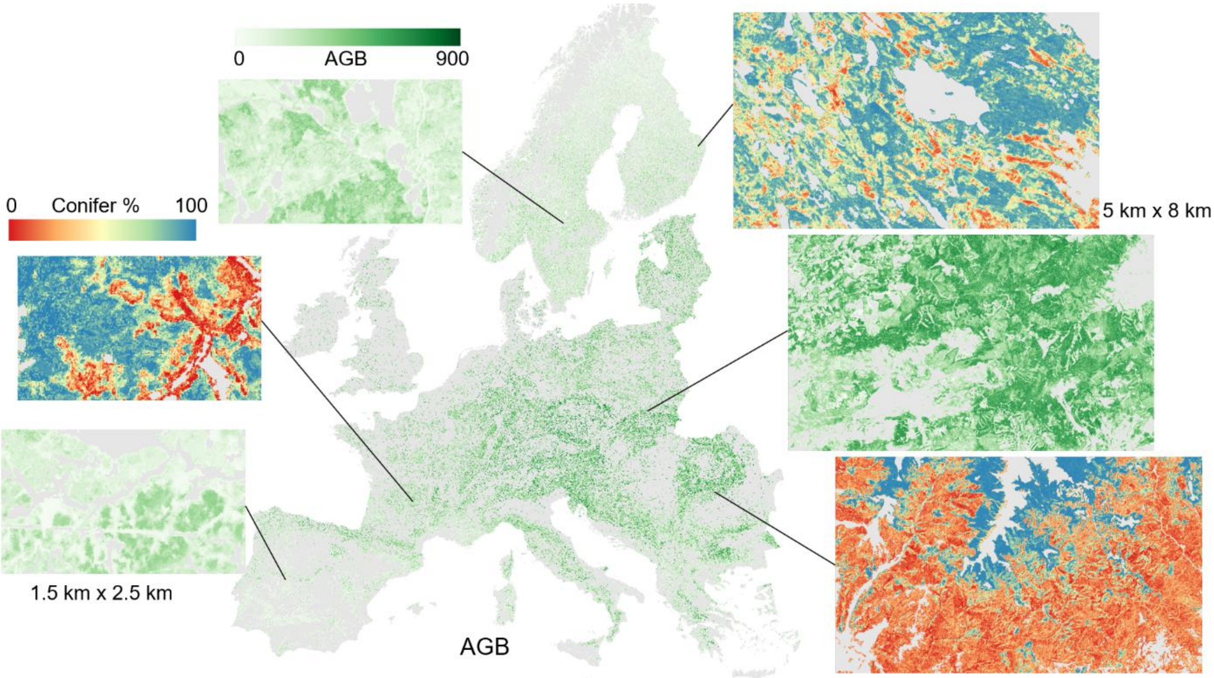
High-Resolution Pan-European Forest Structure Maps: An Integration of Earth Observation and National Forest Inventory Data



Above ground biomass (AGB) map and extracts from the AGB and Deciduous-coniferous proportion maps.

Jukka Miettinen, VTT Technical Research Institute of Finland *

Patricia	Adame	Institute of Forest Sciences (INIA, CSIC)	
Iciar	Alberdi	Institute of Forest Sciences (INIA, CSIC)	
Oleg	Antropov	VTT Technical Research Institute of Finland	
Ólafur	Arnarsson	Land and forest Iceland	Data, Communication and Innovation
Rasmus	Astrup	Norwegian Institute of Bioeconomy Research	Division of Forestry and Forest Resources

Ambros	Berger	Austrian Research Centre for Forests	Department for Forest Inventory
Jón	Bogason	Land and forest Iceland	Data, Communication and Innovation
Gherardo	Chirici	University of Florence	Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies
Piermaria	Corona	Council for Agricultural Research and Analysis of the Agricultural Economy	Research Centre for Forestry and Wood
Giovanni	D'Amico	University of Florence	Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies
Christoph	Fischer	Swiss Federal Research Institute WSL	National Forest Inventory
Florence	Gohon	National Institute of Geographic and Forest Information	National Forest Inventory
Thomas	Gschwantner	Austrian Research Centre for Forests	Department for Forest Inventory
Johannes	Hertzler	Thünen Institute of Forest Ecosystems	National Forest Inventory
Zsofia	Koma	Norwegian Institute of Bioeconomy Research	Division of Forestry and Forest Resources
Kari T.	Korhonen	Natural Resources Institute Finland	Bioeconomy and Environment
Luka	Krajnc	Slovenian Forestry Institute	Department of Forest Yield and Silviculture
Nicolas	Latte	Gembloux Agro-Bio Tech	Forest is life
Philippe	Lejeune	Gembloux Agro-Bio Tech	Forest is life
Andrew	McCullagh	Department of Agriculture, Food and the Marine	Forest Sector Development
Marcin	Mionskowski	Bureau for Forest Management and Geodesy	
Daniel	Moreno	Institute of Forest Sciences (INIA, CSIC)	

