



# FLORE

### Repository istituzionale dell'Università degli Studi di Firenze

# Application of big data analytics in remote sensing supporting sustainable forest management

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Application of big data analytics in remote sensing supporting sustainable forest management / Giovanni D'Amico. - (2022).

Availability:

This version is available at: 2158/1259784 since: 2022-03-03T11:25:07Z

*Terms of use:* Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:

(Article begins on next page)



1	NIVERSITY	1	
2	Ph.D. program in Sustainable Management of		
3	Agricultural, Forestry and Food Resources		
4	Curriculum: Forest Economics, Planning and Wood Science		
5	Cycle XXXIV		
6	Coordinator Prof. Erminio Monteleone		
7	Application of big data analytics in remote sensing		
8	supporting sustainable forest management		
9	ACADEMIC DISCIPLINE		
10	AGR/05 - Forest management and silviculture		
11	Doctoral Candidate	Supervisor	
12	Giovanni D'Amico	Prof. Gherardo Chirici	
13	Со	ordinator	
14	Prof. Erminio Monteleone		
15			
16	Years 2018/2022		

17	Ph.D. Candidate	
18	Giovanni D'Amico	
19	giovanni.damico@unifi.it	
20		
21	Ph.D. supervisor	
22	Prof. Gherardo Chirici	
23	Dep. of Agriculture, Food, Environment and Forestry	
24	University of Florence.	
25	Via San Bonaventura, 13, 50145 Florence	
26		
27	Reviewers	
28	Prof. Anna Barbati	
29	Dept. for innovation in biological, agro-food and forestry systems	
30	University of Tuscia	
31	Via San Camillo de Lellis, 01100 Viterbo, Italy	
32	Prof. María Flor Álvarez Taboada	
33	Escuela de Ingeniería Agraria y Forestal	
34	University of León	
35	Avda. Astorga s/n 24401 Ponferrada (León), Spain	
36	Prof. Piotr Wężyk	
37	Dept. of Forest Resource Management, Faculty of Forestry	
38	University of Agriculture in Krakow	
39	Al. 29 Listopada 46, 31-425 Kraków, Poland	
40		
41	Evaluation Committee	
42	Prof. Gherardo Chirici	
43	University of Florence	
44	Prof. Anne McDonald	
45	Sophia University Graduate School of Global Environmental Studies	
46	Prof. Mats Nilsson	
47	Swedish University of Agricultural Sciences	
48	Dr. Jhon A. Parrotta	
49	US Forest Services Research & Development	
50	2	
	2	

51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	"You can't use an old map to explore a new world"
65	Albert Einstein
66	

### 68 Contents

69	Contents
70	Abstract
71	List of papers
72	Abbreviations
73	1. Introduction
74	1.1. Remote sensing technologies in forestry
75	1.2. Big data in remote sensing27
76	1.2.1. Large scale forest information layers
77	1.2.2. Deep learning approach
78	2. Background motivation and aims
79	3.1. Paper I
80	Abstract
81	1. Introduction
82	2. Materials43
83	2.1. Study area43
84	2.2. Forest mask of Italy44
85	2.3. Italian Airborne Laser Scanner surveys47
86	2.4. Italian National Forest Inventory reference data
87	2.5. Italian national grids50
88	3. Methods
89	3.1. Forest mask

90	3.2. Italian National CHM57
91	4. Results
92	4.1. Forest mask
93	4.2. Italian National CHM62
94	4.3. Forest Information System Web-GIS infrastructure64
95	5. Discussion
96	6. Conclusion72
97	References75
98	3.2. Paper II
99	Abstract
100	1. Introduction
101	2. Materials and Methods91
102	2.1. Study area91
103	2.2. Field Data93
104	2.3. Forest Masks
105	2.4. Overview of the Method102
106	2.5. Wall-to-Wall National GSV Map102
107	2.6. Accuracy Assessment of FMs104
108	2.7. Impact of FMs Accuracy on Model-Assisted GSV
109	Estimation105
110	3. Results

111	3.1. Forest Masks Accuracy Assessment108
112	3.2. GSV Model-Assisted Estimations110
113	4. Discussion114
114	5. Conclusions120
115	References123
116	3.3. Paper III
117	Abstract
118	1. Introduction
119	2. Materials139
120	2.1. Study area
121	2.2. National Forest Inventory data140
122	2.3. Remotely sensed data140
123	3. Methods144
124	3.1. Methods overview144
125	3.2. NFI plot selection
126	3.3. Nonlinear power model147
127	3.4. Stratified estimator (Approach 0)149
128	3.5. Stratified estimation with model-assisted estimation within
129	strata (Approach 1)151

130	3.6. Stratified estimation with model-assisted estimation within
131	strata using pseudo-plots for model construction (Approach 2)
132	
133	3.7. Relative efficiency153
134	4. Results and discussion154
135	4.1. Nonlinear power model154
136	4.2. Stratified estimator with the simple expansion estimator
137	within strata (Approach 0)156
138	4.3. Stratified estimation with model-assisted estimation within
139	strata (Approach 1)156
140	4.4 Stratified estimation with model-assisted estimation within
141	strata using pseudo-plots for model construction (Approach 2)
142	
143	4.5. Relative Efficiency
144	4.6. Summary162
145	5. Conclusion165
146	6. Annex167
147	References170
148	3.4. Paper IV
149	Abstract178
150	1. Introduction179
151	2. Materials and Methods

152	2.1. Survey area
153	2.2. Sentinel-2 imagery
154	2.3. Reference Dataset
155	2.4. Moving Window approach
156	2.5. Fully Connected Neural Network
157	2.6. Logistic Regression
158	2.7. Moving Window calibration and performance assessment
159	
160	2.8. 2018 mapping update197
161	3. Results
162	4. Discussion
163	4.1. Summary and aim of the research
164	4.2. Sentinel-2 pre-processing
165	4.3. Classification algorithm205
166	4.4. Models accuracy208
167	4.5. Procedure replicability
168	5. Conclusions210
169	Acknowledgment211
170	References
171	4. Conclusion
172	Reference

173	Other publications and contributions	235
174	1. Paper	235
175	2. Conference talks and seminars	236
176	Acknowledgements	239
177		
178		

# Application of big data analytics in remote sensing supporting sustainable forest management

#### 181 Abstract

182 Sustainable forest management requires detailed forest information 183 for planning accurate treatments. The information is expected to be 184 accurate enough and preferably obtained at a low cost and with 185 periodic updates. Such spatial scale information is nowadays 186 provided by remote sensing data. On the one hand, the development 187 and use of aerial laser scanning for estimating forest variables has 188 been a game-changer in recent decades for forest management. On 189 the other hand, satellite remote sensing technologies, generated a constant flow of data from different platforms, in different formats 190 and with different purposes. Combined with this ongoing remote 191 192 sensing data stream, the development of computer technology has 193 provided forest management with many new tools for data capture, data representation, data visualization, and management planning 194 195 applications. Today, new computing power makes it possible to 196 tackle the complex problem of managing and processing big data 197 from remote sensing with new strategies that have revolutionized the 198 way of understanding the use of these data sources.

199 This thesis is aimed at assessing big data analytics for practical cases 200 of forest monitoring especially in the Italian context, where large-201 scale aggregated forest remote sensing data have always been a 202 structural lack. Four main studies were covered in the thesis. Study I

involved the review and aggregation of remotely sensed forestry data 203 at the national scale. The available Italian airborne laser scanning 204 205 data were aggregated to develop a consistent mosaic of canopy heigh 206 model, while different local forest maps were used to develop for the 207 first time a high-resolution forest mask of Italy which was validated 208 against the official statistics of the Italian National Forest Inventory. An online geographic forest information system was implemented to 209 store and facilitate the access and analysis of both spatial datasets. 210 211 The two information layers were explored in operational cases, 212 through the integration of remote sensing and inventory data in 213 studies II and III. In the former, the forest mask produced mosaicking the Italian regional local forest maps was compared with four other 214 215 forest masks available for the entire area of Italy to examine their effects on the estimation of growing stock volume and to clarify 216 which product is best suited for this purpose. Non-forest pixels in 217 each forest mask were removed from a national wall-to-wall growing 218 stock volume map constructed using inventory and remote sensing 219 data. The estimated Growing stock volume from each mask was 220 221 compared with the official national forest inventory estimates. In the 222 III study, airborne laser scanning coverage and the forest mask were used in combination with Landsat spectral data for large-scale 223 224 volume estimation. Estimates were performed considering different 225 proportions between airborne laser scanning and Landsat coverage. 226 The integration between satellite spectral data and airborne laser 227 scanning information is particularly critical in countries like Italy, 228 where wall-to-wall airborne laser scanning coverage is still lacking.

In the last study (IV), Sentinel-2 multitemporal data were used to 229 identify poplar plantations, which are the primary source of Italian 230 231 industrial timber. The study area was the dynamic agricultural area 232 of Pianura Padana where most of the Italian poplar plantations are concentrated. The capabilities of the Sentinel-2 data were integrated 233 with a deep learning approach that provided better results compared 234 to traditional logistic regression. The map we produced can allow the 235 poplar plantation monitoring, which requires frequent updating, not 236 237 feasible with traditional forest inventories.

In so doing, these studies, aimed at enhancing knowledge about
missing information layers at the national scale, attempting to close
the gaps underlined by previous studies.

#### 243 List of papers

#### 244 Paper I

- 245 D'Amico G., Vangi E., Francini S., Giannetti F., Nicolaci A.,
- 246 Travaglini D., Massai L., Giambastiani Y., Terranova C., Chirici G.
- 247 (2021). Are we ready for a National Forest Information System?
- 248 State of the art of forest maps and airborne laser scanning data
- 249 availability in Italy. iForest 14: 144-154.
- 250 https://doi.org/10.3832/ifor3648-014

#### 251 Paper II

- 252 Vangi E., D'Amico G., Francini S., Giannetti F., Lasserre B.,
- 253 Marchetti M., McRoberts RE., Chirici G. (2021). The Effect of
- 254 Forest Mask Quality in the Wall-to-Wall Estimation of Growing
- 255 Stock Volume. Remote Sensing. 13(5):1038.
- 256 https://doi.org/10.3390/rs13051038

#### 257 Paper III

- 258 D'Amico G., McRoberts R.E., Giannetti F., Vangi E., Francini S.,
- 259 Chirici G. Effect of LiDAR coverage and field plot data numerosity
- 260 for the forest growing stock volume estimation. Submitted to
- 261 European Journal of Remote Sensing on December 6<sup>th</sup>, 2021.

#### 262 Paper IV

- 263 D'Amico G., Francini S., Giannetti F., Vangi E., Travaglini D.,
- 264 Chianucci F., Mattioli W., Grotti M., Puletti N., Corona P., Chirici
- 265 G. (2021). A deep learning approach for automatic mapping of poplar
- 266 plantations using Sentinel-2 imagery. GIScience & Remote Sensing.
- 267 https://doi.org/10.1080/15481603.2021.1988427.

### 268 Abbreviations

3D	Three dimensional		
ABA	Area-based approach		
ALOS	Advanced Land Observing Satellite		
ALS	Airborne laser scanning		
C&I	Criteria and Indicators		
CHM	Canopy height model		
CLMS	Copernicus Land Monitoring Service		
DL	Deep learning		
EC	European commission		
EEA	European environment agency		
ESA	European Space Agency		
FAO	Food and Agriculture Organization of the United Nations		
FIS	Forest information system		
FRA	Forest resource assessment		
GMES	Global Monitoring for Environment and Security programme		
GSV	Growing stock volume		
HRL	High-resolution layer		
IIASA	International Institute for Applied Systems Analysis		
JAXA	Japanese Aerospace Exploration Agency		
LiDAR	Light detection and ranging		
NFI	National forest inventory		
NFN	National forest mask		
NN	Artificial neural network		
PALSAR	Phased Array type L-band Synthetic Aperture Radar		
	Reducing emissions from deforestation and forest		
KEDD+	degradation projects		
S2	Sentinel-2		
SAR	Synthetic Aperture Radars		
SFM	Sustainable Forest Management		

#### 271 **1. Introduction**

272 Forests are complex environmental systems, characterized by high biological and genetic biodiversity (Dinerstein et al., 1995), 273 274 generating multifunctional services to satisfy social, cultural, environmental, and economic demands (FOREST EUROPE, 2020; 275 276 O'Farrell and Anderson, 2010). In this context of multifunctional 277 services, forest management planning aims to produce timber, maintaining biodiversity, and developing other services required in 278 279 specific situations, where all are of equal importance. To ensure this balance in strategic forest planning many influencing parameters 280 281 must be considered. For this purpose, decision support systems, 282 based on sustainable forest management (SFM) principles, have to 283 be developed aiming at maintaining and preserving the capacity to 284 generate ecosystem services for future generations. In particular, 285 SFM aims to promote better practices over time and foster the 286 development of healthier and more productive forests, taking into 287 account the environmental, economic, social, cultural, and spiritual 288 needs of the full range of stakeholder groups in the countries 289 involved. Considering the difficulty of quantifying and monitoring these aspects at different local and temporal scales, specific criteria 290 291 and indicators (C&I) have been developed. Among the various sets 292 of C&I developed and used in the world (FOREST EUROPE, 2020; ITTO/FAO, 1995, Montreal Process 1995; FAO, 2020), the pan-293 European C&I represents the consensus achieved by European 294 295 countries on the most important aspects of SFM and provide guidance for developing policies and help assess progress on SFM.
It was internationally recognized that C&I are tools for describing,
monitoring, and evaluating national trends in forest condition and
management while also providing an implicit definition of what SFM
means.

The structure of the set is formed by an overarching policy framework of the set, named "Forest policy and governance" (5 indicators), followed by the set of indicators under the six pan-European criteria for SFM, comprising a qualitative part (6 indicators), aligning the specific policies and instruments under each Criterion, and the related quantitative indicators (34 indicators) (Table 1).

Table 1. Pan-European Sustainable Forest Management Criteria &
 Indicators (FOREST EUROPE, 2020)

No.	Indicator	Full Text	
Fore	Forest policy and governance		
1	National Forest Programmes or equivalent		
2	Institutional frameworks		
3	Legal/regulatory framework: National (and/or sub-national) and international commitments		
4	Financial and economic instruments		
5	Information and comunication		
C 1: Forest Resources and Contribution on Global Carbon Cycles			
1.1	Forest area	Area of forest and other wooded land, classified by forest type and by availability for wood supply, and share of forest and other wooded land in total land area	
1.2	Growing stock	Growing stock on forest and other wooded land, classified by forest type and by availability for wood supply	

1.3	Age structure and/or diameter distribution	Age structure and/or diameter distribution of forest and other wooded land, classified by availability for wood supply	
1.4	Carbon stock	Carbon stock and carbon stock changes in forest biomass, forest soils and in harvested wood products	
C 2:	<b>Maintenance of Fo</b>	rest Ecosystem Health and Vitality	
2.1	Deposition and concentration of air pollutants	Deposition and concentration of air pollutants on forest and other wooded land	
2.2	Soil condition	Chemical soil properties (pH, CEC, C/N, organic C, base saturation) on forest and other wooded land related to soil acidity and eutrophication, classified by main soil types	
2.3	Defoliation	Defoliation of one or more main tree species on forest and other wooded land in each of the defoliation classes	
2.4	Forest damage	Forest and other wooded land with damage, classified by primary damaging agent (abiotic, biotic and human-induced)	
2.5	Forest land degradation	Trends in forest land degradation	
C 3:	C 3: Productive Functions of Forests (Wood and Non-Wood)		
3.1	Increment and fellings	Balance between net annual increment and annual fellings of wood on forest available for wood supply	
3.2	Roundwood	Quantity and market value of roundwood	
3.3	Non-wood goods	Quantity and market value of non-wood goods from forest and other wooded land	
3.4	Services	Value of marketed services on forest and other wooded land	
C 4: Biological Diversity in Forest Ecosystems			
4.1	Diversity of tree species	Area of forest and other wooded land, classified by number of tree species occurring	
4.2	Regeneration	Total forest area by stand origin and area of annual forest regeneration and expansion	
4.3	Naturalness	Area of forest and other wooded land by class of naturalness	
4.4	Introduced tree species	Area of forest and other wooded land dominated by introduced tree species	

4.5	Deadwood	Volume of standing deadwood and of lying deadwood on forest and other wooded land	
4.6	Genetic resources	Area managed for conservation and utilisation of forest tree genetic resources (in situ and ex situ genetic conservation) and area managed for seed production	
4.7	Forest fragmentation	Area of continuous forest and of patches of forest separated by non-forest lands	
4.8	Threatened forest species	Number of threatened forest species, classified according to IUCN Red List categories, in relation to total number of forest species	
4.9	Protected forests	Area of forest and other wooded land protected to conserve biodiversity, landscapes and specific natural elements, according to MCPFE categories	
4.1	Common forest bird species	Occurrence of common breeding bird species related to forest ecosystems	
C 5: Protective Functions in Forest Management			
5	Protective forests: 5.1 soil, water and other ecosystem functions; 5.2 infrastructure and managed natural resources	Area of forest and other wooded land designated to prevent soil erosion, preserve water resources, maintain other protective functions, protect infrastructure and managed natural resources against natural hazards	
C 6: Socioeconomic functions and conditions			
6.1	Forest holdings	Number of forest holdings, classified by ownership categories and size classes	
6.2	Contribution of forest sector to GDP	Contribution of forestry and manufacturing of wood and paper products to gross domestic product	
6.3	Net revenue	Net revenue of forest enterprises	
6.4	Investments in forest and forestry	Total public and private investments in forest and forestry	
6.5	Forest sector workforce	Number of persons employed and labour input in the forest sector, classified by gender and age group, education and job characteristics	

6.6	Occupational safety and health	Frequency of occupational accidents and occupational diseases in forestry
6.7	Wood consumption	Consumption per head of wood and products derived from wood
6.8	Trade in wood	Imports and exports of wood and products derived from wood
6.9	Wood energy	Share of wood energy in total primary energy supply, classified by origin of wood
6.1	Recreation in forest	The use of forests and other wooded land for recreation in terms of right of access, provision of facilities and intensity of use

In support of SFM, a thorough understanding of forest and up-to-date forest data are required to assess the composition, structure, and distribution of forest vegetation that, in turn, can be used as base information for management decisions developing effective forest plans that span across a range of spatial and temporal scales (Wulder et al., 2008).

318 Accordingly, forest information is essential for multiple purposes, including national and international forest monitoring programs, 319 320 reporting activities such as in the context of international agreements on forest resource assessment (e.g., Kyoto Protocol) (Corona et al, 321 322 2011; FAO, 2020), restoration programs (e.g., Reducing emissions from deforestation and forest degradation projects - REDD+) 323 (UNCCD, 2015; Smith et al, 2016), biodiversity monitoring (Chirici 324 325 et al., 2012), and the aforementioned local-scale management to improve decision-making, silvicultural measures, and harvesting and 326 327 conservation activities.

328 Typically, this type of data is collected using sample-based National

of forest parameters for large areas such as countries or regions. The 330 most common forest variables needed to assess SFM indicators, as 331 required by national and international agreements at the national and 332 regional levels, are the following: forest area, growing stock volume, 333 334 biomass, and increments (Brosofske et al., 2014; Kangas et al., 335 2018). These aggregated statistics are essential to evaluate the state 336 of forests, but also to support decision making and to develop strategies at different scales and different time horizons. Recently, to 337 338 increase efficiency and accuracy, both in terms of time and cost, 339 remote sensing data have been an NFI crucial component.

#### **1.1. Remote sensing technologies in forestry**

The term "remote sensing", introduced in the 1960s, describes the 341 342 acquisition of information about an object or phenomenon without physical contact with the object and is thus in contrast to in situ 343 344 observation. Remote sensing technologies, which nowadays provide high-quality geospatial information, are considered a key to improve 345 346 repeatable measurements, actions, and processes in forestry 347 (Holopainen et al., 2014; Kovácsová and Antalová, 2010). Many authors have already pointed out that remote sensing technologies 348 are essential for monitoring, quantifying, and mapping forest 349 variables (Hansen et al. 2013, Waser et al. 2017, Kangas et al. 2018, 350 Chirici et al. 2020). 351

## In forestry applications, the availability of remotely sensed data hassteadily increased. Spectral data are collected in many forms and

scales by satellite, aircraft, and drones, with a spatial resolution 354 ranging from tens of meters to a few centimeters. Some data are 355 356 collected daily or at regular intervals across the whole globe, while 357 other data may be collected on an as-needed basis. In addition, 358 structural or three-dimensional (3D) information is gathered from 359 laser, radar, and optical data, allowing forests to be measured in ways 360 that were not previously possible. Among these technologies, Light Detection And Ranging (LiDAR) data collected by airplane or 361 362 helicopter platforms (i.e., Airborne Laser Scanning, ALS), is 363 considered the most useful technology to map forest ecosystems 364 (Figure 1).

ALS data has the ability to collect highly detailed data of large areas, 365 366 giving information on ground elevation and detailed characterization of forests (Holopainen et al., 2014; Hyyppä et al., 2008), on the basis 367 of laser pulses, it is possible to model and detect the 3D structure of 368 forests and to easily estimate biophysical forest variables (e.g. tree 369 heights, vertical structure, growing stock volume, carbon stock) 370 (Dubayah & Drake, 2000; Babcock et al., 2015). In the last decades, 371 372 many studies demonstrated the utility of ALS to monitor forest resources (Nelson, 2013; Kangas et al., 2018), biodiversity (Corona 373 et al., 2011; Lefsky et al., 2002; Lim et al., 2003; Mura et al., 2015; 374 375 Valbuena et al., 2016, 2013; Wulder et al., 2008) to characterize 376 wildlife habitats, and thoroughly in the context of local (Bottalico et al., 2017) and NFI (McRoberts et al., 2013; Næsset, 2007; Næsset et 377 378 al., 2004). Given its proven capabilities in mapping forest variables, the use of ALS data is increasing rapidly worldwide (Zolkos et al., 379

- 380 2013), and in many countries, ALS data are specifically acquired to
- 381 support forest inventory programs.



Figure 1. Airborne laser scanning dataset of a spruce forest with high pulse
density (source: McRoberts et al., 2010b).

386 In traditional NFIs, remote sensing is initially used to stratify 387 sampling units according to their land uses, commonly through the 388 use of high-resolution imagery (McRoberts et al., 2009, McRoberts 389 et al., 2010a,b; Corona, 2010). Countries with a long NFI tradition 390 such as Sweden, Finland, Denmark (Næsset et al., 2004; Nord-391 Larsen and Schumacher, 2012; Tomppo et al., 2008), Canada 392 (Boudreau et al., 2008; Matasci et al., 2018), Austria (Hollaus et al., 2009) and Switzerland (Waser et al., 2017, 2015), forest inventories 393 394 are now integrated with remote sensing technology to construct wallto-wall spatial estimates of forest variables (McRoberts and Tomppo, 395 2007). In operational wall-to-wall forest inventories, a two-stage 396

procedure using ALS data and field plots, i.e. an area-based approach 397 (ABA, Næsset, 2002), has become particularly common and used to 398 399 estimate forest variables such as growing stock volume (Nilsson et 400 al., 2017; Nord-Larsen and Schumacher, 2012), biomass (Nord-401 Larsen and Schumacher, 2012), forest cover (Waser et al., 2015), or 402 forest changes (Næsset et al., 2013). Moreover, to support forest 403 management, spatial data produced by NFIs are commonly implemented in geographic Forest Information Systems (FIS) which 404 405 allow forest managers, forest owners, and government authorities to 406 query forest data through online web-based systems. Examples are 407 available for Norway (https://kilden.nibio.no/), Sweden 408 (https://kartor.skogsstyrelsen.se/kartor/?startapp=skogligagrunddata 409 ), (https://kartta.paikkatietoikkuna.fi/), Spain Finland (http://lidarrioja.agrestaweb.org/#!/map) 410 France or 411 (https://www.geoportail.gouv.fr/carte).

Despite the significant need for wall-to-wall forest maps, especially 412 413 in Mediterranean areas, where forests are considered more 414 vulnerable to climate change scenarios and natural and anthropogenic disturbances (Giannetti et al., 2021; Ogaya & 415 416 Peñuelas, 2021), several critical data needed to accurately estimate 417 forest variables are still missing. The Italian case is emblematic, 418 where the NFI program does not provide wall-to-wall maps as the 419 Enhanced Forest Inventories do (White et al., 2016), but only aggregates estimates of forest variables over large geographic 420 421 regions. Furthermore, a national overview of ALS datasets available 422 in Italy, and an homogeneous national forest mapping process is still

missing, although multiple mapping projects have been carried out at 423 a local scale. Such information is crucial to integrate ALS data with 424 425 other data, such as field surveys conducted by the NFI or to plan 426 future ALS acquisitions. The availability of a forest mask is an essential prerequisite for spatial estimates of forest variables, both to 427 428 limit the establishment of field plots and to determine the area where 429 to apply models for forest variable estimations. Consequently, several countries developed independently their forest maps, such as 430 Sweden (Nilsson et al. 2017), Norway (Naesset 2007), Finland 431 (Maltamo et al. 2014), Switzerland (Waser et al. 2017), Spain 432 433 (Alberdi et al. 2017), United Kingdom (Smith et al. 2010), USA (McRoberts et al. 2005), France (Garnier et al. 2019). 434

435 In recent years, several improvements opened new prospects in remote sensing Earth observation. The main changes concern: i. new 436 satellite mission, ii. more satellites in orbit per mission; iii. the 437 increased spectral, spatial, and temporal resolution of satellites, and 438 iv. the free-and-open data policy of Earth observation programs. 439 Crucial in the increase of remote sensing data are the Sentinel 440 441 missions in the framework of the Copernicus program, an initiative 442 led by the European Commission (EC) in collaboration with the European Space Agency (ESA), previously known as Global 443 444 Monitoring for Environment and Security programme GMES. 445 Fundamental to forest monitoring is the Sentinel-2 (S2) mission given systematic global acquisitions of high-resolution multispectral 446 imagery at high revisit rates. The S2 mission was developed to 447 provide multispectral imagery in continuity with those of the USGS 448

449 Landsat Thematic Mapper instrument. At the same time, the Landsat program, which provides the longest continuous spatially based 450 451 record of the Earth's landmass in existence (Landsat 1 launched in 452 1972), since 2008 adopted an open data policy (Woodcock et al., 453 2008). Additionally, the Landsat program continued its development, with the launch of the new Landsat 9 satellite on September 16<sup>th</sup>, 454 455 2021. All these aspects guide big data availability and the need to 456 develop new tools able to process such large datasets.

#### 457 **1.2. Big data in remote sensing**

Big data refers to a collection of data sets so large and complex that 458 it is difficult to employ traditional data processing algorithms and 459 models (Manyika et al., 2011). Challenges include the acquisition, 460 storage, searching, sharing, transfer, analysis, and visualization of the 461 462 data. In short, big data can be reported as advanced analysis techniques on large volumes of data. However, as technology 463 advances over time, the size of datasets that qualify as big data will 464 grow, regardless of size in terms of terabits. 465

466 Remote sensing big data computing is a challenging task due to the 467 extensive nature of the analysis, combined with the large amount of 468 data handled (Ma et al. 2015). Big Data Analytics in the earth 469 observation field relies on processing, analyzing, and merging 470 multiple images with other data sources, in order to create previously 471 unavailable information that requires heavy computing power. In the 472 meantime, supercomputers, and high-performance computing 473 systems, frequently provided by cloud platforms universally474 available, are becoming abundant (Gorelik et al., 2017).

The unprecedented proliferation of data, together with high
computing powers, allowed the development of numerous largescale forest information layers, as well as enabling new machine
learning approaches.

479

#### 1.2.1. Large scale forest information layers

The big data availability led to an exponential increase in the number 480 of forest maps made available at different spatial scales for global or 481 482 continental forest resources, produced independently by different 483 agencies. For instance, Italian information about forest area can be 484 estimated from any of several forest/non-forest maps (masks), that are all potentially referring to the FAO Forest Resource Assessment 485 486 (FRA) forest definition (FAO, 2020), including i. the CORINE Land Cover project (Büttner et al., 2004), started in 1990 and updated in 487 488 2000, 2006, 2012, and 2018 to monitor land-use changes in the 39 participating countries carried out by the European Environmental 489 490 Agency (EEA, 2007); ii. in the framework of Copernicus Land 491 Monitoring Service (CLMS) coordinated by ESA a forest/non-forest mask in grid format covering entire Europe for the years 2012 and 492 493 2015 is achievable from the forest type, available among the so-494 called High-Resolution Layers (HRL), in which the main input sources of the forest layers are S2 and Landsat 8 time series, 495 complemented by SPOT-5 and Resource- Sat-2 satellite data 496 (Langanke, 2017); iii. the International Institute for Applied Systems 497

Analysis (IIASA) constructed a global forest mask for 2000 by 498 combining through a hybrid approach multiple data sets, calibrated 499 500 with FAO FRA country statistics at the national level 501 (Schepaschenko, 2015), and iv. the Japanese Aerospace Exploration 502 Agency (JAXA), that for the years 2007, 2008, 2009, 2010, and 2015 provides a forest/non-forest mask in a grid format with a 25 m 503 504 resolution for the entire globe, by automatic processing of multipolarization backscatter signals acquired by the two Synthetic 505 506 Aperture Radars (SAR), PALSAR and PALSAR 2 (Phased Array 507 type L-band Synthetic Aperture Radar), which are mounted on the 508 two satellites ALOS and ALOS-2 (Advanced Land Observing Satellite) (JAXA, 2016). 509

510 The above maps, developed according to big data analysis approaches, were designed for different purposes. Therefore, each 511 512 map has specific characteristics, useful for monitoring forest resources on a global or continental scale (Hansen et al. 2013). 513 514 However, they can be affected by consistent errors at national or regional level (Giannetti et al. 2020). Indeed, despite individual 515 516 weaknesses and strengths, spatial differences among these products 517 are relevant at the national scale and can lead to substantial variations 518 in their accuracies (Schepaschenko, 2015; Seebach, 2012). Creating 519 doubts about which is best suited for multiple purposes such as to 520 infer forest statistics in NFIs (Di Biase et al. 2018), to assess forest variables at national scale, and supporting forest owners in planning 521 522 silvicultural interventions at local scale (Kangas et al. 2018), to 523 quantify forest ecosystems services (Vizzarri et al. 2017) or to 524 support precision forestry (Corona et al. 2017). Anyway, only a few 525 studies analyzed the effects of using different forest masks on the 526 uncertainty of forest variables estimates. Furthermore, no study has 527 examined in the Mediterranean environment the impacts of the 528 accuracies of different forest masks on the estimation of growing 529 stock volume (GSV).

530

#### 1.2.2. Deep learning approach

531 The advent of more frequent and more detailed remotely sensed data acquisition, such as the S2 data, with high revisit time (5 days at the 532 equator with two satellites under cloud-free conditions which 533 534 resulted in 2-3 days at mid-latitudes), spectral (13 spectral bands) and 535 spatial resolution (10 to 60 m depending on the wavelength), offers unprecedented perspectives for a wide range of applications in 536 environment and agricultural field (Kussul et al., 2017), led also to 537 the beneficial use of deep learning (DL) approaches (Zhu et al., 2017; 538 539 Ma et al., 2019).

DL is a powerful machine learning technique that obtains great 540 success in several practical applications and attracted interest in 541 542 academic and industrial communities (LeCun et al., 2015). The core idea behind DL is to simulate the human ability to deal with big data 543 544 problems, using all data available to learn and process information to 545 provide the output. In particular, DL represents the procedure of training the artificial neural networks (NN) that are inspired to 546 biological ones where neurons connections and signals strength 547 548 control all brain processes.

549 Three layers of "neurons" called nodes, the input, the hidden, and the output layers compose the NN basic structure. Information flows 550 551 from the input layer, through the hidden layer (one or more than one) to the output layer and then out. The NN parameters are associated 552 with connections and nodes. Each connection between nodes has a 553 554 number associated with it called connection weight, each node has a 555 number, and a formula associated called respectively threshold value and activation function. The NN learning process implicates 556 557 reiterated adjustment of weights and threshold until the produced 558 outputs are as close as possible than expected outputs.

559 Many studies have explored DL for remote sensing tasks, using 560 various NN architectures that have demonstrated good capabilities, 561 mainly attributed to automatic extraction of meaningful features, eliminating the need to identify case-specific characteristics 562 563 (Tsagkatakis, 2019). DL approaches, coupled with S2 data, allowed the identification of land uses previously difficult to separate 564 spectrally. An example of this is the identification of poplar 565 566 plantations, featured by short rotation, high spatial and temporal 567 variability, as well as being localized, at least in Italy, in agricultural environments, with large interannual variations. 568

569 Poplar plantations mapping appeared particularly important to 570 support management and to increase knowledge about Italian poplar 571 production, the primary domestic source of wood for industrial use. 572 The information needed to support poplar plantations management is 573 increasingly complex due to their specific features and to the 574 expansion through the years due to the increased market value of

poplar timber (Corona et al., 2018). In these conditions, conventional 575 national forest inventories, with a typical time frame of 10 years are 576 577 not able to produce such rapid updates. Such limitations may be potentially overcome by adopting robust DL automatic classification 578 579 methods of remotely sensed data, which at the same time are 580 objective and cheaper than traditional approaches and can be 581 repeated to produce near-real-time information due to the vast availability of imagery (Francini et al., 2020, Vaglio et al., 2021). 582

#### 583 **2. Background motivation and aims**

The increased availability of remotely sensed data from multiple 584 sources and the Italian lack of aggregate data and large-scale 585 spatialized information inspired the research work of this thesis. This 586 587 research addresses the synergistic integration of multiple remotely 588 sensed data sources to develop and evaluate the accuracy of new 589 forest information layers currently unavailable at a large scale and 590 facilitate its use in supporting sustainable forest management 591 planning. In this thesis, remote sensing big data analytics are used for the aim of reducing gaps in information layers needed to support 592 593 sustainable forest management. In particular, attention was focused on developing new information layers and procedures based on big 594 595 data.

596 The specific objectives of the papers are:

- to develop two missing information layers at the national level 597 such as the canopy height model (CHM) derived by ALS and the 598 forest mask, making them freely available in a FIS (Paper I); 599 600 to assess the impacts of the accuracies of different national forest 601 masks on the estimation of GSV based on the integration of field information and remotely sensed data (Paper II); 602 603 to assess the impact of multiple sources of information such as • 604 measured Forest Inventory data, LiDAR metrics, and Landsat 605 indices, available in different proportions, for estimating GSV at the national scale (Paper III); 606 to assess a new DL approach based on S2 multitemporal images 607 to identify and map poplar plantations in northern Italy (Paper 608 609 IV).
- 610

#### 612 **3.1. Paper I**

613 Are we ready for a National Forest Information System? State of

614 the art of forest maps and airborne laser scanning data615 availability in Italy

- 616 D'Amico G.<sup>1,\*</sup>, Vangi E.<sup>1,2</sup>, Francini S.<sup>1,2,3</sup>, Giannetti F.<sup>1,4,5</sup>, Nicolaci
- 617 A.<sup>6</sup>, Travaglini D.<sup>1</sup>, Massai L.<sup>4,5</sup>, Giambastiani Y.<sup>4,5,7</sup>, Terranova C.<sup>8</sup>,
  618 Chirici G.<sup>1,5</sup>
- 619
- <sup>1</sup> Department of Agriculture, Food, Environment and Forestry
  (DAGRI), University of Florence. Florence 50145, Italy
- <sup>2</sup> Department of Bioscience and Territory (DiBT), University of
  Molise. Pesche 86090, (IS), Italy
- 624 <sup>3</sup> Department of Innovation in Biological, Agro-Food and Forest
- 625 System (DIBAF), University of Tuscia. Viterbo 01100, (VT), Italy
- <sup>4</sup> Bluebiloba startup Innovativa S.R.L. Florence 50126, Italy
- <sup>5</sup> ForTech, University of Florence joint laboratory, Florence, Italy
- <sup>6</sup>Department of Computer Engineering, Modeling, Electronics, and
- 629 Systems Science (DIMES), University of Calabria. Rende, (CS),
- 630 Italy
- 631 <sup>7</sup> LAMMA Consortium Environmental Modelling and Monitoring
- 632 Laboratory for Sustainable Development, Florence, Italy
- 633 <sup>8</sup> Geoportale Nazionale, Italian Environment Ministry. Rome, Italy
- 634 \* corresponding author giovanni.damico@unifi.it
- 635 iForest Biogeosciences and Forestry
- 636 https://doi.org/10.3832/ifor3648-014
#### 637 Abstract

Forest planning, forest management, and forest policy require
updated, reliable, and harmonized spatial datasets. In Italy a national
geographic Forest Information System (FIS) designed to store and to
facilitate accessing and analysis of spatial datasets is still missing.

Within the different information layers which are useful to start
populating a FIS, two of them are essential for their multiple use in
the assessment of forest resources: *i*) forest mapping, and *ii*) data
from Airborne Laser Scanning (ALS). Both of them are not available
wall-to-wall in Italy even if different potentially useful sources of
information for their implementation exist already.

The objectives of this work are: (i) to review forest maps and ALS 648 data availability in Italy; (ii) to develop for the first time in Italy a 649 high resolution forest mask that was validated against the official 650 statistics of the Italian National Forest Inventory; (iii) to develop the 651 652 first mosaic of all the main ALS data available in Italy producing a 653 consistent Canopy Height Model (CHM); (iv) we finally developed 654 for demonstration scope an on-line geographic FIS where we provide 655 free access to both the layers from (ii) and (iii).

The total area of forest and other wooded lands computed from the forest mask we created was 102,608.82 km<sup>2</sup> (34% of the Italian territory), 1.9% less than the NFI benchmark estimate, this map resulted for the moment the best forest mask available wall-to-wall in Italy. We also found that only the 63% of the Italian territory (the 60% of the forest area) is covered by ALS data. These results underline once more the urgent need for a national strategy tocomplete the availability of forest data in Italy.

664

665 Keywords: National datasets, Forest inventory, Forest monitoring,

666 Forest mask, Airborne Laser Scanning, LiDAR.

667

# 668 **1. Introduction**

669 Forest mapping is an important source of information for assessing 670 woodland resources and needs to be a key issue for any National 671 Forest Inventory (NFI) programme (Waser et al. 2017). Nowadays 672 global and nationwide wall-to-wall raster-type forest resources maps, 673 based on either satellite images, laser scanning, aerial ortomosaic and photogrammetric point cloud data are considered essential to monitor 674 and to quantify forest variables (Hansen et al. 2013, Waser et al. 675 2017, Kangas et al. 2018, Chirici et al. 2020). In fact, forest maps are 676 produced on the basis of remote sensing technologies at different 677 spatial scales for global or continental forest resources monitoring 678 already. Here below we provide a short review of the most important 679 680 and recent efforts for forest mapping at European or Global scale 681 level.

682 Copernicus, the European programme for Earth observation 683 (https://www.copernicus.eu), developed several layers potentially 684 useful for forest monitoring, in particular we refer to the European 685 Forest High Resolution Layers (HRL), and more specifically: *i*) the

Tree cover density (TCD) expressed in percent of tree cover; *ii*) the 686 Dominant leaf type (DLT) based on the domination of broadleaved 687 or coniferous species, and *iii*) the Forest type products (FTY) a forest 688 mask which mimic as close as possible the FAO forest definition. 689 690 These layers, developed primarily with Sentinel-2 time series for the 691 year 2018, and complemented for the years 2012 and 2015 by 692 Landsat 8, SPOT-5 and ResourceSat-2 satellite data (Langanke 693 2017), are available with resolutions ranging between 10 and 100 694 meters. The Copernicus Global Land Service has also recently 695 released the 2015 global land cover map at 100 m resolution (Buchhorn et al. 2019), updating the harmonized global land cover 696 classification for the year 2000 based on SPOT4 images, originally 697 698 produced by the Global Vegetation Monitoring unit of the Join Research Centre of the European Commission. 699

The Japan Aerospace Exploration Agency (JAXA) produced for the
reference years 2007 and 2009 a global forest/non forest map based
on the classification of ALOS and ALOS-2 satellite radar images
with a resolution of 10 meters and a declared accuracy of 84%
compared to the ground base data set (JAXA 2016).

