Int J Appl Earth Obs Geoinformation 84 (2020) 101959

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data

Gherardo Chirici[a], Francesca Giannetti[a], Ronald E. McRoberts[b,c], Davide Travaglini[a], Matteo Pecchi[a], Fabio Maselli[d], Marta Chiesi[d], Piermaria Corona[e]

[a] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università degli Studi di Firenze, 50145, Firenze, Italy
[b] Department of Forest Resources, University of Minnesota, Saint Paul, MN, 55108, USA
[c] Northern Research Station, U.S. Forest Service, Saint Paul, MN, 55108, USA
[d] CNR-IBE, Via Madonna del Piano, 10, 50019, Sesto Fiorentino, FI, Italy
[e] CREA, Research Centre for Forestry and Wood, viale Santa Margherita 80, 52100, Arezzo, Italy

ARTICLE INFO

Keywords:
National Forest Inventory
Spatial estimation
Growing stock
Landsat
Italy
Growing stock volume

ABSTRACT

Spatial predictions of forest variables are required for supporting modern national and sub-national forest planning strategies, especially in the framework of a climate change scenario. Nowadays methods for constructing wall-to-wall maps and calculating small-area estimates of forest parameters are becoming essential components of most advanced National Forest Inventory (NFI) programs. Such methods are based on the assumption of a relationship between the forest variables and predictor variables that are available for the entire forest area. Many commonly used predictors are based on data obtained from active or passive remote sensing technologies. Italy has almost 40% of its land area covered by forests. Because of the great diversity of Italian forests with respect to composition, structure and management and underlying climatic, morphological and soil conditions, a relevant question is whether methods successfully used in less complex temperate and boreal forests may be applied successfully at country level in Italy.

For a study area of more than 48,657 km^2 in central Italy of which 43% is covered by forest, the study presents the results of a test regarding wall-to-wall, spatially explicit estimation of forest growing stock volume (GSV) based on field measurement of 1350 plots during the last Italian NFI. For the same area, we used potential predictor variables that are available across the whole of Italy: cloud-free mosaics of multispectral optical satellite imagery (Landsat 5 TM), microwave sensor data (JAXA PALSAR), a canopy height model (CHM) from satellite LiDAR, and auxiliary variables from climate, temperature and precipitation maps, soil maps, and a digital terrain model.

Two non-parametric (random forests and k-NN) and two parametric (multiple linear regression and geographically weighted regression) prediction methods were tested to produce wall-to-wall map of growing stock volume at 23-m resolution. Pixel level predictions were used to produce small-area, province-level model-assisted estimates. The performances of all the methods were compared in terms of percent root mean-square error using a leave-one-out procedure and an independent dataset was used for validation. Results were comparable to those available for other ecological regions using similar predictors, but random forests produced the most accurate results with a pixel level $R^2 = 0.69$ and $RMSE_{\%} = 37.2\%$ against the independent validation dataset. Model-assisted estimates were more precise than the original design-based estimates provided by the NFI.

1. Introduction

Forest data are essential for multiple purposes including international and national forest monitoring programs, reporting and assessing forest resource distribution (e.g. Kyoto protocol) (Corona et al., 2011; FAO, 2010), monitoring biodiversity (Chirici et al., 2012; FOREST EUROPE, 2015), improving restoration programs (FAO and UNCCD, 2015; Smith et al., 2016) and managing at local scales to improve decision-making processes, silvicultural measures, harvesting and conservation activities.

Usually, in the context of international and national programs, this type of data is collected using sample-based National Forest Inventories (NFIs) that are designed to provide aggregated estimates of forest parameters such as forest area, growing stock volume, biomass, increments at national and regional levels (Brosofske et al., 2014; Kangas et al., 2018). These aggregated statistics are essential to support decision-making processes and to develop strategies over large areas only, because they just provide limited explicit geographic spatial detail, such as large sub-national regions. In these traditional NFIs, remote sensing is used for purposes such as initial stratification of sampling units

https://doi.org/10.1016/j.jag.2019.101959
Received 17 May 2019; Received in revised form 8 August 2019; Accepted 2 September 2019
Available online 03 October 2019

according to their main land uses, most commonly through the use of fine resolution remotely sensed imagery (McRoberts et al., 2009, McRoberts et al., 2010a,b; Corona, 2010).

In countries characterized by longer NFI traditions and/or stronger interests in the operational implementation of sustainable forest management practices such as in Sweden, Finland, Denmark (Næsset et al., 2004; Nord-Larsen and Schumacher, 2012; Tomppo et al., 2008), Canada (Boudreau et al., 2008; Matasci et al., 2018), Austria (Hollaus et al., 2009) and Switzerland (Waser et al., 2017, 2015), traditional inventories are now integrated with a more advanced use of remote sensing technology for mapping forest variables (McRoberts and Tomppo, 2007).

Most frequently these methods are applied to construct wall-to-wall spatial estimates of forest variables such as growing stock volume (Nilsson et al., 2017; Nord-Larsen and Schumacher, 2012), biomass (Nord-Larsen and Schumacher, 2012), forest cover (Waser et al., 2015), or forest changes (Næsset et al., 2013).

Wall-to-wall forest mapping in these modern forest inventories, sometimes characterized as Enhanced Forest Inventories (EFI) (Stinson and White, 2018), is considered an essential component of the forest inventory project aimed at producing forest parameter estimates at multiple spatial scales: traditional aggregated statistics useful for national planning, and at the same time, consistent small-area estimates for sub-national planning or even pixel-level raw data to support local forest management (Matasci et al., 2018; McRoberts et al., 2010a,b; Næsset et al., 2004; Nilsson et al., 2017; Tomppo et al., 2008; Waser et al., 2015).

The EFI approach produces a variety of benefits: it is able to provide detailed information to support decision-making and reduce the costs for numerous forest activities including silvicultural treatments (frequently in the framework of precision forestry), quantification of forest ecosystem services, wood harvesting, and conservation strategies (Kangas et al., 2018). The costs of the shift from a traditional NFI to an EFI are limited, because the major required investment, the field activity, remains the same or it may be even reduced if remote sensing is used for the optimization of the sampling strategy. Major costs may be related to the acquisition and elaboration of remotely sensed data.

Research activities carried out in the last 20 years demonstrated that 3D pulses from airborne laser scanning (ALS) are the most valuable data source for enhancing of growing stock volume and other forest structural variables estimates (Kangas et al., 2018; McRoberts et al., 2010a,b; Næsset, 2007; Nilsson et al., 2017; Montaghi et al., 2013; Nord-Larsen and Schumacher, 2012). The optimal option for the implementation of an EFI is thus the use of ALS data acquired in the same period as the field survey.

ALS acquisition is still expensive, but ALS data are useful for a vast array of applications in land planning, thus its cost can be shared among multiple stakeholders and agencies. However, wall-to-wall ALS data at country level are not yet available in several regions of Europe such as Italy (Giannetti et al., 2018b), Spain (Fernández-Landa et al., 2018), and most developing countries.

Together with ALS, or in case ALS is not available, satellite multispectral data can also be useful, with only small costs because they are nowadays available online for free. Barrett et al. (2016) reported in their review that when NFI data are linked with remotely sensed data, the most frequently used satellite systems are medium-resolution satellites with Landsat the most used. Medium-resolution satellite images (pixel size between 20 and 30 m) permit the prediction of forest variables with spatial detail relevant for forest inventories and sustainable forest management, and also as reported by Nilsson et al. (2017), for forest plans although forest agencies, forest companies, and forest owner associations would prefer as fine resolution as possible (in the range 10–30 m).

Several methods produce wall-to-wall maps of forest variables from field observations (Corona et al., 2014). Such methods are based on the assumption that a model of the relationship between the forest variables to be predicted and predictor variables that are available for the entire forest area can be constructed. These methods include both parametric (i.e. multiple linear regression, geographically weighted regression) and non-parametric (i.e. k-NN, random forests, Artificial Network Analysis) techniques (Barrett et al., 2016; Brosofske et al., 2014; Chirici et al., 2016; Moser et al., 2017) and have already been tested across different forest types and regions (Chirici et al., 2016).

All these methods have been widely applied with remote sensing-based predictors such as 3D data (from ALS data, microwave, or photogrammetry) (e.g. McRoberts et al., 2010a,b; Næsset, 2007; Nilsson et al., 2017; Nord-Larsen and Schumacher, 2012; Persson et al., 2017; Waser et al., 2017, 2015; Hobi and Ginzler, 2012; Ginzler and Hobi, 2015; Breidenbach and Astrup, 2012; Rahlf et al., 2014) or multispectral images from aerial, manned or unmanned, or satellite platforms, (e.g. Brosofske et al., 2014; Fernández-Landa et al., 2018; Matasci et al., 2018; Reese et al., 2002).

All these approaches have already become operational for boreal forests (Kangas et al., 2018), while in Mediterranean areas experiences are yet limited, most probably because wood production is economically less relevant and forest composition and structure is more complex, and thus more difficult to model.

Maselli et al. (2014) tested moderate resolution imagery from global 1 km resolution forest canopy height data from the Geoscience Laser Altimeter System (GLAS) onboard the ICES at satellite (Ice, Cloud, and land Elevation Satellite) for enhancing of growing stock volume estimates at country-level in Italy. Fernández-Landa et al. (2018) enhanced the estimates of the main forest inventory variables (i.e. stand density, basal area and growing stock volume) acquired in the Spanish NFI with Landsat images and ALS in a small study area in La Rioja (Spain). Condés and McRoberts (2017) developed an accurate method for updating NFI estimates of mean growing stock volume (m^3ha^{-1}) using models to predict annual plot-level volume change, and for estimating the associated uncertainties using four monospecific forest types and Landsat images for two study areas in Spain.

Mura et al. (2015) and Bottalico et al. (2017) used ALS for enhancing the estimates of structural diversity in different test areas in Italy (i.e. Molise, Tuscany and Sardinia) using remote sensing-base estimation, while Mura et al. (2018) used Sentinel-2 imagery to enhancing the estimates of growing stock volume for two test areas in Italy.

To our knowledge country-level experiences in Mediterranean areas have not yet been reported in the literature.

However, in Mediterranean areas there is an increasing need for wall-to-wall forest maps because these forests are considered more vulnerable to climate change scenarios and to natural and anthropogenic disturbances such as forest fires and urban sprawl (FAO, 2013; Scarascia-Mugnozza et al., 2000).

The current study aims at constructing wall-to-wall estimates of forest growing stock (GSV) for a large test area (i.e. 48,657 km^2) in central Italy by combining NFI plot data, remotely sensed and auxiliary variables. In particular, the research evaluated the most accurate imputation approach for mapping GSV at fine spatial resolution (23 × 23 m) and calculating small area estimates using a model-assisted approach. The results of this experimental test are aimed at identifying the optimal procedure for the operational GSV and biomass estimation at country-level in Italy.

2. Materials

2.1. Study area

To test possible wall-to-wall spatial estimation alternatives at country-level in Italy we selected a large region in central Italy including the whole of Tuscany and most of the Emilia-Romagna and part of the Liguria Regions for a total extent of 48,657 km^2 (Fig. 1). The area is characterized by large geographical and topographical variability from flat coastal areas, to gentle hills, to steep mountains with elevation

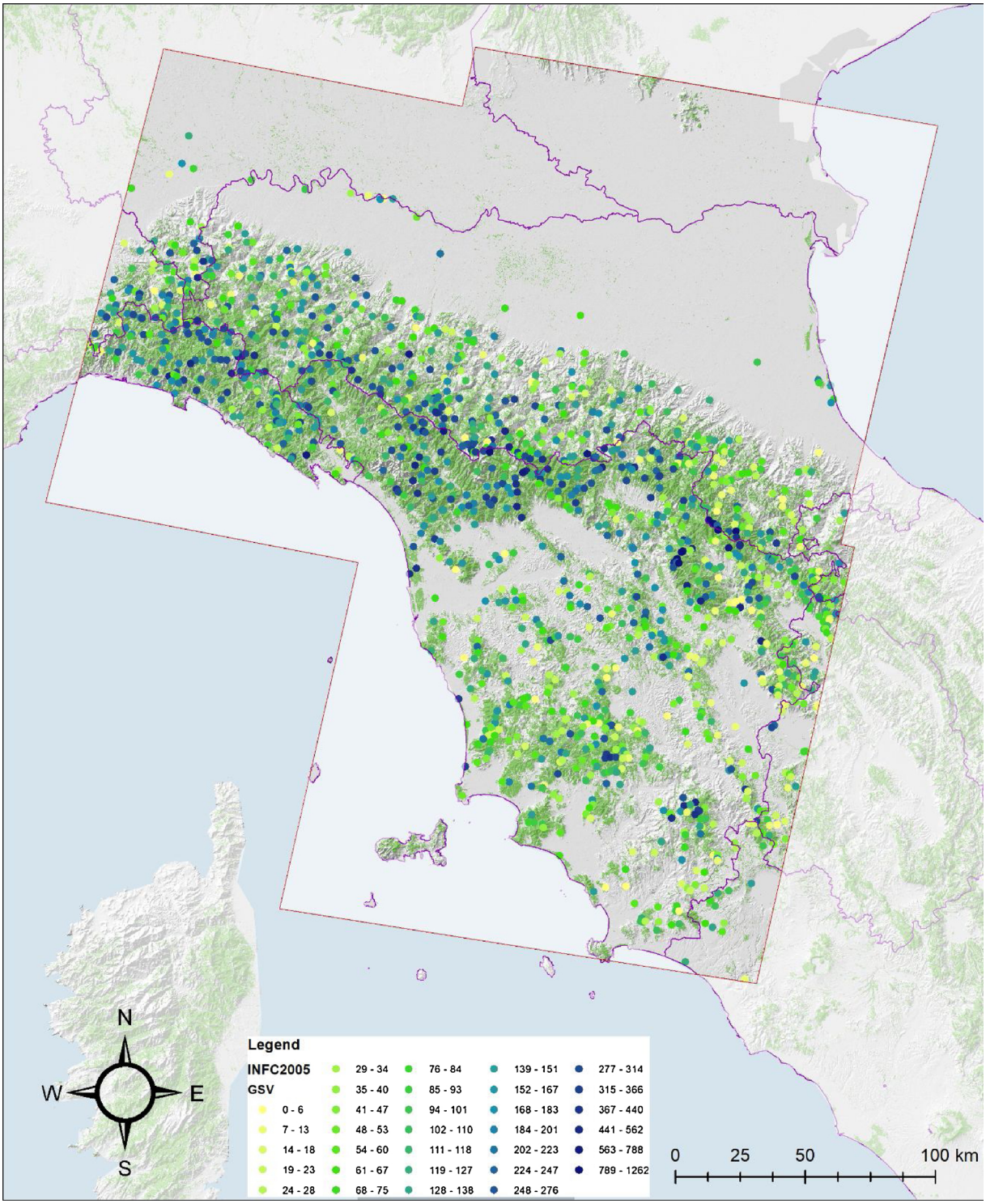


Fig. 1. Study area location (red boundary) and spatial distribution of NFI plots (INFC, 2004). Values of GSV in m^3 ha^{-1}. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

up to 2000 m a.s.l. Total precipitation per year ranges between 3000 mm in Alpi Apuane to 600 mm in the Maremma area (south of Tuscany), while mean temperature ranges from 6 °C in Abetone Mountain and Camaldoli to 17 °C along the coast.

Broadleaf species such as downy oak (*Quercus pubescens* Willd.), pedunculated oak (*Q. robur* L.), Turkey oak (*Q. cerris* L.) and sessile oak (*Q. petraea* Liebl.) (Pecchi et al., 2019) comprise 88% of the total forest area. The coppice management system is applied in 63% of the forests in the study area. Dominant coniferous species, mainly in artificial plantations, are maritime pine (*Pinus pinaster* Ait.) and black pine (*P. nigra* Arnold). Six out of the 14 European Forest Types (Barbati et al., 2014; Giannetti et al., 2018) are represented in the study area.

2.2. Italian national forest inventory data

The field reference data for the study area were acquired for 1350 plots measured in the framework of the 2nd Italian NFI (INFC, 2004) (Fig. 1) which is based on a three-phase, non-aligned, systematic

sampling design (Fattorini et al., 2006). Sampling units are located randomly within 1-km x 1-km grid cells, and in the first phase are classified on the basis of land use using aerial photos. For a subsample of the first-phase forest sampling units, qualitative information such as forest type, management, and property is collected during a terrestrial survey. For a subsample of the second-phase units, a quantitative survey is carried out in the field using circular 13 m radius plots (i.e., 530 m^2). The first two phases are aimed at estimating the forest area and classifying it into forest categories, while the third phase is aimed at collecting observations and measurements for biophysical variables. The plot data used for this study were acquired in the third phase (INFC, 2004). The plot geolocation available for this study has been the target coordinate of the sampling unit, i.e. the theoretical center of the plot that the field crew should reach. Several studies reported in the literature have evaluated the impact of inexact plot location for the estimation of forest growing stock volume or biomass. All of them relate to the use of Airborne Laser Scanning (ALS) pulses, which resulted to be very sensitive to plot location accuracy (McRoberts et al., 2018). On the other hand, in this study we predict the GSV for pixels of 23 m resolution and we expect that the error of Global Navigation Satellite System (GNSS) receivers should be much smaller than the pixel size and for this reason in this study we ignored potential positional inaccuracy of NFI plots.

For each field plot, the predicted GSV per hectare for all callipered trees is freely available online via a spatial database at https://www.inventarioforestale.org/ (Borghetti and Chirici, 2016; Pecchi et al., 2019). The GSV of each tree was predicted using species-specific allometric models developed in the framework of the NFI using tree DBH and tree height as independent variables (Tabacchi et al., 2011). The GSV per hectare of each plot was predicted as the aggregation of volume predictions for all the trees callipered in the plot. The uncertainty of allometric model predictions was considered negligible and ignored following previous results (McRoberts et al., 2016a,b). In Fig. 1 we report the spatial distribution of sample plots, while in Fig. 2 we report the GSV distribution for the 1350 field plots used in this study.

2.3. Validation data

To validate the results of our estimation we used independent field data from 332 circular plots for a different dataset, of which 297 plots of 1256.4 m^2 were measured between 2004 and 2009 to support forest management in forest areas in Tuscany and 35 are ICP level I circular plots measured in 2005 in the framework of the BioSoil Forest Biodiversity project (Galluzzi et al., 2019). The plots are representative of all forest types in the study area. The plots measured to support forest management activities are located in: Vallombrosa, Cerventosa, Lucignano, Chianti, Muraglione, Rincine and Cecina (Fig. 3).

The centers of these plots were georeferenced using a Trimble Juno 3B GNSS system and post-processed with sub-meter accuracy with the closest GNSS national base station and for each plot we applied the same field protocol developed for the Italian INFC. The GSV of each tree and the GSV per hectare of the validation plots were predicted using the same approach described in the previous paragraph for INFC plots. The GSVs of ICP BioSoil Forest Biodiversity plots were estimates using international allometric models as reported in Galluzzi et al. (2019).

The mean GSV in the validation dataset is 350.57 $m^3 ha^{-1}$, with a minimum of 6.8, a maximum of 1288.2 $m^3 ha^{-1}$ and a standard deviation of 254.79 $m^3 ha^{-1}$. The average GSV in the validation data is therefore consistently greater than the GSV registered in the INFC dataset. This was expected since the validation dataset is related to forests located in productive sites where the main forest management objective is nature and landscape conservation. This means that wood removals are generally less than the increments and the GSV tends to accumulate.

2.4. Predictor variables

The rationale for choosing the predictors is based on two elements: *i*) the availability for the whole Italy, since this test is aimed at evaluating different approaches for a country level wall-to-wall GSV spatial estimation, and *ii*) that the predictor can be at least potentially related to GSV from the results of previous investigations or from literature.

2.4.1. Remotely sensed variables

2.4.1.1. Landsat. After having evaluated other possible imagery (Chirici, 2019), to cover the study area we used imagery for three Landsat 5 Thematic Mapper (TM) scenes, 192,030 and 192,029 acquired the 23rd of June 2005, and an image for scene 193,029 acquired the 30th of June 2005. The three images are cloud-free for the forest part of the study area. Level-1 data products in Digital Numbers (DN) where transformed to top of atmosphere (TOA) radiance using

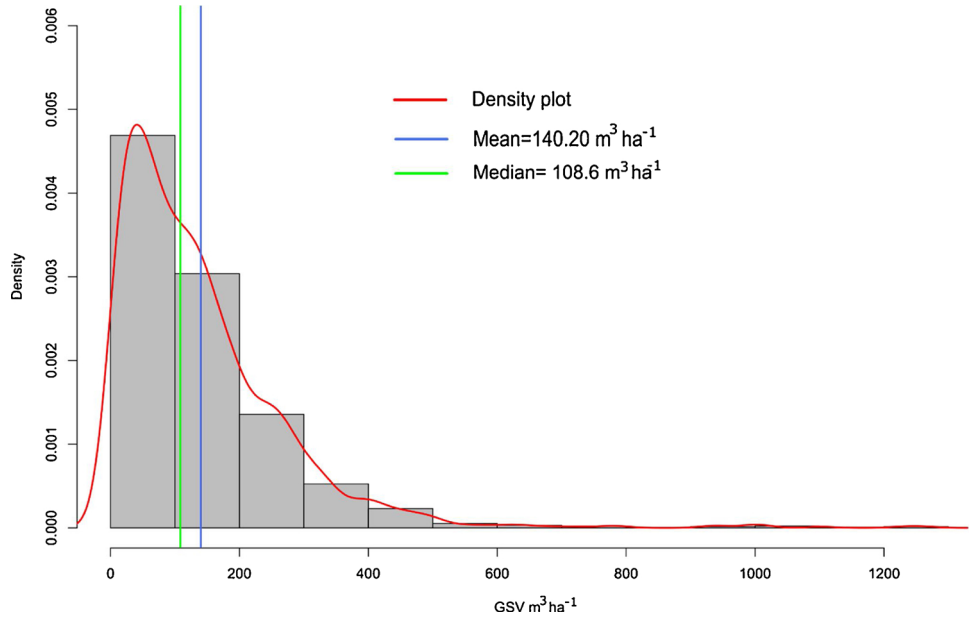


Fig. 2. GSV distribution measured in 1350 INFC plots. The red line is the density distribution, the green line is the median value and the blue line is the mean value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

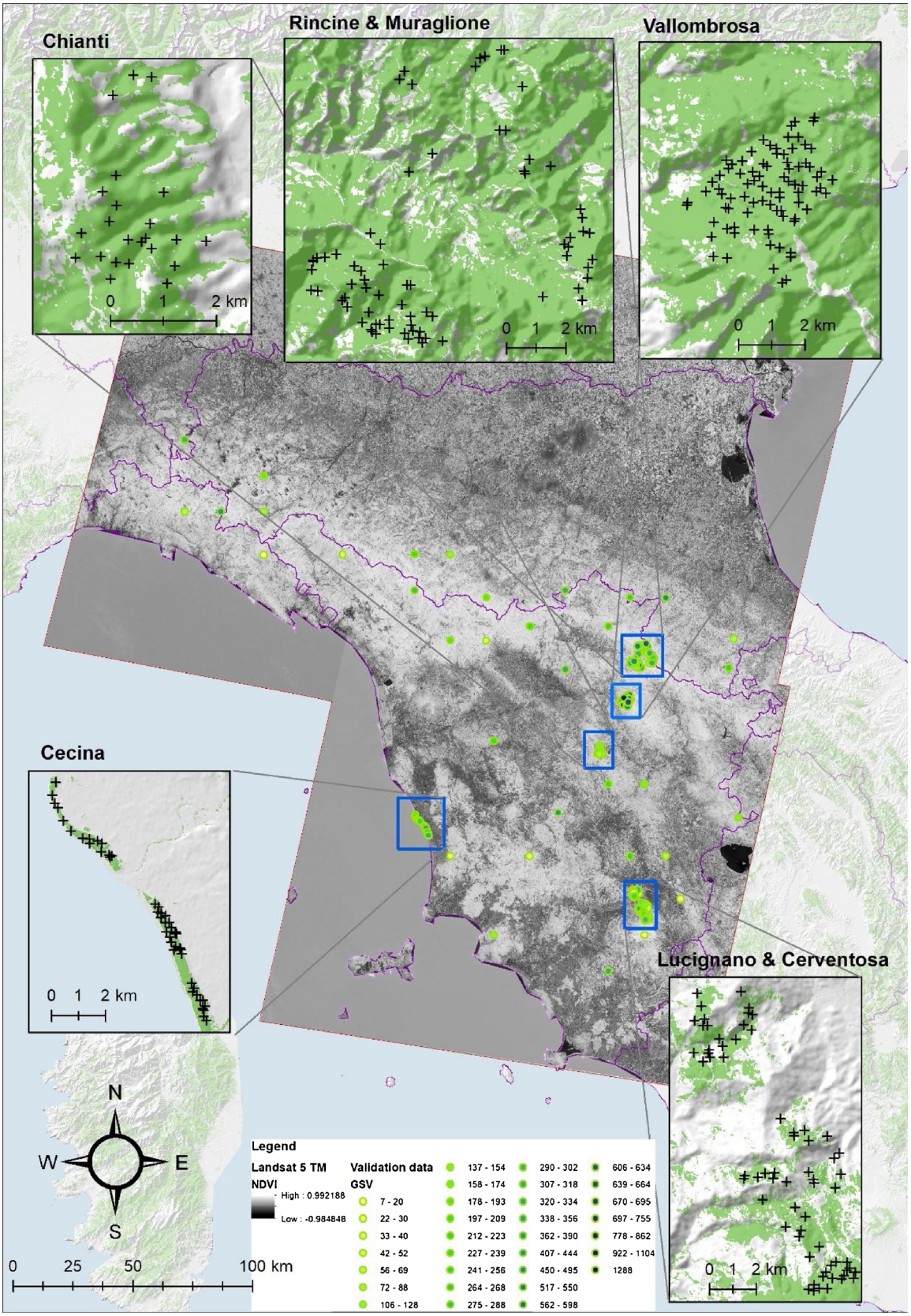


Fig. 3. Validation data used in the study on the basis of the Landsat 5 T M NDVI imagery.

Table 1
Predictors based on remotely sensing and auxiliary data used to predict GSV.

Spatial Database	Band/information	Name of predictors variables	Original spatial resolution
Landsat 5 T M	Band 1	Landsat_B1	30 m
Landsat 5 T M	Band 2	Landsat_B2	30 m
Landsat 5 T M	Band 3	Landsat_B3	30 m
Landsat 5 T M	Band 4	Landsat_B4	30 m
Landsat 5 T M	Band 5	Landsat_B5	30 m
Landsat 5 T M	Band 6	Landsat_B6	60 m
Landsat 5 T M	Band 7	Landsat_B7	30 m
Global PALSAR/PALSAR-2	HH polarization	SAR_HH	25 m
Global PALSAR/PALSAR-2	HV polarization	SAR_HV	25 m
TIN Italy	DTM	DTM	10 m
TIN Italy	SLOPE based on DTM	SLOPE	10 m
Regional land use/land cover map	Forest/non-Forest map	Forest mask	Vector 1:10.000
Climate data	Total annual precipitation	prec	1 km
Climate data	Mean annual temperature	temp_mean	1 km
Climate data	Maximum annual temperature	temp_max	1 km
Climate data	Minimum annual temperature	temp_min	1 km
European Soil Database v2.0	Subsoil available water capacity	AWC_SUB_P	1 km
European Soil Database v2.0	Topsoil available water capacity	AWC_TOP_P	1 km
European Soil Database v2.0	Volume of stones	VS_P	1 km
European Soil Database v2.0	Depth to rock	DR_P	1 km
European Soil Database v2.0	Subsoil cation exchange capacity	CEC_SUB_P	1 km
European Soil Database v2.0	Topsoil cation exchange capacity	CEC_TOP_P	1 km
European Soil Database v2.0	Soil exchange capacity	DIMP_P	1 km
Wall-to-wall Canopy Height Map	Mean Vegetation Height	CHM	1 km

radiometric rescaling coefficients provided with the Level-1 products (Fig. 3).

2.4.1.2. Global PALSAR/PALSAR-2. The SAR data used are the global 25 m resolution PALSAR-2/PALSAR mosaic available for the year 2007 as free open spatial dataset at Japanese Aerospace Exploration Agency (JAXA). Images are available as backscattering coefficient for each polarization HH and HV using the L-band Synthetic Aperture Radars (PALSAR and PALSAR-2) on Advanced Land Observing Satellite (ALOS) and Advanced Land Observing Satellite-2 (ALOS-2). The global 25 m resolution PALSAR/PALSAR-2 mosaic is processed for the geometric correction and radiometric correction to reduce topographic effects on image intensity (i.e. slope correction). The observation mode is FBD (HH, HV) and the off-nadir angle is 34.3 degrees.

2.4.2. Auxiliary variables

2.4.2.1. Digital elevation model. We used the 10 m resolution DEM TINITALY which is the finest and most accurate DEM currently available in Italy (Fornaciai et al., 2012; Tarquini et al., 2007; Tarquini and Nannipieri, 2017). TINITALY is available at http://tinitaly.pi.ingv.it/ in grid format.