Mari	Myllymäki	Natural Resources Institute Finland	Bioeconomy and Environment
Mats	Nilsson	Swedish University of Agricultural Sciences	National Forest Inventory
Jérôme	Perin	Gembloux Agro-Bio Tech	Forest is life
Juho	Pitkänen	Natural Resources Institute Finland	Bioeconomy and Environment
John	Redmond	Department of Agriculture, Food and the Marine	Forest Sector Development
Thomas	Riedel	Thünen Institute of Forest Ecosystems	National Forest Inventory
Johannes	Schumacher	Norwegian Institute of Bioeconomy Research	Division of Forestry and Forest Resources
Lauri	Seitsonen	VTT Technical Research Institute of Finland	
Laura	Sirro	VTT Technical Research Institute of Finland	
Mitja	Skudnik	1) Slovenian Forestry Institute 2) Biotechnical Faculty, University of Ljubljana	1) Department for Forest and Landscape Planning and Monitoring 2) Department of Forestry and Renewable Forest Resources
Arnór	Snorrason	Land and forest Iceland	Research and Developement
Radosław	Sroga	Bureau for Forest Management and Geodesy	
Berthold	Traub	Swiss Federal Research Institute WSL	National Forest Inventory
Bertil	Westerlund	Swedish University of Agricultural Sciences	Department of Forest Resource Management
Stephanie	Wurpillot	National Institute of Geographic and Forest Information	National Forest Inventory

Johannes Breidenbach, NIBIO Norwegian Institute of Bioeconomy Research, Division of Forestry and Forest Resources *

The author list is sorted by last name except for the first and last authors who also serve as corresponding authors.

* corresponding authors: Jukka.Miettinen@vtt.fi, Johannes.Breidenbach@nibio.no

Abstract

We developed Pan-European maps of timber volume (V), above-ground biomass (AGB), and deciduous-coniferous proportion (DCP) with a pixel size of 10 x 10 m^2 for the reference year 2020 using a combination of a Sentinel 2 mosaic, Copernicus layers, and National Forest Inventory (NFI) data. In addition, maps of pixel-level uncertainty are provided.

For mapping, we used the k-Nearest Neighbor (kNN, k=7) approach with a harmonized database of species-specific V and AGB from 14 NFIs across Europe. This database encompasses approximately 151,000 sample plots, which were intersected with the above-mentioned Earth observation data. The maps cover 40 European countries, forming a continuous coverage of the western part of the European continent.

A sample of 1/3 of NFI plots was left out for validation, whereas 2/3 of the plots were used for mapping. Maps were created independently for 13 multi-country processing areas. Root-mean-squared-errors (RMSEs) for AGB ranged from 53 % in the Nordic processing area to 73 % in the South-Eastern area.

The created maps are the first of their kind as they are utilizing a huge amount of harmonized NFI observations and consistent remote sensing data for high-resolution forest attribute mapping. While the published maps can be useful for visualization and other purposes, they are primarily meant as auxiliary information in model-assisted estimation where model-related biases can be mitigated, and field-based estimates improved. Therefore, additional calibration procedures were not applied, and especially high V and AGB values tend to be underestimated. Summarizing map values (pixel counting) over large regions such as countries or whole Europe will consequently result in biased estimates that need to be interpreted with care.

Keywords: European Forest Monitoring System, Remote Sensing, In Situ data

Material and methods

NFI field data

Detailed descriptions of the volume (V), above-ground biomass (AGB) definitions can be found at: https://gitlab.com/nfiesta/pathfinder_demo_study/-/wikis/Codelist%20of%20target%20variables (variable ids 2 and 4). The deciduous-coniferous proportion (DCP) was calculated on plot level as the proportion of AGB of coniferous species (https://gitlab.com/nfiesta/pathfinder_demo_study/-/wikis/Codelists%20of%20subpopulations) at the plot. The deciduous proportion is the inverse of the DCP value.

A Total of 151190 plots were used for mapping and validation. Plots measured 2019-2021 were used to minimize the temporal difference between the field measurement and the Sentinel-2 mosaic (2020). However, in some countries the temporal range was extended either to include a sufficient number of plots or to include measurements from geographic areas that otherwise would have been uncovered. Because the Sentinel-2 and Copernicus data are organized in overlapping tiles (Figure 1), some sample plots are used more than one time for modelling. The spatial distribution of the plots is shown in Figure 1.

