LIFE 14ENV/IT/000414
Demonstrating Remote Sensing integration in sustainable forest management
FRESh Life

ACTION B3

Mapping SFM indicators

Report on the technical and economic viability of using geostatistical methods and techniques for the spatial estimation of growing stock and above ground biomass, at the forest compartment level

Viterbo, 31/10/2017

Summary

1. Goals 3

2. Milestones and Deliverables 3

3. Methodology 3

3.1 Field sample data processing 5

3.2 LIDAR point cloud processing and regression analysis 5

4. Concluding remarks on technical feasibility of spatial estimation of growing stock and above ground biomass 10

5. Economic viability considerations: time and costs to process indicators maps 11

6. References 12

Annex 1 - SFM Indicators Maps 14

1. Goals

University of Tuscia is the Beneficiary responsible for implementation of Action B3 - Mapping SFM indicators. The goals of the Action B3 are to test and evaluate methods coupling remote sensed information collected from RPAS with plot-level data to derive:

- maps of European Forest Types for the pilot study areas
- maps of selected Forest Europe SFM indicators

This report summarizes the technical results achieved in the production of maps of the SFM indicators “#1.3 Growing stock” and “#1.4 Above ground biomass” in the pilot study areas of Rincine, Caprarola and Bosco Pennataro. This task, which concerns the application of statistical methods for the spatial estimation of these variables of interest, has been carried out with the support of two external assistance service contracts (Università degli Studi di Siena-DEPS and CREA). By accomplishing this task, the Action B3 project milestone is reached (see Section 2).

2. Milestones and Deliverables

The B3’s Project Milestone is

Milestone name	*Deadline*
Report on the technical and economic viability of coupling remote sensed information, collected from RPAS, with plot-level data to map selected Forest Europe SFM indicators at operational scale	09/2017

The B3’s Project Deliverable Products are

Deliverable name	*Deadline*
Maps of European Forest Types for the pilot study areas	12/2016
Report on the technical and economic viability of using high spatial resolution optical data to stratify by European Forest Types (EFTs) medium- to large size forest management units	2/2017
Maps of SFM indicators “Defoliation (# 2.3)”, “Forest damage (# 2.4)”, “Number of tree species (# 4.1)” and “Area covered by introduced tree species (# 4.4)” for the pilot study areas	3/2017
Report on the technical and economic viability of using very high spatial resolution optical data for mapping forest health and tree species related SFM indicators at the forest compartment level	4/2017
Maps of SFM indicators: “Growing stock (# 1.3)” and “Above ground biomass (# 1.4)” for the pilot study areas	6/2017
Report on the technical and economic viability of using geostatistical methods and techniques for the spatial estimation of growing stock and above ground biomass, at the forest compartment level	7/2017

3. Methodology

The bulk of this task was to process the huge amount of information from LIDAR point cloud data, in order to derive a number of LIDAR variables (called metrics) that can be tested for correlation with the variables of interest, i.e. growing stock and aboveground biomass. To this end, the sampling surface of the 23x23 m sampling units (529 m2) covering the full spatial extent of each test site was used as reference grid to process LIDAR metrics. Reference values of the variables of interest were devised by processing sample data collected in 50 sample units of the sampling surface during Action B2. The design-based Tessellation Stratified Sampling (TSS) combined with One-Per-Stratum Stratified sampling (OPSS) approach, allowed to the achieve the so-called *spatially balanced sample* that is, a sample in which units are well spread throughout the test sites, being uniformly placed (i.e., selected with uniform probability density) in (50) equal size strata of the study areas (Figure 1). This task of Action B3, by combining field and LIDAR data acquired in Action B2, attempted to spatially estimate the variables of interest by linear regression (albeit with imperfect accuracy) over the sampling surface covered by LIDAR data acquisition in the three sites.

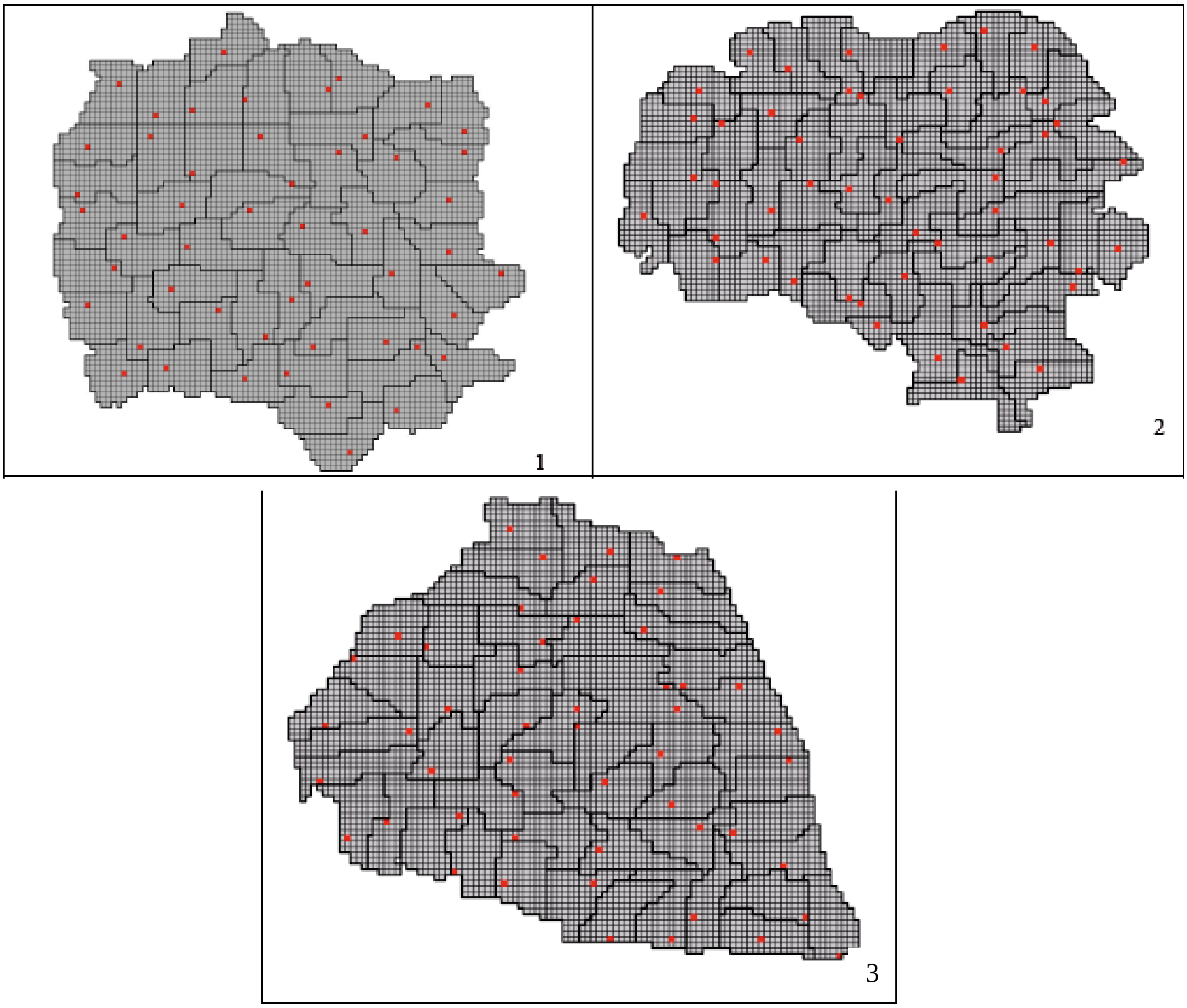


Figure 1. Spatial distribution by OPSS of the 50 sample units in the three test sites: 1 – Rincine; 2 – Caprarola; 3 – Bosco Pennataro.