2.4.2.2. Climate data. Climate data were derived from 1-km downscaled climatological surfaces released for Italy by Maselli et al. (2012). This dataset was obtained through application of geographically weighted regression to the Pan-European E-OBS database, which has a 0.25° spatial resolution (Haylock et al., 2008). The Italian dataset is representative for the period 1981–2010 and includes total annual rainfall and minimum and maximum temperatures, from which mean temperature was currently estimated. The downscaled E-OBS dataset over-estimates minimum temperature and under-estimates maximum temperature and, most importantly, rainfall (Maselli et al., 2012). For this reason we used a version of the rainfall dataset that was corrected as described in Fibbi et al. (2016).

2.4.2.3. Soil data. The soil data used were derived from the European Soil Database v2.0 (2004) (Panagos, 2006). This spatial dataset is the only geographically harmonized soil database available for Europe. It contains a soil geographic database (SGDBE) (i.e. polygons) to which a number of essential soil attributes are attached. From this database we used the quantitative information related to: (i) subsoil available water soil capacity; (ii) topsoil available water soil capacity; (iii) volume of stones; (iv) depth to rock; (v) subsoil cation exchange capacity; (vi) topsoil cation exchange capacity; (vii) soil exchange capacity.

2.4.2.4. World canopy height model. We used the vegetation height available in the wall-to-wall Canopy Height Map (Simard et al., 2011) estimated at 1-km spatial resolution from the ICESat GLAS.

2.4.3. Forest mask

A forest mask was needed to limit the spatial estimation to pixels with predicted forest land cover only. As far as possible the forest mask should mimic the same standard FAO definition used in the Italian NFI (INFC, 2004) and should be dated as close as possible to the reference year 2005 used for the acquisition of the inventory field plot data. After several tests we decided to use local fine resolution land use/land cover maps constructed at a 1:10,000 scale. We used maps from regional geoportals of Liguria, available for the year 2009 (https://geoportal.regione.liguria.it); Tuscany, available for the year 2007 (http://dati.toscana.it/dataset/ucs); and Emilia Romagna, available for the year 2008 (http://geoportale.regione.emilia-romagna.it). We rasterized the original fine resolution maps obtaining a 23 m resolution forest mask of approximately 21,327 km^2, 43% of the study area (Fig. 3).

3. Methods

Imputation methods facilitate prediction of a *response variable Y* measured for a sample of size *n* selected from a finite population of size N. X is used to denote a vector of auxiliary variables with observations for all population units.

The terminology developed for remote sensing applications in forest inventory may vary with respect to the estimation method. When regression models are used, the auxiliary variables are designated as *independent variables* and the response variable is the *dependent variable* (Mardia et al., 1979). For k-Nearest Neighbors (k-NN), the auxiliary variables are designated *feature variables* and the space defined by the feature variables is designated the *feature space;* the set of sample population units for which observations of both response and feature variables are available is designated the *reference set*; and the set of population units for which predictions of response variables are desired

is designated the *target set* (Chirici et al., 2016). For random forests, Breiman (2001) used the term *predictors* to denote the auxiliary variables.

The test area was tessellated into 23 × 23 m pixels whose size mimicked the area of the field plots measured in the field in the NFI program. All the predictors were resampled using a cubic convolution filter of 3 × 3 pixels to the final pixel of resolution of 23 m.

Thus, the population size of $N = 40{,}317{,}260$ was equal to the number of forest pixels in the study area. For each 23 × 23 m pixel a vector of 24 predictors was available from the remote sensing platforms and other auxiliary sources (Table 1). The response variable was GSV ($m^3\,ha^{-1}$) measured in the field for $n = 1350$ INFC plots and an independent validation set of $n = 332$ plots measured for forest management purposes and for the BIOSOIL project.

We tested four imputation approaches for predicting GSV. Two are non-parametric, random forests and k-NN, and two are parametric, multiple linear regression model and geographically weighted regression model. We optimized the four methods using a leave-one-out (LOO) procedure based on the 1350 NFI plots, with the most accurate approach used to predict GSV for all 40,317,260 forest pixels, hereafter characterized as estimation of the GSV map. Predictions were compared to data for the 332 plots of the independent validation set and were used for small-scale aggregated estimation with a model-assisted approach.

In the next sections, we present details for:

(i) the different imputation approaches for predicting GSV and how we optimized these methods with a LOO cross validation technique;
(ii) estimation of the GSV map applying the most accurate approach formerly identified and assessment of its accuracy using the independent validation set;
(iii) small-scale GSV estimation at study area, region (NUT-2) and province (NUT-3) levels.

3.1. Modelling methods and prediction of growing stock volume

3.1.1. Random forests

Random forests (RF) is a decision tree algorithm and nowadays is among the most popular ensemble methods for classifying and predicting forest variables. The algorithm was introduced by Breiman (1996), and its application for the spatial prediction of forest variables using remotely sensed data is well-documented (Baccini et al., 2012; Evans and Cushman, 2009; Falkowski et al., 2009; Houghton, 2007; Stumpf and Kerle, 2011; Yu et al., 2011). RF generates a set of regression trees (n_{tree}) that are aggregated to produce predictions without overfitting the data (Breiman, 2001). To build and grow trees, RF uses a randomly chosen subset of predictors at each splitting node (m_{try}), and trees are grown without the need of pruning. To grow trees, RF uses a procedure called out-of-bag samples (OOB) where each tree is built independently to arrive at the maximum size based on bootstrap samples from the training dataset (i.e., two-thirds of the data), while the remaining one-third of the sample are randomly left out. The OOB allow calculation of an OOB error rate and variable importance measured by calculating the percent increase in the mean square error when the OOB data for each variable are permuted (Breiman, 2001). The predictors that produce the most accurate splits are chosen from a random subset (m_{try}) of the entire predictor set (p).

Following the OOB sample procedure, the prediction error (OBB error) for each of the individual trees can be estimated as:

$$OOB_{error} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

where $\hat{y}_i$ is the predicted output of an OOB sample and y_i is the actual output and n is the total number of OOB sample units.

Among the 24 predictors variables (Table 1), RF was optimized for the number of predictors, n_{tree} and m_{try}. We optimized the number of predictor variables (p) to eliminate irrelevant variables. The cross validation error rate (CV_e) was calculated to assess the performance of each value of p adopted in the model with predictors being removed at each step using various m_{try} functions ($m_{try} = p, p/2, p/3, p/5, p/6 \ldots p/n$) using the same procedure described by Li et al. (2017).

RF was optimized by searching for the combination of n_{tree} and m_{try} that minimized the OOB error. More details on RF imputation can be found in the review of Belgiu and Drăgu (2016) and in the research article of Li et al. (2017). All analyses in this study were performed using the *randomForest* package within the statistical software package R 3.2.0 (Liaw and Wiener, 2002) (https://www.r-project.org).

3.1.2. k-Nearest neighbors

With the k-Nearest Neighbors (k-NN) technique, predictions are calculated as linear combinations of observations for sample units that are nearest to population units for which predictions are desired with respect to a selected distance metric in a space of feature (auxiliary) variables. Chirici et al. (2016) provided a detailed description of the k-NN method and documented more than 250 k-NN forestry applications based on remote sensing for more than 25 countries on six continents. Optimization included consideration of all possible combination of feature variables and selection of the subset that minimized RMSE. For the selected feature variables, we adopted an equal weighting approach. Simultaneously with the selection of feature variables, we searched for *i)* the optimal number of nearest neighbors, k, used for prediction between a minimum of $k = 1$ and a maximum of $k = 40$; and *ii)* the optimal distance metric among unweighted Euclidean, weighted Euclidean, and Canonical Correlation Analysis (CCA) (McRoberts et al., 2016a,b).

3.1.3. Multiple linear regression

Multiple linear regression (MLR) techniques entail the use of models of the form:

$$y_i = \beta_0 + \beta_1 \bullet x_{1i} + \bullet\bullet\bullet + \beta_p \bullet x_{pi} + \varepsilon_i \quad (2)$$

where i indexes sample units, y_i denotes the single response variable, $p \geq 1$ denotes the number of predictor variables, $j = 1, \ldots, p$ indexes the predictor variables, $\beta\ j$ is the respective regression coefficient, and ε_i denotes a random residual term assumed to be distributed $N(0, \sigma_i^2)$. The model was optimized by comparing all possible combinations of all numbers of predictors with coefficients estimated using ordinary least square. Negative GSV predictions were set to 0, and the cross-validation accuracy assessment was performed after this transformation.

3.1.4. Geographically weighted regression

Geographically Weighted Regression (GWR) is a variant of locally weighted regression, which was originally developed by Cleveland and Devlin, 1998, proposed for geographical applications by Brunsdon et al. (1996), and introduced into the remote sensing community by Maselli (2002). Mathematically, GWR entails constructing a linear regression model for each target unit by weighting the values of the reference units according to the Euclidean (geographic) distance between the target unit and the reference units used for prediction. GWR can, therefore, be easily used for forest inventory applications where reference units (plot) are regularly distributed in geographical space (Maselli, 2002).

Using the same notation as for multiple linear regression, the GWR model can be written in the form:

$$y_i = \beta_0^* + \beta_1^* \bullet x_{1i} + \bullet\bullet\bullet + \beta_p^* \bullet x_{pi} + \varepsilon_i \quad (3)$$

where β^* are the geographically weighted regression coefficients, which are estimated for each target unit from relevant statistics (mean vectors and variance-covariance matrices) computed by giving different weights to the N reference units.

A fundamental step for the application of GWR is therefore the

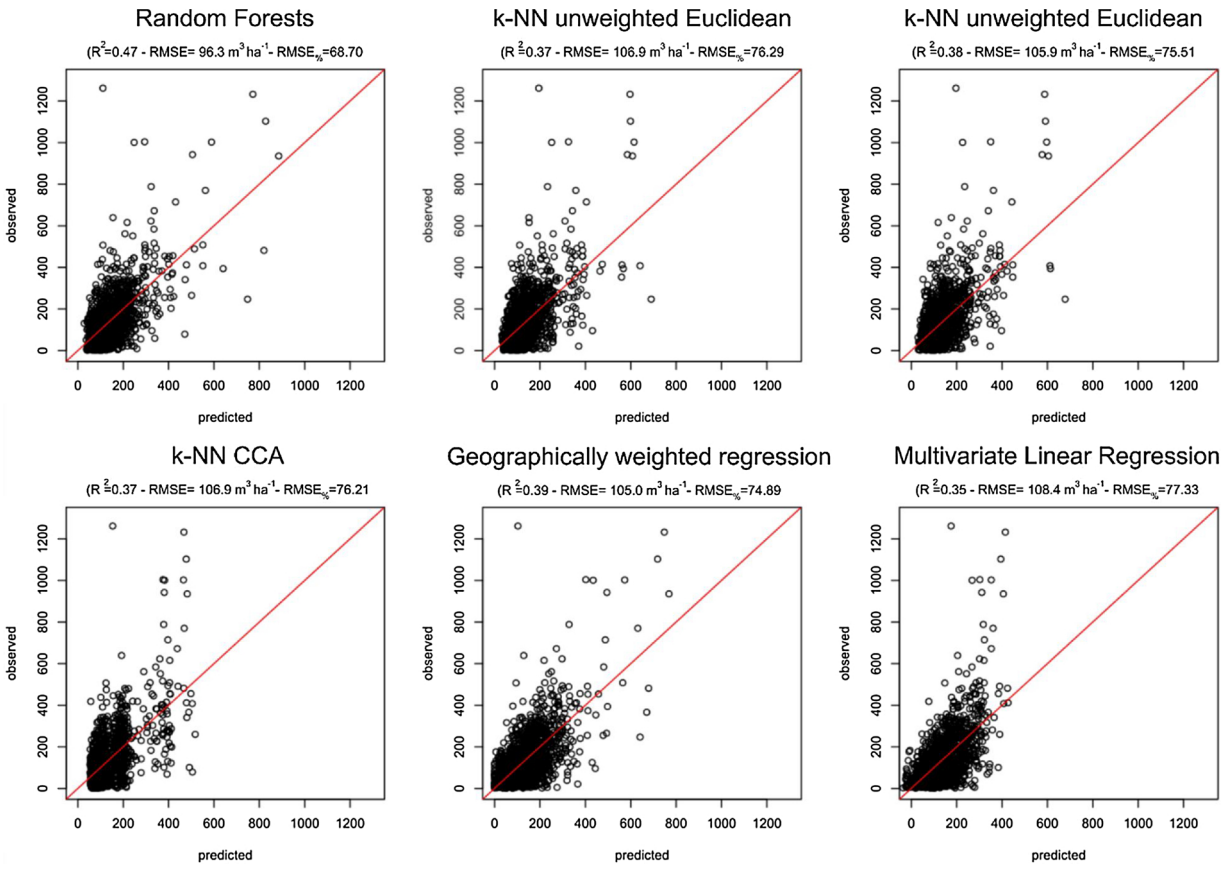


Fig. 4. Scatterplots of GSV observations versus predictions for all the imputation approaches. R^2, RMSE and $RMSE_{\%}$ are based on LOO cross-validation during the optimization phase.

definition of a suitable function to compute these weights. An efficient option is given by a negative exponential function of the spatial Euclidean Distance (ED), i.e. $exp^{(-ED/EDR)}$, which is regulated by the distance range (EDR). The model was optimized as in 3.1.3 using a LOO cross validation strategy, which also served to identify the optimum EDR (see Maselli, 2002, for details).

3.2. Model optimization

During the optimization phase the performance of the different configurations of the four imputation methods was evaluated using the LOO cross validation technique. Each reference set unit is deleted in sequence and predicted using the remaining reference set units (McRoberts et al., 2015).

For each method, we calculated the coefficient of determination (R^2) between the measured and predicted values, the root main square error (RMSE), and the relative RMSE ($RMSE_{\%}$). The RMSE was calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where n is equal to 1350 (the number of field plots), y_i is the value of the GSV observed in the field, and $\hat{y}_i$ is the predicted value of the GSV. $RMSE_{\%}$ was calculated as the percent of RMSE against the mean value of the GSV observations in the 1350 NFI plots. The optimization was finalized by selecting the most accurate method based on RMSE for the estimation phase.

3.3. Mapping and small-scale estimation

The most accurate imputation approach was used to construct a regular 23 m resolution GSV map.

We assessed the accuracy of the GSV map by comparing map unit estimates and field observations for the independent validation set of 332 plots. Again, following the same approach used in the optimization phase described in § 3.2, we estimated the coefficient of determination (R^2), the root mean square error (RMSE) and the relative RMSE ($RMSE_{\%}$). RMSE was calculated as reported in Eq. 4, where n this time is equal to 332.

To construct an inference for the mean value of the GSV for the whole study area, the model-assisted, generalized regression estimators were used (Särndal et al., 1992; Särndal et al., 2003; Breidt and Opsomer, 2009; McRoberts et al., 2016a,b). Before doing so we deleted from the GSV map all the non-forest pixels on the basis of the forest mask (§ 2.4.3).

The map-based estimate of the mean GSV in the forest area was:

$$\hat{\mu}_{map} = \frac{1}{N}\sum_{j=1}^{N}\hat{y}_i \quad (5)$$

where N was the number of 23 m x 23 m forested population units in the study area and $\hat{y}_i$ is the model prediction for the *i-th* population or map unit. However, the map-based estimate must be adjusted for systematic prediction errors using a bias estimate calculated as:

$$\hat{Bias}(\hat{\mu}_{map}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (6)$$

where n is the sample size of INFC (i.e. 1350 plots), $\hat{y}_i$ is the model prediction for the *i-th* sample INFC plot and y_i is the observed value for the *i-th* INFC plot. The model-assisted estimate is the map estimate with the estimated bias subtracted:

$$\hat{\mu}_{model-assisted} = \hat{\mu}_{map} - \hat{Bias}(\hat{\mu}_{map}) \quad (7)$$

while the standard error (SE) of $\hat{\mu}_{model-assisted}$ is:

$$SE\ (\hat{\mu}_{model-assisted}) = \sqrt{V\hat{a}r(\hat{\mu}_{model-assisted})} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(e_i - \bar{e})^2} \quad (8)$$

where $e_i = (\hat{y}_i - y_i)$ and $\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i$

In addition, to assess the efficiency of the model-assisted estimator we compared it with the original design-based estimates produced by the INFC and its relative efficiency coefficient (RE) calculated as:

$$RE = \frac{V\hat{a}r(\hat{\mu}_{NFI})}{V\hat{a}r(\hat{\mu}_{model-assisted})} \quad (9)$$

Because RE coefficient is the ratio between the variances of $V\hat{a}r(\hat{\mu}_{NFI})$ and $V\hat{a}r(\hat{\mu}_{model-assisted})$, values greater than 1 are evidence of greater precision in the model-assisted estimates (Moser et al., 2017). RE coefficient can be interpreted as the factor by which the original sample size would have to be increased to achieve the same precision as that achieved using the remotely sensed auxiliary data.

4. Results

4.1. Optimization

All the four imputation methods produced comparable results with only limited differences. Independently of the parameter used for evaluating the results, RF always achieved the greatest accuracy and MLR the least accuracy. R^2 ranged between 0.35 and 0.47; RMSE between 96.3 $m^3\,ha^{-1}$ and 108.42 $m^3\,ha^{-1}$; and $RMSE_{\%}$ between 68.70% and 77.3% (Fig. 4)

The three different k-NN configurations achieved very similar results with R^2 ranging between 0.369 and 0.382, RMSE ranging between 105.86 $m^3\,ha^{-1}$ and 106.96 $m^3\,ha^{-1}$, and $RMSE_{\%}$ ranging between 75.51% and 76.29% with $k = 21$ for the Euclidean methods and $k = 54$ for the CCA approach.

For the GWR approach we found an optimal EDR of 0.107° with performances very similar to those achieved for k-NN with R^2 of 0.396, RMSE of 105.0 $m^3\,ha^{-1}$ and $RMSE_{\%}$ of 74.89, and always more accurate than MLR.

Of the 24 available predictors considered during the optimization phase, only 15 variables were ever selected with nine predictors never selected. In terms of usefulness of the predictors, the variables derived from Landsat images were the most frequently selected; band 5 was the only one selected by all six models, followed by band 3 selected by five models. The HV polarization of radar backscattering was selected by four models, the rest of the Landsat bands were selected by three models with the exception of band 4 that was selected for two models; similar results were found for HH polarization of radar, precipitation and AWC of top soil. The other variables that were selected at least once were the average annual temperature, the maximum annual temperature, vegetation height, and the volume of rocks in the soil. In terms of number of predictors, k-NN with weighted Euclidean distance metric, k-NN with the unweighted Euclidean distance metric, and GWR all selected five; RF selected six; k-NN with the CCA distance metric selected seven, and MLR selected 10 (Table 2). The full list of the optimization results is reported in Table 2.

Considering these results RF based on six predictors and 300 regression trees was selected for the following estimation phase.

4.2. Estimation

The RF model was used to predict GSV for each of the 4,031,726 23 m x 23 m resolution forest target units in the study area (Fig. 6). GSV predictions ranged between 0 and 1021.54 $m^3\,ha^{-1}$ with a standard deviation of 70.32 $m^3\,ha^{-1}$. For each of the 332 plots in the independent validation set, we predicted GSV using RF and compared it with field observations. We found $R^2 = 0.68$ and RMSE% = 38.2%

Table 2
Parameters used for the different imputation approaches and results reported in terms of R2, RMSE, RMSE%.

Imputation	Type of imputation	Selected predictors	Optimization parameters	R^2	RMSE $m^3\,ha^{-1}$	$RMSE_{\%}$
Random Forests	Non Parametric	LANDSAT_B1 LANDSAT_B3 LANDSAT_B5 LANDSAT_B7 prec SAR_HV CHM	$n_{tree} = 300$	0.47	96.3	68.70
k-NN unweighted based on Euclidean distance	Non Parametric	LANDSAT_B1 LANDSAT_B3 LANDSAT_B5 LANDSAT_B6 AWC_TOP_P	k = 21	0.369	106.96	76.29
k-NN weighted Euclidean based on Euclidean distance	Non Parametric	LANDSAT_B1 LANDSAT_B3 LANDSAT_B5 LANDSAT_B6 AWC_TOP_P	k = 21	0.382	105.86	75.51
k-NN CCA	Non Parametric	LANDSAT_B2 LANDSAT_B4 LANDSAT_B5 LANDSAT_B7 SAR_HV SAR_HHtemp_mean	k = 54	0.370	106.88	76.23
Geographically weighted regression	Parametric	LANDSAT_B2 LANDSAT_B3 LANDSAT_B5 SAR_HV prec	Euclidean distance range (EDR) 0.107°	0.396	105.0	74.89
Multiple Linear Regression	Parametric	LANDSAT_B2 LANDSAT_B3 LANDSAT_B4 LANDSAT_B5 LANDSAT_B6 LANDSAT_B7 SAR_HV SAR_HH temp_mean VS_P		0.352	108.42	77.33

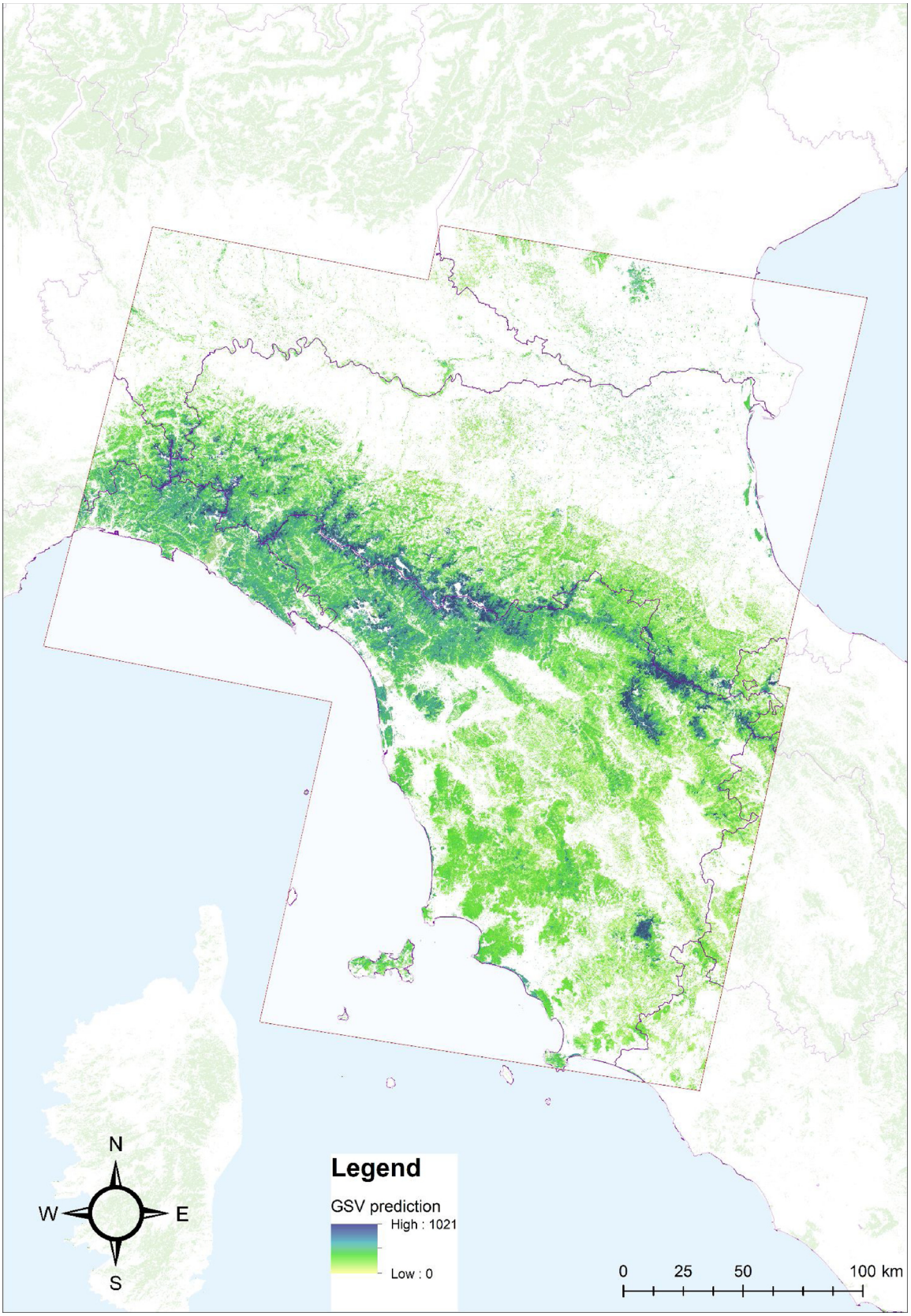


Fig. 5. Growing stock map of the study area generated with Random Forest Imputation. GSV in $m^3\,ha^{-1}$.

(Fig. 5) demonstrating a performance that was greater than achieved using LOO cross-validation.

On the basis of RF estimation for the entire study area, $\hat{\mu}_{model-assisted} = 126.17\ m^3\,ha^{-1}$ with $SE(\hat{\mu}_{model-assisted}) = 2.78\ m^3\,ha^{-1}$, while at regional level $\hat{\mu}_{model-assisted} = 131.58\ m^3\,ha^{-1}$ with $SE(\hat{\mu}_{model-assisted}) = 4.19\ m^3\,ha^{-1}$ for Tuscany, and $\hat{\mu}_{model-assisted}$ $135.42\ m^3\,ha^{-1}$ with $SE(\hat{\mu}_{GREG}) = 5.55\ m^3\,ha^{-1}$ for Emilia Romagna. These regional model-assisted estimates are in line with the official design-based estimates from INFC plots (Gasparini and Tabacchi, 2011; INFC, 2008) which are $128.8\ m^3\,ha^{-1}$ with SE = $4.6\ m^3\,ha^{-1}$ for Tuscany and 128.4 with SE = $7.12\ m^3\,ha^{-1}$ for Emilia-Romagna. These results revealed a RE of 1.09 for Tuscany Region and a RE of 1.28 for Emilia-Romagna Region.

Moreover, the model-assisted estimate of GSV was calculated at province administrative level (Annex 1). Such estimates are not provided by official NFI aggregated statistics.

5. Discussion

The study focused on three objectives: *(i)* to demonstrate that even in large complex Mediterranean landscapes, without the availability of ALS, it is possible to produce spatial wall-to-wall estimates of GSV measured in the field in the National Forest Inventory (INFC, 2004) on the basis of predictors from remotely sensed images and other auxiliary variables, (*ii*) to understand the relative importance of possible predictors available wall-to-wall in Italy and the performance of the different estimation approaches, and *(iii)* to suggest a methodology that can be applied at country level in Italy to produce wall-to-wall predictions of forest variables to support forest planning and management.

To achieve these results for a large study area of $45{,}438\ km^2$ in central Italy, we acquired 24 potential predictors which are available wall-to-wall in Italy and that may directly or indirectly be related to forest biomass and GSV. We compared six different prediction techniques, all of which comparable accuracies but with RF producing the greatest accuracy.

Among the other imputation approaches, GWR yielded the greatest accuracy, in particular outperforming conventional multiple regression. This can be explained considering that the relationships between GSV and virtually all predictors currently considered are affected by several factors which can vary spatially (Lu, 2006). GWR can account for this spatial variability by allowing the per-pixel computation of different regression models. This is particularly relevant in heterogeneous Mediterranean environments, where GWR has already been proficiently applied to Landsat TM/ETM + imagery for forest GSV prediction (Maselli and Chiesi, 2006; Maselli et al., 2014a, b).

Landsat bands of which B5 acquired in short-wave infrared between 1.55 and 1.75 μm was most important, and climate variables of which precipitation was most important, emerged as the most influential predictors. The resulting 23 m resolution GSV map, when compared against an independent set of field measures, demonstrated a good relationship between observed and predicted values ($R^2 = 0.69$ and $RMSE_{\%} = 37.2\%$). However our results are less accurate than those obtained in boreal forests using ALS in Sweden by Nilsson et al. (2017) and in the review of Næsset et al. (2004) for which RMSE usually ranged between 15% and 25% of the average real value measured in the field.

The relatively larger $RMSE_{\%}$ we obtained can be due to several reasons.

Firstly, we did not used metrics from ALS data which are usually the best candidate predictors for GSV estimation. This is confirmed if we compare our results with results reported for studies where ALS was not used. For example Reese et al. (2002), using Landsat data in Sweden, reported pixel-level $RMSE_{\%}$ in the range of 59% and 80%, and Immitzer et al. (2016) in Germany using WorldView-2 imagery report a $RMSE_{\%}$ between 46% and 37%.

Secondly, GSV is relatively small for our forests, we observed a field GSV average of $139\ m^3\,ha^{-1}$, less than half of the $287\ m^3\,ha^{-1}$ reported by Nilsson et al. (2017) in Sweden.

Thirdly, Italy has a heterogeneous landscape, and Mediterranean forests are characterized by considerable complexity in tree species composition and structure relative to temperate and boreal forests.

