FOREST CARBON MONITORING

CCN2

Algorithm Theoretical Basis Document (ATBD), Update

Funded by

esa

COORDINATOR

VTT

VTT Technical Research Centre of Finland Ltd
Tekniikantie 21, Espoo
P.O. Box 1000
FI-02044 VTT
Finland

PARTNERS

EUROPEAN FOREST INSTITUTE

NIBIO
NORWEGIAN INSTITUTE OF BIOECONOMY RESEARCH

GAMMA REMOTE SENSING

south pole

GFZ
Helmholtz Centre Potsdam

YUCA

Terramonitor

Document details:

Project full title:	Forest Carbon Monitoring
ESA Contract No.	4000135015/21/I-NB
VTT project reference	131034
Document reference	VTT-CR-00187-25
Deliverable No.	CCN2-D07
Issue:	2.0
Date:	05.05.2025
Dissemination level:	Project team
Authors:	Jukka Miettinen, VTT Oleg Antropov, VTT Jorma Kilpi; VTT Laura Sirro, VTT Maurizio Santoro, Gamma RS Oliver Cartus, Gamma RS Martin Herold, GFZ Arnan Araza, GFZ/WUR Johannes Breidenbach, NIBIO Zsofia Koma, NIBIO Basanta Gautam; South Pole Alejandra Monsalve; South Pole Lauri Häme, Terramonitor Tuomas Häme, Terramonitor Francesco Minunno, Yucatrote Ritika Srinet, Yucatrote

Distribution:

ESA
Forest Carbon Monitoring Consortium

Change log:

Issue	Date	Status (draft / updated / final)	Remarks
0.0	6.4.2022	Draft	Template and table of contents (ToC) created
0.1	19.6.2022	Draft	Ready for internal review
1.0	23.6.2022	Final FCM D2.1	To be submitted to ESA
1.1	31.3.2025	Draft of CCN2 update	Template and ToC of the updated version
1.1	27.4.2025	Draft of CCN2 update	Ready for internal review
2.0	5.5.2025	Final	To be submitted to ESA

Table of Contents

List of figures

List of tables

Symbols and acronyms

ALOS	The Advanced Land Observing Satellite
AGB	Above Ground Biomass
ALS	Airborne Laser Scanning
ATBD	Algorithm Theoretical Basis Document
CDF	Cumulative Distribution Function
CEOS	Committee on Earth Observation Satellites
CoSSC	Coregistered Single look Slant range Complex
DEM	Digital Elevation Model
EO	Earth Observation
ESA	European Space Agency
FCM	Forest Carbon Monitoring
GLAS	The Geoscience Laser Altimeter System
GRD	Ground Range Detected
GSV	Growing Stock Volume
HH	Horizontal transmit Horizontal receive polarizations
HV	Horizontal transmit Vertical receive polarizations
ICESat	Ice, Cloud and land Elevation Satellite
IPCC	Intergovernmental Panel on Climate Change
InSAR	Interferometric SAR
IW	Interferometric Wide swath mode
LPV	Land Product Validation
JAXA	Japan Aerospace Exploration Agency
LiDAR	Light Detection And Ranging
NICFI	Norway's International Climate and Forest Initiative
NFI	National Forest Inventory
PALSAR	Phased-Array L-band Synthetic Aperture Radar
SAR	Synthetic Aperture Radar
SRS	Stratified Random Sampling
SRTM	Shuttle Radar Topography Mission
TanDEM-X	TerraSAR-X add-on for Digital Elevation Measurements
UTM	Universal Transverse Mercator
VH	Vertical polarization transmit and Horizontal polarization receive
VV	Vertical polarization transmit and Vertical polarization receive

1. Introduction

The Forest Carbon Monitoring (FCM) project developed remote sensing-based, user-centric approaches for forest carbon monitoring. The project implemented a set of tools for monitoring of forest structural variables, biomass and carbon stock. In the main project (July 2021 - June 2023), a prototype platform was successfully implemented and its functionalities demonstrated on nine use cases. In the continuation project (FCM CCN2; May 2024 – November 2025) the selection of available tools was widened, and two additional use case demonstration areas were added. The continuation of the project addressed shortcomings identified by the users during the main project, to close gaps between user expectations and the available tools.

The document at hand, *'CCN2-D07 Algorithm Theoretical Basis Document (ATBD), Update'*, provides scientific basis for the tools offered in the FCM toolbox. In addition to the description of the algorithms, also the uncertainty estimation methods and main datasets used during the project are described. For each tool, examples of the performance in the FCM use case demonstrations have been reported to provide an indication of the level of expected uncertainty of the output products.

In addition to this introduction, the 'CCN2-*D07 Algorithm Theoretical Basis Document (ATBD)'* contains five main sections:

- *FCM concept and tools*, which provides an overview of the Forest Carbon Monitoring concept and available tools.
- *Primary datasets*, which describes the primary datasets that were used in the development and demonstration of the tools.
- *Statistical approaches*, which describes the statistical approaches used in the evaluation of the uncertainty and utilisation of the of the output products.
- *Algorithm descriptions*, which provides descriptions of the algorithms and approaches underlying the FCM tools, including the levels of uncertainty reached in the use case demonstrations.
- *Conclusion,* which wraps up the main message of the ATBD document and provides users with advice on how to select algorithm for their use case.