The World Resources Institute, in the framework of Global Forest
Watch developed an online forest monitoring data set based on the
analysis of Landsat multitemporal series mapping global tree cover
density for the reference year 2010 and a forest gain/loss map for the
period 2001-2019 (Hansen et al. 2013).

710 These maps are considered useful to monitor forest resources at 711 global or continental scales (Hansen et al. 2013), however at local 712 level (i.e., national, regional) they can be affected by large errors (Giannetti et al. 2020). For this reason none of the existing forest 713 714 maps implemented at continental or global level are considered 715 reliable for operational purposes at National level. Consequentially 716 several NFIs developed independently their maps. See for instance 717 the examples for Sweden (Nilsson et al. 2017), Norway (Næsset 718 2007), Finland (Maltamo et al. 2014), Switzerland (Waser et al. 2017), Spain (Alberdi et al. 2017), United Kingdom (Smith et al. 719 720 2010), USA (McRoberts et al. 2005), France (Garnier et al. 2019). 721 These maps are considered essential to infer forest statistics in NFIs 722 (Di Biase et al. 2018), to assess forest variables at national scale and 723 at forest management scale supporting forest owners in their strategic 724 planning and silvicultural measures (Kangas et al. 2018), to quantify forest ecosystems services (Vizzarri et al. 2017), or to support 725 726 precision forestry (Corona et al. 2017). Usually, forest resources 727 maps developed in the context of NFIs have a scale congruent with the size of the sampling units used to acquire the information in the 728 field, in order to reduce the costs of management activities for forest 729 730 owners (Kangas et al. 2018).

The spatial data produced by modern NFIs are nowadays routinely developed in the framework of geographic Forest Information Systems (FIS) to query forest data through on-line web-based system that can be used by forest managers, forest owners and government authorities to support forest management or planning. Examples are available for e.g. Norway (https://kilden.nibio.no/), Sweden (https://kartor.skogsstyrelsen.se/kartor/?startapp=skogligagrunddata 738 ), Finland (https://kartta.paikkatietoikkuna.fi/?lang=en), Spain
739 (http://lidarrioja.agrestaweb.org/#!/map) or France
740 (https://www.geoportail.gouv.fr/carte).

In Italy the NFI program is still designed in a more traditional way,
according to the classification from White et al. (2016), it is not yet
an Enhanced Forest Inventory since it only provides estimates of
forest variables aggregated for large geographical regions and not
wall-to-wall maps.

746 As a result, in Italy a forest mapping process within the NFI program is still missing, even if multiple projects were carried out locally. 747 748 However, in 2018 a new National Forest law was adopted by the Italian Parliament stating clearly that to set up a national forest 749 750 strategy a national high resolution forest map is essential. For this reason, it is very relevant to start a first recognition of existing forest 751 maps in order to understand their consistency in terms of forest 752 definitions used and nomenclature systems adopted. At least to 753 understand if these maps can be useful to support the creation of a 754 forest/non-forest map (forest mask) congruent with the official forest 755 756 area estimations provided by the NFI.

To create wall-to-wall maps and small area estimations of forest
variables, Remotely Sensed (RS) data are essential (Chirici et al.
2020). For example to estimate growing stock volume (Saarela et al.
2016), biomass (Næsset et al. 2011), forest structural variables and
diversity indices (Mura et al. 2016).

Within the different types of RS technologies, Airborne LaserScanner (ALS) emerged as the most viable to derive such maps and

764 to support the development and parameterization of models for a broad range of information needed to enhance NFIs (Maltamo et al. 765 766 2014, White et al. 2016). The advantages of ALS in mapping forest 767 variables is well documented, in the context of NFIs (Næsset 2007, 768 McRoberts et al. 2013), local forest inventories (Mura et al. 2016, 769 Bottalico et al. 2017), supporting biodiversity monitoring (Wulder et 770 al. 2008, Valbuena et al. 2013), or for the characterization of wildlife habitats (Vogeler et al. 2014). ALS data have, in fact, the ability to 771 772 capture highly detailed structural properties of forests (Hyyppä et al. 773 2008, Holopainen et al. 2014). In operational wall-to-wall forest inventories, a two-stage procedure using ALS data and field plots, 774 with the so called Area-Based Approach - ABA (Næsset 2002), has 775 become particularly common, and several countries (e.g. Norway, 776 Sweden, and Finland) already use this technology in the operational 777 implementations of their NFIs (Maltamo et al. 2007, Næsset 2007, 778 779 Nilsson et al. 2017).

In Italy too, after the first studies in early years 2000 (Barilotti et al. 780 2005), several investigations demonstrated that ALS is the most 781 782 important data to calculate predictors for the estimation of forest 783 variables. For an excursus on the first decade of ALS applications in Italy we refer to Montaghi et al. (2013). While more recently Mura 784 785 et al. (2016), developed a methodological approach to map a multiple variation index of forest structural diversity with a statistically 786 787 rigorous inference approach. Chirici et al. (2016) compared four 788 model-assisted estimates of total forest aboveground biomass 789 obtained using ALS echo-based and CHM metrics. Chirici et al. (2018) assessed forest windthrow damaged in Tuscany after the
storm of March 2015 using a post-event ALS data, while Giannetti
et al. (2018) assessed single-tree attributes in a complex mixed
Mediterranean forests by the integration of ALS and terrestrial laser
scanner.

795 However, despite a fairly rapid growth of these techniques, ALS data 796 in Italy are not yet available wall-to-wall (Montaghi et al. 2013, Scrinzi et al. 2017) while several local acquisitions are instead 797 798 available. A complete overview of all the ALS datasets available in Italy is still missing, this information is instead crucial to identify the 799 800 best way to integrate ALS with other source of information such as those from field surveys conducted by the Italian NFI or to plan 801 802 future ALS acquisitions (Corona et al. 2017).

In such a framework the general objective of this work is to present
the activities carried out to better understand the consistency of forest
maps and ALS data in Italy, these activities can be considered
preliminary to a future possible implementation of a National Forest
Information System.

More specifically this paper is aimed at: i) investigating if existing 808 local maps can be aggregated to create an innovative national high 809 resolution forest/non-forest map (forest mask) able to provide forest 810 811 area figures consisting with official forest area estimates from the 812 Italian NFI, for comparative purposes we assessed the quality of 813 other two forest masks developed in the framework of international 814 processes; ii) presenting the creation of an innovative CHM based on 815 the aggregation of all the major ALS dataset available in Italy; and

816 iii) to develop for demonstration purposes an on-line geographic FIS817 to provide free access to these two layers.

To do so in the paper we first provide an overview of the available forest maps (local, continental and global) and ALS data available in Italy, while in the second part we present the procedures used for preparing the forest mask and the CHM, finally in the last part we introduce the on-line geographic FIS.

## 823 2. Materials

824

# 2.1. Study area

825 Italy covers 301,338 km<sup>2</sup>, centered at latitude 42° 30' and longitude 826 12° 30', and it is divided into 21 local administrations called Regions 827 and Autonomous Provinces (the Nomenclature of Territorial Units 828 for Statistics (NUTS) level 2, following the European Statistical Office (Eurostat) classification). Italy has large climate and 829 830 topographical variability with coastal flat areas, hills and two main mountain chains, the Apennines from North-West to South-East, and 831 832 the Alps in the northern part of the country from West to East, with 833 elevations ranging between 0 and 4800 m a.s.l. Based on the last 834 available Italian NFI statistics (INFC 2007), forests and other wooded lands cover 34.7% of the national land territory, with 835 104,675 Km<sup>2</sup>. Forests are dominated by broadleaved species (68% of 836 total forest area), with the presence of 7 out the 14 European Forest 837 838 Types (Barbati et al. 2014).

## 839 **2.2. Forest mask of Italy**

In this paragraph we present the data used to create and validate a
new high resolution forest mask of Italy from local maps, as well as
the other two forest masks used to contrast the quality of this new
map.

844 **2.2.1. Forest mask from local maps** 



- Figure 1. Acquisition year of local maps used to create the high resolutionforest mask.
- 848

849 To create a high resolution forest mask of Italy we mosaicked maps

- 850 of different types. When available we used high resolution forest
- 851 maps (for a total of 16 maps produced with a nominal reference scale

varying between 1:5,000 and 1:25,000), based on forest types
classification systems. For five regions we used land use maps based
on the Corine Land Cover classification system, and one map (i.e.
Sardegna) based on the classification systems of the CORINE
Biotopes. All the maps were produced on the basis of manual
photointerpretation of aerial orthophotos between 2000 and 2016.

The maps were downloaded from local on-line geoportals, when multiple versioning at different years were available, we used the version for the year closest to the year 2005, the reference year of the Italian NFI (Figure 1). The details of the local maps used to create the high resolution forest masks are in Table 1.

863

#### 2.2.2. Forest masks from international layers

Even if a large number of global or continental forest maps are available we decided to use only two products provided by JAXA globally and by COPERNICUS for Europe since they are the only ones that declare to officially mimic the FAO definition of "forest and other wooded land", which is adopted by the Italian NFI too.

JAXA, for the years 2007, 2008, 2009, 2010, and 2015 provides a 869 870 forest/non-forest mask in grid format with a 25 m resolution for the entire globe (JAXA 2016). The JAXA map is produced by automatic 871 processing of multi polarization backscatter signals acquired by the 872 two Synthetic Aperture Radars (SAR), PALSAR and PALSAR 2 873 874 (Phased Array type L-band Synthetic Aperture Radar), which are 875 mounted on the two satellites ALOS and ALOS-2 (Advanced Land 876 Observing Satellite). The JAXA map, here-in-after named JAXA, 877 was downloaded for the year 2007 from
878 https://developers.google.com/earth-

engine/datasets/catalog/JAXA ALOS PALSAR YEARLY SAR. 879 In the framework of Copernicus Land Monitoring Service (CLMS) 880 881 coordinated by the European Space Agency (ESA) a forest/nonforest mask in grid format covering the entire Europe for the years 882 883 2012 and 2015 is derived from the FTY, available among the so-884 called HRL, in which the mainly input sources of the forest layers 885 are Sentinel-2 and Landsat 8 time series, complemented by SPOT-5 886 and ResourceSat-2 satellite data. The FTY map for the year 2012 887 derived through a spatial intersection of the two primary status layers TCD and DLT (Langanke 2017), was downloaded from: 888 889 https://land.copernicus.eu/pan-european/high-resolution-

layers/forests/forest-type-1/status-maps/2012?tab=download. 890 As reported in the metadata, the FTY allows to mimic as close as 891 892 possible the FAO forest and other wooded land definition. In its original spatial resolution (20 m) it consists of two products: 1) a 893 forest types product with a MMU of 0.5 ha, as well as a 10% tree 894 895 cover density threshold applied, and 2) a support layer mapping trees under agricultural uses and in urban contexts on the basis of the land 896 uses from the Corine Land Cover (CLC 2012) project, and the 2012 897 898 degree of imperviousness product (available among the HRL). In the 899 final 20 m spatial resolution product that we used, trees 900 predominantly used for agricultural practices and trees in urban 901 context that are distinguished in the forest additional support layer

- are excluded from the map (Langanke 2017). We referred to this final
- 903 map as HRL-FM.

Tab. 1. Main characteristics of local maps used to create the high resolution

905 forest mask. Map type FM: forest mask, LUM: Land use map

906

Administrative unit	Map type	Production year	Scale	Minimum Mapping Unit (ha)
Abruzzo	FM	2009	1:10000	0.5
Basilicata	FM	2015	1:10000	0.2
Auton. Prov. of Bolzano	FM	2011	1:25000	0.5
Calabria	LUM	2012	1:10000	0.2
Campania	LUM	2009	1:25000	0.5
Emilia-Romagna	LUM	2014	1:10000	0.2
Friuli Venezia Giulia	FM	2013	1:5000	0.2
Lazio	FM	2011	1:25000	0.5
Liguria	FM	2013	1:25000	0.5
Lombardia	FM	2015	1:10000	0.25
Marche	FM	2000	1:25000	0.5
Molise	FM	2004	1:10000	0.5
Piemonte	FM	2016	1:10000	0.2
Puglia	FM	2011	1:10000	0.25
Sardegna	LUM	2013	1:200000	0.5
Sicilia	FM	2010	1:10000	0.5
Toscana	LUM	2013	1:25000	0.2
Auton. Prov. of Trento	FM	2015	1:10000	0.2
Umbria	FM	2012	1:25000	0.5
Valle d'Aosta	FM	2011	1:10000	0.5
Veneto	FM	2006	1:10000	0.5

907

## 2.3. Italian Airborne Laser Scanner surveys

We searched for all the ALS datasets available in Italy collected fromlocal, regional, and national authorities. In total, we found 29 ALS

datasets acquired in the period 2004-2017 by means of 12 local and
national different authorities (Table 2, Figure 2). The data are
available free of charge, or in some cases can be acquired upon
request to the owner or after payment of storage fees.

914 The largest dataset was collected by the Italian Ministry of
915 Environment (MATTM), which acquired ALS data at national level
916 along the Italian coast and rivers for hydraulic risk assessment.

917



918

Fig. 2. ALS datasets and data provider (on the left side), and year of
acquisition (on the right side).

922 The remaining datasets were acquired by Regions and other local 923 authorities (i.e., municipalities, provinces, catchment management 924 authorities, and research institutions). In some areas multitemporal 925 acquisitions are also available, mainly located in Regions with wide 926 local ALS coverage where multiple MATTM surveys were carried 927 out. The Regions with multitemporal dataset are Liguria (49% of the Region), Valle d'Aosta (40%), Molise (36%), Piemonte (33%),
Trentino Alto Adige (32%), Basilicata (26%) (Figure 2). In several
Regions new ALS dataset are going to be acquired or have been
acquired already but are not yet distributed.

932 Table 2 reports the main characteristics of ALS data used in this 933 study. It is important to note that in some cases we were able to 934 collect point clouds as raw or classified data, while for some datasets we were able to collect only pre-processed data such as raster grid 935 Digital Terrain Models (DTMs), which provide the elevation of the 936 937 ground terrain above sea level, and Digital Surface Models (DSM), 938 which provide the elevation above see level of Earth surface including trees, buildings, and other features above the ground. 939

940

## 2.4. Italian National Forest Inventory reference data

The 2<sup>nd</sup> Italian National Forest Inventory (INFC 2007) is based on a
three-phase, non-aligned, systematic sampling design, results are
referred to the year 2005. For more details on Italian NFI we refer to
Fattorini et al. (2006) and Chirici et al. (2020).

To assess the accuracy of the forest masks we used the official estimates of total forest area (i.e., forest + other wooded land) available on line at https://www.sian.it/inventarioforestale/ aggregated at national (NUTS1), regional (NUTS2), and province levels (NUTS3) (INFC 2007).

950 **2.5. Italian national grids** 

Resampling all the different layers (forest masks and ALS) to
produce an harmonized spatial datasets with a common spatial
extension and resolution is a standard procedure when maps have to
be compared with the information collected in the NFIs (Kangas et
al. 2018).

For this study we generated for Italy two reference grids, both 956 projected using the coordinate system WGS 84 / UTM zone 32 North 957 (EPSG:32632), at two different resolutions: 1 m and 23 m. The 958 959 tessellation at 23 m was chosen to mimic the size of the field plots measured in the framework of the Italian NFI (Chirici et al. 2020) 960 and it generated 569,769,690 cells. Following this approach all the 961 962 raster layers potentially included in the geographical FIS should be 963 resampled to the same 23 m resolution.

964 The 23 m cells were then subdivided to create the 1 m grid, consisting
965 of 301,408,166,010 cells, used for the following harmonization
966 process of local forest maps.

Tab. 2. Main characteristics of ALS surveys available in Italy. In spatial resolution column, "Raw data" refers to the
 availability of point clouds with or without ground/non-ground classification.

ID	Data provider	Survey year	Survey area	Km <sup>2</sup>	Flight altitude	Density (pulse/m <sup>2</sup> )	Spatial res. (m)	Sensor	
1	Basilicata	2013	Basilicata	14382	900m	4	5x5	Riegl LMS Q680i	
2	Autonomous Province of Bolzano	2004 - 2006 -	South Tyrol	7411	850- 1100m	0.6	2,5x2,5	TopoSys Falcon II and Optech Gemini ALTM 3033	
3	Bosco Fontana	2006	Bosco Fontana	3	-	5.6	1x1 + raw data	Optech ALTM 3100	
4	Municipality of Firenze	2017	Florence	102	915m	4	1x1	Riegl LMS-Q680i	
5	Autonomous Region of Fruli Venezia- Giulia	2006 - 2010 -	Fruli Venezia- Giulia	10420	180- 3000m	4	1x1 + classifie d raw data	Optech Gemini ALTM 3033	
6	LaMMA	2015	Tuscany forest windthrows 4-5/03/15	436	1100m	4.4	1x1 + classifie d raw data	Riegl LMS-Q680i	
7	MATTM Contracts: 140, 145, 155, 172, 204, 208	2007 - 2016	National Rivers	24154339	-	-	1x1 + raw data	ALTM Gemini, ALTM 3100, Pegasus	

ID	Data provider	Survey year	Survey area	Km <sup>2</sup>	Flight altitude	Density (pulse/m <sup>2</sup> )	Spatial res. (m)	Sensor	
8	MATTM Contracts: 140, 176	2008 - 2012 -	Coast line	1926671	-	-	2x2 + raw data	ALTM Gemini, ALTM 3100, Pegasus	
9	Piemonte	2009 - 2011 -	Piemonte	291792692	4500m	0.5	5x5 + raw data	LEICA ALS50-II (Leica Geosystems 2006)	
10	Autonomous Region of Sardegna	2008	Alghero	666	800m	1	5x5	Riegl LMS-Q560	
11	Autonomous Region of Sardegna	2008	Coast	5579	1400m	1	1x1	Optech Gemini ALTM	
12	Autonomous Region of Sardegna	2009	Ogliastra	318	800m	5	1x1 + raw data	Riegl LMS-Q560	
13	Autonomous Region of Sardegna	2013	Urban centers	15415	700m	4	1x1	Riegl LMS Q680i	
14	Toscana, Province of Arezzo	2004	Arno, Tevere.	89	1200m	0.5-1.5	2x2	Optech Gemini ALTM 3033	

ID	Data provider	Survey year	Survey area	Km <sup>2</sup>	Flight altitude	Density (pulse/m <sup>2</sup> )	Spatial res. (m)	Sensor	
15	Toscana Serchio basin authority	2005	Canale Ozzeri, Rio Guappero	31	1200m	1	1x1	Optech Gemini ALTM 3032	
16	Toscana Serchio basin authority	2006	Serchio and main tributaries	12435	1200m	1	1x1	Optech Gemini ALTM 3033	
17	Toscana	2006- 2007	Mugello, Sieve	305	1200m	1	1x1	Optech Gemini ALTM 3033	
18	Toscana, Province of Siena	2007	Ombrone, Arbia	35	1500m	1	1x1	Optech Gemini	
19	Toscana, Arno basin authority	2008	Elsa, Ombrone, Bisenzio, Sieve	913	1200m	1.50	1x1	ALTM Gemini	
20	Toscana, Arno basin authority	2008	Monti della Calvana	314	2300m	0.40	3x3	ALTM Gemini	
21	Toscana, Arno basin authority	2009	Monti della Calvana	314	2300m	0.40	2x2	ALTM Gemini	
22	Toscana	2010	Lunigiana, Pistoia, Lucca, Scarlino	1923	-	0.50	1x1	Optech Gemini ALTM and Optech Pegasus ALTM	

ID	Data provider	Survey year	Survey area	Km <sup>2</sup>	Flight altitude	Density (pulse/m <sup>2</sup> )	Spatial res. (m)	Sensor	
23	Toscana	2011	Aulla	85	1800- 1900m	0.5	1x1	Optech Gemini	
24	Toscana	2012	Carrara, Pienza, Minucciano , Vagli	101	1600m	1.7	1x1	Optech Gemini ALTM and Optech Pegasus ALTM	
25	Toscana	2012	Magra	54	1400m	1.5	1x1	Optech Pegasus	
26	Toscana	2012	Teglia, Osca, Mangiola	26	1050m	1.5	1x1	Optech Gemini	
27	Autonomous Province of Trento	2006	Trentino excluded Adige river	6702	1000- 1800m	1.8	1x1 + raw data	Optech ALTM 3100	
28	Autonomous Province of Trento	2009	Adige river	636	1500m	0.5	1x1 + raw data	TopoSys	
29	Autonomous Region of Valle d'Aosta	2008	Valle d'Aosta	3620	2700- 4700m	2	2x2 + raw data	Optech Gemini ALTM	

### **3. Methods**

973

3.1. Forest mask

974 **3.1.1. Harmonization of local forest maps** 

The local forest maps were all reprojected in the same coordinate system (WGS 84 / UTM zone 32 North, EPSG:32632), merged and rasterized as grid layers using the national grid with 1 m x 1 m spatial resolution, and reclassified into Boolean masks using the code 1 for pixels classified as "forest", and the code 0 for pixels classified as "non-forest". Please note that hereinafter with the term "forest" we intend the FAO definition of "forest and other wooded areas".

982 Since the different local maps were developed on the basis of different
983 resolutions we decided to apply a simple procedure to harmonize all
984 the maps to the FAO forest definition based on the minimum tree
985 cover of 5%.

986 For each 23 x 23 m cell we then calculated the forest cover ratio  $(F_i)$ 987 as:

988 
$$F_i = \frac{\sum_{i=1}^n y_i}{n} x 100$$
(1)

989 where  $y_i$  is the number of forest pixels resulting from the 1x1 m forest 990 mask, and n = 529, that is the total number of 1 x 1 m pixels in the 23 991 x 23 m pixels.

All the 23 x 23 m cells with  $F_i \ge 5\%$  were classified as "forests" and labelled with code 1, and the remaining cells were classified as "non-

- 994 forest" and labelled with code 0. After the merging process we 995 obtained a national layer hereinafter identified as REG-FM (REGional
- 996 Forest Mask).

997 For comparison the JAXA and HRL maps were resampled from their

- 998 original resolution (25 and 20 meters respectively) to the 23 m national999 grid creating Boolean maps as well.
- 1000 The results are three national raster forest masks REG-FM, JAXA-
- FM, and HRL-FM with the spatial resolution of the national grid at 23x 23 m.
- 1003

## **3.1.2.** Accuracy assessment of the forest masks

To assess the accuracy of the three forest masks (REG-FM, JAXAFM, and HRL-FM) we used as benchmark the forest area reported by
the official statistics of the 2<sup>nd</sup> Italian NFI (INFC 2007).

For each national forest mask, we computed the forest area for all Italy
(NUTS1), for 20 Regions (administrative units at NUTS2 level,
considering the Autonomous Provinces of Trento and Bolzano as a
unique Region) and 103 Provinces (administrative units at NUTS3
level).

- 1012For each administrative unit we computed the percentage difference1013(*diff*%) between the official NFI forest area estimate, and the resulting
- 1014 forest area from the different forest masks as:

1015 
$$dif f_{\%} = \frac{(A_{mask} - A_{NFI})}{A_{NFI}}$$
 (2)

1016 where  $A_{mask}$  is the forest area calculated on the basis of the forest masks

1017 (REG-FM, JAXA-FM, HRL-FM), and A<sub>NFI</sub> is the forest area provided

1018 by official NFI statistics.

1019 We also calculated the coefficient of determination ( $\mathbb{R}^2$ ), and the root 1020 mean square error, both in absolute ( $\mathbb{R}MSE$ ) and relative values 1021 ( $\mathbb{R}MSE_{\%}$ ) against the  $A_{NFI}$ , between NFI statistics and the forest area 1022 from the different forest masks as:

1023 
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (A_{mask_i} - A_{NFI_i})^2}{n}}$$
 (3)

$$1024 \qquad RMSE_{\%} = \frac{RMSE}{A_{NFI}} \tag{4}$$

1025 where  $A_{maski}$  is the forest area calculated on the basis of the forest 1026 masks in the *i*-th administrative unit,  $A_{NFIi}$  is the forest area provided 1027 by official NFI statistics for the *i*-th administrative unit, *n* is the 1028 number of administrative units.

1029 In the results section we reported the forest area estimates from the1030 NFI for the same administrative units together with their estimated1031 standard error.

1032 **3.2. Ital** 

## **3.2. Italian National CHM**

1033 The available ALS datasets were derived from several flight 1034 campaigns and were provided with different specifications and 1035 formats. Therefore, the generation of a homogeneous CHM required 1036 specific pre-processing steps depending of data characteristics. 1037 For those ALS datasets where raw data were available (IDs 3, 5, 9, 12, 27, 28 of Table 2) we firstly classified the ALS point clouds and then 1038 1039 we produced DTMs and DSMs. In LasTools (Isenburg 2017) we eliminated errors in returns with the "lasnoise" algorithm, we 1040 1041 classified the pulses corresponding to ground and non-ground with 1042 "lasground" and "lasclassify" and then we generated the 1 meter 1043 resolution DTM from ground pulses and the DSM from non-ground 1044 pulses with the adaptive TIN model algorithm using "las2dem" in 1045 LasTools. For those datasets (IDs 1, 2, 8, 10, 14, 20, 21, 29) where 1046 point clouds were not available, and DTM and DSM were at a 1047 resolution different from 1 m, the datasets were resampled using a 1048 cubic convolution.

1049 For the remaining dataset (IDs 4, 6, 7, 11, 13, 15, 16, 17, 18, 19, 22,

1050 23, 24, 25, 26) the layers were used as they are since provided already1051 at 1 m resolution.

We then computed a 1 m resolution CHM subtracting the DTM from
the DSM in the coordinate reference system WGS 84 / UTM zone 32

1054 North (EPSG:32632), and lastly, for each 23 x 23 m cell of the national

- 1055 grid we calculated the mean value from the 1 m resolution CHM.
- 1056 **4. Results**
- **4.1. Forest mask**

1058 On the basis of the forest mask resulting from the aggregation of local

1059 forest maps (REG-FM) the total area of forest and other wood lands

1060	resulted of 102,610.9 km <sup>2</sup> , the 34% of land area of Italy, the 1.9% less
1061	than the estimated official statistics of the NFI (104,675.33 $\rm km^2$ with
1062	an estimated standard error of 0.3%). Both the other two forest masks
1063	were less congruent with NFI estimates. With JAXA-FM we resulted
1064	a total of 100,177.8 $\rm km^2$ (33.2% of land area), with an underestimation
1065	of almost the 4% compared to NFI estimates; with HRL-FM we
1066	obtained a total of 112,133.1 $\rm km^2$ (37.2% of land area), with an
1067	overestimation of approximately the 16% if compared to NFI
1068	estimates (Table 3).
1069	
1070	
1071	
1072	
1073	
1074	T = 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 +

1074Tab. 3. Accuracy assessment of national forest masks (REG-FM, JAXA-FM,1075HRL-FM) at NUTS1 (Italy) and NUTS2 (Region) levels. The forest and other1076wooded land area and the percentage difference (Diff%) are reported for1077forest masks; the forest area and its standard error (ES%) are reported for NFI1078estimates.

NUTS		Forest owl area (km <sup>2</sup> )			Diff% (%)			Forest area (km <sup>2</sup> )
1 11		REG	JAXA	HRL	REG	JAXA	HRL	
1	Italy	102610.9	100177.8	112133.1	-2	-4	7	104675.3 (0.3)
	Abruzzo	4556.0	3913.8	5118.1	4	-11	17	4385.9 (1.3)
	Basilicata	3239.5	2402.2	3216.8	-9	-33	-10	3564.27 (1.5)
	Calabria	7841.4	6211.4	6715.5	28	1	10	6129.31 (1.1)
	Campania	4460.5	4644.4	5942.5	0	4	33	4452.75 (1.5)
	Emilia-Romagna	6202.7	6026.1	6915.4	2	-1	14	6088.17 (1.2)
	Friuli Venezia Giulia	3244.7	3473.4	3531.8	-9	-3	-1	3572.24 (1.3)
	Lazio	6200.5	6624.5	6849.3	2	9	13	6058.59 (1.2)
	Liguria	3925.8	4692.9	3963.5	5	25	6	3751.34 (1.1)
	Lombardia	6203.0	8487.4	7029.5	-7	27	6	6657.01 (1.2)
2	Marche	2618.6	2452.5	3305.0	-15	-20	7	3080.76 (1.6)
2	Molise	1583.4	1038.3	1847.1	7	-30	24	1486.4 (2.3)
	Piemonte	9326.8	11266.3	10165.4	-1	20	8	9401.15 (1)
	Puglia	1735.3	1749.3	4390.0	-3	-2	145	1790.4 (2.6)
	Sardegna	8943.8	5636.2	9115.8	-26	-54	-25	12132.51 (0.8)
	Sicilia	5134.5	2082.7	3941.2	52	-38	17	3381.71 (1.9)
	Toscana	11687.8	11645.2	12626.2	1	1	10	11515.38 (0.7)
	Trentino-Alto Adige	7503.8	7958.6	7311.3	-4	2	-6	7797.05 (1.2)
	Umbria	3411.2	3400.3	4160.8	-13	-13	7	3902.55 (1.2)
	Valle d'Aosta	981.1	1435.6	966.3	-7	36	-9	1059.28 (2.7)
	Veneto	4131.2	5036.9	5021.7	-8	13	12	4468.56 (1.4)

1080	The REG-FM forest mask resulted the most accurate also at Regional
1081	(NUTS2) and Province (NUTS3) level. For NUTS2 we found that
1082	REG-FM matches the NFI estimates better than other masks in 11 out
1083	of 20 Regions, against the 6 of HRL-FM, and the 3 of the JAXA-FM
1084	(Table 3). With REG-FM the greatest underestimations were for
1085	Marche (-15%) and Sardegna (-26%), while the greatest
1086	overestimations were for Sicilia (+52%) and Calabria (+28%). For
1087	NUTS3, the REG-FM was the best one in 50 Provinces out of 103, 28
1088	with HRL-FM and 25 with JAXA-FM. At the REG-FM at Province
1089	level demonstrated similar behaviour registered at Regional level,
1090	with a stronger overestimation in Puglia, Sicilia, and Calabria, and
1091	underestimation in Lombardia, Veneto, Basilicata, and Sardegna.



Maala	RMS	SE	RMS	Е%	<b>R</b> <sup>2</sup>		
Mask	Province Region		Province Region		Province	Region	
HRL	24943	112940	24%	21%	0.93	0.90	
JAXA	41340	154774	40%	29%	0.82	0.80	
REG	23254	103133	23%	19%	0.95	0.91	

1097The  $R^2$  and RMSE results confirmed these findings since both for1098Regions (NUTS2) and Provinces (NUTS3) the REG-FM appeared the1099best forest masks between those analyzed with  $R^2 = 0.91$  and RMSE1100= 19% and  $R^2 = 0.95$  and RMSE = 23% for NUTS2 and NUTS31101respectively, while HRL-FM and JAXA-FM ranged between 0.80 and

1102 0.90 in terms of  $\mathbb{R}^2$ , and between 21% and 40% in terms of RMSE



1103 (Table 4 and Figure 3).

1105 Fig. 3. Correlation between the forest masks (REG-FM, JAXA-FM, HRL-1106 FM) area and the NFI estimates at NUTS2 (Region) and NUTS3 (Province) 1107 level. The dotted blue line is the y = x line.

1108 **4.2. Italian National CHM** 

1109 As a result of the mosaicking activity of ALS we produced a 23 m 1110 resolution CHM which covers an area of 191,076.52 km<sup>2</sup>, which

- 1111 represents the 63% of the territory of Italy, and the 59% of the Italian
- 1112 forest area (based on the REG-FM) (Figure 4).



1114 Fig. 4. On the left side the CHM we generated, in the middle the forest area

- 1115 covered by CHM, on the right side the high resolution forest mask.
- 1116



- 1117
- 1118 Fig.5. Area distribution per year of ALS surveys.
- 1119
- 1120 Most of ALS data were acquired between the years 2008 and 2011
- 1121 (the 68% of the total area covered by ALS) (Figure 5).
- 1122 At Regional level (NUTS2) we found that the CHM fully covers the
- 1123 forest area of 4 Regions: Trentino-Alto Adige, Friuli Venezia Giulia,

Basilicata, and Valle d'Aosta. In three more Regions, Piemonte,
Calabria and Liguria, the forest area is almost fully covered with 99%,
91% and 86% coverage respectively. Emilia-Romagna was the region
with the lowest percentage (only the 7% of the forest area), followed
by Lombardia (21%), Lazio (23%), and Marche (37%) (Figure 6).



1130 Fig.6. ALS cover for each Region in Italy.

1131

1129

## 1132 4.3. Forest Information System Web-GIS infrastructure

1133 To give open-access to the harmonized geographic layers, we 1134 developed a first demonstrating web-GIS service, which could be 1135 considered as a possible example for the future development of a 1136 geographic Forest Information System. The platform is an easy 1137 infrastructure which permits to view and query the Italian National 1138 Forest Mask and the National CHM we developed. The infrastructure

- 1139 is designed to be scalable and updated constantly. The first version of
- 1140 the Web-GIS platform (v. 1.0) is available at www.forestinfo.it
- 1141 (Figure 7).



- 1142
- 1143 Fig. 7. Forest Information System Web-GIS interface.
- 1144

1145 Through the Web-GIS, the users can interact in form of GIS-layers 1146 with the two datasets which at the moment are not downloadable. The infrastructure was developed using free-open source geospatial 1147 1148 libraries. At present the Web-GIS is based on GeoServer (Java) and Lizmap®, with data stored on a PostgreSQL database, implemented 1149 1150 with the extension PostGIS, which allows to select data by query and 1151 to create maps. The Web-GIS was structured as a system where the 1152 data management and the data processing are separated. The 1153 infrastructure was designed to collect both raster and vector layers.

#### 1154 **5. Discussion**

1155 In this study we presented: (i) the creation of a new high resolution 1156 Italian National Forest Mask developed on the basis of the aggregation 1157 of local forest maps, the mask was evaluated against the aggregated 1158 forest area estimates from the Italian NFI and in comparison of two 1159 other forest masks created on the basis of radar remotely sensed data available globally by JAXA (JAXA-FM) and for Europe by the 1160 COPERNICUS services on the basis of optical imagery (HRL-FM); 1161 (ii) a new National high resolution CHM generated as aggregation of 1162 1163 harmonized datasets available locally in Italy; and (iii) the first National forest Web-GIS, a platform designed to store and navigate 1164 1165 through geographic forest layers.

Three years were necessary to collect the data used in this project, fora total of approximately 24.7 Tera Byte.

1168 To create the Italian National Forest Mask from local forest maps the 1169 major problems we encountered were related to the different forest 1170 definitions and classification systems used by the different Authorities 1171 (Table 1). In fact, the Italian National Forest Mask (REG-FM) was 1172 obtained by merging 20 local forest maps (considering the 1173 Autonomous Provinces of Trento and Bolzano as a unique Region) 1174 created by photointerpretation. Among these, 12 used the FAO forest definitions, and 8 the regional forest definitions. For these reasons, the 1175 first phase of our work was the harmonization of the different forest 1176 maps deleting forest areas that did not respect the FAO forest and other 1177

1178 wooded land definition. In general, the study revealed that the REG-FM we produced, is more congruent with NFI forest area estimates 1179 than JAXA-FM and HRL-FM, for all the considered spatial scales 1180 (i.e., NUTS1, NUTS2 and NUTS3). This was true especially for north 1181 1182 and central Italy since REG-FM showed more consistent deviations 1183 from NFI figures in the islands (i.e., Sardegna and Sicilia) and in the 1184 southern regions (i.e., Calabria, Basilicata, and Puglia). This tendency 1185 can be due to the large presence in these Regions of olive groves, 1186 orchards, and abandoned pastures or crops with sparse tree coverage 1187 that can be easily confused with other forest types or shrubby 1188 formations typical of the Mediterranean "macchia".

The mean area difference with the NFI estimates across the different Regions was 0.6%, with area differences  $\geq 10\%$  for Sicilia (+52%), Calabria (+28%), Sardinia (-26%), Marche (-15%) and Umbria (-13%). Moreover, it is important to note that we found larger differences between REG-FM and NFI official statistics in those Italian regions where NFI had larger standard errors too (i.e., between 1.1% and 1.9%) (INFC 2007).

For Sardegna, such discrepancies may be due to the classification system used to develop the local map based on habitats and not specifically for forests. In addition, such map was produced with a very small nominal scale (1:200,000), which is not consistent with a minimum mapping unit of 0.5 ha that should be adopted to be consistent with the FAO forest definition. Finally in this region forests 1202 and other wooded lands are frequently characterized by different types of Mediterranean macchia that is complex to classify through remotely 1203 1204 sensing data, even by manual photointerpretation (Hüttich et al. 2014). After all we observed in Sardegna large discrepancies for JAXA-FM 1205 1206 and HRL-FM maps too. In the northern Regions, the forest area from 1207 the HRL-FM map was typically underestimated when compared to NFI figures (from -0.8% of Piemonte to -9% of Friuli Venezia Giulia) 1208 1209 (Table 3), probably for the presence of numerous high-altitude forest 1210 edges, with tree cover between 5 and 10% where the transition 1211 between shrublands, bushlands or other wooded lands is difficulty to 1212 assess by photo interpretation. It should be noted that many northern regions do not use a specific class for other wooded land in their 1213 1214 nomenclature systems, making the harmonization of these maps 1215 difficult

1216 The satellite-derived forest masks (i.e., JAXA-FM and HRL-FM) 1217 were less accurate than the REG-FM, especially in the southern 1218 Regions, but with different behaviors. The JAXA-FM underestimated the forest area in the southern Regions where the forest is 1219 1220 characterized by low vegetation and a limited accumulation of 1221 growing stock volume. Most probably because the sensibility of SAR 1222 backscatter in HV-polarization (JAXA 2016) is relatively poor in 1223 these types of vegetation (Hüttich et al. 2014, Bartsch et al. 2020). The 1224 HRL-FM overestimated the forest area for all the considered spatial scales, even if the vegetation in urban and agricultural contests was 1225

1226 masked out with the Copernicus Forest Additional Support Layer. This procedure may be a possible source of error since it is based on 1227 1228 two other maps: i) the 2012 Imperviousness Degree layer, available 1229 among the HRL, and the CLC 2012. Both these maps may be source 1230 of errors. Especially if we consider the MMU of 25 ha adopted by the 1231 Corine Land Cover map, probably too coarse to capture the 1232 fragmented mosaic of the Italian landscape. For more details on the 1233 Imperviousness Degree layer and the Forest Additional Support Layer 1234 please refer to Langanke (2017). Consequently, for example, we found that in Puglia the HRL-FM classified as forest most of the olive 1235 groves, leading to a discrepancy of 145% between HRL-FM and NFI 1236 1237 data in Puglia.

Our results show that REG-FM is the national forest mask most 1238 1239 congruent with the estimates of the Italian NFI, despite the limitations 1240 found in some of the Regions. The Italian National Forest Mask in the 1241 future could be useful for several applications, for example to create wall-to-wall spatial estimates of forest variables (Chirici et al. 2020), 1242 to mask out non-forest areas when monitoring forest disturbances 1243 1244 from clear-cuts (Giannetti et al. 2020, Francini et al. 2020) or 1245 windthrow damages, as well as for studying forest fragmentation and 1246 ecological networks at national scale level.

As mentioned already Italy does not have an ALS wall-to-wall
coverage yet (Chirici et al. 2020) and, before this study, the exact area
of the Italian land covered by ALS data, as well as a state of the art of

1250 all the main ALS data acquisition, was unknown. To the best of our knowledge, this is the first study reporting an exhaustive description 1251 of the ALS data available in Italy. Moreover, ALS datasets collected 1252 by different authorities resulted to have some common characteristics 1253 1254 which are considered by other authors suitable for forestry 1255 applications (Goodwin et al. 2006, Wulder et al. 2008): flight altitudes 1256 of the acquisition between 500 m and 3000 m, spatial resolution of derived DTM and DSM raster ranging between 1 m and 5 m, pulse 1257 density is between 0.4 and 5 (pulses per  $m^2$ ). Low-pulse ALS (0.4 - 1 1258 pulses per  $m^2$ ), usually aimed at the creation of digital elevation 1259 1260 models (i.e. DTM or DSM), still allow a reliable estimation of typically forest structure metrics at the plot level (~23 m pixel size) 1261 1262 (Jakubowski et al. 2013).