Moreover, we found that the accuracy of the pixel-level estimation evaluated with the independent dataset was greater than those we found with the LOO procedure in the optimization phase. The result was not expected but it is probably due to the fact that the GSV measured in the independent validation dataset has a more normal distribution around the mean values (Fig. 5) than those from the INFC (Fig. 2) and that the average GSV in the independent validation dataset is also greater ($351\ m^3\,ha^{-1}$) than those measured in INFC plots ($140\ m^3\,ha^{-1}$).

In line with previous results from the literature, we observed an underestimation for large GSV observations, independently of the prediction approach. This effect has anyhow a limited impact when the comparison was done with LOO against the INFC plots because just a few of them have very large GSV observations (Fig. 4). This saturation effect with under-predictions for plots with GSV greater than $600\ m^3\,ha^{-1}$ was well-known because spectral reflectance values are not sensitive, for example, to multilayer canopy forest or dense forests (Zhao et al., 2016). Moreover, some authors have reported that areas characterized by very complex topographic features (i.e. from flat terrain to mountains up to 2000 m a.s.l.) affect the spectral signature and the data saturation values of forest aboveground biomass and growing stock volume (Lu et al., 2012, 2016; Foody et al., 2003; Nichol and Sarker, 2011). However, the saturation effect was reported in the literature even when ALS data were used (Nilsson et al., 2017; Giannetti et al., 2018a,2018b; Lefsky et al., 2005).

Even if RF was found to be the most accurate method, only small differences in prediction accuracies were found across the different non-parametric and parametric methods. Nilsson et al. (2017) reported similar conclusions for Sweden using ALS data.

Regarding the model-assisted estimates calculated on the basis of the GSV map, with the use of our approach it was possible to increase the precision of INFC predictions at regional level (RE = 1.09 in Tuscany and RE = 1.28 in Emilia Romagna) and to provide for the first time growing stock estimates at Province level.

It is important to remember that the use of pixel level estimates of map products similar to those we presented in Fig. 6 is discouraged since GSV predictions in single pixels may be affected by a consistent bias (McRoberts and Tomppo, 2007). We therefore suggest aggregation of predictions from several pixels (Areas Of Interests – AOI), since in case the pixel prediction errors are independent and distributed with zero mean, then when the AOI increases, then averaged value of the pixels tend to equal the real value (McRoberts and Tomppo, 2007). Users could aggregate GSV pixel level estimates to create estimates for different AOIs, for example related to ecological regions, municipality boundaries, or forest management units.

6. Conclusions

Forest tree monitoring and assessment are rapidly evolving as new information needs arise and new techniques and tools become available. However, the exploitation of the latter, as well as their implementation within operative management processes, should be evidence-based (Corona, 2018).

Under this perspective, several conclusions can be drawn from the study. Firstly, Landsat data are confirmed as a reliable and efficient source of information for modeling GSV, even in large and complex Mediterranean forest areas. Secondly, we found that in the Mediterranean area, predictors derived from climate data are a valid spatial data source for modeling GSV most probably because they can describe different growing season conditions. Thirdly, all the tested modelling approaches have the capability to predict GSV with

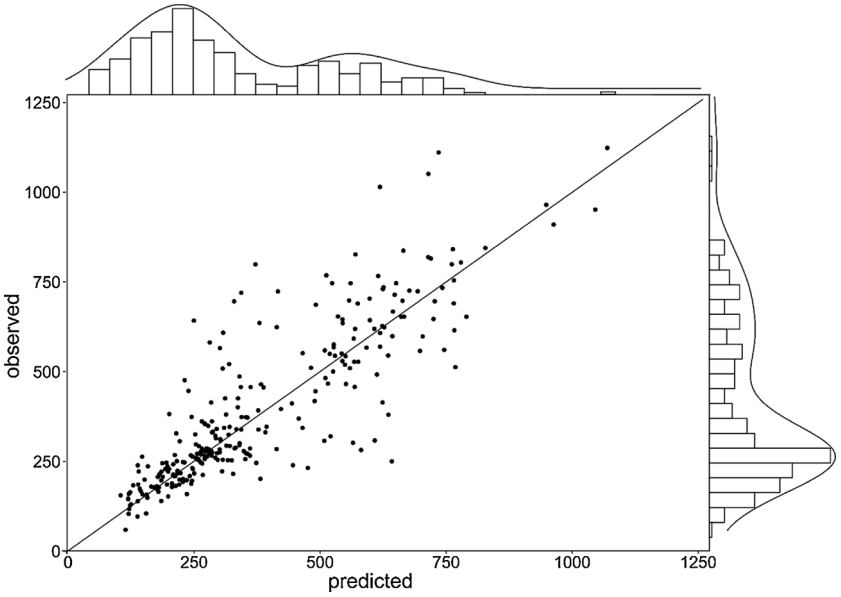


Fig. 6. Scatterplot of GSV observations versus predictions obtained by RF for the 332 units of the independent dataset.

comparable results. Fourthly, the GSV map is confirmed as a valid tool for model-assisted inference at regional and province levels.

We can affirm that the 23 m resolution GSV map we produced can be useful and practical to support the requirements of national and regional forest bodies, forest companies and forest owners. This map could be the basis for decision support systems as proposed by Puletti et al. (2017) for a test area in south Italy, as a tool to assess wood production and harvesting activities in forest proprieties, thereby contributing to improving the Mediterranean forest economy and, if used at forest management scale, reducing the cost for data acquisition needed for the implementation of management plans.

Moreover, the GSV map can be used to produce model-assisted estimates at province level (NUT-3), augmenting the spatial resolution of traditional NFI design-based estimates which are currently available only for administrative Regions (NUT-2) and thus adding value to the INFC. Under this point of view the proposed methodology is now ready for a wall-to-wall application in Italy to move the traditional NFI program to a more modern EFI, in line with achievements in other countries.

Under this point of view it is also strongly recommended that in the future the Italian NFI could evolve in a permanent monitoring system, where a sample of the total number of field plots is visited in the field every year in order to complete the revisit of all the plots in 5–10 years.

In the future we hope that ALS will be finally available wall-to-wall in Italy to facilitate prediction of forest variables estimates with even greater accuracy. In such a context satellite LiDAR data from the Global Ecosystem Dynamics Investigations (GEDI space laser data) and from the ICESAT-2 (Geoscience Laser Altimeter System - GLAS) are potentially extremely important in Italy if ALS will not be available sooner.

Appendix A

Small-scale estimates of mean GSV ($m^3\ ha^{-1}$) obtained with RF model at Province (NUT-3) level. For each Province we also report the forest area estimation from the second Italian National Forest Inventory (INFC, 2007, 2008; Gasparini and Tabacchi, 2011).

Region	Province	Province Area (km^2)	Total Forest Area (km^2) (INFC)	SE Total Forest Area (%) (INFC)	n_i	GSV $\hat{\mu}_{GREG}(m^3\ ha^{-1})$	GSV $SE(\hat{\mu}_{GREG})(\%)$
Tuscany	Arezzo	323300	1792,19	4.2	127	111.13	8.5
	Firenze	351369	1785,00	4.2	117	151.89	12.3
	Grosseto	450312	1979,61	4.0	116	98.35	8.6
	Livorno	121371	473,64	8.6	23	108.88	16.7
	Lucca	177322	1210,44	5.2	64	198.86	13.1
	Massa Carrara	115468	867,13	6.2	30	148.69	14.69
	Pisa	244472	950,53	6.0	54	98.82	11.28
	Pistoia	96412	506,40	8.3	32	214.30	43.91
	Prato	36572	233,34	12.3	13	186.87	24.1
	Siena	38298	1717,10	4.3	115	85.81	6.09
Emilia-Romagna	Bologna	370232	1007,61	5.6	56	112.60	11.12
	Forlì-Cesena	237840	1066,21	5.5	70	86.08	15.24
	Modena	268802	686,95	7.0	49	123.97	14.97
	Parma	344748	1525,42	4.4	85	170.81	9.94
	Piacenza	258586	848,37	6.2	51	111.66	13.07
	Ravenna	185944	213,32	13.0	19	80.55	12.39
	Reggio Emilia	229126	635,18	7.3	58	126.75	11.26
Liguria	La Spezia	88135	542,29	7.6	46	144.55	16.34

References

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., Houghton, R.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Chang. 2, 182.

Barbati, A., Marchetti, M., Chirici, G., Corona, P., 2014. European Forest Types and Forest Europe SFM indicators: tools for monitoring progress on forest biodiversity conservation. For. Ecol. Manage. 321, 145–157. https://doi.org/10.1016/j.foreco.2013.07.004.

Barrett, F., McRoberts, R.E., Tomppo, E., Cienciala, E., Waser, L.T., 2016. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. Remote Sens. Environ. 174, 279–289. https://doi.org/10.1016/j.rse.2015.08.029.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote. Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Borghetti, M., Chirici, G., 2016. Raw data from the Italian National Forest Inventory are on-line and publicly available. For. - Riv. Di Selvic. Ed Ecol. For. 13, 33–34. https://doi.org/10.3832/efor0083-013.

Bottalico, F., Chirici, G., Giannini, R., Mele, S., Mura, M., Puxeddu, M., McRoberts, R.E., Valbuena, R., Travaglini, D., 2017. Modeling Mediterranean forest structure using airborne laser scanning data. Int. J. Appl. Earth Obs. Geoinf. 57, 145–153. https://doi.org/10.1016/j.jag.2016.12.013.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sens. Environ. 112, 3876–3890. https://doi.org/10.1016/j.rse.2008.06.003.

Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. 131, 1255–1267. https://doi.org/10.1007/s10342-012-0596-7.

Breidt, F.J., Opsomer, J.D., 2009. . Chapter 27 - nonparametric and semiparametric estimation in complex surveys. In: Rao, C.R. (Ed.), Handbook of Statistics, Handbook of Statistics. Elsevier, pp. 103–119. https://doi.org/10.1016/S0169-7161(09)00227-2.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1023/A:1018054314350.

Brosofske, K.D., Froese, R.E., Falkowski, M.J., Banskota, A., 2014. A review of methods for mapping and prediction of inventory attributes for operational forest management. For. Sci. 60, 733–756. https://doi.org/10.5849/forsci.12-134.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28, 281–298.

Chirici, G., 2019. A preliminary comparison between IMAGE2006 and landsat 5 TM imagery for the wall-to-wall spatial estimation of growing stock volume in Italy (in press). Proceedings of the 11TH Session of the Winter Forest School at the Forest Research Institute „Application of Geoinformatics in Forestry" 12-14 March.

Chirici, G., McRoberts, R.E., Winter, S., Bertini, R., Bröändli, U.-B., Asensio, I.A., Bastrup-Birk, A., Rondeux, J., Barsoum, N., Marchetti, M., 2012. National forest inventory contributions to forest biodiversity monitoring. For. Sci. 58, 257–268. https://doi.org/10.5849/forsci.12-003.

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., Travaglini, D., McRoberts, R.E., 2016. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. Remote Sens. Environ. 176, 282–294. https://doi.org/10.1016/j.rse.2016.02.001.

Cleveland, W.S., Devlin, S.J., 1998. Locally weighted regression: an approach to regression analysis by local fitting. J. Comput. Graph. Stat. 83, 596–610.

Condés, S., McRoberts, R.E., 2017. Updating national forest inventory estimates of growing stock volume using hybrid inference. For. Ecol. Manage. 400, 48–57.

Corona, P., 2010. Integration of forest mapping and inventory to support forest management. IForest 3, 59–64. https://doi.org/10.3832/ifor0531-003.

Corona, P., 2018. Communicating facts, findings and thinking to support evidence-based strategies and decisions. Ann. Silvicult. Res. 42, 1–2. https://doi.org/10.12899/asr-1617.

Corona, P., Chirici, G., McRoberts, R.E., Winter, S., Barbati, A., 2011. Contribution of large-scale forest inventories to biodiversity assessment and monitoring. For. Ecol. Manage. 262, 2061–2069. https://doi.org/10.1016/j.foreco.2011.08.044.

Corona, P., Fattorini, L., Franceschi, S., Chirici, G., Maselli, F., Secondi, L., 2014. Mapping by spatial predictors exploiting remotely sensed and ground data: a comparative design-based perspective. Remote Sens. Environ. 152, 29–37. https://doi.org/10.1016/j.rse.2014.05.011.

Evans, J.S., Cushman, S.A., 2009. Gradient modeling of conifer species using random forests. Landsc. Ecol. 24, 673–683. https://doi.org/10.1007/s10980-009-9341-0.

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T., 2009. Characterizing forest succession with lidar data: an evaluation for the Inland Northwest, USA. Remote Sens. Environ. 113, 946–956. https://doi.org/10.1016/j.rse.2009.01.003.

FAO, 2013. Strategic Framework on Mediterranean Forests. High Level Segment of the Third Mediterranean Forest Week, Tlemcen March 21, 2013.

FAO, 2010. Global Forest Resources Assessment 2010. FAO Forestry Paper 163, 350 pp. doi:ISBN 978-92-5-106654-6.

FAO, F. and A.O., UNCCD, G.M. of the, 2015. Sustainable Financing for Forest and Landscape Restoration: the Role of Public Policy Makers. pp. 12.

Fattorini, L., Marcheselli, M., Pisani, C., 2006. A three-phase sampling strategy for large-scale multiresource forest inventories. J. Agric. Biol. Environ. Stat. 11, 296–316. https://doi.org/10.1198/108571106X130548.

Fernández-Landa, A., Fernández-Moya, J., Tomé, J.L., Algeet-Abarquero, N., Guillén-Climent, M.L., Vallejo, R., Sandoval, V., Marchamalo, M., 2018. High resolution forest inventory of pure and mixed stands at regional level combining National Forest Inventory field plots, Landsat, and low density lidar. Int. J. Remote Sens. 0, 1–15. https://doi.org/10.1080/01431161.2018.1430406.

Fibbi, L., Chiesi, M., Moriondo, M., Bindi, M., Chirici, G., Papale, D., Maselli, M., 2016. Correction of a 1 km daily rainfall dataset for modelling forest ecosystem processes in Italy. Meteorol. Appl. https://doi.org/10.1002/met.1554.

Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive Relations of Tropical Forest Biomass from Landsat TM Data and Their Transferability between Regions 85. pp. 463–474. https://doi.org/10.1016/S0034-4257(03)00039-7.

FOREST EUROPE, 2015. FOREST EUROPE Liaison Unit MadridState of Europe'S Forests 2015., Ministerial Conference on the Protection of Forests in Europe2015. State of Europe'S Forests 2015., Ministerial Conference on the Protection of Forests in Europe.

Fornaciai, A., Favalli, M., Karátson, D., Tarquini, S., Boschi, E., 2012. Morphometry of scoria cones, and their relation to geodynamic setting: A DEM-based analysis. J. Volcanol. Geotherm. Res. 217–218, 56–72. https://doi.org/10.1016/j.jvolgeores.2011.12.012.

Galluzzi, M., Giannetti, F., Puletti, N., Canullo, R., Rocchini, D., Bastrup, A., Gherardo, B., 2019. A plot level exploratory analysis of European forest based on the results from the BioSoil Forest Biodiversity project. Eur. J. For. Res. https://doi.org/10.1007/s10342-019-01205-2.

L'inventario nazionale delle foreste e dei serbatoi forestali di carbonio INFC 2005. In: Gasparini, P., Tabacchi, G. (Eds.), Secondo Inventario Forestale Nazionale Italiano. Metodi E Risultati. Ministero Delle Politiche Agricole, Alimentari E Forestali; Corpo Forestale Dello Stato. Consiglio Per La Ricerca E La Sperimentazione in Agricoltura, Unità Di Ricerca Per Il Monitoraggio E La Pianificazione Forestale. Edagricole-Il Sole 24 Ore, Bologna, 653 pp.

Giannetti, F., Barbati, A., Mancini, L.D., Travaglini, D., Bastrup-Birk, A., Canullo, R., Nocentini, S., Chirici, G., 2018a. European Forest Types: toward an automated classification. Ann. For. Sci. https://doi.org/10.1007/s13595-017-0674-6.

Giannetti, F., Chirici, G., Gobakken, T., Næsset, E., Travaglini, D., Puliti, S., 2018b. A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. Remote Sens. Environ. 213. https://doi.org/10.1016/j.rse.2018.05.016.

Ginzler, C., Hobi, M., 2015. Countrywide stereo-image matching for updating digital surface models in the framework of the swiss national forest inventory. Remote Sens. 7, 4343–4370. https://doi.org/10.3390/rs70404343.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res. Atmos. 113, D20119.

Hobi, M.L., Ginzler, C., 2012. Accuracy assessment of digital surface models based on WorldView-2 and ADS80 stereo remote sensing data. Sensors Basel (Basel) 12, 6347–6368. https://doi.org/10.3390/s120506347.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B., Maier, B., 2009. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. Int. J. Remote Sens. 30, 5159–5175. https://doi.org/10.1080/01431160903022894.

Houghton, R.A., 2007. Balancing the global carbon budget. Annu. Rev. Earth Planet. Sci. 35, 313–347. https://doi.org/10.1146/annurev.earth.35.031306.140057.

Immitzer, M., Stepper, C., Böck, S., Straub, C., Atzberger, C., 2016. Forest Ecology and Management Use of WorldView-2 stereo imagery and National Forest Inventory data for wall-to-wall mapping of growing stock. For. Ecol. Manage. 359, 232–246. https://doi.org/10.1016/j.foreco.2015.10.018.

INFC, 2008. Le stime di superficie – risultati per macroaree e Province. In: Gasparini, A.P., Di Cosmo, L., Gagliano, C., Tabacchi, G.M.G. (Eds.), Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAAF – Ispettorato Generale Corpo Forestale dello Stato, CRA-MPF, Trento.

INFC, 2007. Le stime di superficie 2005 - seconda parte. In: Tabacchi, A.G., De Natale, F., Di Cosmo, L., Floris, A., Gagliano, C., Gasparini, P., Salvadori, I., Scrinzi, G., Tosi, V. (Eds.), Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAF – Corpo Forestale dello Stato - Ispettorato Generale, CRA - ISAFA, Trento. Available from: http://www.infc.it [on line].

INFC, 2004. Il disegno di campionamento. Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAF - Direzione Generale per le Risorse Forestali Montane e Idriche, Corpo Forestale dello Stato, ISAFA, Trento 36 p. http://www.isafa.it/scientifica/.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., Olsson, H., 2018. Remote sensing and forest inventories in Nordic countries – roadmap for the future. Scand. J. For. Res. 7581, 1–16. https://doi.org/10.1080/02827581.2017.1416666.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., Acker, S.A., 2005. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. Remote Sens. Environ. 95, 532–548. https://doi.org/10.1016/j.rse.2005.01.010.

Li, wang, Z., XIN, ping, X., TANG, H., YANG, F., CHEN, B., ZHANG, B., 2017. Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. J. Integr. Agric. 16, 286–297. https://doi.org/10.1016/S2095-3119(15)61303-X.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. Nucleic Acids Res. 5, 983–999. https://doi.org/10.1023/A:1010933404324.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. Int. J. Digit. Earth 0, 1–43. https://doi.org/10.1080/17538947.2014.990526.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Laurin, G.V., Saah, D., 2012. Aboveground Forest Biomass Estimation With Landsat and LiDAR Data and

Uncertainty Analysis of the Estimates. https://doi.org/10.1155/2012/436537.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27 (7), 1297–1328.
Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press ISBN 0-12-471252-5.
Maselli, F., 2002. Improved estimation of environmental parameters through locally calibrated multivariate regression analyses. Photogramm. Eng. Remote Sensing 68, 1163–1171.
Maselli, F., Pasqui, M., Chirici, G., Chiesi, M., Fibbi, L., Salvati, R., Corona, P., 2012. Modeling primary production using a 1 km daily meteorological data set. Clim. Res. 54, 271–285.
Maselli, F., Chiesi, M., Mura, M., Marchetti, M., Corona, P., Chirici, G., 2014a. Combination of optical and LiDAR satellite imagery with forest inventory data to improve wall-to-wall assessment of growing stock in Italy. Int. J. Appl. Earth Observ. Geoinform. 26, 377–386. https://doi.org/10.1016/j.jag.2013.09.001.
Maselli, F., Chiesi, M., 2006. Evaluation of statistical methods to estimate forest volume in a Mediterranean region. IEEE Trans. Geosci. Remote. Sens. 44 (8), 2239–2250.
Maselli, F., Chiesi, M., Corona, P., 2014b. Use of geographically weighted regression to enhance the spatial features of forest attribute maps. J. Appl. Remote Sens. 8 (1), 083533. https://doi.org/10.1117/1.JRS.8.083533.
Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S.J., 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sens. Environ. 209, 90–106. https://doi.org/10.1016/j.rse.2017.12.020.
McRoberts, R., Tomppo, E., Schadauer, K., Vidal, C., Stahl, G., Chirici, G., Lanz, A., Cienciala, E., Winter, S., Smith, B., 2009. Harmonizing national forest inventories. J. For. 179–187.
McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V., Tomppo, E.O., 2010b. Using remotely sensed data to construct and assess forest attribute maps and related spatial products. Scand. J. For. Res. 25, 340–367. https://doi.org/10.1080/02827581.2010.497496.
McRoberts, R.E., Domke, G.M., Chen, Q., Næsset, E., Gobakken, T., 2016a. Using genetic algorithms to optimize k-Nearest Neighbors configurations for use with airborne laser scanning data. Remote Sens. Environ. 184, 387–395. https://doi.org/10.1016/j.rse.2016.07.007.
McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. 110, 412–419. https://doi.org/10.1016/j.rse.2006.09.034.
McRoberts, R.E., Tomppo, E.O., Næsset, E., 2010a. Advances and emerging issues in national forest inventories. Scand. J. For. Res. 25, 368–381.
McRoberts, R.E., Næsset, E., Gobakken, T., 2015. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. Remote Sens. Environ. 163, 13–22.
McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., Westfall, J.A., 2016b. Hybrid estimators for mean aboveground carbon per unit area. For. Ecol. Manage. 378, 44–56.
McRoberts, R.E., Chen, Q., Walters, B.F., Kaisershot, D.J., 2018. The effects of global positioning system receiver accuracy on airborne laser scanning-assisted estimates of aboveground biomass. Remote Sens. Environ. 207, 42–49.
Montaghi, A., Corona, P., Dalponte, M., Gianelle, D., Chirici, G., Olsson, H., 2013. Airborne Laser Scanning of Forest Resources: an overview of research in Italy as a commentary case study. Int. J. Appl. Earth Obs. Geoinf. 23, 288–300. https://doi.org/10.1016/j.jag.2012.10.002.
Moser, P., Vibrans, A.C., McRoberts, R.E., Næsset, E., Gobakken, T., Chirici, G., Mura, M., Marchetti, M., 2017. Methods for variable selection in LiDAR-assisted forest inventories. Forestry 90, 112–124. https://doi.org/10.1093/forestry/cpw041.
Mura, M., Bottalico, F., Giannetti, F., Bertani, R., Giannini, R., Mancini, M., Orlandini, S., Travaglini, D., Chirici, G., 2018. Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. Int. J. Appl. Earth Obs. Geoinf. 66, 126–134. https://doi.org/10.1016/j.jag.2017.11.013.
Mura, M., McRoberts, R.E., Chirici, G., Marchetti, M., 2015. Estimating and mapping forest structural diversity using airborne laser scanning data. Remote Sens. Environ. 170, 133–142. https://doi.org/10.1016/j.rse.2015.09.016.
Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. Scand. J. For. Res. 22 433–422.
Næsset, E., Bollandsås, O.M., Gobakken, T., Gregoire, T.G., Ståhl, G., 2013. Model-assisted estimation of change in forest biomass over an 11year period in a sample survey supported by airborne LiDAR: a case study with post-stratification to provide "activity data.". Remote Sens. Environ. 128, 299–314. https://doi.org/10.1016/j.rse.2012.10.008.
Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., Söderman, U., 2004. Laser scanning of forest resources: the nordic experience. Scand. J. For. Res. 19, 482–499. https://doi.org/10.1080/02827580410019553.
Nichol, J.E., Sarker, L.R., 2011. Improved biomass estimation using the texture parameters of two high-resolution. Optical Sensors 49, 930–948.
Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., Olsson, H., 2017. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. Remote Sens. Environ. 194, 447–454. https://doi.org/10.1016/j.rse.2016.10.022.
Nord-Larsen, T., Schumacher, J., 2012. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. Remote Sens. Environ. 119, 148–157. https://doi.org/10.1016/j.rse.2011.12.022.
Panagos, P., 2006. The European soil database. GEO: connexion 5, 32–33.
Pecchi, M., Marchi, M., Giannetti, F., Bernetti, I., Binidi, M., Moriondo, M., Maelli, F., Fibbi, L., Corona, P., Travaglini, D., Chirici, G., 2019. Reviewing climatic traits for the main forest tree species in Italy Matteo. iForest Biogeosci. For. 12, 173–180. https://doi.org/10.3832/ifor2835-012.
Persson, H.J., Olsson, H., Soja, M.J., Ulander, L.M.H., Fransson, J.E.S., 2017. Experiences from large-scale forest mapping of Sweden using TanDEM-X data. Remote Sens. 9. https://doi.org/10.3390/rs9121253.
Puletti, N., Floris, A., Scrinzi, G., Chianucci, F., Colle, G., Michelini, T., Pedot, N., Penasa, A., Scalercio, S., Corona, P., 2017. CFOR: a spatial decision support system dedicated to forest management in Calabria. For. - Riv. Di Selvic. Ed Ecol. For. 14, 135–140. https://doi.org/10.3832/efor2363-014.
Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R., 2014. Comparison of four types of 3D data for timber volume estimation. Remote Sens. Environ. 155, 325–333. https://doi.org/10.1016/j.rse.2014.08.036.
Reese, H., Nilsson, M., Sandström, P., Olsson, H., 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. Comput. Electron. Agric. 37, 37–55. https://doi.org/10.1016/S0168-1699(02)00118-7.
Särndal, C.-E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer. ed., Berlin.
Scarascia-Mugnozza, G., Oswald, H., Piussi, P., Radoglou, K., 2000. Forests of the Mediterranean region: gaps in knowledge and research needs. For. Ecol. Manage. 132 (1), 97–109. https://doi.org/10.1016/S0378-1127(00)00381-9.
Smith, P., Bustamente, H., Ahammad, H., Clark, H., Dong, E.A., Elsiddig, H., Haberl, R., Harper, J., House, M., Jafari, O., Masera, C., Mbow, N.H., Racindranath, C.W., Vermeulen, S.J., Campbell, B.M., Ingram, J.S.I., Aide, T.M., Clark, M.L., Grau, H.R., López-Carr, D., Levy, M.A., Redo, D., Bonilla-Moheno, M., Riner, G., Andrade-Núñez, M.J., Muñiz, M., Yale Center for Environmental Law and Policy, FAO, Spracklen, B.D., Kalamandeen, M., Galbraith, D., Gloor, E., Spracklen, D.V., 2016. STATE OF THE WORLD'S FORESTS - agriculture, forestry and other Land use challanges and opportunities, climate change 2014: mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. https://doi.org/10.1146/annurev-environ-020411-130608.
Stinson, G., White, J., 2018. What's the Difference Between EFI and NFI? Demystifying current acronyms in forest inventory in Canada.
Stumpf, A., Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. Remote Sens. Environ. 115, 2564–2577. https://doi.org/10.1016/j.rse.2011.05.013.
Särndal, C.-E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York 694p.
Tabacchi, G., Di Cosmo, L., Gasparini, P., Morelli, S., 2011. Stima Del Volume E Della Fitomassa Delle Principali Specie Forestali Italiene, Equazioni Di Previsione, Tavole Del Volume E Tavole Della Fitomassa Arborea Epigea.
Tarquini, S., Isola, I., Favalli, M., Mazzarini, F., Bisson, M., Pareschi, M.T., Boschi, E., 2007. TINITALY/01: a new triangular irregular network of Italy. Ann. Geophys. 50, 407–425.
Tarquini, S., Nannipieri, L., 2017. The 10 m-resolution TINITALY DEM as a trans-disciplinary basis for the analysis of the Italian territory: current trends and new perspectives. Geomorphology. https://doi.org/10.1016/j.geomorph.2016.12.022.
Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. 112, 1982–1999. https://doi.org/10.1016/j.rse.2007.03.032.
Waser, L.T., Fischer, C., Wang, Z., Ginzler, C., 2015. Wall-to-wall forest mapping based on digital surface models from image-based point clouds and a NFI forest definition. Forests 6, 4510–4528. https://doi.org/10.3390/f6124386.
Waser, L.T., Ginzler, C., Rehush, N., 2017. Wall-to-Wall tree type mapping from countrywide airborne remote sensing surveys. Remote Sens. 9. https://doi.org/10.3390/rs9080766.
Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. Isprs J. Photogramm. Remote. Sens. 66, 28–37. https://doi.org/10.1016/j.isprsjprs.2010.08.003.
Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., 2016. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. Remote Sens. https://doi.org/10.3390/rs8060469.

remote sensing

MDPI

Article

The Effect of Forest Mask Quality in the Wall-to-Wall Estimation of Growing Stock Volume

Elia Vangi [1,2], Giovanni D'Amico [1], Saverio Francini [1,2,3], Francesca Giannetti [1,4,*], Bruno Lasserre [2], Marco Marchetti [2], Ronald E. McRoberts [5] and Gherardo Chirici [1,4,6]

1 Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies, University of Florence, 50145 Firenze, Italy; elia.vangi@unifi.it (E.V.); giovanni.damico@unifi.it (G.D.); saverio.francini@unifi.it (S.F.); gherardo.chirici@unifi.it (G.C.)
2 Department of Biosciences and Territory, University of Molise, 86100 Campobasso, Italy; lasserre@unimol.it (B.L.); marchettimarco@unimol.it (M.M.)
3 Department for Innovation in Biological, Agri-Food and Forestry Systems, University of Tuscia, 01100 Viterbo, Italy
4 ForTech Joint Laboratory, University of Florence, 50145 Firenze, Italy
5 Department of Forest Resources, University of Minnesota, Saint Paul, MN 55108, USA; mcrob001@umn.edu
6 Research Unit COPERNICUS, University of Florence, 50145 Firenze, Italy
* Correspondence: francesca.giannetti@unifi.it

Abstract: Information about forest cover and its characteristics are essential in national and international forest inventories, monitoring programs, and reporting activities. Two of the most common forest variables needed to support sustainable forest management practices are forest cover area and growing stock volume (GSV m^3 ha^{-1}). Nowadays, national forest inventories (NFI) are complemented by wall-to-wall maps of forest variables which rely on models and auxiliary data. The spatially explicit prediction of GSV is useful for small-scale estimation by aggregating individual pixel predictions in a model-assisted framework. Spatial knowledge of the area of forest land is an essential prerequisite. This information is contained in a forest mask (FM). The number of FMs is increasing exponentially thanks to the wide availability of free auxiliary data, creating doubts about which is best-suited for specific purposes such as forest area and GSV estimation. We compared five FMs available for the entire area of Italy to examine their effects on the estimation of GSV and to clarify which product is best-suited for this purpose. The FMs considered were a mosaic of local forest maps produced by the Italian regional forest authorities; the FM produced from the Copernicus Land Monitoring System; the JAXA global FM; the hybrid global FM produced by Schepaschencko et al., and the FM estimated from the Corine Land Cover 2006. We used the five FMs to mask out non-forest pixels from a national wall-to-wall GSV map constructed using inventory and remotely sensed data. The accuracies of the FMs were first evaluated against an independent dataset of 1,202,818 NFI plots using four accuracy metrics. For each of the five masked GSV maps, the pixel-level predictions for the masked GSV map were used to calculate national and regional-level model-assisted estimates. The masked GSV maps were compared with respect to the coefficient of correlation (ρ) between the estimates of GSV they produced (both in terms of mean and total of GSV predictions within the national and regional boundaries) and the official NFI estimates. At the national and regional levels, the model-assisted GSV estimates based on the GSV map masked by the FM constructed as a mosaic of local forest maps were closest to the official NFI estimates with ρ = 0.986 and ρ = 0.972, for total and mean GSV, respectively. We found a negative correlation between the accuracies of the FMs and the differences between the model-assisted GSV estimates and the NFI estimate, demonstrating that the choice of the FM plays an important role in GSV estimation when using the model-assisted estimator.