2. Forest Carbon Monitoring concept and tools

2.1 Forest Carbon Monitoring concept

The Forest Carbon Monitoring concept (Figure 1) aims to provide a toolset to support monitoring of forest structural variables, biomass and carbon stock. The underlying idea of the concept is that a set of tools is needed to meet the highly varying requirements by different types of stakeholders. The goal of the toolset is to be able to provide optimal tools for satellite-based forest monitoring tasks depending on the available datasets and specific user requirements. Key aspects of the FCM approach include:

- Maximising the integration of in-situ data whenever available
- Integration of process-based forest ecosystem carbon modelling into the system
- Flexibility to user needs ranging from private company area monitoring to continental analyses

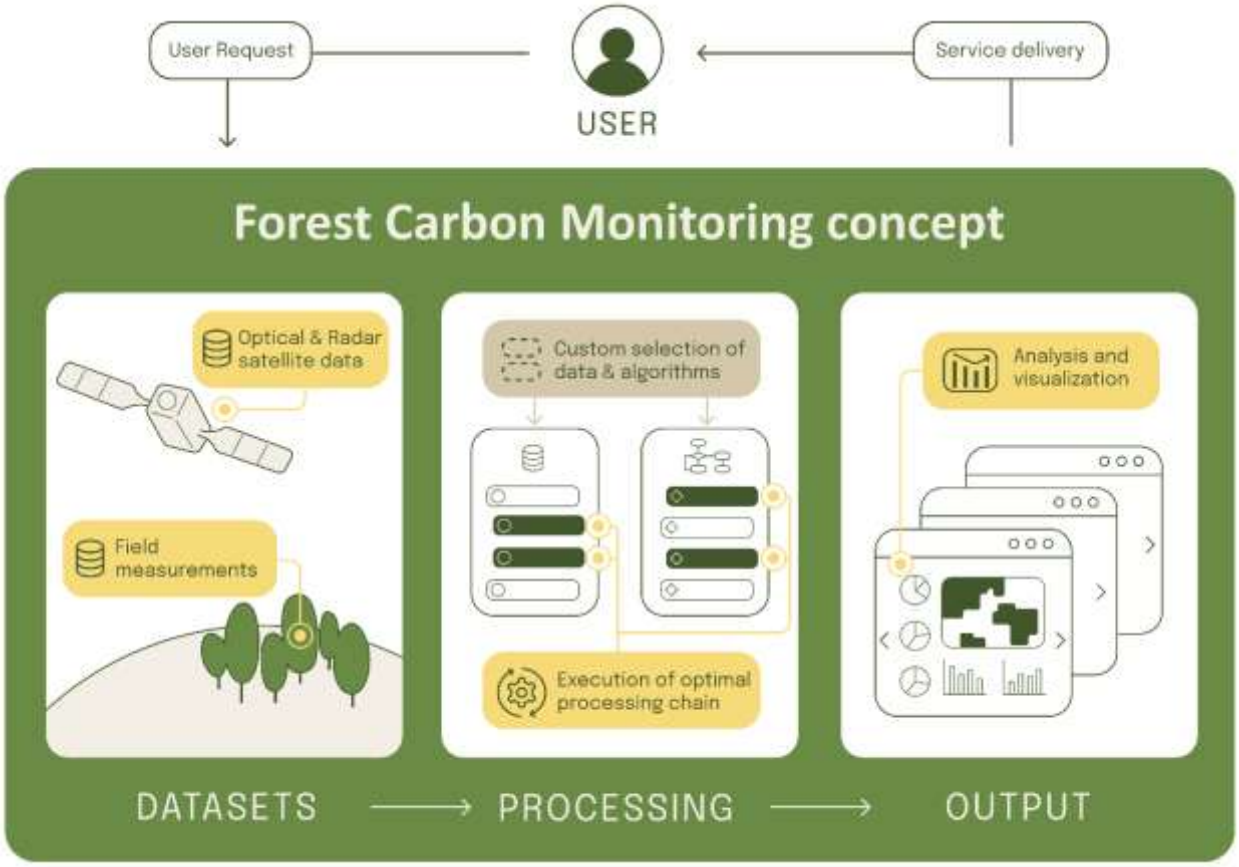

Figure 1. High-level illustration of the Forest Carbon Monitoring concept.

The algorithms and tools were developed together with the use partners and each tool was implemented and evaluated in one of the 11 use cases described in the next section. The findings from the use case demonstrations provided valuable information on the performance of the algorithms in varying ecosystems and with varying availability of Earth Observation (EO) and reference datasets.

2.2 Available tools

While the FCM project mainly concentrated on prediction of forest biomass and carbon variables, other forest variables were also needed in the prediction. Traditional forest inventory variables were used as inputs for biomass modelling, or growing stock volume was converted to biomass estimates using conversion factors. Furthermore, many users required information on traditional forest inventory variables as well (such as basal area, diameter, height). These basic forest variables are needed to support forest management decisions but also allow biomass or carbon flux prediction when required.

