a}_{k'l}) = -\frac{R^2}{NQ-1} \left\{ \sum_{k>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl} w_{k'hl} + p_{kl} p_{k'l} \right\},$$

$$k' \neq k = 1, \dots, K, l = 1, \dots, L \quad (5.5)$$

while the covariance between \hat{a}_{kl} and $\hat{a}_{k'l'}$ should be estimated using

$$\hat{c}(\hat{a}_{kl}, \hat{a}_{k'l'}) = \frac{\hat{a}_{kl} \hat{a}_{k'l'}}{NQ-1}, k, k' = 1, \dots, K, l' \neq l, \dots, L \quad (5.6)$$

Although Eqs. 5.1 and 5.2 represent unbiased estimators of the areal extents for all the combinations of land category and region, the sum of the $K \times L$ estimates \hat{a}_{kl} is not precisely equal to A . It is indeed equal to R times the share of points collected during the first sampling phase that fall within the borders of the national territory. Analogously, for each kl stratum, $\sum_{k=1}^K \hat{a}_{kl}$ is not equal to A_l since it corresponds to R times the share of points falling in the region l .

Since the values of A and A_l are known, the problem can be solved by calibrating the \hat{a}_{kl} -values so that the territorial totals correspond to the actual values. The calibration can be done using the following calibration factor,

$$p_{kl}^{cal} = \hat{a}_{kl} / \sum_{k=1}^K \hat{a}_{kl}, l = 1, \dots, L \quad (5.7)$$

which is characterised by the fact that $\sum_{k=1}^K p_{kl}^{cal} = 1$ for any region l .

The calibrated estimates of areal extents can then be obtained as

$$\hat{a}_{kl}^{cal} = A_l \times p_{kl}^{cal}. \quad (5.8)$$

The formulas to estimate the variances and covariances for \hat{a}_{kl}^{cal} can be found in Fattorini et al. (2006).

5.3 Estimation of Total and Density Values of Quantitative Variables

During the third phase of the sampling procedure, for any of the $(V+1) \times L \times F$ second phase stratum concerning a Forest category, if $n_{khl} > 0$, a sample Q_{khl} of m_{khl} points is selected through simple random sampling without replacement. Each sampled point is then observed on the field and the amount of any variable of interest is measured within a circular area centred on it (cf. Chaps. 2 and 4). According to Fattorini et al. (2006), the obtained data can then be used to estimate t_{kl} , for any variable, through the following unbiased estimator

$$\hat{t}_{kl} = NQ \sum_{h>NF} w_{hl} w_{khl} \bar{x}_{khl}, k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.9)$$

where \bar{x}_{khl} is the sample mean of the Horvitz-Thompson total estimates of the variable observed in the sample points of stratum khl .

Fattorini et al. (2006) also provide conservative estimators of the variances and covariances of \hat{t}_{kl} , that is

$$\begin{aligned} \hat{v}(\hat{t}_{kl}) = \frac{NQ^2}{NQ-1} & \left\{ \sum_{h>NF} w_{hl} \frac{N_{hl}-1}{n_{hl}-1} w_{khl} (n_{khl}-1) \frac{s_{khl}^2}{m_{khl}}, \right. \\ & + \sum_{h>NF} w_{hl} \frac{N_{hl}-1}{n_{hl}-1} w_{khl} (1-w_{khl}) \bar{x}_{khl}^2 + \sum_{h>NF} w_{hl} w_{khl}^2 \bar{x}_{khl}^2 - \bar{x}_{hl}^2 \left. \right\} \\ k = NF + 1, \dots, K, l = 1, \dots, L, \end{aligned} \quad (5.10)$$

$$\begin{aligned} \hat{c}(\hat{t}_{kl}, \hat{t}_{k'l}) = -\frac{NQ^2}{NQ-1} & \left\{ \sum_{k>NF} \frac{N_{hl}-n_{hl}}{n_{hl}-1} w_{hl} w_{khl} w_{k'hl} \bar{x}_{khl} \bar{x}_{k'hl} + \bar{x}_{khl} \bar{x}_{k'hl} \right\}, \\ k' \neq k = NF + 1, \dots, K, l = 1, \dots, L, \end{aligned} \quad (5.11)$$

and

$$\hat{c}(\hat{t}_{kl}, \hat{t}_{k'l'}) = \frac{\hat{t}_{kl} \hat{t}_{k'l'}}{NQ-1}, k, k' = NF + 1, \dots, K, l' \neq l, \dots, L \quad (5.12)$$

where s_{khl}^2 is the sample variance of the estimated values of the variable observed in the sample points of stratum khl .