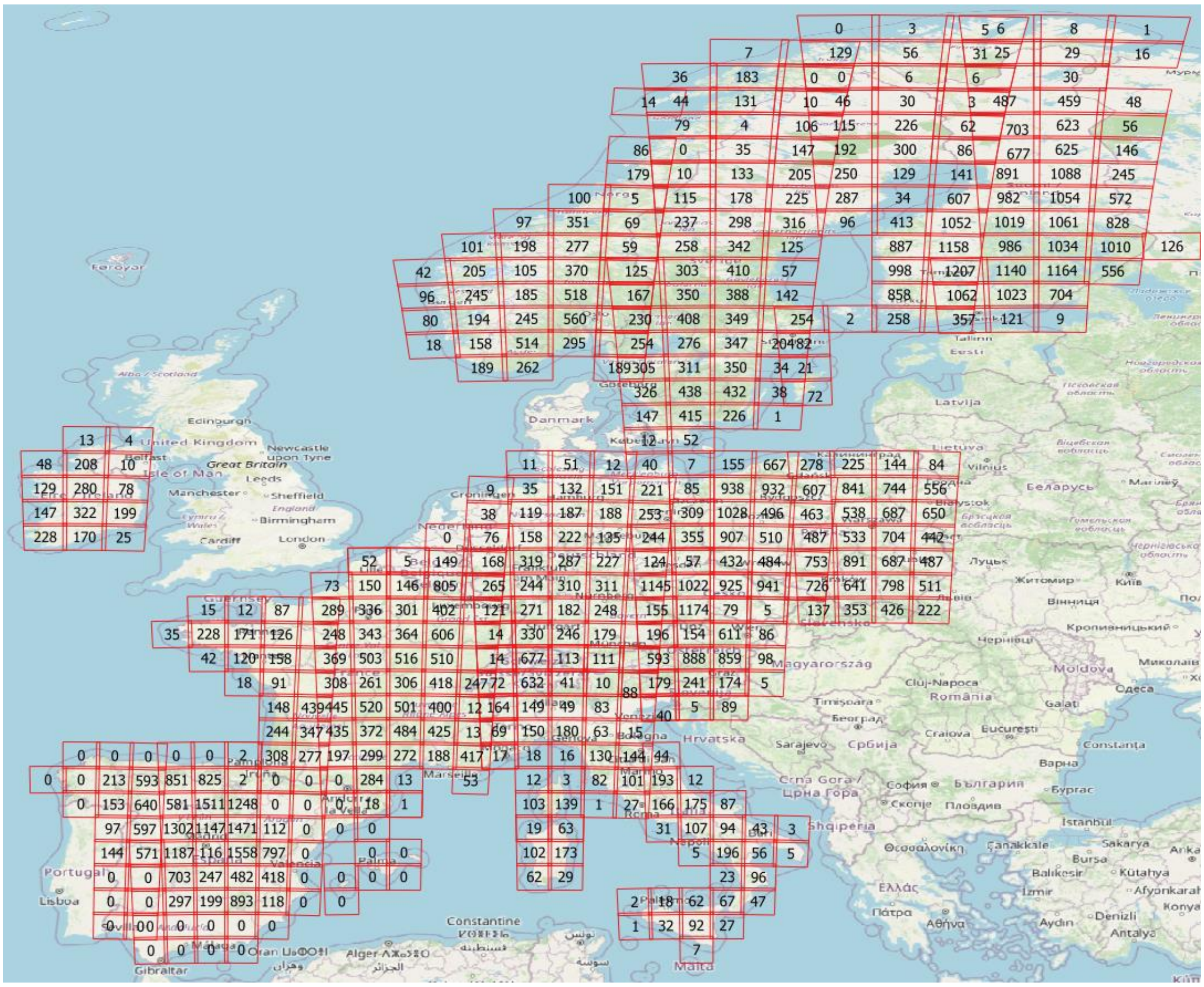


Figure 1: Number of field plots covered by Sentinel-2 tiles in countries that provided NFI data.

Remote sensing data and products

The following remotely sensed data were used in the production of the maps:

- A European-wide Sentinel-2 mosaic of composite images for the year 2020 (https://portal.forestcarbonplatform.org/) was used as the primary remote sensing data source. The image compositing algorithm using Level 2A surface reflectance products is described in Miettinen et al. (2021). The final composite images included seven spectral bands (B2, B3, B4, B5, B8, B11 and B12), all resampled into 10 m spatial resolution.
- Copernicus High Resolution Layers for forest in 10 m resolution with reference year 2018 are products derived from Sentinel-2 data. We used Forest Type (FTY; coniferous forest, broadleaved forest, non-forest) and Tree Cover Density (TCD; 0-100%) (EEA 2021).
- Global Forest Change (GFC, (Hansen et al. 2013)) is a Landsat-based product (ca. 30 m resolution) providing the year of forest canopy loss. We used version 1.9 that included the years of forest canopy losses between and including 2001 and 2021.

The Sentinel-2 and Copernicus data created the feature space when choosing the nearest neighbours (see more details below). The Global Forest Change product was used only in screening plots that had experienced changes after the field measurements.

The Sentinel-2 mosaic and thus the final maps cover a total of 40 European countries. Besides the 27 EU countries, this includes Albania, Andorra, Bosnia and Herzegovina, Holy See, Liechtenstein, Monaco, Montenegro, North Macedonia, Norway, San Marino, Serbia, Switzerland and United Kingdom.

Extracting remotely sensed data at NFI plots

The digital numbers of the Sentinel-2 and the TCD data are continuous values and the weighted mean for a circle with a size of 100 m^2 centered on each NFI sample plot was calculated from them. The pixel proportions covering the circle were used as weights. For the categorical variables (FTY and GFC), the weighted mode was calculated for the same circle. The weighted mode is the category with the greatest sum of weights. The calculations were done using R by each NFI organization, in order to use the exact plot coordinates. A 100 m^2 circle was chosen because it corresponds to the pixel size of the Sentinel-2 data which reduces distortions resulting from the calculation of weighted means.

Application of the kNN approach

The k Nearest Neighbour (kNN) approach was used with ‘feature banks' including target variables (V, AGB, and DCP) and remotely sensed variables (seven Sentinel-2 bands, TCD, and FTY) for each NFI plot. Furthermore, 1 km INSPIRE grid cell locations (northing and easting) were included in the feature space to enable utilization of the plot location when selecting the nearest neighbours.

The data were screened to 1) exclude cloud contaminated observations (based on composite image quality band accepting only digital numbers greater than 4000) and 2) exclude plots where changes had been detected by GFC since 2018 (or a year before the earliest plot measurements; Table 1). The final data were sorted by volume and every third plot was extracted into a validation feature bank. The remaining two thirds of the plots were used for the mapping feature bank.