Figure 2 provides an overview of the algorithms used in the Forest Carbon Monitoring tools. The tools can be divided into four main groups: 1) Pre-processing, 2) Forest structure mapping, 3) Ecosystem modelling and 4) Change detection. Although most of the tools developed in the FCM project are flexible regarding the input data, the integrated pre-processing tools make the implementation of processing workflows more fluent. All the forest structure mapping tools can take single date EO imagery or analysis ready products as inputs, but the Sentinel-1 and Sentinel-2 compositing tools enable creation of feasible input data in cases where suitable single date imagery or analysis ready products are not readily available.

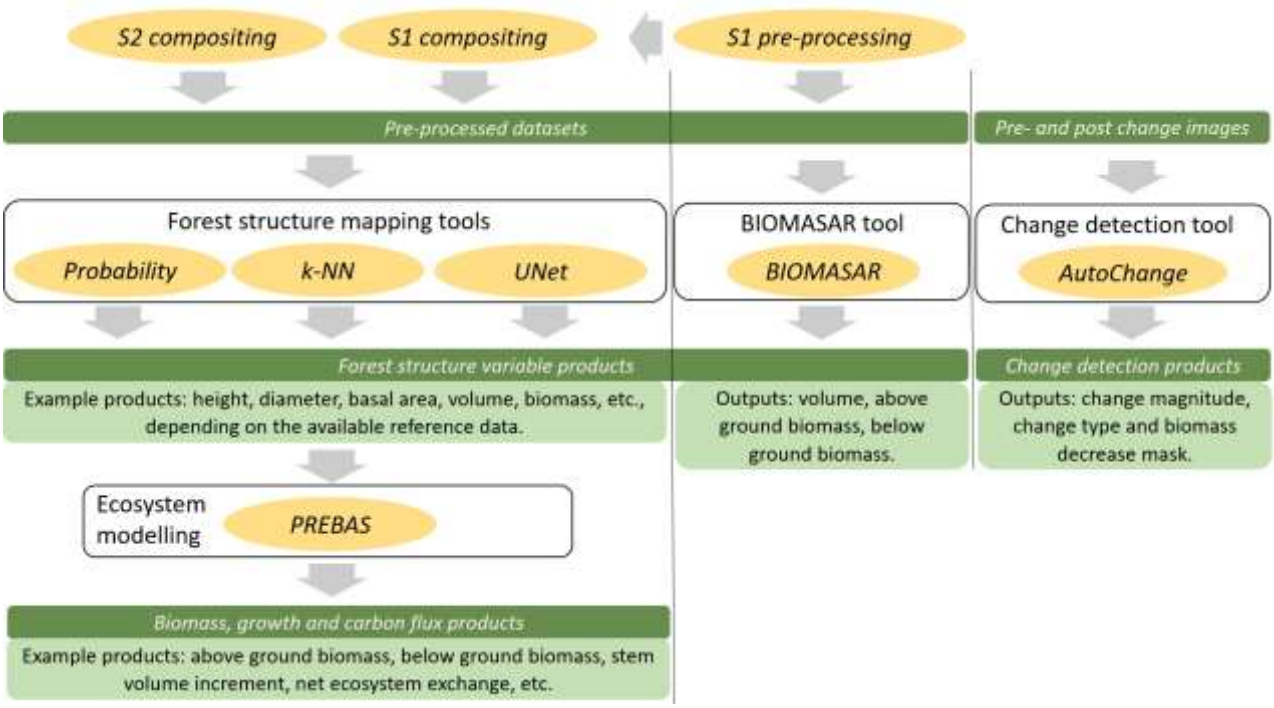

Figure 2. Overall workflow of algorithms (orange ovals) used in the Forest Carbon Monitoring.

The Forest structure mapping tools form the core of the FCM toolbox. Three of the tools (namely Probability, k-NN and UNet) are highly versatile tools that can work with multi-sensor datasets and produce predictions of a wide range of variables, depending on the availability of datasets and user requirements. Naturally, correlation of the EO data features with the target variable features is a pre-requisite for any meaningful forest variable prediction. The BIOMASAR tool is a more specialised tool designed for growing stock volume and biomass prediction using radar datasets (typically a combination of C and L-band data).

The ecosystem modelling tool PREBAS enables modelling of biomass and carbon stocks and fluxes for current situation as well as future forecasting. As a process-based ecosystem model, variations of climatic and other environmental or anthropogenic factors can be taken into account while producing forecasts with varying future scenarios.

In addition to the forest structure mapping and ecosystem modelling tools, also a change detection tool called Autochange is provided in the FCM toolbox. This is a versatile generic image-to-image change detection algorithm that accepts a wide range of input data types and is resistant to general level differences in the pre- and post-change imagery. The tool provides change magnitude as its main output.

The tools described above have been extensively tested in 11 use case demonstration sites during the FCM project (Figure 3 and Table 1). Each of these sites had a dedicated user partner with specific requirements. The availability of EO and reference datasets varied drastically between the use cases. These variations gave an excellent opportunity to evaluate the usability of tools in a wide range of situations.