Most of the ALS data available in forest area were acquired in three years, i.e., 2008 (18.4%), 2009 (17.6%), and 2010 (21.9%). Several studies demonstrated that a gap larger than 5 years between field measures and ALS data is problematic when ABA approach is used to estimate forest variables (Wulder et al. 2008, Tompalski et al. 2019). These means that when the new NFI data for 2015 will be available, most of the existing ALS data will be useless.

In addition to new ALS surveys, new data from NASA's Global
Ecosystem Dynamics Investigation (GEDI) mission, which is a
waveform LiDAR sensor mounted on the International Space Station
that is designed to provide a sample of ground-based and canopy

LiDAR metrics for large-scale analysis in the Mediterranean forest as
well (Dubayah et al. 2020), should be considered for future
applications and implementations nationwide.

In the near future, the availability of these two national geographic 1277 1278 layers will allow the possibility of producing others national forest 1279 layers. For example, it will be possible to derive spatial estimation of NFI forest variables using model-assisted estimators, which require 1280 1281 the availability of a forest mask (McRoberts et al. 2014, Mura et al. 2016, Bottalico et al. 2017, Chirici et al. 2020) or hierarchical models, 1282 which are specifically designed to use partial CHM coverage (Saarela 1283 1284 et al. 2016). Moreover, the two geographic layers can be used to study forest structure (Wulder et al. 2008, Valbuena et al. 2013, Mura et al. 1285 2016, Bottalico et al. 2017), species characterization (Maltamo et al. 1286 2015), habitat modeling (Vihervaara et al. 2015), or mapping forest 1287 1288 disturbances using optical remote sensing data (Giannetti et al. 2020, 1289 Francini et al. 2020).

1290 To ensure a wide use of the two national layers, and of other national forest layers that will be released in the future, a FIS Web-GIS 1291 1292 platform was developed. The platform has a free-access and allows 1293 users to perform query on specific areas of interest using map products 1294 consistent and aligned with a national grid with cell size of 23 x 23 m. 1295 Expected future improvement of the FIS Web-GIS are: to implement 1296 additional tools specifically designed for forest applications to overcame the limitation of the regional geo-portal, which are usually 1297
1298 designed just for cartographic purposes; to release other national forest

1299 geographic layers, such as the growing stock volume map based on

1300 NFI data (Chirici et al. 2020).

1301 Finally, it is important to note that national spatial forest datasets such 1302 as forest/non-forest mask, forest types, forest roads, and growing stock 1303 volume are basic requirements to develop precision forestry 1304 applications (Corona et al. 2017) and forest decision support systems 1305 at national level, similar to the ones already tested at regional scale in 1306 Italy to support forest planning and forest management (Puletti et al. 2017), and to map and value forest ecosystem services (Vizzarri et al. 1307 1308 2017).

1309 **6.** Conclusion

1310 Four main conclusions can be drawn by this work:

We generated a first high resolution forest mask for Italy (REG-FM) based on the aggregation of local forest and land use maps. Even if the input original dataset were created with different forest definitions and at different dates, the resulting forest mask undestimated for less than 2% the official estimation of the total forest area from the Italian NFI.

- 1317 Even if the REG-FM resulted more congruent with NFI figures than
- 1318 the forest masks based on radar (JAXA-FM) and optical (HRL-FM)
- 1319 imagery at National level, in some Regions and Provinces the REG-
- 1320 FM was affected by strong underestimations and overestimations,

most probably for a mix of different causes (differences in forest definitions, characteristics of the vegetation especially in Mediterranean macchia and on forest edges). This indicates that the REG-FM, even if it represents the best forest mask currently available in Italy, cannot yet be adopted as an official layer for reporting purposes and an operational revision of REG-FM by manual photointerpretation should start as soon as possible.

1328 The harmonized CHM we produced aggregating all the ALS data 1329 currently available in Italy covers only the 59% of Italian forests and 1330 start to be quite old already. These data are essential for forest 1331 monitoring and should be routinely acquired together with aerial 1332 images.

1333 Through the development of a demonstration FIS Web-GIS online 1334 we demonstrated how this information can be widely distributed to 1335 all the potential stakeholders (i.e., forest owners, managers, and 1336 technicians of local and national authorities). It is important that an 1337 operational project for the implementation of a National Geographic 1338 Forest Information System on-line can start as soon as possible in 1339 Italy.

1340 Moreover, we hope that the different forest mapping and monitoring 1341 programs currently active in Italy will converge on a common 1342 nomenclature system in order to produce harmonized maps 1343 (Chiavetta et al. 2016). For this purpose, we suggest to adopt the

European Forest Types nomenclature systems (Barbati et al. 2014),which covers all the forest types in Italy.

1346 It is also important to remember that to develop future forest layers (such as maps of growing stock volume or biomass) with ABA 1347 1348 approach using NFI data and remote sensing data (i.e., CHM and 1349 optical remote sensing data) it is necessary to have the correct coordinate of field plots. For this reason, we hope that the Italian NFI 1350 1351 in the framework of the 3rd National Forest Inventory will release, at 1352 least for research purposes, both NFI data and the exact coordinates of plots measured in the field by the crews with GNSS. 1353

Finally, we strongly suggest the evolution of the Italian NFI program into a permanent monitoring system, in order to update the ground data over a period of 5-10 years by visiting a sub-sample of the field plots each year, as it is done in other EU and non-EU NFI programs.

- 1359 References
- Alberdi I, Cañellas I, Vallejo Bombín R (2017). The Spanish National
  Forest Inventory: history, development, challenges and
  perspectives. Pesquisa Florestal Brasileira. 37, 361. https://doi.org/10.4336/2017.pfb.37.91.1337
- 1364Barbati A, Marchetti M, Chirici G, Corona P (2014). European Forest1365Types and Forest Europe SFM indicators: Tools for monitoring1366progress on forest biodiversity conservation. Forest Ecology and1367Management.1368https://doi.org/10.1016/j.foreace.2013.07.004
- 1368 https://doi.org/10.1016/j.foreco.2013.07.004
- Barilotti A, Turco S, Napolitano R, Bressan E (2005). La tecnologia
  LiDAR per lo studio della biomassa negli ecosistemi forestali
  [LiDAR technology for biomass study in forest ecosystems]. In
  Proceedings of 15th Meeting of the Italian Society of Ecology,
  12-14. [In Italian]
- Bartsch A, Widhalm B, Leibman M, Ermokhina K, Kumpula T,
  Skarin A, Wilcox EJ, Jones BM, Frost GV, Höfler A, Pointner
  G (2020). Feasibility of tundra vegetation height retrieval from
  Sentinel-1 and Sentinel-2 data. Remote Sensing of Environment
  237, 111515. https://doi.org/10.1016/j.rse.2019.111515
- 1379Bottalico F, Chirici G, Giannini R, Mele S, Mura M, Puxeddu M,1380McRoberts RE, Valbuena R, Travaglini D (2017). Modeling1381Mediterranean forest structure using airborne laser scanning1382data. International Journal of Applied Earth Observation and1383Geoinformation57,1284https://doi.org/10.1016/j.jag.2016.12.012
- 1384 https://doi.org/10.1016/j.jag.2016.12.013
- Buchhorn M, Smets B, Bertels L, Lesiv M, Tsendbazar NE, Herold
  M, Fritz S (2019). Copernicus Global Land Service: Land Cover
  100m: Epoch 2015: Globe. Version V2. 0.2.
- 1388 Carte Géoportail, URL https://www.geoportail.gouv.fr/carte
   1389 (accessed 9.5.20)
- Chiavetta U, Camarretta N, Garfi V, Ottaviano M, Chirici G, Vizzarri
  M, Marchetti M (2016). Harmonized forest categories in central
  Italy. Journal of Maps, 12, 98-100. https://doi.org/10.1080/17445647.2016.1161437

- Chirici G, McRoberts RE, Fattorini L, Mura M, Marchetti M (2016).
  Comparing echo-based and canopy height model-based metrics
  for enhancing estimation of forest aboveground biomass in a
  model-assisted framework. Remote Sensing of Environment
  174, 1–9. https://doi.org/10.1016/j.rse.2015.11.010
- Chirici G, Bottalico F, Giannetti F, Del Perugia B, Travaglini D,
  Nocentini S, Kutchartt E, Marchi E, Foderi C, Fioravanti M,
  Fattorini L, Bottai L, McRoberts RE, Næsset E, Corona P,
  Gozzini B (2018). Assessing forest windthrow damage using
  single-date, post-event airborne laser scanning data. Forestry 91,
  27–37. https://doi.org/10.1093/forestry/cpx029
- Chirici G, Giannetti F, McRoberts RE, 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. International Journal of Applied Earth Observation and Geoinformatics 84, 101959. - https://doi.org/10.1016/j.jag.2019.101959
- 1411 Corona P, Chianucci F, Quatrini V, Civitarese V, Clementel F, Costa
  1412 C, Floris A, Menesatti P, Puletti N, Sperandio G, Verani S, Turco
  1413 R, Bernardini V, Plutino M, Scrinzi G (2017). Precision forestry:
  1414 concepts, tools and perspectives in Italy. Forest@ 14, 1–12 [In
  1415 Italian]. https://doi.org/10.3832/efor2285-014
- 1416 Di Biase RM, Fattorini L, Marchi M (2018). Statistical inferential
  1417 techniques for approaching forest mapping. A review of
  1418 methods. Annals of Silvicultural Research 42, 46–58. 1419 http://dx.doi.org/10.12899/asr-1738
- 1420 Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, 1421 Hofton M, Hutt G, Kellner J, Luthcke S, Armstrong J, Tang H, 1422 Duncanson L, Hancock S, Jantz P, Marselis S, Patterson PL, Qi
  1423 W, Silva C (2020). The Global Ecosystem Dynamics 1424 Investigation: High-resolution laser ranging of the Earth's 1425 forests and topography.Science of Remote Sensing, 1 100002. -1426 https://doi.org/10.1016/j.srs.2020.100002
- Fattorini L, Marcheselli M, Pisani C (2006). A three-phase sampling
  strategy for large-scale multiresource forest inventories. Journal
  of Agricultural, Biological, and Environmental Statistics 11,

1430	296-316 https://doi.org/10.1198/108571106X130548
1431	Forest Type 2012 - Copernicus Land Monitoring Service, URL
1432	https://land.copernicus.eu/pan-european/high-resolution-
1433	layers/forests/forest-type-1/status-maps/2012?tab=download
1434	(accessed 9.5.20)
1435	Francini S, McRoberts RE, Giannetti F, Mencucci M, Marchetti M,
1436	Scarascia-Mugnozza G, Chirici G (2020). Near-real time forest
1437	change detection using PlanetScope imagery. European Journal
1438	of Remote Sensing 53, 233-244
1439	https://doi.org/10.1080/22797254.2020.1806734
1440	Garnier M, Bastick C, Colin A, Commagnac L, Lallemant T,
1441	Maisonneuve B, Mazepa F, Simon M, Vega C (2019). La BD
1442	Forêt® version 2. L'IF - Synthèse périodique de l'inventaire
1443	forestier, 46 -Available on line at https://inventaire-
1444	forestier.ign.fr/IMG/pdf/lif_46_poster.pdf
1445	Giannetti F, Puletti N, Quatrini V, Travaglini D, Bottalico F, Corona
1446	P, Chirici G (2018). Integrating terrestrial and airborne laser
1447	scanning for the assessment of single-tree attributes in
1448	Mediterranean forest stands. European Journal of Remote
1449	Sensing 51, 795–807
1450	https://doi.org/10.1080/22797254.2018.1482733
1451	Giannetti F, Pegna R, Francini S, McRoberts RE, Travaglini D,
1452	Marchetti M, Scarascia-Mugnozza G, Chirici G (2020). A new
1453	method for automated clear-cut disturbance detection in
1454	Mediterranean coppice forests using Landsat time series.
1455	Remote Sensing, 12(22), 3720
1456	https://doi.org/10.3390/rs12223720
1457	Global PALSAR-2/PALSAR Yearly Mosaic   Earth Engine Data
1458	Catalog, URL https://developers.google.com/earth-
1459	engine/datasets/catalog/JAXA_ALOS_PALSAR_YEARLY_S
1460	AR (accessed 9.5.20)
1461	Goodwin NR, Coops NC, Culvenor DS (2006). Assessment of forest
1462	structure with airborne LiDAR and the effects of platform
1463	altitude. Remote Sensing of Environment 103, 140-152
1464	https://doi.org/10.1016/j.rse.2006.03.003
1465	Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA,

- Tyukavina A. (2013). High-Resolution Global Maps of 21stCentury Forest Cover Change. Science (80-. ). 342, 850–854.
  https://doi.org/10.1126/science.1244693
- Holopainen M, Vastaranta M, Hyyppä J (2014). Outlook for the next
  generation's precision forestry in Finland. Forests 5, 1682–1694.
   https://doi.org/10.3390/f5071682
- Hüttich C, Korets M, Bartalev S, Zharko V, Schepaschenko D,
  Shvidenko A, Schmullius C (2014). Exploiting growing stock
  volume maps for large scale forest resource assessment: Crosscomparisons of ASAR- and PALSAR-based GSV estimates with
  forest inventory in Central Siberia. Forests 5, 1753–1776. https://doi.org/10.3390/f5071753
- Hyyppä J, Hyyppä H, Leckie D, Gougeon F, Yu X, Maltamo M
  (2008). Review of methods of small-footprint airborne laser
  scanning for extracting forest inventory data in boreal forests.
  International Journal of Remote Sensing. https://doi.org/10.1080/01431160701736489
- 1483 INFC (2007). Le stime di superficie 2005 Prima parte. In Tabacchi
  1484 G, De Natale F, Di Cosmo L, Floris A, Gagliano C, Gasparini P,
  1485 Genchi L, Scrinzi G, Tosi V. Inventario Nazionale delle Foreste
  1486 e dei Serbatoi Forestali di Carbonio. MiPAF Corpo Forestale
  1487 dello Stato Ispettorato Generale, CRA ISAFA, Trento.
  1488 Available from: http://www.infc.it [on line].
- Isenburg M (2017). LAStools efficient LiDAR processing software, obtained from http://rapidlasso.com/LAStools.
- 1491Jakubowski MK, Guo Q, Kelly M (2013). Tradeoffs between LIDAR1492pulse density and forest measurements accuracy. Remote1493Sensing of Environment 103: 245-253.1494https://doi.org/10.1016/j.rse.2012.11.024
- 1495 JAXA (2016). Global 25m Resolution PALSAR-2 / PALSAR Mosaic
  1496 and Forest / Non-Forest Map (FNF) Dataset Description Japan
  1497 Aerospace Exploration Agency (JAXA) Earth Observation
  1498 Research Center (EORC).
- Kangas A, Astrup R, Breidenbach J, Fridman J, Gobakken T,
  Korhonen KT, Maltamo M, Nilsson M, Nord-Larsen T, Næsset
  E, Olsson H (2018). Remote sensing and forest inventories in

1502	Nordic countries-roadmap for the future. Scandinavian Journal
1503	of Forest Research 33, 397–412
1504	https://doi.org/10.1080/02827581.2017.1416666
1505	Kartor, URL
1506	https://kartor.skogsstyrelsen.se/kartor/?startapp=skogligagrund
1507	data (accessed 9.5.20)
1508	Kilden - Arealinformasjon, URL
1509	https://kilden.nibio.no/?lang=nb&X=7195706.12&Y=284337.7
1510	5&zoom=0.050089679614182224&topic=arealinformasjon&b
1511	gLayer=graatone_cache (accessed 9.5.20)
1512	Langanke T (2017). Copernicus Land Monitoring Service - High
1513	Resolution Layer Forest: Product Specifications Document.
1514	Copernicus team at EEA, 38.
1515	Liang X, Kankare V, Hyyppä J, Wang Y, Kukko A, Haggrén H, Yu
1516	X, Kaartinen H, Jaakkola A, Guan F, Holopainen M, Vastaranta
1517	M (2016). Terrestrial laser scanning in forest inventories. ISPRS
1518	Journal of Photogrammetry and Remote Sensing 115, 63-77.
1519	https://doi.org/10.1016/J.ISPRSJPRS.2016.01.006
1520	Lidar Rioja   Visor Cartográfico, URL
1520 1521	Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)
1520 1521 1522	Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20) Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A,
1520 1521 1522 1523	LidarRiojaVisorCartográfico,URLhttp://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based
1520 1521 1522 1523 1524	Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20) Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop
1520 1521 1522 1523 1524 1525	LidarRiojaVisorCartográfico,URLhttp://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)MaltamoM, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A,Hyyppä J (2007).Experiences and possibilities of ALS basedforest inventory in Finland.In Proceedings of ISPRS Workshopon Laser Scanning, 270-279.
1520 1521 1522 1523 1524 1525 1526	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017-</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scandinavian Journal of Forest</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scandinavian Journal of Forest Research 30, 336–345</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scandinavian Journal of Forest Research 30, 336–345 https://doi.org/10.1080/02827581.2014.986520</li> </ul>
1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536	<ul> <li>Lidar Rioja   Visor Cartográfico, URL http://lidarrioja.agrestaweb.org/#!/map (accessed 9.5.20)</li> <li>Maltamo M, Packalén P, Peuhkurinen J, Suvanto A, Pesonen A, Hyyppä J (2007). Experiences and possibilities of ALS based forest inventory in Finland. In Proceedings of ISPRS Workshop on Laser Scanning, 270-279.</li> <li>Maltamo M, Næsset E, Vauhkonen J (2014). Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Springer Netherlands, Dordrecht https://doi.org/10.1007/978-94-017- 8663-8</li> <li>Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015). Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scandinavian Journal of Forest Research 30, 336–345 https://doi.org/10.1080/02827581.2014.986520</li> <li>McRoberts RE, Bechtold WA, Patterson PL, Scott CT, Reams GA</li> </ul>

- 1538the USDA Forest Service: Historical perspective and1539announcement of statistical documentation. Journal of Forestry1540103, 304–308. https://doi.org/10.1093/jof/103.6.304
- 1541McRoberts RE, Næsset E, Gobakken T (2013). Inference for lidar-1542assisted estimation of forest growing stock volume. Remote1543Sensing of Environment 128, 268–275. -1544https://doi.org/10.1016/j.rse.2012.10.007
- 1545McRoberts RE, Liknes GC, Domke GM (2014). Using a remote1546sensing-based, percent tree cover map to enhance forest1547inventory estimation. Forest Ecology and Management 331, 12–154818. https://doi.org/10.1016/j.foreco.2014.07.025
- 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. International
  Journal of Applied Earth Observation and Geoinformation 23,
  288–300. https://doi.org/10.1016/j.jag.2012.10.002
- Mura M, McRoberts RE, Chirici G, Marchetti M (2016). Statistical
  inference for forest structural diversity indices using airborne
  laser scanning data and the k-Nearest Neighbors technique.
  Remote Sensing of Environment 186, 678–686. https://doi.org/10.1016/j.rse.2016.09.010
- 1559Næsset E (2002). Predicting forest stand characteristics with airborne1560scanning laser using a practical two-stage procedure and field1561data. Remote Sensinf of Environment 80, 88–99. -1562https://doi.org/10.1016/S0034-4257(01)00290-5
- 1563 Næsset E (2007). Airborne laser scanning as a method in operational
  1564 forest inventory: Status of accuracy assessments accomplished
  1565 in Scandinavia. Scandinavian Journal of Forest Research 22,
  1566 433–422. https://doi.org/10.1080/02827580701672147
- Næsset E, Gobakken T, Solberg S, Gregoire TG, Nelson R, Ståhl G,
  Weydahl D (2011). Model-assisted regional forest biomass
  estimation using LiDAR and InSAR as auxiliary data: A case
  study from a boreal forest area. Remote Sensing of Environment
  115, 3599–3614. https://doi.org/10.1016/j.rse.2011.08.021
- 1572 Nilsson M, Nordkvist K, Jonzén J, Lindgren N, Axensten P,
  1573 Wallerman J, Egberth M, Larsson S, Nilsson L, Eriksson J,

- 1574Olsson H (2017). A nationwide forest attribute map of Sweden1575predicted using airborne laser scanning data and field data from1576the National Forest Inventory. Remote Sensing of Environment1577194, 447–454. https://doi.org/10.1016/j.rse.2016.10.022
- 1578 Paikkatietoikkuna, URL https://kartta.paikkatietoikkuna.fi/?lang=en
  1579 (accessed 9.5.20)
- 1580 Puletti N, Floris A, Scrinzi G, Chianucci F, Colle G, Michelini T, 1581 Pedot N, Penasa A, Scalercio S, Corona P (2017). CFOR: a 1582 spatial decision support system dedicated to forest management Forest@ Italian]. 1583 in Calabria. 14, 135 - 140[In https://doi.org/10.3832/efor2363-014 1584
- 1585 Saarela S, Holm S, Grafström A, Schnell S, Næsset E, Gregoire TG,
  1586 Nelson RF, Ståhl G (2016). Hierarchical model-based inference
  1587 for forest inventory utilizing three sources of information.
  1588 Annals of Forest Science 73, 895–910. 1589 https://doi.org/10.1007/s13595-016-0590-1
- Scrinzi G, Floris A, Clementel F, Bernardini V, Chianucci F, Greco S, 1590 1591 Michelini T, Penasa A, Puletti N, Rizzo M, Turco R, Corona P (2017). Models of stand volume and biomass estimation based 1592 on LiDAR data for the main forest types in Calabria (southern 1593 1594 Italv). Forest@ 14, 175 - 187[In Italian]. 1595 https://doi.org/10.3832/efor2399-014
- 1596 Smith S, Gilbert J, Bull G, Gillam S, Whitton E (2010). National
  1597 inventory of woodland and trees (1995–99): methodology.
  1598 Forestry Commission Research Report. Forestry Commission
  1599 Scotland, Edinburgh. i–iv 1–60.
- 1600 Tompalski P, White JC, Coops NC, Wulder MA (2019).
  1601 Demonstrating the transferability of forest inventory attribute
  1602 models derived using airborne laser scanning data. Remote
  1603 Sensing of Environment 227, 110–124. 1604 https://doi.org/10.1016/j.rse.2019.04.006
- Valbuena R, Packalen P, Mehtätalo L, García-Abril A, Maltamo M
  (2013). Characterizing forest structural types and shelterwood
  dynamics from Lorenz-based indicators predicted by airborne
  laser scanning. Canadian Journal of Forest Research 43, 1063–
  1074. https://doi.org/10.1139/cjfr-2013-0147

- Vihervaara P, Mononen L, Auvinen AP, Virkkala R, Lü Y, Pippuri I,
  Packalen P, Valbuena R, Valkama J (2015). How to integrate
  remotely sensed data and biodiversity for ecosystem assessments
  at landscape scale. Landscape Ecology 30, 501–516. https://doi.org/10.1007/s10980-014-0137-5
- Vizzarri M, Sallustio L, Travaglini D, Bottalico F, Chirici G, Garfi V,
  Lafortezza R, Veca DSLM, Lombardi F, Maetzke F, Marchetti
  M (2017). The MIMOSE approach to support sustainable forest
  management planning at regional scale in mediterranean
  contexts. Sustainability 9. https://doi.org/10.3390/su9020316
- Vogeler JC, Hudak AT, Vierling LA, Evans J, Green P, Vierling KT
  (2014). Terrain and vegetation structural influences on local
  avian species richness in two mixed-conifer forests. Remote
  Sensing of Environment 147, 13–22. https://doi.org/10.1016/j.rse.2014.02.006
- Waser LT, Ginzler C, Rehush N (2017). Wall-to-Wall tree type
  mapping from countrywide airborne remote sensing surveys.
  Remote Sensing 9. https://doi.org/10.3390/rs9080766
- White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski
  P (2016). Remote Sensing Technologies for Enhancing Forest
  Inventories: A Review. Canadian Journal of Remote Sensing 42,
  619–641. https://doi.org/10.1080/07038992.2016.1207484
- Wulder MA, Bater CW, Coops NC, Hilker T, White JC (2008). The
  role of LiDAR in sustainable forest management. Forest
  Chronicle 84, 807–826.

### **3.2. Paper II**

1637	The Effect of Forest Mask Quality in the Wall-to-Wall Estimation
1638	of Growing Stock Volume.
1639	Vangi E. <sup>1,2</sup> , D'AmicoG. <sup>1</sup> , Francini S: <sup>1,2,3</sup> , Giannetti F: <sup>1,4,*</sup> , Lasserre B.
1640	<sup>2</sup> , Marchetti M. <sup>2</sup> , McRoberts R.E. <sup>5</sup> , Chirici G. <sup>1,4,6</sup>
1641	
1642	<sup>1</sup> Department of Agricultural, Food, Environmental and Forestry
1643	Sciences and Technologies, University of Florence, Firenze 50145,
1644	Italy;
1645	<sup>2</sup> Department of Biosciences and Territory, University of Molise,
1646	Campobasso, 86100, Italy;
1647	<sup>3</sup> Department for Innovation in Biological, Agri-food and Forestry
1648	systems, University of Tuscia, Viterbo, 01100, Italy
1649	<sup>4</sup> ForTech Joint Laboratory, University of Florence, Firenze 50145,
1650	Italy
1651	<sup>5</sup> Department of Forest Resources, University of Minnesota, Saint
1652	Paul, MN, 55108, USA
1653	<sup>6</sup> Research Unit COPERNICUSUniversity of Florence, Firenze
1654	50145, Italy
1655	* corresponding author: francesca.giannetti@unifi.it
1656	Remote Sensing
1657	https://doi.org/10.3390/rs13051038

#### 1658 Abstract

1659 Information about forest cover and its characteristics are essential in national and international forest inventories, monitoring programs, 1660 1661 and reporting activities. Two of the most common forest variables 1662 needed to support sustainable forest management practices are forest 1663 cover area and growing stock volume (GSV m3 ha-1). Nowadays, national forest inventories (NFI) are complemented by wall-to-wall 1664 1665 maps of forest variables which rely on models and auxiliary data. The spatially explicit prediction of GSV is useful for small-scale 1666 1667 estimation by aggregating individual pixel predictions in a model-1668 assisted framework. Spatial knowledge of the area of forest land is an 1669 essential prerequisite. This information is contained in a *forest mask* 1670 (FM). The number of FMs is increasing exponentially thanks to the 1671 wide availability of free auxiliary data, creating doubts about which is 1672 best-suited for specific purposes such as forest area and GSV 1673 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 1674 1675 product is best-suited for this purpose. The FMs considered were a mosaic of local forest maps produced by the Italian regional forest 1676 authorities; the FM produced from the Copernicus Land Monitoring 1677 1678 System; the JAXA global FM; the hybrid global FM produced by Schepaschencko et al., and the FM estimated from the Corine Land 1679 Cover 2006. We used the five FMs to mask out non-forest pixels from 1680 1681 a national wall-to-wall GSV map constructed using inventory and 1682 remotely sensed data. The accuracies of the FMs were first evaluated against an independent dataset of 1,202,818 NFI plots using four 1683 accuracy metrics. For each of the five masked GSV maps, the pixel-1684 level predictions for the masked GSV map were used to calculate 1685 1686 national and regional-level model-assisted estimates. The masked 1687 GSV maps were compared with respect to the coefficient of 1688 correlation ( $\rho$ ) between the estimates of GSV they produced (both in 1689 terms of mean and total of GSV predictions within the national and 1690 regional boundaries) and the official NFI estimates. At the national and regional levels, the model-assisted GSV estimates based on the 1691 1692 GSV map masked by the FM constructed as a mosaic of local forest maps were closest to the official NFI estimates with  $\rho = 0.986$  and  $\rho =$ 1693 1694 0.972, for total and mean GSV, respectively. We found a negative correlation between the accuracies of the FMs and the differences 1695 1696 between the model-assisted GSV estimates and the NFI estimate, 1697 demonstrating that the choice of the FM plays an important role in 1698 GSV estimation when using the model-assisted estimator.

1699

# 1700 Keywords: forest mask; spatial estimation; growing stock volume;1701 Italy

#### 1702 **1. Introduction**

1703 Information about forest cover and its characteristics are essential in national and international forest inventories, monitoring programs, 1704 1705 and reporting activities (Schepaschenko et al., 2015; FAO, 2010) such 1706 as in the context of international agreements (e.g., Kyoto protocol), and restoration programs (e.g., Reducing emissions from deforestation 1707 and forest degradation projects (REDD+))(FAO UNCCD, 2015). Two 1708 1709 of the most common forest variables needed to estimate sustainable forest management indicators as required by the national and 1710 1711 international framework and agreements relate to forest cover area 1712 (generally according to the international definition adopted by the 1713 Food and agriculture organization (FAO) and the total growing stock 1714 volume (GSV, m3) (McRoberts et al., 2013; Witke et al., 2019). These 1715 data are usually provided by national forest inventory (NFI) programs 1716 which use probability-based approaches to infer the estimates for large 1717 areas such as countries and regions within countries. (McRoberts et al., 2013; Hansen et al., 1983; McRoberts et al., 2006). In several 1718 countries with long NFI histories such as Norway (Næsset et al., 1719 2004), Finland (Tomppo et al., 2008), Austria (Hollaus et al., 2010), 1720 and Switzerland (Waser et al., 2006; 2015), the typical NFI ground 1721 1722 survey is nowadays complemented by continuous spatial predictions, characterized as wall-to-wall maps of forest variables which rely on 1723 models and wall-to-wall auxiliary data such as remotely sensed data 1724 1725 (Kangas et al., 2018; White et al., 2016; Næsset, 2014).

1726 Wall-to-wall GSV data are useful because they can be integrated into 1727 decision support systems to assess wood production and harvesting activities at small scales (i.e., in forest properties) (Puletti et al., 2017; 1728 Chirici et al., 2020; Giannetti et al., 2020; D'Amico et al., 2021) and 1729 1730 to produce small-scale estimates by aggregating individual pixel 1731 predictions (Särndal et al., 1992; 2003; Breidt et al., 2009; McRobetrs et al., 2016). In the probability-based framework, multiple estimators 1732 1733 including the stratified, post-stratified, and model-assisted estimators can be used. The latter is considered asymptotically unbiased in the 1734 sense that the mean of estimates obtained using the estimator for all 1735 1736 possible samples approaches the true value as the sample size 1737 increases (McRobetrs et al., 2016).

GSV and above-ground biomass are known to be strongly correlated 1738 1739 with three-dimensional (3D) data such as those acquired through 1740 airborne laser scanning (ALS) or photogrammetric techniques (Wittke 1741 et al., 2019; White et al., 2016; Næsset et al., 2008; McRoberts et al., 1742 2010; Giannetti et al., 2018; Goodbody et al., 2018). However, acquiring these data is still expensive, and some countries such as Italy 1743 1744 still do not have wall-to-wall ALS coverage (D'Amico et al., 2021). 1745 Multispectral satellite data are often used instead of or with 3D data to 1746 predict GSV, thanks to their free availability over large areas (Barrett 1747 et al., 2016; Saarela et al., 2016; Holm et al., 2017; Nilsson et al., 2017). 1748

Several types of models can be used to produce wall-to-wall 1749 predictions of forest attributes in a model-assisted approach. These 1750 models include both parametric and non-parametric techniques (White 1751 1752 et al., 2016; Chirici et al., 2020; Goodbody et al., 2019; Barrett et al., 1753 2016; Immitzer et al., 2016), with the recent prevalence of multiple 1754 linear regression and random forests (White et al., 2016; Karlson et 1755 al., 2015; Belgiu et al., 2016). Regardless of the estimation approach, spatial knowledge of the area covered by forest land is an essential 1756 1757 prerequisite, both to restrict the establishment of field plots and to 1758 restrict the application of the models. A forest mask (FM) indicates 1759 the location of forest land and is often in a raster or a spatial polygon database format. FMs are conventionally obtained by manual 1760 delineation of aerial images, or by supervised or unsupervised 1761 1762 classification of satellite imagery, from both optical or radar imagery (Stankiewicz et al., 2003; Hansen et al., 2013; Dostálová et al., 2016), 1763 1764 and more recently ALS data (Eysn et al., 2012; Dalponte et al., 2014; 1765 Rudjord et al., 2016; Øivind et al., 2018). Remotely sensed data suitable for forest mapping are nowadays frequently and freely 1766 1767 available(Woodcock et al., 2008; Wulder et al., 2019; Olofsson et al., 1768 2020). For this reason, the number of FMs has increased 1769 exponentially, creating doubts about which is best-suited for specific 1770 purposes such as forest area and GSV estimation. National 1771 information about forest extent can be estimated from any of several FMs produced independently by different research agencies globally 1772

1773 or for large areas, including the European Environmental Agency (EEA) (European Environmental Agency, 2007), the European Space 1774 Agency (ESA) (Langanke, 2017), the International Institute for 1775 Applied Systems Analysis (IIASA) (Schepaschenko et al., 2015), and 1776 1777 the Japanese Aerospace Exploration Agency (JAXA) (JAXA, 2016). 1778 Despite individual weaknesses and strengths, spatial differences 1779 among these products are evident and can lead to substantial variation 1780 in their accuracies (Schepaschenko et al., 2015; Seebach et al., 2012). 1781 Furthermore, these FMs were developed for different aims and thus have different characteristics in terms of minimum mapping unit 1782 1783 (MMU) and minimum mapping width (MMW), reference forest definition, and year of production. 1784

Multiple studies have compared land cover maps at global and local 1785 1786 levels. Fritz and See (2005) and Giri et al. (2005) compared the Global 1787 Land Cover 2000 data set and the MODIS global land cover product 1788 and highlighted areas with strong disagreements. Hovos et al. (2017) 1789 compared four global satellite-based land cover maps and showed a worsening of area agreements as the spatial scale increases. Neumann 1790 1791 et al. (2007) provided an assessment of compatibilities and differences 1792 between the CORINE2000 and GLC2000 datasets and reported 1793 general disagreement due to the combination of thematic similarities, 1794 spatial heterogeneity, and classification accuracy. Seebach et al. 1795 (2011) compared the advantages and limitations of four pan-European forest cover maps for the reference years 2000, demonstrating that the 1796

1797 spatial agreement between the maps ranged between 50% to 70% within a large study area in Europe. The authors found the greatest 1798 1799 spatial differences among all maps in the Alpine and Mediterranean regions. Here, the vulnerability to climate change and anthropogenic 1800 1801 disturbance is extremely large and will cause an increased demand for 1802 accurate wall-to-wall maps (Chirici et al., 2020). Only a few studies have analyzed the effects of using different FMs on the uncertainty of 1803 1804 forest parameter estimates. Rodríguez-Veiga et al. (2016) reported a 1805 large impact on estimates of national carbon stocks in Mexico caused by discrepancies in forest extent estimated from different FMs. In their 1806 study, Li et al. (2017) considered the uncertainty of the MODIS land 1807 cover products, finding substantial differences in the regional climate 1808 modeling outputs when the uncertainty was not considered. Esteban et 1809 1810 al. (2020) estimated the effects of the uncertainty of forest species 1811 maps used in the sampling and forest parameter estimation processes 1812 in a Spanish study area. Their study revealed that the effects of map uncertainty are not negligible, especially for less common 1813 Mediterranean forest species. 1814

1815 The choice of FM can heavily impact the estimation of forest 1816 parameters in two different manners: i) it affects the number of plots 1817 selected for the construction of the predictive model and ii) it affects 1818 the total area to which the model is applied (Esteban et al., 2020).

1819 The aim of this paper is to evaluate the impacts of the accuracies of 1820 different FMs on the estimation of GSV based on the integration of 1821 field information and remotely sensed data. We constructed a national wall-to-wall GSV map with an optimized procedure based on a 1822 random forests model with remotely sensed imagery and auxiliary 1823 data as predictors (Chirici et al., 2020). We used five different FMs to 1824 1825 mask out non-forest areas from the GSV map and then used the model-1826 assisted regression estimator to estimate total and mean GSV (m3 1827 ha-1) for the forest portion of the GSV map. We then investigated the 1828 relationship between mask accuracy and agreement between the 1829 model-assisted total GSV estimates and the official NFI estimates. The test was carried out for the entire area of Italy. Finally, we clarified 1830 which product was best-suited for total and mean GSV estimation, 1831 1832 both at national and regional levels.

1833

#### **2. Materials and Methods**

**2.1. Study area** 

The study was carried out in Italy which covers 301,408 km2 (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.

1839 The territory falls within the temperate zone of a Mediterranean 1840 climatic region (Pinna, 1970). On the coasts of the main islands, the 1841 average annual rainfall is 250 mm but reaches more than 3000 mm in 1842 the Alpine and pre-Alpine belts. Average yearly temperatures vary 1843 between 16 °C in the southern coastal areas to 10 °C in the inner

- 1844 central regions and the pre-Alps, with temperatures less than 5 °C in
- 1845 the mountain ranges and on the highest peaks.



1846

Figure 1. The study area with the distribution of the national forest inventories
(NFI) plots colored by growing stock volume (GSV) expressed in m3 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 2.2.2)
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

- 1860 sylvestris L., P. nigra A., P. pinae L., P. pinaster A.), which are mainly
- 1861 artificial plantations located in mountain areas or near the coast
- 1862 (Figure 1). Seven of the 14 European forest types occur in Italy, of
- 1863 which the most common is the thermophilous deciduous forest (White
- 1864 et al., 2016, Barbati et al., 2014).
- 1865 Italy is divided into 20 administrative regions (NUTS2) for each of
- 1866 which the NFI produces estimates of forest area, total and mean GSV,
- 1867 and their standard errors (SEs). The average GSV is 144 m3 ha-1
- 1868 (Gasparini et al., 2009).
- 1869

#### 2.2. Field Data

1870

#### 2.2.1. Second Italian National Forest Inventory

1871 The field reference data for the wall-to-wall spatial prediction of GSV 1872 were acquired in the framework of the second Italian NFI (INFC, 1873 2007) based on a three-phase, systematic, unaligned sampling design 1874 with 1 x 1 km grid cells (Fattorini et al., 2006). In the first phase, N =1875 301,300 points were selected and classified with respect to 10 coarse 1876 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 1877 points, qualitative information such as forest type, management, and 1878 property were collected during a field survey. In the third phase, for a 1879 sub-sample of 6782 points extracted from the second-phase points, a 1880 1881 quantitative survey was carried out for circular plots of 13 m radius (530 m2). All tree stems with a DBH of at least 2.5 cm were callipered, 1882

1883 and for a subsample, height was measured. For all 6782 third-phase plots, allometric models (Tabacchi et al., 2011) were used to predict 1884 1885 GSV (m3) which was then aggregated at plot-level and scaled to a per unit area basis. For this study, allometric model prediction uncertainty 1886 1887 and uncertainty due to Global Navigation Satellite System (GNSS) 1888 position error were expected to be negligible for the spatial resolution adopted (McRobetrs et al., 2015; Chirici et al., 2020; McRoberts et al., 1889 1890 2016; McRoberts et al., 2018). The third-phase plots have a mean GSV 1891 of 145.75 m3 ha-1, with a median value of 102.82 m3 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: 02-10-2020)
(McRoberts et al., 2018), for the reference year 2005.

1896 The study area was tessellated into a 23 x 23 m national grid whose 1897 pixel area matched the area of the NFI ground plots, for a total of 1898 569,769,690 pixels (D'Amico et al., 2021). The national grid was used 1899 as a spatial reference grid for resampling the predictor variables and 1900 the FM to 23 x 23 m resolution.

1901

#### 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

1907 climate change (Penman et al., 2003; Romano et al., 2011; Corona et al., 2012). IUTI is a permanent monitoring system that estimates the 1908 extent of six land use categories identified in the GPG-LULUCF. The 1909 IUTI is based on a systematic unaligned sampling design with 0.5 x 1910 1911 0.5 km grid cells which is an intensification of the NFI sample grid, 1912 for a total of 1,202,828 points of which 301,300 are the first-phase 1913 points of the NFI. The six categories reported by IUTI are urban, 1914 agriculture, forest land, grassland, wetland, other (Masek et al., 2006). 1915 Each point is photo-interpreted in three time periods (1990, 2008, 1916 2012) for estimating land-use change using aerial orthophotos with 1917 spatial resolution ranging between 1 x 1 m for 1990 and 0.5 x 0.5 m 1918 for 2008. We combined the six land use categories into forest and nonforest and assigned the value 1 to all the points classified as forest 1919 1920 (class 1.1, 1.2) and 0 to all other categories. Subsequently, the forest 1921 class included 32% of the total observations with 387.085 of 1922 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.