Keywords: forest mask; spatial estimation; growing stock volume; Italy

Citation: Vangi, E.; D'Amico, G.; Francini, S.; Giannetti, F.; Lasserre, B.; Marchetti, M.; McRoberts, R.E.; Chirici, G. The Effect of Forest Mask Quality in the Wall-to-Wall Estimation of Growing Stock Volume. *Remote Sens.* **2021**, *13*, 1038. https://doi.org/10.3390/rs13051038

Academic Editor: Inge Jonckheere

Received: 9 February 2021
Accepted: 4 March 2021
Published: 9 March 2021

1. Introduction

Information about forest cover and its characteristics are essential in national and international forest inventories, monitoring programs, and reporting activities [1,2] such as in the context of international agreements (e.g., Kyoto protocol), and restoration programs (e.g., Reducing emissions from deforestation and forest degradation projects, REDD+) [3]. Two of the most common forest variables needed to estimate sustainable forest management indicators as required by the national and international framework and agreements relate to forest cover area (generally according to the international definition adopted by the Food and Agriculture Organization (FAO) and the total growing stock volume (GSV, m^3) [4,5]. These data are usually provided by national forest inventory (NFI) programs which use probability-based approaches to infer the estimates for large areas such as countries and regions within countries. [4,6,7]. In several countries with long NFI histories such as Norway [8], Finland [9], Austria [10], and Switzerland [11,12], the typical NFI ground survey is nowadays complemented by continuous spatial predictions, characterized as wall-to-wall maps of forest variables which rely on models and wall-to-wall auxiliary data such as remotely sensed data [13–15].

Wall-to-wall GSV data are useful because they can be integrated into decision support systems to assess wood production and harvesting activities at small scales (i.e., in forest properties) [16–19] and to produce small-scale estimates by aggregating individual pixel predictions [20–23]. In the probability-based framework, multiple estimators including the stratified, post-stratified, and model-assisted estimators can be used. The latter is considered asymptotically unbiased in the sense that the mean of estimates obtained using the estimator for all possible samples approaches the true value as the sample size increases [23].

GSV and above-ground biomass are known to be strongly correlated with three-dimensional (3D) data such as those acquired through airborne laser scanning (ALS) or photogrammetric techniques [5,14,24–27]. However, acquiring these data is still expensive, and some countries such as Italy still do not have wall-to-wall ALS coverage [19]. Multi-spectral satellite data are often used instead of or with 3D data to predict GSV, thanks to their free availability over large areas [28–31].

Several types of models can be used to produce wall-to-wall predictions of forest attributes in a model-assisted approach. These models include both parametric and non-parametric techniques [14,17,27,28,32] with the recent prevalence of multiple linear regression and random forests [14,33,34]. Regardless of the estimation approach, spatial knowledge of the area covered by forest land is an essential prerequisite, both to restrict the establishment of field plots and to restrict the application of the models. A forest mask (FM) indicates the location of forest land and is often in a raster or a spatial polygon database format. FMs are conventionally obtained by manual delineation of aerial images, or by supervised or unsupervised classification of satellite imagery, from both optical or radar imagery [35–37], and more recently ALS data [38–41]. Remotely sensed data suitable for forest mapping are nowadays frequently and freely available [42–44]. For this reason, the number of FMs has increased exponentially, creating doubts about which is best-suited for specific purposes such as forest area and GSV estimation. National information about forest extent can be estimated from any of several FMs produced independently by different research agencies globally or for large areas, including the European Environmental Agency (EEA) [45], the European Space Agency (ESA) [46], the International Institute for Applied Systems Analysis (IIASA) [1], and the Japanese Aerospace Exploration Agency (JAXA) [47]. Despite individual weaknesses and strengths, spatial differences among these products are evident and can lead to substantial variation in their accuracies [1,48]. Furthermore, these FMs were developed for different aims and thus have different characteristics in terms of minimum mapping unit (MMU) and minimum mapping width (MMW), reference forest definition, and year of production.

Multiple studies have compared land cover maps at global and local levels. Fritz and See [49] and Giri et al. [50] compared the Global Land Cover 2000 data set and

the MODIS global land cover product and highlighted areas with strong disagreements. Hoyos et al. [51] compared four global satellite-based land cover maps and showed a worsening of area agreements as the spatial scale increases. Neumann et al. [52] provided an assessment of compatibilities and differences between the CORINE2000 and GLC2000 datasets and reported general disagreement due to the combination of thematic similarities, spatial heterogeneity, and classification accuracy. Seebach et al. [53] compared the advantages and limitations of four pan-European forest cover maps for the reference year 2000, demonstrating that the spatial agreement between the maps ranged between 50% to 70% within a large study area in Europe. The authors found the greatest spatial differences among all maps in the Alpine and Mediterranean regions. Here, the vulnerability to climate change and anthropogenic disturbance is extremely large and will cause an increased demand for accurate wall-to-wall maps [17]. Only a few studies have analyzed the effects of using different FMs on the uncertainty of forest parameter estimates. Rodríguez-Veiga et al. [54] reported a large impact on estimates of national carbon stocks in Mexico caused by discrepancies in forest extent estimated from different FMs. In their study, Li et al. [55] considered the uncertainty of the MODIS land cover products, finding substantial differences in the regional climate modeling outputs when the uncertainty was not considered. Esteban et al. [56] estimated the effects of the uncertainty of forest species maps used in the sampling and forest parameter estimation processes in a Spanish study area. Their study revealed that the effects of map uncertainty are not negligible, especially for less common Mediterranean forest species.

The choice of FM can heavily impact the estimation of forest parameters in two different manners: (i) it affects the number and locations of plots selected for the construction of the predictive model and (ii) it affects the total area to which the model is applied [56].

The aim of this paper is to evaluate the impacts of the accuracies of different FMs on the estimation of GSV based on the integration of field information and remotely sensed data. We constructed a national wall-to-wall GSV map with an optimized procedure based on a random forests model with remotely sensed imagery and other auxiliary data as predictors [17]. We used five different FMs to mask out non-forest areas from the GSV map and then used the model-assisted regression estimator to estimate total and mean GSV (m^3 ha^{-1}) for the forest portion of the GSV map. We then investigated the relationship between mask accuracies and agreement between the model-assisted total GSV estimates and the official NFI estimates. The test was carried out for the entire area of Italy. Finally, we clarified which product was best-suited for total and mean GSV estimation, both at national and regional levels.

2. Materials and Methods

2.1. Study Areas

The study was carried out in Italy which covers 301,408 km^2 (Figure 1). Italy has extreme variations in climatic conditions due to proximity to the sea and elevation ranges between coastal areas and the Alpine region with elevations as great as 4000 m asl.

The territory falls within the temperate zone of a Mediterranean climatic region (Pinna, 1970). On the coasts of the main islands, the average annual rainfall is 250 mm but reaches more than 3000 mm in the Alpine and pre-Alpine belts. Average yearly temperatures vary between 16 °C in the southern coastal areas to 10 °C in the inner central regions and the pre-Alps, with temperatures less than 5 °C in the mountain ranges and on the highest peaks.

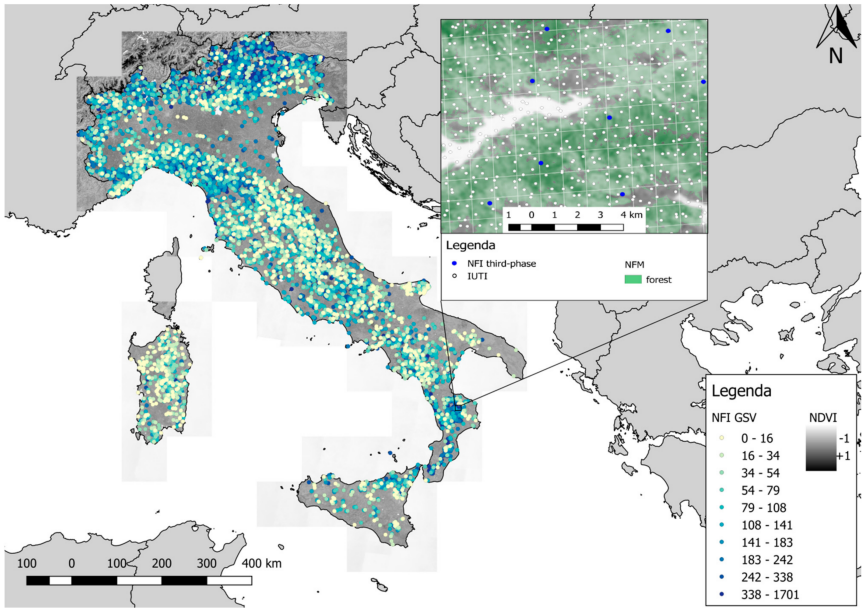


Figure 1. The study area with the distribution of the national forest inventory (NFI) plots colored by growing stock volume (GSV) expressed in m^3 ha^{-1}. On the right, a detail of the distribution of sample points used in the study within the NFI 1 x 1 km grid where the third-phase NFI plots (Section 2.2.1) are depicted in blue and the Inventario dell'Uso delle Terre in Italia (IUTI) points (Section 2.2.2) in white.

According to the last Italian NFI (INFC, 2007), forest vegetation and other wooded lands occupy 10,467,533 ha, about 34% of the national territory. Forests are dominated by deciduous trees (68%), mainly Quercus oak (*Q. petrea* (M.) L., *Q. pubescens* W., *Q. robur* L., *Q. cerris* L.), and European beech (*Fagus sylvatica* L.). The dominant conifers are Norway spruce (*Picea abies* K.) and pines (*Pinus sylvestris* L., *P. nigra* A., *P. pinae* L., *P. pinaster* A.), which are mainly artificial plantations located in mountain areas or near the coast (Figure 1). Seven of the 14 European forest types occur in Italy, of which the most common is the thermophilous deciduous forest [14,57].

Italy is divided into 20 administrative regions (NUTS2) for each of which the NFI produces estimates of forest area, total and mean GSV, and their standard errors (SEs). The average GSV is 144 m^3 ha^{-1} [58].

2.2. Field Data

2.2.1. Second Italian National Forest Inventory

The field reference data for the wall-to-wall spatial prediction of GSV were acquired in the framework of the second Italian NFI [59] based on a three-phase, systematic, unaligned sampling design with 1 × 1 km grid cells [60]. In the first phase, N = 301,300 points were selected and classified with respect to 10 coarse land-use strata using aerial orthophotos. In the second phase, for an $n < N$ sub-sample of the points in the "forest" stratum of the first-phase points, qualitative information such as forest type, management, and property were collected during a field survey. In the third phase, for a sub-sample of 6782 points extracted from the second-phase points, a quantitative survey was carried out for circular plots of 13 m radius (530 m^2). All tree stems with DBH of at least 2.5 cm were callipered, and for a subsample, height was measured. For all 6782 third-phase plots, allometric models [61] were used to predict GSV (m^3) which was then aggregated at plot-level and scaled to a per unit area basis. For this study, allometric model prediction uncertainty and uncertainty due to Global Navigation Satellite System (GNSS) position error were expected to be negligible for the spatial resolution adopted [4,17,23,62]. The third-phase plots have mean GSV of 145.75 m^3 ha^{-1}, with median value of 102.82 m^3 ha^{-1}.

Official, design-based NFI estimates of total forest area and mean and total GSV at national and regional NUTS2 levels were acquired online at https://www.sian.it/inventarioforestale/ (accessed on: 2 October 2020) [62], for the reference year 2005.

The study area was tessellated into a 23 × 23 m national grid whose pixel area matched the area of the NFI ground plots, for a total of 569,769,690 pixels [19]. The national grid was used as a spatial reference grid for resampling the predictor variables and the FM to 23 × 23 m resolution.

2.2.2. Inventory of Land Use in Italy

To evaluate the accuracy of the FMs, we used the sample points from the Italian land use inventory (Inventario dell'Uso delle Terre in Italia, IUTI). The IUTI has adopted the methodology of approach number three of the Good Practices Guidance for Land Use, Land Use Change, and Forestry (GPG-LULUCF) of the Intergovernmental Panel on climate change [63–65]. IUTI is a permanent monitoring system that estimates the extent of six land use categories identified in the GPG-LULUCF. The IUTI is based on a systematic unaligned sampling design with 0.5 × 0.5 km grid cells which is an intensification of the NFI sample grid, for a total of 1,202,828 points of which 301,300 are the first-phase points of the NFI. The six categories reported by IUTI are urban, agriculture, forest land, grassland, wetland, other [65]. Each point is photo-interpreted in three time periods (1990, 2008, 2012) for estimating land-use change using aerial orthophotos with spatial resolution ranging between 1 × 1 m for 1990 and 0.5 × 0.5 m for 2008. We combined the six land use categories into forest and non-forest and assigned the value 1 to all the points classified as forest (class 1.1, 1.2) and 0 to all other categories. Subsequently, the forest class included 32% of the total observations with 387,085 of 1,202,818 points.

For this study, we used the IUTI points as an independent dataset to evaluate the accuracies of the FMs. We used the 2008 photointerpretation to be as consistent as possible with the 2005 NFI ground surveys.

2.2.3. Predictor Variables

To predict GSV as described in Section 3.1, we used predictors obtained from multiple sources including remotely sensed variables from multiple sensors, climate, and soil characteristics (Table 1). The variables were selected based on their availability throughout the national territory as reported by [17]. All variables were resampled from the original resolution to the 23 × 23 m pixel size of the national grid. A more detailed description of the database is provided by [17].

Table 1. Predictor variables based on remotely sensed and auxiliary data.

Database	Band/Information	Predictor Variables	Original Pixel Width
Landsat 7 ETM+	Three years median of Band 1	Landsat_B1	30 m
Landsat 7 ETM+	Three years median of Band 2	Landsat_B2	30 m
Landsat 7 ETM+	Three years median of Band 3	Landsat_B3	30 m
Landsat 7 ETM+	Three years median of Band 4	Landsat_B4	30 m
Landsat 7 ETM+	Three years median of Band 5	Landsat_B5	30 m
Landsat 7 ETM+	Three years median of Band 6	Landsat_B6	30 m
Landsat 7 ETM+	Three years median of Band 7	Landsat_B7	30 m
Global PALSAR/PALSAR-2	HH polarization	SAR_HH	25 m
Global PALSAR/PALSAR-3	HV polarization	SAR_HV	25 m
Climate data	Total annual precipitation	Prec	1 km
Climate data	Mean annual temperature	temp_mean	1 km
Climate data	Maximum annual temperature	temp_max	1 km
Climate data	Minimum annual temperature	temp_min	1 km
European Soil Database v2.0	Subsoil available water capacity	AWC_SUB	1 km
European Soil Database v2.1	Topsoil available water capacity	AWC_TOP	1 km

2.2.4. Landsat Composite Image

We constructed a cloud-free composite image across Italy based on 848 Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images acquired in the same year as the field survey (2005) +/− 1 year (Figure 2).

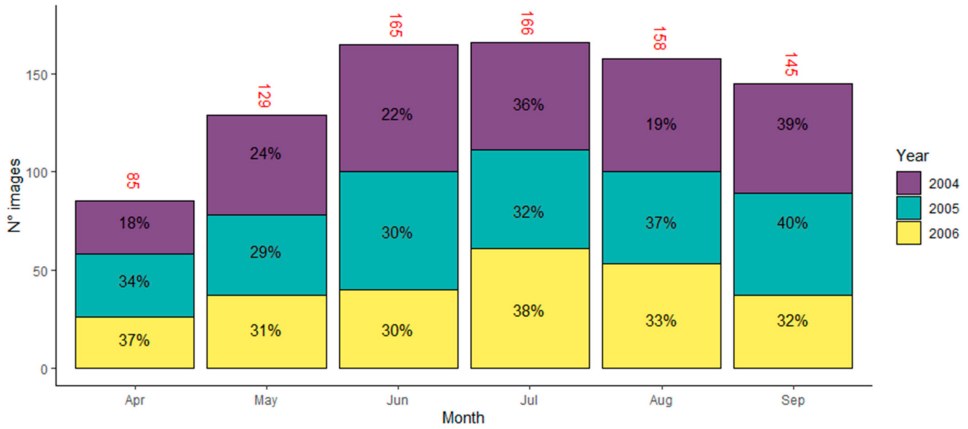


Figure 2. Distribution of Landsat 7 ETM+ images per month, divided by acquisition years.

We used Landsat 7 Surface Reflectance Tier 1 imagery from the Earth Engine Data Catalog, acquired in the vegetation period (1 April–30 September), atmospherically corrected using Landsat Ecosystem Disturbance Adaptive Processing System LEDAPS [66]. We masked out cloud pixels based on the quality assessment (QA) band provided with the Landsat 7 database, using the C function of mask algorithm (CFMask) [67]. Finally, for each 23 × 23 m national grid pixel, we calculated the median values for each Landsat band [68].

2.2.5. SAR Variables

We used SAR data from PALSAR-2/PALSAR from the Advanced Land Observing Satellite (ALOS) and Advanced Land Observing Satellite-2 (ALOS-2) freely available at the global level online from the Japan Aerospace Exploration Agency (JAXA) at 25 × 25 m resolution. We rescaled the raw backscattering coefficients for each polarization HH and HV for the year 2007 to the 23 × 23 m pixel of the national grid. For more information on this data we refer to https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm (accessed on: 5 November 2019)

2.2.6. Climate and Soil Variables

We derived climate data from the 1 x 1 km downscaled climatological maps obtained by Maselli et al. [69] which is representative of the period 1981–2010. The dataset includes the following variables: total annual precipitation, mean annual temperature, maximum annual temperature, minimum annual temperature. For more details on these climate data, we refer to Chirici et al. [17].

Soil variables were from the harmonized soil geodatabase of Europe (European Soil Database v2.0—2004) [70]. The subsoil available water capacity and topsoil available water capacity soil variables used for this study were selected using the optimization phase described in Chirici et al. [17].

2.3. Forest Masks

We obtained five FMs available for the entire Italian territory that potentially reflect the forest FAO Forest Resource Assessment (FRA) definition [2]. These masks can be divided into two main categories: (i) FMs obtained by semi-automated classification of remotely sensed data; (ii) FMs obtained by manual delineation and classification of fine-resolution images. All the FMs were first reprojected in the WGS 84 / UTM zone 32

North (EPSG:32632) reference system to make them comparable and then resampled at the 23 × 23 m resolution of the national grid resulting to produce five comparable FMs.

2.3.1. National Forest Mask (NFM)

We used the national forest mask (NFM) which is based on the mosaic of local forest maps produced by manual photointerpretation by the Italian regional forest authorities [19]. The mosaic was constructed by merging 16 fine resolution forest maps with nominal reference scales varying between 1:5000 and 1:25,000 and five land use maps specifically filtered to produce forest cover maps. All the maps were based on manual photointerpretation of aerial orthophotos. The local forest maps were reclassified into Boolean masks using code 1 for pixels classified as "forest", and code 0 for pixels classified as "non-forest". The NFM is a mosaic of 20 fine-resolution regional forest maps resampled at the 23 × 23 m national grid resolution. The mask is also available on-line at www.forestinfo.it

2.3.2. Copernicus Land Monitoring System (CLMS) Forest Mask

To construct the Copernicus FM, we first used the 2012 Forest Type map (https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/2012?tab=download) (accessed on: 5 November 2020) that uses the Tree Cover Density layer (https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density/status-maps/2012?tab=download) (accessed on: 5 November 2020) to classify all 20 × 20 m pixels of European lands as forest when the tree cover density is at least 10% and when such pixels are aggregated into a continuous patch of at least 0.52 hectares [46]. We excluded pixels in agricultural and urban contexts from the Forest Type map, using the Forest Additional Support Layer also available from Copernicus at https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/2012?tab=download (accessed on: 5 November 2020). The resulting map reflects as closely as possible the international forest definition in a raster layer having 23 × 23 m resolution

2.3.3. JAXA Forest Mask

JAXA constructed an FM for the reference years 2007 ± 1 with a spatial resolution of 25 × 25 m based on the HV-polarization backscatter images acquired by the PALSAR and PALSAR 2 sensors carried by the ALOS and ALOS2 satellites. JAXA adopted the FAO forest definition [47] and is available online at https://developers.google.com/earthengine/datasets/catalog/JAXA_ALOS_PALSAR_YEARLY_SAR (accessed on: 5 Novemeber 2020).

2.3.4. Hybrid Global Forest Mask 2000 (FM00)

Schepaschenko et al. [1] constructed a global FM using a hybrid approach combining multiple local, national, and global datasets into a single product. This map was constructed by converting the global forest probability map into a forest/non-forest map using a threshold calculated for each country. The threshold selected for this study produced area estimates that matched as closely as possible the official FAO forest area statistics. We characterized this map as "FM00". The map has a spatial resolution of 1 × 1 km, was produced for the reference year 2000, and is available online at https://application.geo-wiki.org/branches/biomass/ (accessed on: 5 November 2020).

2.3.5. Corine Land Cover 2006 (CLC06)

The CORINE Land Cover (CLC) project was initiated in 1990 by the European Environmental Agency (EEA) [71] and has been updated in 2000, 2006, 2012, and 2018 to monitor land-use changes in the 39 participating countries [45]. It consists of land cover maps based on a nomenclature system of 44 classes produced by photointerpretation of fine-resolution satellite imagery. CLC uses a MMU of 25 hectares and a MMW of 100 m. For this study, we acquired the CLC map for the reference year 2006 ± 1 (referred to as "CLC06") obtained by photo-interpretation of SPOT-4/5 and IRS P6 LISS III dual data images (EEA, 2007) [45]

and available online in vector format at https://land.copernicus.eu/pan-european/corine-land-cover/clc-2006?tab=download (accessed on: 5 November 2020). To derive the CLC mask, we first rasterized the vector product to the 23 × 23 m spatial resolution of the national grid, and then we assigned the categories 2.4.4, 3.1.1, 3.1.2, 3.1.3, 3.2.3, 3.2.4 to the "forest" class and all the remaining categories to the "non-forest" class.

2.4. Overview of the Method

A concise overview of the methodology followed is presented: (i) a wall-to-wall GSV map was constructed using a random forests model with the NFI plot GSV data and the predictor variables; (ii) the accuracies of the five FMs were assessed; (iii) the wall-to-wall GSV map was masked in turn with each of the five FMs, obtaining five masked GSV maps; (iv) for each masked GSV map we estimated the mean and total GSV with the model-assisted regression estimator, at the national and regional levels; (v) we compared model-assisted estimations for each FM with the official estimate from the Italian NFI, in terms of correlation coefficient; (vi) we assessed the relationship between FM accuracies and GSV estimates in terms of the correlation coefficient.

2.5. Wall-to-Wall National GSV Map

To estimate the effects of FM accuracy on the model-assisted GSV estimates, we constructed a GSV map consisting of GSV predictions for all 23 × 23 m pixels of the national grid (569,769,690 pixels) using the random forests (RF) prediction technique with the NFI plot GSV data and the predictor variables described in Table 1. RF was optimized following Chirici et al. [17] by selecting the combination of predictor variables and parameter values (n_{tree} and m_{try}) that minimized the root mean square error (RMSE) calculated using the leave one out cross-validation (LOOCV) technique [72]. RMSE was calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where n is the number of third-phase NFI plots (i.e., 6782), y_i is the i-th GSV associated with the plots and $\hat{y}_i$ is the i-th GSV predicted by the random forests model. The most accurate combination resulting from LOOCV was used to predict the GSV for all N pixels of the study area to produce a 23 × 23 m resolution GSV map. The model fitting and optimization phase was performed using the *randomForest* package within the statistical software package R 3.6.3 [73] (https://www.r-project.org, accessed on: 5 November 2020). For the 6782 NFI plots, the pixel-level GSV predictions ranged between 0 and 690 m^3 ha^{-1} with a standard deviation of 68.5 m^3 ha^{-1} while the original NFI values ranged between 0.3 and 701 m^3 ha^{-1} with a standard deviation of 147 m^3 ha^{-1}. The map was found to have a mean deviation of −4.3 m^3 ha^{-1}.

2.6. Accuracy Assessment of FMs

We first assessed the five FMs with respect to thematic accuracy using the IUTI dataset as reference data. For each of the 1,202,828 points of the IUTI database, we extracted the forest/non-forest classification from the five FMs and constructed the respective five confusion matrices. For each matrix we calculated four metrics:

$$\text{OverallAccuracy} = \frac{\sum True\ positive + \sum True\ negative}{\sum Total\ population} \quad (2)$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (3)$$

where:

$$p_0 = \text{Overall Accuracy and } p_e = \frac{1}{N^2}\sum_k \sum True\ positive * \sum True\ negative \quad (4)$$

for k categories and N observations.

$$\text{Precision} = \frac{\sum True\ positive}{\sum True\ positive + \sum False\ positive} \quad (5)$$

$$\text{Recall} = \frac{\sum True\ positive}{\sum True\ positive + \sum False\ negative} \quad (6)$$

These metrics need to be used together to correctly describe the quality of classification in the case of unbalanced datasets. This is the case for forest masks when the forest and non-forest classes cover the land area with very different proportions. In such cases, many classification performance indicators including overall accuracy may provide misleading information [74,75]. For this reason, the mask accuracy comparison should focus on recall as per Equation (6) and, most importantly, precision as per Equation (5).