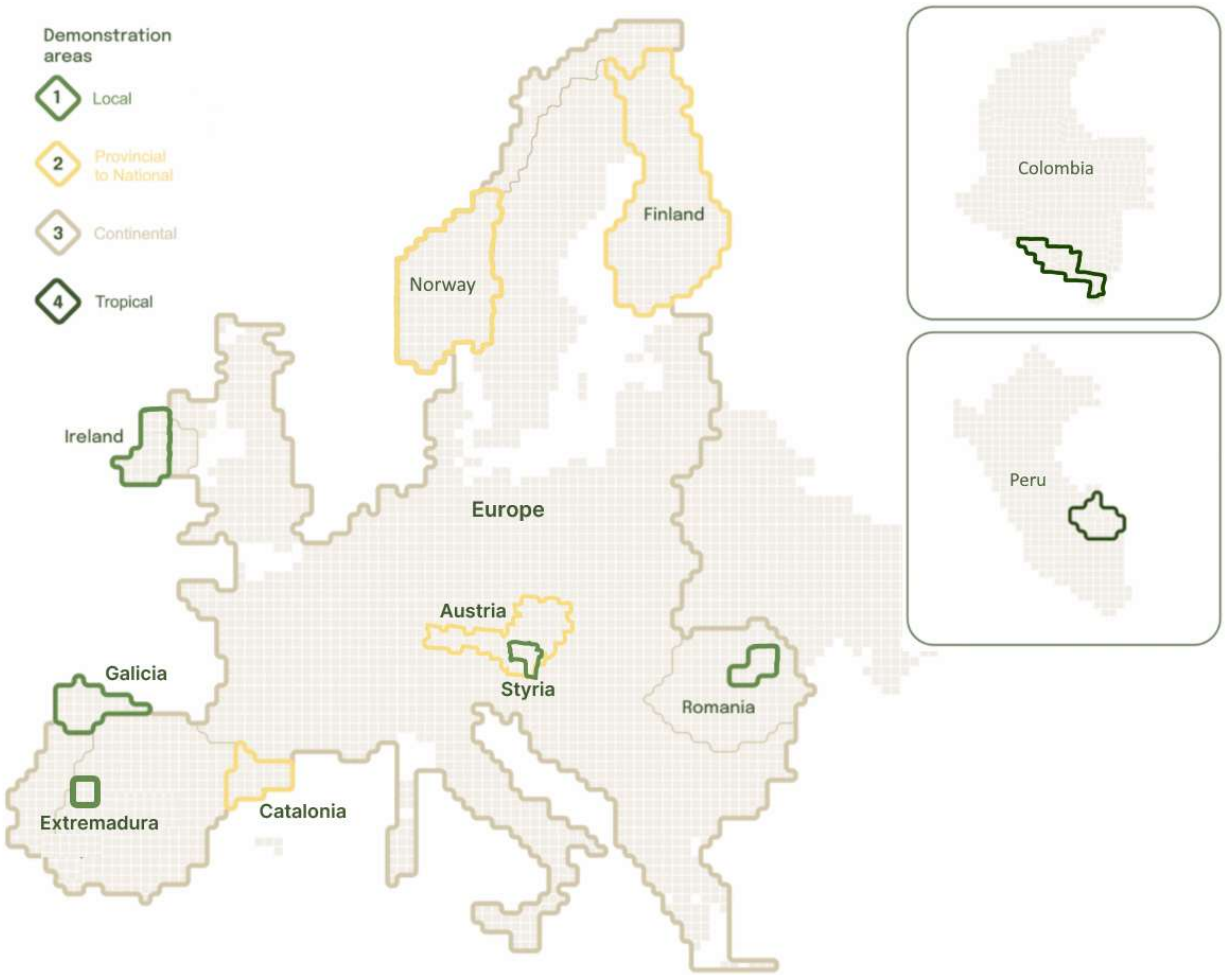

Figure 3. Geographic distribution of FCM use case demonstrations.

Table 1. Characteristics of FCM use case demonstrations.

Area	Size (S2 tiles)	Primary datasets[1]	Primary algorithms	Output variables[2]	Years
Galicia	5	S2, S1 Private field plots	Probability PREBAS	D, G, H, GSV, AGB, BGB, SVI	2019 + 2020 +2021
Ireland	8	S2, S1 Private field plots	Probability PREBAS	N, D, G, H, GSV, Species% (4), AGB, BGB, SVI	2019 + 2020 +2021
Romania	3	S2, S1 Private field plots	k-NN PREBAS	D, G, H, GSV, Species% (2), Site, AGB, BGB, SVI	2019 + 2020 +2021
Extremadura	1	S2 Private field plots	Probability PREBAS	D, G, H, AGB, BGB, SVI	2017 + 2022
Styria	3	S2, S1 Private field plots	Probability PREBAS	D, G, H, GSV, Species% (2), AGB, BGB, SVI	2015 + 1018 + 2021
Catalonia (+ Andorra)	8	S2, S1[3] NFI + private field plots	kNN UNet	N, D, G, H, GSV, AGB	2020 + 2021 + 2023 + 2024
Norway	35	S2, S1, P2 NFI field plots	kNN Unet PREBAS	D, G, H, GSV, Species% (3), AGB, BGB, SVI	2017 + 2019 + 2021 + 2023
Finland	63	NFI multisource maps	PREBAS	AGB, BGB, SVI	2017 + 2019
Europe	746	S1, P2, IceSat-2 NFI field plots LiDAR reference	BIOMASAR	GSV, AGB, BGB	2017 + 2020 + 2021 + 2023
Colombia	16	S2, P2, NICFI Field campaign	Two-step sampling	AGB	2023/2024
Peru	16	S2, P2 NFI field plots	Probability	D, G, H, GSV, AGB	2020 + 2021