The overall data processing workflow leading to the production of maps is displayed in Figure 2.

In the subsequent sections, technical details on each data processing step are given.

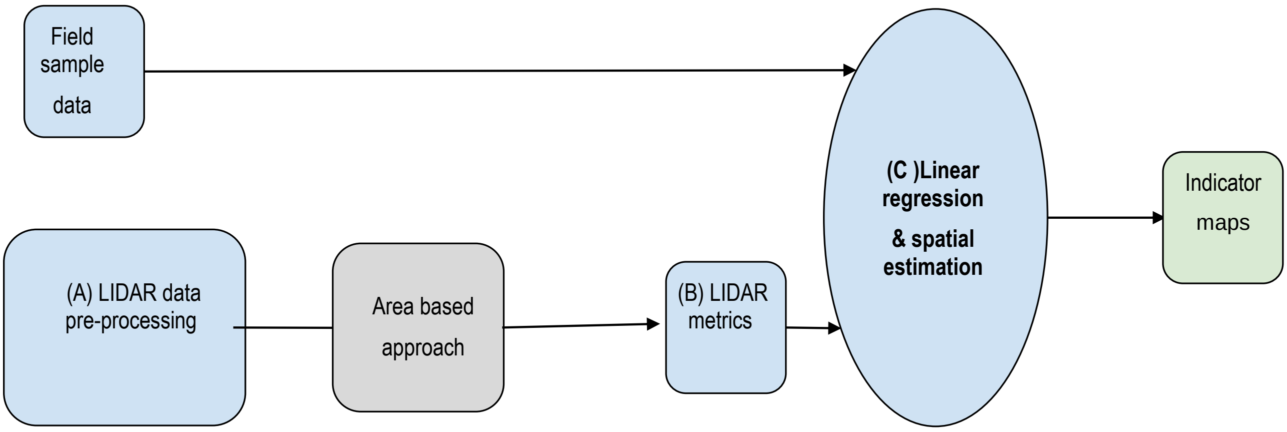


Figure 2. Methodological workflow for the spatial estimation of the SFM indicators of interest.

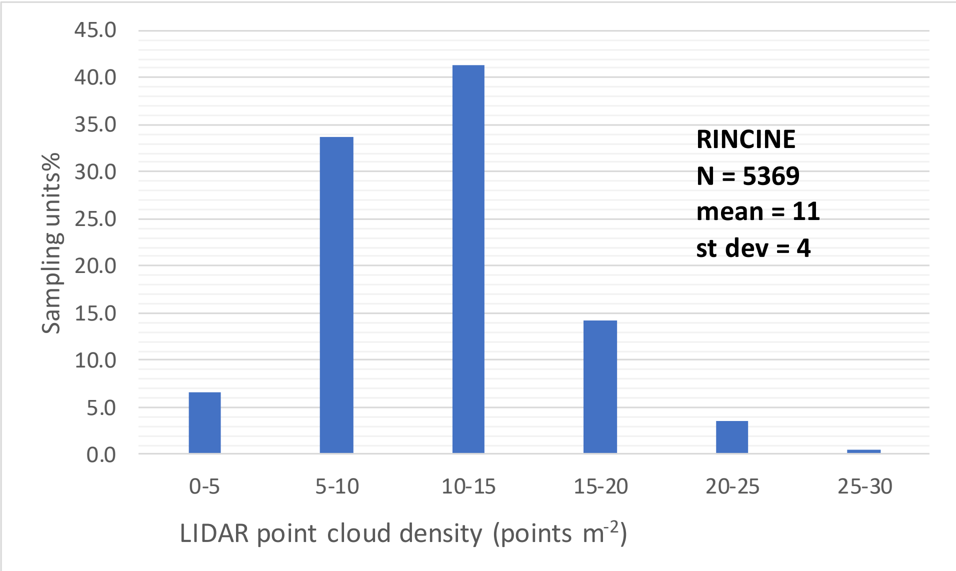
3.1 Field sample data processing

The growing stock and aboveground biomass indicators were estimated at each sample site location using species-specific volume and biomass models devised for the national forest inventory in Italy, using tree DBH and measured height as predictor variables (Tabacchi et al., 2011). Accordingly, the values of the reference variables y_j were calculated for the one sample-site location surveyed in each stratum (y_1, y_2, ...y_{50}).

3.2 LIDAR point cloud processing and regression analysis

LIDAR data acquired in Action B2 was available for the entire area covered by the field inventory in Rincine and Caprarola and for about 70% of Bosco Pennataro test site (nearly 200 ha). In any case, the areas covered by laser scanning largely exceed the size foreseen in the proposal (100 ha).

Multiple-returns point clouds were available for the three test sites, as shown in Figure 3.