Table 1: Number of sample plots used for mapping and validation by country.

Country code	Measurement years	Number of plots used for mapping	Number of plots used for validation
AT	2019-2021	3816	1908
BE	2019-2021	934	466
CH	2019-2022	2168	1083
CZ	2018-2020	6902	3450
DE	2017	5735	2867
ES	2018-2021	12592	6296
FI	2019-2021	21432	10716
FR	2019-2021	12390	6195
IE	2020-2022	1185	592
IT	2018-2019	2813	1406
NO	2019-2021	5908	2954
PL	2019-2021	15754	7877
SE	2019-2021	8602	4300
SI	2018	566	283
TOTAL	-	**100797**	**50393**

For the production of the maps, 13 processing areas were created taking into account the geographical areas of Europe and the availability of field sample plots (Figure 2). Six of the processing areas contained NFI plots, while the remaining processing areas did not have any NFI plots. For those areas without NFI plots, feature banks were created using plots from ecologically similar sourcing areas (Table 2).

In regions with NFI plots, the processing area borders were located in the middle of countries, but all plots from the countries that were (partially) covered by a processing area were used for mapping. This approach ensured that in adjacent processing areas, plots from both sides of the processing area border were used, resulting in a smooth transition between the processing areas with no visible changes in the map attributes at the borders of the processing areas.

For areas where NFI plots were available, the INSPIRE 1 km grid northing and easting were used as features in the selection of the nearest neighbours. For processing areas with no NFI plots, the INSPIRE 1 km grid locations were not used among the attributes to search for the closest neighbours. An exception was processing area 2 (Figure 2), where northing was used.

Euclidean distance was used to select the nearest neighbors in the feature space. The k-NN prediction $\hat{y}_p$ for pixel p is given by

$$\hat{y}_p = \sum_l w_l y_l \qquad (1)$$

where y_l is the vector of observations and w_l is the weight of the l'th nearest neighbor with l=1,…,k. Weights inverse to the Euclidian distance were used. In addition to the prediction, we calculated the standard deviation $\hat{s}_p$ of the nearest neighbors for each pixel as a measure of uncertainty

$$\hat{s}_p = \sqrt{\frac{\sum_l (y_l - \hat{y}_p)^2}{k}}. \quad (2)$$

All processing was conducted in the Forestry TEP (https://f-tep.com/) by Sentinel-2 tiles. A total of 745 Sentinel-2 tiles were processed.