[1)] S2 = Sentinel-2, S1 = Sentinel-1, P2 = PALSAR 2, NICFI = NICFI Planet mosaic, NFI = National Forest Inventory

[2)] N = stem density, D = diameter, G = basal area, H = height, Species% (x) = species proportions (of basal area) for x species or species groups, Site = site type, GSV = growing stock volume, AGB = above ground biomass, BGB = below ground biomass, SVI = stem volume increment

[3)] S1 not used in project continuation phase

The lessons learned in these use case demonstrations regarding the uncertainty of the output products are reported in this document under the 'Performance'-subsections within each algorithm description. Further information regarding output product uncertainties in specific conditions or general applicability of the tools can be obtained by contacting the FCM team through the website (https://www.forestcarbonplatform.org/) or through the Forestry TEP platform (https://f-tep.com/).

3. Primary data sources

3.1 General note regarding input data

It is important to note that most of the algorithms utilised in the FCM tools are capable of using multiple data sources. The primary data sources presented in this chapter are the ones that were used in the testing phase or in the use case demonstrations during the Forest Carbon Monitoring project. Three FCM tools have also been created for preprocessing of Sentinel-1 and Sentinel-2 data. These are also the primary data sources for the FCM tools. As most of the tools are flexible regarding input data, the best set of input data sources can be decided case-by-case depending on the availability of datasets and the objectives of the mapping.

The primary data sources presented in this chapter include:

1. Sentinel-2 satellite imagery, with a compositing algorithm
2. Sentinel-1 satellite data, with description of pre-processing steps and compositing algorithm
3. ALOS-2 PALSAR-2 data, with description of the pre-processing steps
4. TanDEM-X data
5. Spaceborne LiDAR data

3.2 Sentinel-2

3.2.1 Sentinel-2 imagery

The Sentinel-2 mission is designed to provide global acquisitions of fine high-resolution, multispectral optical imagery in fine temporal resolution. It has three satellites in orbit: S2A launched on 23 Jun 2015, S2B on 7 Mar 2017 and S2C on 5 Sep 2024. The satellites have a wide (290 km) imaging swath width and 10 days revisit time at the equator. With two satellites in orbit (the target number to be maintained), this enables five days imaging frequency at the equator and 2-3 days imaging frequency at mid-latitudes. With coverage limits between 56° south and 84° north latitudes, the data cover all forested areas of the world. The Multi-Spectral Instrument (MSI) on board Sentinel-2 satellites has 13 spectral bands, four of which have 10 m and six of which have 20 m spatial resolution. The remaining three bands with 60 m spatial resolution are mainly used for atmospheric correction.

Sentinel-2 is the main optical satellite data source for the FCM tools. Due to its open and free data policy and global coverage, it is an optimal choice for operational forest monitoring. Furthermore, it provides sufficient spatial resolution to meet most user requirements and a suitable selection of wavelengths for forest variable prediction. Its high imaging frequency improves the probability of obtaining cloud free observations.

The Level 2A surface reflectance product is systematically generated by ESA and distributed in tiles of 110 x 110 km^2. This has been the main Sentinel-2 data product used in the FCM project. Seven spectral bands have been typically used (Table 2), based on earlier findings on the importance of bands for forest monitoring (Astola et al. 2019, Miettinen et al. 2021).

Table 2. Sentinel-2 spectral bands typically used in FCM tool demonstrations.

Sentinel-2 band	B02	B03	B04	B05	B08	B11	B12
Wavelength	Blue 0.49 µm	Green 0.56 µm	Red 0.67 µm	Red Edge 1 0.71 µm	NIR 0.84 µm	SWIR 1.61 µm	SWIR 2.19 µm
Original spatial resolution	10 m	10 m	10m	20 m	10 m	20 m	20 m

3.2.2 Sentinel-2 multi-temporal compositing

In case where suitable single date imagery is not available, the Sentinel-2 compositing can be conducted with a tool developed by Terramonitor. The objective of the compositing process is to create a cloud-free image from many observations. To this end, each pixel is evaluated according to four criteria: cloudiness, resemblance to usual pixels observed in the location (based on a reference mosaic), haze and shadows. A weight is then given for each pixel according to the four criteria. These weights are used to average the observations given as input and produce the final image. The weighted average merging algorithm is defined mathematically as follows. Let $X=(x_1, x_2,\ldots,x_t)$ denote a time series of observations for a given geographical point, where each element x_i denotes an observation from the Sentinel-2 satellite. Each observation $x=(x(0), x(1),\ldots,x(12))$ consists of the band values $x(0), x(1),\ldots,x(12)$, where $x(0)$ is the value of the scene classification band provided with the Sentinel-2 Level 2A product and $x(1),\ldots,x(12)$ correspond to the values of the Sentinel-2 bands 2, 3, 4, 5, 8, 11 and 12, respectively.