In applying Eq. 5.9, it is necessary that if $n_{khl} > 1$ then $m_{khl} \geq 2$ while if $n_{khl} = 1$ then $m_{kl} \geq 1$.

The density values, d_{kl} , can be straightforwardly estimated with

$$\hat{d}_{kl} = \hat{t}_{kl} / \hat{a}_{kl}^{cal}. \quad (5.13)$$

Unfortunately, the variance of \hat{d}_{kl} , which is a ratio estimator, is intractable. Approximate unbiased estimates of the variances and covariances of (5.13) can however be obtained using the common approach of linearising the ratio using the first leading term of its Taylor series expansion (Särndal et al., 1992). The reliability and precision of these approximated estimates depend on the level of precision of \hat{t}_{kl} and \hat{a}_{kl}^{cal} . For this reason, in some circumstances the error of density estimates has not been reported.

For the survey campaign of INFC2005, the estimators described by Eqs. (5.9) and (5.10) were also modified to obtain the estimates of the total values of the quantitative variables for subsets of population units (e.g., the trees) identified by relevant qualitative attributes, such as the tree species. Let us consider M non-overlapping subsets. The total value, t_{kml} , of a quantitative variable of interest for subset m , forest category k , and region l can be properly estimated with

$$\hat{t}_{kml} = NQ \sum_{h>NF} w_{hl} w_{khl} \bar{x}_{kmhl}$$

$$k = NF + 1, \dots, K - 1, \quad m = 1, \dots, M, \quad l = 1, \dots, L. \quad (5.14)$$

The estimated variance of \hat{t}_{kml} is therefore

$$\hat{v}(\hat{t}_{kml}) = \frac{NQ^2}{NQ - 1} \left\{ \sum_{h>NF} w_{hl} \frac{N_{hl} - 1}{n_{hl} - 1} w_{khl} (n_{khl} - 1) \frac{s_{kmhl}^2}{m_{khl}} \right.$$

$$+ \sum_{h>NF} w_{hl} \frac{N_{hl} - 1}{n_{hl} - 1} w_{khl} (1 - w_{khl}) \bar{x}_{kmhl}^2$$

$$\left. + \sum_{h>NF} w_{hl} w_{khl}^2 \bar{x}_{kmhl}^2 - \bar{x}_{klm}^2 \right\}$$

$$k = NF + 1, \dots, K - 1, \quad m = 1, \dots, M, \quad l = 1, \dots, L, \quad (5.15)$$

where \bar{x}_{kmhl} and s_{kmhl}^2 are, respectively, the sample mean and sample variance of the estimated values of the variable observed in the sample points of stratum khl for the population units that belong to subset m .

Further modifications of the estimators have allowed also to provide estimates for two specific cases. On the one hand, they have been modified to estimate the areal extents of forest categories for subsets, $m = 1, \dots, M$, of sample points identified by qualitative attributes measured during the third phase of the sampling plan. On the other hand, they have been modified to estimate a_{kml} and t_{kml} in those circumstances in which subset m is identified during the second phase of the sampling plan. See Fattorini et al. (2011) for further details about these estimators.

5.4 Comparison Between the Two Forest Inventories

The presented estimation strategy has been used during the estimation process of the parameter of interest for INFC2005. Such strategy also remained valid for the estimation process of INFC2015. Indeed, the modifications occurred among the two forest inventories have no impact on the use of the estimation strategy, as explained in the following.

The sampling plan adopted for INFC2005 consisted of a three-phase structure (cf. Chap. 2). In the first phase, carried out using aerial photointerpretation, the area was divided into polygons of equal size (1 km²), from which a point was randomly selected (one from each polygon). Then the population of such selected points was divided into 21 strata corresponding to the territorial districts of the national territory (regions), and into strata corresponding to the land use and cover categories. In the second phase, carried out by surveys on the ground, a stratified sample of points was selected from the strata defined in the first phase only for land use and cover categories of interest for INCF2005. The selected sample points were assigned then to strata corresponding to the forest types. In the third phase, from each Forest stratum, an additional sample of points was selected. Then, plots were laid out around each of these points, in order to define the area on which the measurements of the variables were carried out (cf. Chap. 4).