2.7. Impact of FMs Accuracy on Model-Assisted GSV Estimation

The five FMs were used to mask out all non-forest pixels in the national GSV map. The pixel-level predictions for the resulting five masked GSV maps were used with a model-assisted, generalized regression estimator to infer mean and total GSV at both national (NUTS1) and regional levels (NUTS2) [20–23]. An initial estimate of GSV can be calculated from the masked GSV maps as,

$$\hat{\mu}_{\text{initial}} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i \quad (7)$$

where N is the number of forest pixels within the masked GSV map and $\hat{y}_i$ is the GSV prediction obtained using the RF model for the i-th pixel. However, this estimator may be biased because of systematic prediction error. The bias can be estimated as,

$$\hat{\text{Bias}}(\hat{\mu}_{\text{initial}}) = \frac{1}{n}\sum_{j=1}^{n}(\hat{y}_j - y_j) \quad (8)$$

where n is the NFI sample size, i.e., the number of plots used for constructing the model, $\hat{y}_j$ is the GSV model prediction for the j-th plot and y_j the observed value of GSV for the j-th plot. Subtracting the estimated bias from the initial estimate yields the model-assisted estimator as,

$$\hat{\mu}_{ma} = \hat{\mu}_{\text{initial}} - \hat{\text{Bias}}(\hat{\mu}_{\text{initial}}) = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i - \frac{1}{n}\sum_{j=1}^{n}\left(\hat{y}_j - y_j\right) \quad (9)$$

where ma denotes model-assisted, $\hat{\mu}_{ma}$ is the estimate of mean GSV for the given masked GSV map, N is the number of forest pixels within the masked GSV map, $\hat{y}_i$ is the GSV prediction obtained using the RF model for the i-th pixel. The standard error (SE) for the estimator is:

$$SE(\hat{\mu}_{ma}) = \sqrt{\frac{1}{n(n-1)}\sum_{j=1}^{n}(e_j - \bar{e})^2} \quad (10)$$

where n is the NFI sample size, $e_j = \hat{y}_j - y_j$ and $\bar{e} = \frac{1}{n}\sum_{j=1}^{n} e_j$.

Similarly, the model-assisted estimator for the GSV total was:

$$\hat{\tau}_{ma} = \sum_{i=1}^{N} y_i - \frac{N}{n}\sum_{j=1}^{n}\left(\hat{y}_j - y_j\right) \quad (11)$$

where $\hat{\tau}_{ma}$ is the estimate of total GSV for the given GSV-masked map, N the number of pixels within the masked GSV map, $\hat{y}_i$ the GSV prediction obtained using the RF model for i-th pixel. The SE for the $\hat{\tau}_{ma}$ is given by d'Oliviero et al. [76]:

$$SE(\hat{\tau}_{ma}) = \sqrt{N^2\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{j=1}^{n}\frac{(e_j - \bar{e})^2}{n-1}} \quad (12)$$

where N is the population size, n is the NFI sample size, $e_j = \hat{y}_j - y_j$ and $\bar{e} = \frac{1}{n}\sum_{j=1}^{n} e_j$.

It is important to note that correction for estimated bias compensates for GSV map inaccuracy and makes the model-assisted estimator asymptotically unbiased.

Using the SEs, it was possible to construct confidence intervals for both estimates of mean and total GSV for the entire study area. These intervals are expressed as

$$\hat{E}_{ma} \pm t_n * SE(\hat{E}_{ma}) \quad (13)$$

where $\hat{E}_{ma}$ denotes either the model-assisted estimate of mean GSV or total GSV, $SE(\hat{E}_{ma})$ is the SE of $\hat{E}_{ma}$, and the factor t_n depends on the desired significance level and the distribution of the response variable. For most distributions and applications, t_n = 2 produces an approximate 95% confidence interval [23]. For purposes of constructing confidence intervals, the focus of the study was estimation of mean and total GSV and the SEs using the model-assisted regression estimators. To compare the GSV estimates produced with the five masked GSV maps and the NFI estimates at national and regional levels, we used the t statistic calculated as follows:

$$t = \frac{\hat{E}_{ma} - \hat{E}_{NFI}}{\sqrt{SE^2(\hat{E}_{ma}) + SE^2(\hat{E}_{NFI})}} \quad (14)$$

where $\hat{E}_{ma}$ denotes either the model-assisted estimate of mean GSV or total GSV for the masked GSV maps, $\hat{E}_{NFI}$ denotes either the *NFI* estimate of mean GSV or total GSV, and $SE^2(\hat{E}_{ma})$ and $SE^2(\hat{E}_{NFI})$ are the squares of the SEs of the estimates. Values of $|t| > 2$ indicate that the two estimates are statistically significantly different. Correlations for estimates of both mean and total estimates and the corresponding *NFI* estimates in terms of Pearson correlation coefficient ($\hat{\rho}_{Mean}$, $\hat{\rho}_{Total}$) were also calculated.

In addition, we calculated relative efficiency (*RE*) to assess the quality of the model-assisted estimators, compared to the SE obtained by the NFI [17], both at national and regional scales. *RE* was calculated as:

$$RE = \frac{\hat{Var}(\hat{E}_{NFI})}{\hat{Var}(\hat{E}_{ma})} \quad (15)$$

where $\hat{Var}(\hat{E}_{NFI})$ and $\hat{Var}(\hat{E}_{ma})$ are the estimated variances of the *NFI* estimates and the model-assisted estimates, respectively.

Values of RE greater than 1.0 are evidence of greater precision in the model-assisted estimates [77]. RE could be interpreted as the factor by which the original sample size would have to be increased to achieve the same precision as that achieved using the remotely sensed auxiliary data [17].

Finally, we evaluated the relationship between the accuracies of the FMs (in terms of overall accuracy, κ, precision and recall) and the SEs of the model-assisted estimates for the NUTS2 administrative level using the Pearson correlation coefficient ($\hat{\rho}$).

3. Results

3.1. Forest Mask Accuracy Assessment

At the national level, the most accurate FM was the NFM with an underestimation against the NFI estimates of only −2%, followed by the CLC06 with −3%, JAXA with

−4%, CLMS with +16%, and FM00 with +51%. The same ranking was obtained from the comparison with IUTI in terms of OA, κ, and precision (Table 2). For 17 of the 20 regions, the NFM was the most accurate, followed by the CLMS FM in two regions, and CLC06 in the remaining region. The confusion matrices for each one of the five FMs are shown in Figure 3.

Table 2. Accuracy assessment for the five forest masks (FMs) based on the confusion matrices with the IUTI.

Mask	Accuracy			
	OA (Equation (2))	κ (Equation (3))	Precision (Equation (5))	Recall (Equation (6))
CLMS	0.88	0.73	0.73	0.92
JAXA	0.85	0.61	0.71	0.74
FM00	0.76	0.51	0.55	0.91
CLC06	0.87	0.70	0.77	0.81
NFM	0.91	0.79	0.84	0.90

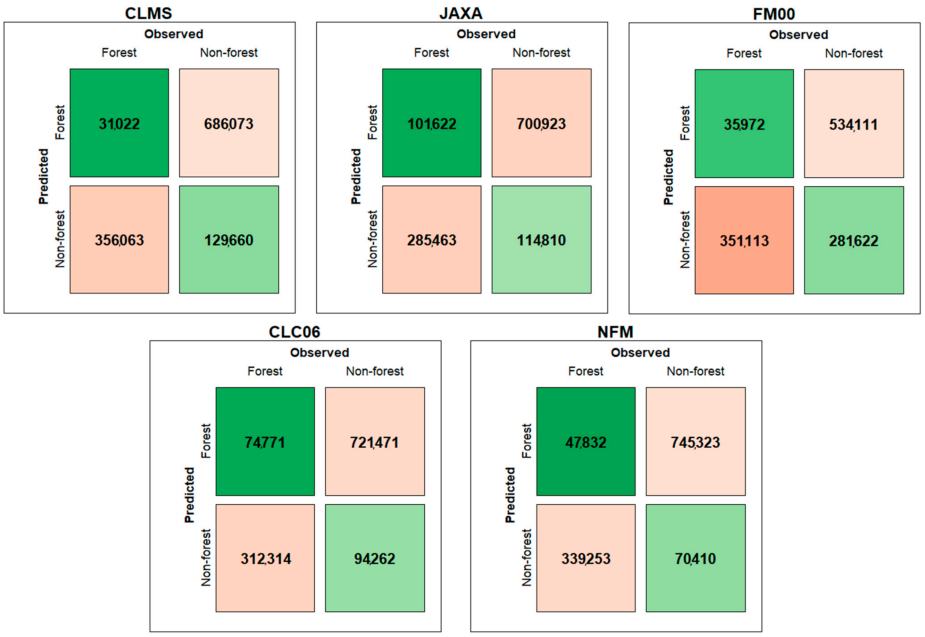


Figure 3. Confusion matrices of each forest mask.

We also noted that regardless of the FM used, the islands (Sicilia and Sardegna) and some of the southern regions (Calabria, Campania, Puglia) were characterized by small precision and recall (sensitivity), leading to numerous misclassifications of non-forest as forest (commission errors) (Figure 4).

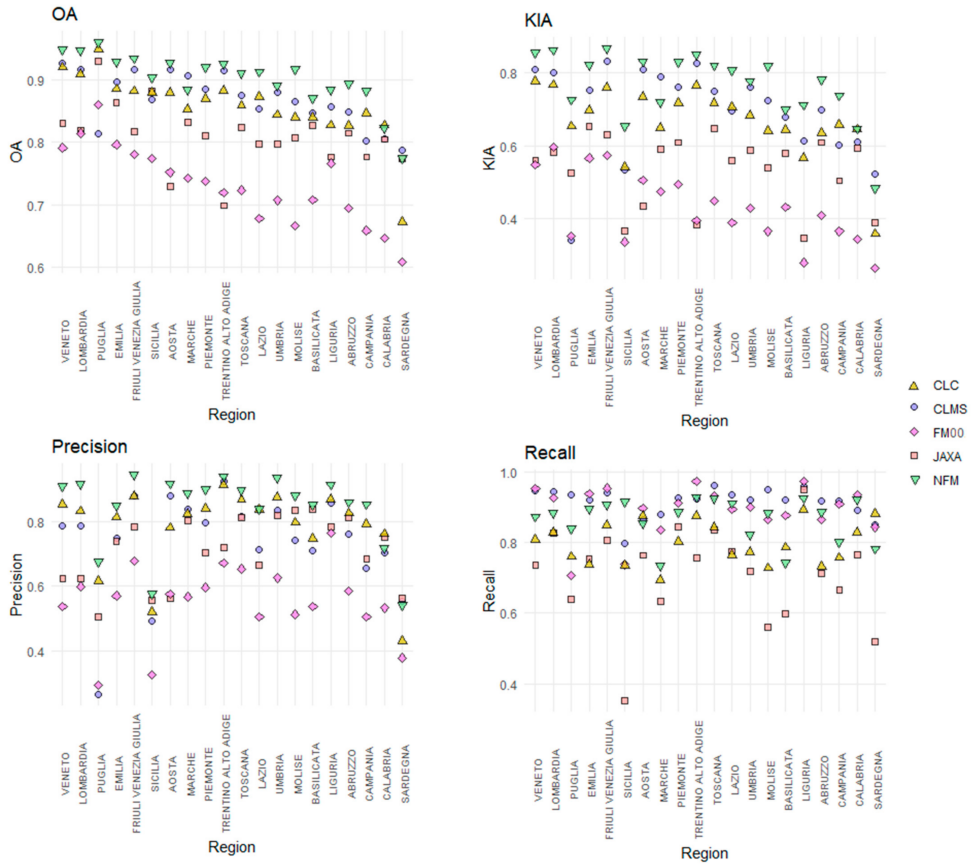


Figure 4. Comparison of four accuracy metrics among the FMs, calculated at regional level (NUTS2).

3.2. GSV Model-Assisted Estimations

In Figure 5, the GSV map of Italy produced with the RF model is reported.

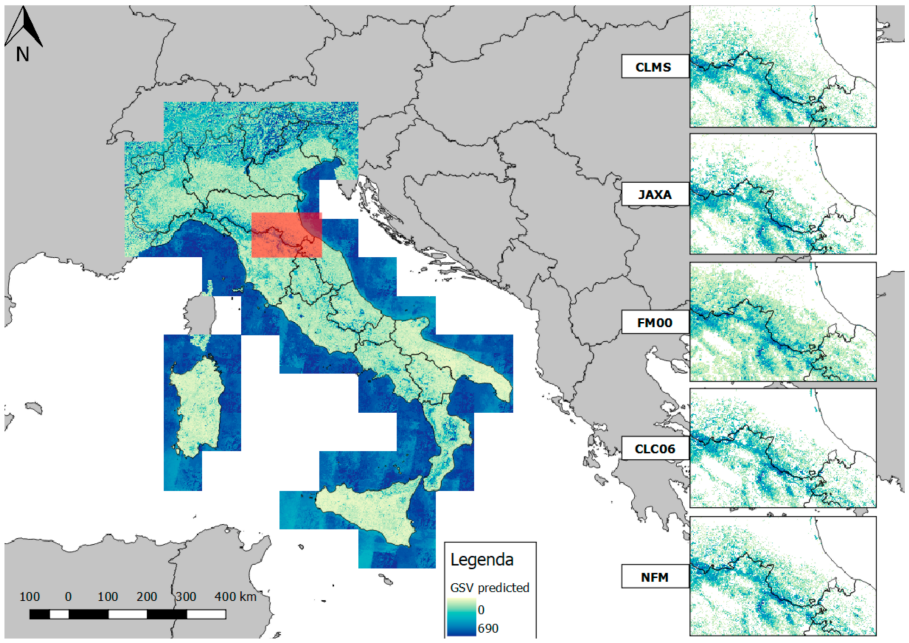


Figure 5. Growing stock map of Italy generated with random forests model. GSV in m^3 ha^{-1}. On the right, a detail of the GSV map masked with the five forest masks.

For the five masked GSV maps, $\hat{\mu}_{ma}$ ranged between 125 (CLMS) and 135 (NFM), $m^3\ ha^{-1}$ with $SE(\hat{\mu}_{ma})$ between 1.1 and 1.3 $m^3\ ha^{-1}$. For comparison, the design-based estimation of mean GSV from the NFI was 131 $m^3\ ha^{-1}$ with SE of 1.6 $m^3\ ha^{-1}$. Three of the five GSV-masked maps (NFM, CLC06, JAXA) produced estimates that were not statistically significantly different from the NFI estimate. The value of $\hat{\tau}_{ma}$ ranged between 1321 (JAXA) and 1525 (CLMS) millions m^3, with $SE(\hat{\tau}_{ma})$ between 13 (NFM) and 17 (JAXA) million m^3, while the official estimate from the NFI was 1366 million m^3 with SE of 14 million m^3, demonstrating a general trend towards overestimation of total volume (Table 3). The differences between the total GSV estimate for two of the five masked GSV maps (NFM, CLC06) and the NFI estimate were not statistically significantly different from 0.

Table 3. Model-assisted regression estimates for the five maps. The last row has the Italian NFI estimates.

Forest Mask	**Model-Assisted and NFI Estimates**						
	$\hat{u}_{ma}$	$SE(\hat{\mu}_{ma})$ ($m^3\ ha^{-1}$)	$t(\hat{u})$ (Equation (14_)	$\hat{\tau}_{ma}$ (m^3)	$SE(\hat{\tau}_{ma})$	$t(\hat{\tau})$ (Equation (14))	RE
CLMS	125	1.2	−3	1,525,000,000	14,487,500	7.9	1.17
JAXA	131	1.3	0	1,321,000,000	13,342,100	−2.3	1.09
FM00	113	1.1	−9.5	1,791,000,000	17,014,500	19.3	1.15
CLC06	135	1.3	1.94	1,387,000,000	13,572,900	1.0	1.12
NFM	134	1.2	1.5	1,371,000,000	13,037,800	0.26	1.16
INFC (NFI)	131	1.6	0	1,366,000,000	13,959,000	0	1

For the 20 NUTS2 administrative regions, the greatest correlation with the NFI estimates was achieved by the GSV map masked with the NFM mask with $\hat{\rho}$ = 0.972 and $\hat{\rho}$ = 0.986 for the mean and total GSV, respectively (Table 4). The GSV maps masked with the CLMS and FM00 masks, despite their large values of $\hat{\rho}$, show a systematic overestimation of the $\hat{\tau}_{ma}$.

Table 4. Coefficient of correlation between the mean and total model-assisted estimate and NFI estimates for administrative NUTS2 regions (* p-value = 0; ** p-value < 0.001).

Forest Mask	$\hat{\rho}_{Total}$	$\hat{\rho}_{Mean}$
CLMS	0.978 *	0.963 **
JAXA	0.968 **	0.971 **
FM00	0.979 *	0.949 **
CLC	0.977 **	0.970 **
NFM	0.986 *	0.972 *

Regarding $\hat{\mu}_{ma}$, for 16 of 20 regions, the differences between the model-assisted estimates and the NFI estimate were not statistically significantly different from 0 for the NFM masked GSV map, for 15 regions for CLMS and JAXA, for 14 regions for CLC06, and for 10 regions for FM00. Similar results were obtained for $\hat{\tau}_{ma}$ for which the differences for 16 of 20 regions were not statistically significantly different from 0 for the NFM masked GSV map, 15 for CLC06 and JAXA, six for CLMS, and two for FM00. The regions that always showed a statistically significant difference between the model-assisted estimates and the official NFI turned out to be the islands (Sardegna, Sicilia) and two regions (Puglia, Umbria), while those for which there were never significant differences were seven, distributed in northern and central Italy.

RE exceeded 1 for most regions, regardless of the FM used. RE < 1 was observed in one region for the CLMS and FM00 masks (Toscana), two regions for the CLC06 mask (Toscana, Emilia Romagna), and four regions for the JAXA mask (Toscana, Emilia Romagna, Sardegna, Umbria). The only masked GSV map that leads to RE coefficient always >1 was the NFM.

3.3. Relationship Between FMs Accuracy and GSV Estimates

The relationship between the accuracies of the FMs and the SEs of the estimates with the model-assisted estimator is presented in Table 5. The correlation was calculated for the 20 administrative regions.

Table 5. Correlation coefficient between the accuracy metrics and the SEs of estimates for each FM. The overall values were calculated based on all five FMs together.

Forest Mask	$\hat{\rho}$			
	Overall Accuracy	**κ**	**Precision**	**Recall**
CLMS	−0.26	−0.43	−0.48	−0.25
JAXA	0.26	−0.27	−0.36	−0.62
FM00	0.12	−0.24	−0.57	−0.68
CLC	0.09	−0.20	−0.39	−0.29
NFM	0.09	−0.26	−0.26	−0.58
Overall	0.03	−0.20	−0.32	−0.42

4. Discussion

The aim of this study was to assess the effects of using different FMs available for Italy for the area-based estimation of GSV. We first constructed a pixel-level GSV map for the entirety of Italy based on the procedure recently proposed by [17]. We then acquired five different FMs and, after evaluating their accuracies against an independent dataset (IUTI), we used them to mask out non-forest areas from the national GSV map produced with the random forest model. We then compared the five resulting model-assisted GSV estimates aggregated at regional levels with the official design-based NFI estimates.

Four of the five FMs achieved overall accuracies > 85%, based on the 2008 land use classification of IUTI points, with the CLC06 and NFM outperforming the other products. At the national level, the FM that achieved the greatest overall accuracy, κ and precision was the NFM, followed by the CLC06. Despite the greatest recall (0.91) achieved, the FM00 was affected by systematic overestimation of the regional forest area due to the original coarse resolution [1] which made this FM unsuitable for GSV estimation.

In contrast, the JAXA FM produced the smallest recall (0.74), most probably because the SAR backscatter in the HV polarization is relatively insensitive to Mediterranean vegetation [19,78] which probably caused an underestimation of the forest area. The photointerpreted FMs, CLC06 and NFM, had the greatest precision. This is an expected result because forest land use identification is typically done by local experts. However, CLC06 produced less precision than the NFM because it was implemented for monitoring land cover, not land uses, adopting a MMU and a crown cover threshold greater than that adopted by the INFC 2005 [53,79]. In fact, the CLC project did not map forest clear-cuts and other natural or anthropic disturbances as forest land use, but rather as bare soil or other non-forest classes, affecting the estimation of forest area. Conversely, the NFM, as a mosaic of local forest maps, is designed to monitor forest land use, such as the NFI. However, the small precision of the accuracy showed that false positives were the majority of classification errors.

At the regional level, OA was greater than 85% for 18 regions for the NFM mask, followed by the CLMS mask (14 regions), the CLC06 mask (12 regions), the JAXA mask (3 regions), and the FM00 mask (1 region). Regardless of the FM used, the greatest uncertainty was found in the southern regions and the islands (Campania, Calabria, Abruzzo, Basilicata, Sardegna, Sicilia), most probably because of the complex Mediterranean formations and complex agroforestry landscape tiles that characterized these regions where the NFI estimates also have larger associated SEs.

The greatest accuracies were achieved for regions characterized by greater forest cover (Liguria, Trentino-Alto Adige, Friuli-Venezia Giulia, Umbria, Toscana). These regions are characterized by extensive forests with continuous coverage and greater accumulation of

GSV, as in the Apennine and Alpine Mountains, which probably reduces the likelihood of forest misclassifications, regardless of the FM considered.

Conversely, the forests bordering other land uses, along rivers, and in the coastal and rural contexts are typically characterized by a sparse canopy, which makes them more difficult to correctly classify, even by manual photointerpretation.

In conclusion, regarding the qualities of the FMs, the most accurate was the NFM, which was comparable with the CLC06, but with the advantage of a finer MMU which makes it more suitable for regional and local scale applications.

Regarding the model-assisted GSV estimates, although all the masked GSV maps overestimated total GSV, the NFM masked GSV map was most accurate as a trade-off between the national and regional GSV and the SE of estimates. The general overestimation was caused by the trend of the prediction model to overpredict GSV for pixels with small observed GSV values. (i.e., GSV < 250 m^3 ha^{-1}). This evidence, along with the limited GSV that characterizes Italian forests, caused the general overestimation at the national level. One possible solution is to increase the performance of the model, for example, by integrating ALS metrics which is a well-established data source for enhancing GSV predictions [8,13,15]. Both the CLMS and FM00 masked GSV maps suffered from systematic prediction error which caused the overestimation of $\hat{\tau}_{ma}$, both nationally and regionally. For the CLMS masked GSV map, this can be caused by the inclusion of many agricultural and rural areas that occur in Italy [46], and for FM00 because of the original coarse spatial resolution (1 × 1 km). The differences between the model-assisted total GSV estimates and the official NFI estimate for two of the five masked GSV maps (NFM, CLC06) were statistically significantly different from 0. At the national level, the mean GSV estimates were comparable for all maps, except for the GSV map masked with the FM00 mask. The JAXA masked GSV map produced the same estimate as the NFI for mean GSV but underestimated the total due to the underestimation of forest area. However, the SEs were almost comparable for all the GSV-masked maps considered. The SE is mainly affected by the number of NFI plots used for building the model and calculation of the correction term in the estimator. Despite the differences among the FMs, the NFI plots falling within the forested portions of the FMs were similar, ranging between 6100 (CLMS) and 5800 (JAXA). Differences in the number of plots selected by each FM are likely to be concentrated at the forest edge, where maps are more prone to classification errors. These results confirm the findings of Esteban et al. [56], suggesting that the FM effects on area estimates are more important than the effects of field plot sampling variability on the uncertainty of the mean and total estimates.

At the regional level, the NFM produced the greatest $\hat{\rho}$ relative to the NFI estimates, both for $\hat{\mu}_{ma}$ and $\hat{\tau}_{ma}$, with the largest number of regional estimates in accordance with the NFI (16 regions out of 20). The NFM was also the only FM that led consistently to RE > 1. The CLC06 achieved similar results, with the major exception of Sardegna and in general in the southern regions, where, as we reported before, the MMU of the CLC project is not fine enough to discern the complex patchwork in the landscape of a rural region.

$SE(\hat{\tau}_{ma})$ was smaller than $SE(\hat{\tau}_{NFI})$ for 16 regions, which represent 70% of the Italian territory. The regions with the greatest $SE(\hat{\tau}_{ma})$ were Puglia, Valle d'Aosta, Molise, Basilicata, and Marche ($SE(\hat{\tau}_{ma})$ > 5%) probably because of the small number of NFI plots in these regions. Nevertheless, with the use of the model-assisted estimation approach, it was possible to decrease the error of the estimates with respect to the NFI estimates, both at the national (NUTS1), and regional levels (NUTS2).

Regarding the relationship between the FM accuracy and the SEs of the estimates, we found small correlation coefficients, in particular with the overall accuracy. The SE depends primarily on the sample size, which is less affected by the accuracy of the FMs, as reported by Esteban et al. [56]. The accuracy metric was more correlated with the SE of the estimates than was the recall, followed by the precision. This is an expected result because these metrics are strictly related to the area classified as forest which, in turn, affects the number

of NFI plots included in the FMs. Of interest, the FM with the greatest recall (CLMS) was also the FM that included the greatest number of NFI plots.

However, the negative correlation with the other accuracy metrics demonstrated that a more accurate FM leads to a smaller $SE(\hat{\tau}_{ma})$.

It would be interesting to combine the available maps by aggregating their beneficial features to overcome the problems associated with each FM as per McRoberts et al. [23]. Another option would be to calibrate the FMs using the NFI data as per Næsset et al. [80].

In conclusion, the differences in the accuracies of the FMs led to different GSV estimates, although the SEs were almost comparable. The smallest GSV difference against the official NFI estimate was obtained by the most accurate FMs, i.e., the NFM. This is likely due to the correct classification of the main, dense forests, which have the largest amount of volume and subsequently make the greatest contribution in the model-assisted estimation. Presumably, forest misclassification occurs mainly along the margins and in boundary areas between different land uses.

5. Conclusions

This paper presents a comparative analysis of the impacts of different forest masks on model-assisted estimation of GSV. Several conclusions can be drawn from this study.

- At national and regional levels, the masked GSV map constructed using the NFM mask produced GSV estimates that were most similar to the official NFI estimates. Regardless of the forest mask, the major disagreement with the official estimate was found in the southern regions and islands, most probably because of the presence of the Mediterranean macchia, which is difficult to classify correctly, even by manual photointerpretation of fine-resolution images. These were the regions with the least classification accuracies. Regions with abundant forest components (central and northern regions) produced the most accurate masks and the most accurate and most precise GSV estimates.
- Despite the small correlation coefficients, we found a negative relationship between forest mask accuracy and the standard error of the GSV estimate, demonstrating that the accuracy of the FM must be considered in the GSV estimation through the model-assisted estimator.
- The quality of the model-assisted estimation mostly depends on the performance of the prediction model. A more accurate FM can compensate for systematic model prediction errors, leading to a greater agreement with official NFI GSV estimates, both at national and regional levels.

In conclusion, we recommend using the NFM, both at regional and national levels, for studies aimed at estimating forest parameters related to variables such as forest area, GSV, AGB, and carbon stock. However, it is plausible to assume that as the accuracy of the model predictions increases thanks to the growing availability of 3D data, the NFM will always produce more accurate and precise estimates of total GSV. In this regard, we hope that in the future, wall-to-wall ALS coverage will be finally available in Italy, to enhance the prediction of forest variables with even greater accuracy.

Finally, we strongly recommended that the different forest mapping and monitoring programs currently active in Italy converge on a common method and lead to harmonized, multiscale systems in line with the international forest definition.