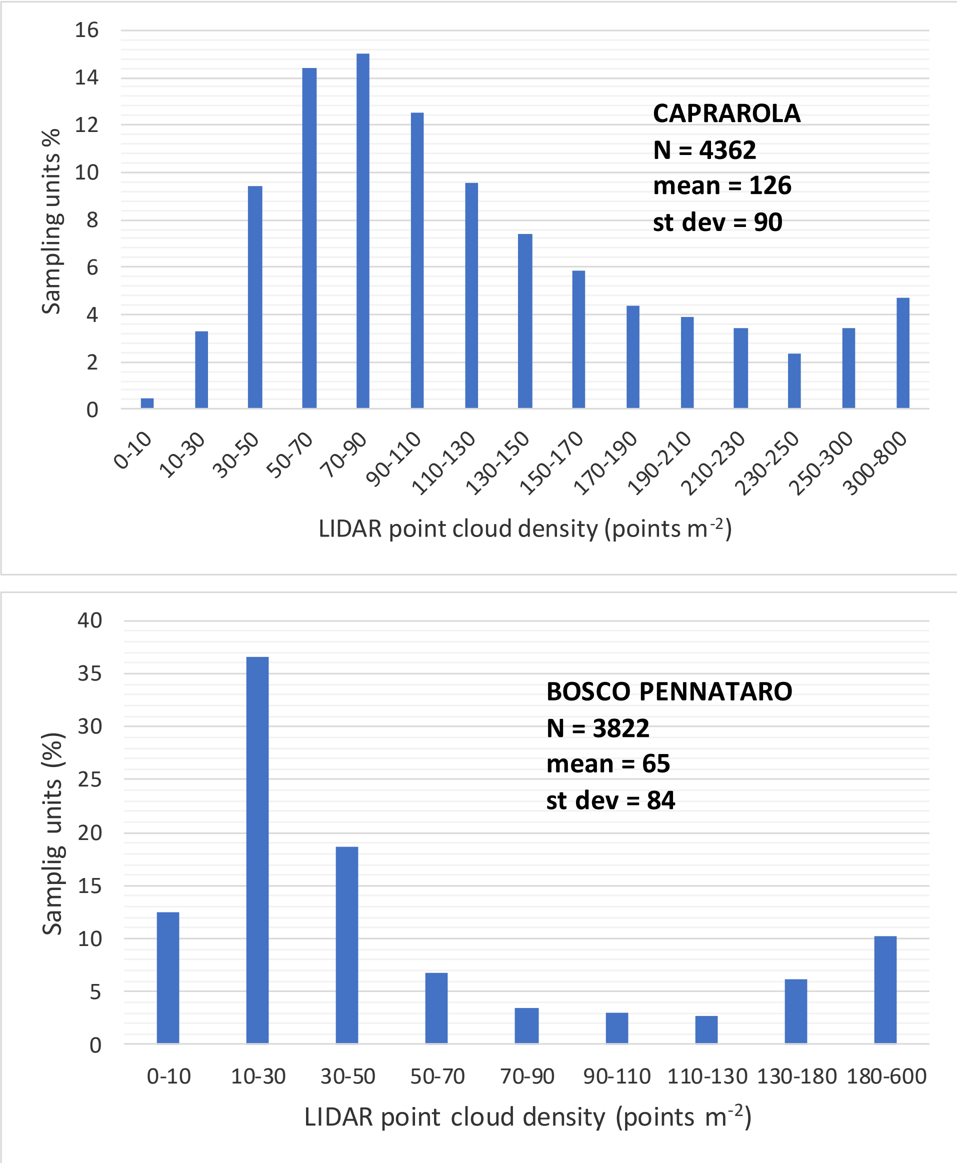
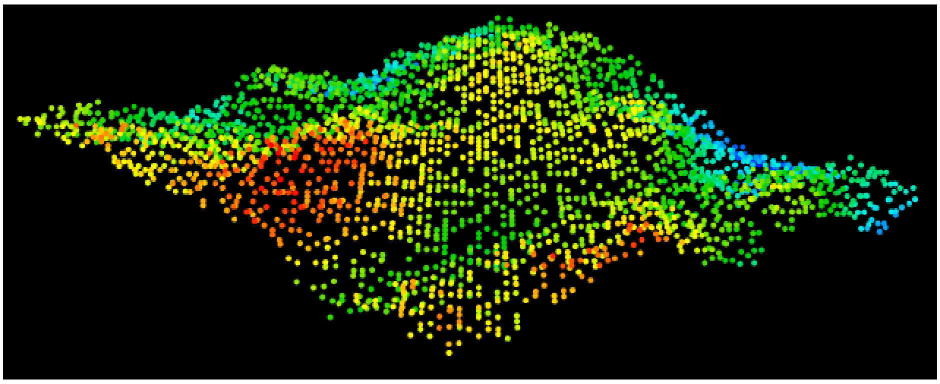


Figure 3. Frequency distribution and statistics of LIDAR point cloud density in the three test sites. N=number of sampling units scanned by LIDAR sensor per test site.

LIDAR point cloud has been processed following these steps:

A – Pre-processing

- Point clouds derived from different scans over the same study area have been co-registrated and merged into one;
- The point cloud has been divided into tiles according to the OPSS sampling grid shown Figure 1;
- Following a routine written in R and available at the github-repository of CREA-Forest Geomatics Laboratory (https://github.com/ForGeoLab/FreshLIFE) , that mainly uses the lidR package, the point cloud of each tile has been cropped to the sampling grid and classified into ground and not-ground returns (Figure 4); statistics of the proportion of ground returns are shown in table 1.
- From ground returns a raster-DTM with 1 meter of resolution has been derived and used to normalize the cropped point cloud itself.



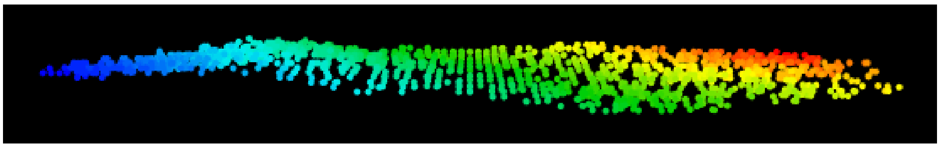


Figure 4. Top: not-ground (upper canopy) returns, bottom: ground returns.

Table 1. Statistics of the proportion of ground returns out of total returns per sampling unit.

Test area	*Average proportion (%)*	*Standard deviation*
RINCINE	16	8
CAPRAROLA	2	2
BOSCO PENNATARO	3	6

B- LIDAR Variable extraction by sampling units

The LIDAR normalized point cloud has been further processed to derive the 27 LIDAR metrics listed in Table 2, for each 23x23 m sampling unit of the sampling surface covered by LIDAR data.

Table 2. List of LIDAR derived metrics for each sampling unit.

	LIDAR metrics	*Definition*
Height-based (23 metrics)	Maximum (h_{max}) and mean (h_{mean}) heights	The maximum and mean heights above the ground of all first returns
	Quadratic Mean of Heights (*qmh*)	The quadratic mean of heights of all first returns
	Coefficient of variation of heights (h_{cv})	Coefficient of variation of heights of all first returns
	Percentile heights (h_{P10}, h_{P20}, h_{P30}, h_{P40}, h_{P50}, h_{P60}, h_{P70}, h_{P75}, h_{P80}, h_{P90})	The percentiles of the heights distributions of all first returns
	Skewness (h_s) and kurtosis (h_k) of heights	The skewness and kurtosis of the normalised heights of all first returns
	Mean heights within three layers (h_{mean1}, h_{mean2}, h_{mean3})	Mean of heights lower than 1/3 (h_{mean1}), between 1/3 and 2/3 (h_{mean2}), and higher than 2/3 (h_{mean3}) of h_{max}

	Square of percentile heights (h_{QP75}, h_{QP90})	The squared value of percentiles of the canopy heights distributions of first returns
	Percentile heights of canopy points (h_{PC75},h_{PC90})	The percentiles of the canopy heights distributions of first returns
Density-based (4 metrics)	Percentage of points over the ground (*ogp*)	The number of first returns classified as no-ground over the total first returns
	Percentage of points within three layers(d_1, d_2, d_3)	Percentage of points in three layers lower than 1/3 (d_1), between 1/3 and 2/3 (d_2), and higher than 2/3 (d_3) of h_{max}

C - Regression model and spatial estimation

Following the so-called "area-based approach" (Naesset, 2002), empirical relationships were searched between growing stock or aboveground biomass reference values estimated from the ground sample data and the variety of LIDAR-derived metrics for the matching locations, using the simple linear regression model:

$$y_j = a + b*x_j + e_j$$

Where:

y_j is an observed reference value of the dependent variable (growing stock or above ground biomass) at the j sampling unit (50 sampling units for Rincine and Caprarola, 36 for Bosco Pennataro);

x_j is the independent variable i.e. the LIDAR-derived metric available for all the sampling surface;

e_j is the residual error of the model at that location.