1927

#### 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 (Chirici et al., 2020). All variables
  were resampled from the original resolution to the 23 x 23 m pixel size
  of the national grid. A more detailed description of the database is
  provided by (Chirici et al., 2020).
- 1936 Table 1. Predictor variables based on remotely sensed and auxiliary data.

Database	Band/information	Predictor variables	Original spatial resolution
	3 years median of Band 1	Landsat_B1	30 m
	3 years median of Band 2	Landsat_B2	30 m
	3 years median of Band 3	Landsat_B3	30 m
Landsat 7 ETM+	3 years median of Band 4	Landsat_B4	30 m
	3 years median of Band 5	Landsat_B5	30 m
	3 years median of Band 6	Landsat_B6	30 m
	3 years median of Band 7	Landsat_B7	30 m
Global DALSAR/	HH polarization	SAR_HH	25 m
PALSAR	HV polarization	SAR_HV	25 m
	Total annual precipitation	Prec	1 km
	Mean annual temperature	temp_mean	1 km
Climate data	Maximum annual temperature	temp_max	1 km
	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

#### 1937 **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).

We used Landsat 7 Surface Reflectance Tier 1 imagery from the Earth 1942 1943 Engine Data Catalog, acquired in the vegetation period (1st April-30th September), atmospherically corrected using Landsat Ecosystem 1944 Disturbance Adaptive Processing System LEDAPS (Masek et al., 1945 2006). We masked out cloud pixels based on the quality assessment 1946 (QA) band provided with the Landsat 7 database, using the C function 1947 1948 of mask algorithm (CFMask) (Foga et al., 2017). Finally, for each 23 1949 x 23 m national grid pixel, we calculated the median values for each Landsat band (Kennedy et al., 2018). 1950



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

#### **2.2.5. SAR Variables**

1955 We used SAR data from PALSAR-2/PALSAR from the Advanced 1956 Land Observing Satellite (ALOS) and Advanced Land Observing 1957 Satellite-2 (ALOS-2) freely available at the global level online from the Japan Aerospace Exploration Agency (JAXA) at 25 x 25 m 1958 resolution. We rescaled the raw backscattering coefficients for each 1959 1960 polarization HH and HV for the year 2007 to the 23 x 23 m pixel of the national grid. For more information on this data we refer to 1961 https://www.eorc.jaxa.jp/ALOS/en/palsar fnf/fnf index.htm 1962

1963 (accessed on: 05-11-2019)

1964

#### 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. (2012) 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. (2020).

1971 Soil variables were from the harmonized soil geodatabase of Europe 1972 (European Soil Database v2.0 - 2004) (Penagis et al., 2004). The 1973 subsoil available water capacity and topsoil available water capacity 1974 soil variables used for this study were selected using the optimization 1975 phase described in Chirici et al. (2020).

#### **2.3. Forest Masks**

1977 We obtained five FMs available for the entire Italian territory that potentially reflect the forest FAO FRA definition (FAO, 2010). These 1978 masks can be divided into two main categories: i) FMs obtained by 1979 1980 semi-automated classification of remotely sensed data; ii) FMs obtained by manual delineation and classification of fine-resolution 1981 1982 images. All the FMs were first reprojected in the WGS 84 / UTM zone 1983 32 North (EPSG:32632) reference system to make them comparable 1984 and then resampled at the 23 x 23 m resolution of the national grid 1985 resulting to produce five comparable FMs.

1986

#### 2.3.1. National Forest Mask (NFM)

We used the national forest mask (NFM) which is based on the mosaic 1987 1988 of local forest maps produced by manual photointerpretation by the 1989 Italian regional forest authorities (D'Amico et al., 2021). The mosaic 1990 was constructed by merging 16 fine resolution forest maps with nominal reference scales varying between 1:5,000 and 1:25,000 and 1991 five land use maps specifically filtered to produce forest cover maps. 1992 All the maps were based on manual photointerpretation of aerial 1993 orthophotos. The local forest maps were reclassified into Boolean 1994 1995 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-1996 1997 resolution regional forest maps resampled at the 23 x 23 m national grid resolution. The mask is also available on-line at www.forestinfo.it 1998

## 19992.3.2. Copernicus Land Monitoring System (CLMS)2000Forest Mask

To construct the Copernicus FM, we first used the 2012 Forest Type map (https://land.copernicus.eu/pan-european/high-resolutionlayers/forests/forest-type-1/status-maps/2012?tab=download) that uses the Tree Cover Density layer (https://land.copernicus.eu/paneuropean/high-resolution-layers/forests/tree-cover-density/status-

2006 maps/2012?tab=download) (accessed on: 05-11-2020) to classify all 2007 20 x 20 m pixels of European lands as forest when the tree cover 2008 density is at least 10% and when such pixels are aggregated into a 2009 continuous patch of at least 0.52 hectares (Langanke et al., 2017). We 2010 excluded pixels in agricultural and urban contexts from the Forest Type map, using the Forest Additional Support Layer also available 2011 2012 from Copernicus at https://land.copernicus.eu/pan-european/high-2013 resolution-layers/forests/forest-type-1/status-

2014 maps/2012?tab=download (accessed on: 05-11-2020). The resulting
2015 map reflects as closely as possible the international forest definition in
2016 a raster layer having 23 x 23 m resolution

2017

#### 2.3.3. JAXA Forest Mask

2018JAXA constructed an FM for the reference years  $2007\pm1$  with a spatial2019resolution of 25 x 25 m based on the HV-polarization backscatter2020images acquired by the PALSAR and PALSAR 2 sensors carried by2021the ALOS and ALOS2 satellites. The JAXA declares to adopt the FAO2022forest definition (JAXA, 2016) and is available online at

- https://developers.google.com/earthengine/datasets/catalog/JAXA A 2023 LOS PALSAR YEARLY SAR (accessed on: 05-11-2020).
- 2024

#### 2025 2.3.4. Hybrid Global Forest Mask 2000 (FM00)

2026 Schepaschenko et al. (2015) constructed a global FM using a hybrid approach combining multiple local, national, and global datasets into 2027 2028 a single product. This map was constructed by converting the global 2029 forest probability map into a forest/non-forest map using a threshold 2030 calculated for each country. The threshold selected for this study 2031 produced area estimates that matched as closely as possible the official FAO forest area statistics. We characterized this map as "FM00". The 2032 2033 map has a spatial resolution of 1 x 1 km, was produced for the reference year 2000, and is available online at https://application.geo-2034 wiki.org/branches/biomass/ (accessed on: 05-11-2020). 2035

2036

#### 2.3.5. Corine Land Cover 2006 (CLC06)

2037 The CORINE Land Cover (CLC) project was initiated in 1990 by the 2038 European Environmental Agency (EEA) (Büttner et al., 2004) and has been updated in 2000, 2006, 2012, and 2018 to monitor land-use 2039 2040 changes in the 39 participating countries (EEA, 2007). It consists of land cover maps based on a nomenclature system of 44 classes 2041 2042 produced by photointerpretation of fine-resolution satellite imagery. 2043 CLC uses a MMU of 25 hectares and a MMW of 100 m. For this study, 2044 we acquired the CLC map for the reference year 2006±1 (referred to 2045 as "CLC06") obtained by photo-interpretation of SPOT-4/5 and IRS

P6 LISS III dual data images (EEA, 2007) and available online in vector format at https://land.copernicus.eu/pan-european/corine-landcover/clc-2006?tab=download (accessed on: 05-11-2020). To derive the CLC mask, we first rasterized the vector product to the 23 x 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.

2053

#### 2.4. Overview of the Method

A concise overview of the methodology followed is presented: i) a 2054 2055 wall-to-wall GSV map was constructed using a random forests model 2056 with the NFI plot GSV data and the predictor variables; ii) the 2057 accuracies of the five FMs were assessed; iii) the wall-to-wall GSV 2058 map was masked in turn with each of the five FMs, obtaining five 2059 masked GSV maps; iv) for each masked GSV maps we estimated the 2060 mean and total GSV with the model-assisted regression estimator, at 2061 the national and regional level; v) we compared model-assisted 2062 estimations for each FM with the official estimation from the Italian NFI, in terms of correlation coefficient; vi) we assessed the 2063 relationship between FMs accuracy and GSV estimates in terms of the 2064 2065 correlation coefficient.

2066

#### 2.5. Wall-to-Wall National GSV Map

2067 To estimate the effects of FM accuracy on the model-assisted GSV2068 estimates, we constructed a GSV map consisting of GSV predictions

for all 23 x 23 m pixels of the national grid (569,769,690 pixels) using 2069 the random forests (RF) prediction technique with the NFI plot GSV 2070 data and the predictor variables described in Table 1. RF was 2071 2072 optimized following Chirici et al. (2020) by selecting the combination 2073 of predictor variables and parameter values (ntree and mtry) that 2074 minimized the root mean square error (RMSE) calculated using the 2075 leave one out cross-validation (LOOCV) technique (McRoberts et al., 2076 2015). RMSE was calculated as:

2077 
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$
 (1)

2078 where n is the number of third-phase NFI plots (i.e., 6782),  $y_i$  is the i-2079 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 2080 from LOOCV was used to predict the GSV for all N pixels of the study 2081 area to produce a 23 x 23 m resolution GSV map. The model fitting 2082 and optimization phase was performed using the randomForest 2083 package within the statistical software package R 3.6.3 (Devarriva et 2084 2085 al., 2020) (https://www.r-project.org, accessed on: 05-11-2020). For 2086 the 6,782 NFI plots, the pixel-level GSV predictions ranged between 0 and 690 m3 ha-1 with a standard deviation of 68.5 m3 ha-1 while 2087 2088 the original NFI values ranged between 0.3 and 701 m3 ha-1 with a 2089 standard deviation of 147 m3 ha-1. The map was found to have a 2090 mean deviation of -4.3 m3 ha-1.

#### 2091 **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:

2097 Overall Accuracy = 
$$\frac{\sum True \ positive + \sum True \ negative}{\sum \ Total \ population}$$
 (2)

2098 
$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$
 (3)

2099 Where:

$$2100 \quad p_0 = \text{Overall Accuracy}$$

2101 
$$p_e = \frac{1}{N^2} \sum_k \sum True \text{ positive } * \sum True \text{ negative}$$
 (4)

2102 for k categories and N observations.

2103 Precision = 
$$\frac{\sum True \ positive}{\sum True \ positive + \sum False \ positive}$$
 (5)

2104 Recall = 
$$\frac{\sum True \ positive}{\sum True \ positive + \sum False \ negative}$$
 (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 (Devarriva et al., 2020; Jaafor et al., 2012). For this reason, the model accuracy comparison should focus on recall as per Equation (6) and, most importantly, precision as per Equation(5).

### 2114 2.7. Impact of FMs Accuracy on Model-Assisted GSV 2115 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) (Särndal et al., 1992; 2003; Breidt et al., 2009). An initial estimate of GSV can be calculated from the masked GSV maps as,

2123 
$$\hat{\mu}_{\text{initial}} = \frac{1}{n} \sum_{i=1}^{N} \hat{y}_i \tag{7}$$

2124 where *N* is the number of forest pixels within the masked GSV map 2125 and  $\hat{y}_t$  is the GSV prediction obtained using the RF model for the i-th 2126 pixel. However, this estimator may be biased because of systematic 2127 prediction error. The bias can be estimated as,

2128 
$$\widehat{\text{Bias}}(\widehat{\mu}_{\text{initial}}) = \frac{1}{n} \sum_{j=1}^{n} (\widehat{y}_j - y_j)$$
(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,

2134 
$$\hat{\mu}_{ma} = \hat{\mu}_{initial} - \widehat{Bias}(\hat{\mu}_{initial}) = \frac{1}{N} \sum_{i=1}^{N} y_i - \frac{1}{n} \sum_{j=1}^{n} (\hat{y}_j - y_j)$$
 (9)

- 2135 where ma denotes model-assisted,  $\hat{\mu}_{ma}$  is the estimate of mean GSV
- 2136 for the given masked GSV map, N is the number of forest pixels within

the masked GSV map,  $\hat{y}_l$  is the GSV prediction obtained using the RF

2138 model for the i-th pixel. The standard error (SE) for the estimator is:

2137

2139 
$$SE(\hat{\mu}_{ma}) = \sqrt{\frac{1}{n(n-1)} \sum_{j=1}^{n} (e_j - \hat{e}_j)^2}$$
 (10)

2140 where *n* is the NFI sample size,  $e_j = \hat{y}_j - y_j$  and  $\hat{e}_j = \frac{1}{n} \sum_{j=1}^{n} e_j$ .

2141 Similarly, the model-assisted estimator for the GSV total was:

2142 
$$\hat{\tau}_{ma} = \sum_{i=1}^{N} y_i - \frac{N}{n} \sum_{j=1}^{n} (\hat{y}_j - y_j)$$
(11)

2143 where  $\hat{\tau}_{ma}$  is the estimate of total GSV for the given GSV-masked 2144 map, N the number of pixels within the masked GSV map,  $y_i$  the GSV 2145 prediction obtained using the RF model for i-th pixel. The SE for the 2146  $\hat{\tau}_{ma}$  is given by d'Oliviero et al. (2012):

2147 
$$SE(\hat{\tau}_{ma}) = \sqrt{N^2 \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{j=1}^n \frac{(e_j - \hat{e}_j)^2}{n-1}}$$
 (12)

2148 where N is the population size, n is the NFI sample size,  $e_j = \hat{y}_j - y_j$ 2149 and  $\hat{e}_j = \frac{1}{n} \sum_{j=1}^n e_j$ .

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

$$2156 \qquad \hat{E}_{ma} \pm t_n * SE(\hat{E}_{ma}) \tag{13}$$

where  $\hat{E}_{ma}$  denotes either the model-assisted estimate of mean GSV 2157 or total GSV,  $SE(\hat{E}_m)$  is the SE of  $\hat{E}_{ma}$ , and the factor th depends on 2158 2159 the desired significance level and the distribution of the response 2160 variable. For most distributions and applications, tn = 2 produces an 2161 approximate 95% confidence interval (McRoberts et al., 2008). For 2162 purposes of constructing confidence intervals, the focus of the study 2163 was the estimation of mean and total GSV and the SEs using the 2164 model-assisted regression estimators. To compare the GSV estimates produced with the five masked GSV maps and the NFI estimates at 2165 2166 national and regional levels, we used the t statistic calculated as 2167 follows:

2168 
$$t = \frac{\hat{E}_{ma} - \hat{E}_{NFI}}{\sqrt{SE^2(\hat{E}_{ma}) + SE^2(\hat{E}_{NFI})}}$$
(14)

2169 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 2170 estimate of mean GSV or total GSV, and  $SE^2(\hat{E}_{ma})$  and  $SE^2(\hat{E}_{NFI})$ 2171 are the squares of the SEs of the estimates. Values of |t| > 2 indicates 2172 that the two estimates are statistically significantly different. 2173 2174 Correlations for estimates of both mean and total estimates and the 2175 corresponding NFI estimates in terms of Pearson correlation coefficient ( $\hat{\rho}_{Mean}, \hat{\rho}_{Total}$ ) were also calculated. 2176

In addition, we calculated relative efficiency (RE) to assess the quality
of the model-assisted estimators, compared to the SE obtained by the
NFI (Chirici et al., 2020), both at national and regional scales. RE was
calculated as:
2181 
$$RE = \frac{V\widehat{a}r(\widehat{E}_{NFI})}{V\widehat{a}r(\widehat{E}_{ma})}$$
(15)

2182 where  $Var(\hat{E}_{NFI})$  and  $Var(\hat{E}_{ma})$  are the estimated variances of the 2183 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 (Moser et al., 2016). 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 (Chirici et al., 2020).

2189 Finally, we evaluated the relationship between the accuracies of the 2190 FMs (in terms of overall accuracy,  $\kappa$ , precision and recall) and the SEs 2191 of the model-assisted estimates for the NUTS2 administrative level 2192 using the Pearson correlation coefficient ( $\hat{\rho}$ ).

#### 2193 **3. Results**

2194

# **3.1. Forest Masks Accuracy Assessment**

2195 At the national level, the most accurate FM was the NFM with an 2196 underestimation against the NFI estimates of only -2%, followed by 2197 the CLC06 with -3%, JAXA with -4%, CLMS with +16%, and FM00 2198 with +51%. The same ranking was obtained from the comparison with 2199 IUTI in terms of OA, k, and precision (Table 2). For 17 of the 20 regions, the NFM was the most accurate, followed by the CLMS FM 2200 in two regions, and CLC06 in the remaining region. The confusion 2201 matrices for each one of the five FMs are shown in Figure 3. 2202



- 2204 Figure 3. Confusion matrices of each forest mask.
- 2205

2206Table 2. Accuracy assessment for the five forest masks (FMs) based on the2207confusion matrices with the IUTI.

Mask	Accuracy					
	OA	к	Precision	Recall		
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		

2208

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).



2215Figure 4. Comparison of four accuracy metrics among the FMs, calculated at<br/>regional level (NUTS2).

2217

2214

# 2218 **3.2. GSV Model-Assisted Estimations**

In Figure 5, the GSV map of Italy produced with the random forestsmodel is reported.



2221

Figure 5. Growing stock map of Italy generated with random forests model.
GSV in m3 ha-1. On the right, a detail of the GSV map masked with the five forest masks.

2226 For the five masked GSV maps,  $\hat{\mu}_{ma}$  ranged between 125 (CLMS) and 135 (NFM), m3 ha-1 with a  $SE(\hat{\mu}_{ma})$  between 1.1 and 1.3 m3 2227 2228 ha-1. For comparison, the design-based estimation of mean GSV from the NFI was 131 m3 ha-1 with a SE of 1.6 m3 ha-1. Three of the five 2229 2230 GSV-masked maps (NFM, CLC06, JAXA) produced estimates that 2231 were not statistically significantly different from the NFI estimate. The 2232 value of  $\hat{\tau}_{ma}$  ranged between 1321 (JAXA) and 1525 (CLMS) millions m3, with  $SE(\hat{\tau}_{ma})$  between 13 (NFM) and 17 (JAXA) 2233 2234 million m3, while the official estimate from the NFI was 1366 million 2235 m3 with SE of 14 million m3, demonstrating a general trend towards

- 2236 overestimation of total volume (Table 3). The differences between the
- 2237 total GSV estimate for two of the five masked GSV maps (NFM,
- 2238 CLC06) and the NFI estimate were not statistically significantly
- different from 0.
- 2240

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

Forest mask			Model-assisted and NFI estimates (m <sup>3</sup> )				
	μ̂ <i>ma</i>	$SE(\hat{\mu}_{ma})$	$t(\hat{\mu})$	τ̂ <sub>ma</sub>	$SE(\hat{\tau}_{ma})$	$t(\hat{\tau})$	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

2244

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  $\hat{\tau}_{ma}$ .

2251Table 4. Coefficient of correlation between the mean and total model-assisted2252estimate and NFI estimates for administrative NUTS2 regions (\*p-value=0;2253\*\*p-value < 0.001).</td>

2	2	5	Λ
4	L	J	4

Forest mask	$\widehat{\rho}_{Total}$	$\widehat{\mathbf{\rho}}_{Mean}$
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 2256 2257 model-assisted estimates and the NFI estimate were not statistically significantly different from 0 for the NFM masked GSV map, for 15 2258 regions for CLMS and JAXA, for 14 regions for CLC06, and for 10 2259 regions for FM00. Similar results were obtained for  $\hat{\tau}_{ma}$  for which the 2260 differences for 16 of 20 regions were not statistically significantly 2261 2262 different from 0 for the NFM masked GSV map, 15 for CLC06 and 2263 JAXA, six for CLMS, and two for FM00. The regions that always 2264 showed a statistically significant difference between the model-2265 assisted estimates and the official NFI turned out to be the islands 2266 (Sardegna, Sicilia) and two regions (Puglia, Umbria), while those which never did were seven, distributed in northern and central Italy. 2267 2268 RE exceeded 1 for most regions, regardless of the FM used. RE < 12269 was observed in one region for the CLMS and FM00 masks (Toscana), 2270 two regions for the CLC06 mask (Toscana, Emilia Romagna), and 2271 four regions for the JAXA mask (Toscana, Emilia Romagna,

- 2272 Sardegna, Umbria). The only masked GSV map that leads to RE coefficient always >1 was the NFM. 2273
- 2274 3.3. Relationship Between FMs Accuracy and GSV Estimates
- The relationship between the accuracies of the FMs and the SEs of the 2275
- 2276 estimates with the model-assisted estimator is presented in Table 5.
- 2277 The correlation was calculated for the 20 administrative regions.
- 2278

2279 Table 5. Correlation coefficient between the accuracy metrics and the SEs of 2280 estimates for each FM. The overall values were calculated based on all five FMs together.

22	ð	2

Forest mask	ρ					
r or est mask	Overall Accuracy	к	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		

#### 2283 4. Discussion

The aim of this study was to assess the effects of using different FMs 2284 available for Italy for the area-based estimation of GSV. We first 2285 2286 constructed a pixel-level GSV map for the entirety of Italy based on 2287 the procedure recently proposed by Chirici et al. (2020). We then acquired five different FMs and, after evaluating their accuracies 2288

against an independent dataset (IUTI), we used them to mask out nonforest 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 designbased NFI estimates.

2294 Four of the five FMs achieved overall accuracies > 85%, based on the 2295 2008 land use classification of IUTI points, with the CLC06 and NFM outperforming the other products. At the national level, the FM that 2296 achieved the greatest overall accuracy,  $\kappa$  and precision was the NFM, 2297 followed by the CLC06. Despite the greatest recall (0.91) achieved, 2298 2299 the FM00 was affected by systematic overestimation of the regional forest area due to the original coarse resolution (Schepaschenko et al., 2300 2301 2015) which made this FM unsuitable for GSV estimation.

2302 In contrast, the JAXA FM produced the smallest recall (0.74), most 2303 probably because the SAR backscatter in the HV polarization is 2304 relatively insensitive to Mediterranean vegetation (D'Amico et al., 2305 2021; Bartsch et al., 2020) which probably caused an underestimation of the forest area. The photointerpreted FMs, CLC06 and NFM, had 2306 2307 the greatest precision. This is an expected result because forest land 2308 use identification is typically done by local experts. However, CLC06 2309 produced less precision than the NFM because it was implemented for monitoring land cover, not land uses, adopting a MMU and a crown 2310 2311 cover threshold greater than that adopted by the INFC 2005 (Seebach et al., 2011; Vizzarri et al., 2015). In fact, the CLC project did not map 2312

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.

2319 At the regional level, OA was greater than 85% for 18 regions for the 2320 NFM mask, followed by the CLMS mask (14 regions), the CLC06 2321 mask (12 regions), the JAXA mask (3 regions), and the FM00 mask 2322 (1 region). Regardless of the FM used, the greatest uncertainty was 2323 found in the southern regions and the islands (Campania, Calabria, Abruzzo, Basilicata, Sardegna, Sicilia), most probably because of the 2324 2325 complex Mediterranean formations and complex agroforestry landscape tiles that characterized these regions where the NFI 2326 2327 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 probability of forest misclassifications, regardless of the FM considered.

Conversely, the forests bordering other land uses, along rivers, and inthe coastal and rural contexts are typically characterized by a sparse

canopy, which makes them more difficult to correctly classify, evenby 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.

2343 Regarding the model-assisted GSV estimates, although all the masked 2344 GSV maps overestimated total GSV, the NFM masked GSV map was 2345 most accurate as a trade-off between the national and regional GSV 2346 and the SE of estimates. The general overestimation was caused by the trend of the prediction model to overpredict GSV for pixels with small 2347 observed GSV values. (i.e., GSV < 250 m3 ha-1). This evidence, 2348 along with the limited GSV that characterizes Italian forests, caused 2349 2350 the general overestimation at the national level. One possible solution is to increase the performance of the model, for example, by 2351 2352 integrating ALS metrics which is a well-established data source for 2353 enhancing GSV predictions (Næsset et al., 2004; Kangas et al., 2018; Næsset et al., 2014). Both the CLMS and FM00 masked GSV maps 2354 2355 suffered from systematic prediction error which caused the 2356 overestimation of  $\hat{\tau}_{ma}$ , both nationally and regionally. For the CLMS 2357 masked GSV map, this can be caused by the inclusion of many 2358 agricultural and rural areas that occur in Italy (Langanke, 2017), and 2359 for FM00 because of the original coarse spatial resolution (1 x 1 km). 2360 The differences between the total GSV model-assisted estimates and 2361 the official NFI estimate for two of the five masked GSV maps (NFM, CLC06) were statistically significantly different from 0. At the 2362 2363 national level, the mean GSV estimates were comparable for all maps, 2364 except for the GSV map masked with the FM00 mask. The JAXA 2365 masked GSV map produced the same value as the NFI for mean GSV 2366 but underestimated the total due to the underestimation of forest area. 2367 However, the SEs were almost comparable for all the GSV-masked 2368 maps considered. The SE is mainly affected by the number of NFI 2369 plots used for building the model and calculation of the correction 2370 term in the estimator. Despite the differences among the FMs, the NFI 2371 plots falling within the forested portions of the FMs were similar, ranging between 6100 (CLMS) and 5800 (JAXA). Differences in the 2372 number of plots selected by each FM are likely to be concentrated at 2373 2374 the forest edge, where maps are more prone to classification errors. 2375 These results confirm the findings of Esteban et al. (2020), suggesting 2376 that the FM effects on area estimates are more important than the 2377 effects of field plot sampling variability on the uncertainty of the mean 2378 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 complexpatchwork in the landscape of a rural region.

2387  $SE(\hat{\tau}_{ma})$  was smaller than  $SE(\hat{\tau}_{NFI})$  for 16 regions, which represent 2388 70% of the Italian territory. The regions with the greatest  $SE(\hat{\tau}_{ma})$ 2389 were Puglia, Valle d'Aosta, Molise, Basilicata, and Marche  $(SE(\hat{\tau}_{ma}) > 5\%)$  probably because of the small number of NFI plots in 2390 2391 these regions. Nevertheless, with the use of the model-assisted estimation approach, it was possible to decrease the error of the 2392 estimates with respect to the NFI estimates, both at the national 2393 2394 (NUTS1), and regional levels (NUTS2).

- 2395 Regarding the relationship between the FM accuracy and the SEs of 2396 the estimates, we found small correlation coefficients, in particular 2397 with the overall accuracy. The SE depends primarily on the sample 2398 size, which is less affected by the accuracy of the FMs, as reported by 2399 Esteban et al. (2020). The accuracy metric was more correlated with 2400 the SE of the estimates than was the recall, followed by the precision. 2401 This is an expected result because these metrics are strictly related to 2402 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 2403 (CLMS) was also the FM that included the greatest number of NFI 2404 2405 plots.
- 2406 However, the negative correlation with the other accuracy metrics
- 2407 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. (2016). Another option would be to calibrate the FMs using the NFI data as per Næsset et al. (2007).

2412 In conclusion, the differences in the accuracies of the FMs led to 2413 different GSV estimates, although the SEs were almost comparable. 2414 The smallest GSV difference against the official NFI estimate was 2415 obtained by the most accurate FMs, i.e., the NFM. This is likely due 2416 to the correct classification of the main, dense forests, which have the 2417 largest amount of volume and subsequently make the greatest contribution in the model-assisted estimation. Presumably, forest 2418 misclassification occurs mainly along the margins and in boundary 2419 2420 areas between different land uses.

#### 2421 **5.** Conclusions

This paper presents a comparative analysis of the impacts of different
forest masks on the 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.

2436 Despite the small correlation coefficients, we found a negative 2437 relationship between forest mask accuracy and the standard error of 2438 the GSV estimate, demonstrating that the accuracy of the FM must be 2439 considered in the GSV estimation through the model-assisted 2440 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.

2446 In conclusion, we recommend using the NFM, both at regional and 2447 national levels, for studies aimed at estimating forest parameters related to variables such as forest area, GSV, AGB, and carbon stock. 2448 2449 However, it is plausible to assume that as the accuracy of the model 2450 predictions increases thanks to the growing availability of 3D data, the 2451 NFM will always produce more accurate and precise estimates of total 2452 GSV. In this regard, we hope that in the future, wall-to-wall ALS 2453 coverage will be finally available in Italy, to enhance the prediction of forest variables with even greater accuracy. 2454

- 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.
- 2459

### 2460 **References**

- 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, doi:10.1016/j.foreco.2013.07.004.
- 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, doi:10.1016/j.rse.2015.08.029.
- 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 Sentinel1 and Sentinel-2 data. Remote. Sens. Environ. 2020, 237, 111515,
  doi:10.1016/j.rse.2019.111515.
- 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, doi:10.1016/j.isprsjprs.2016.01.011.
- 2477 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, 2480 2009; pp. 103–119, doi:10.1016/S0169-7161(09)00227-2.
- 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 at
  http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.618.9
  940&rep=rep1&type=pdf (accessed on 06/02/2021).
- 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, doi:10.1016/j.jag.2019.101959.
- 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 06/11/2020).

- 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.
- 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, doi:10.1016/j.rse.2013.09.006.
- 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, doi:10.1016/j.eswa.2019.112866.
- 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, doi:10.1016/j.rse.2012.05.014.
- 2515 Dostálová, A.; Hollaus, M.; Milenković, M.; Wagner, W. FOREST
  2516 AREA DERIVATION FROM SENTINEL-1 DATA. Isprs Ann.
  2517 Photogramm. Remote. Sens. Spat. Inf. Sci. 2016, III-7, 227–233,
  2518 doi:10.5194/isprs-annals-iii-7-227-2016.
- 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,
  doi:10.3390/rs12203360.
- European Environmental Agency. Environmental Statement; Office for
  Official Publications of the European Communities:
  Luxembourg, 2007; ISBN 978-92-9167-936-2.
- Eysn, L.; Hollaus, M.; Schadauer, K.; Pfeifer, N. Forest Delineation
  Based on Airborne LIDAR Data. Remote. Sens. 2012, 4, 762–
  783, doi:10.3390/rs4030762.
- FAO, UNCCD. Sustainable Financing for Forest and Landscape
  Restoration: The Role of Public Policy Makers; FAO, Rome
  2015; p. 12.
- FAO. Global forest resources assessment 2010. Terms and definition.
  Working paper 144/E. 2010. Available online:

- 2535
   http://www.fao.org/3/am665e/am665e00.pdf
   (accessed on 07/10/2020).
- 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,
  doi:10.1198/108571106x130548.
- 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, doi:10.1016/j.rse.2017.03.026.
- 2546 Fritz, S.; See, L. Comparison of land cover maps using fuzzy
  2547 agreement. Int. J. Geogr. Inf. Sci. 2005, 19, 787–807,
  2548 doi:10.1080/13658810500072020.
- Giannetti, F.; Chirici, G.; Gobakken, T.; Næsset, E.; Travaglini, D.; 2549 Puliti, S. A new approach with DTM-independent metrics for 2550 forest growing stock prediction using UAV photogrammetric 2551 2552 Remote. Sens. Environ. 2018. 213, data. 195-205. 2553 doi:10.1016/j.rse.2018.05.016.
- 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,
  doi:10.1016/j.ecolind.2020.106513.
- 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, doi:10.1016/j.rse.2004.09.005.
- 2562Goodbody, T.R.H.; Coops, N.C.; White, J.C. Digital Aerial2563Photogrammetry for Updating Area-Based Forest Inventories: A2564Review of Opportunities, Challenges, and Future Directions.2565Curr. For. Rep. 2019, 5, 55–75, doi:10.1007/s40725-019-00087-25662.
- 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 21stCentury Forest Cover Change. Science 2013, 342, 850–853.

- 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.
- 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, doi:10.1080/01431160903022894.
- 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.
- 2582 doi:10.1016/j.rse.2017.04.004.
- Hoyos, A.P.; Rembold, F.; Kerdiles, H.; Gallego, J. Comparison of
  Global Land Cover Datasets for Cropland Monitoring. Remote.
  Sens. 2017, 9, 1118, doi:10.3390/rs9111118.
- 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, doi:10.1016/j.foreco.2015.10.018.
- INFC. 2007. 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. Available online: http://www.infc.it
  (accessed on 05/11/2019).
- 2597 Jaafor, O.; Birregah, B. KNN-LC: Classification in Unbalanced 2598 using KNN-Based Algorithm and Datasets а Local 2599 Centralities In: Adjallah K., Birregah B., Abanda H. (eds) Data-Driven Modeling for Sustainable Engineering. ICEASSM 2017. 2600 Lecture Notes in Networks and Systems, vol 72. Springer, Cham. 2601 2602 https://doi.org/10.1007/978-3-030-13697-0 7D'Oliveira, M.V.; Reutebuch, S.E.; McGaughey, R.J.; Andersen, H.-E. Estimating 2603 forest biomass and identifying low-intensity logging areas using 2604 airborne scanning lidar in Antimary State Forest, Acre State, 2605 Western Brazilian Amazon. Remote. Sens. Environ. 2012, 124, 2606 2607 479-491, doi:10.1016/j.rse.2012.05.014.

- 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), Japan, 2016.
- 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, doi:10.1080/02827581.2017.1416666.
- 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.
- 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,
  doi:10.3390/rs10050691.
- Langanke, T. Copernicus Land Monitoring Service–High Resolution
   Layer Forest: Product Specifications Document 38;Copernicus
   team at EEA, 2017.
- 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, doi:10.1016/s20953119(15)61303-x.
- Liaw, A.; Wiener, M. Classification and regression by randomForest.
  Nucleic Acids Res. 2002, 5, 983–999,
  doi:10.1023/A:101093340432.
- 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.
- 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,
  doi:10.3354/cr01121.
- 2644 McRoberts, R.E. Probability- and model-based approaches to 2645 inference for proportion forest using satellite imagery as ancillary

- 2646data. Remote. Sens. Environ. 2010, 114, 1017–1025,2647doi:10.1016/j.rse.2009.12.013.
- 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, doi:10.1016/j.rse.2017.09.036.
- 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, doi:10.1109/JSTARS.2012.2227299.
- 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.
- McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. Remote. Sens. Environ. 2007, 110, 412–419, doi:10.1016/j.rse.2006.09.034.
- 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, doi:10.1139/cjfr-2016-0064.
- 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, doi:10.1093/forestry/cpw041.
- 2672 Næsset, E. Airborne laser scanning as a method in operational forest
  2673 inventory: Status of accuracy assessments accom-plished in
  2674 Scandinavia. Scand. J. For. Res. 2007, 22, 433–422.
- 2675 Næsset, E. Area-Based Inventory in Norway–From Innovation to an
  2676 Operational Reality. In Forestry Applications of Airborne Laser
  2677 Scanning Concepts Case Stud.; Maltamo, M., Næsset, E.,
  2678 Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands,
  2679 2014; pp. 215–240, doi:10.1007/978-94-017-8663-8\_11.
- 2680 Næsset, E.; Gobakken, T. Estimation of above- and below-ground
  2681 biomass across regions of the boreal forest zone using airborne
  2682 laser. Remote. Sens. Environ. 2008, 112, 3079–3090,
  2683 doi:10.1016/j.rse.2008.03.004.

- Næsset, E.; Gobakken, T.; Holmgren, J.; Hyyppä, H.; Hyyppä, J.; 2684 Maltamo, M.; Nilsson, M.; Olsson, H.; Persson, Å.; Söderman, 2685 2686 U. Laser scanning of forest resources: The nordic experience. 2687 Scand. L For Res 2004. 19, 482-499, doi:10.1080/02827580410019553. 2688
- 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, doi:10.1016/j.jag.2007.02.004.
- 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, doi:10.1016/j.rse.2016.10.022.
- Øivind, D.T.; Salberg, A.-B.; Kermit, M.; Øystein, R.; Gobakken, T.; 2699 2700 Næsset, E.; Aarsten, D. Tree species classification in Norway 2701 from airborne hyperspectral and airborne laser scanning data. 2702 Eur. J. Remote. Sens. 2018. 51, 336-351. 2703 doi:10.1080/22797254.2018.1434424.
- 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, doi:10.1016/j.rse.2019.111492.
- P.; Gasparini, F.; De Natale, L.; Di Cosmo, C.; Gagliano, I.; Salvadori,
  G.; Tabacchi e.V.; Tosi. INFC, 2009–I caratteri quantitativi–parte
  1, vers. 2. Inventario Nazionale delle Foreste e dei Serbatoi
  Forestali di Carbonio; MiPAAF–Ispettorato Generale Corpo
  Forestale dello Stato, CRA-MPF: Trento, Italy.,2009
- 2713 Panagos, P. The European Soil Database; GEO: 2006; pp. 32–33.
- Penman, J.; Gytarsky, M.; Hiraushi, T.; Krug, T.; Kruger, D.; Pipatti, 2714 2715 R.; Buendia, L.; Miwa, K.; Ngara, T.; Tanabe, K.; et al. Good Practice Guidance for Land Use, Land Use Change and Forestry. 2716 Chapter 3: Annex 3A.1 Biomass Default Tables for Section 3.2 2717 Forest Land Good Practice Guidance for Land Use, Land-Use 2718 Change and Forestry; The Institute for Global Environmental 2719 2720 Strategies for the IPCC and the Intergovernmental Panel on Climate Change, Hayama: Kanagawa, Japan, 2003; p. 21. 2721

- 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, doi:10.3832/efor2363-014.
- 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,
  doi:10.1016/j.rse.2016.06.004.
- 2731 Romano D, Arcarese C, Bernetti A, Caputo A, Condor RD, Contaldi
  2732 M, Lauretis R, Di Cristofaro E, Federici S, Gagna A, Gonella B,
  2733 Lena F, Liburdi R, Taurino, E., Vitullo, M. Italian Greenhouse
  2734 Gas Inventory 1990–2009. National Inventory Report; ISPRA:
  2735 Rome, Italy, 2011.
- 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, 12/02/2016; p. 4, doi:10.1109/whispers.2016.8071665.
- 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, doi:10.1007/s13595-016-0590-1.
- 2745 Särndal, C.-E.; Swensson, B.; Wretman, J. Model Assisted Survey
  2746 Sampling; Springer: New York, NY, USA, 1992; p. 402, p. 694.
- 2747 Särndal, C.-E.; Swensson, B.; Wretman, J. Model Assisted Survey
  2748 Sampling; editor, Springer: Berlin, Germany, 2003.
- 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,
  doi:10.1016/j.rse.2015.02.011.
- 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, doi:10.1016/j.envsci.2012.04.010.

- Seebach, L.M.; Strobl, P.; Miguel-Ayanz, J.S.; Gallego, J.; BastrupBirk, A. Comparative analysis of harmonized forest area
  estimates for European countries. Forests 2011, 84, 285–299,
  doi:10.1093/forestry/cpr013.
- 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 IV, 17
  March 2003; doi:10.1117/12.462423.
- 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 2011.
- 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, doi:10.1016/j.rse.2007.03.032.
- Vizzarri, M.; Chiavetta, U.; Chirici, G.; Garfi, V.; Bastrup-Birk, A.;
  Marchetti, M. Comparing multisource harmonized forest types
  mapping: A case study from central Italy. Iforest-Biogeosci. For.
  2782 2015, 8, 59–66. Available
  online:http://www.sisef.it/iforest/contents/?id=ifor1133-007
  (accessed on 05/11/2020).
- 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, doi:10.3390/f6124386.
- 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, doi:10.1016/j.jag.2005.10.001.
- 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, doi:10.1080/07038992.2016.1207484.