Author Contributions: Conceptualization, G.C., E.V. and F.G.; methodology, E.V., G.D., S.F.; software, E.V.; validation, E.V.; formal analysis, E.V., G.D.; investigation, E.V., G.D.; resources, G.C., M.M. and B.L.; data curation, E.V., G.D.; writing—original draft preparation, E.V., G.C., R.E.M.; writing—review and editing, G.C., R.E.M., F.G., M.M. and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schepaschenko, D.; See, L.; Lesiv, M.; McCallum, I.; Fritz, S.; Salk, C.; Moltchanova, E.; Perger, C.; Shchepashchenko, M.; Shvidenko, A.; et al. Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics. *Remote Sens. Environ.* **2015**, *162*, 208–220. [CrossRef]
2. FAO. Global Forest Resources Assessment 2010. Terms and Definition. Working paper 144/E. 2010. Available online: http://www.fao.org/3/am665e/am665e00.pdf (accessed on 7 October 2020).
3. FAO; UNCCD. *Sustainable Financing for Forest and Landscape Restoration: The Role of Public Policy Makers*; FAO: Rome, Italy, 2015; p. 12.
4. McRoberts, R.E.; Naesset, E.; Gobakken, T. Accuracy and Precision for Remote Sensing Applications of Nonlinear Model-Based Inference. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 27–34. [CrossRef]
5. Wittke, S.; Yu, X.; Karjalainen, M.; Hyyppä, J.; Puttonen, E. Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *76*, 167–178. [CrossRef]
6. Hansen, M.H.; Madow, W.G.; Tepping, B.J. An evaluation of model dependent and probability-sampling inferences in sample surveys. *J. Am. Stat. Assoc.* **1983**, *78*, 776–793. [CrossRef]
7. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [CrossRef]
8. Næsset, E.; Gobakken, T.; Holmgren, J.; Hyyppä, H.; Hyyppä, J.; Maltamo, M.; Nilsson, M.; Olsson, H.; Persson, Å.; Söderman, U. Laser scanning of forest resources: The nordic experience. *Scand. J. For. Res.* **2004**, *19*, 482–499. [CrossRef]
9. Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999. [CrossRef]
10. Hollaus, M.; Dorigo, W.; Wagner, W.; Schadauer, K.; Höfle, B.; Maier, B. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *Int. J. Remote Sens.* **2009**, *30*, 5159–5175. [CrossRef]
11. Waser, L.T.; Schwarz, M. Comparison of large-area land cover products with national forest inventories and CORINE land cover in the European Alps. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 196–207. [CrossRef]
12. Waser, L.T.; Fischer, C.; Wang, Z.; Ginzler, C. Wall-to-Wall Forest Mapping Based on Digital Surface Models from Image-Based Point Clouds and a NFI Forest Definition. *Forests* **2015**, *6*, 4510–4528. [CrossRef]
13. Kangas, A.; Astrup, R.; Breidenbach, J.; Fridman, J.; Gobakken, T.; Korhonen, K.T.; Maltamo, M.; Nilsson, M.; Nord-Larsen, T.; Næsset, E.; et al. Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scand. J. For. Res.* **2018**, *33*, 397–412. [CrossRef]
14. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [CrossRef]
15. Næsset, E. Area-Based Inventory in Norway–From Innovation to an Operational Reality. In *Forestry Applications of Airborne Laser Scanning Concepts Case Study*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 215–240. [CrossRef]
16. Puletti, N.; Floris, A.; Scrinzi, G.; Chianucci, F.; Colle, G.; Michelini, T.; Pedot, N.; Penasa, A.; Scalercio, S.; Corona, P.; et al. CFOR: A spatial decision support system dedicated to forest management in Calabria. *For. Riv. Selvic. Ed Ecol. For.* **2017**, *14*, 135–140. [CrossRef]
17. Chirici, G.; Giannetti, F.; McRoberts, R.E.; Travaglini, D.; Pecchi, M.; Maselli, F.; Chiesi, M.; Corona, P. Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *84*, 101959. [CrossRef]
18. Giannetti, F.; Puletti, N.; Puliti, S.; Travaglini, D.; Chirici, G. Modelling Forest structural indices in mixed temperate forests: Comparison of UAV photogrammetric DTM-independent variables and ALS variables. *Ecol. Indic.* **2020**, *117*, 106513. [CrossRef]
19. D'Amico, G.; Vangi, E.; Francini, S.; Giannetti, F.; Nicolaci, A.; Travaglini, D.; Massai, L.; Giambastiani, Y.; Terranova, C.; Chirici, G. Are We Ready for a Web-Based National Forest Information System? State of the Art of for-Est Maps and Airborne Laser Scanning Data Availability in Italy. *iForest*. [CrossRef]
20. Särndal, C.-E.; Swensson, B.; Wretman, J. *Model Assisted Survey Sampling*; Springer: New York, NY, USA, 1992; pp. 402, 694.
21. Särndal, C.-E.; Swensson, B.; Wretman, J. *Model Assisted Survey Sampling*; Springer: Berlin, Germany, 2003. Available online: https://www.amazon.com/Assisted-Survey-Sampling-Springer-Statistics/dp/0387406204 (accessed on 6 November 2020).
22. Breidt, F.J.; Opsomer, J.D. Chapter 27-nonparametric and semiparametric estimation in complex surveys. In *Handbook of Statistics*; Rao, C.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2009; pp. 103–119. [CrossRef]
23. McRoberts, R.E.; Vibrans, A.C.; Sannier, C.; Næsset, E.; Hansen, M.C.; Walters, B.F.; Lingner, D.V. Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area. *Can. J. For. Res.* **2016**, *46*, 924–932. [CrossRef]

24. Næsset, E.; Gobakken, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090. [CrossRef]
25. McRoberts, R.E. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sens. Environ.* **2010**, *114*, 1017–1025. [CrossRef]
26. Giannetti, F.; Chirici, G.; Gobakken, T.; Næsset, E.; Travaglini, D.; Puliti, S. A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. *Remote Sens. Environ.* **2018**, *213*, 195–205. [CrossRef]
27. Goodbody, T.R.H.; Coops, N.C.; White, J.C. Digital Aerial Photogrammetry for Updating Area-Based Forest Inventories: A Review of Opportunities, Challenges, and Future Directions. *Curr. For. Rep.* **2019**, *5*, 55–75. [CrossRef]
28. Barrett, F.; McRoberts, R.E.; Tomppo, E.; Cienciala, E.; Waser, L.T. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sens. Environ.* **2016**, *174*, 279–289. [CrossRef]
29. Saarela, S.; Holm, S.; Grafström, A.; Schnell, S.; Næsset, E.; Gregoire, T.G.; Nelson, R.F.; Ståhl, G. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Ann. For. Sci.* **2016**, *73*, 895–910. [CrossRef]
30. Holm, S.; Nelson, R.; Ståhl, G. Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne lidar, and space lidar. *Remote Sens. Environ.* **2017**, *197*, 85–97. [CrossRef]
31. Nilsson, M.; Nordkvist, K.; Jonzén, J.; Lindgren, N.; Axensten, P.; Wallerman, J.; Egberth, M.; Larsson, S.; Nilsson, L.; Eriksson, J.; et al. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sens. Environ.* **2017**, *194*, 447–454. [CrossRef]
32. Immitzer, M.; Stepper, C.; Böck, S.; Straub, C.; Atzberger, C. Use of WorldView-2 stereo imagery and National Forest Inventory data for wall-to-wall mapping of growing stock. *For. Ecol. Manag.* **2016**, *359*, 232–246. [CrossRef]
33. Karlson, M.; Ostwald, M.; Reese, H.; Sanou, J.; Tankoano, B.; Mattsson, E. Mapping tree canopy cover and above-ground biomass in Sudano-Sahelian woodlands using landsat 8 and random forest. *Remote Sens.* **2015**, *7*, 10017–10041. [CrossRef]
34. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *Isprs J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
35. Stankiewicz, K.; Dąbrowska-Zielińska, K.; Gruszczynska, M.; Hoscilo, A. Mapping vegetation of a wetland ecosystem by fuzzy classification of optical and microwave satellite images supported by various ancillary data. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology, Crete, Greece, 17 March 2003. [CrossRef]
36. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef]
37. Dostálová, A.; Hollaus, M.; Milenković, M.; Wagner, W. Forest area derivation from sentinel-1 data. *Isprs Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-7*, 227–233. [CrossRef]
38. Eysn, L.; Hollaus, M.; Schadauer, K.; Pfeifer, N. Forest Delineation Based on Airborne LIDAR Data. *Remote Sens.* **2012**, *4*, 762–783. [CrossRef]
39. Dalponte, M.; Ørka, H.O.; Ene, L.T.; Gobakken, T.; Næsset, E. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens. Environ.* **2014**, *140*, 306–317. [CrossRef]
40. Rudjord, O.; Trier, O.D. Tree species classification with hyperspectral imaging and lidar. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 12 February 2016; p. 4. [CrossRef]
41. Øivind, D.T.; Salberg, A.-B.; Kermit, M.; Øystein, R.; Gobakken, T.; Næsset, E.; Aarsten, D. Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. *Eur. J. Remote Sens.* **2018**, *51*, 336–351. [CrossRef]
42. Woodcock, C.E.; Allen, R.; Anderson, M.; Belward, A.; Bindschadler, R.; Cohen, W.; Gao, F.; Goward, S.N.; Helder, D.; Helmer, E.; et al. Free Access to Landsat Imagery. *Science* **2008**, *320*, 1011a. [CrossRef] [PubMed]
43. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **2019**, *225*, 127–147. [CrossRef]
44. Olofsson, P.; Arévalo, P.; Espejo, A.B.; Green, C.; Lindquist, E.; McRoberts, R.E.; Sanz, M.J. Mitigating the effects of omission errors on area and area change estimates. *Remote Sens. Environ.* **2020**, *236*, 111492. [CrossRef]
45. European Enviromental Agency. *Enviromental Statement*; Office for Official Publications of the European Communities: Luxembourg, 2007; ISBN 978-92-9167-936-2.
46. Langanke, T. *Copernicus Land Monitoring Service–High Resolution Layer Forest: Product Specifications Document 38*; Copernicus team at EEA, 2017. Available online: https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-high (accessed on 5 November 2020).
47. JAXA. *Global 25m Resolution PALSAR-2/PALSAR Mosaic and Forest/Non-Forest Map (FNF) Dataset Description*; Japan Aerospace Exploration Agency (JAXA), Earth Observation Research Center (EORC), 2016. Available online: https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/DatasetDescription_PALSAR2_Mosaic_FNF_revE.pdf (accessed on 5 November 2020).
48. Seebach, L.; McCallum, I.; Fritz, S.; Kindermann, G.; LeDuc, S.; Böttcher, H.; Fuss, S. Choice of forest map has implications for policy analysis: A case study on the EU biofuel target. *Environ. Sci. Policy* **2012**, *22*, 13–24. [CrossRef]
49. Fritz, S.; See, L. Comparison of land cover maps using fuzzy agreement. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 787–807. [CrossRef]
50. Giri, C.; Zhu, Z.; Reed, B. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sens. Environ.* **2005**, *94*, 123–132. [CrossRef]

51. Hoyos, A.P.; Rembold, F.; Kerdiles, H.; Gallego, J. Comparison of Global Land Cover Datasets for Cropland Monitoring. *Remote Sens.* **2017**, *9*, 1118. [CrossRef]
52. Neumann, K.; Herold, M.; Hartley, A.; Schmullius, C. Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 425–437. [CrossRef]
53. Seebach, L.M.; Strobl, P.; Miguel-Ayanz, J.S.; Gallego, J.; Bastrup-Birk, A. Comparative analysis of harmonized forest area estimates for European countries. *Forests* **2011**, *84*, 285–299. [CrossRef]
54. Rodríguez-Veiga, P.; Saatchi, S.; Tansey, K.; Balzter, H. Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. *Remote Sens. Environ.* **2016**, *183*, 265–281. [CrossRef]
55. Li, Z.-W.; Xin, X.-P.; Tang, H.; Yang, F.; Chen, B.-R.; Zhang, B.-H. Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. *J. Integr. Agric.* **2017**, *16*, 286–297. [CrossRef]
56. Esteban, J.; McRoberts, R.E.; Fernández-Landa, A.; Tomé, J.L.; Marchamalo, M. A Model-Based Volume Estimator that Accounts for Both Land Cover Misclassification and Model Prediction Uncertainty. *Remote Sens.* **2020**, *12*, 3360. [CrossRef]
57. Barbati, A.; Marchetti, M.P.; Chirici, G.; Corona, P. European Forest Types and Forest Europe SFM indicators: Tools for monitoring progress on forest biodiversity conservation. *For. Ecol. Manag.* **2014**, *321*, 145–157. [CrossRef]
58. Gasparini, P.; De Natale, F.; Di Cosmo, L.; Gagliano, C.; Salvadori, I.; Tabachi, G.; Tosi, V. INFC, 2009–I caratteri quantitativi–parte 1, vers. 2. In *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio*; MiPAAF–Ispettorato Generale Corpo Forestale dello Stato, CRA-MPF: Trento, Italy, 2009.
59. INFC. Le stime di superficie 2005-seconda parte. In *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio*; Tabacchi, A.G., De Natale, F., Di Cosmo, L., Floris, A., Gagliano, C., Gasparini, P., Salvadori, I., Scrinzi, G., Tosi, V., Eds.; MiPAF–Corpo Forestale dello Stato-Ispettorato Generale, CRA-ISAFA: Trento, Italy, 2007. Available online: http://www.infc.it (accessed on 5 November 2019).
60. Fattorini, L.; Marcheselli, M.; Pisani, C. A three-phase sampling strategy for large-scale multiresource forest inventories. *J. Agric. Biol. Environ. Stat.* **2006**, *11*, 296–316. [CrossRef]
61. Tabacchi, G.; Di Cosmo, L.; Gasparini, P.; Morelli, S. *Stima Del Volume E Della Fitomassa Delle Principali Specie Forestali Italiene, Equazioni Di Previsione, Tavole Del Volume E Tavole Della Fitomassa Arborea Epigea*; Consiglio per la Ricerca e Sperimentazione in Agricoltura, Unità di Ricerca per Il Monitoraggio e la Pianificazione Forestale: Trento, Italy, 2011.
62. McRoberts, R.E.; Chen, Q.; Walters, B.F.; Kaisershot, D.J. The effects of global positioning system receiver accuracy on airborne laser scanning-assisted estimates of aboveground biomass. *Remote Sens. Environ.* **2018**, *207*, 42–49. [CrossRef]
63. Penman, J.; Gytarsky, M.; Hiraushi, T.; Krug, T.; Kruger, D.; Pipatti, R.; Buendia, L.; Miwa, K.; Ngara, T.; Tanabe, K.; et al. *Good Practice Guidance for Land Use, Land Use Change and Forestry. Chapter 3: Annex 3A.1 Biomass Default Tables for Section 3.2 Forest Land Good Practice Guidance for Land Use, Land-Use Change and Forestry*; The Institute for Global Enviromental Strategies for the IPCC and the Intergovernmental Panel on Climate Change, Hayama: Kanagawa, Japan, 2003; p. 21.
64. Romano, D.; Arcarese, C.; Bernetti, A.; Caputo, A.; Condor, R.D.; Contaldi, M.; Lauretis, R.; Di Cristofaro, E.; Federici, S.; Gagna, A.; et al. *Italian Greenhouse Gas Inventory 1990–2009. National Inventory Report*; ISPRA: Rome, Italy, 2011.
65. Corona, P.; Barbati, A.; Tomao, A.; Bertani, R.; Valentini, R.; Marchetti, M.; Fattorini, L.; Perugini, L. Land use inventory as framework for environmental accounting: An application in Italy. *Iforest Biogeosci. For.* **2012**, *5*, 204–209. Available online: http://www.sisef.it/iforest/contents?id=ifor0625-005 (accessed on 6 November 2020). [CrossRef]
66. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.-K. A Land-sat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72. [CrossRef]
67. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]
68. Kennedy, R.E.; Yang, Z.; Gorelick, N.; Braaten, J.; Cavalcante, L.; Cohen, W.B.; Healey, S. Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sens.* **2018**, *10*, 691. [CrossRef]
69. Maselli, F.; Pasqui, M.; Chirici, G.; Chiesi, M.; Fibbi, L.; Salvati, R.; Corona, P. Modeling primary production using a 1 km daily meteorological data set. *Clim. Res.* **2012**, *54*, 271–285. [CrossRef]
70. Panagos, P. *The European Soil Database*; GEO, 2006; pp. 32–33. Available online: https://www.researchgate.net/publication/224842031_The_European_Soil_Database (accessed on 5 November 2020).
71. Büttner, G.; Feranec, J.; Jaffrain, G.; Mari, L.; Maucha, G.; Soukup, T. The corine land cover 2000 project. EARSeL eProceedings 3, 3/2004 331. 2004. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.618.9940&rep=rep1&type=pdf (accessed on 6 February 2021).
72. McRoberts, R.E.; Næsset, E.; Gobakken, T. Optimizing the k-Nearest Neighbors technique for estimating forest above-ground biomass using airborne laser scanning data. *Remote Sens. Environ.* **2015**, *163*, 13–22. [CrossRef]
73. Liaw, A.; Wiener, M. Classification and regression by randomForest. *Nucleic Acids Res.* **2002**, *5*, 983–999. [CrossRef]
74. Devarriya, D.; Gulati, C.; Mansharamani, V.; Sakalle, A.; Bhardwaj, A. Unbalanced breast cancer data classification using novel fitness functions in genetic programming. *Expert Syst. Appl.* **2020**, *140*, 112866. [CrossRef]
75. Jaafor, O.; Birregah, B. KNN-LC: Classification in Unbalanced Datasets using a KNN-Based Algorithm and Local Centralities. In *Data-Driven Modeling for Sustainable Engineering*; Adjallah, K., Birregah, B., Abanda, H., Eds.; Springer: Cham, Switzerland, 2017; Volume 72. [CrossRef]

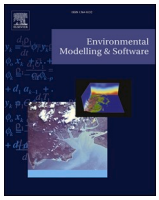
76. D'Oliveira, M.V.; Reutebuch, S.E.; McGaughey, R.J.; Andersen, H.-E. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sens. Environ.* **2012**, *124*, 479–491. [CrossRef]
77. Moser, P.; Vibrans, A.C.; McRoberts, R.E.; Næsset, E.; Gobakken, T.; Chirici, G.; Mura, M.; Marchetti, M. Methods for variable selection in LiDAR-assisted forest inventories. *Forestry* **2016**, *90*, 112–124. [CrossRef]
78. Bartsch, A.; Widhalm, B.; Leibman, M.; Ermokhina, K.; Kumpula, T.; Skarin, A.; Wilcox, E.J.; Jones, B.M.; Frost, G.V.; Höfler, A.; et al. Feasibility of tundra vegetation height retrieval from Sentinel-1 and Sentinel-2 data. *Remote Sens. Environ.* **2020**, *237*, 111515. [CrossRef]
79. Vizzarri, M.; Chiavetta, U.; Chirici, G.; Garfì, V.; Bastrup-Birk, A.; Marchetti, M. Comparing multisource harmonized forest types mapping: A case study from central Italy. *Iforest-Biogeosci. For.* **2015**, *8*, 59–66. Available online: http://www.sisef.it/iforest/contents/?id=ifor1133-007 (accessed on 5 November 2020). [CrossRef]
80. Næsset, E. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accom-plished in Scandinavia. *Scand. J. For. Res.* **2007**, *22*, 433–442. [CrossRef]

Environmental Modelling and Software 159 (2023) 105580

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



Large-scale high-resolution yearly modeling of forest growing stock volume and above-ground carbon pool

Elia Vangi [a,b], Giovanni D'Amico [a,c,*], Saverio Francini [a,d], Costanza Borghi [a], Francesca Giannetti [a], Piermaria Corona [c], Marco Marchetti [b], Davide Travaglini [a], Guido Pellis [e], Marina Vitullo [e], Gherardo Chirici [a]

[a] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università degli Studi di Firenze, Italy
[b] Dipartimento di Bioscienze e Territorio, Università degli Studi del Molise, Italy
[c] CREA Research Centre for Forestry and Wood, Italy
[d] Fondazione per il Futuro delle Città, Firenze, Italy
[e] Istituto Superiore per la Protezione e la Ricerca Ambientale, Italy

ARTICLE INFO

Keywords:
National forest inventory
GSV
Carbon stock
Forest modeling
Spatial modeling
Italy

ABSTRACT

Within the Paris Agreement's Enhanced Transparency Framework, consistent data collections are the prerequisite for a successful reporting of GHG emissions. For such purposes, NFIs are usually the primary source of information, even if they are frequently not designed for producing estimations on a yearly basis and in the form of wall-to-wall high-resolution maps. In this framework, we present a new spatial model to produce yearly growing stock volume (GSV), above-ground biomass (AGB), and carbon stock wall-to-wall estimates. We tested the model in Italy for the period 2005–2018, obtaining a time-series of yearly maps at 23 m spatial resolution. Results were validated against the 2015 Italian NFI reaching an average RMSE% of 19% for aggregated areas. Results were also compared against data reported by the Italian GHG inventory, reaching an RMSE% of 28% and 20% for GSV and carbon stock respectively.

We demonstrated that the modeling approach can be successfully used for setting up a forest monitoring system to meet the interests of governments in inventories of GHG emissions and private entities in carbon offset investments.

1. Introduction

Under the enhanced transparency framework of the Paris Agreement, each country Party must report every two years an inventory of their anthropogenic greenhouse gases (GHGs) emissions by sources and removals by sinks following the Intergovernmental Panel on Climate Change (IPCC) guidelines and guidance (IPCC et al., 2006). The GHG emission inventory has to fulfill the IPCC key principles: transparency, accuracy, completeness, consistency, and comparability while providing helpful information for assessing the climate impacts. The "Land Use, Land-Use Change and Forestry" (LULUCF) is exceptionally demanding, dealing with natural carbon dynamics and aiming to assess emissions and removals related to the impact of anthropogenic activities. The LULUCF sector is responsible for significant GHG emissions globally, mainly due to deforestation activities. In this framework, forests are pivotal ecosystems, being a substantial and growing atmospheric carbon sink (Sellers et al., 2018). Forests are estimated to sequester 30% of the total global CO_2 released into the atmosphere annually (Houghton and Nassikas, 2017), corresponding to 7.6 Gt CO_2 y^{-1}, reflecting a balance between gross carbon removals and gross emissions from deforestation and other disturbances (Harris et al., 202; Xu et al., 2021). Increasing the carbon stored in the above and below-ground forest biomass is a mitigation mechanism to fight climate change and offset anthropogenic emissions worldwide (Di Cosmo et al., 2016).

Despite the UNFCCC requirements related to the provision by Parties of biennial forestry-related carbon stock change, many National Forest Inventories (NFI) are not designed for continuous yearly reports and cannot cope with the required reporting frequency due to longer update cycles (McRoberts et al., 2018). Estimating carbon stock changes between consecutive NFIs is a pivotal step in accomplishing the reporting requirements. The methodology should be based on year-to-year measured forest variables or prediction models to extend NFI-based

* Corresponding author. Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università degli Studi di Firenze, Italy.
E-mail address: giovanni.damico@unifi.it (G. D'Amico).

https://doi.org/10.1016/j.envsoft.2022.105580
Received 23 March 2022; Received in revised form 4 November 2022; Accepted 8 November 2022
Available online 11 November 2022

estimates to assessment years rather than a simple interpolation between estimates produced by NFIs at different years (Federici et al., 2008).

Even if the main source of information for such reporting activities are NFIs (Tomppo et al., 2010; Condés and McRoberts, 2017; Kulbokas et al., 2019), in recent times, considerable efforts have been laid out to integrate remotely sensed (RS) data in the process. Examples are available to provide spatially continuous (also referred to as wall-to-wall maps) and updated estimations of several forest variables such as: the growing stock volumes (GSV), the above-ground biomass (AGB) (Kangas et al., 2018; Chirici et al., 2020; Vangi et al., 2021), and the rate of forest disturbances (Hansen et al., 2013; van der Werf et al., 2017; Francini et al., 2021; Francini et al., 2022a, a; Francini et al., 2022b, b). Coupling traditional NFI information acquired in the field with such wall-to-wall maps based on remotely sensed data is the basis for evolving from traditional NFIs to the new so called Enhanced Forest Inventory (EFI) framework (White et al., 2016). This has already been carried out by Countries with a long history in NFIs, such as those in the Scandinavian area (Næsset et al., 2004; Nord-Larsen and Schumacher, 2012; Tomppo et al., 2008), Canada (White et al., 2016), Austria (Hollaus et al., 2009) and Switzerland (Waser et al., 2017). The EFI approach has several benefits (Chirici et al., 2020): it enables the estimation of forest attributes from a local to national scale to support local management and national planning; it can provide estimates of forest removals due to logging and other disturbances, which are essential in the context of carbon cycle assessments (Francini et al., 2021). But evolving from traditional NFIs to EFIs requires elaborating a huge amount of remotely sensed (RS) data which in turn requires investments in software and hardware resources for their processing (D'Amico et al., 2021). Conversely, field activities can be reduced by optimizing the sampling strategy by integrating RS data (Corona, 2010). For example biomass density maps constructed from remotely sensed data can be used to enhance the stratification of ground inventories, to supply carbon stock changes estimates in poorly-sampled or unapproachable areas, or for verification purposes (ISPRA, 2021a).

Coming more specifically to the problem of how to estimate forest carbon stock changes, there are at least three approaches reported in literature that are based on different level of data availability (Williams et al., 2012). The first one is the gain-loss method, a process-based approach, which estimates emissions and removals from changes in carbon stocks due to forest land and related land-use changes. Default data are provided in each land-use category chapter to allow the estimation of biomass carbon stock changes in case of missing country-specific data. This is the method recommended only for countries without an NFI and poor data collection.

In the second approach, forest carbon sinks are estimated by coupling estimates of forest age with age-specific carbon sequestration models. These models are derived from yield tables, expressing carbon stocks as a function of age stand.

The third approach, called the stock-change approach, requires bitemporal biomass carbon stock measurements; therefore, its application is suitable in countries having NFI systems and other land-use categories, where stocks of different biomass pools are surveyed with a regular frequency. This method results in considerably less uncertainty (McRoberts et al., 2018; ISPRA, 2021a).

Examples of these approaches are presented by Harris et al. (2021) and Xu et al. (2021), they both integrated spatially explicit datasets and ground-measured forest inventories data to provide global estimates of temporally averaged global forest carbon emissions and removals for the 21st century, founding that woody carbon stocks increased slowly but significantly at a local and regional scale.

Saatchi et al. (2011) presented a benchmark map of biomass carbon content across the world's tropical forests for 2000 by combining ground data with airborne laser scanning (ALS), multispectral, and radar data: their map provided estimations of carbon stocks for countries where prior estimates were scarce or not complete. With the stock-change approach, Paul et al. (2021) assessed the carbon stock and changes in New Zealand using the NFI data from 2002 to 2014, showing that national forests are carbon-neutral but with wide variation in carbon stocks between different forest categories. In Russia, Shepashenko et al. (2021) used multiple RS-based maps and NFI data to estimate CO_2 sequestration founding figures 47% higher than the national GHG inventory.

In Italy, Dalponte and Coomes (2016) developed an approach to map the carbon density of the Italian Alps through ALS and hyperspectral data. Nonini and Fiala (2021) developed a model to assess the forest biomass and carbon stock at stand-level with a gain-loss approach in a northern region in Italy.

Federici et al. (2008) developed a model to estimate carbon stock change data for the carbon pools to be reported under the forest land category in the LULUCF sector in the GHG inventory. The *For-est* (***For***est – ***est***imates) is a bookkeeping model that calculates the above-ground biomass pool C stock annually by adding the annual net increment and subtracting yearly losses associated with harvest (industrial roundwood and fuelwood), forest fires, and other mortality. A detailed description of the modeling approach is reported in Italian Ministry for the EnvironmentLand and Sea, 2019 (section 3.3). The annual GSV is converted to AGB and then to carbon stock by species-specific parameters. The model is currently used by Italy to estimate carbon stock changes for the national GHG inventory under UNFCCC (ISPRA, 2021b; b).

The aim of this study is to present the development of a spatial approach for the wall-to-wall estimation of GSV and carbon stock to fill the information gaps left by the long updating cycle of the periodic Italian NFI, under the framework for a new EFI that better fits the international reporting requirements.

To do so, we propose a new methodology to produce a yearly high-resolution (23 m) forest above-ground carbon pools and GSV maps. Our approach is initiated by a 23 m resolution wall-to-wall GSV map of 2005 constructed by combining Landsat imagery with NFI data (Vangi et al., 2021). We then applied yearly increments with species-specific growth models derived from yield tables driven by the forest GSV to estimate the annual current increment (Federici et al., 2008). We take into account removals due to forest disturbances predicted using Landsat imagery and the 3I3D forest disturbance detection algorithm (Francini et al., 2021; Francini et al., 2022a; a). The approach was tuned against a set of independent field observations, and the final pixel-level estimates were aggregated at the regional level and validated against the design-based estimation from the last Italian NFI completed in 2015, obtaining an RMSE% at a regional level of 19% and 17% for GSV and carbon stock, respectively. Our estimates were also compared with official data reported in the Italian GHG inventory.

To the best of the author's knowledge, this represents the first attempt to provide high-resolution wall-to-wall yearly time-series maps of forest growing stock volume and carbon stock in Italy. These new products allow the spatial analysis of the annual Italian forest carbon stock changes, consistently with the IPCC guidelines.