The predicted value of y_j is:

$$\hat{y}_j = a + b*x_j$$

Regression analysis estimated the values of *a* and *b* that minimized the sum of errors $e_j = y_j - \hat{y}_j$.

In order to evaluate how well linear regression models from different LIDAR metrics fitted the data in each test site, the coefficient of determination (R^2) was used as goodness-of-fit metric to compare fits across different linear models. The resulting best simple linear regression model for each indicator and test site is displayed in Table 3.

The selected linear regression models have been used to produce the spatial estimation the SFM indicators over the entire area covered by LIDAR data acquisition. This information has been further processed into statistics of each indicator at the forest compartment unit level (Table 4), so as to map the average value of the indicators at the spatial scale most relevant for operational purposes (Annex 1).

Table 3. Parameter estimates and associated coefficient of determination for the best fit simple linear regression models devised for the spatial estimation of growing stock and aboveground biomass in the three test sites.

	Indicator # 1.3 Growing stock				*Indicator # 1.4 Aboveground biomass*			
Study area	*Selected metric*	*model equation*	*R^2*	*RMSE ($m^3\ ha^{-1}$)*	*Selected metric*	*model equation*	*R^2*	*RMSE ($t\ ha^{-1}$)*
Rincine	*qmh*	-52.559 + 36.647x	0.80	111	*qmh*	62031 + 15350x	0.68	64.7
Caprarola	h_{mean2}	-155.069 + 33.093x	0.26	167	h_{mean2}	-112684 + 25747x	0.25	134.4
Bosco Pennataro	h_{mean}	-34.938 + 22.427x	0.44	84	h_{mean}	-34232 + 17997x	0.42	69.9

Table 4. Growing stock and aboveground biomass statistics by forest compartment units in the three test sites.

RINCINE	*Growing stock volume*				*Aboveground biomass*			
Forest compartment #	*Average ($m^3\ ha^{-1}$)*	*St. dev. ($m^3\ ha^{-1}$)*	*CV%*	*Total (m^3)*	*Average ($t\ ha^{-1}$)*	*St. dev. ($t\ ha^{-1}$)*	*CV%*	*Total (t)*
29	300	103	34	3445	210	43	21	2408
30	312	97	31	4610	215	41	19	3172
32	282	87	31	7251	202	37	18	5198
47	327	112	34	6683	221	47	21	4516
48	357	141	40	6169	233	59	25	4038
49	430	89	21	9805	264	37	14	6023
50	401	150	37	5871	252	63	25	3691
54	504	113	22	4024	295	47	16	2357
55	752	244	32	3501	399	102	26	1858
71	649	202	31	12599	355	88	25	6891
72	491	243	49	6930	290	102	35	4090
73	607	262	43	6034	338	110	32	3363
74	326	157	48	7218	221	66	30	4877
75	307	173	56	1784	212	72	34	1236
79	377	190	51	3807	241	81	33	2439
80	213	112	53	2302	173	47	27	1871
82	532	231	43	8751	307	97	32	5048
83	547	197	36	6745	313	83	26	3861
84	445	185	42	11612	270	78	29	7051
85	484	136	28	7117	287	57	20	4217
Private	170	99	58	1227	155	41	27	1118
Study area	**413**	**204**	**49**	**127486**	**257**	**86**	**33**	**79324**

CAPRAROLA	*Growing stock volume*				*Aboveground biomass*			
Forest compartment #	*Average ($m^3\ ha^{-1}$)*	*St. dev. ($m^3\ ha^{-1}$)*	*CV%*	*Total (m^3)*	*Average ($t\ ha^{-1}$)*	*St. dev. ($t\ ha^{-1}$)*	*CV%*	*Total (t)*
54	448	93	21	5855	357	72	20	4659
55	530	93	17	5382	421	72	17	4274
56	482	63	13	7808	383	49	13	6203
57	510	70	14	11834	405	55	14	9398
58	512	89	17	13077	406	69	17	10378
59	502	91	18	8094	398	70	18	6426
60	530	96	18	14709	420	75	18	11665
61	458	66	14	9223	364	51	14	7336
62	427	68	16	5396	340	53	16	4299
63	435	109	25	4976	347	86	25	3968
64	522	80	15	8445	414	62	15	6700
65	537	85	16	9175	426	66	15	7274
66	510	73	14	6633	404	57	14	5264
67	426	80	19	4237	339	62	18	3376
68	467	76	16	7121	372	59	16	5662
69	475	100	21	5103	378	77	21	4056
70	430	49	11	1364	342	38	11	1087
Study area	**491**	**90**	**18**	**128432**	**390**	**70**	**18**	**102023**

BOSCO PENNATARO	Growing stock volume				Aboveground biomass			
Forest compartment #	Average (m^3 ha^{-1})	St. dev. (m^3 ha^{-1})	CV%	Total (m^3)	Average (t ha^{-1})	St. dev. (t ha^{-1})	CV%	Total (t)
4	383	90	24	11565	301	72	24	9094
5	383	97	25	13471	302	78	26	10592
6	377	84	22	11078	296	67	23	8708
7	372	92	25	5495	293	73	25	4318
8	393	93	24	10277	309	75	24	8085
9	386	95	25	9119	303	76	25	7171
10	397	92	23	1113	312	74	24	876
11	390	95	24	4916	307	76	25	3867
12	402	86	21	10162	316	69	22	7998
13	390	101	26	8092	307	81	26	6365
14	410	86	21	2104	323	69	21	1656
16	411	97	23	566	324	77	24	445
17	314	128	41	183	245	102	42	143
Study area	**387**	**93**	**24**	**88138**	**304**	**74**	**24**	**69319**

4. Concluding remarks on technical feasibility of spatial estimation of growing stock and above ground biomass

Summarizing results achieved in the three test areas we can draw the general conclusion that the spatial estimation of the growing stock and aboveground biomass indicators using LIDAR height-based metrics as predictors led to disparate results in the three test sites. Moderate to strong linear relationships were found in Bosco Pennataro and Rincine, with R^2 from linear regression ranging from 0.44 to 0.80 for the growing stock and from 0.42 to 0.68 for aboveground biomass respectively. The independent variable x is in both cases related to mean of heights of all first returns. In Caprarola test site the goodness of fit of the prediction models is much lower, though the R^2 indicate that one quarter of the variation of growing stock or aboveground biomass is explained by the LIDAR height-based metric, specifically the mean heights of returns between 1/3 and 2/3 of the maximum height of all first returns.