The weight for an observation x is given by the formula $w(x)=m_c(x)\, m_d(x)\, m_h(x)\, m_s(x)$, where $m_c(x), m_d(x), m_h(x)$ and $m_s(x)$ represent multiplier functions that are based on scene classification, spectral distance, haze and shadows, respectively. Each multiplier function produces a value between 0 and 1 that describes the validity of the observation with respect to one of the criteria. For example, if the multiplier function m_h gives the value 1, it means that the observation is assumed to be totally valid with respect to haze, i.e., haze-free. If the weight of an observation is close to one, it means that the validity of the observation is large with respect to all of the criteria.

The scene classification multiplier is defined by the formula:

$$m_c(x) = \begin{cases} 1 \; if \; x_0 \in V \\ 0 \; otherwise \end{cases}, \qquad (3.2.2.1)$$

where $V = \{2,4,5,6,7\}$ denotes the set of valid classes for the scene classification band (2: dark area pixels, 4: vegetation, 5: not vegetated, 6: water, 7: unclassified).

The spectral distance multiplier is used to evaluate the resemblance of the pixel to cloud-free pixels observed in the location, and is defined using the formula:

$$m_d(x) = \left(\min\left(\max\left(1 - \frac{d(x,L)}{d_{max}}, 0\right), 1\right)\right)^{p_d}, \qquad (3.2.2.2)$$

where $L=\{l_1,\ldots,l_n\}$ is a collection of n cloud-free reference observations and $d(x,L)$ denotes the minimum spectral distance between observation x and the observations in L, that is, $d(x,L)=min\{d_e(x,l) \mid l \in L\}$, where d_e represents the Euclidean distance function. d_{max} and p_d are constants whose values are set to 3000 and 6, respectively. The values were set through visual evaluation of preliminary test results for one tile in Finland (35VLJ) and earlier experiences with the aim of maximizing the number of observations without including any noticeable haze in the final product.

The haze multiplier is defined using the formula:

$$m_{\mathrm{h}}(x) = \left(\min\left(\max\left(1 - \frac{x_1}{h_{\max}}, 0\right), 1\right)\right)^{p_{\mathrm{h}}}, \quad (3.2.2.3)$$

where x_1 represents the value of the Sentinel-2 band 1, $h_{\max} = 3000$ and $p_{\mathrm{h}} = 6$.

The shadow multiplier is defined by the formula:

$$m_{\mathrm{s}}(x) = \begin{cases} \min\left(max\left(1 - \frac{\mathrm{x_s} - \mathrm{c_0}}{\mathrm{c_1} - \mathrm{c_0}}, 0\right), 1\right) \; if \; x_{\mathrm{s}} \leq c_1, \\ \min\left(max\left(1 - \frac{x_s - c_1}{c_2 - c_1}, 0\right), 1\right) \; otherwise \end{cases}, \quad (3.2.2.4)$$

where x_{s} is the value of the Sentinel-2 band 8 (near infrared), $c_0 = 100$, $c_1 = 250$ and $c_2 = 2000$.

Finally, given the time series of observations $X=(x_1,x_2,\ldots,x_t)$ and the corresponding weights $w(x_1),w(x_2),\ldots,w(x_t)$, the weighted average a_X of the observations is given by the formula:

$$a_X = \frac{\sum_{i=1}^{t} x_i w(x_i)}{\sum_{i=1}^{t} w(x_i)} \quad (3.2.2.5)$$

Seven spectral bands are output into the resulting composite images (Table 1). In addition to the seven spectral bands, a quality parameter is calculated. The quality band value described the probability of at least one good observation, which is calculated per pixel using the formula $P = 1 - \prod(1-p_i)$, where p_i denotes the probability that observation i was good for $i \in \{1,\ldots,n\}$, where n denotes the number of observations for the pixel. For the final composite images, all bands are resampled to match the 10 m resolution bands using nearest neighbour resampling.

3.3 Sentinel-1

3.3.1 Sentinel-1 data

Sentinel-1 (S1) is a spaceborne mission operated by the European Commission in the Copernicus framework and consists, as of year 2025, of three identical satellites (1A, 1B, and 1C), each operating a C-band SAR. Sentinel-1A was launched in 2014 and began routine observations in 2015. Sentinel-1B was launched in 2016 and became operational at the beginning of 2017. Operation of Sentinel-1B ended in 2021 because of a hardware failure. Sentinel-1C was launched in 2025. Each satellite has a 12-day repeat-pass interval. When combined, two satellites provide for a six-day repeat coverage and even more frequent observations when considering the overlap between adjacent orbits, in particular at high latitudes. Over land, the Interferometric Wide Swath (IWS) mode is the primary acquisition mode, which allows for single- or dual-polarization acquisitions (VV or VV/VH over most of the Earths' land area, HH or HH/HV over the Arctic and Antarctic) with a spatial resolution of approximately 20 m in range and 5 m in azimuth. Being a Copernicus mission, the greatest priority is given to acquisitions over Europe, where each satellite acquires continuously along both ascending and descending orbital tracks (Figure 4). Combined, Sentinel-1A and 1B acquired a total of about 60 000 scenes per year with 60 observations from ascending and descending orbital tracks each (relative orbit in ESA

terminology) over the European demonstration area. Sentinel-1A and 1C will provide for a similar data amount in the coming years.