In INFC2015, the adopted sampling plan did not suffer substantial changes. Indeed, the three-phase structure was maintained: the first was designed to classify the land use and cover; the second was aimed at definitively classifying the land use and cover and forest types in correspondence of sampling points; sought to define the areas for the survey of interest variables. Therefore, the sampling plan did not undergo any changes that altered its original design. The changes made in the INFC2015 have regarded exclusively the definition of strata and the sample sizes. Specifically:

- i. The number of points in the national territory was equal to 301,306 for INFC2005 and equal to 301,271 for INFC2015;
- ii. The classification scheme of the land use and cover used for the INFC2005 aerial photointerpretation included 12 classes/subclasses, while the scheme used for the INFC2015 included 13 classes/subclasses (including the class of non-classified points);
- iii. To estimate the INFC2015 areal extent, 3 new classes of forest inventory interest must be added to the 13 classes/subclasses, which following the aerial photointerpretation of the INFC2005, belonged to the classes/subclasses not of interest for the subsequent phases of the inventory;
- iv. The number of classes/subclasses from aerial photointerpretation, which were not of interest for the forest inventory went from seven (for INFC2005) to eight (for INFC2015);
- v. The number of classes of forest inventory interest was equal to four in the INFC2005, which became eight in the INFC2015.

As explained, the modifications occurred as a result of slightly different systems of aerial photointerpretation adopted during the INFC2015 and of changes in the classes/subclasses of interest. All this resulted in a different number of strata among INCF2005 and INCF2015, as well changes to some units from a stratum to another in the two survey waves. Nevertheless, from a methodological point of view, the sampling plan adopted during the two waves remained the same, since it was a three-phase design, with the definition of areas at each phase based on stratification. This fact confirmed the possibility of using the same estimation techniques adopted

during the INCF2005 for INCF2015, both for the estimation of areal extents and for the values of the interest variables.

The estimators used for INCF2015 were implemented in the open source software R (R Core Team, 2020), a free environment for statistical computing.

Appendix (Italian Version)

Riassunto Questo capitolo si occupa di illustrare le procedure statistiche adottate per stimare i valori incogniti dei parametri di interesse per l’inventario forestale. In particolare, viene innanzitutto descritto come i dati raccolti durante la seconda fase del piano di campionamento sono stati utilizzati per stimare le superfici delle diverse categorie di uso e copertura del suolo. In secondo luogo, vengono illustrate le procedure per stimare in maniera appropriata i valori totali e di densità delle grandezze rilevate nel corso della terza fase dell’indagine campionaria. Tali procedure sono state sviluppate in occasione dell’INFC2005 e, come spiegato in questo capitolo, sono valide anche per l’INFC2015.

Introduzione

Per stimare i valori incogniti dei parametri oggetto di interesse dell’inventario forestale, utilizzando i dati rilevati mediante l’indagine campionaria, è necessario impiegare le procedure statistiche adeguate. In altre parole, si devono identificare gli appropriati stimatori corretti. Questo capitolo descrive gli stimatori corretti che sono stati sviluppati in occasione dell’indagine campionaria per l’INFC2005 e spiega perché possono essere utilizzati anche per l’indagine riguardante l’INFC2015.

Innanzitutto è importante definire quali sono i parametri oggetto di stima dell’inventario forestale (NFI). A tale fine, si deve considerare che il territorio nazionale è suddiviso in L distretti territoriali corrispondenti alle 21 regioni e province autonome italiane, chiamate “regioni” nel testo che segue. Le superfici delle L regioni sono indicate, rispettivamente, con $A_1, A_2, \dots, A_i, \dots, A_L$. La superficie complessiva dell’Italia è dunque uguale a $A = \sum_{l=1}^L A_l$. Ciascuna regione è caratterizzata da $K = NF + F + 1$ categorie, in cui NF è il numero delle categorie non facenti parte della Superficie forestale (in INFC posto uguale a uno per semplificare il calcolo), F indica il numero di categorie facenti parte della Superficie forestale, mentre la classe residua comprende le categorie che sono state escluse dall’inventario.