A possible explanation of the different performances of the predictive models in the three test sites can be the different canopy penetration of the LIDAR, which affected the amount of ground returns and, accordingly, the quality of the derived raster DTM used for the normalization of the LIDAR point cloud. Despite the density of the LIDAR point cloud in Caprarola and Bosco Pennataro was much higher than in Rincine, only a very small percentage of returns were classified as ground (on average 2-3%). In Rincine the average proportion of ground returns was more consistent (16%), leading to a more accurate DTM, a more accurate estimation of the canopy height, variable to which both growing stock and aboveground biomass are correlated, and therefore a better overall fit of the regression model to the data.

It can be concluded from this test that the quality of maps of growing stock and aboveground biomass SFM indicators in the three test areas seems to be heavily influenced by the density of ground returns. Scanning during the leaf-on season, combined with dense forest stands like those covering the test areas, caused ground surfaces hidden below crown foliage to be difficult to acquire, since the light hitting the leaves rarely reached the ground in the first place.

But even so, the test demonstrates that is technically feasible to derive reliable spatial estimates of growing stock and above ground biomass by LIDAR-assisted inference, considering that optimal results were achieved in the most difficult conditions of Rincine test site, which has the highest heterogeneity in terms of forest types and spatial variability of the variables of interest. The limitations arising from a single flight in leaf-on conditions suggest that a LIDAR acquisition also during leaf-off season, by providing a better identification of the ground surface, would ultimately result in better LIDAR-assisted predictive models.

When spatial estimates of the variables of interest are derived with a good model fitting, important benefits arise from an operational point of view:

- The final user may be able to process estimates of means, variability or total value of the indicators for areas of interest, e.g. forest compartments at a much higher accuracy or lower costs than using field survey only (see Table 4);
- The final user may be able to map the variables of interest, with the limits of a predictive model, but at a much higher spatial resolution than would be possible by field survey.

5. Economic viability considerations: time and costs to process indicators maps

A quantification of the time required to produce the maps of the growing stock and above ground biomass indicator is reported in Table 5. This time is compared with the time required for field data collection in the sample plots for the same indicators, as recorded in the specific data collection sheet.

Table 5. Quantification of time required to process indicators maps in the three test sites.

Test site	*Time required to process indicators maps (hours)*				*Time spent for field survey(hours) 50 plots*
	LIDAR Pre-processing	*Modelling*	*Spatial prediction*	*TOTAL*	
Rincine	3	1	~6	10	135
Caprarola	3	1	~5	9	140
Bosco Pennataro	3	1	~4	8	289

An evaluation of costs of mapping growing stock and above ground biomass indicators is reported in Tables 6-9.

The hourly rate of a Technician for LIDAR data processing is based on market prices (60 €/hour).

The hourly rate of a Junior technician for field sample data collection is calculated based on the additional staff contract costs (15 €/hour).

The costs of each step of the process (Field sample data collection, Lidar acquisition, LIDAR Processing and map production) are detailed in Tables 6- 7. The total cost of producing maps of the growing stock and above ground biomass indicators is the sum of the costs of these three stages of the mapping process. The total cost is estimated for each test site and it is reported in Table 8. The total cost is significantly affected by the cost of LIDAR data acquisition and, to a lesser extent, by the cost of field survey.

Applying the same sampling intensity used in the project (1 plot of 529 m^2 every 5 ha), the cost of field survey on 50 plots, calculated on the basis of the hourly rate of an experienced professional as result of market prices (40 €/hour), would amount to 18,000 € (Table 9). Based on this assessment, one can argue that the cost of business as usual scenarios, i.e. traditional forest inventory, reaches the (average) cost of Lidar derived indicators maps when the number of sampling plots equals to 70 units ca (point B, Figure 5). At this sampling intensity, the total area covered by sample plots is 3.7 ha, i.e. the 1.4% of the test area. At the same cost, the LIDAR based approach provides maps of the indicators and related benefits for the users (see § 4).

Table 6. Costs of field sample data collection in 50 plots, in the three test sites.

Test site	*Time*	*Staff*	*Total time*	*Unit cost*	*Total cost*
	Hours per person	Number	Hours	€/hour	€
Rincine	135	3	405	15	6075
Caprarola	140	3	420	15	6300
Bosco Pennataro	289	3	867	15	13005

Table 7. Costs assessed for Lidar acquisition, LIDAR Processing and map production in the three test sites.

Test site	*Lidar acquisition*			*LIDAR Processing and map production*		
	Area	*Unit cost*	*Total cost*	*Time*	*Unit cost*	*Total cost*
	ha	€/ha	€	hour	€/hour	€
Rincine	250	70	17500	10	60	600
Caprarola	250	70	17500	9	60	540
Bosco Pennataro	200	70	14000	8	60	480

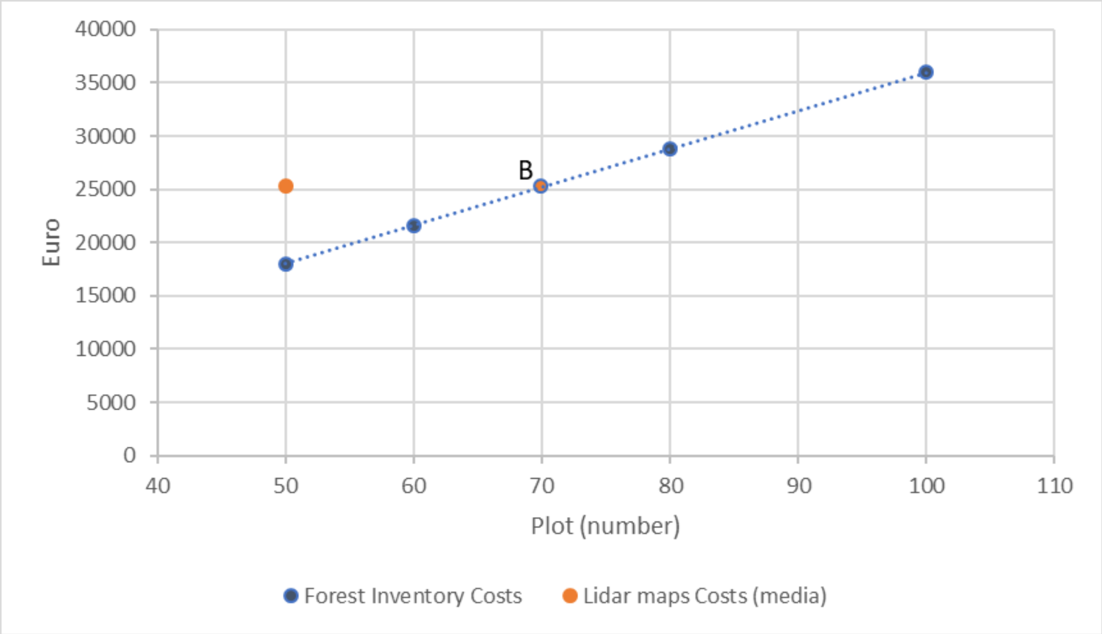
Table 8. Total costs of the maps of the growing stock and above ground biomass indicators in the three test sites.