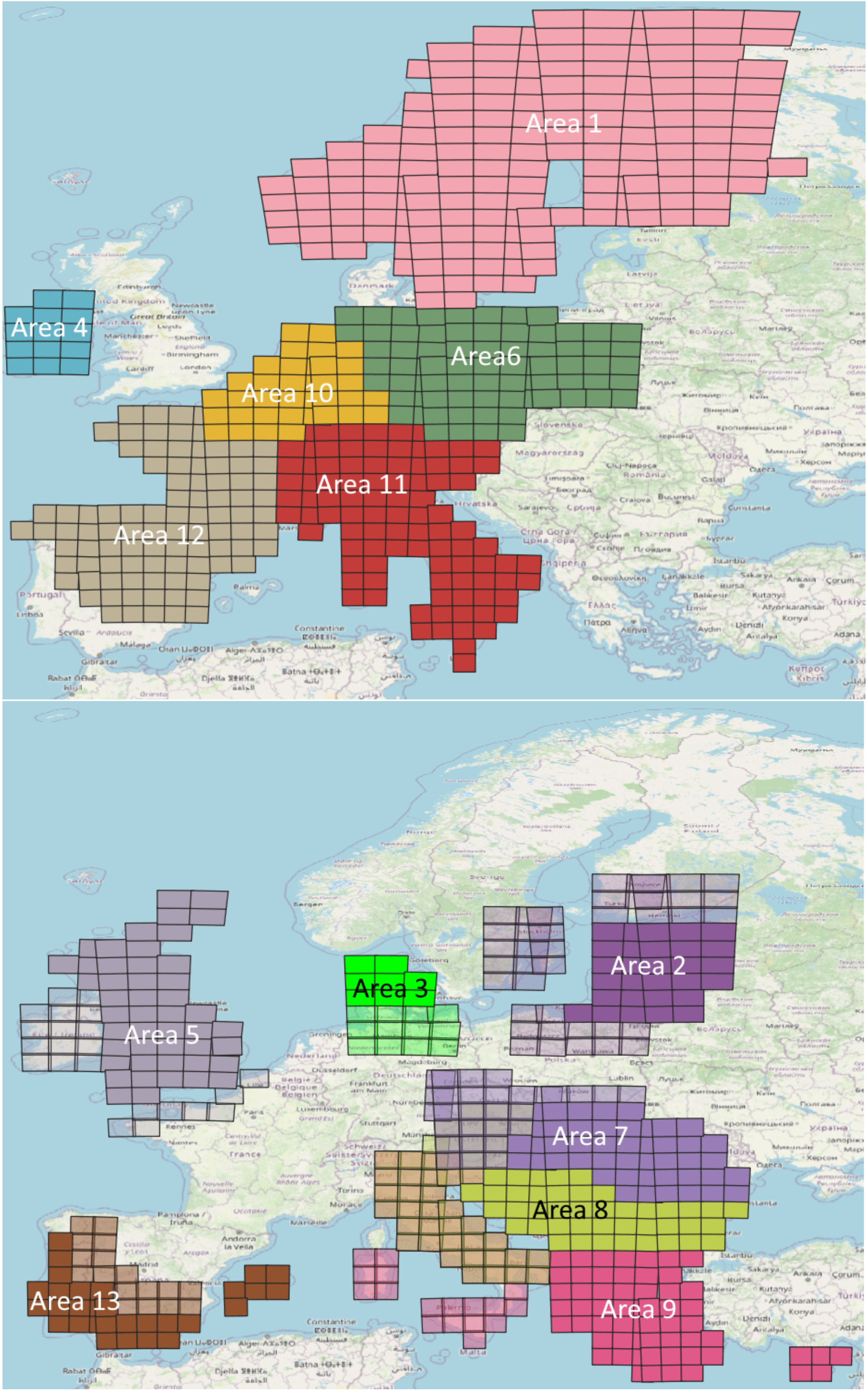


Figure 2: Sentinel-2 tiles colorized by processing area. Top: processing areas where NFI plots were available and location information was used in modelling. Bottom: processing areas where NFI plots were not available and location information was not used in modelling (except for northing in Area 2). Plot sourcing areas are shown in transparent color.

Table 2: Key statistics of the NFI data used for mapping in each processing area (sourcing area in case of processing areas without NFI plots).

Processing area	Number of mapping plots	Mean V	SD V	Max V	Mean AGB	SD AGB	Max AGB	Mean DCP (%)	SD DCP (%)
1	35942	130	111	950	75	59	546	74	32
2	15836	216	158	1266	127	90	918	68	37
3	913	339	193	1091	180	107	684	58	44
4	1185	200	189	959	142	105	579	71	42
5	1598	212	189	1015	147	107	598	57	46
6	32207	313	210	1360	190	121	1753	60	42
7	6062	324	235	1316	201	140	1238	60	42
8	4302	246	229	1390	153	127	1000	42	45
9	2181	165	166	1218	118	104	681	18	35
10	19059	273	201	1358	164	111	862	34	43
11	27488	277	215	1360	166	118	1101	40	44
12	24982	154	160	1292	110	96	1044	39	45
13	5221	59	78	632	62	60	614	41	47

Error metrics

Root-mean-square-error (RMSE) and bias were used to evaluate the mapping results. These metrics were calculated by predicting the response variables for the plots in the validation feature bank.

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$Bias = \frac{\sum_i (y_i - \hat{y}_i)}{n} \tag{4}$$

where *y* represents the observed values, $\hat{y}$ represents the predicted values, i=1,…,*n* indexes the observations, and *n* is the number of observations in the validation feature bank. Both of these values were also compared to the mean value of the variable in the validation feature bank, deriving relative metrics (RMSE and bias in per-cent).

Maps

Maps of the three different target variables are made available including V, AGB, and DCP. While the DCP map provides the percentage of conifers, this allows straightforward calculation of the broadleaf proportion as 100-DCP. Figure 3 illustrates the AGB map with subsets of AGB and DCP maps from different parts of Europe.

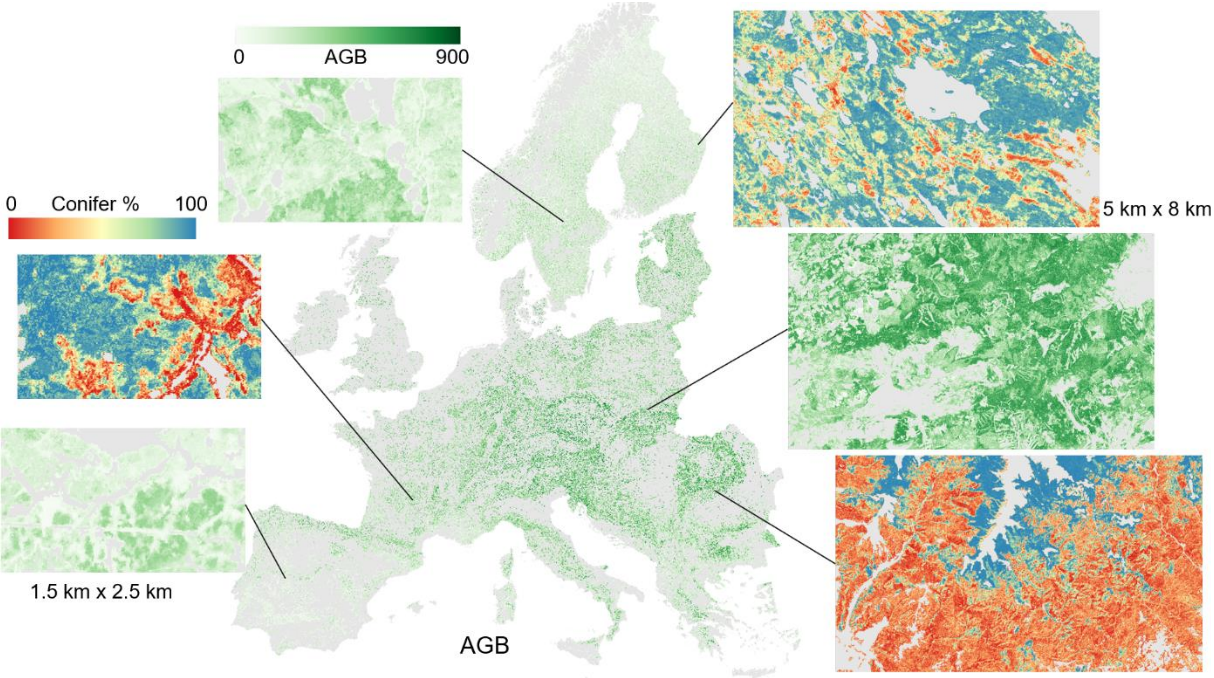


Figure 3: European wide AGB map and extracts from the AGB and DCP maps.