I parametri principali oggetto di interesse che possono essere stimati utilizzando i dati rilevati durante la seconda fase del piano di campionamento sono le $K \times L$ superfici a livello di singola regione delle diverse categorie di copertura, qui indicate con $\{a_{kl} : k = 1, \dots, K, l = 1, \dots, L\}$. Differenti aggregazioni di queste grandezze—quali la superficie della categoria k nell’intero territorio nazionale, $A_k = \sum_{l=1}^L a_{kl}$, la superficie complessiva di un sottoinsieme C di alcune categorie in una data

regione l , $a_{Cl} = \sum_{k \in C} a_{kl}$, o nell'intero paese, $a_C = \sum_{l=1}^L a_{Cl}$ —forniscono ulteriori informazioni utili.

I parametri principali oggetto di interesse che possono essere stimati al termine della terza fase del piano di campionamento riguardano, invece, le variabili quantitative rilevate al suolo, quali il numero di alberi, il volume o la biomassa. In particolare, per ognuna di queste variabili, è possibile stimare gli $F \times L$ valori totali, t_{kl} , per tutte le combinazioni di categoria forestale e regione. Come per i parametri a_{kl} , anche le aggregazioni dei valori di t_{kl} possono essere di particolare interesse. Nello specifico, per una data variabile, la grandezza $T = \sum_{k=1}^K \sum_{l=1}^L t_{kl}$ corrisponde al valore totale nel territorio nazionale; $T_l = \sum_{k=1}^K t_{kl}$ corrisponde al valore totale nella regione l ; e $T_k = \sum_{l=1}^L t_{kl}$ indica il valore totale per la categoria k nel territorio nazionale. Per un dato sottoinsieme C di categorie, è inoltre interessante conoscere il valore della variabile nella regione l , $t_{Cl} = \sum_{k \in C} t_{kl}$, e nell'intero territorio nazionale, $t_C = \sum_{l=1}^L t_{Cl}$. Un altro insieme di parametri rilevanti è costituito dagli $F \times L$ valori di densità $d_{kl} = t_{kl}/a_{kl}$ e da aggregazioni dei medesimi, ossia $d_k = t_k/a_k$, $d_l = t_l/a_l$, $d_{Cl} = t_{Cl}/a_{Cl}$ e $d_C = t_C/a_C$.

Stime appropriate di tutti questi parametri e delle rispettive varianze possono essere ottenute utilizzando gli stimatori sviluppati per l'indagine campionaria dell'INFC2005 (Fattorini et al., 2011). Tali stimatori sono spiegati nelle sezioni successive di questo capitolo. In particolare, la Sect. 5.2 fornisce una descrizione essenziale degli stimatori delle superfici mentre la Sect. 5.3 illustra brevemente le procedure di stima dei valori totali e di densità delle variabili quantitative. Infine, la Sect. 5.4 conclude il capitolo spiegando perché questi stimatori rimangono validi anche per l'indagine campionaria dell'INFC2015.

Stima delle superfici

In base al disegno campionario adottato per l'INFC2005 durante la prima fase dell'indagine campionaria (cfr. Chap. 2), per ognuno degli NQ quadrati del reticolo, viene selezionato un punto in maniera casuale. Gli $N_0 \leq NQ$ punti selezionati sono poi classificati, mediante fotointerpretazione, in $H = NF + V + 1$ strati di uso e copertura del suolo, in cui i $V < F$ strati si riferiscono a una classificazione meno dettagliata delle F categorie di interesse forestale di cui si intende stimare la superficie (cfr. Sect. 5.1). L'ultimo strato residuo comprende tutti i punti per i quali la fotointerpretazione non ha dato esiti sufficientemente precisi per quanto riguarda l'uso e la copertura del suolo.

Gli N_0 punti sono classificati anche in base alla regione di appartenenza. Si ha dunque una doppia stratificazione caratterizzata da $H \times L$ strati, ciascuno indicato con U_{hl} e di dimensione pari a N_{hl} ($h = 1, \dots, H, l = 1, \dots, L$).

Per ogni strato hl relativo ad una delle V categorie o alla categoria residua dei punti non classificabili, ossia in cui $h = NF + 1, \dots, H$, se $N_{hl} > 0$, viene selezionato

un campione S_{hl} di n_{hl} punti mediante campionamento casuale semplice senza ripetizione. I punti campionati vengono poi visitati a terra allo scopo di correggere eventuali errori di classificazione commessi durante la fase della fotointerpretazione. I punti, invece, appartenenti agli strati con $h = 1, \dots, NF$, relativi alle categorie NFWL, non vengono campionati (cf. Chap. 2).

