Comparing model-assisted estimators of structural variables in forest surveys

Ivan Sciascia[1], Matteo Garbarino[2], Giorgio Vacchiano[2], Renzo Motta

Abstract We propose a selection of common model-assisted estimators and their relative efficiencies in assessing different forest structural variables in the Italian Alps. Field measurements were coupled to remotely sensed data as ancillary information. The novelty of this approach relies in the compared estimation of several independent structural variables.

1 Introduction

Assessment of natural resources requires sampling. The problem of estimating population parameters (e.g., means or totals) from a sample applies to spatially-explicit forest assessment, where only a limited set of points among the land-total (population) can be surveyed. Expansion estimators were initially proposed, where units were weighted by the inverse of their inclusion probability.

After remote sensing had greatly increased the opportunities for spatially-explicit environmental sampling, procedures to incorporate auxiliary data were designed. This procedures adjust sampling weights by multipliers calibration factors that make the estimates agree with known totals. These calibration weights will generally result in estimates that are design consistent, and that have a smaller variance than the Horvitz-Thompson estimator.

[1] Ivan Sciascia, Dip. di Scienze Economico-Sociali e Matematico-Statistiche, Università di Torino; email: ivan.sciascia@unito.it

[2] Matteo Garbarino, Dip. Agro.Selvi.Ter., Università di Torino, email: matteo.garbarino@unito.it

[2] Giorgio Vacchiano, Dip. Agro.Selvi.Ter., Università di Torino, email: giorgio.vacchiano@unito.it

[2] Renzo Motta, Dip. Agro.Selvi.Ter., Università di Torino, email: renzo.motta@unito.it

The auxiliary and response variable can be linked by a working model. Model-assisted estimators are approximately (asymptotically) design-unbiased, and are particularly efficient if a working model is correct. Recent progresses were the development of k-NN estimators, that evolved since in a well-established but divergent research field [1], and the application of multivariate regression estimators [5]. Nonlinear / nonparametric assisting models require the auxiliary variable be known for all the population units, an approach labelled "model-calibration" by Wu and Sitter [7].

The aims of this study are: a) to provide a model-assisted calibration estimate of traditional and newly proposed forest parameters; b) to compare the efficiency of the expansion estimator to that of model-calibrated ones; c) to compute estimator efficiency across three different inclusion probabilities, in order to assess the most efficient sampling ratio in the general context of forest inventories.

2 Study area and field data sampling

Field data were collected in two watersheds (Musella; Ventina) in Lombardy region (Italy). Total land area is 1150 and 1124 ha respectively. Forests are dominated by European larch (*Larix decidua* Mill.) with Norway spruce (*Picea abies* (L.) H. Karst) as a co-dominant species at lower elevations. Mountain pine (*Pinus uncinata* Mill.) and Swiss stone pine (*Pinus cembra* L.) are more abundant at Ventina [3].

We applied a stratified random sampling (SRS) design [2], using homogeneous landscape units as a stratification variable. We sampled a total of 68 circular plots (radius = 12m); measurements included forest structure, topographic and anthropogenic variables (Tab.1). Details of sampling methods are given in Garbarino et. al [2].
The comparison of the results of the estimators proposed in this work are conditioned by the previous sampling design described in [2] for the choice of the 68 plots.

3 Calibration and model-assisted estimation

The efficiencies of the estimators were assessed on a subset of field data. The samples were extracted by using a two stage re-sampling design based on three different sampling ratios (0,2; 0,25; 0,3) for each stage.

3.1 Two-stage expansion estimator

The two-stage expansion estimator ($\hat{Y}_{TS-EXP}$) is:

$$\hat{Y}_{TS-EXP} = \frac{N}{n}\sum_{i\in c}\hat{Y}_i \quad \text{where} \quad \hat{Y}_i = \frac{M_i}{m_i}y_i \tag{1}$$

3.2 Two-stage calibration estimator

We considered the calibration procedure to obtain a set of calibrated weights according to the known of population total of the auxiliary variable.

The calibration constraint is $\sum c_i x_i = X$ where *X* is the population total of the auxiliary variable.

The two-stage calibration estimator ($\hat{Y}_{TS-C}$) is:

$$\hat{Y}_{TS-C} = \sum c_i y_i = \hat{Y}_{TS-EXP} + \hat{\beta}_C (X - \hat{X}_{TS-EXP}) \text{ where } \hat{\beta}_C = \frac{\sum w_i x_i y_i}{\sum w_i x_i^2} \quad (2)$$

3.3 Model-assisted estimator

We consider the superpopulation model: $\hat{f}_{MCi} = \beta x_i + c + \varepsilon$

The calibration constraint is: $\sum_{i \in U} \hat{f}_{MCi} = \sum_{i \in s} \hat{f}_{MCi} w_i$