In addition to the target variable maps, also standard deviation layers are provided for each target variable. All of the maps were masked with the FTY forest extent. The other areas have been masked as non-forest and no-data (Table 3).

Table 3: Output map technical characteristics.

Variable	Variable naming	Pixel values	Format
Volume	vol stdev_vol	65535: No data 65534: Non-forest Other values: m³/ha	10 m UInt16 GeoTiff
Above Ground Biomass	agb stdev_agb	65535: No data 65534: Non-forest Other values: t/ha	10 m UInt16 GeoTiff
Conifer proportion	P_agb_conifers stdev_P_agb_conifers	65535: No data 65534: Non-forest Other values: %	10 m UInt16 GeoTiff

All maps are provided in 10 m spatial resolution in 500 x 500 km tiles in the EPSG:3035 - ETRS89-extended / LAEA Europe projection (Figure 3). The file naming follows the following pattern:

'Year'_'variable'_'tile'.tif, with variable names as defined in Table 3.

For example, the standard deviation map of volume for the 500 x 500 km grid with the lower left corner E 3 900 000 m and N 2 400 000 m is ‘2020_stdev_vol_E39_N24.tif’.

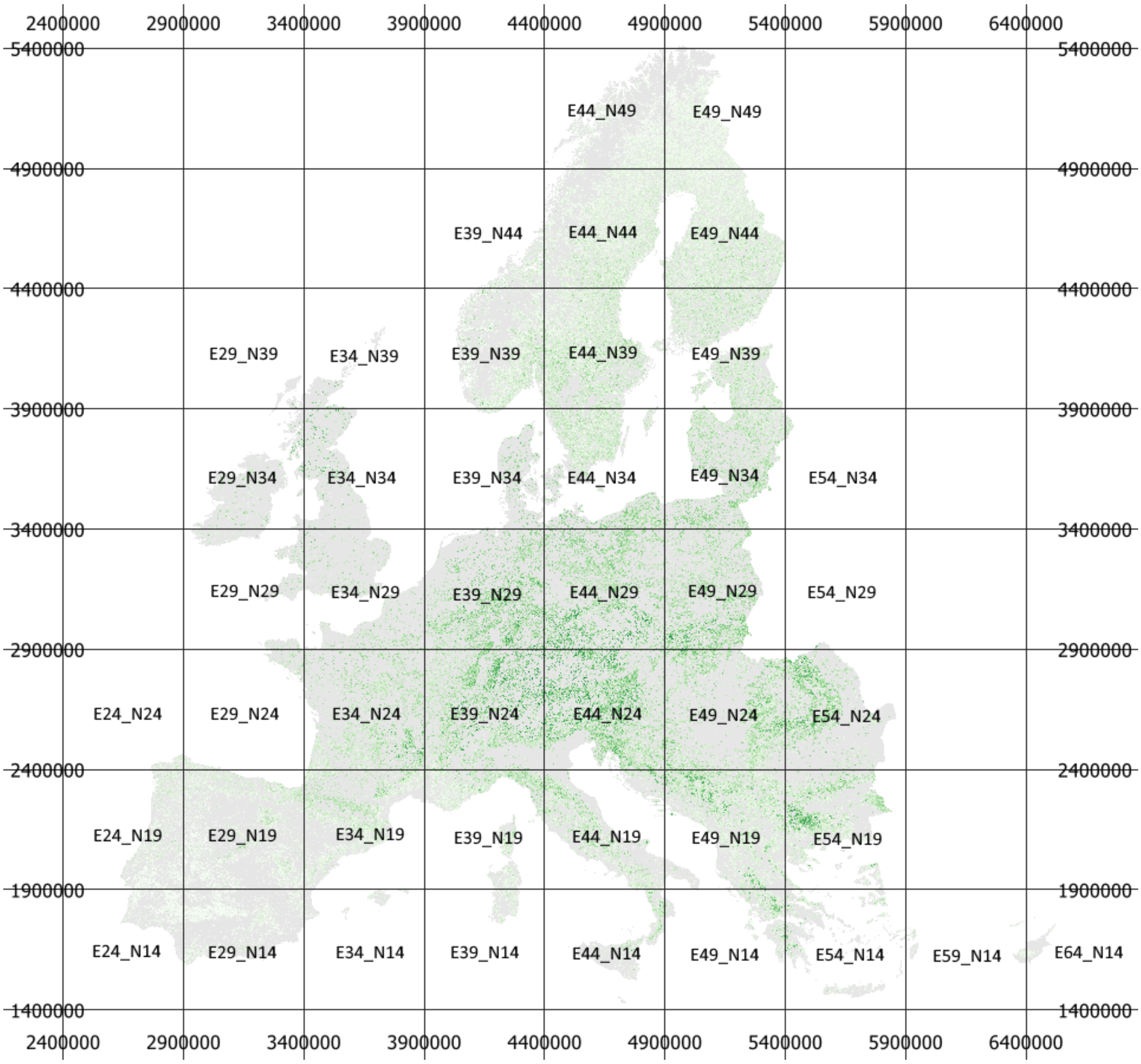


Figure 4: Maps are provided in 10 m spatial resolution in 500 x 500 km tiles in the EPSG:3035 - ETRS89-extended / LAEA Europe projection. Volume map as background in the image

Error metrics are provided in Table 4 and Figure 5.