Secondo Fattorini et al. (2006) uno stimatore corretto di a_{kl} è

$$\hat{a}_{kl} = R \left\{ w_{kl} + \sum_{h>NF} w_{hl} w_{khl} \right\}, k = 1, \dots, NF, l = 1, \dots, L \quad (5.16)$$

oppure

$$\hat{a}_{kl} = R \sum_{h>NF} w_{hl} w_{khl}, k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.17)$$

dove R è la superficie complessiva degli NQ quadrati, $w_{hl} = N_{hl}/NQ$ è il peso dello strato hl , $w_{kl} = N_{kl}/NQ$ è il peso dello strato kl relativo alle categorie non facenti parte della Superficie forestale i cui punti non sono stati campionati durante la seconda fase e $w_{khl} = n_{khl}/n_{hl}$ è la quota di punti di S_{hl} appartenenti alla categoria k .

Fattorini et al. (2006) dimostrano, inoltre, che uno stimatore conservativo della varianza di \hat{a}_{kl} è

$$\hat{v}(\hat{a}_{kl}) = \frac{R^2}{NQ - 1} \left\{ w_{kl} + \sum_{h>NF} \frac{N_{hl} - 1}{n_{hl} - 1} w_{hl} w_{khl} - \sum_{h>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl}^2 - p_{kl}^2 \right\},$$

$$k = 1, \dots, NF, l = 1, \dots, L \quad (5.18)$$

oppure

$$\hat{v}(\hat{a}_{kl}) = \frac{R^2}{NQ - 1} \left\{ \sum_{h>NF} \frac{N_{hl} - 1}{n_{hl} - 1} w_{hl} w_{khl} - \sum_{h>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl}^2 - p_{kl}^2 \right\},$$

$$k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.19)$$

dove $p_{kl} = \hat{a}_{kl}/R$.

Nell'utilizzare gli stimatori descritti dalle Equazioni 5.18 e 5.19, se $N_{hl} > 1$ allora è necessario che $n_{kl} \geq 2$; mentre se $N_{hl} = 1$ allora deve valere che $n_{kl} \geq 1$.

Per poter stimare le varianze delle diverse aggregazioni di \hat{a}_{kl} , come ad esempio, $\hat{A}_k = \sum_{l=1}^L \hat{a}_{kl}$, è necessario ottenere anche le covarianze delle stime coinvolte nelle aggregazioni. Fattorini et al. (2006) dimostrano che la covarianza tra \hat{a}_{kl} e $\hat{a}_{k'l}$ può essere stimata in maniera appropriata con

$$\hat{c}(\hat{a}_{kl}, \hat{a}_{k'l}) = -\frac{R^2}{NQ-1} \left\{ \sum_{k>NF} \frac{N_{hl} - n_{hl}}{n_{hl} - 1} w_{hl} w_{khl} w_{k'hl} + p_{kl} p_{k'l} \right\},$$

$$k' \neq k = 1, \dots, K, l = 1, \dots, L \quad (5.20)$$

mentre la covarianza tra \hat{a}_{kl} and $\hat{a}_{k'l}$ dovrebbe essere stimata ricorrendo allo stimatore seguente,

$$\hat{c}(\hat{a}_{kl}, \hat{a}_{k'l}) = \frac{\hat{a}_{kl} \hat{a}_{k'l}}{NQ-1}, k, k' = 1, \dots, K, l' \neq l, \dots, L \quad (5.21)$$

Sebbene le Equazioni 5.16 e 5.17 rappresentino stimatori corretti delle superfici per tutte le combinazioni di categoria e regione, la somma delle $K \times L$ stime di \hat{a}_{kl} non corrisponde esattamente ad A . Tale somma, infatti, è uguale a R volte la quota di punti, rilevati durante la prima fase di campionamento, che cadono entro i confini del territorio nazionale. In modo analogo, per ciascun strato kl , $\sum_{k=1}^K \hat{a}_{kl}$ non è uguale a A_l poiché corrisponde a R volte la quota di punti che cadono nella regione l .