2. Materials

2.1. Study area

The study was carried out in Italy, covering 301,408 km^2 (Fig. 1). Italy has a wide range of climatic conditions due to its proximity to the sea and the presence of two main mountain belts with elevations ranging between sea level up to 4000 m a.s.l. Italy has mainly a temperate Mediterranean climate (Pinna, 1970). According to the 2015 Italian NFI (INFC, 2021), forest vegetation and other wooded lands occupy 11,054, 458 ha, about 36% of the national land. Deciduous species cover 68% of the forest area and are represented mainly by Quercus oak (*Q. petraea* (Matt.) Liebl., *Q. pubescens* Willd., *Q. robur* L., *Q. cerris* L.), and European beech (*Fagus sylvatica* L.). Coniferous species, such as Norway spruce

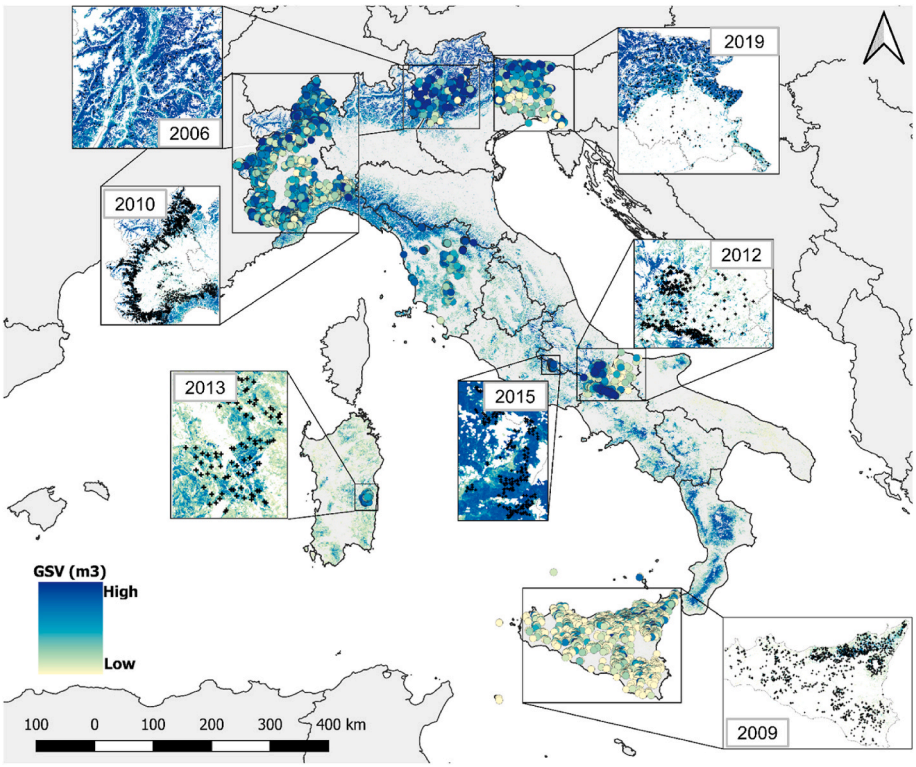


Fig. 1. The study area (the whole of Italy) with the location of the independent plots was used for optimizing and validating the model.

(*Picea abies* (L.) H. Karst.) and pines (*Pinus sylvestris* L., *P. nigra* J.F. Arnold, *P. pinae* L., *P. pinaster* Aiton), form vast plantations, especially in the northern regions and coastal areas (Fig. 1).

Italy is divided into 20 administrative regions (NUTS2); the NFI produces every ten years regional estimates for several variables including forest area, total and average GSV, and biomass with the relative associated standard errors (SE) with a traditional design-based approach. According to NFIs, at the country level, the average GSV was 121 $m^3\ ha^{-1}$ and 135 $m^3\ ha^{-1}$ in 2005 and 2015, respectively.

2.2. Growing stock volume baseline map

For the assessment of forest GSV and above-ground carbon stock in the years following the last NFI, we used the 2005 GSV map produced by Vangi et al. (2021) for Italy as the initial GSV baseline data (GSV_{2005}). This map consists of GSV predictions by Landsat and other RS imagery at 23 × 23 m resolution for all forest pixels. The full description of the methodology is available from Chirici et al. (2020). The model fitting and tuning steps were carried out using the *randomForest* package in the statistical software R 4.0.5 (Liaw and Wiener, 2002) (https://www.r-project.org, accessed on: June 16th, 2021). The pixel-level estimations of the GSV range between 0 and 690 m^3ha^{-1} with a mean value of 134 m^3ha^{-1} and a standard deviation of 41.5 m^3ha^{-1} (for comparison the official NFI estimates range between 0 and 950 with a mean value of 145 and a standard deviation of 69 m^3ha^{-1}). Using a model-assisted estimation approach (Corona, 2010), the 2005 growing stock volume map led to a standard error of 1.2% and 1% for the mean and total national GSV estimation, respectively (Vangi et al., 2021).

2.3. Forest category maps

In Italy, a forest category map with a spatial resolution consistent with the input GSV map used in this study is not yet available. For this reason, the distribution of forest categories was derived from the Corine Land Cover (CLC) maps, which are available for the reference years 2006, 2012, and 2018. In Italy, CLC is the only spatial source that provides consistent information on forest category distribution across different years on a national scale. The CLC project was started in 1990 by the European Environmental Agency (Büttner et al., 2004) and consists of a European-scale land-use monitoring program with a 44-class nomenclature system produced by photointerpretation of high-resolution satellite imagery. CLC uses a minimum mapping unit (MMU) of 25 ha and a minimum mapping width (MMW) of 100 m (EEA, 2007). The original CLC nomenclature system classifies the forest into three classes: broadleaves, coniferous, and mixed forests. In the Italian implementation the CLC maps produced by the *Istituto Superiore per la Protezione e la Ricerca Ambientale* (ISPRA) classify forests into 28 classes (Bologna et al., 2004). In this study, we re-classified forests into 18 classes (Annex I). Forest category maps were obtained from the original CLC vector products by rasterizing at the same spatial resolution as the baseline GSV map. Then we masked out the non-forest categories by assigning them to the "non-forest" class.

Our spatial approach requires for each year a newly updated forest category map. Since the CLC project is not updated yearly, we used the forest category map of 2006 for the years 2005–2009, that of 2012 for 2010–2014, and that of 2018 for 2015–2018 (Fig. 2). This procedure was considered appropriate since the percentual change of forest area

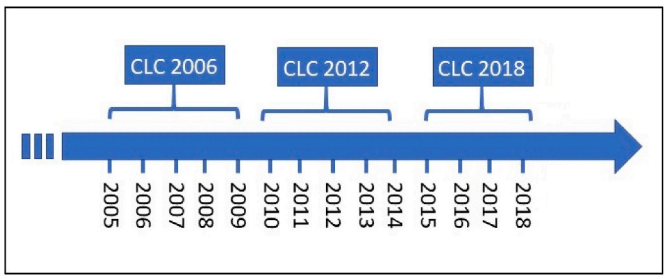


Fig. 2. Corine Land Cover (CLC) forest category maps used for each year of the study period.

based on the CLC maps is limited. In the period 2006–2012, considering both increments and decrements, the forest area changed only by 741 km^2 (the 0.74% of the forest area) while in the period 2012–2018, it changed only by 705 km^2 (the 0.70% of the forest area).

2.4. National collection of yield tables

The national collection of yield tables from Federici et al. (2001) was used to model the current increment of forests as a function of GSV and the GSV as a function of forest age. Yield tables reported the GSV and current increment as a function of forest age for 27 species within 13 genera. The 27 species were linked to the 18 forest categories derived from CLC (see section 2.3), with *ad-hoc* harmonization bridges developed for this study. Bridges preserve data attributes based on different definitions, allowing their comparison at a higher hierarchical level (Annex I).

2.5. Forest disturbances time-series maps

Data on the spatial distribution of forest disturbances in the period 2004–2018 were needed to account for forest harvesting and other disturbances in GSV and the carbon stock estimation process. These data were produced with the 3I3D algorithm (Francini et al., 2021) recently implemented in Google Earth Engine (GEE) (Francini et al., 2022b), a cloud-based platform that can process massive amounts of remotely sensed data (Gorelick et al., 2017).

3I3D is an unsupervised algorithm that predicts forest disturbances requiring no input parameters or calibration. It analyses the pattern over three sequential years of three indexes (3I) of photosynthetic activity used as three-dimensional (3D) space axes. 3I3D was applied using yearly cloud-free composites of Landsat surface reflectance images atmospherically corrected with LEDAPS (Wolfe et al., 2004) and acquired with a solar zenith angle smaller than 76°. Candidate images were acquired during the vegetative season (between Jun and Aug), with a cloud cover lower than 50% in the scene. As a result, we obtained a collection of about 800 images per year. We excluded those pixels covered by clouds, shadows, water, and snow (Foga et al., 2017) and pixels with an opacity value greater than 0.3. For each year, we then selected the "best" pixels among the remaining ones using the Best Available Pixel procedure (BAP) (Griffiths et al., 2013; White et al., 2016), obtaining a BAP-collection of cloud-free composite for each year between 2004 and 2018. Specifically, the BAP pixel selection is based on a set of scores, among which i) the sensor, ii) the day of the year, iii) the distance to cloud or cloud shadows, and iv) the opacity. The BAP was recently implemented in GEE, with the full code openly available. A detailed description of the application, guidance, and suggestions on BAP parameters setting is provided on GitHub (https://code.earthengine.google.com/?accept_repo=users/sfrancini/bap).

We used BAP cloud-free composites as input for the 3I3D algorithm to predict forest disturbances with a MMU of 500 m^2 over the study period. Official forest data on burned areas, annually produced and released for the same period by the Italian Forest Service (Comando Unità Forestali, Ambientali e Agroalimentari of Carabinieri), have been also used. This dataset includes burnt areas from forest fires acquired through a ground survey with the Global Navigation Satellite System (GNSS).

We merged the official national database of forest fires with the forest disturbances map produced by 3I3D (with an OR logical operator), classifying forest pixels for each investigated year in "disturbed" or "undisturbed". Based on these maps, we finally produced the "age" of disturbed forests for each investigated year as the number of years since the last disturbance event (YSLD).

2.6. Calibration data

To optimize and calibrate the procedure, we used 9258 circular plots where the GSV was measured in the field between 2006 and 2019 in the framework of local forest inventories (Fig. 1). The plots are distributed under different environmental conditions and forest categories over the whole country. The same survey protocol of the Italian NFI was adopted in all these plots. The tree-level GSV was determined by the allometric models used for the Italian NFI (Tabacchi et al., 2011), and then tree-level data were aggregated at the plot level. For this study, allometric model prediction and GNSS position uncertainties are expected to be negligible for the spatial resolution adopted (McRoberts et al., 2013, 2016, 2018; Chirici et al., 2020). The mean GSV in this calibration dataset is 216 $m^3\ ha^{-1}$, with a maximum of 1482.4 and a standard deviation of 155 $m^3\ ha^{-1}$. To find the most appropriate solution, we evaluated the models in terms of RMSE% at the plot level, comparing GSV estimates with the observed one.

2.7. Validation data

To validate the results, we compared aggregated regional GSV and carbon stock estimates for 2015 based on the pixel level values we produced (the GSV_{2015} 23 m resolution map) with official regional estimates from the Italian NFI. We also compared aggregated values of GSV and carbon stock produced by our method with the official estimates reported in the Italian 2006–2019 GHG inventory. Just as in the calibration, we compared the accuracy of our results in terms of RMSE%, calculated as the percent of RMSE against the mean official values.

3. Methods

3.1. Overview of the spatial approach

For the years not covered by the periodic Italian NFI, our spatial approach for predicting GSV and carbon stocks was carried out differently for pixels belonging to disturbed and undisturbed forests. Annual stock changes were predicted for each 23 m pixel, using the GSV2005 mapped for 2005 and the YSLD as unique drivers. The GSV_{2005} was selected since it is strictly related to the above-ground biomass and carbon stock, and it is directly measured by the NFI in the field. While the YSLD was selected because it can be easily obtained based on change detection algorithms, such as the 3I3D, and it is the primary driver of forest variables in yield tables.

Our approach uses the first derivative of the Richards function (Eq. (1)) to calculate the current increment (Eq. (2)) as a function of the GSV in undisturbed forests for each of the 18 CLC forest categories (Federici et al., 2008). The following equation defines the Richards function:

$$\frac{dy}{dt} = \frac{k}{v} \bullet y\left[1 - \left(\frac{y}{a}\right)^{v}\right] + y_0 \text{ first derivative} \quad (1)$$

Its analytical solution defines the Richards growth curve:

$$y = a \bullet \left[1 - e^{(\beta - kt)}\right]^{-\frac{1}{v}} \quad (2)$$

where the general constraints for the parameters are a, $k > 0$; $-1 \leq \beta \leq \infty$; $v \neq 0$.

The curve is bounded and monotonic, highly flexible thanks to its four parameters. It can be efficiently approximated to a logistic ($a \to \infty$, $v > 0$), exponential ($v > 1$), or other most used growth curves. However, due to the number of parameters and their high covariance, the curve is difficult to fit and can cause problems during the non-linear regression (Federici et al., 2008). The current increment represents the dependent variable, while the independent variable is the GSV map.

In disturbed forest areas, different potential models were evaluated for estimating GSV as a function of YSLD for each forest category. Using the data in the yield tables collection, we tested four regression models, two non-parametric, random forests and Support Vector Machine (SVM), and two parametric, polynomial, and linear regression. The GSV represents the dependent variable, while the independent variable is the

forest age from yield tables.

The optimization was fine-tuned by picking the most accurate model based on the correlation coefficient (r^2). All four regression models yielded comparable results with only slight differences. The SVM model slightly outperformed other approaches, with an average r^2 among the forest categories of 0.91, against 0.90 of random forests and 0.88 of polynomial and linear regression.

After the model fitting for each category in undisturbed and disturbed forest areas, the above-ground carbon stock is predicted in the five steps described below:

1. Starting from the initial GSV_{2005} map, the current increment is estimated via the Richards function in undisturbed forest areas for each year and for each forest category, based on the relationships between GSV and current increment derived from yield tables.
2. Similar to point 1, GSV in disturbed forest areas is estimated for each year, and forest category with the corresponding SVM model, based on the relationships between YSLD and GSV derived from the yield tables.
3. For each year and forest category, the GSV is calculated as the sum of the previous year's GSV and the estimated current increment, subtracting the losses due to natural mortality and adding the GSV in disturbed forest areas calculated in step 2.
4. For each year and forest category, the GSV ($m^3\ ha^{-1}$) is converted in above-ground biomass (AGB) (Mg d.m. ha^{-1}) with the equation:

$$AGB = GSV^{*}BEF^{*}WBD \quad (3)$$

where GSV is the growing stock volume calculated in step 3, BEF is the category-specific biomass expansion factor (dimensionless), and WBD is the wood basal density (Mg d.m. m^{-3}).

5. Carbon stocks are derived from AGB by applying the default carbon fraction factor of 0.47 (IPCC et al., 2006).

Following the IPCC Good Practice Guidance for LULUCF (IPCC et al., 2006), the average rate of natural mortality was set equal to 0.116% for evergreen categories, 0.117% for deciduous categories, and 0.1165% for mixed categories, while BEF and WBD are those applied by the *For-est* model (Federici et al., 2008).

The detailed methodology is described here below.

3.2. GSV estimation in undisturbed forests

In undisturbed forests, the GSV for the year *n* was computed at pixel level by adding the current increment of the year *n-1* to the GSV of the year *n-1* and subtracting losses due to natural mortality. We used an age-independent model to create annual maps of the current increment. Category-specific growth models were constructed using the data of the national yield tables collection to calculate the current increment ($m^3\ ha^{-1}\ y^{-1}$) as a function of GSV ($m^3\ ha^{-1}$), using the first derivative of the Richards function (eq. (1)). The GSV represents the independent variable *x*, while the dependent variable *y* is the correspondent current increment. The forest category-specific Richards functions were fitted using all the fertility classes of the yield tables. The parameterization was based on a 25 iterations-bootstrap-cross-validation procedure. For each bootstrap iteration and species, the RMSE was calculated, and the model which reported the lowest RMSE was chosen as the final model,. RMSE was calculated as:

$$RMSE_{sp} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n}} \quad (4)$$

where *n* is the number of observations in the yield tables for the species *sp*, y_i is the current increment value reported in the yields table for the i-th observation and $\widehat{y_i}$ is the current increment predicted from the model for the i-th observation.

We started the process based on the GSV_{2005} map produced by Vangi et al. (2021), and we produced the updated GSV_{2006} map applying, for each 23 × 23 m undisturbed forest pixel, the current increment per hectare predicted for each forest category with the corresponding growth curves and subtracting the natural mortality. Then, the process was repeated for 2007 based on GSV_{2006} and so on until 2018.

By applying the Richard first derivative approach, the current increment was estimated with an average RMSE (as per eq. (4)) of 49.4% across all forest categories, with significant variations among forest categories, mainly due to the number of observations and fertility classes available in the yield tables. Some of the most frequent forest categories in Italy obtained the best results, such as the maple-ash-hornbeam mixed forests (RMSE = 22%), the Mediterranean maquis (RMSE = 13%), and the chestnut forests (RMSE = 26%), which altogether represent more than 25% of the national forest area.

The best results were obtained by the exotic plantations category with an RMSE of 3.2%, most probably because of their homogeneous growth behaviour. In comparison, the mixed conifers category obtained the worst result with an RMSE of 95% (but they cover only 4.4% of the forest area), most probably because of their heterogeneous composition.

3.3. GSV estimation in disturbed forests

We already know that forest age is not an appropriate predictor for estimating the productivity of undisturbed forests in Italy since they are mainly uneven-aged and are characterized by a complex mosaic of different ages or cohorts (Federici et al., 2008; Frate et al., 2015). Instead, in most disturbed forests, trees regrowing after the disturbance results in even-aged stands, at least for the first years after the disturbance. This is particularly true for clearcuts in coppice forest (Chirici et al., 2020), which represent the most common forest disturbance in Italy, based on Francini et al. (2022, a), representing 80% of all forest loggings in Italy. In such a situation, forest age can be used to predict GSV growth in disturbed stands using the data in the national yields table collection. We used forest categories-specific SVM models to predict the GSV based on forest age with a radial basis kernel function. SVM approaches are known to be robust against outliers and overfitting and are well-suited for approaching problems with a limited amount of training data. These algorithms can generate non-linear decision surfaces by mapping the data into a high-dimensional space through non-linear mapping functions called kernel functions (Cortes and Vapnik, 1995; Pal and Mather, 2005), allowing the separation of the data through linear hyperplanes (Dixon and Candade, 2008). Among the kernel functions, one of the most used is the radial basis function, which has two tuning parameters *C* (regularization parameter) and γ (kernel width) (Kavzoglu and Colkesen, 2009). An in-depth explanation of SVM-based models and kernels is presented in Smola and Schölkopf (2004) and Kavzoglu and Colkesen (2009). Implementations of SVM models in RS can be found in Mountrakis et al. (2011). In this study, the parameters of SVM and radial kernel (C, γ) were determined by bootstrap cross-validation with 25 iterations using the grid search method, by selecting the pairs of parameters that produce the lowest cross-validation RMSE among an exponentially growing sequence of the parameters ($C = 2^{1/2}, 2^1, \ldots, 2^5$; $\gamma = 2^{-5}, 2^{-4}, \ldots, 2^{\circ}$). For each bootstrap iteration and species, the RMSE was calculated as per eq. (4), and the model which obtained the lowest RMSE was chosen as the final model.

In disturbed areas identified by the 3I3D algorithm, the GSV was computed for each year and forest category by applying the category-specific SVM models fitted from the yield tables data. The YSLD for the year n was used as the independent variable to predict the GSV_n in each disturbed pixel, obtaining a GSV map of forest disturbances for each year between 2005 and 2018.

The complete GSV_n map was produced by overlaying the GSV_n maps of disturbed and undisturbed forests.

The SVM models led to an average RMSE of 35.9%, with a maximum

of 64% for the mixed forests with the prevalence of coniferous and a minimum of 5% for the maple-ash-hornbeam mixed forest. As for the Richard models, we observed significant variations depending on the number of fertility classes in yield tables.

3.4. Carbon stock conversion

Once estimated the GSV, amounts of AGB are consequently assessed. For every forest typology, starting from the GSV, the AGB (Mg d.m. ha^{-1}) is calculated, through equation (3), following the approach presented in Federici et al. (2008).

Carbon stock maps were derived from AGB maps by applying the default factor for carbon fractions of 0.47 (IPCC et al., 2006).

The pixel-level predictions of GSV and stocked carbon were aggregated at the regional level for each year.

4. Results

4.1. GSV and carbon stock estimation

The spatial approach for estimating annual GSV and above-ground carbon pool was applied to produce 23 m resolution yearly pixel-level estimates from 2005 to 2018. Based on our results, the GSV increased in Italy by 522 million m^3 moving from an average of 130 m^3 ha^{-1} to 180 m^3 ha^{-1}. GSV and above-ground carbon stocks time series are reported in Annex II. Carbon stock increased by 206 million of t in the same period, moving from 36.9 Mg C ha^{-1} to 59.3 Mg C ha^{-1}, with an average accumulation rate of 14.7 mln Mg C y^{-1}. Regionally, most of the GSV and carbon gains dominate mountain landscapes of the Alps and Apennines mountains. In the years 2005–2018, among all forest categories, beech forests accumulated the most GSV, with about 3926 mln of m^3 corresponding to 54 mln Mg C of above-ground carbon stored (about 28.3% of the total carbon absorbed by national forests), followed by mixed broadleaf forests (34 mln Mg C, about 18% of the total) and the fir/spruce forests (23 mln Mg C, 12% of the total). Regardless of the forest category, in the study period, carbon accumulation is reflected mainly in the increase of the carbon density rather than the increase of the total forest area, which amounts to 145,000 ha according to the CLC maps.

Northern regions (Trentino-Alto Adige, Piemonte, Lombardia, Veneto, Friuli Venezia Giulia) have the highest GSV accumulation in terms of absolute and per hectare figures, accounting for 54% of the national total. Other regions with significant GSV accumulation are Toscana, Sardegna, and Emilia-Romagna, contributing 20% to the national GSV growth (each up to 20 mln m^3 in the study period). In contrast, most southern regions (Molise, Campania, Puglia, Basilicata, Sicilia) show the least accumulation of GSV, less than 8 mln m^3 between 2005 and 2018. Carbon uptake has similar patterns, exhibiting higher absolute and per hectare storage in many northern regions (Trentino-Alto Adige, Piemonte, Lombardia, Veneto) and lower in the southern ones (Molise, Puglia, Umbria) (Fig. 4). Also, at the regional level, the accumulation of GSV and carbon stock is primarily driven by the increase of GSV and carbon density rather than the total forest area. This is probably due to a decrease in the harvested area over the last two decades, which allowed for significant growth in GSV per unit area (Francini et al., 2022a; b).

In Fig. 3 is reported the overall absolute accumulation of GSV and carbon stock at the regional level over the study period.

4.2. Validation and comparison of our results

During the optimization phase the 23 m resolution pixel level estimations of GSV estimates obtained applying the best configuration of our models where compared against the GSV measured in the field in 9258 independent plots acquired in different years in the period 2006–2018. From such comparison the average RMSE% was 57% ranging between 89.6% in 2009 and 34% in 2015, (Fig. 5). The bias across years was −3.7 m^3 ha^{-1}, with the minimum in 2010 (−0.2 m^3 ha^{-1}) and the maximum in 2013 and 2015 (−70.3 and −60.2 m^3 ha^{-1}). These values are in the range of previous experiences (Immitzer et al., 2016; Chirici et al., 2020; Vangi et al., 2021).

Pixel level estimations for the year 2015 where aggregated for administrative Regions and compared with the official 2015 NFI estimates (INFC, 2021), resulting in a 6.2% and 1.1% difference at the national level for GSV and carbon stock, respectively (calculated as the mean value of the difference between predicted and observed results). We obtained an RMSE of 19.5% and 17.8% at the regional level and an r^2 of 0.94 and 0.92 for GSV and carbon stock, respectively (Fig. 6). RMSE was calculated as:

$$RMSE_{NFI} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{NFI\,i} - \widehat{y_i}\right)^2}{n_{rg}}} \quad (5)$$

where n_{rg} is the number of Italian regions, y_{NFIi} is the official NFI value (of GSV and carbon stock) for the i-th region and $\widehat{y_i}$ is the aggregated estimation (of GSV and carbon stock) produced by the spatial approach for the i-th region.

The data for the comparison against the 3rd Italian NFI (INFC, 2021) are presented in Annex III.

Finally, following the same procedure, our GSV predictions aggregated for Italian Regions were compared with official estimates of Italian GHG inventories for 2005–2018, obtaining an overall RMSE% of 28.6% and an r^2 of 0.77 with a growing trend over time. Here the RMSE was calculated as:

$$RMSE_{GHG} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{GHG\,i} - \widehat{y_i}\right)^2}{n_{rg}}} \quad (6)$$

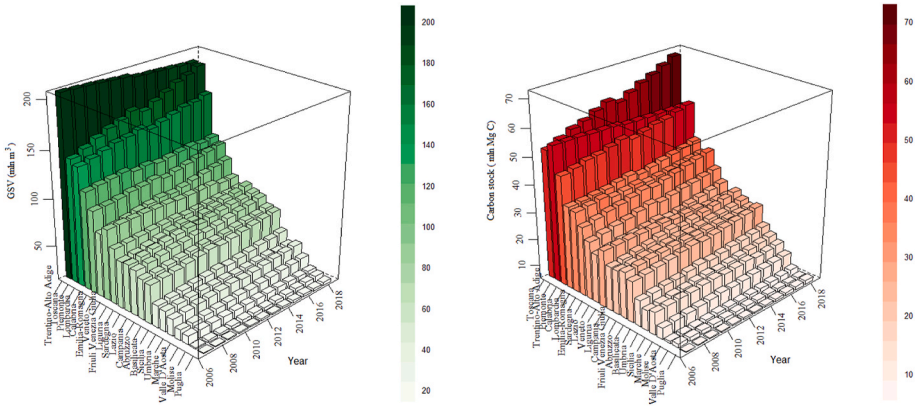


Fig. 3. Annual GSV (left) and carbon stock (right) at the regional level from 2005 to 2018.

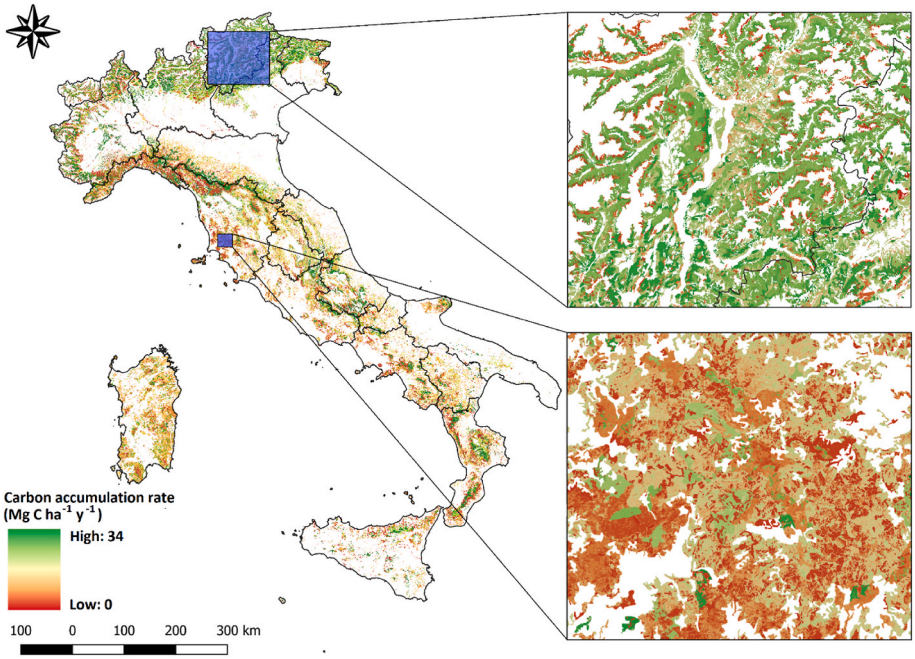


Fig. 4. Left: Pixel-level carbon accumulation rate in forests (Mg C ha^{-1} y^{-1}) for 2005–2018.

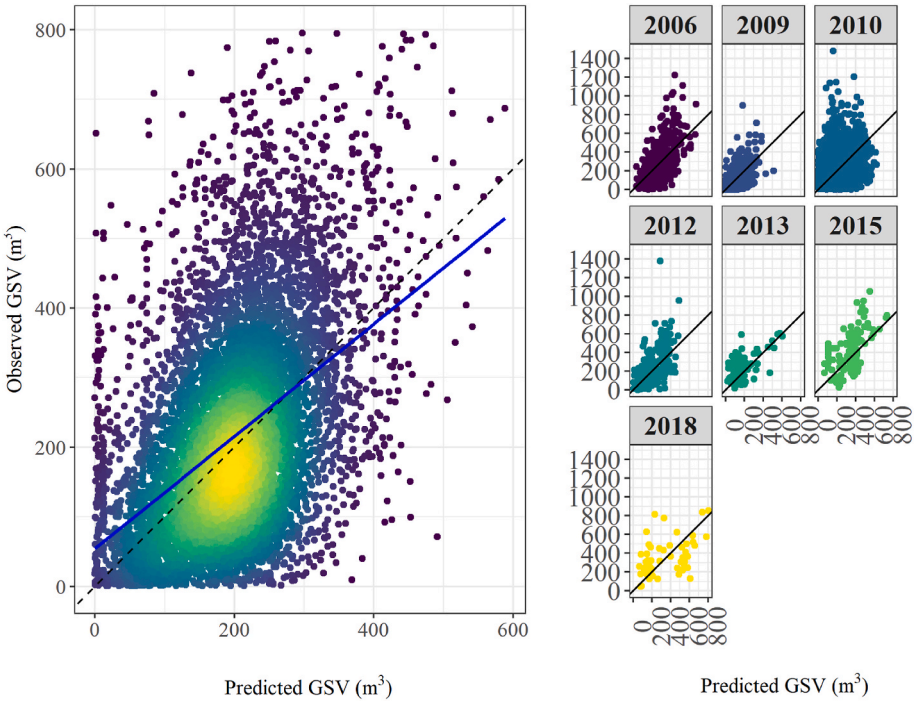


Fig. 5. Left: observed GSV in the field plot against predicted GSV; Right: observed against predicted GSV for each year. Blue is the regression line.

where n_{rg} is the number of Italian regions, y_{GHG_i} is the official GHG inventoriy value (of GSV and carbon stock) for the i-th region and $\hat{y}_i$ is the aggregated estimation (of GSV and carbon stock) produced by the spatial approach for the i-th region.