The model assisted two-stage estimator ($\hat{Y}_{TS-MC}$) is:

$$\hat{Y}_{TS-MC} = \hat{Y}_{TS-EXP} + \left\{ \sum_{i \in U} \hat{f}_{MCi} - \sum_{i \in s} \hat{f}_{MCi} w_i \right\} \hat{\beta}_{MC} \text{ where } \hat{\beta}_{MC} = \frac{\sum_{i \in s} \hat{f}_{MCi} w_i y_i}{\sum_{i \in s} \hat{f}^2_{MCi} w_i} \quad (3)$$

The estimated variance of two-stage model assisted estimator is:

$$v(\hat{Y}_{TS-C}) = N^2 \frac{1 - f_1}{n} S^2_y (1 - \rho^2_{\hat{f}MCy}) + \frac{N}{n} \sum_{i \in c} M^2_i \frac{1 - f_{2i}}{m_i} s^2_{iy} (1 - \rho^2_{i\hat{f}MCy}) \quad (4)$$

4 Results and discussions

The two-stage calibration estimator proved to be the best one especially when forest canopy cover and basal area are used as response variable and elevation, solar radiation and proximity to buildings are treated as auxiliaries (Table 2).

The superpopulation models, both linear and quadratic were used to build two model-assisted estimators by using canopy cover and elevation variables. The model-assisted estimators proved to be less efficient than the calibrated one (Table 3).

The use of sampling strategies that couple remote sensing with field data allowed to improve the estimation efficiencies in design based estimations. A substantial improvement of the model-assisted efficiencies can be obtained through the development of a multivariate model.

References

1. Baffetta F., Fattorini L., Corona P. Estimation of small woodlot and tree row attributes in large-scale forest inventories. Environ. Ecol. Stat. 18, 147--167 (2009)
2. Garbarino M., Weisberg P.J., Motta R. Interacting effects of physical environment and anthropogenic disturbances on the structure of European larch *(Larix decidua Mill.)* forest. For. Ecol. Manag., 257, 1794–1802 (2009)

3. Garbarino M., Lingua E., Martinez Subira' M., Motta R. The larch wood pasture: structure and dynamics of a cultural landscape. Eur. J. Forest Res., 130, 491--502 (2011)
4. Montanari G.E., Ranalli M.G. Nonparametric model-calibration estimation in survey sampling. J. Am. Stat. Assoc., 100, 1429--1442 (2005)
5. Opsomer J.D., Breidt F.J., Moisen G.G., Kauermann G. Model-Assisted Estimation of Forest Resources With Generalized Additive Models. J. Am. Stat. Assoc., 102, 400--409 (2007)
6. Wu C. Optimal calibration estimators in survey sampling. Biometrika, 90, 937--951 (2003)
7. Wu C., Sitter R.R. A Model-Calibration approach to using complete auxiliary information from survey data. J. Am. Stat. Assoc., 96, 185--193 (2001)

Table 1: Mean values, standard errors (SE) and range values of response (Res.) and ancillary (Anc.) variables collected at Musella and Ventina sites.

Variable	***Cat.***	***Code***	***Unit***	***Mean (± SE)***	***Range***
Diameter	Res.	DBH	cm	18,16 (± 0,39)	5 - 105
Height	Res.	He	m	9,11 (± 0,19)	1,3 - 30,5
Basal area	Res.	BA	m^2	22,72 (± 0,42)	0,6 - 110,8
Canopy cover	Res.	CC	%	50,9 (± 0,52)	1 - 86
Larch prop.	Res.	LDo	%	0,70 (± 0,01)	0 - 1
Reg. density	Res.	RDe	n/ha	290,93 (± 8,03)	0 - 1238,7
Species	Res.	Sp	-	-	-
Elevation	Anc.	El	m a.s.l.	2000,22 (± 4,18)	1676,4 - 2293,4
Slope	Anc.	Sl	°	25,99 (± 0,25)	0 - 45,2
Aspect	Anc.	As	-	0,05 (± 0,02)	-1 - 1
Solar rad.	Anc.	Sr	kJ/m^2year	69685,42 (± 285,77)	40452 - 83555
Buildings pro.	Anc.	Bu	m	21,03 (± 0,27)	7,2 - 45,5
Roads pro.	Anc.	Ro	m	98,76 (± 2,05)	2,5 - 370

Table 2: Relative efficiencies (Calibrated VS Expansion) at Musella and Ventina sites.

Variable	***Aux***	***Efficiency*** $\frac{MSE_{CAL}}{MSE_{EXP}}$		
	f	**0.20**	**0.25**	**0.30**
Basal area	El	0,52	0,94	0,70
	Sr	0,53	0,90	0,66
	Bu	0,79	0,90	0,63
	Ro	0,81	1,37	0,83
Canopy Cover	El	0,28	0,21	0,31
	Sr	0,26	0,25	0,31
	Bu	0,56	0,65	0,39
	Ro	0,53	0,86	0,74

Table 3: Relative efficiencies (Model-assisted VS Calibrated) at Musella and Ventina sites.

Variable	***Model***	***Efficiency*** $\frac{MSE_{MC}}{MSE_{CAL}}$
Canopy Cover	Linear (El)	1,94
	Quadratic (El)	1,49