Dato che i valori di A e A_l sono noti, il problema può essere risolto calibrando i valori di \hat{a}_{kl} in modo tale che i totali per territorio corrispondano ai valori effettivi. Ciò può essere ottenuto utilizzando il fattore di calibrazione seguente,

$$p_{kl}^{cal} = \hat{a}_{kl} / \sum_{k=1}^K \hat{a}_{kl}, l = 1, \dots, L \quad (5.22)$$

rispetto al quale $\sum_{k=1}^K p_{kl}^{cal} = 1$ per ogni regione l .

Si possono dunque ottenere le stime calibrate delle superfici con

$$\hat{a}_{kl}^{cal} = A_l \times p_{kl}^{cal}. \quad (5.23)$$

Le formule per la stima delle varianze e covarianze di \hat{a}_{kl}^{cal} sono riportate in Fattorini et al. (2006).

Stima dei valori totali e di densità delle variabili quantitative

Durante la terza fase della procedura di campionamento, per gli strati riferibili ad una categoria del Bosco tra i $(V+1) \times L \times F$ strati di seconda fase, se $n_{khl} > 0$ allora viene selezionato un campione Q_{khl} di m_{khl} punti mediante campionamento casuale semplice senza ripetizione. Ciascun punto selezionato viene poi visitato a terra così da rilevare l'ammontare di ogni variabile di interesse entro un'area circolare centrata su di esso (cfr. Chaps. 2 e 4). Secondo Fattorini et al. (2006), i dati ottenuti possono essere utilizzati per stimare il parametro t_{kl} , per ogni variabile, utilizzando il seguente

stimatore corretto

$$\hat{t}_{kl} = NQ \sum_{h>NF} w_{hl} w_{khl} \bar{x}_{khl}, k = NF + 1, \dots, K, l = 1, \dots, L \quad (5.24)$$

dove \bar{x}_{khl} è la media cam-pio-na-ria delle stime di Horvitz-Thompson dei totali della variabile osservata nei punti campionati dello strato khl .

Fattorini et al. (2006) indicano anche quali sono gli stimatori conservativi delle varianze e covarianze di \hat{t}_{kl} , ossia

$$\begin{aligned} \hat{v}(\hat{t}_{kl}) = \frac{NQ^2}{NQ-1} & \left\{ \sum_{h>NF} w_{hl} \frac{N_{hl}-1}{n_{hl}-1} w_{khl} (n_{khl}-1) \frac{s_{khl}^2}{m_{khl}}, \right. \\ & + \sum_{h>NF} w_{hl} \frac{N_{hl}-1}{n_{hl}-1} w_{khl} (1-w_{khl}) \bar{x}_{khl}^2 + \sum_{h>NF} w_{hl} w_{khl}^2 \bar{x}_{khl}^2 - \bar{x}_{hl}^2 \left. \right\} \\ k = NF + 1, \dots, K, l = 1, \dots, L, \end{aligned} \quad (5.25)$$

$$\begin{aligned} \hat{c}(\hat{t}_{kl}, \hat{t}_{k'l'}) = -\frac{NQ^2}{NQ-1} & \left\{ \sum_{k>NF} \frac{N_{hl}-n_{hl}}{n_{hl}-1} w_{hl} w_{khl} w_{k'hl} \bar{x}_{khl} \bar{x}_{k'hl} + \bar{x}_{khl} \bar{x}_{k'hl} \right\}, \\ k' \neq k = NF + 1, \dots, K, l = 1, \dots, L, \end{aligned} \quad (5.26)$$

e

$$\hat{c}(\hat{t}_{kl}, \hat{t}_{k'l'}) = \frac{\hat{t}_{kl} \hat{t}_{k'l'}}{NQ-1}, k, k' = NF + 1, \dots, K, l' \neq l, \dots, L \quad (5.27)$$

dove s_{khl}^2 è la varianza cam-pio-na-ria dei valori stimati della variabile osservata nei punti campionati dello strato khl .

Nell'applicare l'Equazione 5.24, è necessario che se $n_{khl} > 1$ allora valga che $m_{khl} \geq 2$; mentre se $n_{khl} = 1$ allora si deve avere che $m_{kl} \geq 1$.