Test site	*Field sample data collection*	*Lidar acquisition*	*LIDAR Processing and map production*	***TOTAL COST***	
	€	€	€	€	€/ha
Rincine	6075	17500	600	24175	97
Caprarola	6300	17500	540	24340	97
Bosco Pennataro	13005	14000	480	27485	110

Table 9. Total cost of field data collection in forest inventory plots.

Plots	*Time Plot*	*Cost*	*Total cost*
Number	Hours	€/hours	€
50	9	40	18000

Figure 5. Comparison of forest inventory costs by increasing the number of sampling units, with the average cost of Lidar derived Maps growing stock and above ground biomass indicator (based on 50 plots).



In conclusion, despite the cost of business as usual, i.e. traditional forest inventory in sample plots, is lower than the cost of LIDAR derived maps of growing stock and aboveground biomass indicators for sample sizes < 70 units, it must be emphasized that the final benefits of the two approaches cannot be compared. In fact, in ordinary field-work the value of the indicators is known only for a relatively small fraction of the sampling surface, while (good) regression models derived from LIDAR data allow the spatial estimation of these variables over all the sampling units of this surface.

6. References

Fattorini L., Corona P., Chirici G., Pagliarella M.C. (2015). Design-based strategies for sampling spatial units from regular grids with applications to forest surveys, land use and land cover estimation. Environmetrics, 26, 216-228.

Nasset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two - stage procedure and field data. Remote Sens. Environ. 80:88−99.

Tabacchi G., Di Cosimo L., Gasparini P., Morelli S. (2011). Stima del volume e della fitomassa delle principali specie forestali italiane. Equazioni di previsione, tavole del volume e tavole della fitomassa arborea epigea. Consiglio per la Ricerca e la sperimentazione in Agricoltura, Unità di Ricerca per il Monitoraggio e la Pianificazione Forestale. Trento. 412 pp.

Annex 1 - SFM Indicators Maps

Figure 1. Average value of growing stock volume by forest compartments units of the study area of Rincine (blanks correspond to non-forest areas within forest compartments).

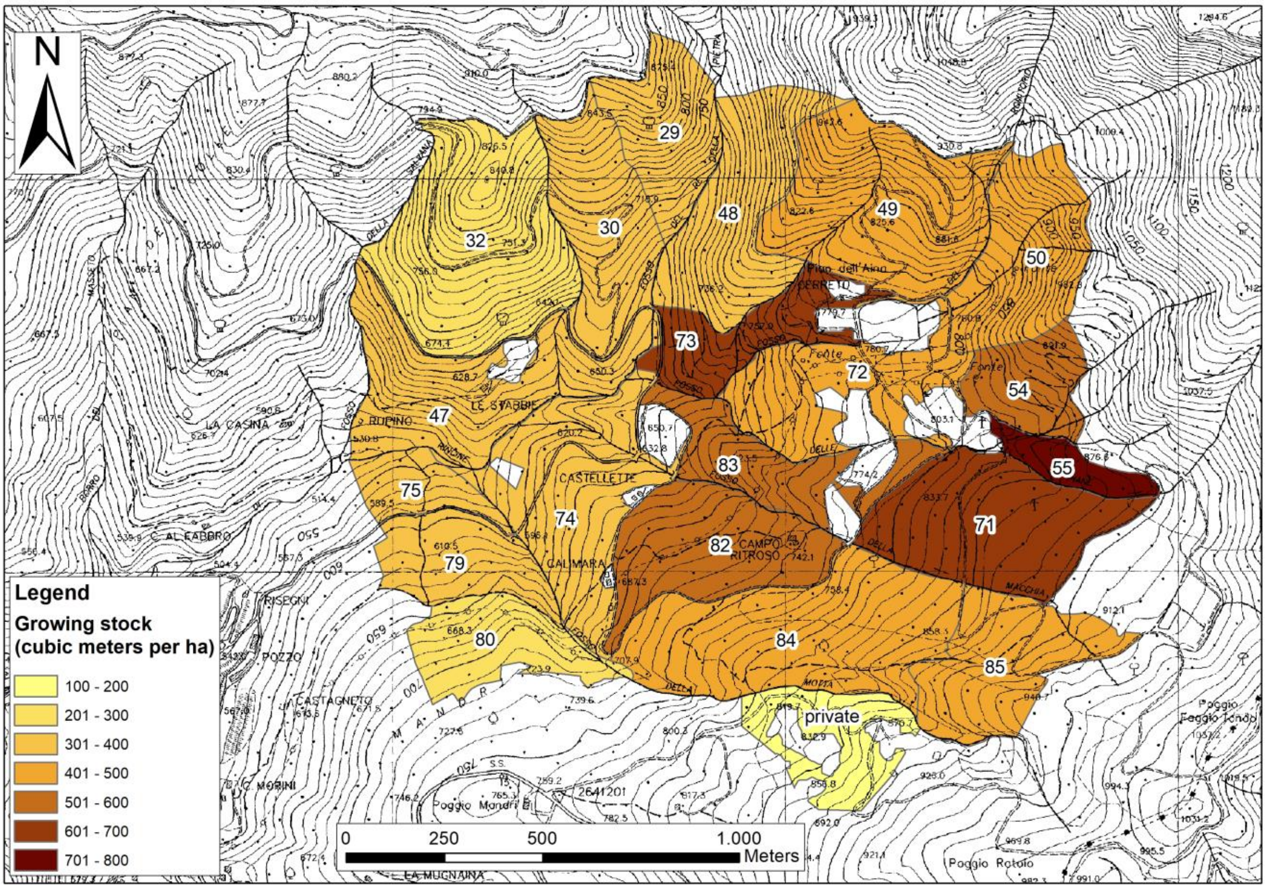


Figure 2. Average value of aboveground biomass by forest compartments units of the study area of Rincine (blanks correspond to non-forest areas within forest compartments).

