Unit-level Small Area Estimation of Forest Inventory with GEDI Auxiliary Information

Shaohui Zhang[1,2], Cédric Vega[3], Olivier Bouriaud[3], Sylvie Durrieu[4], Jean-Pierre Renaud[2,3]

[1]University of Eastern Finland, Yliopistokatu 7, 80130 Joensuu, Finland
Email: shaohui.zhang@uef.fi

[2]Office National des Forêts, 8 Allée de Longchamp, 54600 Villers lès Nancy, France
Email: jean-pierre.renaud-02@onf.fr

[3]The Institut National de L'information Géographique et Forestière, 14 Rue Girardet, 54000 Nancy, France
Email: {cedric.vega; olivier.bouriaud}@ign.fr

[4]UMR TETIS, INRAE, Univ Montpellier, 500 Rue Jean-François Breton, 34196 Montpellier, France
Email: sylvie.durrieu@inrae.fr

1. Introduction

National Forest Inventories (NFIs) play an important role in understanding the state of forests at the national and regional levels. Forest inventory for small territorial areas, such as municipalities, is also important for decision-makers. However, information is relatively limited at this level. As a result, developing small area estimation (SAE) approaches has gained increasing popularity in the field of forest inventory. It enables prediction of forest attributes for sub-populations using regression models based on auxiliary data commonly derived from remote sensing techniques over an area of interest (AOI). It has been reported that SAE can improve the precision of forest inventory without increasing costs (Mandallaz, Breschan and Hill 2013) and may produce reliable predictions of forest attributes locally, even when field plots are not available (Rao 2014).

Tomppo (2006) is a pioneer in the use of auxiliary data for multi-source forest inventory. Previously, common sources of auxiliary data often came from satellite-based imagery (McRoberts et al. 2007), digital aerial photogrammetry (Breidenbach et al. 2018), and airborne laser scanning (Magnussen et al. 2014). NASA's newly-launched Global Ecosystem Dynamics Investigation (GEDI) is a full waveform LiDAR instrument aboard the International Space Station (ISS). Its products consist of footprint measurements projected to cover 4% of the global land surface by the end of its mission (Dubayah et al. 2020). This will provide an unprecedented opportunity to systematically collect samples of forest information that can be used in SAE on a large scale.

The objective of this study is to explore the possibility of using GEDI auxiliary data to improve the accuracy of forest inventory for a large natural area in central France (Sologne), as well as for smaller sub-areas defined by French administrative boundaries (departments). The results will then be compared against estimates obtained from simple random sampling (SRS), to assess the efficiency of the auxiliary data.

2. Data and Methods

2.1 Study Area

Our study is based in Sologne, central France, which covers an area of approximately 6000km^2. The topography is mostly flat, with most elevations within the range of 110–250m, except the south-eastern part, where undulating terrain reaches 400m. Forests cover approximately 48% of the area and are dominated by pure broadleaved species (75.3%). Conifer and mixed stands account for 15.5% and 9.2% of the forest areas respectively. The climate is temperate Atlantic, with mean annual temperature and precipitation of 11°C and 725mm.

2.2 NFI Data

For the study area, 902 permanent NFI plots, surveyed between 2015 and 2019, were available. This five-year timeframe is routinely used in official French NFI statistics. Each plot contains detailed inventory information, including density (trees/ha), quadratic mean diameter (cm), basal area (m2/ha),

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.
Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.1941

dominant height (m), and volume (m3/ha). However, this paper focuses solely on the aspect of forest volume estimations. Details of inventory schemes and methods can be found in Hervé et al. (2014).

2.3 Auxiliary Data Sources and Processing

GEDI Level 2A products, acquired between 2019-04-22 and 2020-04-14, were used as auxiliary data. Each footprint is of 25-meter diameter and has the information of beam types (i.e., full or half power), sensitivity, geo-located elevation and height metrics.

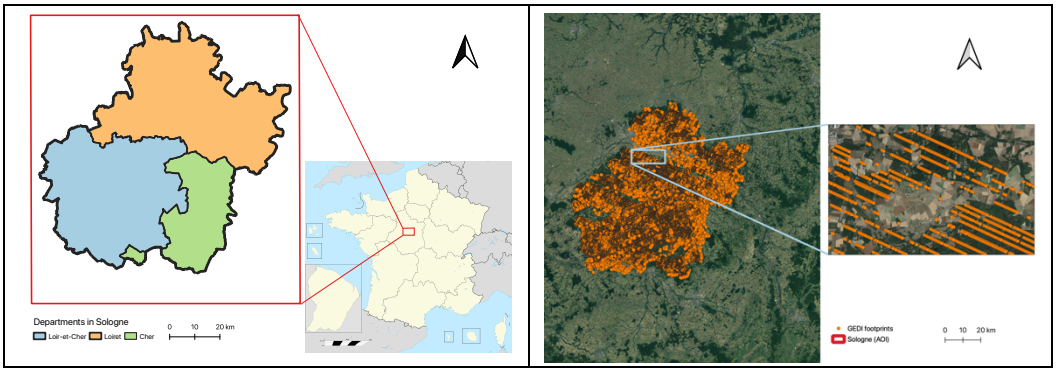

Figure 1: Location of the Study Area (departments have different colours) and availability of GEDI footprints over the AOI.