I valori di densità, d_{kl} , posso essere stimati direttamente con

$$\hat{d}_{kl} = \hat{t}_{kl} / \hat{a}_{kl}^{cal}. \quad (5.28)$$

Sfortunatamente, la varianza di \hat{d}_{kl} , che costituisce lo stimatore di un rapporto, è matematicamente intrattabile. Stime approssimativamente corrette delle varianze e covarianze di (5.28) possono però essere ottenute mediante l'approccio, comune in queste circostanze, che consiste nel linearizzare il rapporto utilizzando il termine del primo ordine dell'espansione in serie di Taylor (Särndal et al., 1992). L'affidabilità e la precisione di queste stime approssimate dipendono dal livello di precisione di \hat{t}_{kl}

e \hat{a}_{kl}^{cal} . Per questa ragione, in alcuni casi, l'errore della stima di densità non è stato riportato.

Per l'indagine campionaria dell'INFC2005, gli stimatori descritti dalle Equazioni (5.24) e (5.25) sono stati inoltre modificati affinché possano fornire le stime dei valori totali delle variabili quantitative per sottoinsiemi di unità della popolazione (ad es. gli alberi). Tali sottoinsiemi sono identificati da attributi qualitativi rilevanti come, ad esempio, la specie. Si considerino M sottoinsiemi non sovrapposti. Il valore totale, t_{kml} , di una variabile quantitativa di interesse per il sottoinsieme m , la categoria del Bosco k e la regione l può essere correttamente stimato con

$$\hat{t}_{kml} = NQ \sum_{h>NF} w_{hl} w_{khl} \bar{x}_{kmlh}$$

$$k = NF + 1, \dots, K - 1, \quad m = 1, \dots, M, \quad l = 1, \dots, L. \quad (5.29)$$

La varianza stimata di \hat{t}_{kml} è dunque data da

$$\hat{v}(\hat{t}_{kml}) = \frac{NQ^2}{NQ - 1} \left\{ \sum_{h>NF} w_{hl} \frac{N_{hl} - 1}{n_{hl} - 1} w_{khl} (n_{khl} - 1) \frac{s_{kmlh}^2}{m_{khl}} \right.$$

$$+ \sum_{h>NF} w_{hl} \frac{N_{hl} - 1}{n_{hl} - 1} w_{khl} (1 - w_{khl}) \bar{x}_{kmlh}^2$$

$$\left. + \sum_{h>NF} w_{hl} w_{khl}^2 \bar{x}_{kmlh}^2 - \bar{x}_{klm}^2 \right\}$$

$$k = NF + 1, \dots, K - 1, \quad m = 1, \dots, M, \quad l = 1, \dots, L, \quad (5.30)$$

dove \bar{x}_{kmlh} e s_{kmlh}^2 sono, rispettivamente, la media campionaria e la varianza campionaria dei valori stimati della variabile osservata nei punti campionati dello strato khl per le unità della popolazione che appartengono al sottoinsieme m .

Ulteriori modifiche degli stimatori hanno permesso di ottenere le stime riguardanti due casi particolari. Da un lato, gli stimatori sono stati modificati per stimare le superfici delle categorie forestali per i sottoinsiemi, $m = 1, \dots, M$, di punti di campionamento identificati da attributi qualitativi rilevati durante la terza fase del piano di campionamento. Dall'altro lato, sono stati modificati per stimare a_{kml} e t_{kml} nelle circostanze in cui il sottoinsieme m è stato identificato durante la seconda fase del piano di campionamento. Si veda Fattorini et al. (2011) per ulteriori dettagli sulla costruzione di tali stimatori.

Confronto tra i due inventari forestali

La strategia presentata è stata utilizzata durante la stima dei parametri di interesse per l'INFC2005. Tale strategia rimane valida anche per la procedura di stima dell'INFC2015. Infatti, le modifiche occorse tra i due inventari forestali non hanno avuto impatto sull'uso della strategia di stima, come di seguito esplicitato.

Il piano di campionamento adottato per l'INFC2005 è consistito in una struttura a tre fasi (cfr. Chap. 2). In una prima fase, condotta mediante fotointerpretazione, l'area è stata suddivisa in poligoni di eguale dimensione (1 km²), da cui veniva selezionato casualmente un punto da ogni poligono. La popolazione costituita dai punti selezionati è stata poi divisa in 21 strati, corrispondenti ai distretti territoriali del territorio nazionale (le regioni), e poi in strati corrispondenti alle diverse categorie di uso e copertura del suolo. Nella seconda fase, condotta mediante campionamento al suolo, un campione stratificato di punti è stato selezionato dagli strati definiti in prima fase, esclusivamente per le categorie di uso e copertura del suolo di interesse per l'INFC2005. I punti selezionati sono stati attribuiti a strati corrispondenti alle categorie forestali. Nella terza fase, da ognuno dei campioni selezionati in seconda fase e attribuiti a categorie del Bosco, è stato selezionato un campione aggiuntivo di punti. Dunque, attorno ad ognuno di questi punti è stato disegnato un insieme di aree di saggio allo scopo di definire le aree in cui effettuare la misurazione delle variabili di interesse (cfr. Chap. 4).