- Wittke, S.; Yu, X.; Karjalainen, M.; Hyyppä, J.; Puttonen, E. 2797 Comparison of two-dimensional multitemporal Sentinel-2 data 2798 2799 with three-dimensional remote sensing data sources for forest 2800 inventory parameter estimation over a boreal forest. Int. J. Appl. Obs. 2801 Geoinf. 2019, 76, 167-178, Earth 2802 doi:10.1016/j.jag.2018.11.009.
- 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,
  doi:10.1126/science.320.5879.1011a.
- 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, doi:10.1016/j.rse.2019.02.015.
- 2812

# **3.3. Paper III**

2814	Effects of lidar coverage and field plot data numerosity on forest
2815	growing stock volume estimation
2816	Giovanni D'Amico <sup>1,*</sup> , Ronald E. McRoberts <sup>2</sup> , Francesca Giannetti <sup>1</sup> ,
2817	Elia Vangi, <sup>1,3</sup> , Saverio Francini <sup>1</sup> , Gherardo Chirici <sup>1</sup>
2818	
2819	<sup>1</sup> Department of Agricultural, Food, Environmental and Forestry
2820	Sciences and Technologies, University of Florence, Firenze 50145,
2821	Italy;
2822	<sup>2</sup> Department of Forest Resources, University of Minnesota, Saint
2823	Paul, MN, 55108, USA;
2824	<sup>3</sup> Department of Biosciences and Territory, University of Molise,
2825	Campobasso, 86100, Italy
2826	* corresponding author: giovanni.damico@unifi.it
2827	Submitted to European Journal of Remote Sensing.

#### 2829 Abstract

2830 Forest parameter estimation is required to support the sustainable management of forest ecosystems, especially in the climate change 2831 2832 context. Currently, forest resource assessment is increasingly linked 2833 to auxiliary information obtained from active or passive remote 2834 sensing (RS) technologies. In forest parameter estimation, airborne 2835 laser scanning (ALS) data have been demonstrated to be an invaluable source of information. However, ALS data are often not available for 2836 2837 the entire forest area, whereas images from multiple satellite systems 2838 that are available free of charge offer new opportunities for large-scale 2839 forest surveys. This study aims to assess and estimate the contribution 2840 of field plot data and ALS data along with Landsat data to the precision 2841 of growing stock volume (GSV) estimates. We compared different 2842 approaches for model-assisted estimation of mean forest GSV per unit 2843 area using different proportions of the field sample data, ALS cover 2844 data, and wall-to-wall Landsat data. Model-assisted estimators were used with NFI sample data in a study area in Italy using 10 remote 2845 2846 sensing predictors, specifically the seven Landsat 7 ETM+ bands and three fine-resolution metrics based on ALS-derived canopy height. 2847 2848 We found that relative to the standard simple expansion estimator, the 2849 model-assisted estimators produced relative efficiency of 1.16 when using only Landsat data and relative efficiencies as great as 1.33 when 2850 2851 using increasing levels of ALS coverage.

- 2852 Keywords: Airborne Laser Scanning, Growing stock volume,
  2853 Landsat 7 ETM+, National Forest Inventory.
- 2854

#### 2855 **1. Introduction**

2856 Forests play an important role in mitigating the effect of climate 2857 change and making recreation and economic contributions to our 2858 society. Worldwide, there is an increasing demand to improve forest 2859 parameter estimation in support of monitoring the state of these 2860 ecosystems. Usually, forest parameter estimates are provided by a 2861 design-based approach with field data collected in the frameworks of traditional national forest inventories (NFIs) (Chirici et al., 2020; 2862 McRoberts et al., 2014; White et al., 2016). Estimates aggregated for 2863 2864 large geographic areas are requested by national and international reporting frameworks such as Forest Europe and FAO (FAO, 2020; 2865 FOREST EUROPE, 2015). However, NFI data are expensive, since 2866 2867 they require long and costly field campaigns, so a major scientific 2868 challenge is to develop new methods to produce useful forest information in a more cost-efficient way (White et al., 2016, Saarela 2869 2870 et al., 2018). In recent decades, advancements in earth observation 2871 technologies have opened the possibilities for using remotely sensed 2872 (RS) data to support forest inventories, so that the strategies to collect 2873 information and produce forest inventory estimates have changed 2874 consequentially (Waser et al., 2017; White et al., 2016; Saarela et al.,

2016; Chirici et al., 2020; Kangas et al., 2018). The exploitation of RS 2875 imagery became mandatory in deforestation and forest degradation 2876 surveys (REDD) because of the large cost of field surveys and forest 2877 inaccessibility in tropical and subtropical countries and remote areas 2878 2879 (e.g., Corona et al., 2014, Pagliarella et al., 2016). RS data in 2880 combination with field data can be used to increase the precision of 2881 large-scale forest inventory estimates in two different stages: the 2882 design stage and the estimation stage or both (Saarela et al., 2018). At the design stage, RS data can be used for stratification (McRoberts et 2883 2884 al., 2002) or probability sampling (Saarela et al., 2015; Lisańczuk et 2885 al., 2020). In the estimation stage, RS data can be used with either model-based inference (Gregoire, 1998; McRoberts et al., 2011) or 2886 design-based inference via stratified, post-stratified or model-assisted 2887 2888 estimation (Särndal et al., 1992).

2889 RS data can be acquired by different platforms, with different sensors, 2890 and at different resolutions, opening a vast array of methods useful for 2891 increasing the efficiency of forest parameter estimation (White et al., 2016; Saarela et al., 2018; Holm et al., 2017; Puliti et al., 2018). At 2892 2893 the national inventory scale, methods based on satellite images are still 2894 the most widely used because of the small cost and temporally high 2895 frequency associated with acquisition of satellite imagery. The main 2896 disadvantages of using satellite imagery compared to airborne laser 2897 scanning (ALS) data, a form of light detection and ranging (lidar) data collected by airplane or helicopter platforms, are spatial accuracy and 2898

resolution, both of which are substantially greater for ALS. In fact, in 2899 recent years, many studies have successfully used ALS data to assist 2900 in the estimation of forest biophysical parameters, as well as to provide 2901 accurate and precise estimates of growing stock volume (GSV) 2902 2903 (Montaghi et al. 2013, Kotivuori et al. 2016, White et al. 2016). ALS 2904 has particularly revolutionized forest inventories with its great 2905 potential to produce three-dimensional (3D) forest vertical structure 2906 data for prediction of a variety of forest attributes (Næsset 2002, 2907 Næsset et al. 2004, Hyyppä et al. 2008, Nilsson et al. 2017, Tompalski et al. 2019). Furthermore, high-performance computer servers 2908 2909 represent a shift toward the possibility of detailed forest mapping, which can accurately identify individual tree crowns useful in support 2910 2911 of forest management (Yun et al., 2021). However, ALS data are often 2912 limited by the cost of acquisition, processing, and data volume. As a 2913 result, complete ALS (or unmanned aerial vehicles) coverage is often impossible for large areas. Therefore, widespread and ready 2914 2915 availability of such data is inhibited (Li et al. 2019, Puliti et al., 2018), in contrast to some RS data that are freely available wall-to-wall (e.g., 2916 2917 satellite data).

2918 Despite the well-documented use of wall-to-wall auxiliary data with 2919 model-assisted (Corona et al., 2014), the use of one wall-to-wall 2920 auxiliary dataset in conjunction with another auxiliary dataset with 2921 only partial coverage in a rigorous uncertainty assessment framework 2922 has been investigated in only a small number of recent papers

(Gregoire et al., 2016; Puliti et al., 2018; Saarela et al., 2016; Saarela 2923 et al., 2018). Holm et al. (2017) demonstrated how the hybrid 3-phase 2924 2925 estimators can be used in situations where ALS data are used to relate ground plot measurements to lidar satellite observations (i.e., GLAS). 2926 2927 Holm et al. (2017) further demonstrated that hybrid 3-phase estimators 2928 facilitate the assessment of mean biomass density and variance that 2929 account for sampling variability and the model prediction uncertainty 2930 associated with two predictive models (i.e., ground plot-ALS models 2931 and ALS-GLAS models). Saarela et al. (2016) illustrate a novel and design-based, model-assisted attempt to exploit wall-to-wall satellite 2932 2933 information together with partial ALS information in the estimation of forest parameters from ground sample surveys. 2934

The objective of this study is to assess the impact of partially available,
fine resolution ALS data for model-assisted estimation of GSV. In
particular, two main research questions are investigated:

What is the effect of varying the ALS data forest coverage on GSVestimates?

What is the effect of pseudo-plots constructed in the ALS stratum usedto construct a GSV-RS model that is then applied to the entire studyarea?

To address these questions, we used an Italian dataset representing by
forests with ALS data, covering the Alpine, and Mediterranean
ecological regions, Landsat 7 ETM+ data, Canopy Height Model

(CHM) derived ALS data, and NFI plots for which GSV was measuredin the field.

#### **2948 2. Materials**

**2949 2.1. Study area** 

2950 The study was conducted in Italy (centered at  $42^{\circ}$  30' N and  $12^{\circ}$  30' E) 2951 in the forests for which ALS data are available, totaling 60,700 km<sup>2</sup> 2952 (D'Amico et al., 2021). The country is characterized by large 2953 vegetation variability due to its specific geographical and 2954 topographical conditions. The Italian peninsula has a typically flat 2955 coastal strip, a hilly part in the hinterland, and two main mountain 2956 ranges, the Alps in the north with peaks over 4800 m above sea level 2957 and the Apennines along the peninsula length.

2958 Italian forests and other wooded lands are mainly distributed in hilly and mountainous areas covering 104,675 km<sup>2</sup>, corresponding to 34% 2959 of the land area (INFC 2004, Tabacchi et al. 2007). Italian forests 2960 consist mainly of broadleaf species (about 68% of the total). The most 2961 2962 dominant tree species are Downy oak (Quercus pubescens), Pedunculated oak (O. robur), Turkey oak (O. cerris), Sessile oak (O. 2963 petraea), European beech (Fagus sylvatica), each exceeding area of 2964 one million hectares. The most common coniferous forests, especially 2965 2966 in the Alps, are those dominated by Norway spruce (Picea abies).

The Italian study area was tessellated into N=97,116,385 square grid cells each with area of 530 m2, equal to the NFI plot size (see section 2969 2.2.). The N grid cells constitute our target population U.

2970 **2.2. National Forest Inventory data** 

The Italian field reference data were measured in 2005 as part of the 2971 2972 2<sup>nd</sup> Italian NFI (Figure 1 A) (INFC 2004, Chirici et al. 2020). Data for 2618 circular, 530 m<sup>2</sup> NFI field plots for which ALS data were 2973 2974 acquired within five years of field measurement were available 2975 (Figure 1A). For each NFI plot, GSV (m<sup>3</sup> ha<sup>-1</sup>) for each callipered tree 2976 was predicted using species-specific allometric models developed by 2977 the NFI using tree DBH and tree height as independent variables 2978 (Tabacchi et al. 2011). The uncertainty of the allometric model 2979 predictions was considered negligible and ignored following previous 2980 results (McRoberts et al., 2016a, 2016b). The GSV of each NFI plot 2981 was predicted by aggregating volume predictions for all the trees callipered in the plot. Chirici et al. (2020) provides a complete 2982 description of GSV prediction for Italian NFI plots. The field plots 2983 selected are denoted as n elements of the population (U) (Figure 2). 2984

2985

# 2.3. Remotely sensed data

2986

#### 2.3.1. Landsat predictors

2987A composite of Landsat 7 ETM+ images was computed using the2988Google Earth Engine (GEE) platform (Gorelick et al. 2017) which

provides the complete, pre-processed archive of Landsat data. 2989 Specifically, we used the Landsat 7 Surface Reflectance Tier 1 images, 2990 i.e. atmospherically corrected and with the surface reflectance values 2991 calculated using the LEDAPS algorithm (Masek et al. 2013). We 2992 2993 selected late-spring and summer images with less than 70% cloud cover (i.e. between April 1st and September 30th) acquired in 2005, the 2994 2995 same period as the NFI field campaign. The image collection resulted 2996 in a total of 106 images. To avoid noise values in the images, pixels 2997 covered by clouds, shadow, water, and snow were masked using the CFMask algorithm implemented in GEE (Foga et al. 2017) and were 2998 2999 not used to calculate the composite image. The composite image was constructed to obtain a unique image in specific time windows using 3000 3001 all available satellite images (Wulder et al. 2019). The pixel values of 3002 the composite image are calculated as a function of the corresponding 3003 pixels of the acquired images in the time windows. In our case, the 3004 median function was used to calculate pixel values for each Landsat 7 3005 band of the composite image (Figure 1 B).

3006 Based on the Italian composite image, seven Landsat predictors were 3007 calculated, one for each Landsat band, specifically, the mean value of 3008 the composite image pixels within the plot area (Table 1). Moreover, 3009 the same Landsat predictors were calculated for all *N* grid cells of the 3010 Italian study area population.

141

#### 3011 2.3.2 Canopy Height Model predictors

In Italy, a national raster grid CHM at 1×1m resolution derived from 3012 3013 available ALS data is available (D'Amico et al., 2021). Based on the 3014 national CHM (Figure 1C), three standard CHM variables were 3015 calculated for all available NFI plots and for the N study area grid cells. The CHM predictor variables were computed using the R-3016 CRAN package "raster" (Hijmans et al., 2012) and were the three 3017 3018 height standard metrics: (i) the mean (CHM Mean), (ii) the 90th 3019 percentile of the canopy height distribution (CHM Prc90), and (iii) the standard deviation (CHM Std), of the 1×1m pixel values that were 3020 inside or intersected by the boundary of the 23×23m pixels (Table 1). 3021

3022 3023

Data source	Variable name	Information			
NFI	GSV	Field measured growing stock volume			
			Wavelength	Resolution	
	Band 1	Blue	0.45-0.52 μm	30 m	
	Band 2	Green	0.52-0.60 μm	30 m	
	Band 3	Red	0.63-0.69 µm	30 m	
Landsat 7	Band 4	Near-infrared	0.77-0.90 μm	30 m	
ETM+	Band 5	Short-wave infrared	1.55-1.75 μm	30 m	
	Band 6	Thermal infrared	10.40-12.50 μm	60 m	
	Band 7	Short-wave infrared	2.09-2.35 μm	30 m	
	CHM_Mean	Mean of 1 m × 1 m CHM pixels			
CHM	CHM_Prc90	90 percentile of $1 \text{ m} \times 1 \text{ m}$ CHM pixels			
	CHM_Std	Standard deviation of $1 \text{ m} \times 1 \text{ m}$ CHM pixels			

Table 1: Landsat and CHM predictors

3024

We used CHM because in Italy ALS point cloud data are not always available and because Chirici et al. (2016) has already demonstrated that GSV prediction accuracies are comparable for CHM metrics and point-based metrics. The CHM was derived from ALS datasets collected by different authorities between 2004 and 2017.



Figure 1. A: Italian field plots; B: RGB Landsat composite image captured to
create annual images with a median value for each pixel in ALS cover; C:
Italian CHM cover.

3030

3035 The ALS datasets shared some common characteristics considered 3036 suitable for forestry applications (Goodwin et al. 2006, Wulder et al. 3037 2008): acquisition flight altitudes between 500 m and 3000 m; spatial resolution of derived CHM ranging between 1 m and 5 m; pulse 3038 density between 0.4 and 5 pulses per  $m^2$ , which even at small values 3039  $(0.4-1.0 \text{ pulses per m}^2)$  facilitate generation of reliable digital elevation 3040 models for plot-level forest estimates (~23 m pixel size - Jakubowski 3041 et al. 2013). However, several studies have demonstrated that a lag 3042 time greater than five years between field measurements and ALS data 3043 3044 can be problematic when predicting forest variables (Wulder et al. 2008, Tompalski et al. 2019), especially when the area-based approach 3045
3046 is used (Næsset, 2002). Thus, we selected CHM metrics derived from 3047 ALS acquired within five years of the NFI field survey. The grid cells 3048 for which the CHM metrics are available were denoted as the *stratum*<sub>2</sub> 3049 elements of the population (U) (Figure 2).

## **3050 3. Methods**

3051

#### 3.1. Methods overview

Full Landsat and CHM coverage, including for all NFI plots, were available for the study area. Our goal was to estimate the effects of varying the CHM and Landsat coverage proportions when estimating GSV. To evaluate the effects of varying CHM and Landsat coverages, we used a stratified estimation approach for which the strata were the Landsat coverage (*stratum*<sub>1</sub>) and the CHM coverage (*stratum*<sub>2</sub>), with proportions respectively denoted as  $w_1$  and  $w_2$  with  $w_1+w_2=1$ .

Estimates obtained using the simple expansion estimator, which is based exclusively on the NFI plot data within strata (Approach 0), were used for comparisons with estimates based on two additional approaches: (Approach 1) the model-assisted estimator within strata, and (Approach 2) the model-assisted estimator within strata using additional CHM-based pseudo-plots for model construction.

We progressively varied the amount of CHM and Landsat coverage using different  $w_1$  and  $w_2$  proportions (i.e.,  $w_1 = 10\%$  and  $w_2 = 90\%$ ;  $w_1 = 20\%$  and  $w_2 = 80\%$ ; ...;  $w_1 = 90\%$  and  $w_2 = 10\%$ ). Based on the proportion, the CHM stratum was constructed by randomly selecting 3069 pixels until the correct proportion  $w_2$  was achieved. All remaining 3070 pixels were then designated as the Landsat stratum. NFI plots were 3071 distributed between the two strata according to their locations. For 3072 each proportion, we repeated the three approaches iteratively, 3073 selecting randomly the strata 50 times. Finally, the average values over 3074 all iterations were used to estimate the effects of CHM coverage 3075 change among approaches (Tables A1, A2, A3).

Strat	um1					
W1	Stratum <sub>2</sub>	PSEUDO-PLOT				
	W2	ESTIMATED TARGET VARIABLES (GSV)	ARGET VARIABLES (GSV)			
		<i>n</i> <sub>2</sub>				
	CHM PREDICTORS	TARGET VARIABLES (G	SV)			
		<b>n</b> 1				
		TARGET VARIABLES (GS	SV)			

3076

3077 Figure 2. Predictors overview. The  $w_1$  and  $w_2$  proportions varied 3078 progressively. 3079

3080 For Approach 0 data for all plots were used with the simple expansion 3081 estimator. For Approach 1, the stratified estimators were used with the 3082 within-strata means and variances estimated using the model-assisted 3083 regression estimators. For Approach 2 we constructed pseudo-plots by 3084 using the CHM variable to predict GSV for some selected plot-size 3085 areas in the CHM stratum. The intent was to determine if the combination of the NFI plot data and the pseudo-plot predictions
would produce a more accurate GSV-Landsat model. We then used
stratified estimation for which within-strata means and variances were
estimated using the model-assisted estimators using data for only the
within-strata NFI plots but no pseudo-plots.

3091

## 3.2. NFI plot selection

3092 In the temporal lag between NFI field plot measurement (2005) and ALS acquisition (D'Amico et al., 2021), some plots were harvested or 3093 otherwise substantially disturbed between the two dates. To alleviate 3094 3095 this discrepancy, we selected ALS data acquired in a range of five 3096 years of the NFI field survey. Moreover, plots disturbed in the period 3097 between the laser scanning and the field inventory or incorrectly 3098 linked to the ALS data due to poor plot locations were identified and 3099 removed in the following way. A residuals analysis was performed 3100 with a weighted estimation of heteroscedastic residual variances (section 3.3). Specifically, plots for which the ratio of the residual 3101 calculated as the difference between the observation and prediction 3102 and the corresponding residual standard deviation estimated through 3103 the weighted method were greater than 2.0 were considered outliers. 3104 In total, 3% of the NFI plots were identified as outliers and removed 3105 from the final dataset (2534 NFI plots). 3106

# **3107 3.3. Nonlinear power model**

3108 A nonlinear power regression model was used to describe the 3109 relationship between GSV for NFI plots and associated CHM metrics. 3110 The simple correlation coefficient between CHM metrics and GSV for 3111 *stratum*<sub>2</sub> (CHM) was performed to select the CHM metric that 3112 produced the most accurate predictions. The model has the 3113 mathematical form,

3114 
$$y_i = \beta_1 * x_i^{\beta_2} + \varepsilon_i \tag{1}$$

where *i* index plots,  $y_i$  is GSV,  $x_i$  is the ALS metric,  $\varepsilon_i$  is a random 3115 3116 residual, and the  $\beta s$  are parameters to be estimated. An advantage of 3117 this model is that when the ALS metrics are zero, as is the case for 3118 many non-forest plots, the prediction will also be zero. Preliminary 3119 analyses indicated that the individual ALS metric that produced the 3120 most accurate predictions was CHM mean (Eq. (1)). All possible 3121 combinations of one, two, and three additional independent variables 3122 together with CHM mean were considered for the model. However, 3123 none contributed to statistically significantly increasing the quality of 3124 model fit to the data.

3125 As with most biological data, residual heteroscedasticity in the form 3126 of greater residual variances for larger observations was evident. Although the mathematical form of Eq. (1) readily lends itself to a log 3127 for either linearization 3128 transformation or reduction of heteroscedasticity, weighted nonlinear least squares were used for 3129 these analyses. The heterogeneous model residual variance,  $\sigma_i^2$ , was 3130

characterized using a 5-step procedure (McRoberts et al., 2016b, Section 3.2.2): (i) calculate the model prediction residuals as  $\varepsilon_i =$  $y_i - \hat{y}$  where  $\hat{y}_i = \hat{\beta}_1 * x_1^{\hat{\beta}_2}$ ; (ii) order the pairs  $(x_i, \varepsilon_i)$  to  $x_i$ ; (iii) aggregate the ordered pairs into groups of size 25; (iv) for each group, g, calculate the mean,  $\overline{x_g}$ , of the predictor variable and the standard deviation,  $\hat{\sigma_g}$ , of the grouped residuals; and (v) model the relationship between  $\hat{\sigma}_g$  and  $\bar{x}_g$  as,

3138 
$$\widehat{\sigma_g} = \lambda * \bar{x}_g + \varepsilon_g,$$
 (2)

3139 where  $\lambda$  is a model parameter estimated using ordinary least squares 3140 techniques. For the weighted nonlinear least squares technique, each 3141 observation was weighted inversely to its residual variance estimated 3142 by substituting the CHM mean for  $\bar{x}_g$  in Eq. (2) (McRoberts et al., 3143 2018).

As for the CHM metric, in *stratum*<sub>1</sub> (non-CHM), a nonlinear power 3144 3145 regression model was used to describe the relationship between GSV from NFI plots and associated Landsat metrics. To select the 3146 3147 combination of independent variables that produced the greatest prediction accuracy, in addition to the seven Landsat predictors, we 3148 3149 calculated the normalized difference for all the predictor combinations 3150 (21 new indices). A subset of the three Landsat predictors, with the greatest correlations to GSV observations, were selected to develop 3151 3152 the model:

3153 
$$y_i = \beta_3 * x_{ij}^{\beta_4} * e^{\beta_5 * x_{ij} + \beta_6 * x_{ij}} + \varepsilon$$
 (3)

3154 where *i* indicates plots or population units,  $x_{ij}$  is the jth Landsat metric,

3155  $\beta_s$  are parameters to be estimated.

# **3156 3.4. Stratified estimator (Approach 0)**

3157 The essence of stratified estimation is to assign population units to 3158 groups or strata, where for this study the strata are the CHM coverage and the non-CHM (Landsat) coverage, estimate the within-strata 3159 sample plot means and variances using the simple expansion 3160 3161 estimators, and then calculate the population estimate as a weighted 3162 average of the within-strata estimates where the weights are proportional to the stratum sizes. With stratified estimation, (1) the 3163 3164 strata weights are calculated as the relative proportions of the 3165 population area corresponding to strata, and (2) the sample unit is 3166 assigned to a single stratum. For this study, we varied the strata 3167 proportions and consequently the strata weights. At the same time, we 3168 assigned NFI plots to strata based on the stratum assignment of the population units containing the plot centers. 3169

3170 Stratified estimates of means and variances are calculated using3171 estimators provided by Cochran (1977):

3172 
$$\hat{\mu}_{STR} = \sum_{h=1}^{H} w_h \hat{\mu}_h,$$
 (4)

3173 and

3174 
$$Var(\hat{\mu}_{STR}) = \sum_{h=1}^{H} w_h^2 \frac{\hat{\sigma}_h^2}{n_h},$$
 (5)

3175 where  $\hat{\mu}_{STR}$  denotes the stratified estimator of the mean; h=1, ..., H3176 denote strata;  $w_h$  denotes the *h*th stratum weight; and  $n_h$  denotes the 3177 number of plots assigned to the *h*th stratum;

3178 
$$\hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \tag{6}$$

3179 denotes the sample mean for the *h*th stratum;  $y_{hi}$  is the *i*th sample 3180 observation of GSV in the *h*th stratum; and

3181 
$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \hat{\mu}_h),$$
 (7)

3182 denotes the sample variance for the *h*th stratum.

The simple expansion estimator, used here within strata and 3183 sometimes referred to as "simple random sample" or "direct" 3184 3185 estimator, has some advantages: (i) simplicity, using only the sample data, without fitting a model or some other statistical procedure, (ii) 3186 3187 intuitiveness, using common arithmetic mean and the Central Limit 3188 Theorem to determine its variance; and (iii) unbiasedness under 3189 simple random and systematic sampling designs (Moser et al., 2017). 3190 The disadvantage of the simple expansion estimator is the possibly large variances, mainly when the sample size is small and/or the 3191 population is highly variable (McRoberts et al., 2013). Nevertheless, 3192 because it is unbiased, this approach was used to compare the different 3193 model-assisted estimators used with the different predictors and strata 3194 proportions. 3195

# 3196**3.5. Stratified estimation with model-assisted**3197estimation within strata (Approach 1)

3198 Model-assisted estimators use models based on auxiliary data to 3199 enhance inferences but rely on probability samples for validity. For 3200 this study, within each stratum, we used the model-assisted regression 3201 estimators of means and variances provided by Särndal et al. (1992). 3202 The population and the corresponding plots are divided into two strata, 3203 sequentially changing their proportions ( $w_1$  and  $w_2$ ).

3204 
$$\hat{\mu}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{y}_i^{Landsat} - \frac{1}{n_1} \sum_{i=1}^{n_1} \left( \hat{y}_i^{Landsat} - y_i^{Landsat} \right)$$
(8)

3205 
$$\hat{\mu}_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \hat{y}_i^{CHM} - \frac{1}{n_2} \sum_{i=1}^{n_2} \left( \hat{y}_i^{CHM} - y_i^{CHM} \right)$$
(9)

3206 where:  $N_i$  is the number of Landsat pixels in the non-CHM stratum; 3207  $N_2$  is the number of CHM cells in the CHM stratum;  $y_i^{Landsat}$  is an 3208 inventory plot observation from the non-CHM stratum;  $\hat{y}_i^{Landsat}$  is a 3209 prediction from the GSV-Landsat model;  $y_i^{CHM}$  is an inventory plot 3210 observation from the CHM stratum;  $\hat{y}_i^{CHM}$  is a prediction from the 3211 GSV-CHM model. The estimate of the overall mean and standard 3212 error are:

3213 
$$\hat{\mu} = \sum_{h=1}^{2} w_h \cdot \hat{\mu}_h$$
 (10)

3214 and

3215 
$$SE(\hat{\mu}) = \sqrt{\widehat{Var}(\hat{\mu})} = \sqrt{\sum_{h=1}^{2} w_h^2 \cdot \frac{\widehat{\sigma}_h^2}{n_h}}$$
(11)

3216 where

3217 
$$\hat{\sigma}_{1}^{2} = \frac{1}{(n_{1}-1)} \sum_{i=1}^{n_{1}} \left( \hat{y}_{i}^{Landsat} - y_{i}^{Landsat} \right)^{2}$$
(12)

3218 and

3219 
$$\hat{\sigma}_2^2 = \frac{1}{(n_2 - 1)} \sum_{i=1}^{n_2} (\hat{y}_i^{CHM} - y_i^{CHM})^2$$
 (13)

The primary advantage of the model-assisted estimators is that they capitalize on the relationship between the sample observations and their model predictions to reduce the variance of the estimate of the within strata means and, by extension, the population mean (McRoberts et al., 2014).

# 32253.6. Stratified estimation with model-assisted3226estimation within strata using pseudo-plots for model3227construction (Approach 2)

3228 In the second approach, we added pseudo-plots to investigate the 3229 possibility of constructing a more accurate GSV-Landsat model. We 3230 first tessellated the study area into 8x8 km grid cells. In grid cells that 3231 had pixels selected for the CHM stratum, we randomly selected one 3232 pixel that was not already associated with an NFI plot. Pseudo-plots 3233 were constructed at the selected locations by using the power model 3234 (Eq. (1)) and the CHM variables to predict GSV. Based on the selection of pseudo-plot locations, the intensity of pseudo-plot 3235 3236 sampling was proportional to the sampling intensity of NFI plots in 3237 the CHM stratum.

Based on the two strata sizes, and thus the number of NFI plots  $(n_2)$  in the CHM stratum, the GSV for the pseudo-plots was estimated. The inventory plots  $(n_1 + n_2)$  and the CHM-based pseudo-plots were used to construct a GSV-Landsat model which was applied to the entire

3242 study area. Next, the model-assisted estimator with NFI plots in the Landsat stratum  $(n_1)$ , was used to estimate the mean GSV within it 3243 (*stratum*<sub>1</sub>) (Eq. (8)). While for computing GSV estimates in the CHM 3244 3245 stratum (*stratum*<sub>2</sub>), presented NFI plots  $(n_2)$  are used. Thus, although 3246 pseudo-plots are also included in the CHM stratum, we used this same 3247 model to predict GSV for the entire CHM stratum for the model-3248 assisted estimator. Consequently, pseudo-plot data do not affect 3249 model-assisted estimation so only NFI plots  $(n_2)$ , were used to estimate the mean GSV within CHM stratum (stratum<sub>2</sub>) (Eq. (9)). The same 3250 3251 effect occurred in calculating the variance and, consequently, the 3252 standard errors, which for the two strata were calculated as Eq. (11). 3253 Also, in this approach, we varied the strata proportions and weights of population units in strata. While in the non-CHM stratum  $(stratum_1)$ 3254 3255 there were the corresponding NFI plots, in the CHM one (*stratum*<sub>2</sub>) there were NFI plots and pseudo-plots, both increased progressively, 3256 3257 simulating greater CHM coverage. Particularly, the number of pseudo-3258 plots increased as the size of the CHM stratum increased, up to a maximum of 567 pseudo-plots selected in the 23x23 grid (Table A3). 3259

3260

# **3.7. Relative efficiency**

To assess the efficiency of the model-assisted estimators, we compared variance estimates obtained with each approach with variance estimates obtained with the simple expansion estimator, taken as reference, and calculated the relative efficiency coefficient(*RE*) as:

$$3266 \qquad RE = \frac{V\widehat{a}r(\widehat{\mu}_{SEE})}{V\widehat{a}r(\widehat{\mu}_{S})} \tag{14}$$

3267 where  $Var(\hat{\mu}_{SEE})$  is the simple expansion estimator variance (Eq. (6)), 3268 and  $Var(\hat{\mu}_s)$  is the variance for Approaches 1 and 2 with model-3269 assisted estimation within strata.

Because RE is the ratio between the values of the variance of 3270  $Var(\hat{\mu}_{SFF})$  and  $Var(\hat{\mu}_{s})$ , values greater than 1 are evidence of greater 3271 3272 precision for the model-assisted estimates (Moser et al., 2017). The 3273 *RE* coefficient can be interpreted as the factor by which the original 3274 sample size would have to be increased to achieve the same precision without using the remotely sensed auxiliary data as that achieved using 3275 3276 the remotely sensed auxiliary data (Chirici et al., 2020; Francini et al., 2020; Francini et l., 2021). 3277

- 3278 4. Results and discussion
- 3279

# 4.1. Nonlinear power model

The analysis of the simple correlation coefficient of Landsat metrics used as independent variables for predicting GSV in *stratum*<sub>1</sub> (non-CHM) and the simple correlation coefficient between the CHM metric and GSV for *stratum*<sub>2</sub> (CHM) are reported in Table 2. The final models for *stratum*<sub>1</sub> (Eq. (3)) and *stratum*<sub>2</sub> (Eq. (1)) reported R<sup>2</sup> of 0.26 and 0.44, respectively (Figure 3).





 $3287 \qquad \mbox{Figure 3. GSV} \ (m^3 \ ha^{-1}) \ observation \ vs \ prediction \ scatters \ plot \ and \ residuals.$ 

3288 On the top: CHM model prediction, on the bottom: Landsat model prediction.

3289 Darker dots are average of aggregations of 25 observations.

3290

3291 Table 2 Indices with the greatest correlation coefficients between the GSV

- 3292 and Landsat and CHM metrics
- 3293

Metric	RMSE	<b>R</b> <sup>2</sup>	MAE
B5_B6	128.03	0.224	94.37
SWIRI (Band5)	128.81	0.214	95.26
B6_B7	129.02	0.212	95.53
CHM mean	109.19	0.435	75.94

# 32944.2. Stratified estimator with the simple expansion3295estimator within strata (Approach 0)

The stratified estimator of the mean with the simple expansion 3296 estimator within strata yielded overall estimates of  $\hat{\mu} = 159.58 \text{ m}^3 \text{ ha}^{-1}$ 3297 with  $SE(\hat{\mu}) = 2.89 \text{ m}^3 \text{ ha}^{-1}$ . Considering each stratum, the estimates of 3298 the mean ranged from  $\hat{\mu}_1 = 157.6 \text{ m}^3 \text{ ha}^{-1}$  to  $\hat{\mu}_1 = 160.4 \text{ m}^3 \text{ ha}^{-1}$ , for 3299 stratum<sub>1</sub> and between  $\hat{\mu}_2 = 158.2 \text{ m}^3 \text{ ha}^{-1}$  and  $\hat{\mu}_2 = 160.67 \text{ m}^3 \text{ ha}^{-1}$  for 3300 3301 stratum<sub>2</sub>. These differences are attributed to differences between strata weights and the proportions of plots assigned to strata (Table A1, 3302 Figure 5). In particular, *stratum*<sub>1</sub> standard errors,  $SE(\hat{\mu}_1)$ , ranged 3303 approximately from 3.0 m<sup>3</sup> ha<sup>-1</sup> to 9.1 m<sup>3</sup> ha<sup>-1</sup> while *stratum*<sub>2</sub> standard 3304 errors,  $SE(\hat{\mu}_2)$ , from 3.0 m<sup>3</sup> ha<sup>-1</sup> to 9.1 m<sup>3</sup> ha<sup>-1</sup> with the greatest 3305 estimates for small stratum proportions and the greatest estimates for 3306 large stratum proportions and numbers of plots. 3307

## 3308 3309

# 4.3. Stratified estimation with model-assisted estimation within strata (Approach 1)

The model-assisted estimates of the mean for the entire population and for the individual strata based on different stratum proportions were similar to the corresponding simple expansion estimates obtained for Approach 0 with greater similarity for increasing *stratum*<sub>2</sub> (CHM) proportions (Table A2, Figure 5). The standard errors for estimates of the means were smaller with increasing stratum proportions, with values from  $SE(\hat{\mu}_1) = 2.5 \text{ m}^3 \text{ ha}^{-1}$  to  $SE(\hat{\mu}_1) = 7.8 \text{ m}^3 \text{ ha}^{-1}$  for *stratum*<sub>1</sub> 3317 and form  $SE(\hat{\mu}_2) = 2.2 \text{ m}^3 \text{ ha}^{-1}$  to  $SE(\hat{\mu}_2) = 7.0 \text{ m}^3 \text{ ha}^{-1}$  for *stratum*<sub>2</sub>.



3318 (Figure 4).

Figure 4. Standard error of the GSV estimate in Approach 1 overall and forboth CHM and Landsat strata.

3322

3319

3323 Similarly, the bias estimates for the MA estimator in both strata were smaller with increasing stratum proportions from  $\widehat{B\iota as_1} = -1.7 \text{ m}^3 \text{ ha}^3$ 3324 <sup>1</sup> to  $\widehat{Bias}_1 = -1.9 \text{ m}^3 \text{ ha}^{-1}$  for stratum<sub>1</sub> and from  $\widehat{Bias}_2 = -0.4 \text{ m}^3 \text{ ha}^{-1}$  to 3325  $\widehat{Bias}_2 = -5.4 \text{ m}^3 \text{ ha}^{-1}$  for *stratum*<sub>2</sub>. Small bias estimates reflect the 3326 3327 means of differences between GSV observations and model predictions, and small variance estimates can be attributed to the good 3328 3329 quality of fit of the power model to the data. However, even with small 3330 bias estimates,  $\hat{\mu}_1$  was underestimated for all stratum proportions. This

3331 spectral saturation effect with underpredictions for plots with GSV greater than 500 m<sup>3</sup> ha<sup>-1</sup> is well-documented in the literature (Chirici 3332 3333 et al., 2020). Indeed, Landsat spectral reflectance values are not 3334 sensitive to multilayer canopy forests or dense forests (Zhao et al., 3335 2016). Moreover, complex topographic features, such as for most of 3336 the Italian forest area, may affect the spectral signature and the data saturation values of forest aboveground biomass and GSV (Lu et al., 3337 2016; Nichol and Sarker, 2011). The saturation effect, although less 3338 severe, has also been reported in the literature for ALS data (Nilsson 3339 et al., 2017; Giannetti et al., 2018; Lefsky et al., 2005). 3340

# 33414.4 Stratified estimation with model-assisted3342estimation within strata using pseudo-plots for model3343construction (Approach 2)

3344 To construct a more accurate model for predicting GSV from the 3345 Landsat auxiliary data, we generated pseudo-plots using the CHM 3346 variable to predict GSV for selected areas in the CHM stratum. The 3347 model-assisted estimates of the means for the entire population and for 3348 the individual stratum based on different stratum proportions were similar to the means estimates for the preceding Approach 1 (Table 3349 A3, Figure 5). For the Landsat  $stratum_1$ , the bias estimates for the 3350 3351 model-assisted estimator in the Landsat stratum were smaller than the estimates for Approach 1 for almost all proportions, with values from 3352  $\widehat{Bias_1} = 1.61 \text{ m}^3 \text{ ha}^{-1}$  to  $\widehat{Bias_1} = -2.8 \text{ m}^3 \text{ ha}^{-1}$ . The smaller values of 3353  $\widehat{Bias}_1$  for Approach 2 relative to those for Approach 1 reflect the 3354

positive influence of the pseudo-plots which neutralized the saturation effect of Landsat data. However, as the Landsat stratum size decreases and, therefore, with fewer NFI plots  $(n_1)$ ,  $\widehat{Buas_1}$  tends to increase in absolute value. This inconsistent bias trend depends on the decreasing numbers of NFI plots, and also on the positive effect of increasing the numbers of pseudo-plots.



Figure 5. GSV and standard error of the estimated GSV distributions, for approaches 1 and 2 over 50 iterations for each strata proportion. The green dashed line represents the mean value of the simple expansion estimator (Approach 0).

For the Landsat *stratum*<sub>1</sub>, the strata standard errors for estimates of the means were smaller for increasing stratum proportions with values ranging from  $SE(\hat{\mu}_1) = 2.6 \text{ m}^3 \text{ ha}^{-1}$  to  $SE(\hat{\mu}_1) = 7.9 \text{ m}^3 \text{ ha}^{-1}$  for *stratum*<sub>1</sub>. In addition, the standard errors for estimates of the means were smaller with increasing numbers of pseudo-plots with values from  $SE(\hat{\mu}) = 2.5 \text{ m}^3 \text{ ha}^{-1}$  to  $SE(\hat{\mu}) = 2.2 \text{ m}^3 \text{ ha}^{-1}$ . For *stratum*<sub>2</sub>, bias  $\widehat{Bias}_2$  and  $SE(\hat{\mu}_2)$  were the same as for Approach 1.

3373

# 4.5. Relative Efficiency

Relative efficiency was calculated for each approach with estimates 3374 obtained with the simple expansion estimator as reference (Eq. (14)) 3375 (Tables A1, A2, A3). RE (Eq. (14)) for the stratified estimation with 3376 model -assisted estimation within strata ranged between 1.17 to 1.31. 3377 3378 For situations with 100% Landsat and 100% CHM, RE were 3379 respectively 1.16 and 1.33. For Approach 2, REs were between 1.17 3380 and 1.31, with relevant implications for inventory applications. 3381 Indeed, given the considerable current expense associated with ground sampling, large REs have the potential to greatly enhance NFI 3382 3383 estimation. However, the RE values obtained using pseudo-plots to construct a more accurate model-assisted estimator using the Landsat 3384 3385 auxiliary data are comparable to those for Approach 1.