Only the best quality footprint data were retained based on the following filtering criteria: Firstly, we only kept footprints whose quality flag is 1, degrade flag is 0, and sensitivity is ≥ 0.9. This is a standard filtering criterion applied to select suitable GEDI footprints, which also ensures that the beam power is strong enough to penetrate the canopy and reach the ground with 90% probability (Hancock et al. 2019). Next, we followed the definition of forest from FAO, which states that "trees in a forest reach a minimum height of 2-5 m at maturity" (FAO 2021). Therefore, we used an average height of 3m as threshold and removed those footprints whose RH_{100} values were smaller than 3 and thus did not qualify as forest. As a result, a total number of 112,569 footprints were included for further analysis. In addition, the footprints were intersected with forest masks to determine in which types of forest they were located. Height metrics and forest types were then extracted from the footprints and formed the auxiliary data frame.

Lastly, seven nearest neighbouring GEDI footprints around individual NFI plots were identified based on their Euclidean distances. We set an additional distance threshold of 200m to filter out those NFI plots that had neighbouring footprints located farther than the threshold and thus may misrepresent the plot information. Based on the identified shot numbers, the remaining 105 NFI plots were then joined with one of the seven neighbouring footprints that shared the closest forest height, as defined by the smallest value of |NFI dominant height – RH_{100}|. This formed the calibration data frame.

2.4 Unit-level Small Area Estimation

Unit-level SAE was performed using the two-phase estimation procedure provided in R package "forestinventory" and described in Hill, Massey and Mandallaz (2021). The first phase of auxiliary GEDI information was used to generate model predictions based on linear regression using the method of ordinary least square. The second phase contains the targeted NFI plot attributes, i.e., forest volume alone in this case, that is used to generate model coefficients and correct bias.

3. Results and Discussion

A total of 101 auxiliary GEDI variables (100 height metrics and one forest type) were available and tested to predict forest volumes. Variable selection was done using an exhaustive search with the help of "randomForestSRC" R-package (Ishwaran and Kogalur 2021). The most relevant variables were manually verified to yield the best model fit and a low variance inflation factor (< 5). The final linear model retained was:

$$\hat{Y} = \beta_0 + \beta_1 * RH_{100} + \beta_2 * RH_{20} + \beta_3 * Forest\ Type \quad (1)$$

Where $\hat{Y}$ is the predicted volume, β_0 is the intercept, β_1, β_2 and β_3 represent coefficients. RH_{100} and RH_{20} represent respectively the relative heights at which the maximum and the 20th percentile of waveform energy above the ground peak are reached.

Based on this model formulation, SAE of forest volume was performed at the Sologne and sub-area levels. Results showed that GEDI auxiliary information significantly improved estimate accuracy compared with results obtained without this auxiliary information (Table 1). At the whole AOI level, the variance was significantly reduced and an increase in relative efficiency by a factor of 2.6 was obtained. In each sub-area, similar results were achieved, with reduced variance and an increase in relative efficiency by factors varying between 1.6 and 2.6. However, the mean forest volume was somehow underestimated using SAE estimation with the help of GEDI auxiliary data at both AOI and sub-area levels. We further calibrated the model and discovered that this was likely caused by model extrapolation, as a considerable number (16%) of predicted forest volumes fell outside the model calibration domain.

Table 1. Volume estimations of SAE and SRS at both AOI and department levels

Area	Plot N	SRS Estimation	SRS Variance	SAE Estimation	SAE Variance	Relative Efficiency
Overall AOI	105	192.0 ± 25.6	164.3	170.2 ± 15.8	62.4	2.6
Cher	20	198.0 ± 51.1	653.0	185.3 ± 40.2	404.0	1.6
Loiret	46	224.1 ± 43.6	475.5	181.7 ± 27.2	224.1	2.6
Loir-et-Cher	39	151.1 ± 34.8	302.2	150.6 ± 22.0	120.4	2.5

4. Conclusions

This paper performed unit-level small area estimation using GEDI Level 2A as auxiliary data. We associated GEDI auxiliary information with NFI plots based on Euclidean distance to assess forest volume estimation. It is shown that GEDI auxiliary information can help improve forest volume estimation significantly when compared to simple random sampling alone. The fact that GEDI data are open-access and cover the entire country makes it a particularly attractive tool for improving forest inventory at regional and local levels.

References

Breidenbach J et al., 2018, Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data, *Remote Sensing of Environment*, 212, 199–211.

Dubayah R et al., 2020, The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, *Science of Remote Sensing*, 1, 100002.

FAO, 2021, Comparative framework and Options for harmonization of definitions: http://www.fao.org/3/Y4171E/Y4171E10.htm. Accessed 16 March 2021.

Hancock S et al., 2019, The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions, *Earth and Space Science*, 6(2): 294–310.

Hervé JC et al., 2014, L'inventaire des ressources forestières en France: un nouveau regard sur de nouvelles forêts, *Revue Forestière Française*, (3): 247-260.

Hill A, Massey A and Mandallaz D, 2021, The R Package forestinventory: Design-Based Global and Small Area Estimations for Multiphase Forest Inventories, *Journal of Statistical Software*, 97(1): 1–40.

Ishwaran H and Kogalur UB, 2021, *RandomForestSRC: Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)*, *R package version 2.11.0*.

Magnussen S et al., 2014, National forest inventories in the service of small area estimation of stem volume, *Canadian Journal of Forest Research*, 44(9): 1079–1090.

Mandallaz D, Breschan J and Hill A, 2013, New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 43(11): 1023-1031.

McRoberts RE et al., 2007, Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery, *Remote Sensing of Environment*, 111(4): 466–480.

Rao JNK, 2014, Small-area estimation, in *Wiley StatsRef: Statistics Reference Online*, 1–8.

Tomppo E, 2006, The Finnish multi-source national forest inventory-small area estimationand map production, in *Forest Inventory – Methodology and Applications*, 195–224.