Il piano di campionamento adottato per l'INFC2015 non ha subito cambiamenti sostanziali. La struttura in tre fasi precedentemente messa in atto è stata mantenuta: la prima fase è stata disegnata per classificare in via preliminare l'uso del suolo; la seconda ha avuto l'obiettivo di classificare in via definitiva le categorie di uso e copertura del suolo in corrispondenza dei punti campionati, di attribuire la categoria forestale e classificare i caratteri qualitativi per la stima delle superfici; la terza ha definito le aree per la misura delle variabili quantitative di interesse. Quindi, il piano di campionamento non ha subito cambiamenti tali da alterarne la struttura e il disegno originali. I cambiamenti occorsi durante la pianificazione dell'INFC2015 hanno riguardato esclusivamente la definizione degli strati e delle dimensioni campionarie. Nello specifico:

- i. il numero di punti nel territorio nazionale è stato pari a 301,306 durante l'INFC2005 e pari a 301,271 durante l'INFC2015;
- ii. lo schema di classificazione delle categorie di uso e copertura del suolo utilizzato per la fotointerpretazione dell'INFC2005 ha incluso 12 classi/sottoclassi, mentre lo schema utilizzato per l'INFC2015 ha incluso 13 classi/sottoclassi (inclusa la classe dei punti non classificati);
- iii. per la stima delle superfici durante l'INFC2015, alle 13 classi/sottoclassi originarie sono state aggiunte 3 nuove classi di interesse inventariale, che, secondo la fotointerpretazione implementata per l'INFC2005, appartenevano a classi/sottoclassi non di interesse per le fasi successive dell'inventario, ritenute invece tali a seguito della nuova fotointerpretazione messa in atto per l'INFC2015;

- iv. il numero di classi/sottoclassi non forestali da fotointerpretazione è stato di 7 classi/sottoclassi nell'INFC2005, mentre è stato di 8 classi/sottoclassi nell'INFC2015;
- v. il numero di classi di interesse inventariale è stato pari a 4 nell'INFC2005, a fronte delle 7 dell'INFC2015.

Come visto, le modifiche sono state dovute a differenti sistemi di fotointerpretazione adottati durante l'INFC2015 e a cambiamenti nelle classi/sottoclassi di interesse. Tutto questo ha condotto ad una differente numerosità di strati definiti rispettivamente per l'INFC2005 e l'INFC2015, così come ad alcuni cambiamenti di unità da uno strato all'altro tra le due occasioni di indagine. Ciò nonostante, da un punto di vista metodologico, il piano di campionamento adottato durante i due inventari è rimasto invariato, dato che esso resta un disegno a tre fasi, con la definizione delle aree ad ogni fase basata sulla stratificazione. Questo conferma la possibilità di utilizzare per l'INFC2015 le stesse tecniche di stima adottate durante l'INFC2005, sia per quanto riguarda la stima delle superfici che per la stima dei valori delle caratteristiche di interesse.

Gli stimatori utilizzati per l'INFC2015 sono stati implementati mediante l'utilizzo del software statistico open source R (R Core Team, 2020), un ambiente software gratuito per il calcolo statistico.

References

- Fattorini, L., Marcheselli, M., & Pisani, C. (2006). A three-phase sampling strategy for large-scale multiresource forest inventories. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(3), 1–21.
- Fattorini, L., Gasparini, P., & De Natale, F. (2011). Descrizione generale delle procedure di stima. In P. Gasparini & G. Tabacchi (Eds.), *L'Inventario Nazionale delle Foreste e dei serbatoi forestali di Carbonio INFC 2005. Secondo inventario forestale nazionale italiano. Metodi e risultati* (pp. 75–81). Edagricole-Il Sole 24 Ore, Milano. ISBN 978-88-506-5394-2.
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Särndal, C. E., Svensson, B., & Wretman, J. (1992). *Model assisted survey sampling* (p. 694). Springer.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