Table 4: Plot level error metrics for the 13 processing areas calculated from the validation feature banks. Areas where plots were available and the plot location was used in the selection of the neighbours are in blue, whereas areas where plots were not available are in green.

Processing area	Vol RMSE	Vol RMSE %	Vol Bias	Vol Bias%	AGB RMSE	AGB RMSE %	AGB Bias	AGB Bias%	DCP RMSE	DCP RMSE %	DCP Bias	DCP Bias%
Area 1	75,71	58,23	-0,77	-0,59	40,17	53,23	-0,70	-0,93	22,30	30,36	1,06	1,45
Area 2	124,53	57,77	0,27	0,12	69,97	55,35	0,03	0,02	22,8	33,7	0,6	0,89
Area 3	187,73	55,38	0,67	0,05	103,39	58,03	0,54	0,30	22,58	37,71	-0,17	-0,28
Area 4	129,80	64,81	-0,52	-0,26	77,81	54,03	-1,05	-0,73	28,35	42,13	0,05	0,08
Area 5	146,38	71,95	9,27	4,56	90,57	63,44	4,33	3,03	26,98	50,56	0,74	1,38
Area 6	177,37	56,60	8,16	2,60	107,50	56,56	4,76	2,50	22,30	36,84	0,29	0,47
Area 7	200,96	61,94	7,89	2,43	127,22	63,36	5,26	2,62	24,83	41,30	-0,40	-0,66
Area 8	181,34	73,64	5,95	2,42	110,37	72,03	2,70	1,76	24,91	59,12	-0,08	-0,18
Area 9	126,97	76,85	2,28	1,38	86,14	73,43	3,33	2,84	19,08	110,3	-0,89	-5,13
Area 10	168,59	61,72	1,65	0,6	98,07	59,45	0,78	0,47	24,11	71,56	0,09	0,27
Area 11	178,18	64,27	5,16	1,86	103,57	62,31	2,69	1,62	24,24	61,82	0,11	0,30
Area 12	114,99	74,56	-0,88	-0,57	74,38	67,55	-0,60	-0,54	25,15	64,77	-0,32	-0,82
Area 13	49,53	83,92	-2,06	-3,48	44,62	71,64	-1,00	-1,60	26,53	65,47	-0,48	-1,20