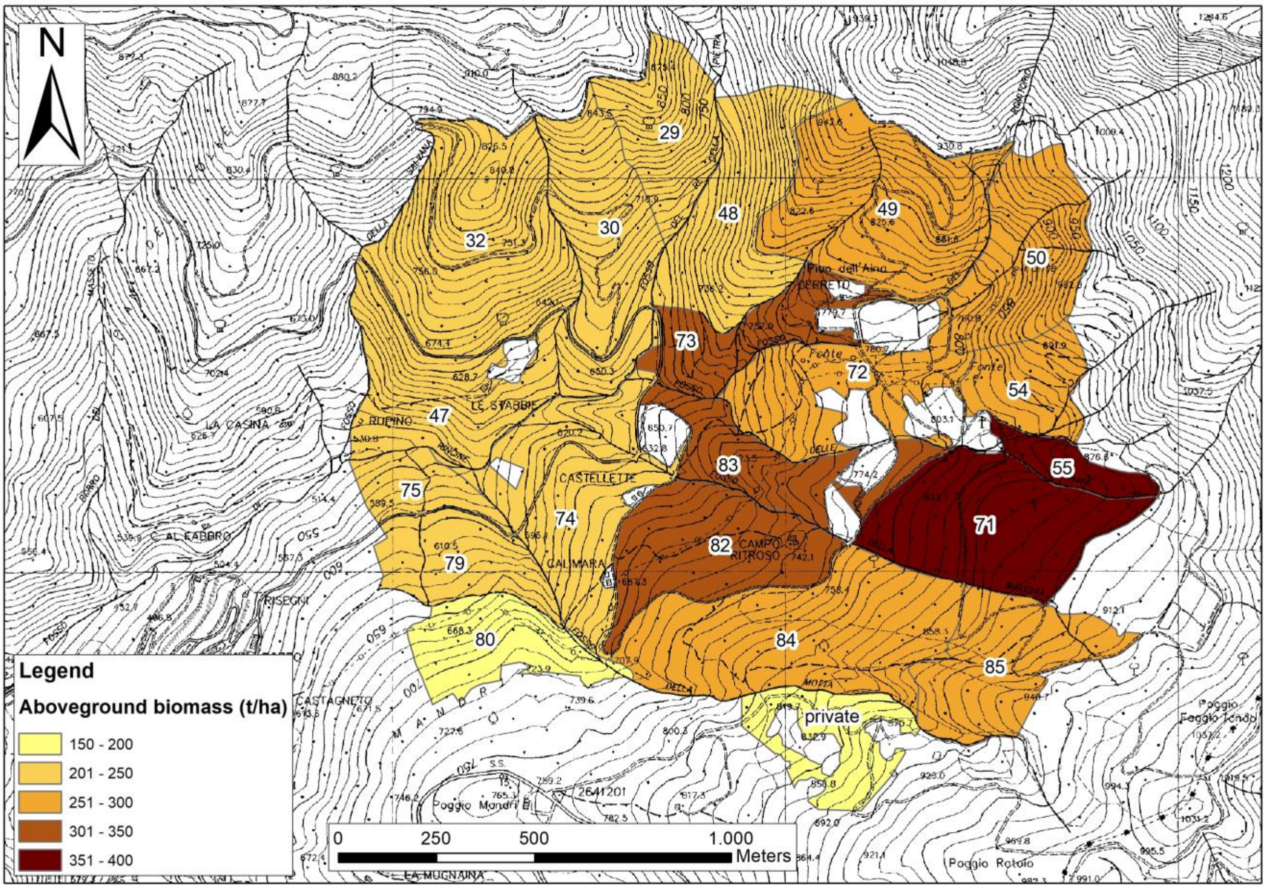


Figure 3. Average value of growing stock volume by forest compartments units of the study area of Caprarola.

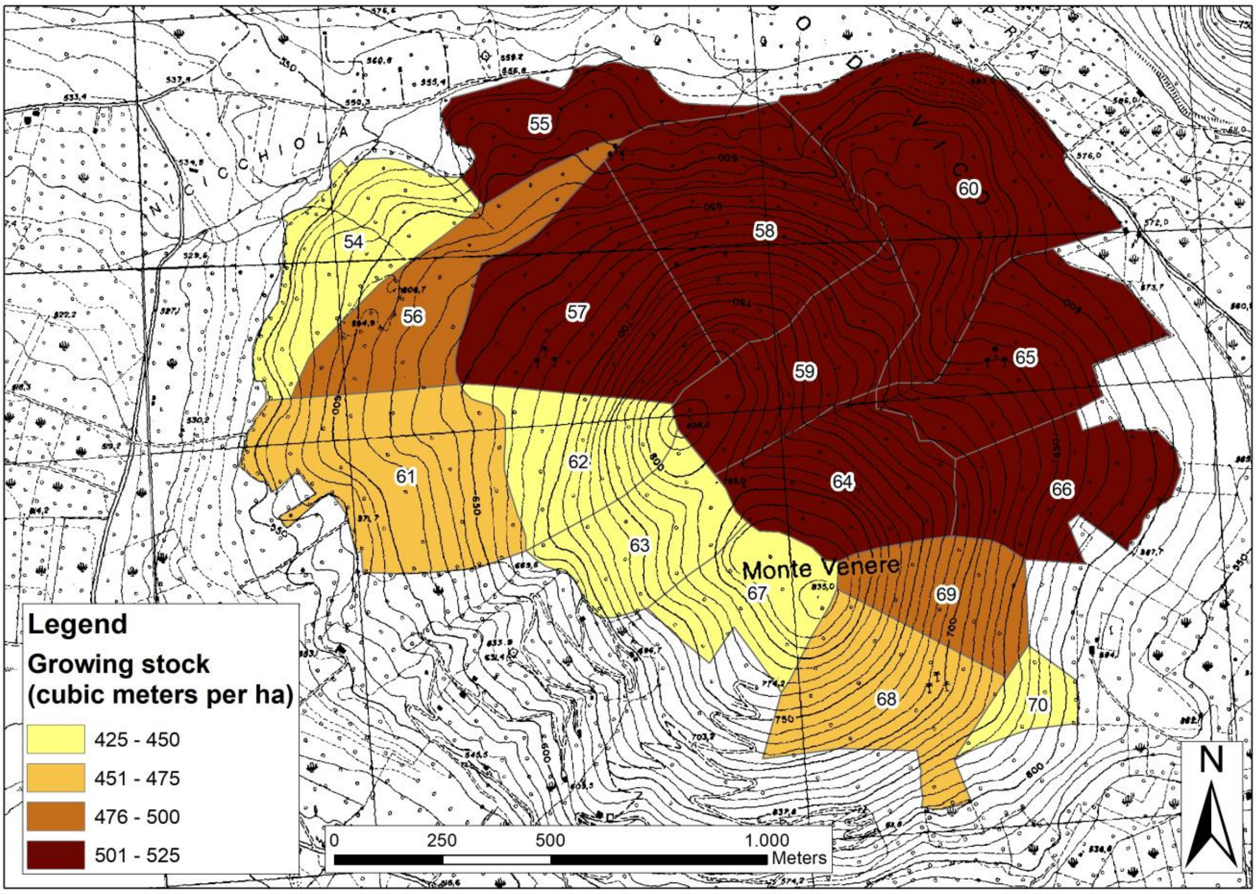


Figure 4. Average value of aboveground biomass by forest compartments units of the study area of Caprarola.