3386 *RE* values for  $w_2=0.5$  are especially relevant for the various 3387 approaches examined, because with the release of data from the 3<sup>rd</sup> 3388 NFI in coming months, approximately 50% of Italian forests will have

ALS coverage. The 3<sup>rd</sup> NFI field surveys have been carried out since 3389 2017. So, unless a nationwide ALS survey is conducted in the short 3390 3391 term to ensure the temporal consistency with the measured field plots, CHM coverage will continue to be characterized by only partial and 3392 3393 fragmentary coverage. The method presented in this work may be 3394 applied operationally. It is worth noting that completion and updating 3395 of the national ALS data cover are planned for the following National 3396 Recovery and Resilience Plan (NRRP) (MITE, 2021).

3397 *RE* for stratified estimation with model-assisted estimation within 3398 strata with equal strata proportions for Approach 1 produced *RE*=1.23. 3399 In Approach 2, with equal strata proportions, *RE*=1.23.

For both approaches, *RE* values were greater for greater CHM stratum 3400 proportions. Nevertheless, even with limited ALS cover, cost 3401 3402 efficiency should not be ignored. For example, RE = 1.234 (equal to 50% ALS coverage in Approach 1) means that to achieve the same 3403 3404 precision levels without the use of the auxiliary information, sample 3405 sizes would have to be increased by a factor of 0.234, i.e., for this study, the sample size of 2534 would have to be increased by  $0.234 \times$ 3406 3407 2534= 593 plots. For a 2021 measurement cost of approximately 3408 500€/plot, the cost savings from using the auxiliary information and 3409 stratified estimation is a non-negligible amount of 296,500€.

# **4.6. Summary**

3411 The reasons that led to the development of this methodology are related to the historical Italian situation and the upcoming release of 3412 the 3<sup>rd</sup> NFI data. Field surveys of the 3<sup>rd</sup> inventory were performed 3413 between 2017 and 2020 (De Laurentis et al., 2021), and the time-3414 consistent, available ALS data are fragmented and distributed over the 3415 3416 whole territory. The approaches developed in this work are geared 3417 toward considering the availability of non-wall-to-wall ALS data and 3418 how their availability affects large-scale volume estimation (Figure 6). 3419 The simple expansion estimator was used as a reference and then 3420 calculated for each stratum proportion, although, as expected, these 3421 changes provided comparable values for both strata (Table A1).

3422 The stratified estimator with model-assisted estimation within strata produced more precise estimates (decrease in  $\widehat{SE}(\hat{\mu})$ ), as the CHM 3423 3424 coverage increased (Table A2). Indeed, the use of ALS data confirms the potential to improve GSV estimation performance because of its 3425 well-known ability to capture canopy information (Lu et al., 2012). Of 3426 note was the underestimation for  $stratum_1$  (Landsat). The Landsat 3427 3428 estimation model (Eq. (3)) produced small estimated model-assisted bias ( $\widehat{B\iota as_1} = -2 \text{ m}^3 \text{ ha}^{-1}$ ), although the average GSV for the study area 3429 was  $\hat{\mu}_1$  154.9 m<sup>3</sup> ha<sup>-1</sup>, substantially less than the average observed plot 3430 GSV of 159.6 m<sup>3</sup> ha<sup>-1</sup>. This difference is ascribed to the saturation 3431 effect of the Landsat predictors in the GSV estimation. Uncertainties 3432 can be attributed to both typical coppices for complex forest structures 3433

and mature forests with volumes greater than 500 m<sup>3</sup> ha<sup>-1</sup> that cannot 3434 be accurately predicted using data from passive optical sensors 3435 (Chirici et al., 2020; Lu et al., 2012). More accurate image composite 3436 methods such as the Best Available Pixel (BAP) (White et al. 2014) 3437 3438 can improve composite image generation with more homogeneous 3439 band values. In addition, other available wall-to-wall satellite optical 3440 data will need to be evaluated (Sentinel-2 (S2), Landsat 8 and Landsat 3441 9). For example, Mura et al. (2018) demonstrated a comparable capability between S2 and Landsat 8 in estimating GSV, while Astola 3442 et al. (2019) found that models based on S2 were more accurate than 3443 3444 Landsat 8 models for estimating multiple forest variables.

In Approach 2, we tried to increase the sample size and precision by 3445 3446 adding more plots. Because no more NFI plots were available, we 3447 constructed pseudo-plots by using the CHM variable to predict GSV 3448 for some selected plot-size areas in the CHM stratum. The CHM data 3449 are distributed in nationwide ALS surveys from 2004 through 2017 3450 (D'Amico et al., 2021). Pseudo-plots were constructed using ALS data distributed throughout Italy acquired before 2010. We used data for 3451 3452 the combination of the measured plots and the pseudo-plots to 3453 construct a more accurate GSV-Landsat model, which, in turn, may 3454 produce greater precision for the model-assisted estimator using the 3455 Landsat auxiliary data. However, considering CHM stratum, pseudo-3456 plots have no effects on the model-assisted estimation of the mean and variance because predictions equal the simulated observations. 3457

Indeed, we used the NFI plots  $(n_2)$  in the CHM stratum, to construct a GSV-CHM model to predict GSV for the entire *stratum*<sub>2</sub> (including pseudo-plots). The results for *stratum*<sub>1</sub> showed more accurate  $\hat{\mu}_1$ estimation by reducing  $\widehat{Bias}_1$ , while for *stratum*<sub>2</sub> the results were the same as those for Approach 1. The *SEs* for the entire population for the two approaches showed no appreciable differences.



Figure 6. Study area Growing Stock Volume prediction map generated with Approach 1 (w1 =0.5, w2 = 0.5).

## **5.** Conclusion

3468 GSV for Italian forest land covered by ALS was estimated with three approaches using existing ALS, Landsat, and NFI data. Three primary 3469 3470 conclusions were drawn from the study. Firstly, CHM and Landsat 3471 data are confirmed as a reliable and efficient sources of information for predicting GSV, even in large and complex Mediterranean forest 3472 areas. Moreover, the power model facilitates accurate estimation of 3473 3474 biological variables such as GSV. Secondly, remotely sensed auxiliary 3475 data may contribute to increasing the precision of GSV estimates. Thirdly, ALS data, although fragmentary and acquired in different 3476 3477 years, contributes to improved GSV estimates. CHM and Landsat data, increased the efficiency of GSV estimation compared with the 3478 3479 standard stratified estimate with the simple expansion estimator within 3480 strata for the two approaches, (i) stratified estimation with model-3481 assisted estimation within strata and (ii) stratified estimation with 3482 model-assisted estimation within strata and CHM-based pseudo-plots. The total ALS coverage provided the greatest improvement in 3483 3484 accuracy with a relative efficiency of 1.33. However, only a portion of forests are covered by ALS currently. Even in a realistic scenario for 3485 Italy, with CHM data in only 50% of forests, their contribution 3486 increases the accuracy (variance) by a factor of 1.24. 3487

From this perspective, we encourage the Italian NFI to publicly release
both NFI plot data, including the exact plot coordinates, for the 3<sup>rd</sup>
National Forest Inventory for purposes of supporting construction of

- 3491 future RS-based forest maps of GSV or biomass. Lastly, in the future
- 3492 we anticipate that ALS will finally be available wall-to-wall in Italy to
- 3493 facilitate prediction of forest variables with even greater accuracy.

# **6. Annex**

 Table A1. Average results of Approach 0 for each different strata portion, after 50 iterations.

		Land	sat stratum			СНМ	Population			
	W1	$n_1$	$\widehat{\mu}_1$	$SE(\hat{\mu}_1)$	W2	$n_2$	$\widehat{\mu}_2$	$SE(\hat{\mu}_2)$	μ	SE(µ̂)
0	1	2534	159.58	2.89	0	0	0	0	159.58	2.89
1	0.9	2278	159.73	3.05	0.1	256	158.28	9.05	159.58	2.89
2	0.8	2023	159.30	3.23	0.2	511	160.72	6.48	159.59	2.89
3	0.7	1779	159.65	3.45	0.3	755	159.42	5.27	159.58	2.89
4	0.6	1521	160.17	3.72	0.4	1013	158.69	4.58	159.58	2.89
5	0.5	1272	158.89	4.07	0.5	1262	160.30	4.10	159.59	2.89
6	0.4	1014	160.29	4.58	0.6	1520	159.11	3.72	159.58	2.89
7	0.3	760	159.00	5.25	0.7	1774	159.82	3.46	159.57	2.89
8	0.2	512	157.58	6.36	0.8	2022	160.07	3.24	159.57	2.89
9	0.1	254	160.41	9.11	0.9	2280	159.50	3.04	159.59	2.89
10	0	0	0	0	1	2534	159.58	2.89	159,58	2,89

Table A2. Average results of Approach 1 for each different strata portions, after 50 iterations.

			Landsa	t stratun	1				CHM	Population					
	<i>w1</i>	<i>n</i> <sub>1</sub>	$\frac{\sum_{i=1}^{N_1} \widehat{y}_i^{Lands}}{N_1}$	<b>Bias</b> <sub>1</sub>	$\widehat{\mu}_1$	$SE(\hat{\mu}_1)$	<b>W</b> 2	<i>n</i> <sub>2</sub>	$\frac{\sum_{i=1}^{N_2} \widehat{y}_i^{CHM}}{N_2}$	<b>Bias</b> <sub>2</sub>	$\widehat{\mu}_2$	$SE(\hat{\mu}_2)$	û	$\widehat{SE}(\widehat{\mu})$	RE
0	1	2534	154.31	-1.77	156.09	2.49	0	0	0	0	0	0	156.09	2.49	1.16
1	0.9	2278	154.35	-1.77	156.13	2.63	0.1	256	153.79	-5.40	158.86	6.91	156.41	2.47	1.17
2	0.8	2023	154.15	-1.77	155.92	2.78	0.2	511	156.31	-2.44	158.75	4.94	156.48	2.44	1.19
3	0.7	1779	154.50	-1.72	156.23	2.99	0.3	755	157.38	-0.86	158.24	3.96	156.83	2.41	1.20
4	0.6	1521	154.87	-1.79	156.66	3.22	0.4	1013	157.74	-0.71	158.45	3.44	157.38	2.37	1.22
5	0.5	1272	153.80	-1.73	155.52	3.52	0.5	1262	159.01	-0.40	159.41	3.09	157.47	2.34	1.23
6	0.4	1014	154.66	-1.89	156.55	3.94	0.6	1520	158.04	-0.59	158.63	2.79	157.80	2.30	1.26
7	0.3	760	154.19	-1.81	156.01	4.53	0.7	1774	158.28	-0.63	158.91	2.61	158.04	2.28	1.27
8	0.2	512	152.82	-1.76	154.57	5.46	0.8	2022	158.68	-0.42	159.10	2.44	158.19	2.24	1.29
9	0.1	254	155.61	-1.80	157.41	7.84	0.9	2280	158.36	-0.47	158.83	2.29	158.69	2.20	1.31
10	0	0	0	0	0	0	1	2534	158.49	-0.49	158.98	2.17	158.98	2.17	1.33

3501Table A3. Average results of Approach 2 for each different strata portions, after 50 iterations. The Pp header3502indicates pseudo-plots amount.

			Landsat	t stratun	1		CHM stratum								Population		
	W1	<i>n</i> <sub>1</sub>	$\frac{\sum_{i=1}^{N_1} \hat{y}_i^{Lands}}{N_1}$	<i>Bias</i> <sub>1</sub>	$\widehat{\mu}_1$	$SE(\hat{\mu}_1)$	<b>W</b> 2	<i>n</i> <sub>2</sub>	Рр	$\frac{\sum_{i=1}^{N_2} \widehat{y}_i^{CHM}}{N_2}$	<b>Bias</b> <sub>2</sub>	$\hat{\mu}_2$	$SE(\hat{\mu}_2)$	û	$\widehat{SE}(\widehat{\mu})$	RE	
1	0.9	2279	155.14	-0.94	156.08	2.63	0.1	255	62	153.46	-5.40	158.86	6.91	156.36	2.5	1.17	
2	0.8	2023	156.64	0.89	155.75	2.78	0.2	511	127	156.31	-2.44	158.75	4.94	156.35	2.4	1.19	
3	0.7	1779	157.62	1.61	156.01	2.99	0.3	755	189	157.38	-0.86	158.24	3.96	156.68	2.4	1.20	
4	0.6	1521	157.44	0.90	156.54	3.22	0.4	1013	253	157.74	-0.71	158.45	3.44	157.30	2.4	1.22	
5	0.5	1272	156.63	1.21	155.43	3.52	0.5	1262	314	159.01	-0.40	159.41	3.09	157.42	2.3	1.23	
6	0.4	1014	156.16	-0.48	156.64	3.95	0.6	1520	379	158.04	-0.59	158.63	2.79	157.83	2.3	1.25	
7	0.3	760	154.17	-1.78	155.94	4.55	0.7	1774	443	158.28	-0.63	158.91	2.61	158.02	2.3	1.27	
8	0.2	512	154.60	0.01	154.59	5.49	0.8	2022	504	158.68	-0.42	159.10	2.44	158.20	2.2	1.29	
9	0.1	254	154.16	-2.80	156.96	7.94	0.9	2280	567	158.36	-0.47	158.83	2.29	158.65	2.2	1.31	

# 3505 References

- Astola, H., Häme, T., Sirro, L., Molinier, M., & Kilpi, J. (2019).
  Comparison of Sentinel-2 and Landsat 8 imagery for forest
  variable prediction in boreal region. Remote Sensing of
  Environment, 223, 257-273. Doi https://doi.org/10.1016/j.rse.2019.01.019
- Chirici G, Giannetti F, McRoberts RE, 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. International
  Journal of Applied Earth Observation and Geoinformation. 84:
  101959. doi: 10.1016/J.JAG.2019.101959
- 3517 Chirici G, McRoberts RE, Fattorini L, Mura M, Marchetti M (2016).
  3518 Comparing echo-based and canopy height model-based metrics
  3519 for enhancing estimation of forest aboveground biomass in a
  3520 model-assisted framework. Remote Sensing of Environment.
  3521 174: 1–9. doi: 10.1016/j.rse.2015.11.010
- Corona, P., Fattorini, L., Franceschi, S., Scrinzi, G., Torresan, C.,
  2014. Estimation of standing wood volume in forest
  compartments by exploiting airborne laser scanning information:
  model-based, design-based, and hybrid perspectives. Can. J. For.
  Res. 44, 1303–1311.
- D'Amico G, Vangi E, Francini S, Giannetti F, Nicolaci A, Travaglini
  D, Massai L, Giambastiani Y, Terranova C, Chirici G, Submitted.
  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/ifor3648014
- 3533 De Laurentis D., Papitto G., Gasparini P., Di Cosmo L., Floris A.,
  3534 2021. Italian Forests, Selected results of the third National Forest
  3535 Inventory INFC 2015. Carabinieri Command for the Protection
  3536 of Biodiversity and Parks. Roma, Italia. Available online:
  3537 https://www.inventarioforestale.org/sites/default/files/datiinvent
  3538 ario/pubb/Sintesi\_INFC2015.pdf

- FAO 2020. Global Forest Resources Assessment 2020: Main report.
  184 p. Rome. https://doi.org/10.4060/ca9825en
- Foga S, Scaramuzza PL, Guo S, Zhu Z, Dilley RD, Beckmann T,
  Schmidt GL, Dwyer JL, Joseph Hughes M, Laue B (2017). Cloud
  detection algorithm comparison and validation for operational
  Landsat data products. Remote Sensing of Environment. 194:
  379–390. doi: 10.1016/j.rse.2017.03.026
- 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.
- 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.
- Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R
  (2017). Google Earth Engine: Planetary-scale geospatial analysis
  for everyone. Remote Sensing of Environment. 202: 18–27. doi:
  10.1016/j.rse.2017.06.031
- Hijmans R.J., van Etten J., 2012. raster: Geographic analysis and
  modeling with raster data. R package version 2.0-12.
  http://CRAN.R-project.org/package=raster
- 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.
  https://doi.org/10.1016/j.rse.2017.04.004
- Hyyppä J, Hyyppä H, Leckie D, Gougeon F, Yu X, Maltamo M 3567 (2008). Review of methods of small-footprint airborne laser 3568 scanning for extracting forest inventory data in boreal forests. 3569 3570 International Journal of Remote Sensing. doi: \_ 3571 10.1080/01431160701736489
- 3572 INFC, 2004. Il disegno di campionamento. Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAF - Direzione 3573 3574 Generale per le Risorse Forestali Montane e Idriche, Corpo 3575 Forestale dello Stato, ISAFA, Trento. 36 p. http://www.isafa.it/scientifica/. 3576

- 3577 Gregoire, T.G., Næsset, E., McRoberts, R.E., Ståhl, G., Andersen, H.3578 E., Gobakken, T., Ene, L., Nelson, R., 2016. Statistical rigor in Lidar-assisted estimation of aboveground forest biomass. Remote Sens. Environ. 173, 98–108.
- 3581Gregoire, T.G., 1998. Design-based and model-based inference in3582survey sampling: appreciating the difference. Canadian Journal3583ofForest3584https://doi.org/10.1139/x98-166
- Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., 3585 Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., 3586 Næsset, E., Olsson, H., 2018. Remote sensing and forest 3587 3588 inventories in Nordic countries - roadmap for the future. Scand. 1 - 16.3589 J. For. Res. 7581. 3590 https://doi.org/10.1080/02827581.2017.1416666
- Kotivuori E, Korhonen L, Packalen P (2016). volume , biomass and
  dominant height in Finland. 50: 1–28.
- 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 S, Quackenbush LJ, Im J (2019). Airborne lidar sampling strategies
  to enhance forest aboveground biomass estimation from Landsat
  imagery. Remote Sensing. 11. doi: 10.3390/rs11161906
- Lisańczuk, M., Mitelsztedt, K., Parkitna, K. et al. Influence of
  sampling intensity on performance of two-phase forest inventory
  using airborne laser scanning. For. Ecosyst. 7, 65 (2020).
  https://doi.org/10.1186/s40663-020-00277-6
- 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.
- Masek, J.G., E.F. Vermote, N. Saleous, R. Wolfe, F.G. Hall, F.
  Huemmrich, F. Gao, J. Kutler, and T.K. Lim. 2013. LEDAPS
  Calibration, Reflectance, Atmospheric Correction Preprocessing
  Code, Version 2. Model product. ORNL DAAC, Oak Ridge,
  Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1146
- McRoberts, R.E., Magnussen, S., Tomppo, E.O., & Chirici, G., 2011.
   Parametric, bootstrap, and jackknife variance estimators for the

- 3615k-Nearest Neighbors technique with illustrations using forest3616inventory and satellite image data. Remote Sensing of3617Environment, 115(12), 3165-3174.3618https://doi.org/10.1016/j.rse.2011.07.002
- McRoberts, R.E., Naesset, E. and Gobakken, T., 2013. Accuracy and
  precision for remote sensing applications of nonlinear modelbased inference. IEEE J. Select. Topics Appl. Earth Observat.
  Remote Sens., 6 (1), 27–34.
- McRoberts, R.E., Liknes, G.C., Domke, G.M., 2014. Using a remote
  sensing-based, percent tree cover map to enhance forest
  inventory estimation. Forest Ecology and Management, 331,
  pp. 12-18. https://doi.org/10.1016/j.foreco.2014.07.025
- 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.
- 3632 McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S.,
  3633 Westfall, J.A., 2016b. Hybrid estimators for mean aboveground
  3634 carbon per unit area. For. Ecol. Manage. 378, 44–56.
  3635 https://doi.org/10.1016/j.foreco.2016.07.007
- 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.
  https://doi.org/10.1016/j.rse.2017.09.036
- MITE Italian Ministry of Ecological Transition, 2021. Ministerial
  decree of 09-29-2021 online:
  https://www.mite.gov.it/sites/default/files/archivio/allegati/PNR
  R/DEC%20398 29 9 2021%20(1).pdf
- 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. International
  Journal of Applied Earth Observation and Geoinformation. 23:
  288–300. doi: 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 in- ventories. Forestry 90,
  112–124. https://doi.org/10.1093/forestry/cpw041.
- Mura, M., Bottalico, F., Giannetti, F., Bertani, R., Giannini, R., 3654 3655 Mancini, M., Orlandini, S., Travaglini, D., Chirici, G. (2018). Exploiting the capabilities of the Sentinel-2 multi spectral 3656 instrument for predicting growing stock volume in forest 3657 ecosystems. International Journal of Applied Earth Observation 3658 3659 and Geoinformation, 66, 126-134. doi 3660 https://doi.org/10.1016/j.jag.2017.11.013
- 3661 Næsset E., 2002. Predicting forest stand characteristics with airborne
  3662 scanning laser using a practical two-stage procedure and field
  3663 data. Remote Sensing of Environment. 80: 88–99. doi:
  3664 10.1016/S0034-4257(01)00290-5
- Næsset, E., Ørka, H.O., Solberg, S., Bollandsås, O.M., Hansen, E.H., 3665 Mauya, E., Zahabu, E., Malimbwi, R., Chamuya, N., Olsson, H., 3666 Gobakken, T., 2016. Mapping and estimating forest area and 3667 3668 aboveground biomass in Miombo woodlands in Tanzania using 3669 data from airborne laser scanning, tandem-X, Rapideye, and global forest maps: a com- parison of estimated precision. 3670 3671 Remote 282 - 300.Sens. Environ. 175. 3672 https://doi.org/10.1016/j.rse.2016.01.006
- 3673 Næsset E, Gobakken T, Holmgren J, Hyyppä H, Hyyppä J, Maltamo
  3674 M, Nilsson M, Olsson H, Persson Å, Söderman U (2004). Laser
  3675 scanning of forest resources: The nordic experience.
  3676 Scandinavian Journal of Forest Research. 19: 482–499. doi:
  3677 10.1080/02827580410019553
- Nichol, J.E., Sarker, L.R., 2011. Improved biomass estimation using
  the texture para- meters of two high-resolution. Optical Sensors
  930–948. doi 10.1109/TGRS.2010.2068574
- 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 Sensing of Environment.
  194: 447–454. doi: 10.1016/j.rse.2016.10.022
- Pagliarella, M. C., Sallustio, L., Capobianco, G., Conte, E., Corona,
  P., Fattorini, L., & Marchetti, M. (2016). From one-to two-phase
  sampling to reduce costs of remote sensing-based estimation of

- 3690land-cover and land-use proportions and their changes. Remote3691Sensing of Environment, 184, 410-417.3692https://doi.org/10.1016/j.rse.2016.07.027
- Puliti S, Saarela S, Gobakken T, Ståhl G, Næsset E (2018). Combining
  UAV and Sentinel-2 auxiliary data for forest growing stock
  volume estimation through hierarchical model-based inference.
  Remote Sensing of Environment. 204: 485–497. doi:
  10.1016/j.rse.2017.10.007
- 3698 Saarela S, Holm S, Grafström A, Schnell S, Næsset E, Gregoire TG,
  3699 Nelson RF, Ståhl G (2016). Hierarchical model-based inference
  3700 for forest inventory utilizing three sources of information. Annals
  3701 of Forest Science. 73: 895–910. doi: 10.1007/s13595-016-05903702 1
- 3703 Saarela S, Holm S, Healey SP, Andersen HE, Petersson H, Prentius
  3704 W, Patterson PL, Næsset E, Gregoire TG, Ståhl G (2018).
  3705 Generalized hierarchical model-based estimation for
  3706 aboveground biomass assessment using GEDI and landsat data.
  3707 Remote Sensing. 10. doi: 10.3390/rs10111832
- 3708 Särndal, C.E., Swensson, B., Wretman J., 1992. Model Assisted
  3709 Survey Sampling. Springer, New York
- Tabacchi G, De Natale F, Di Cosmo L, Floris A, Gagliano C,
  Gasparini P, Genchi L, Scrinzi G, Tosi V (2007). Le stime di
  superficie 2005 Prima parte. Inventario Nazionale delle Foreste
  e dei Serbatoi Forestali di Carbonio. MINISTERO DELLE
  POLITICHE AGRICOLE, ALIMENTARI E FORESTALI.
- 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., pp. 415.
- Tompalski P, White JC, Coops NC, Wulder MA (2019).
  Demonstrating the transferability of forest inventory attribute
  models derived using airborne laser scanning data. Remote
  Sensing of Environment. 227: 110–124. doi:
  10.1016/j.rse.2019.04.006
- Waser LT, Ginzler C, Rehush N (2017). Wall-towall tree type
  mapping from countrywide airborne remote sensing surveys.
  Remote Sensing 9 (8): 766. doi: 10.3390/rs9080766

- White, J.C.; Wulder, M.A.; Hobart, G.W.; Luther, J.E.; Hermosilla,
  T.; Griffiths, P.; Coops, N.C.; Hall, R.J.; Hostert, P.; Dyk, A.; et
  al. 2014. Pixel-based image compositing for large-area dense
  time series applications and science. Can. J. Remote Sens. 40,
  192–21. doi https://doi.org/10.1080/07038992.2014.945827
- White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski
  P (2016). Remote Sensing Technologies for Enhancing Forest
  Inventories: A Review. Canadian Journal of Remote Sensing. 42:
  619–641. doi: 10.1080/07038992.2016.1207484
- Wulder MA, Loveland TR, Roy DP, Crawford CJ, Masek JG, 3736 Woodcock CE, Allen RG, Anderson MC, Belward AS, Cohen 3737 3738 WB, Dwyer J, Erb A, Gao F, Griffiths P, Helder D, Hermosilla 3739 T, Hipple JD, Hostert P, Hughes MJ, Huntington J, Johnson DM, Kennedy R, Kilic A, Li Z, Lymburner L, McCorkel J, Pahlevan 3740 N, Scambos TA, Schaaf C, Schott JR, Sheng Y, Storey J, 3741 Vermote E, Vogelmann J, White JC, Wynne RH, Zhu Z (2019). 3742 Current status of Landsat program, science, and applications. 3743 3744 Remote Sensing of Environment. 225: 127-147. - doi: 10.1016/j.rse.2019.02.015 3745
- Yun, T., Jiang, K., Li, G., Eichhorn, M. P., Fan, J., Liu, F., ... & Cao,
  L. (2021). Individual tree crown segmentation from airborne
  LiDAR data using a novel Gaussian filter and energy function
  minimization-based approach. Remote Sensing of Environment.
  256, 112307. doi: 10.1016/j.rse.2021.112307
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., 2016. Examining
  spectral reflectance sa- turation in landsat imagery and
  corresponding solutions to improve forest above- ground biomass
  estimation. Remote Sens., 8(6), 469; doi: 10.3390/rs8060469
- 3755

# 3756 **3.4. Paper IV**

# A deep learning approach for automatic mapping of poplar plantations using Sentinel-2 imagery

- 3759 D'Amico G.<sup>1</sup>, Francini S.<sup>1,2,3,\*</sup>, Giannetti F.1, Vangi E.<sup>1,3</sup>, Travaglini
- 3760 D.<sup>1</sup>, Chianucci F.<sup>4</sup>, Mattioli W.<sup>5</sup>, Grotti M.<sup>4,6</sup>, Puletti N.<sup>4</sup>, Corona
- 3761 *P*.<sup>2,4</sup>, *Chirici G*.<sup>1</sup>
- 3762

<sup>1</sup> Department of Agriculture, Food, Environment and Forestry,
Università degli Studi di Firenze, Via San Bonaventura, 13, 50145
Firenze, Italy

- <sup>2</sup> Dipartimento per l'Innovazione dei sistemi Biologici, Agroalimentari
- 3767 e Forestali, Università degli Studi della Tuscia, Via San Camillo de
- 3768 Lellis, 01100 Viterbo, Italy
- 3769 <sup>3</sup> Dipartimento di Bioscienze e Territorio, Università degli Studi del
- 3770 Molise, Contrada Fonte Lappone, 86090 Pesche, Italy
- 3771 <sup>4</sup> CREA, Research Centre for Forestry and Wood, Viale Santa
- 3772 Margherita 80, 52100 Arezzo, Italy
- 3773 <sup>5</sup> CREA, Research Centre for Forestry and Wood, Via Valle della
- 3774 Quistione, 27, 00166 Rome, Italy
- 3775 6 ERSAF Regional Agency for Services to Agriculture and Forestry,
- 3776 Via Pola 12, 20124 Milan, Italy
- 3777 \* corresponding author saverio.francini@unifi.it
- 3778 GIScience & Remote Sensing
- 3779 https://doi.org/10.1080/15481603.2021.1988427.

#### 3780 Abstract

3781 Poplars are one of the most widespread fast-growing tree species used for forest plantations. Owing to their distinct features (fast growth and 3782 3783 short rotation) and the dependency on the timber price market, poplar 3784 plantations are characterized by large inter-annual fluctuations in their extent and distribution. Therefore, monitoring poplar plantations 3785 require a frequent update of information - not feasible by National 3786 3787 Forest Inventories due to their periodicity - achievable by remote sensing systems applications. In particular, the new Sentinel-2 mission 3788 3789 with a revisiting period of five days represents a potentially efficient 3790 tool for meeting this need.

3791 In this paper, we present a deep learning approach for mapping poplar plantations using Sentinel-2 time series. A reference dataset of poplar 3792 3793 plantations was available for a large study area of more than 46,000 km2 in Northern Italy and served as training and testing data. Two 3794 3795 classification methods were compared: (1) a fully connected neural network (also called multilayer perceptron), and (2) a traditional 3796 3797 logistic regression. The performance of the two approaches was estimated through bootstrapping procedure with a confidence interval 3798 of 99%. Results indicated for deep learning an omission error rate of 3799 2.77%±2.76%, showing improvements compared to 3800 logistic regression, omission error rate =  $8.91\% \pm 4.79\%$ . 3801

3802 Keywords: big data; multitemporal classification; Fully Connected
3803 Neural Networks; forest tree crops; tree species mapping, deep
3804 learning

3805

# 3806 1. Introduction

Poplar (Populus spp.) plantations for timber production are globally 3807 widespread (Ball et al., 2005) (FAO/IPC, 2018). The genus Populus is 3808 well suited for biomass production due to its fast-growing 3809 3810 performance and wood quality. Poplar cultivation provides environmental benefits too, such as the prevention of erosion and 3811 3812 protection of soil, water quality, habitat for many species (Corona et 3813 al., 2020), and it is also directly used for phytoremediation and climate change mitigation. Since conventional National Forest Inventories are 3814 3815 typically updated every 10 years, they are not able to produce suitable 3816 information to support the management of poplar plantations, that are 3817 instead cultivated with very short rotations: 2 years for bioenergy 3818 production, and 10-12 years for plywood production.

Traditional specific inventories of poplar plantations both based on photointerpretation or on field surveys are expensive and timeconsuming (Chiarabaglio et al., 2018; Mattioli et al., 2019; Corona et al. 2020; Marcelli et al. 2020). When, to reduce their cost, data are acquired on the basis of a sampling design they fail in producing spatially explicit maps (White et al., 2016) that, on the contrary, are
3825 even more required for reliable forest plantation management (Di3826 Biase et al., 2018).

Such limitations may be potentially overcome by adopting robust automatic classification methods of remotely sensed data, which at the same time are objective and cheaper than traditional approaches and can be repeated to produce near-real-time information due to the vast availability of imagery (Francini et al., 2020, Vaglio et al., 2021).

In the last few years, the increasing availability of open-access optical
satellite data and the increased big data analysis capabilities led to a
significant advancement in mapping performance of such methods (Li
et al., 2015).

The advent of more frequent and more detailed imageries (such as 3836 those from Sentinel-2 -S2- constellation) has led to the beneficial use 3837 3838 of deep learning (DL) approaches (Zhu et al., 2017; Ma et al., 2019). Many studies explored DL for RS tasks, using several NN 3839 3840 architectures. Relevant studies for the performance achieved in land 3841 cover classification were conducted by Tong et al. (2020) and Alhassan et al. (2020). They used, respectively, high-resolution RS 3842 3843 imagery in China and Landsat imagery in Canada. Despite good 3844 results, both approaches had limitations. The use of high-resolution 3845 images involves long revisit times. While the lower spatial resolution of Landsat imagery is a limiting factor for detailed mapping of highly 3846 3847 heterogeneous areas where poplar plantations are located.

Differently, the short revisit time and high spatial resolution of S2, 3848 allowed the analysis of vegetative cycles, obtaining good performance 3849 using machine learning approaches for crop classification in test sites 3850 spread all over the globe (Inglada et al., 2015; Belgiu and Csillik, 3851 3852 2018; Vuolo et al., 2018). Furtherore, in tree crop (eucalyptus and oil 3853 palm) mapping studies, some DL approaches, guaranteed high overall 3854 accuracy (> 90%) in most diverse environmental conditions, such as 3855 the Iberian Peninsula (Forstmaier et al., 2020) and Malavsia (Liu et 3856 al., 2021; Zheng et al., 2018).

Although RS imagery has been widely used for land use and crop 3857 3858 classification, only a few studies focused in detail on mapping poplar 3859 plantations. The first agroforestry area mapping and estimation study was carried out in Punjab using LISS IV data (Ahmad et al., 2016). S2 3860 3861 data facilitated new studies, such as the analysis of poplar spectral 3862 reflectance at different ages and map poplar agroforestry in two Indian 3863 States (Rizvi et al., 2020), or in the northwest of Turkey, with a single 3864 S2 image (Tonbul et al., 2020). Hamrouni et al. (2020) combined S2 and SAR imagery (i.e., S1), respectively to map and differentiate into 3865 two main stand ages poplar plantations in three French sites. These 3866 3867 studies, although S2-based, focused on single tiles, with limited 3868 datasets, without exploring the potential of highly frequent satellite 3869 imagery in a big data approach to mapping poplar plantations.

3870 This study was inspired by the idea that the spectral signature of poplar 3871 plantations changes in time in a way that is different from that of other crops in the same agricultural areas. Such temporal dynamic is related
to phenological changes of poplar trees as well as with the
accumulation of biomass during the growing season.

To validate our hypothesis, we developed a DL classification 3875 3876 algorithm using multitemporal S2 imagery and tested it to map poplar 3877 plantations in the large and dynamic Padan Plain where Italian poplar 3878 plantations are concentrated. Three main findings supported the 3879 development of our approach: (i) traditional machine learning 3880 algorithms are not efficient enough to extract the complex and nonlinear patterns generally observed in large datasets (Najafabadi et 3881 3882 al., 2015); (ii) in big data analysis, DL results generally exceed those of machine learning (Chollet et al., 2017), (iii) DL models provide 3883 3884 high performance in classification, ensuring an immediately 3885 understandable structure. The DL model we developed, was compared 3886 in terms of accuracy with a traditional logistic regression (LR) model 3887 based on the same predictors. The analysis was carried out for two study years: 2017 for training and validation and 2018 to demonstrate 3888 the replicability of the procedure. 3889

3890

## 2. Materials and Methods

**2.1. Survey area** 

The survey area coincides with the Padan plain in Northern Italy, where most of the Italian poplar plantations are concentrated (Mattioli et al., 2019; Corona et al. 2020), for more than 46,000 km<sup>2</sup> covering

five administrative Regions (Figure 1). Using local land use maps 3895 (D'Amico et al., 2021) we masked out forests and urban areas. The 3896 remaining agricultural areas are characterized by different crops, 3897 horticultural cultivation, and various forest tree crops (Azar et al., 3898 3899 2016), for a total of 330,000 ha. Among forest tree crops, the 3900 specialized poplar plantations, object of our study, are predominant. 3901 Among others, tree plantations of other broadleaf trees, generally 3902 polycyclic plantations, represent 30% of total plantations. More 3903 sporadic are coppice plantations of broadleaf trees, largely consisting of willows and poplars, while coniferous wood plantations are almost 3904 3905 absent (Mattioli et al., 2019).



3906

3907 Figure 1. Study area: Sentinel-2 summer cloud-free composite image.

3908

In this area, poplar plantations are intensively managed and primarilytargeted to plywood production, with rotations usually about 10-12

3911 years and tree spacing between  $36 \text{ m}^2$  (6 x 6 m) and  $49 \text{ m}^2$  (7 x 7 m)

3912 (Corona et al., 2018b, Puletti et al., 2019). In about three-fourth of the

3913 poplar plantations, the 'I-214' (Populus × euroamericana) hybrid

- 3914 clone is used (Chianucci et al. 2020a, 2020b).
- 3915

# 2.2. Sentinel-2 imagery

**2.2.1. Pre-processing** 

The two twin S2 satellites feature an innovative wide-swath width
(290 km), high-resolution, MSI sensor with 13 spectral bands, and a
spatial resolution ranging between 10 and 60 m depending on the
bands (Drusch et al., 2012).

3921 The 10 m resolution bands (three in the visible wavelengths, and one 3922 in the Near Infrared, NIR) are highly suitable for application in 3923 vegetation mapping with object-based image analysis (OBIA) 3924 approaches (Chirici et al., 2016; Garcia et al. 2018; Mura et al., 2018). S2 satellite images are available in tiles with a fixed size of 100 x 100 3925 3926 km. The study area is covered by 13 S2 tiles (Figure 3). We 3927 downloaded from the ESA Copernicus Open Access Hub all the available Level 1C images (i.e., Top-Of-Atmosphere (TOA) 3928 reflectance values), acquired between October 2016 and March 2019, 3929 3930 with cloud cover less than 80% and 70% for the years 2017 and 2018 respectively, for a total of 3,716 images, of which 2,075 were used for 3931 the 2017 analysis and the remaining 1,641 for the 2018 map update 3932 (Figure 2). 3933



3934

3935 Figure 2. Sentinel-2 image acquisition date for each tile.

3936

# 2.2.2. Sentinel-2 multitemporal predictors calculation

The S2 Level 1C TOA images were first corrected to Bottom of Atmosphere Level 2A reflectance, removing pixels covered by clouds and shadows using *sen2cor* software v2.5.5 (Müller-Wilm et al., 2013), available in *sen2r* RStudio package (Ranghetti et al., 2019). All the bands were resampled at 10 m resolution using the GDAL 'gdalwarp' function (Greenberg and Mattiuzzi, 2020).

Then, we calculated a total of 68 indices to be used as predictors 3943 during the classification, computed on the basis of multiple S2 bands 3944 (Table 1). The study area is characteristically cloudy, given the 3945 frequent presence of fog and proximity to mountains. Therefore, we 3946 3947 tested different time windows to generate a cloud-free composite with suitable observations in almost every pixel (White et al., 2014; 3948 3949 Francini et al., in review). Although large, we selected a four-month 3950 time window that ensured for each pixel the availability of cloud-free observations. Furthermore, the cloud-free composite was computed 3951 using time-distance-weighted averaging which guaranteed accurate 3952 3953 estimates of monthly pixels (Eq. 1), even if calculated using a fourmonth time window. 3954

3956 Table 1. Description of the predictors used in the image classification.

Name	Predictors		
Monthly NDVI (mNDVI)	12		
$mNDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \tag{1}$			
where $\rho_{RED}$ and $\rho_{NIR}$ are the reflectance values of each pixel in the red and near-infrared bands, calculated by averaging the values that the pixel has assumed over a temporal window of four months, weighted over time distance between image acquisition date and the 15 <sup>th</sup> of each target month (i.e. between December, first month available with images downloaded from October and following December: Figure 2).			
Summer Spectral Bands (SSB)			
For each S2 band, we calculated the median value of the images acquired in the period 1st May – 30th September. These 11 spectral bands are not predictors but were used to calculate summer spectral indices as described below.			
Summer Spectral indices (SSI)	55		
Using SSB we calculated a set of 55 normalized differential indices based on the 55 pairs of bands available combining the 11 bands. Mathematically, the S2 normalized differential indices, defined as: $Index_{k} = \frac{S2_{bi} - S2_{bj}}{S2_{bi} + S2_{bj}};$ (2)			
where $i \neq j$ , correspond to k-combinations $(S2_{bi}, S2_{bj})$ of the set composed of S2 bands. The number of k-combinations is equal to 55. It corresponds to the binomial coefficient calculated using the factorials according to:			
$\binom{n}{k} = \frac{n!}{k! (n-k)!} \tag{3}$			
where $n = 11$ and $k = 2$ . After that, we standardized data of different S2 combinations as (Enwright et al., 2019): <i>Index</i> <sub>st i</sub>			
$= \frac{(Index_i - \overline{Index_i})}{DS_{Index_i}} $ (4)			

3958 **2.3. F** 

# 2.3. Reference Dataset

A reference dataset is useful both to find the best classification
procedure (optimization) and to calculate the final performance of the
classification (accuracy assessment).