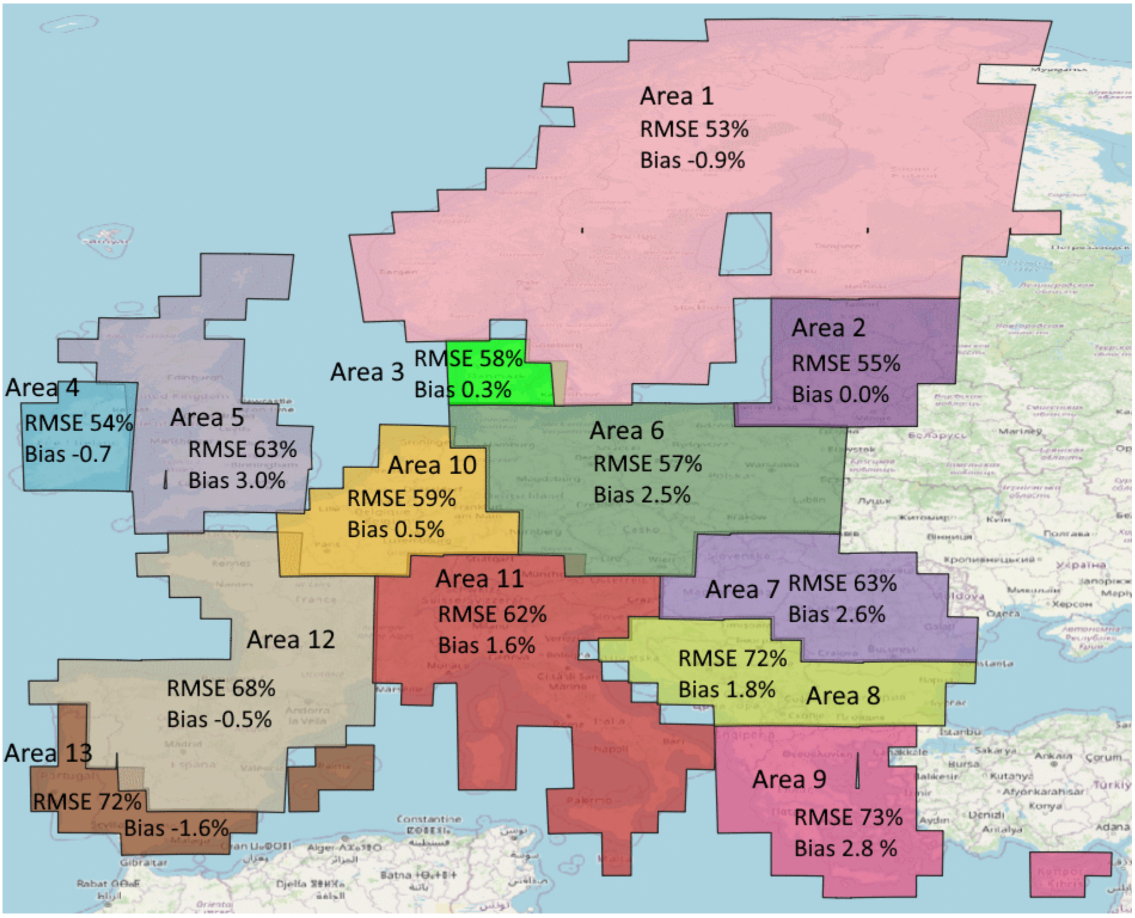


Figure 5: Relative accuracies of AGB by processing areas.

Error metrics have been calculated by using the kNN model based on the mapping feature banks to predict values for the plots in the validation feature banks. For the processing areas where NFI plots are available, this approach can be expected to produce reliable error metrics. However, for the processing areas that do not have plots (i.e. Areas 2, 3, 5, 7, 8, 9 and 13), the error metrics should be treated with caution. For these areas, the mapping and validation feature banks were manually compiled using plots from ecologically similar regions. Although the selection was made with best available knowledge of the areas, it must be assumed that the error metrics for the areas without field plots may be overly optimistic.

Scatterplots based on the based on the validation data show that the bulk of the data follow the 1:1 line. However, high values tend to be underestimated and low values tend to be overestimated (Figure 6).

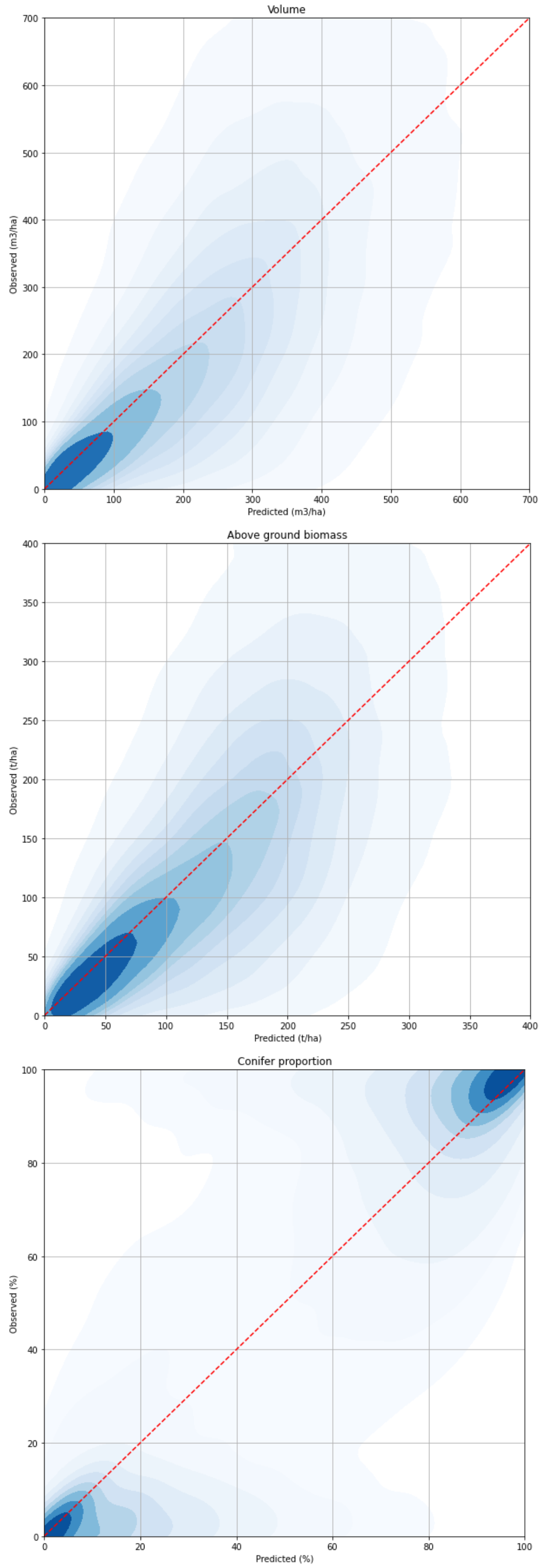


Figure 6: Scatterplots of observed and predicted V, AGB, and DCP based on the validation data. Darker colors indicate a higher density of observations.

Acknowledements

This study received funding from the EU under GA101056907 (PathFinder).

References

EEA. 2021. Copernicus Land Monitoring Service - User Manual - High Resolution land cover characteristics - Tree-cover/forest and change 2015-2018.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S., Goetz, S.J., and Loveland, T.R. 2013. High-resolution global maps of 21st-century forest cover change. science **342**(6160): 850-853.

Miettinen, J., Carlier, S., Häme, L., Mäkelä, A., Minunno, F., Penttilä, J., Pisl, J., Rasinmäki, J., Rauste, Y., and Seitsonen, L. 2021. Demonstration of large area forest volume and primary production estimation approach based on Sentinel-2 imagery and process based ecosystem modelling. International Journal of Remote Sensing **42**(24): 9467-9489.