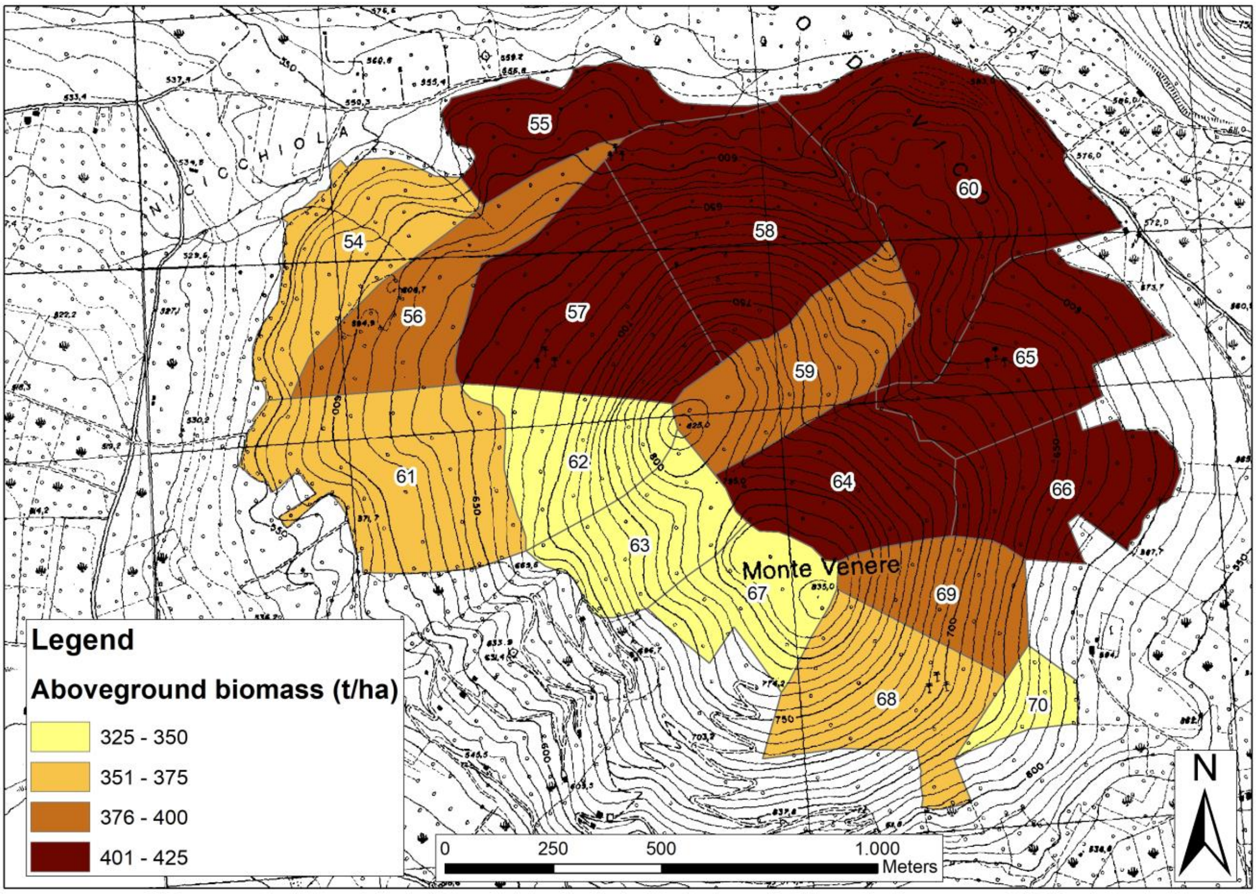


Figure 5. Average value of growing stock volume by forest compartments units of the study area of Bosco Pennataro.

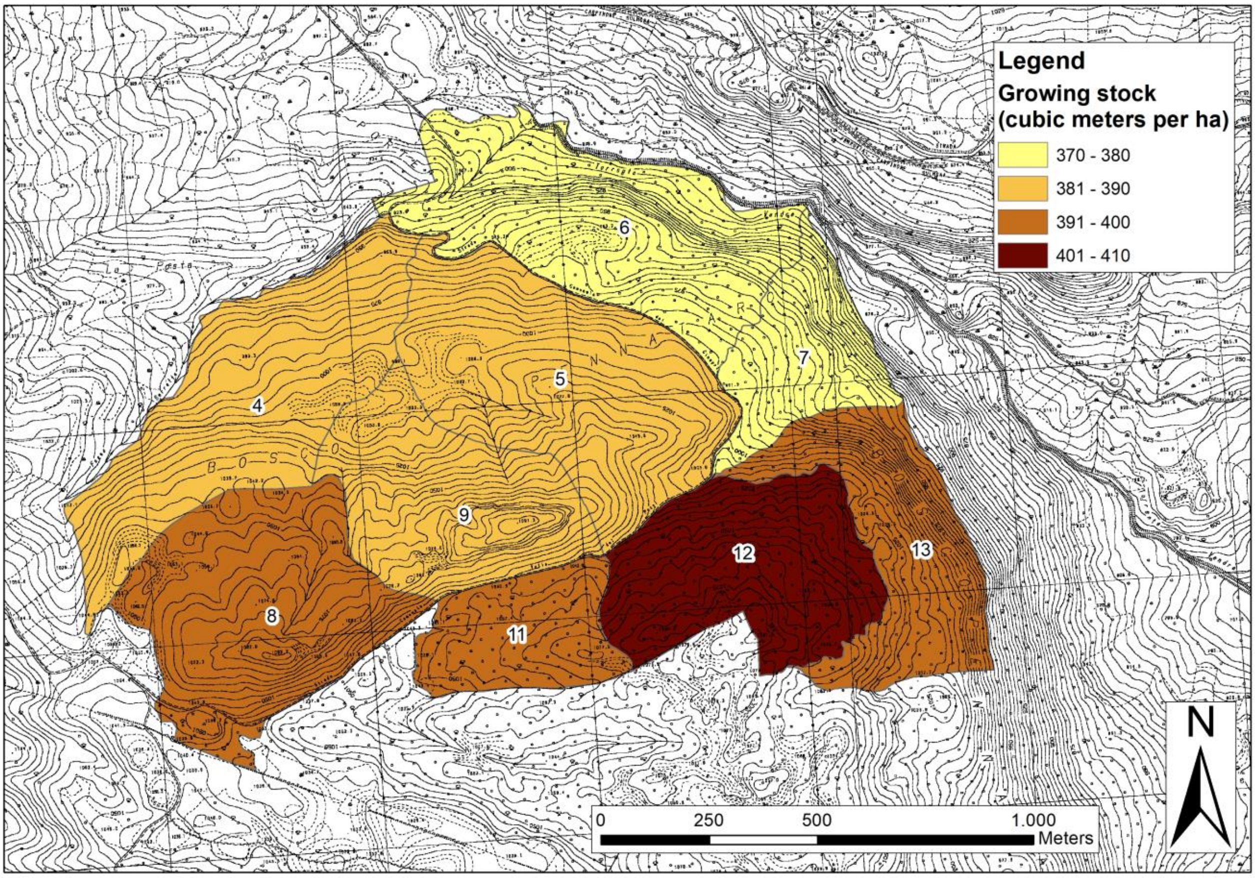


Figure 6. Average value of aboveground biomass by forest compartments units of the study area of Bosco Pennataro.