The carbon stock was also compared against the official Italian GHG inventory (LULUCF sector, forest land remaining forest land category) for the same period (ISPRA, 2021a; b), yielding an r^2 of 0.88 and an overall RMSE% (as per equation (6)) of 23.1% and 17.2% among years and regions, respectively. As for the GSV, consistency with official estimates has worsened over the years, while at the regional level reached the minimum in Piemonte (RMSE 2.5%) and the maximum in Trentino-Alto Adige (RMSE 48.6%). Thirteen out of 20 regions showed an RMSE% less than 15%, with seven regions less than 10%. Fig. 7 reports the regional comparison between our results and the official estimates from the national GHG inventory (ISPRA, 2021b; b) regarding GSV and carbon stock.

5. Discussion

The main objective of the study was to develop a new spatial

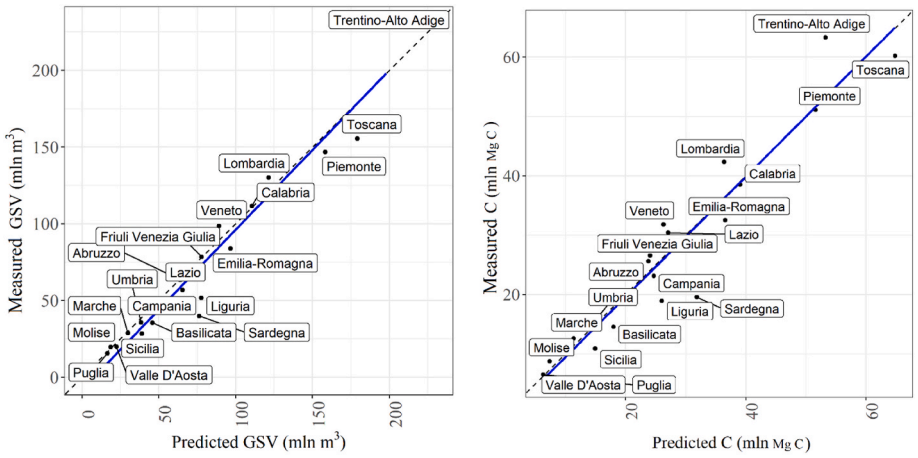


Fig. 6. Left: INFC 2015 GSV against predicted GSV; Right: INFC 2015 carbon stock against predicted carbon stock. The dotted line is the y = x line, and the blue is the regression line.

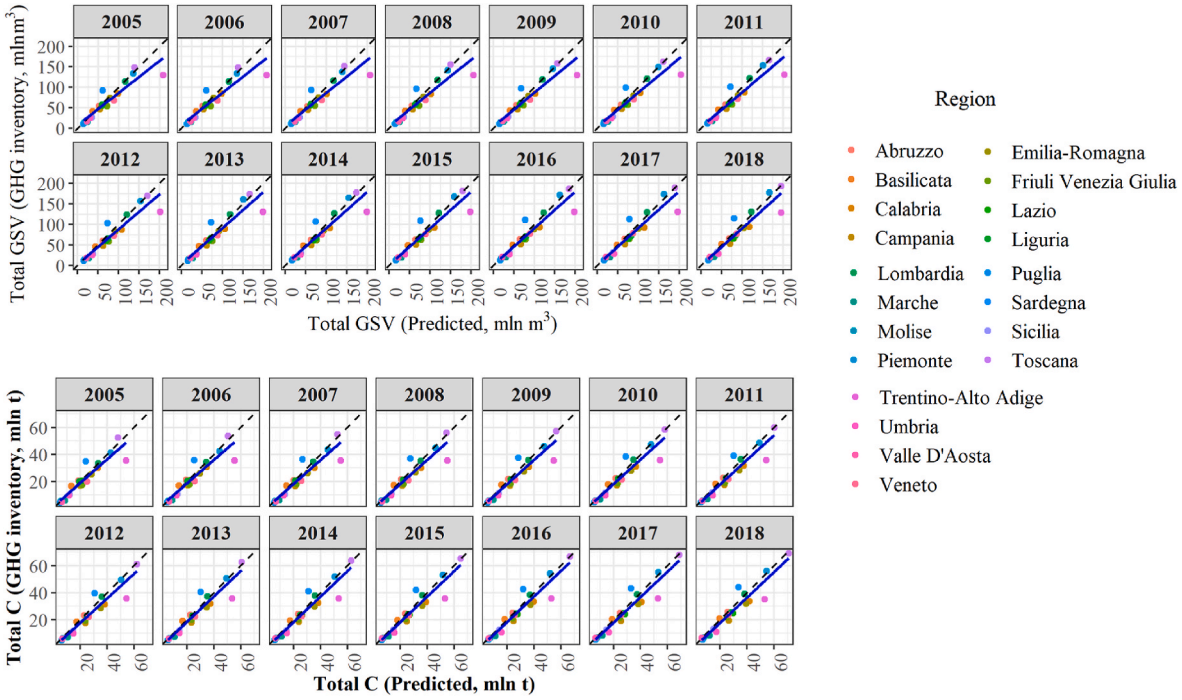


Fig. 7. Comparison of GSV (right) and carbon stock (left) at the regional level against the Italian GHG inventory LULUCF sector, forest land remaining forest land category) for each year of the study period.

approach for producing wall-to-wall high-resolution yearly GSV and carbon stock predictions between consecutive NFI field measurements, exploiting remotely sensed and auxiliary data, which could be used for operational application and to respond to international reporting tasks. Empirical models developed by interpolating data from NFI field plots could produce updated estimates only for short-term predictions, for which growth conditions such as climate and management regimes are expected to be stable (Peng, 2000). Moreover, in such models, stand variables are driven by the age of the forest, but in natural conditions, growth is strictly related to species and local environmental conditions. For these reasons, our novel approach for estimating carbon stocks and changes at the national and regional level in the above-ground carbon pool is driven only by NFI GSV data and yield models.

The yearly wall-to-wall maps of GSV and carbon stock can support reporting activities and forest management at any scale by aggregating pixel-level predictions producing small-area estimations, for example, using the estimators proposed by Chirici et al. (2020).

Our results agreed with those reported by the official Italian GHG inventory and show an increasing trend in the above-ground carbon pool that reflects both the expansion of forest areas (according to NFI, in the period 2005–2015, forest areas increased by 58,692 ha y^{-1}, approximatively 0.5% of total forest area in 2015) and the increased growing stock resulting from a forest harvest rate lower than the current increment. The organic carbon in the above-ground biomass of the Italian forest exceeded 566 million Mg C in 2018, with a different contribution of regions and forest categories in terms of fixed organic carbon. GSV and carbon stock distribution among forest categories comply with the 2020 Italian FAO FRA report for 2005 and 2010, with beech and spruce/fir forests accounting for 40% and 36% of the total GSV and carbon stock, respectively. Due to their limited area nationwide, the forest categories contributing the least to carbon storage (less than 0.5% of the total) are exotic conifers, broadleaf plantations, wood

arboriculture, and riparian formations. It is worth noting that other wooded vegetation in open and shrublands has a relatively significant contribution to terrestrial carbon sinks, storing more than seven mln of Mg C of carbon (approximately 5% of the national total).

Large forest areas characterize regions with the most GSV and carbon stock and highly carbon-storing forest categories (beech, mixed broadleaf, fir/spruce, larch, Mediterranean pines, exotic plantations), despite most of these categories are subject to intense forest harvesting, as resulted from 3I3D disturbances maps. In this regard, it is worth noting that the major disagreement between our approach and the official ISPRA estimates is found in the northern regions and particularly in Trentino-Alto Adige, which leads to the maximum relative RMSE (for which the GHG inventories estimate a decrease in GSV and carbon stock over the study period). This mismatch is primarily due to underestimating the number of forest disturbances, especially in high-forest stands of the main mountain ranges, where silvicultural treatments are based on continuous canopy cover approaches that are difficult to detect by optical satellite imagery. The lack of fire data compounds the underestimation of disturbances before 2007 and after 2017.

Moreover, for the Autonomous Provinces of Trento and Bolzano, the database of forest fires includes only fires greater than 20 ha, potentially increasing the underestimation of the total number of disturbances. Without offsetting for forest harvests and forest fires, the GSV builds up rapidly, driven by the increase of the current increment of highly productive forest categories, leading to significant overestimates and large values of RMSE. However, in Trentino-Alto Adige, the mismatch between our results and the 2015 NFI is less evident, with an underestimation of only 13% and 15% for GSV and carbon stock, respectively.

Another source of uncertainty is the overestimation of the current increment with the Richard first derivative compared to the 2nd NFI field measurement. This finding contrasts with the underestimation reported by Federici et al. (2008) compared to the 1st NFI data. The discrepancy between the predicted and reported current increment of NFI is likely driven by three factors: i) an outdated collection of national yield tables which no longer reflect the country's real average quality of forest sites; ii) Richard's first derivatives were fitted against all species quality class data in yield tables, and since species data were aggregated to match the species composition of each forest category as closely as possible, the predicted current increment of less productive species can be overestimated, especially for mixed forest categories; iii) the overestimation of the GSV in the initial 2005 GSV map, used as the independent variable for Richards functions. This evidence was reported by Vangi et al. (2021). SVM models generally performed better than the Richards function for the same forest categories, except for chestnut, hygrophilous formations, exotic broadleaf plantations, larch, and stone pine categories. For those forest categories, the SVM model underestimates GSV for older age classes, for instance, over ten years for exotic broadleaf plantations and riparian formations, 25 years for chestnut, and more than 80 years for the remaining forest categories. Hence, the effect of underestimation only affects categories of hygrophilous formations and non-native species plantations, as the years since the last disturbance can be at most equal to 13 years (2018–2005).

During the optimization phase using the dataset of independent field plots the differences between pixel-level measured and estimated GSV increased over time. This bias with under-predictions for plots with high values of GSV, especially for years 2009, 2012, and 2015, can be caused by the well-known saturation effect, especially in dense crown cover and steep terrain (Nilsson et al., 2017; Giannetti et al., 2018; Chirici et al., 2020). Such dataset was acquired for those years in forests managed for productive purposes or, more in general, with an average GSV much greater (420 m^3 ha^{-1}) than those measured in INFC plots (140 m^3 ha^{-1}). In this phase the largest bias was observed in southern regions and in the islands, that are characterized by sparse Mediterranean vegetation where modeling the GSV resulted more difficult in previous experiences too (D'Amico et al., 2021; Vangi et al., 2021). But it is important to note that wall-to-wall spatial predictions from NFI field observations should never be used at the pixel level since the single pixel predictions may be affected by a consistent bias (McRoberts and Tomppo, 2007).

However, when we aggregated the pixel level estimations at the regional level, we obtained very satisfying results consistent both with the Italian GHG inventory and the 3rd NFI (INFC, 2021). It is worth noting that, concerning the 3rd NFI in 2015, the carbon stock difference at the regional level was minor than the GSV difference (Fig. 7). This is interesting because it proves that the species-specific BEF and WBD together with the spatial distribution of forest categories can compensate for the overestimation of GSV. Under this point of view, our estimates are conservative since the approach neither overestimates increases nor underestimates decreases in carbon stocks with respect to the NFI official estimations.

As soon as new high-resolution forest types maps will be available, the use of the CLC maps should be reconsidered. Several studies have already highlighted the limitations of CLC maps in forestry, primarily because of the wide MMU of 25 ha (Seebach et al., 2012; Vizzarri et al., 2015; Vangi et al., 2021).

6. Conclusions

To the best of our knowledge, this is the first study to provide yearly high-resolution GSV, AGB, and carbon stock wall-to-wall time-series maps for the whole national territory in Italy, allowing an in-depth analysis of the forest carbon stock changes, consistently with the IPCC guidelines. The spatial nature of our results enables small-scale estimates by aggregating individual pixel predictions, enhancing the spatial resolution of traditional NFI design-based estimates (Chirici et al., 2020), and can be embedded into decision support systems to support sustainable forest management and precision forestry activities. Furthermore, the knowledge of the spatial distribution of carbon among forest categories can be of fundamental importance under the climate mitigation goals of the Paris Agreement (UNFCCC, 2015).

The growing need for new information and technological advances is driving the rapid evolution of forest monitoring and assessment. However, the exploitation of the latter and their implementation within international reporting processes should be evidence-based (Corona, 2018).

This study provides an innovative spatial framework to track GSV and carbon stock changes between NFI surveys at local to national scales, providing a reliable monitoring approach to meet the increasing interests of non-government and private entities in carbon offset investments. Our new method incorporates forest disturbance between surveys, such as forest fires and harvesting, thanks to Landsat-based time-series metrics exploited by the 3I3D unsupervised change detection algorithm, which has already proved to be a better solution in the Mediterranean environment than previous algorithms (Francini et al., 2021). The 3I3D algorithm here demonstrated to be a valid solution for deriving forest age to track the GSV regrowth after disturbances by applying age-dependent relationships in yield tables.

The approach could be improved with 2015 NFI data, as well as updated information on allometric models and yield tables, allowing better model calibration and quality assurance routines. It is worth remembering that although GHG inventories are not measured on the ground, they represent official data sources used for national and international reporting activities.

Nevertheless, the availability of ground-based carbon content data to calibrate and validate the method is desirable. For this reason, as soon as they will be available 2015 NFI carbon stock ground data will allow the evaluation of the pixel-level performance for the carbon stock map for 2015, providing insight into the method's effectiveness. The availability of an official high-resolution national forest map and wall-to-wall multitemporal ALS data could also be fundamental for improving the quality of GSV and carbon stock spatial estimations.

Author's contributions

Conceptualization, G.C., E.V., and G.D.; methodology, E.V., G.D., S·F.; software, E.V.; validation, E.V.; formal analysis, E.V., G.D.; investigation, E.V., G.D.; data curation, E.V., G.D.; writing—original draft preparation, E.V. All authors have contributed to the drafting of the manuscript. All the authors assisted in the quality control, and revisions of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This study was partially supported by the following projects:

MULTIFOR "Multi-scale observations to predict Forest response to pollution and climate change" PRIN 2020 Research Project of National Relevance funded by the Italian Ministry of University and Research (prot. 2020E52THS);

SUPERB "Systemic solutions for upscaling of urgent ecosystem restoration for forest related biodiversity and ecosystem services" H2020 project funded by the European Commission, number 101036849 call LC-GD-7-1-2020;

EFINET "European Forest Information Network" funded by the European Forest Institute, Network Fund G-01-2021.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105580.

References

Bologna, S., Chirici, G., Corona, P., Marchetti, M., Pugliese, A., Munafò, M., 2004. Sviluppo e implementazione del IV livello Corine Land Cover per i territori boscati e ambienti semi-naturali in Italia. Paper presented at the annual meeting for the Society of ASITA, pp. 467–472.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., Soukup, T., 2004. The Corine Land Cover 2000 Project. EARSeL eProceedings 3, 3/2004 331. Available online. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.618.9940&rep=rep1&type=pdf.

Chirici, G., Giannetti, F., McRoberts, R.E., Travaglini, D., Pecchi, M., Maselli, F., Chiesi, M., Corona, P., 2020. Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data. Int. J. Appl. Earth Obs. Geoinf. 84, 101959 https://doi.org/10.1016/J.JAG.2019.101959.

Condés, S., McRoberts, R.E., 2017. Updating national forest inventory estimates of growing stock volume using hybrid inference. For. Ecol. Manag. 400, 48–57. https://doi.org/10.1016/j.foreco.2017.04.046.

Corona, P., 2010. Integration of forest mapping and inventory to support forest management. iFor. Biogeosci. For. 3, 59–64. https://doi.org/10.3832/ifor0531-003.

Corona, P., 2018. Communicating facts, findings and thinking to support evidence-based strategies and decisions. Annals of Silvicultural Research 42, 1–2. https://doi.org/10.12899/ASR-1617.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7 (10), 1236–1245. https://doi.org/10.1111/2041-210X.12575.

Di Cosmo, L., Gasparini, P., Tabacchi, G., 2016. A national-scale, stand-level model to predict total above-ground tree biomass from growing stock volume. For. Ecol. Manag. 361, 269–276. https://doi.org/10.1016/j.foreco.2015.11.008, 0378–1127.

Dixon, B., Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? Int. J. Rem. Sens. 29 (4), 1185–1206. https://doi.org/10.1080/01431160701294661.

D'Amico, G., Vangi, E., Francini, S., Giannetti, F., Nicolaci, A., Travaglini, D., Massai, L., Giambastiani, Y., Terranova, C., Chirici, G., 2021. Are we ready for a National Forest Information System? State of the art of forest maps and airborne laser scanning data availability in Italy. iForest 14, 144–154. https://doi.org/10.3832/ifor3648-014.

EEA, 2007. European Enviromental Agency, environmental statement 2007. Office for Official Publications of the European Communities, Luxembourg, pp. 1–14. ISBN 978-92-9167-936-2.

Federici, S., Quaratino, R., Papale, D., Tulipano, S., Valentini, R., 2001. Sistema Informatico Delle Tavole Alsometriche d'Italia, DiSAFRi – Università Degli Studi Della Tuscia [online] URL: http://gaia.agraria.unitus.it.

Federici, S., Vitullo, M., Tulipano, S., De Lauretis, R., Seufert, G., 2008. An approach to estimate carbon stocks changes in forest carbon pools under the UNFCCC: the Italian case. iForest 1, 86–95 [online: 2008-05-19] URL: http://www.sisef.it/iforest/.

Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Hughes, M.J., Laue, B., 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. Rem. Sens. Environ. 194, 379–390. https://doi.org/10.1016/j.rse.2017.03.026.

Francini, S., McRoberts, Ronald E., Giannetti, F., Marchetti, M., Mugnozza, G.S., Chirici, G., 2021. The Three Indices Three Dimensions (3I3D) algorithm: a new method for forest disturbance mapping and area estimation based on optical remotely sensed imagery. Int. J. Rem. Sens. 42 (12), 4693–4711. https://doi.org/10.1080/01431161.2021.1899334.

Francini, S., McRoberts, R.E., D'Amico, G., Coops, N.C., Hermosilla, T., White, J.C., Wulder, M.A., et al., 2022a. An open science and open data approach for the statistically robust estimation of forest disturbance areas. Int. J. Appl. Earth Obs. Geoinf. 106, 102663 https://doi.org/10.1016/j.jag.2021.102663.

Francini, S., D'Amico, G., Vangi, E., Borghi, C., Chirici, G., 2022b. Integrating GEDI and Landsat: spaceborne lidar and four decades of optical imagery for the analysis of forest disturbances and biomass changes in Italy, 2022 Sensors 22. https://doi.org/10.3390/s22052015, 2015.

Frate, L., Carranza, M.L., Garfì, V., Di Febbraro, M., Tonti, D., Marchetti, M., Ottaviano, M., Santopuoli, G., Chirici, G., 2015. Spatially explicit estimation of forest age by integrating remotely sensed data and inverse yield modeling techniques. iForest 9, 63–71. https://doi.org/10.3832/ifor1529-008.

Giannetti, F, Puletti, N, Quatrini, V, Travaglini, D, Bottalico, F, Corona, P, Chirici, G, 2018. Integrating terrestrial and airborne laser scanning for the assessment of single-tree attributes in Mediterranean forest stands. Eur. J. Remote Sens. 51, 795–807. https://doi.org/10.1080/22797254.2018.1482733.

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D, Moore, R, 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Griffiths, P, van der Linden, S, Kuemmerle, T, Hostert, P, 2013. A pixel-based Landsat compositing algorithm for large area land cover monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 6(5), 2088–2101. https://doi.org/10.1109/JSTARS.2012.2228167.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., et al., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853. https://doi.org/10.1126/science.1244693. American Association for the Advancement of Science.

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., de Bruin, S., Farina, M., Fatoyinbo, L., et al., 2021. Global maps of twenty-first century forest carbon fluxes. Nat. Clim. Change 11, 234–240. https://doi.org/10.1038/s41558-020-00976-6.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B., Maier, B., 2009. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. Int. J. Rem. Sens. 30, 5159–5175. https://doi.org/10.1080/01431160903022894.

Houghton, R.A., Nassikas, A.A., 2017. Global and regional fluxes of carbon from land use and land cover change 1850–2015. Global Biogeochem. Cycles 31 (3), 456–472.

Immitzer, M., Stepper, C., Böck, S., Straub, C., Atzberger, C., 2016. Forest ecology and management use of WorldView-2 stereo imagery and national forest inventory data for wall-to-wall mapping of growing stock. For. Ecol. Manag. 359, 232–246. https://doi.org/10.1016/j.foreco.2015.10.018.

INFC, 2021. Italian Forests. Selected Results of the Third National Forest Inventory, ISBN 978-88-338-5140-2. INFC 2015. Carabinieri Command of Forestry, Environmental and Agri-food units and CREA - Research Centre for Forestry and Wood. Tipografia Supernova (TN). Settembre 2021.

IPCC, 2006. In: Eggleston, H.S., Miwa, K., Srivastava, N., Tanabe, K. (Eds.), IPCC Guidelines for Greenhouse Gases Inventory. A Primer, Prepared by the National Greenhouse Gas Inventories Programme. IGES, Japan.

ISPRA, 2021a. Moving to the 2006 IPCC Guidelines for UNFCCC Reporting. ISPRA Rapporti 358/2021.

ISPRA, 2021b. National Inventory Report 2021 - Italian Greenhouse Gas Inventory 1990-2019. ISPRA Rapporti 341/2021.

Italian Ministry for the Environment, Land and Sea (IMELS), 2019. National Forestry Accounting Plan. Dicember 2019. https://www.minambiente.it/sites/default/files/archivio/allegati/clima/nfap_final_resubmission_2019_clean.pdf.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., et al., 2018. Remote sensing and forest inventories in Nordic countries–roadmap for the future, 2018 Scand. J. For. Res. 33, 397–412.

Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. Int. J. Appl. Earth Obs. Geoinf. 11 (5), 352–359. https://doi.org/10.1016/j.jag.2009.06.002.

Kulbokas, G., Jurevičienė, V., Kuliešis, A., Augustaitis, A., Petrauskas, E., Mikalajūnas, M., Vitas, A., Mozgeris, G., 2019. Fluctuations in gross volume increment estimated by the Lithuanian National Forest Inventory compared with annual variations in single tree increment. Balt. For. 25 (2), 273–280.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. Nucleic Acids Res. 5, 983–999. https://doi.org/10.1023/A:1010933404324.

McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. 110, 412–419. https://doi.org/10.1016/j.rse.2006. 09.034.

McRoberts, R.E., Naesset, E., Gobakken, T., 2013. Accuracy and precision for remote sensing applications of nonlinear model-based inference. IEEE Journal of Selected Topics in Applied Earth Observations and Remote 6, 27–34. https://doi.org/10.1109/JSTARS.2012.2227299.

McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., Westfall, J.A., 2016. Hybrid estimators for mean above-ground carbon per unit area. For. Ecol. Manag. 378, 44–56.

McRoberts, R.E., Næsset, E., Gobakken, T., Chirici, G., Condes, S., Hou, Z., Saarela, S., Chen, Q., Stahl, G., Walters, B.F., 2018. Assessing components of the model-based mean square error estimator for remote sensing assisted forest applications. Can. J. For. Res. 48, 642–649. https://doi.org/10.1139/cjfr-2017-0396.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review, 2011 ISPRS J. Photogrammetry Remote Sens. 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., Söderman, U., 2004. Laser scanning of forest resources:the nordic experience. Scand. J. For. Res. 19, 482–499. https://doi.org/10.1080/02827580410019553.

Nilsson, M, Nordkvist, K, Jonzén, J, Lindgren, N, Axensten, P, Wallerman, J, Egberth, M, Larsson, S, Nilsson, S, Eriksson, J, Olsson, H, et al., 2017. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. Remote Sens. Environ. 194, 447–454. https://doi.org/10.1016/j.rse.2016.10.022.

Nonini, L., Fiala, M., 2021. Estimation of carbon storage of forest biomass for voluntary carbon markets: preliminary results. J. For. Res. 32 (1), 329–338. https://doi.org/10.1007/s11676-019-01074-w.

Nord-Larsen, T., Schumacher, J., 2012. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. Remote Sens. Environ. 119, 148–157. https://doi.org/10.1016/j.rse.2011.12.022.

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Rem. Sens. 26 (5), 1007–1011. https://doi.org/10.1080/01431160512331314083.

Paul, T., Kimberley, M.O., Beets, P.N., 2021. Natural forests in New Zealand–a large terrestrial carbon pool in a national state of equilibrium. Forest Ecosystems 8 (1), 1–21. https://doi.org/10.1186/s40663-021-00312-0.

Peng, C, 2000. Understanding the role of forest simulation models in sustainable forest management. Environ. Impact Assess. Rev. 20 (4), 481–501. https://doi.org/10.1016/S0195-9255(99)00044-X.

Pinna, M., 1970. Contributo alla classificazione del clima d'Italia. Riv. Geogr. Ital. 77 (2), 129–152.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., et al., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. USA 108 (24), 9899–9904.

Schepaschenko, D., Moltchanova, E., Fedorov, S., Karminov, V., Ontikov, P., Santoro, M., See, L., et al., 2021. Russian forest sequesters substantially more carbon than previously reported. Sci. Rep. 11 (1), 1–7. https://doi.org/10.1038/s41598-021-92152-9.

Seebach, L, McCallum, I, Fritz, S, Kindermann, G, LeDuc, S, Böttcher, H, Fuss, A, 2012. Choice of forest map has implications for policy analysis: A case study on the EU biofuel target. Environ. Sci. Policy 22, 13–24. https://doi.org/10.1016/j.envsci.2012.04.010.

Sellers, P.J., Schimel, D.S., Moore, B., Liu, J., Eldering, A., 2018. Observing carbon cycle–climate feedbacks from space. *Proceedings of the National Academy of Sciences*, U.S.A. 115, 7860–7868. https://doi.org/10.1073/pnas.1716613115.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Tabacchi, G., Di Cosmo, L., Gasparini, P., Morelli, S., 2011. Stima Del Volume E Della Fitomassa Delle Principali Specie Forestali Italiene, Equazioni Di Previsione, Tavole Del Volume E Tavole Della Fitomassa Arborea Epigea.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. 112, 1982–1999. https://doi.org/10.1016/j.rse.2007.03.032.

Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., Gabler, K., Schadauer, K., Cienciala, E., et al., 2010. National Forest Inventories. Pathways for Common Reporting, vol. 1. European Science Foundation, pp. 541–553.

UNFCCC, 2015. The Paris Agreement. Report of the Conference of the Parties on its Twenty-First Session held in Paris from 30 November to 11 December 2015. http://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english_.pdf.

Van Der Werf, G.R., Randerson, J.T., Giglio, L., Van Leeuwen, T.T., Chen, Y., Rogers, B. M., Kasibhatla, P.S., et al., 2017. Global fire emissions estimates during 1997–2016. Earth Syst. Sci. Data 9 (2), 697–720. https://doi.org/10.5194/essd-9-697-2017.

Vangi, E., D'Amico, G., Francini, S., Giannetti, F., Lasserre, B., Marchetti, M., McRoberts, R.E., Chirici, G., 2021. The effect of forest mask quality in the wall-to-wall estimation of growing stock volume. Rem. Sens. 13, 1038. https://doi.org/10.3390/rs13051038.

Vizzarri, M, Chiavetta, U, Chirici, G, Garfì, V, Bastrup-Birk, A, Marchetti, M, 2015. Comparing multisource harmonized forest types mapping: A case study from central Italy. Iforest-Biogeosci. For. 8, 59–66. https://doi.org/10.3832/ifor1133-007.

Waser, L.T., Ginzler, C., Rehush, N., 2017. Wall-to-Wall tree type mapping from countrywide airborne remote sensing surveys. Rem. Sens. 9 https://doi.org/10.3390/rs9080766.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: a review. Can. J. Rem. Sens. 42 (5), 619–641. https://doi.org/10.1080/07038992.2016.1207484.

Williams, P.A., Allen, C., Macalady, A., Griffin, D., Woodhouse, C.A., Meko, D.M., Swetnam, T.W., et al., 2012. Temperature as a potent driver of regional forest drought stress and tree mortality. Nat. Clim. Change 3, 292–297. https://doi.org/10.1038/nclimate1693.

Wolfe, R., Masek, J., Saleous, N., Hall, F., 2004. LEDAPS: mapping North American disturbance from the Landsat record. In: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, vol. 1. IEEE.

Xu, L., Saatchi, S.S., Yang, Y., Yu, Y., Pongratz, J., Bloom, A.A., Bowman, C., et al., 2021. Changes in global terrestrial live biomass over the 21st century. Sci. Adv. 7 (27), eabe9829 https://doi.org/10.1126/sciadv.abe9829.