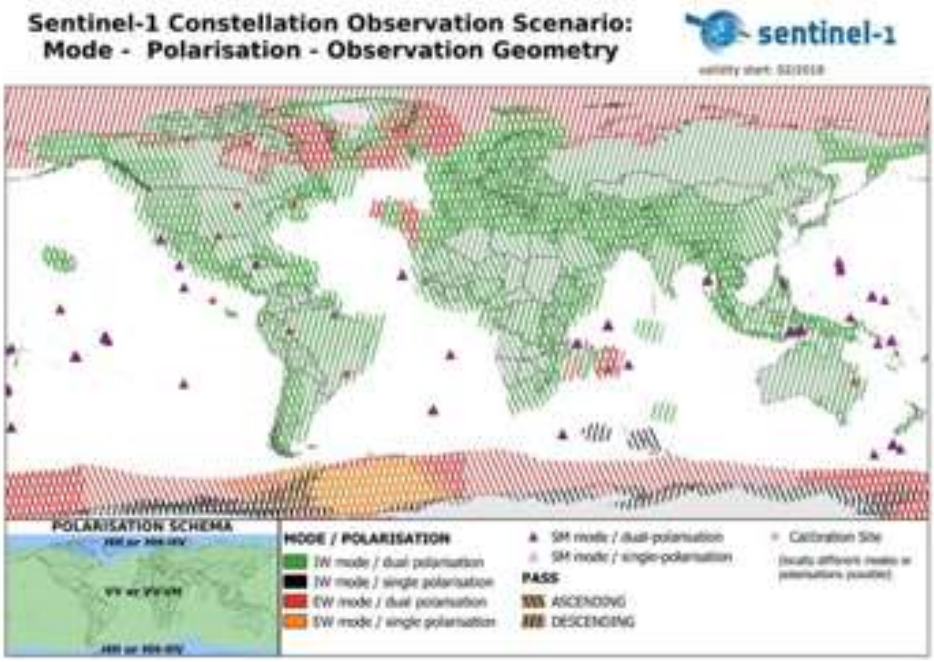

Figure 4. Observation geometry of the Sentinel-1 mission[1].

3.3.2 Sentinel-1 pre-processing

The Sentinel-1 C-band backscatter intensity was identified as a core observable to support the prediction of forest growing stock volume and above-ground biomass variables following the approaches developed in the frame of CCI Biomass. The pre-processing is applied to Sentinel-1 images provided in Ground Range Detected (GRD) format. GRD images consist of ground-range projected images of the SAR backscatter intensity. Scope of pre-processing that is carried out in the FCM project is to generate a stack of terrain geocoded, radiometrically calibrated, speckle-filtered and co-registered Sentinel-1 observations. Pre-processing with the commercial software package by GAMMA Remote Sensing comprises:

1) 2 x 2 multi-looking in range and azimuth to obtain pixels with 20 x 20 m^2 ground pixel posting,
2) compensation for the noise equivalent sigma nought (NESZ),
3) updating of orbit state vectors with precision orbit vectors provided by ESA within 20 days past the image acquisition[2]
4) topographic correction accounting for varying pixel scattering areas dependent on topography as with Frey et al. (2013) to produce “terrain-flattened” g^0 backscatter intensity images,
5) geocoding and orthorectification based on the Copernicus 1-arcsecond Digital Elevation Model (DEM) to the target UTM map grid with 20 x 20 m^2 pixel size.

[1] https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario
[2] https://qc.sentinel1.eo.esa.int/aux_poeorb/

All geocoded images are resampled to the same MGRS/UTM tiling grid to which ESA processes Sentinel-2 data to allow for a joint use/inter-comparison of Sentinel-1 and Sentinel-2 imagery.

Given the large level of correlation among biomass maps generated from Sentinel-1 observations acquired with 6-day repeat intervals and, hence, the limited benefit of considering all available observations, only images acquired in dual-polarization (VV/VH) IWS mode by one of the two satellites, i.e., S1A, were considered in the FCM project. This reduced the amount of data to be processed to 25 to 30 000 GRDs for each of the four years for which forest GSV and AGB maps were produced. Only in a few areas with reduced data availability, e.g., Southern Finland, data from both satellites had to be considered.

For the development and validation of the algorithm to be applied for the pan-European mapping, data from testing sites in Finland Romania, and Catalonia were used. Sentinel-1 GRD images were selected and pre-processed in accordance with the processing workflow discussed above. For each testing site, all images acquired by S1A from one descending and one ascending relative orbit in two years have been considered (Table 3). Only in the case of the testing area in Southern Finland (Finland S) was data from S1B added to the stack to obtain a consistent time series of observations from both, ascending and descending, orbits (Figure 5). For each testing site the time frame covered by the selected Sentinel-1 data was chosen in accordance with the collection of the *in situ* information available for each site.

