MODEL-ASSISTED SMALL-AREA ESTIMATION WITH AUTOMATED MODEL BUILDING FOR SWISS NATIONAL FOREST INVENTORY USING TWO-PHASE SAMPLING

M. Pulkkinen[1], J. Zell[2] and A. Lanz[2]

[1] Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Switzerland &
Laboratory of Forest Inventory, National Institute of Geographic and Forest Information IGN, France
e-mail: minna.pulkkinen@ign.fr

[2] Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Switzerland

Abstract

National Forest Inventories (NFIs) provide estimates of parameters related to the state and change of forest resources within a country. In addition to national and regional estimates, information is increasingly being demanded also for smaller areas and domains. In the Swiss NFI, the estimates are based on the measurements of permanent field plots placed on a rectangular grid over the country. In small areas, however, the current design-based post-stratified estimators are imprecise or impossible to apply due to the insufficient number of field plots. With reasonable auxiliary data available (that covary with target variable data and cover the entire country on a denser grid than target variable data), estimates for small areas can be obtained with design-based model-assisted estimators, which exploit auxiliary data via models between target variables and auxiliary variables.

We employed the model-assisted estimators introduced by Mandallaz (2008, 2013) for two-phase sampling of continuous populations in order to estimate the forest area and growing-stock volume in the Swiss municipalities (n = 2932), forest districts (n = 101) and cantons (n = 25). The first-phase auxiliary data on a 100 m × 100 m grid came from several sources (vegetation height model derived from digital aerial images; Copernicus satellite image products; digital elevation model based on a LiDAR campaign) and comprised several dozen variables (with transformations and interactions included, altogether 80–120 potential explanatory variables). The second-phase target variable data on a √2 km × √2 km sub-grid were obtained from the field plot measurements of the 4th Swiss NFI. For forest area estimation, we used a common external logistic model in all the small areas. For volume estimation, we built internal linear models individually for each small area (i) by fitting a model of a common pre-determined form (established in a separate pre-study), and (ii) by building a linear model in an automated procedure, where the explanatory variables included in the model were selected with lasso. Individual model-building data for each small area consisted of a fixed number of field plots (fulfilling certain quality criteria) closest to the area centroid and were obtained by adding plots from outside the area; the required number of plots varied according to the type of the area (municipality < forest district < canton).

The resulting model-assisted estimates were compared to the standard Swiss NFI post-stratified estimates in forest districts and cantons (where such estimates could be computed). The model-assisted point estimates were found to be close to the NFI estimates and generally more precise than the NFI estimates. Consequently, the model-assisted estimates in municipalities (for which no NFI reference estimates exist) can be considered fairly plausible. The models constructed with the automated procedure generally resulted in more precise estimates than the models of the pre-determined form. Further work is needed, however, to study the effect of the model-building settings (size and quality of model-building data, transformations and interactions of auxiliary variables, elastic net instead of just lasso for variable selection) on estimation precision. In conclusion, design-based model-assisted small-area estimation with automated model building could be developed into an operational system, to which new target parameters and updated auxiliary data sets can easily be

added, as no prior modelling effort is needed. Model-assisted estimation also exhibits a clear potential for improving precision even in areas and domains where standard estimators can be applied.

Keywords: continuous population, design-based estimation, difference estimator, lasso, regression estimator.

References

Mandallaz, D. (2008) *Sampling Techniques for Forest Inventories*. Chapman & Hall/CRC, Boca Raton, USA.

Mandallaz, D. (2013) Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, **43**, 441–449.