3962 To create the reference dataset, we first segmented the 10 m resolution bands of the S2 imagery acquired in the period 1st May - 30th 3963 3964 September (summer spectral bands, SSB, Table 1). We used the Mean 3965 Shift (MS) segmentation algorithm which produces a labelled image 3966 based on the spectral distance of neighboring pixels. Specifically, if 3967 this range distance is below the range radius, the pixels are grouped 3968 into the same cluster. The MS algorithm does not require prior knowledge of the number and shape of the clusters (Boukir et al., 3969 3970 2012), so the best segmentation parameters (i.e. Spatial Radius (hs) 3971 equal to 4 pixels, Range Radius (hr) of 500 and 15 pixels as Minimum size (ms)) were selected by visual evaluation using a trial-and-error 3972 approach of the alignment between the shape of the polygons 3973 generated by segmentation and the boundaries identified in the image 3974 3975 (Mathieu et al., 2007).

Each of 242,893 polygons generated by the segmentation process, for a total of 328,492.5 ha (equal to 32,849,250 S2 pixels), was assigned to (1) poplar class (i.e., 10,189 polygons covering 31,329.3 hectares) or (2) non-poplar class. First, the class assignment was based on data provided by the INARBO (Inventory of forest farming tree crops in Italy) project (Mattioli et al., 2019) and then refined by

photointerpretation of high-resolution aerial orthophotos acquired in 3982 the years 2014 and 2015, where it was straightforward to discriminate 3983 the simple, typical layout of poplar plantations from other agronomic 3984 crops. We found that poplar plantations younger than 3 years old do 3985 3986 not exhibit a canopy cover enough to discriminate their spectral 3987 responses from ground vegetation and soil. For this reason, such 3988 young plantations were not considered in this study. All the doubt 3989 cases (0.02% of the reference dataset) were checked in the field without finding any misclassification so that the established polygons 3990 set can be considered as an error-free field truth. 3991



Figure 3. Sentinel-2 tiles processed with the reference poplar polygonsdataset.

# 3995 **2.4. Moving Window approach**

3996 We implemented a window locally calibrated approach, i.e., our algorithm divided the large survey area into 25 km x 25 km 3997 "windows", resulting in 208 windows among which just 79 included 3998 3999 poplar plantation polygons in the reference data to be used for 4000 developing the classification models. We calibrated a different model 4001 for each window because such 79 window areas, compared to the 4002 entire survey area, are expected to have more homogeneous (i) environmental conditions, (ii) weather conditions, (iii) land use and 4003 land cover, (iv) amount of available cloud -free images, and 4004 4005 consequently monthly composites.

4006

# 2.5. Fully Connected Neural Network

4007 DL models consist of N stacked layers composed of M nodes that facilitate learning through successive representations of the input data 4008 (Heaton et al., 2018). We developed a fully connected neural network, 4009 also called Multilayer Perceptron (MLP) where all nodes or neurons 4010 4011 in one layer are connected to the nodes in the next layer. The data are transformed in each layer using weights, which are specific parameters 4012 that link the nodes of subsequent layers (Hawryło et al., 2020). The 4013 4014 MLP is applied to the pixels of the 2017 segmented polygons to 4015 classify poplar plantations using data derived by multitemporal 4016 satellite imagery (i.e., a total of 68 predictors described in Table 1).

4017 The MLP method was implemented and optimized by TensorFlow, an4018 open-source platform for Machine Learning (Abadi et al., 2017).

We configured the MLP using a trial-and-error approach based on a 4019 small subsample (2K polygons) of the reference dataset (i.e., 1K 4020 4021 poplar polygons and 1K non-poplar polygons). The MLP model configuration that achieved larger accuracy consists of 17 layers, with 4022 4023 different nodes and activation functions, which are functions used in 4024 neural networks (NN) to compute the weighted sum of input and 4025 biases to decide if a neuron can be fired or not (Nwankpa et al., 2018). 4026 The MLP is structured as five consecutive sequences of hidden layers 4027 with the same activation functions but a different number of neurons per layer. For each of the consecutive sequence, the MLP uses a 4028 4029 Rectified Linear Unit (ReLU) function (Eq. 5) for the first layers, 4030 which performs a threshold operation to each input element where 4031 negative values are set to zero,

4032 
$$f(x) = max(0, x) = \begin{cases} x_i, & \text{if } x_i \ge 0\\ 0, & \text{if } x_i < 0 \end{cases}$$
(5)

4033 Then, in the second hidden layer, a Hyperbolic Tangent (Tanh) 4034 function (Eq. 6) is applied. Tanh is an S-shaped curve passing through 4035 the origin that, in this case, modifies the positive values produced by 4036 Relu, returning a rapid increase for small values and an asymptotic 4037 flattening to 1 for large ones.

4038 
$$f(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)$$
 (6)

To prevent overfitting and to ensure model generalizability, the MLP 4039 applies the dropout layers function that serves to discard randomly 4040 some nodes from the network during each training session, with 4041 different probability for each sequence (i.e., respectively 20% for the 4042 1st sequence and 10% for other four). Because each training sub-4043 4044 network is different, it is possible to prevent the MLP from overfitting 4045 the training data, to improve the generalization and its ability 4046 (Srivastava et al., 2014). For loss function optimization, we used a Sigmoid (Eq. 7), 4047

$$f(x) = \left(\frac{1}{(1+e^{-x})}\right) \tag{7}$$

4049 that rescale the input in a fuzzy value (Benz et al., 2004) between 0 and 1, which, in our case, can be interpreted as the probability of a 4050 4051 pixel to be a poplar plantation. Then, we calculated the median value of the probability to be poplar of the pixels included in each 4052 4053 segmented polygon to attribute it a unique probability value. Finally, by applying a cutoff of 0.1 to the *probability to be poplar* value, the 4054 polygons are assigned to the poplar or non-poplar class. This cutoff 4055 4056 value was chosen after several tests, to limit omission errors since it 4057 excludes only polygons with a probability of representing poplar less 4058 than 10% (Figure 5).

4059 In total the proposed NN has 3,576 neurons and 1,108,295 weights
4060 arranged in 15 fully connected hidden layers: (68 / 500 / 500 / 400 /
400 / 400 / 360 / 200 / 200 / 180 / 100 / 100 / 90 / 50 / 50 / 45 / 1)

4062 (Figure 4). The NN was trained with the Adam gradient-based

4063 algorithm and categorical cross entropy as a loss function, over 150
4064 epochs using a batch dimension equal to the size of the training
4065 dataset.



Figure 4. MLP model architecture. Down: the number of nodes for each layer.
Up: the activation functions used and the percentage of per-layer dropped out
nodes.

4070 **2.6. Logistic Regression** 

4066

4071 For comparison on the same dataset, we tested also the well-known4072 parametric approach based on logistic regression (LR).

4073 The relationship between a categorical dichotomous variable Y, 4074 representing the poplar or non-poplar classes, and the independent 4075 variables X from the 68 predictors (Table 1) can be expressed in the 4076 form:

4077 
$$p_i = f(X_i; \beta) + \varepsilon_i \tag{8}$$

4078 Where  $p_i$  is the probability that the *i*th y = 1,  $\beta$  is the vector parameters 4079 to be estimated and  $\varepsilon_i$  is a vector of residual assumed to be distributed 4080 with 0 mean (Agresti, 2007; McRoberts et al., 2013).

4081 The statistical expectation of Y can be formulated by the logistic4082 model:

4083 
$$p_i = \frac{exp(\sum_{j=1}^J \beta_j x_{ij})}{1 + exp(\sum_{j=1}^J \beta_j x_{ij})} + \varepsilon_i$$
(9)

4084 where *j* indices the independent variables and  $\beta$  can be estimated by 4085 maximizing the log-likelihood L:

4086 
$$ln(L) = \sum_{i=1}^{n} f(X_i; \beta)^{yi} \left[1 - f(X_i; \beta)\right]^{(1-yi)}$$
(10)

4087 As for the MLP model, the LR model provided the fuzzy probability
4088 to be poplar of each pixel. At the polygon level, the probability value
4089 was then attributed by the median value of the included pixels,
4090 applying the same cutoff of MLP equal to 0.1 (Figure 5).

#### 4091 4092

# 2.7. Moving Window calibration and performance assessment

4093 We performed a moving window calibration approach, so, for each *i*th 4094 window (window<sub>*i*</sub>), a different model was calibrated and assessed in 4095 terms of performance. To calibrate the MLP model, the polygons 4096 included in the window<sub>*i*</sub> were split into three sets: (1) training<sub>*i*</sub> (60%), 4097 (2) validation<sub>*i*</sub> (30%) and (3) test<sub>*i*</sub> (10%), maintaining the real, albeit 4098 unbalanced proportion of poplar plantation and non-poplar plantation 4099 polygons present in the window<sub>*i*</sub>. For each *i*th window, the *i*th model

- 4100 uses (1) the training<sub>i</sub> to adjust nodes weights, (2) the validation<sub>i</sub> to 4101 avoid overfitting, thus to evaluate the loss function during training and 4102 (3) the test<sub>i</sub> to assess the accuracy of the model (Laurin et al., 2021). 4103 In this way, the performance of the method is evaluated using never-4104 seen-before data.
- 4105 In contrast to the MLP, the LR calibration is not iterative, and a 4106 validation set is not required. For this reason, the LR model was 4107 trained using both the training<sub>*i*</sub> and the validation<sub>*i*</sub> while its 4108 performance, for the sake of comparability, was assessed using the 4109 same data (test<sub>*i*</sub>) used for the MLP.
- 4110 The performance of models was assessed using a bootstrapping 4111 procedure (Bradleyel and Tibshirani, 1993; Hawryło et al., 2020). For 4112 one hundred thousand iterations we selected the 20% of the test data -4113 obtained aggregating the 79 test<sub>i</sub> - to calculate three parameters of 4114 performance: the overall accuracy (eq. 11), the omission error rate (eq. 4115 12) and the commission error rate (eq. 13).
- 4116  $Overall Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$  (11)
- 4117  $Omission Error rate = \frac{FN}{TP+FN}$  (12)

4118 Commission Error rate = 
$$\frac{FP}{TP+FP}$$
 (13)

4119 where TP = true positives count, corresponding to pixels correctly 4120 classified as poplar plantation; TN = true negative count, 4121 corresponding to pixels correctly classified as non-poplar plantation; 4122 FP = false positive count (or commission errors), corresponding to 4123 pixels incorrectly classified as poplar plantation; and FN = false 4124 negative count (or omission errors), corresponding to pixels4125 incorrectly classified as non-poplar plantation (Francini et al., 2021).





Figure 5. Model flow chart. Each window is selected to locally calibrate the
model using local polygons. For each window, each local polygon's pixel is
classified independently by MLP or LR. The polygon probability of being
poplar is evaluated from the median of the probability of being poplar for
each own pixel.

The Overall Accuracy represents the ratio between the sum of
correctly classified polygons and the total number of polygons,
Omission Error rate refers to poplar polygons erroneously predicted
as non-poplar plantation and Commission Error rate are calculated by
reviewing the classified sites for incorrect classifications.

4137

#### 2.8. 2018 mapping update

4138 Since the RS data we used as input (Table 1) can be calculated for each 4139 year, our procedure is repeatable. Moreover, using the pre-trained 4140 models developed for 2017, our algorithm can be applied do not 4141 requiring a reference dataset. To prove this, we used our algorithm to 4142 map poplar plantations in 2018, using as a dataset the 242,255 4143 polygons obtained from the segmentation of the S2 imagery acquired in the period  $1^{st}$  May –  $30^{th}$  September 2018 (as described in section 4144 4145 2.3.). We finally compared the differences between 2017 and 2018 poplar plantations predicted maps, in terms of area and changes. 4146

#### 4147 **3. Results**

In Table 2 we reported the confusion matrices of both the MLP and
the LR models calculated considering just the test-sets, i.e., 10% of the
reference data kept as an independent dataset.

- 4152 Table 2. Models confusion matrices. For both MLP (A) and LR (B) we report
- 4153 the number of TP,TN, FP and FN.
- 4154

A Pred		Predic	ted MLP	
		poplar plantation	non-poplar plantation	
ence	poplar plantation	TP 1124	FN 32	1156
Refer	non-poplar plantation	FP 2307	TN 22375	24682
		3431	22407	

В		Prec	]	
		poplar plantation	non-poplar plantation	
ence	poplar plantation	TP 1053	FN 103	1156
Refer	non-poplar plantation	FP 1701	TN 22981	24682
		2754	23084	

4155

The results showed that out of 25,838 test polygons, the amount of poplar plantations correctly classified (i.e., TP) was greater for MLP predictions than for LR predictions. Furthermore, in MLP prediction very few poplar plantation polygons were missed (i.e., FN), representing only 0.1% of the total test-set, while LR miss three times as many polygons compared to MLP. The amount of commission errors rate (i.e., FP) was 8.9% for MLP and 6.6% for LR.

4163 Models accuracy and relative confidence intervals of the three4164 parameters of performance were evaluated with a confidence interval

4165 of 99% using one hundred thousand bootstrap iterations in the same 4166 test-set for both models. The MLP results showed an omission error 4167 rate of  $2.8\% \pm 2.8\%$ , a commission errors rate of  $67.2\% \pm 4.6\%$  and 4168 an overall accuracy of  $91.0\% \pm 1.0\%$ . LR provided an omission error 4169 rate of  $8.9\% \pm 4.8\%$ , a commission errors rate of  $61.8\% \pm 5.3\%$  and 4170 an overall accuracy of  $93.0\% \pm 0.9\%$  (Figure 6).



4171

4172 Figure 6. Density distributions by overall accuracy, omission error rate and

- 4173 commission error rate in one hundred thousand bootstrapping iterations for
- 4174 MLP (above) and LR (bottom).

- 4176 Figure 7 shows the number of poplar polygons in test-sets and the
- 4177 performance parameters results of the models for each window4178 derived by the moving window calibration approach.



4179

Figure 7. On the top, the number of poplar plantations in test-set per window.
On the bottom, overall accuracy, omission error rate and commission error
rate for each window with respect to MLP (above) and LR (bottom).

4183

Where, the MLP omission errors rate remained below 10% in 87% of
the cases (i.e., 69 windows), exceeding over 50% only in four cases
due to the low presence of poplar polygons in the test-set, the
commission errors rate were less than 75% in 80% of the cases (i.e.,
windows), and the overall accuracy was higher than 90% in 60%

- 4189 of the cases (i.e., 48 windows). The LR results per window showed an
- 4190 omission errors rate that were less than 10% in 53% of cases (i.e., 42
- 4191 windows), a commission errors rate that were less than 75% in 85%
- 4192 of cases (i.e., 67 windows) and an overall accuracy of more than 90%
- 4193 in 72% of cases (i.e., 57 windows).
- 4194 After the optimization procedure and the selection of the best model,
- 4195 we applied the algorithm to map poplar plantations in the subsequent
- 4196 year (2018). While for 2017 the total coverage of predicted poplar
- 4197 plantations was 48,638.98 ha, the update to 2018 presents a predicted
- 4198 poplar plantations area of 51,846.14 ha (Table 3, Figure 8).



- 4199
- Figure 8. Map of poplar plantation dynamics, with plantation added and lostbetween 2017 and 2018 resulting from MLP model.
- 4202
- 4203 At Regional (European Union NUTS2) level (Table 3), the area of
- 4204 poplar plantations for the year 2018 increased sensibly in Veneto

(+16%) and Emilia-Romagna (+13%), while for Friuli Venezia Giulia
there was a loss of 15% of area. Lombardy and Piedmont with over
76% of poplar plantations area show an increase in area of 9% and 6%
respectively. In absolute terms, the change in area in Lombardy
(+2131 ha) represents 46% of the total, followed by Piedmont (16%),
Friuli Venezia Giulia (15%), Emilia-Romagna (12%) and Veneto
(11%) (Figure 1).

4212 Table 3. 2017 and 2018 poplar and non-poplar cover per administrative

- 4213 Region (European Union NUTS2).
- 4214

	2017		2018			
NUTS2	Poplar (ha)	Non- poplar (ha)	Poplar (ha)	Non- poplar (ha)	Gain (ha)	Loss (ha)
Emilia- Romagna	4133	32076	4662	31607	1545	1016
Friuli Venezia Giulia	4504	37348	3812	37774	902	1594
Lombardy	24062	65283	26193	62981	5300	3169
Piedmont	12986	53225	13742	52345	5019	4263
Veneto	2954	91920	3436	91417	1568	1085
Total	48639	279854	51846	276124	14334	11127

## 4215 **4. Discussion**

4216

# 4.1. Summary and aim of the research

4217 In this work we presented a classification algorithm based on a DL 4218 approach to map poplar plantations in Northern Italy, using S2 4219 multitemporal data. The work aims to evaluate the distribution of 4220 poplar plantations, which are the first source of wood products for 4221 industrial use in Italy. Despite the importance of plantations, there is a 4222 lack of spatialized information needed for management and planning. 4223 The S2 multitemporal data and indices proved to be effective 4224 predictors to overcome this issue, allowing to efficiently differentiate 4225 between poplar plantations and different land cover classes. The DL 4226 classification algorithm was tested and compared with a traditional LR classifier, showing better accuracy, and minimizing omission errors 4227 4228 (2.77%±2.76%). Our work confirms that S2 allows mapping land 4229 cover categories (Bruzzone et al., 2017; Belgiu and Csillik, 2018; 4230 Vuolo et al., 2018), such as poplar plantations, mapped here in a large 4231 and dynamic study area. The herein proposed MLP algorithm 4232 represents an efficient tool, able to provide annual statistics that are 4233 not obtainable by traditional inventory approaches.

4234

# 4.2. Sentinel-2 pre-processing

The temporal resolution of S2 satellite imagery allows increasing thenumber of available cloud-free images compared to Landsat (Belgiu

and Csillik, 2018; Gómez et al., 2016), overcoming the limitations of 4237 missing data-pixels with no observations due to clouds or cloud 4238 shadows (White et al., 2014). Temporal NDVI patterns represent the 4239 most widely used tool to track phenological changes and vegetation 4240 4241 signatures (D'Odorico et al., 2013; Hagolle et al., 2015). Accordingly, 4242 we produced monthly cloud-free NDVI composites using, for each 4243 month, a temporal window of four months. It is important to note that 4244 with the launch of the twin S2 satellite (July 2017) the number of 4245 images doubled, consequently the revisitation time of the mission 4246 went from 10 days to 5 days. This is the reason why we used two different clouds thresholds for 2017 and 2018, respectively 70% and 4247 80%. However, for future investigations, narrower temporal windows 4248 (e.g., two months) could be tested. This aspect may convey an increase 4249 4250 in the classification accuracy of cultivated tree species (Vrieling et al., 4251 2018). Another aspect that could contribute to increasing the accuracy 4252 of the classification model in the future is the availability of longer S2 4253 time series: Vuolo et al. (2018) demonstrated that the accuracy of crops classification in Austria by random forest (RF) increased with 4254 4255 time frame of the analysis (using imagery from January to May, the 4256 overall accuracy was about 50% while with imagery over two years 4257 the overall accuracy reached the 95%).

4258 **4.3. Classification algorithm** 

The MLP is confirmed as a "universal approximator" capable of learning any function, such as poplar plantation classification, on the condition that the dataset is large enough to train and validate a highperformed DL algorithm. Attention must be paid to overfitting as each fully connected layer requires many parameters or weights. To avoid this, the dropout layers used in each block of the MLP have proven to be effective and no additional regularization strategies were needed.

4266 We tested several activation functions and we obtained the best 4267 performances using two consecutive activation functions: (i) Relu, 4268 which set to 0 the negative input values, and (ii) Tanh, an S-shaped function. In particular, Tanh is a zero-center function whose range lies 4269 4270 between -1 to 1. It is smoother than the traditional sigmoid activation 4271 function (Eq. 7), giving a rapid increase for small positive values and a flattening for large ones (Nwankpa et al., 2018), enhancing Relu 4272 outputs. Among the infinite structures and activation functions that 4273 4274 can be developed, ours guaranteed better results than traditional (LR) 4275 systems. However, there are many other activation functions, and their 4276 combinations should also be investigated for these classification 4277 purposes in future research.

4278 Our dataset, as commonly in data-based modeling, was unbalanced.
4279 We tried several approaches to balance the poplar and non-poplar
4280 classes size. However, each one had some cons. The under-sampling
4281 of the non-poplar class causes information lost for the NN while

4282 oversampling the poplar class increases the database size (465,408 polygons) which led to a longer computation time while do not 4283 introduce any additional information. Although there were many other 4284 approaches to managing dataset unbalance, in our work we did not 4285 4286 focus on this but tried to develop a model that would work with the 4287 available data. Particularly, we aimed to optimize the right 4288 performance parameter. This is because, with unbalanced dataset, 4289 most classification performance indicators, including overall 4290 accuracy, may provide misleading information (Devarriya et al., 2020; 4291 Jaafor and Birregah, 2020). Consequently, we made the greatest effort 4292 to minimize omission errors (i.e., poplar plantation polygons wrongly classified as non-poplar plantations). While commission errors could 4293 4294 be removed in post-process, omission errors represent undetected 4295 poplar plantations that would not be mapped. Thus, making the 4296 classification algorithm and the predicted map useless.

4297 Moreover, we developed a Moving Window Calibration approach 4298 which locally preserving the proportion between classes. Therefore, the proposed methodology allowed the capture of local environmental 4299 4300 and meteorological differences by calibrating the model for each 4301 window. On the other hand, to replicate this methodology, the 4302 reference dataset must be well distributed throughout the study area 4303 and training polygons may be available for each window. Based on 4304 local models, we classified the individual pixels of each polygon in the window, aggregating the values by the median. This approach, 4305

although computationally intensive, allowed us to achieve higher
accuracies by limiting the effect of the rare misclassified pixels within
a polygon. Moreover, this object-based approach results in a vector
file, which is easier to use from an operational point of view.

4310 Regarding the S2 bands normalization, we adopted a procedure that to 4311 our knowledge was never applied before. Specifically, we calculated 4312 55 standardized spectral indices SSI based on the 55 normalization 4313 differences calculable through a pairwise combination of the 11 S2 4314 bands. The SSI robustly represented the spectral properties of pixels, 4315 properly helping the DL model to capture more comprehensive data 4316 and satisfying the demand for artificial NN training data (Hu et al., 2018). While this normalization process produces some indices 4317 4318 common in RS applications (among which the NDVI, the NBR, etc.) 4319 some of the SSI were never reported in the literature before and future 4320 research should further explore their potential and limitations such as 4321 their autocorrelation.

Finally, an advantage of DL models compared to other machine
learning methods is their greater ability to characterize the diversity in
big data and the fact that a variable selection is not required (Yu et al.,
2017). For these reasons we did not analyze predictors' importance but
further investigations should be performed in future research.

#### 4327 **4.4. Models accuracy**

The whole procedure was tested using MLP and LR (Figure 5) and major differences were reported. The MLP had an overall accuracy of 91.0%  $\pm$  1.0%, a commission errors rate of 67.2%  $\pm$  4.6% and an omission errors rate of 2.8%  $\pm$  2.8%. The LR results showed a commission errors rate of 61.8%  $\pm$  5.3% and an omission errors rate of 8.9%  $\pm$  4.8%. The overall accuracy reached by the LR was slightly larger than that obtained by the MLP (i.e., 93.0%  $\pm$  0.9%).

4335 Due to unbalanced class sample sizes, the model accuracy comparison 4336 focused on commission and, most importantly, omission errors. 4337 Therefore, the fact that the LR omission errors rate  $(8.9\% \pm 4.8\%)$  was 4338 three times higher than that obtained by MLP ( $2.8\% \pm 2.8\%$ ) is a 4339 critical issue that highlights the superiority of the MLP model, also evident from the analysis of the accuracy per window, where the 4340 4341 omission error rate of MLP was always smaller than that of LR. At the 4342 edges of the survey area, where fewer poplar plantations were located (Figure 7), the omission error rate values were over 50% in four 4343 4344 windows for MLP models and 16 for LR models. On the other hand, 4345 maintaining the omission errors low implies more commission errors. As mentioned, by optimizing the omission error rate, the commission 4346 4347 error rate showed, even under per-window analysis, MLP values 4348 larger than LR, especially in the western part of the survey area.

4349 Commission errors are more abundant than omission errors because,4350 although the S2 data led to an increase in performance, several non-

poplar plantation polygons still appear to be identical in spectral 4351 behaviour to that of poplar plantations (Figure 6). Similarly, it is not 4352 possible to discriminate young plantations' spectral responses from 4353 ground vegetation and soil by remote sensing, due to the limited 4354 4355 canopy structure for the first three years after planting. On the other 4356 hand, commission error polygons can be easily removed by photo-4357 interpretation but, to make the process completely automatic, 4358 additional data should be tested in the future. In particular, for future research, we suggest testing hyperspectral and/or three-dimensional 4359 data as additional predictors. In this sense, the two polar-orbiting 4360 4361 Sentinel-1 satellites, performing synthetic aperture radar imaging, may represent a crucial game-changer. 4362

4363

## 4.5. Procedure replicability

4364 The procedure we developed for the year 2017 allowed to map the 4365 cover of poplar plantations for a total of 48,638.98 ha. To update the map at the following year, we applied the same pre-trained model used 4366 for 2017 but inputting the S2 predictors assessed at the year 2018. The 4367 predicted area of poplar plantations in 2018 was 51,846.14 ha, about 4368 6.6% greater than that of 2017. Such remarkable variations for the two 4369 4370 years were registered also at regional scale (Table 3) with the greatest 4371 increase reported in Veneto (+16%): those changes are due to the 4372 increasing market value of poplar wood in the last decade, which has 4373 boosted new investments in poplar plantations (Coaloa and4374 Chiarabaglio, 2019).

#### 4375 **5.** Conclusions

4376 Forest tree monitoring and assessment are rapidly evolving as new 4377 information needs arise and new techniques and tools become 4378 available. Among these, the most widely applied and promising 4379 approaches today are ensemble methods and DL (Mazzia et al., 2020; 4380 Vuolo et al., 2018). However, the exploitation of these tools, as well 4381 as their implementation within operative management processes, should be evidence-based (Corona, 2018a). The major contribution of 4382 4383 this study is the set-up of an efficient automatic approach to map forest 4384 tree plantations on farmland using S2 multitemporal imagery. Poplar 4385 plantations in Northern Italy have been considered as a key case study. 4386 As a result of the study, three primary conclusions can be drawn:

the S2 mission proved to be an efficient tool to classify forest tree
crops on farmland; the revisitation time of 5 days and the spectral
range are two key aspects for efficiently pinpointing both the
temporal and the spectral behaviors of poplar plantations: to this
end we used monthly NDVI cloud free composites, creating 55
spectral indices;

the classification method that incorporates MLP provides accurate
 classification prediction of poplar plantations: the method is
 reliable and efficient, ensuring the near absence of omission

- 4396 errors; accordingly, compared with logistic regression, the MLP 4397 allowed to reduce the omission errors from  $8.9\% \pm 4.8\%$  to 2.8%4398  $\pm 2.8\%$ ;
- the procedure here developed and tested provide automatically
  good results and can be applied to different reference datasets; to
  prove this, we applied the algorithm over the year 2018,
  identifying an increase in poplar plantation area of about 6.6%
  compared to 2017.
- 4404 Acknowledgment
- The study has been financially supported by the research project
  "Sistema di monitoraggio multiscalare a supporto della pioppicoltura
  di precisione nella Regione Lombardia" (PRECISIONPOP) funded by
  Lombardy Region (Grant number: E86C18002690002).

#### 4409 **References**

- Abadi, M., Isard, M., Murray, D.G., 2017. A computational model for
  TensorFlow an introduction, in: MAPL 2017 Proceedings of the
  1st ACM SIGPLAN International Workshop on Machine
  Learning and Programming Languages, Co-Located with PLDI
  2017. https://doi.org/10.1145/3088525.3088527
- 4415 Agresti, A. 2007. An introduction to categorical data analysis. Second
  4416 edition. Wiley series in probability and statistics. Available at
  4417 https://mregresion.files.wordpress.com/2012/08/agresti4418 introduction-to-categorical-data.pdf
- Ahmad, T., Sahoo, P.M., Jally, S.K., 2016. Estimation of area under agroforestry using high resolution satellite data. Agrofor. Syst. 90, 289–303. https://doi.org/10.1007/s10457-015-9854-2
- Alhassan, V., Henry, C., Ramanna, S., Storie, C., 2020. A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. Neural Comput.
  Appl. 32, 8529–8544. https://doi.org/10.1007/s00521-019-04349-9
- 4427 Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio,
  4428 P.A., 2016. Assessing in-season crop classification performance
  4429 using satellite data: A test case in Northern Italy. Eur. J. Remote
  4430 Sens. 49, 361–380. https://doi.org/10.5721/EuJRS20164920
- Ball, J., Carle, J., Del Lungo, A., 2005. Contribution of poplars and
  willows to sustainable forestry and rural development in:
  Unasylva No. 221, Vol. 56 2005/2. Poplars and willows 56, 3–
  9.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using
  pixel-based and object-based time-weighted dynamic time
  warping analysis. Remote Sens. Environ. 204, 509–523.
  https://doi.org/10.1016/j.rse.2017.10.005
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen,
  M., 2004. Multi-resolution, object-oriented fuzzy analysis of
  remote sensing data for GIS-ready information 58, 239–258.
  https://doi.org/10.1016/j.isprsjprs.2003.10.002
- Boukir, S., Jones, S., Reinke, K., 2012. Fast mean-shift based
  classification of very high resolution images: application to forest
  cover mapping. ISPRS Ann. Photogramm. Remote Sens. Spat.

- 4446 Inf. Sci. 1, 111–116. https://doi.org/10.5194/isprsannals-I-7-111-4447 2012 4448 Bradleyel, E., Tibshirani, R., 1993. An introduction to the Bootstrap. 4449 Adv. Pediatr. https://doi.org/10.1016/j.yapd.2013.04.017 4450 Bruzzone, L., Bovolo, F., Paris, C., Solano-Correa, Y.T., Zanetti, M., Fernandez-Prieto, D., 2017. Analysis of multitemporal Sentinel-4451 4452 2 images in the framework of the ESA Scientific Exploitation of 4453 Operational Missions. 2017 9th Int. Work. Anal. Multitemporal 4454 Remote Sens. Images. MultiTemp 2017 2-5.4455 https://doi.org/10.1109/Multi-Temp.2017.8035230 Chianucci, F., Puletti, N., Grotti, M., Bisaglia, C., Giannetti, F., 4456 4457 Romano, E., Brambilla, M., Mattioli, W., Cabassi, C., Bajocco, S., Li, L., Chirici, G., Corona, P., Tattoni, C. 2020a. Influence of 4458 image pixel resolution on canopy cover estimation in poplar 4459 plantations from field, aerial and satellite optical imagery. Annals 4460 4461 of Silvicultural Research (early view). 46 (1).4462 http://dx.doi.org/10.12899/asr-2074 4463 Chianucci, F., Puletti, N., Grotti, M., Ferrara, C., Giorcelli, A., Coaloa, 4464 D., Tattoni, C. 2020b. Nondestructive Tree Stem and Crown Volume Allometry in Hybrid Poplar Plantations Derived from 4465 Terrestrial Laser Scanning. Forest Science, 66: 737-746. 4466 4467 https://doi.org/10.1093/forsci/fxaa021 4468 Chiarabaglio, P.M., Bergante, S., Sciré, M., Coaloa, D., 2018. Low 4469 cost poplar inventory in the plain of Piemonte (Italy). Ann. Silvic. 4470 Res. 42, 39–42. https://doi.org/10.12899/asr-1494 Chirici G, Mura M, McInerney D, Py N, Tomppo EO, Waser LT, 4471
  - 4471 Chirici G, Mura M, McInerney D, Py N, Tomppo EO, Waser LT,
    4472 Travaglini D, McRoberts RE (2016). A meta-analysis and review
    4473 of the literature on the k-Nearest Neighbors technique for forestry
    4474 applications that use remotely sensed data. Remote Sensing of
    4475 Environment. 176: 282-294.
  - 4476 https://doi.org/10.1016/j.rse.2016.02.001
  - 4477 Chollet, F., Allaire, J.J, 2017. Deep Learning with R. Manning
    4478 Publications, Manning Early Access Program, Ed. 1st. MEA,
    4479 New Delhi, India.
  - 4480 Coaloa D., Chiarabaglio P.M. (2019) Produzione vivaistica
    4481 pioppicola. Sherwood 243: 33-35. In Italian

- 4482 Corona P., 2018a. Communicating facts, findings and thinking to
  4483 support evidence-based strategies and decisions. Ann Silv Res
  4484 42: 1-2. http://doi.org/10.12899/asr-1617
- Corona, P., Bergante, S., Castro, G., Chiarabaglio, P.M., Coaloa, D.,
  Facciotto, G., Gennaro, M., Giorcelli, A., Rosso, L., Vietto, L.,
  Nervo, G., 2018b. Linee di indirizzo per una pioppicoltura
  sostenibile. Rete Rurale Nazionale, Consiglio per la ricerca in
  agricoltura e l'analisi dell'economia agraria, Roma. ISBN: 97888-99595-96-8. In Italian
- 4491 Corona, P., Chianucci, F., Marcelli, A., Gianelle, D., Fattorini, L., Grotti, M., Puletti, N., Mattioli, W., 2020. Probabilistic sampling 4492 4493 and estimation for large-scale assessment of poplar plantations in 4494 Northern Italy. Eur. J. For. Res. 139(6): 981-988. 4495 https://doi.org/10.1007/s10342-020-01300-9
- D'Amico G, Vangi E, Francini S, Giannetti F, Nicolaci A, Travaglini
  D, Massai L, Giambastiani Y, Terranova C, Chirici G, Submitted.
  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/ifor3648014
- 4502Devarriya D., Gulati C., Mansharamani V., Sakalle A. 2020.4503Unbalanced breast cancer data classification using novel fitness4504functions in genetic programming. Expert Systems with4505Applications140:112866.
- 4506 https://doi.org/10.1016/j.eswa.2019.112866
- 4507 Di Biase R.M., Fattorini L., Marchi M. 2018. Statistical inferential
  4508 techniques for approaching forest mapping. A review of methods.
  4509 Annals of Silvicultural Research 42: 46-58.
  4510 http://dx.doi.org/10.12899/asr-1738
- 4511 D'Odorico, P., Gonsamo, A., Damm, A., Schaepman, M.E., 2013.
  4512 Experimental evaluation of sentinel-2 spectral response functions
  4513 for NDVI time-series continuity. IEEE Trans. Geosci. Remote
  4514 Sens. 51, 1336–1348.
  4515 https://doi.org/10.1109/TGRS.2012.2235447
- 4516 Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V.,
  4517 Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P.,
  4518 Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012.
  4519 Sentinel-2: ESA's Optical High-Resolution Mission for GMES

- 4520 Operational Services. Remote Sens. Environ. 120, 25–36.
  4521 https://doi.org/10.1016/J.RSE.2011.11.026
- Enwright, N.M., Wang, L., Wang, H., Osland, M.J., Feher, L.C.,
  Borchert, S.M., Day, R.H., 2019. Modeling barrier island habitats
  using landscape position information. Remote Sens. 11.
  https://doi.org/10.3390/rs11080907
- 4526 FAO/IPC, 2018. Report on the 49th Session of the Executive4527 Committee of the International Poplar Commission.
- Francini, S., McRoberts, R.E., Giannetti, F., Mencucci, M., Marchetti,
  M., Scarascia Mugnozza, G., Chirici, G., 2020. Near-real time
  forest change detection using PlanetScope imagery. Eur. J.
  Remote Sens. 53, 233-244.
  https://doi.org/10.1080/22797254.2020.1806734
- Francini, S., McRoberts, R., E., Giannetti, F., Marchetti, M.,
  Scarascia, G., M., & 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, International Journal of Remote
  Sensing, 42:12, 4693-4711, doi:
- 4539 10.1080/01431161.2021.1899334
- 4540 Francini S., McRoberts R.E., D'Amico G., Coops N.C., Hermosilla
  4541 T., White J.C., Wulder M.A., Marchetti M., Scarascia
  4542 MugnozzaG., Chirici G., (Submitted). An open science-open data
  4543 approach for statistically robust estimation of forest disturbance
  4544 area. International Journal of Applied Earth Observation
- 4545 Garcia, J., Barbedo, A., Alvim, L., Romani, S., Ribeiro, R., Gonçalves,
  4546 V., 2018. A Review on the Automatic Segmentation and
  4547 Classification of Agricultural Areas in Remotely Sensed Images.
  4548 Embrapa Doc. Ser. 156.
- 4549 Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed
  4550 time series data for land cover classification: A review. ISPRS J.
  4551 Photogramm. Remote Sens. 116, 55–72.
  4552 https://doi.org/10.1016/j.isprsjprs.2016.03.008
- 4553Greenberg, JA, Mattiuzzi, M. 2020. gdalUtils: Wrappers for the4554Geospatial Data Abstraction Library (GDAL) Utilities. R4555package version 2.0.3.2. https://CRAN.R-4556project.org/package=gdalUtils.
- Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2015. A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VENµS and Sentinel-2 images. Remote Sens. 7, 2668–2691. https://doi.org/10.3390/rs70302668
- Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C., Sheeren, D.,
  2020. Synergistic use of sentinel-1 and sentinel-2 time series for
  poplar plantations monitoring at large scale. Int. Arch.
  Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch. 43,
  1457–1461. https://doi.org/10.5194/isprs-archives-XLIII-B32020-1457-2020
- 4568 Hawryło, P.; Francini, S.; Chirici, G.; Giannetti, F.; Parkitna, K.; Krok, G.; Mitelsztedt, K.; Lisańczuk, M.; Stereńczak, K.; 4569 Ciesielski, M.; Wężyk, P.; Socha, J. The Use of Remotely Sensed 4570 Data and Polish NFI Plots for Prediction of Growing Stock 4571 4572 Volume Using Different Predictive Methods. Remote 4573 Sens. 2020, 12, 3331.
- 4574 Heaton, J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville:
  4575 Deep learning. Genet Program Evolvable Mach 19, 305–307
  4576 (2018). https://doi.org/10.1007/s10710-017-9314-z
- Hu, Y., Zhang, Q., Zhang, Y., Yan, H., 2018. A deep convolution
  neural network method for land cover mapping: A case study of
  Qinhuangdao, China. Remote Sens. 10, 1–17.
  https://doi.org/10.3390/rs10122053
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., 4581 Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 4582 4583 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite 4584 4585 12356-12379. optical imagery. Remote Sens. 7. 4586 https://doi.org/10.3390/rs70912356
- Jaafor, O., Birregah, O. 2020. KNN-LC: Classification in Unbalanced
  Datasets using a KNN-Based Algorithm and Local Centralities.
  Lecture Notes in Networks and Systems. 10.1007/978-3-03013697-0\_7
- Laurin G.V., Francini S., Luti T., Chirici G., Pirotti F., Papale D.,
  2021. Satellite open data to monitor forest damage caused by
  extreme climate-induced events: a case study of the Vaia storm
  in Northern Italy. Forestry: An International Journal of Forest

- 4595 Research, 94, 3, 407–416. 4596 https://doi.org/10.1093/forestry/cpaa043
- Li, X., Gong, P., Liang, L., 2015. A 30-year (1984 2013) record of annual urban dynamics of Beijing City derived from Landsat data. Remote Sens. Environ. 166, 78–90. https://doi.org/10.1016/j.rse.2015.06.007
- 4601 Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep
  4602 learning in remote sensing applications: A meta-analysis and
  4603 review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.
  4604 https://doi.org/10.1016/j.isprsjprs.2019.04.015
- 4605 Marcelli, A., Mattioli, W., Puletti, N., Chianucci, F., Gianelle, D.,
  4606 Grotti, M., Chirici, G., Amico, G.D., Francini, S., Travaglini, D.,
  4607 Fattorini, L., 2020. Large-scale two-phase estimation of wood
  4608 production by poplar plantations exploiting Sentinel-2 data as
  4609 auxiliary information. Silva Fennica 54(2): article id 10247.
  4610 https://doi.org/10.14214/sf.10247
- 4611 Mathieu, R., Aryal, J., Chong, A.K., 2007. Object-based classification
  4612 of ikonos imagery for mapping large-scale vegetation
  4613 communities in urban areas. Sensors 7, 2860–2880.
  4614 https://doi.org/10.3390/s7112860
- 4615 Mattioli, W., Puletti, N., Coaloa, D., Rosso, L., Chianucci, F., Grotti,
  4616 M., Corona, P., 2019. INARBO.IT: inventario degli impianti di 4617 arboricoltura da legno in Italia. Sherwood 239: 7-10. In Italian
- 4618 Mazzia, V., Khaliq, A., Chiaberge, M., 2020. Improvement in land
  4619 cover and crop classification based on temporal features learning
  4620 from Sentinel-2 data using recurrent-Convolutional Neural
  4621 Network (R-CNN). Appl. Sci. 10, 1–23.
  4622 https://doi.org/10.3390/app10010238
- 4623 McRoberts, R.E., Naesset, E., Gobakken, T., 2013. Accuracy and
  4624 precision for remote sensing applications of nonlinear model4625 based inference. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.
  4626 6, 27–34. https://doi.org/10.1109/JSTARS.2012.2227299
- 4627 Müller-Wilm, U., Louis, J., Richter, R., Gascon, F., Niezette, M.,
  4628 2013. Sentinel-2 Level 2a Prototype Processor: Architecture,
  4629 Algorithms and First Results. ESA Living Planet Symp. 2013,
  4630 Edinburgh, UK 2013, 3–10.
- 4631 Mura, M., Bottalico, F., Giannetti, F., Bertani, R., Giannini, R.,
  4632 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
- 4637 Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N.,
  4638 Wald, R., Muharemagic, E., 2015. Deep learning applications
  4639 and challenges in big data analytics. Journal of Big Data 2, 1.
  4640 https://doi.org/10.1186/s40537-014-0007-7
- 4641 Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018.
  4642 Activation Functions: Comparison of trends in Practice and
  4643 Research for Deep Learning 1–20.
- 4644 Puletti, N., Grotti, M., Scotti, R., 2019. Evaluating the eccentricities
  4645 of poplar stem profiles with terrestrial laser scanning. Forests 10,
  4646 1–9. https://doi.org/10.3390/f10030239
- 4647 Ranghetti, L., Busetto, L., aceresi, 2019. ranghetti/sen2r: Version
  4648 1.1.0. https://doi.org/10.5281/ZENODO.3406962
- 4649 Rizvi, R.H., Handa, A.K., Sridhar, K.B., Singh, R.K., Dhyani, S.K.,
  4650 Rizvi, J., Dongre, G., 2020. Spatial analysis of area and carbon
  4651 stocks under Populus deltoides based agroforestry systems in
  4652 Punjab and Haryana states of Indo-Gangetic Plains. Agrofor.
  4653 Syst. 3. https://doi.org/10.1007/s10457-020-00540-3
- 4654 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I.,
  4655 Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent
  4656 Neural Networks from Overfitting. J. Mach. Learn. Res. 15,
  4657 1929–1958.
- Tonbul, H., Colkesen, I., Kavzoglu, T., 2020. Classification of poplar trees with object-based ensemble learning algorithms using Sentinel-2A imagery. J. Geod. Sci. 10, 14–22. https://doi.org/10.1515/jogs-2020-0003
- Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L.,
  2020. Land-cover classification with high-resolution remote
  sensing images using transferable deep models. Remote Sens.
  Environ. 237, 1–35. https://doi.org/10.1016/j.rse.2019.111322
- Vaglio L., G., Francini, S., Luti, T., Chirici, G., Pirotti, F., Papale, D.,
  2021. Satellite open data to monitor forest damage caused by
  extreme climate-induced events: a case study of the Vaia storm
  in Northern Italy, Forestry: An International Journal of Forest

- 4670
   Research,
   94,
   3,
   Pages
   407 

   4671
   416, https://doi.org/10.1093/forestry/cpaa043
   407 407
- Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang,
  T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M.,
  2018. Vegetation phenology from Sentinel-2 and field cameras
  for a Dutch barrier island. Remote Sens. Environ. 215, 517–529.
  https://doi.org/10.1016/j.rse.2018.03.014
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.T.,
  2018. How much does multi-temporal Sentinel-2 data improve
  crop type classification? Int. J. Appl. Earth Obs. Geoinf. 72, 122–
  130. https://doi.org/10.1016/j.jag.2018.06.007
- 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. Remote Sens.
  4684 42, 619–641. https://doi.org/10.1080/07038992.2016.1207484
- White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T.,
  Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A.,
  Guindon, L., 2014. Pixel-based image compositing for large-area
  dense time series applications and science. Can. J. Remote Sens.
  4689
  40, 192–212. https://doi.org/10.1080/07038992.2014.945827
- 4690Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote4691sensing scene classification: a data augmentation enhanced4692convolutional neural network framework. GIScience Remote4693Sens.54,4694https://doi.org/10.1080/15481602.2017.1222277
- 4694 https://doi.org/10.1080/15481603.2017.1323377
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F.,
  Fraundorfer, F., 2017. Deep learning in remote sensing: a review.
  https://doi.org/10.1109/MGRS.2017.2762307
- 4698

# 4700 **4. Conclusion**

Forest monitoring and assessment are rapidly evolving as new 4701 information needs arise and new techniques and tools become 4702 available. The continuous stream of remotely sensed data, acquired by 4703 4704 any type of sensor: satellite, airborne, or drone, allowing the 4705 acquisition of new information even in areas historically lacking in 4706 forest maps. This scenario of big data availability was the basis of my 4707 Ph.D., in which, the main objectives of my research were addressed in 4708 four scientific papers. In each paper, I focused on specific preliminary questions, attempting to answer and contribute to the increase of 4709 4710 scientific community knowledge. Specifically, the aims of my Ph.D. 4711 improvement of the Italian forest resources involved the 4712 understanding, through the availability of remote sensing data and techniques. Indeed, spatialized data, homogeneously available at 4713 4714 large scales, are increasingly critical to support sustainable forest 4715 management. Furthermore, no less important is the simple sharing of homogenized information layers, even at a large scale, through 4716 4717 modern web GIS, as done for the national FIS.

The first high-resolution forest mask of Italy (NFM) produced by combining local forest and land use maps, resulted more congruent with NFI statistics than forest masks based on radar (JAXA) and optical (HRL) imagery, underestimating for less than 2% of the official NFI estimation of the total forest area. At national and regional levels, the masked GSV map constructed using the NFM produced 4724 GSV estimates that were most in line with the official NFI estimates. A major disagreement with the official NFI estimates was found in the 4725 4726 southern regions and islands, most probably because of the presence of the Mediterranean macchia, which is more difficult to accurately 4727 4728 map. The negative relationship between forest mask accuracy and the 4729 standard error of the GSV estimate demonstrated that the accuracy of 4730 the forest mask must be considered in the GSV estimation through the 4731 model-assisted estimator. Indeed, a more accurate forest mask can compensate for systematic model prediction errors, leading to greater 4732 agreement with official NFI GSV estimates at both the national and 4733 4734 regional levels. Despite these results, due to the non-homogeneous origin, the NFM developed cannot be currently adopted as an official 4735 4736 layer for reporting purposes. Therefore, an operational review of the 4737 mask using remote sensing data and manual photointerpretation is 4738 necessary.

4739 The harmonized CHM produced by combining all the ALS data sets 4740 currently available in Italy covers 59% of Italian forests. These kinds of data are essential for forest monitoring and should be routinely 4741 4742 acquired together with aerial images. In the future, wall-to-wall ALS 4743 coverage in Italy would improve the prediction of forest variables. To 4744 date, the most effective way to employ ALS national coverage for forest GSV estimation resulted from integration with Landsat spectral 4745 4746 data, in conjunction with NFI 2005 field measurements. Among the 4747 different approaches of estimates, we tested a stratified model-assisted within strata, represented respectively by Landsat and ALS coverage,
and a stratified model-assisted approach which involve LiDAR-based
pseudo-plots to create a more accurate GSV-Landsat model for the
Landsat stratum. The study confirmed that LiDAR and Landsat data
are a reliable and efficient source of information to enhance GSV
estimates, even in large and complex Mediterranean forest areas.

4754 LiDAR data, although fragmentary and acquired in different years, 4755 allowed to improve GSV estimates. However, to improve forest 4756 variable predictions it is strongly recommended that in the future the 4757 Italian NFI evolves into a permanent monitoring system, where a 4758 sample of the total number of plots is visited in the field every year to complete the revisit of all plots in 5-10 years. In addition, achieve the 4759 4760 wall-to-wall lidar coverage with surveys planned simultaneously with 4761 NFI surveys would facilitate the prediction of forest variable estimates 4762 with even greater precision.

Based on S2 multitemporal imagery, an efficient automatic approach
to map and update forest tree plantations on farmland was set up in the
Padan Plain, the most suitable area for poplar production in Italy. The
results highlighted the great potential of S2 data in agricultural and
forest species identification, and how the use of large data sets with a
DL approach leads to more accurate mapping results than traditional
methods, reducing errors of omission by approximately two-thirds.

4770

A common thread throughout the studies included in this thesis was 4771 the use of remote sensing big data: i. for their identification and 4772 homogenization of information layers at a national scale (Paper I); ii. 4773 for large-scale evaluation of the effect of different forest masks in 4774 4775 volume estimation (Paper II); iii. in the GSV estimation with different 4776 sources of information such as CHM and Landsat metrics at a national 4777 scale (Paper III), or iv. through a DL approach for the classification of 4778 poplar plantations in the Padan Plain (Paper IV).

4779 In this thesis, the main aim was to create and to increase knowledge 4780 about new information layers, augmenting the Italian forestry 4781 availability of consistent and reliable forest spatial data, potentially useful to support sustainable forest management. Crucial to clarify 4782 4783 some of the aims of the thesis was the period spent at the Swedish 4784 University of Agricultural Sciences (SLU) in Umeå. The economic 4785 and social importance of Sweden's forests appeared from the 4786 numerous forest monitoring programs based on remote sensing data. Moreover, information on the distribution and status of forests are 4787 periodically updated with freely available data in the framework of 4788 4789 NFI. In the context of research and knowledge development, it is 4790 critical to have consistent up-to-date data and a well-structured NFI, 4791 with raw data shared with researchers, technicians, and stakeholders.

4792 In the future, more consistent integration of remote sensing4793 applications for forest mapping of different forest variables in the4794 framework of the NFI should be promoted in Italy.

#### 4795 **Reference**

- 4796 Alberdi I., Canellas I., Vallejo Bombin R.,2017. The Spanish national
  4797 forest inventory: history, development, challenges and
  4798 perspectives. Pesquisa Florestal Brasileira 37: 361. doi:
  4799 10.4336/2017.pfb.37.91.1337.
- Babcock C., Finley A.O., Bradford J.B., Kolka R., Birdsey R., Ryan
  M.G., 2015. LiDAR based prediction of forest biomass using
  hierarchical models with spatially varying coefficients. Remote
  Sensing of Environment 169, 113-127.
  doi:10.1016/j.rse.2015.07.028.
- 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.
  International Journal of Applied Earth Observation and
  Geoinformation 57, 145–153. doi:10.1016/j.jag.2016.12.013.
- 4810 Boudreau J., Nelson R.F., Margolis H.A., Beaudoin A., Guindon L.,
  4811 Kimes D.S., 2008. Regional aboveground forest biomass using
  4812 airborne and spaceborne LiDAR in Québec. Remote Sensing of
  4813 Environment 112, 3876–3890. doi: 10.1016/j.rse. 2008.06.003.
- 4814 Brosofske K.D., Froese R.E., Falkowski M.J., Banskota A., 2014. A
  4815 review of methods for mapping and prediction of inventory
  4816 attributes for operational forest management. Forest Science 60,
  4817 733–756. doi: 10.5849/forsci.12-134.
- Büttner, G.; Feranec, J.; Jaffrain, G.; Mari, L.; Maucha, G.; Soukup, 4818 4819 Τ. The CORINE Land Cover project 2000. EARSeLeProceedings 3, 3/2004 331. 2004. Available online: 4820 4821 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.618. 4822 9940&rep=rep1&type=pdf (accessed on September 2021).
- 4823 Chirici G., Giannetti F., McRoberts R.E., Travaglini D., Pecchi M.,
  4824 Maselli F., Chiesi M., Corona P., 2020. Wall-to-wall spatial
  4825 prediction of growing stock volume based on Italian National
  4826 Forest Inventory plots and remotely sensed data. International
  4827 Journal of Applied Earth Observation and Geoinformatics 84,
  4828 101959. doi: 10.1016/j.jag.2019.101959.
- 4829 Chirici G., McRoberts R.E., Winter S., Bertini R., Bröändli U.-B.,
  4830 Asensio I.A., Bastrup-Birk A., Rondeux J., Barsoum N.,
  4831 Marchetti M., 2012. National forest inventory contributions to

- 4832 forest biodiversity monitoring. Forest Science 58, 257–268.
  4833 doi:10.5849/forsci.12-003.
- 4834 Corona P., Chianucci F., Quatrini V., Civitarese V., Clementel F.,
  4835 Costa C., Floris A., Menesatti P., Puletti N., Sperandio G., Verani
  4836 S., Turco R., Bernardini V., Plutino M., Scrinzi G., 2017.
  4837 Precision forestry: concepts, tools and perspectives in Italy.
  4838 Forest@ 14: 1-12. [In Italian with English summary] doi:
  4839 10.3832/efor2285-014.
- 4840 Corona P., Bergante S., Castro G., Chiarabaglio P.M., Coaloa D.,
  4841 Facciotto G., Gennaro M., Giorcelli A., Rosso L., Vietto L.,
  4842 Nervo G., 2018. Linee di indirizzo per una pioppicoltura
  4843 sostenibile. Rete Rurale Nazionale, Consiglio per la ricerca in
  4844 agricoltura e l'analisi dell'economia agraria, Roma. ISBN: 9784845 88-99595-96-8. In Italian.
- 4846 Corona P., Chirici G., McRoberts R.E., Winter S., Barbati A., 2011.
  4847 Contribution of large-scale forest inventories to biodiversity
  4848 assessment and monitoring. Forest Ecology and Management
  4849 262, 2061–2069. doi:10.1016/j.foreco.2011.08.044.
- 4850 Di Biase R.M., Fattorini L., Marchi M., 2018. Statistical inferential
  4851 techniques for approaching forest mapping. A review of methods.
  4852 Annals of Silvicultural Research 42: 46-58. doi: 10.12899/asr4853 1738.
- 4854 Dinerstein E., Olson D.M., Graham D.J., Webster A.L., Primm S.A.,
  4855 Bookbinder M.P., Ledec G., 1995. Conservation Assessment of
  4856 the Terrestrial Ecoregions of Latin America and the Caribbean.
  4857 The World Bannk Washington (DC).
- 4858 Dubayah R.O., Drake J.B., 2000. Lidar remote sensing for forestry.
  4859 Journal of forestry 98.6: 44-46. doi: 10.1093/jof/98.6.44.
- 4860 European Enviromental Agency, EEA, 2007. Enviromental
  4861 Statement; Office for Official Publications of the European
  4862 Communities: Luxembourg; ISBN 978-92-9167-936-2.
- 4863 FAO, UNCCD. Sustainable Financing for Forest and Landscape
  4864 Restoration: The Role of Public Policy Makers; FAO: Rome,
  4865 Italy, 2015.
- 4866 FAO, 2020.Global Forest Resources Assessment 2020: Main report.
  4867 Rome. https://doi.org/10.4060/ca9825en.

- 4868 FOREST EUROPE, 2020. State of Europe's Forests 2020.,
  4869 Ministerial Conference on the Protection of Forests in Europe,
  4870 FOREST EUROPE Liaison Unit Bratislava.
- 4871 Francini S., McRoberts R.E., Giannetti F., Mencucci M., Marchetti M., Scarascia Mugnozza G., Chirici G., 2020. Near-real time 4872 forest change detection using PlanetScope imagery. European 4873 4874 Journal of Remote Sens. 53. 233-244. doi: 4875 10.1080/22797254.2020.1806734.
- 4876 Garnier M., Bastick C., Colin A., Commagnac L., Lallemant T.,
  4877 Maisonneuve B., Mazepa F., Simon M., Vega C., 2019. La BD
  4878 ForetR version 2. L'IF Synthese periodique de l'inventaire
  4879 forestier no. 46, Institut National de L'information Geographique
  4880 et Forestiere IGN, Saint-Mande, France. [in French] [online]
  4881 URL: http://inventaire-

4882 forestier.ign.fr/IMG/pdf/lif\_46\_poster.pdf.

- 4883 Giannetti F., Pecchi M., Travaglini D., Francini S., D'Amico G.,
  4884 Vangi E., Cocozza C., Chirici G., 2021. Estimating VAIA
  4885 Windstorm Damaged Forest Area in Italy Using Time Series
  4886 Sentinel-2 Imagery and Continuous Change Detection
  4887 Algorithms. Forests, 12, 680. doi: 10.3390/f12060680
- 4888 Giannetti F., Puletti N., Puliti S., Travaglini D., Chirici G., 2020.
  4889 Modelling Forest structural indices in mixed temperate forests:
  4890 comparison of UAV photogrammetric DTM-independent
  4891 variables and ALS variables. Ecological Indicators. 2020, 117,
  4892 106513. doi: 10.1016/j.ecolind.2020.106513.
- 4893 Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore
  4894 R., 2017. Google Earth Engine: Planetary-scale geospatial
  4895 analysis for everyone. Remote sensing of Environment 202, 184896 27. doi: 10.1016/j.rse.2017.06.031.
- Hansen M.C., Potapov P.V., Moore R., Hancher M., Turubanova S.A.,
  Tyukavina A., 2013. High-resolution global maps of 21st-century
  forest cover change. Science 342 (6160): 850-853. doi:
  10.1126/science.1244693.
- 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.
  International Journal Remote Sens. 30, 5159–5175. doi:
  10.1080/01431160903022894.

- Holopainen M., Vastaranta M., Hyyppa J., 2014. Outlook for the next generation's precision forestry in Finland. Forests 5: 1682-1694.
   doi: 10.3390/f5071682.
- Hyyppa J., Hyyppa H., Leckie D., Gougeon F., Yu X., Maltamo M.,
  2008. Review of methods of small-footprint airborne laser
  scanning for extracting forest inventory data in boreal forests.
  International Journal of Remote Sensing 29 (5): 1339-1366. doi:
  10.1080/01431160701736489.
- 4914 ITTO/FAO, 1995. Report on Harmonization of Criteria and Indicators
  4915 for Sustainable Forest Management, FAO/ITTO Expert
  4916 Consultation, 13 16 February 1995. Rome, Italy.
- 4917 JAXA. Global 25m Resolution PALSAR-2/PALSAR Mosaic and Forest/Non-Forest Map (FNF) Dataset Description; Japan 4918 4919 Aerospace Exploration Agency (JAXA), Earth Observation 2016. 4920 Research Center (EORC). Available online: https://www.eorc.jaxa.jp/ALOS/en/palsar fnf/DatasetDescriptio 4921 n PALSAR2 Mosaic FNF revE.pdf (accessed on September 4922 4923 2021).
- 4924 Kangas A., Astrup R., Breidenbach J., Fridman J., Gobakken T., Korhonen K.T., Maltamo M., Nilsson M., Nord-Larsen T., 4925 Olsson H., 2018. Remote sensing and forest inventories in Nordic 4926 4927 countries - Roadmap for the future. Scandinavian Journal of 4928 Forest Research 33: 397-412. doi 10.1080/02827581.2017.1416666. 4929
- 4930 Kovácsová P., Antalová M., 2010. Precision Forestry Definition and
  4931 Technologies. Šumarski List 143, 603–611.
- Kussul N., Lavreniuk M., Skakun S., Shelestov A., 2017. Deep
  Learning Classification of Land Cover and Crop Types Using
  Remote Sensing Data. IEEE Geoscience and Remote Sensing
  Letters, 14(5), 778–782.
  https://doi.org/10.1109/LGRS.2017.2681128.
- 4937 Langanke T 2017. Copernicus Land Monitoring Service-High Resolution Layer Forest: Product Specifications Document 38; 4938 4939 2017. team at EEA, Available online: Copernicus 4940 https://www.eea.europa.eu/data-and-maps/data/copernicus-landmonitoring-servicehigh (accessed on September 2021). 4941
- 4942 LeCun Y., Bengio Y., Hinton G., 2015. Deep Learning. Nature 521,
  4943 436-444. doi:10.1038/nature14539.

- 4944 Lefsky M., Cohen W.B., Parker G.G., Harding D.J., 2002. Lidar
   4945 Remote Sensing for Ecosystem Studies. Bioscience 52, 19–30.
   4946 doi:10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2.
- Lim K.P., Treitz P., Wulder M.A., St-Onge B.A., Flood M., 2003.
  LiDAR remote sensing of forest structure. Progress in Physical Geography 27, 88–106. doi: 10.1191/0309133303pp360ra.
- Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson B.A., 2019. Deep
  learning in remote sensing applications: A meta-analysis and
  review. ISPRS Journal of Photogrammetry and Remote Sening.
  152, 166–177. doi: 10.1016/j.isprsjprs.2019.04.015.
- Ma Y., Wu H., Wang L., Huang B., Ranjan R., Zomaya A., Jie W.,
  2015. Remote sensing big data computing: Challenges and
  opportunities. Future Generation Computer Systems, 51, 47-60.
  doi: 10.1016/j.future.2014.10.029.
- Maltamo M., Naesset E., Vauhkonen J., 2014. Forestry applications of
  airborne laser scanning: concepts and case studies. Springer,
  Dordrecht, Netherlands, pp. 464. doi: 10.1007/978-94-017-86638.
- 4962 Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C.,
  4963 Byers A.H., 2011. Big data: The next frontier for innovation,
  4964 competition, and productivity. McKinsey Global Institute.
- 4965 Matasci G., Hermosilla T., Wulder M.A., White J.C., Coops N.C.,
  4966 Hobart G.W., Zald H.S.J., 2018. Large-area mapping of Canadian
  4967 boreal forest cover, height, biomass and other structural attributes
  4968 using Landsat composites and lidar plots. Remote Sensing of
  4969 Environmet 209, 90–106. doi: 10.1016/j.rse.2017.12.020.
- 4970 McRoberts R.E., Bechtold W.A., Patterson P.L., Scott C.T., Reams
  4971 G.A., 2005. The enhanced Forest Inventory and Analysis
  4972 program of the USDA Forest Service: historical perspective and
  4973 announcement of statistical documentation. Journal of Forestry
  4974 103: 304-308. doi: 10.1093/jof/103.6.304.
- 4975 McRoberts R.E., Cohen W.B., Næsset E., Stehman S.V., Tomppo 4976 E.O., 2010b. Using remotely sensed data to construct and assess forest attribute maps and related spatial products. Scandinavian 4977 4978 Forest Research 340-367. Journal of 25, doi: 10.1080/02827581.2010.497496. 4979
- 4980 McRoberts R.E., Næsset E., Gobakken T., 2013. Inference for lidar 4981 assisted estimation of forest growing stock volume. Remote

- 4982Sensing ofEnvironment128,268–275.4983doi:10.1016/j.rse.2012.10.007.
- 4984 McRoberts R.E., Tomppo E.O., 2007. Remote sensing support for national forest inventories. Remote Sensing of Environment 110, 4986 412–419. doi: 10.1016/j.rse.2006. 09.034.
- 4987 McRoberts R.E., Tomppo E.O., Næsset E., 2010a. Advances and
  4988 emerging issues in national forest inventories. Scandinavian
  4989 Journal of Forest Research. 25, 368–381. doi:
  4990 10.1080/02827581.2010.496739.
- 4991 Montreal Process. 1995. Criteria and Indicators for the Conservation4992 and Sustainable Management of Temperate and Boreal Forest.
- Mura M., McRoberts R.E., Chirici G., Marchetti M., 2015. Estimating
  and mapping forest structural diversity using airborne laser
  scanning data. Remote Sensing of Environment 170, 133–142.
  doi: 10.1016/j.rse.2015.09.016.
- 4997 Naesset E., 2002. Predicting forest stand characteristics with airborne
  4998 scanning laser using a practical two-stage procedure and field
  4999 data. Remote Sensing of Environment 80: 88-99. doi:
  5000 10.1016/S0034-4257(01)00290-5.
- Næsset E., 2007. Airborne laser scanning as a method in operational
  forest inventory: Status of accuracy assessments accomplished in
  Scandinavia. Scandinavian Journal of Forest Research 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 Sensing of Environment 128, 299–314. doi:
  10.1016/j.rse. 2012.10.008.
- 5011 Næsset E., Gobakken T., Holmgren J., Hyyppä H., Hyyppä J.,
  5012 Maltamo M., Nilsson M., Olsson H., Persson Å., Söderman U.,
  5013 2004. Laser scanning of forest resources: the nordic experience.
  5014 Scandinavian Journal of Forest Research 19, 482–499.
  5015 doi:10.1080/02827580410019553
- Nelson R., 2013. How did we get here? An early history of forestry
  lidar1. Canadian Journal of Remote Sensing 39, S6–S17. doi:
  10.5589/m13-011.

- 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 Sensing of Environment
  194, 447–454. doi: 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 Sensing of Environment 119, 148–157.
  doi: 10.1016/j.rse.2011.12.022.
- 5029O'Farrell P.J., Anderson P.M.L., 2010. Sustainable multifunctional5030landscapes: A review to implementation. Current Opinion in5031Environmental503210.1016/j.cosust.2010.02.005.
- 5033Ogaya R., Peñuelas J., 2021. Climate Change Effects in a5034Mediterranean Forest Following 21 Consecutive Years of5035Experimental Drought. Forests, 12, 306.5036https://doi.org/10.3390/f12030306
- Schepaschenko D., See L., Lesiv M., McCallum I., Fritz, S.; Salk, C.; 5037 Moltchanova, E.; Perger, C.; Shchepashchenko, M.; Shvidenko, 5038 A.; et al. Development of a global hybrid forest mask through the 5039 synergy of remote sensing, crowdsourcing and European 5040 Enviromental Agency. Enviromental Statement; Office for 5041 5042 Official Publications of the Communities: European Luxembourg, 2007; ISBN 978-92-9167-936-2. 5043
- Seebach L., McCallum I., Fritz S., Kindermann G., LeDuc S., Böttcher
  H., Fuss S., 2012. Choice of forest map has implications for
  policy analysis: A case study on the EU biofuel target.
  Environmental Science & Policy 2012, 22, 13–24,
  doi:10.1016/j.envsci.2012.04.010.
- 5049 Smith S., Gilbert J., Bull G., Gillam S., Whitton E., 2010. National inventory of woodland and trees (1995-99): methodology.
  5051 Forestry Commission Research Report, Forestry Commission 5052 Scotland, Edinburgh, UK, vol. i-iv, pp. 1-60.
- 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 Sensing of
  Environment 112, 1982–1999. doi: 10.1016/j.rse.2007.03.032.

- Tsagkatakis G., Aidini A., Fotiadou K., Giannopoulos M., Pentari A.,
  Tsakalides P., 2019. Survey of deep-learning approaches for
  remote sensing observation enhancement. Sensors, 19 (18), 3929.
  doi: 10.3390/s19183929.
- Vaglio L., G., Francini, S., Luti, T., Chirici, G., Pirotti, F., Papale, D.,
  2021. Satellite open data to monitor forest damage caused by
  extreme climate-induced events: a case study of the Vaia storm
  in Northern Italy, Forestry: An International Journal of Forest
  Research, 94, 3, Pages 407–416. doi: 10.1093/forestry/cpaa043.
- Valbuena R., Eerikäinen K., Packalen P., Maltamo M., 2016. Gini
  coefficient predictions from airborne lidar remote sensing display
  the effect of management intensity on forest structure. Ecological
  Indicators 60, 574–585. doi: 10.1016/j.ecolind.2015.08.001.
- Valbuena R., Packalen P., Mehtätalo L., García-Abril A., Maltamo
  M., 2013. Characterizing forest structural types and shelterwood
  dynamics from Lorenz-based indicators predicted by airborne
  laser scanning. Canadian Journal of Forest Research 43(11):
  1063-1074. doi: 10.1139/cjfr-2013-0147.
- 5075 Vizzarri M., Sallustio L., Travaglini D., Bottalico F., Chirici G., Garfi
  5076 V., Lafortezza R., Veca D.S.L.M., Lombardi F., Maetzke F.,
  5077 Marchetti M., 2017. The MIMOSE approach to support
  5078 sustainable forest management planning at regional scale in
  5079 Mediterranean contexts. Sustainability 9 (2): 316. doi:
  5080 10.3390/su9020316.
- Waser L.T., Ginzler C., Rehush N., 2017. Wall-to-wall tree type
  mapping from countrywide airborne remote sensing surveys.
  Remote Sensing 9 (8): 766. doi: 10.3390/rs9080766
- 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. doi: 10.3390/f6124386.
- 5088 White J.C., Coops N.C., Wulder M.A., Vastaranta M., Hilker T.,
  5089 Tompalski P., 2016. Remote sensing technologies for enhancing
  5090 forest inventories: a review. Canadian Journal of Remote Sensing
  5091 42: 619-641. doi: 10.1080/07038992.2016.12074 84.
- 5092 Woodcock C.E., Allen R., Anderson M., Belward A., Bindschadler R.,
  5093 Cohen W., Gao F., Goward S.N., Helder D., Helmer E., Namani
  5094 R., Oreopoilos L., Schott J., Thenkabail P.S., Vermote E.F.,

- 5095Vogelman J., Wulder M.A., Wynne R., 2008. Free access to5096Landsat imagery. Science 320: 1011. doi:509710.1126/science.320.5879.1011a.
- Wulder M.A., Bater C.W., Coops N.C., Hilker T., White J., 2008. The
  role of LiDAR in sustainable forest management. The Forestry
  Chronicle 84, 807–826. doi: 10.5558/tfc84807-
- 5101 Zhu X.X., Tuia D., Mou L., Xia G.S., Zhang L., Xu F., Fraundorfer
  5102 F., 2017. Deep learning in remote sensing: a review. IEEE
  5103 Geoscience and Remote Sensing Magazine 5, 4. doi:
  5104 10.1109/MGRS.2017.2762307.
- 5105 Zolkos S., Goetz S., Dubayah R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing.
  5107 Remote Sensing of Environment 128, 289–298. doi: 10.1016/j.rse.2012.10.017.
- 5109 5110

## 5112 Other publications and contributions

#### 5113 **1. Paper**

Chirici G, Giannetti F, Travaglini D, Nocentini S, Francini S, 5114 1. D'Amico G, Calvo E, Fasolini D, Broll M, Maistrelli F, Tonner J, 5115 Pietrogiovanna M, Oberlechner K, Andriolo A, Comino R, 5116 Faidiga A, Pasutto I, Carraro G, Zen S, Contarin F, Alfonsi L, 5117 5118 Wolvnski A, Zanin M, Gagliano C, Tonolli S, Zoanetti R, Tonetti 5119 R, Cavalli R, Lingua E, Pirotti F, Grigolato S, Bellingeri D, Zini E, Gianelle D, Dalponte M, Pompei E, Stefani A, Motta R, 5120 5121 Morresi D, Garbarino M, Alberti G, Valdevit F, Tomelleri E, 5122 Torresani M, Tonon G, Marchi M, Corona P, Marchetti M (2019). 5123 Stima dei danni della tempesta "Vaia" alle foreste in Italia. 5124 Forest@ 16: 3-9. (In Italian) https://doi.org/10.3832/efor3070-016 Marcelli A., Mattioli W., Puletti N., Chianucci F., Gianelle D., 5125 2. Grotti M., Chirici G., D'Amico G., Francini S., Travaglini D., 5126 Fattorini L., Corona P. (2020). Large-scale two-phase estimation 5127 of wood production by poplar plantations exploiting Sentinel-2 5128 data as auxiliary information. Silva Fennica. 54-2, 10247. 5129 https://doi.org/10.14214/sf.10247 5130 3. Vangi E., D'Amico G., Francini S., Giannetti F., Lasserre B., 5131

5132 Marchetti M., Chirici G. (2021). The New Hyperspectral Satellite

5133 PRISMA: Imagery for Forest Types Discrimination. Sensors, 21,

5134 1182. https://doi.org/10.3390/s21041182

5135 4. Francini S, D'Amico G, Mencucci M, Seri G, Gravano E, Chirici
5136 G (2021). Telerilevamento e procedure automatiche: validi
5137 strumenti di supporto al monitoraggio delle utilizzazioni forestali.
5138 Forest@ 18: 27-34. (In Italian) https://doi.org/10.3832/efor38355139 018

- Giannetti, F.; Pecchi, M.; Travaglini, D.; Francini, S.; D'Amico,
   G.; Vangi, E.; Cocozza, C.; Chirici, G. (2021). Estimating VAIA
   Windstorm Damaged Forest Area in Italy Using Time Series
   Sentinel-2 Imagery and Continuous Change Detection
   Algorithms. Forests, 12, 680. https://doi.org/10.3390/f12060680
   Francini S., McRoberts R.E., D'Amico G., Coops N.C.,
- Hermosilla T., White J.C., Wulder M.A., Marchetti M., Scarascia
  MugnozzaG., Chirici G., (Submitted). An open science-open data
  approach for statistically robust estimation of forest disturbance
  area. International Journal of Applied Earth Observation
- 5150 **2. Conference talks and seminars**

D'Amico G., Del Perugia B., Chirici G., Giannetti F., Nocentini
 S., Travaglini D. - "Stima della provvigione delle pinete litoranee
 di pino domestico della Toscana con dati telerilevati laser
 scanning - Spatial estimation of standing volume in Italian stone
 pine forests along the Tuscan coast with airborne laser scanning
 4<sup>th</sup> Italian

- 5157 National Congress of Silviculture, November 5 9, 2018, Turin -5158 Italy.
- D'Amico G., Francini S., Giannetti F., Vangi E., Travaglini D.,
   Mattioli W., Chianucci F., Grotti M., Puletti N., Corona P., Azzi
   N., Chirici G., Use of multitemporal Sentinel-2 imagery for
   semiautomatic classification of poplar plantations: a deep learning
   approach. Sessione Parallela 11: Innovazione nel monitoraggio
   forestale. XII Congresso Nazionale SISEF. Palermo, 12-15
   November 2019
- Borghi C., Francini S., Pollastrini M., Bussotti F., Travaglini D.,
   Marchetti M., Munafò M., Scarascia-Mugnozza G., Tonti G.,
   Ottaviano M., Giuliani C., Cavalli A., Vangi E., D'Amico G.,
   Giannetti F., Chirici G., Monitoring thirty-five years of Italian
   forest disturbance using Landsat time series. X International
   Conference AIT. Virtual Cagliari, 13-15 September 2021.
- 5172 4. Giannetti F., Tattoni C., D'Amico G., Francini S., Chirici G.,
  5173 Romano E., Brambilla M., Travaglini D., Vangi E., Chianucci F.,
  5174 Multiscale monitoring of poplar plantations using proximal and
  5175 remotely-sensed imagery. X International Conference AIT.
  5176 Virtual Cagliari, 13-15 September 2021.
- 5177 5. D'Amico G., Giannetti F., Vangi E., Borghi C., Francini S.,
  5178 Travaglini D., Chirici G., Multitemporal LiDAR data for forest
  5179 carbon monitoring in Mediterranean Forest. X International
  5180 Conference AIT. Virtual Cagliari, 13-15 September 2021.

- 5181
  6. Fanara V., Chirici G., Cocozza C., D'Amico G., Giannetti F.,
  5182
  Francini S., Salbitano F., Speak A., Vangi E., Travaglini D.,
  5183
  Estimation of Multitemporal Dry Deposition of Air Pollution by
  5184
  Urban Forests at City Scale. X International Conference AIT.
  5185
  Virtual Cagliari, 13-15 September 2021.
- Giannetti F., Giambastiani Y., Fiorentini S., Travaglini D.,
   Francini S., Vangi E., D'Amico G., Chiesi M., Maselli F., Chirici
   G., The Key Role of Multiscale Remote Sensing Data to Develop
   Forest Decision Support Systems. X International Conference
   AIT. Virtual Cagliari, 13-15 September 2021.
- 8. Vangi E., D'Amico G., Francini S., Giannetti F., Chirici G.,
  Lasserre B., Marchetti M., The New Hyperspectral Satellite
  PRISMA: Imagery for Forest Types Discrimination. X
  International Conference AIT. Virtual Cagliari, 13-15 September
  2021.
- Vangi E., D'Amico G., Francini S., Borghi C., Giannetti F.,
   Travaglini D., Chirici G., A Spatial approach for multi-temporal
   spatial estimation of forest GSV and aboveground carbon pool. A
   case of study in Tuscany (Italy). X International Conference AIT.
   Virtual Cagliari, 13-15 September 2021.
- 5201

### 5202 Acknowledgements

5203 After many years of education, it is now time to take the next step and finish my doctorate degree. Many people have supported me on this 5204 journey. Gherardo Chirici, my supervisor and co-author on all the 5205 papers, giving me the freedom to find my own research interests, and 5206 5207 who guided and supported me. The GeoLAB professors Davide 5208 Travaglini, Susanna Nocentini and Claudia Cocozza for teaching me 5209 the meaning of scientific research. I am very grateful for being given 5210 complete intellectual freedom in finding my research interests. The past three years have been an invaluable learning and work 5211 5212 experience.

5213 My everyday support is Francesca Giannetti. She is unparalleled for 5214 the countless activities she does and in which she has involved me 5215 over the years. Special thanks go to my colleagues, co-authors and 5216 good friends Dr. Saverio Francini and Dr. Elia Vangi who helped me 5217 to grow and understand the research world, but also what's outside.

Thank you to all my co-authors: Prof. Piermaria Corona, Dr.
Francesco Chianucci, Dr. Nicola Puletti, Dr. Walter Mattioli, Dr.
Mirko Grotti colleague and friend in many experiences, and also to
Prof. Antonino Nicolaci, Dr. Lorenzo Massai, Dr. Yamuna
Giambastiani, Dr. Carlo Terranova and the invaluable contribution of
Prof. Ronald McRoberts, for their help in guiding me in the complex
work of writing papers. This thesis is the result of their support.

5225 I am also grateful to Prof. Mats Nilsson and the whole SLU forest remote sensing group for their kind help during my Swedish PhD 5226 period. Pandemic postponed the departure, but the foreign period was 5227 a key formative moment. It has been very precious to work with you 5228 5229 and to learn many new things. In Umeå I found an outstanding 5230 environment and definitely multicultural, as a matter of fact I would 5231 like to thank Alex, Benoît, Chris, Ingrid, Langning, Raul and Ritwika, 5232 for the great moments even outdoors.

Thank you also to the reviewers of this thesis, Prof. Anna Barbati,
Prof. Flor Álvarez and Prof. Piotr Wężyk for their valuable comments.
I would like to thank the other GeoLAB colleagues, with whom I
shared lunches and coffee and enjoyed moments in the forest. Special
thanks to Patrizia Rossi, always present, helpful, and kind.

Thanks to Pier Carlo, Arianna, Jacopo and Chiara, faithful and
irreplaceable friends. Thanks to my basketball teammates, with whom
we shared victories and defeats, spending every week few hours of
fun.

I want to thank my parents for always supporting me. You have taught
me to follow my dreams but also to be fair and take care of others.
Irene, with Andrea and my nephews Pietro and Alessandro, you filled
these years with lots of smiles. Lorenzo e Carolina, you are an
example, a fantastic and happy couple, although for your wedding was
too hot.

- Finally, my deepest gratitude goes to Federica, a constant partner in
  my life. The last years we had many things to do, work, travel, new
  house until my thesis. So many hours of work when I should have
  spent more time together. You have supported me by helping me with
  all the other things in life, making me a better man.
- 5253
- 5254 Firenze, February 2022
- 5255
- 5256 Giovanni D'Amico