Table 3. Sentinel GRD imagery processed for the six test areas in Europe.

Site	Satellite	Relative orbit	Years	Number of images
Finland N	S1A	80/116	2018/2019	118
Finland S	S1A/S1B	87/153	2018/2019	89
Romania	S1A	109/131	2019/2020	240
Catalonia	S1A	37/132	2015/2016	143

3.3.3 Sentinel-1 compositing

The forest structural variable retrieval approaches described in Sections 5.2-5.4 benefit from using multi-temporal composites of the Sentinel-1 backscatter images. A Forestry TEP tool was therefore created, which calculates the average backscatter at VV or VH polarization for each Sentinel-1 orbital track. For an annual composite, for example, given the 12-day repeat cycle of Sentinel-1A, the Forestry TEP tool computes the average backscatter across 30 to 31 images acquired in a given year. Each multitemporal composite is accompanied by a single map of the local incidence angle and layover/shadow mask.

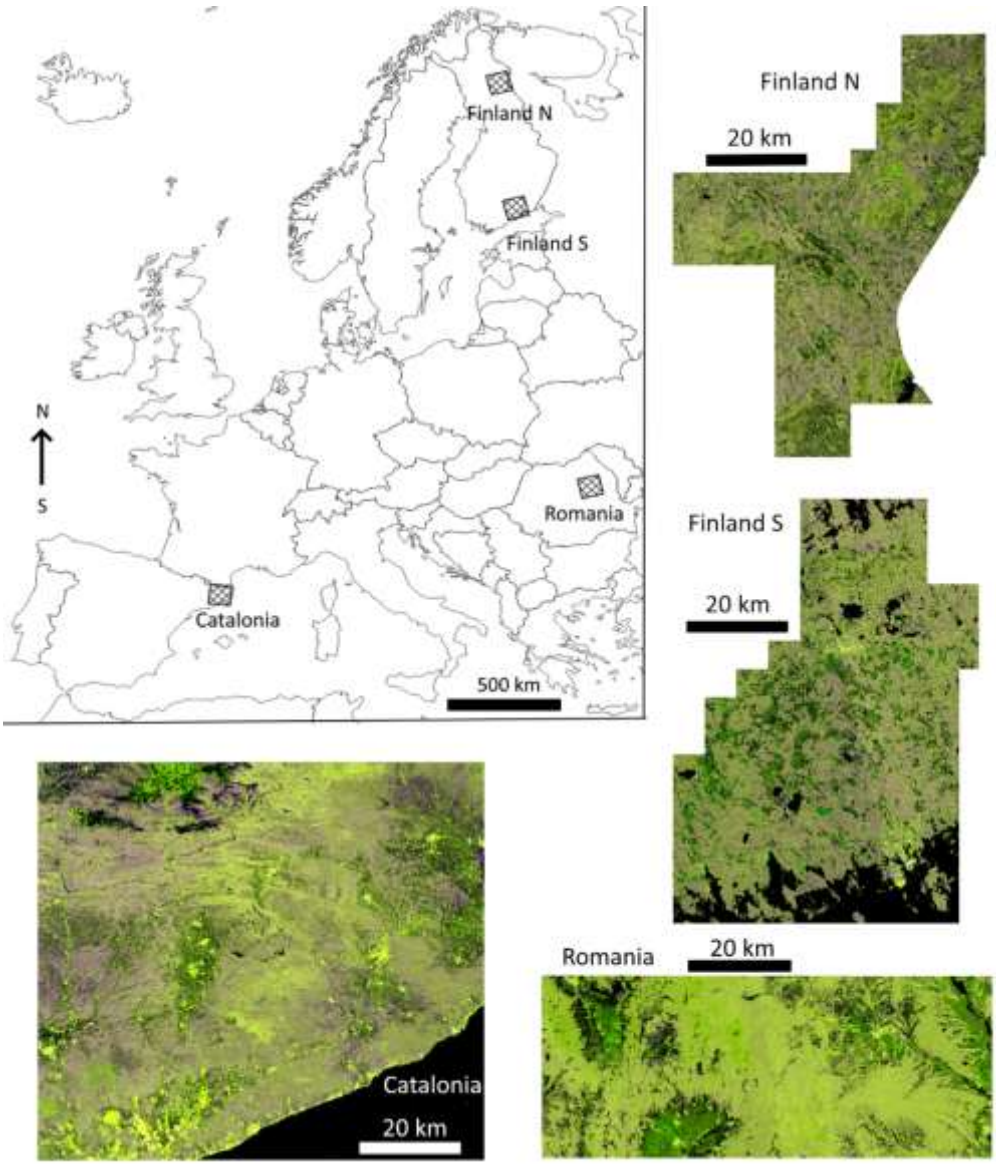

Figure 5. Location of the study sites. Each site is illustrated with a colour composite of Sentinel-1 imagery (Red: VV-polarized backscatter; Green: VH-polarized backscatter; Blue: difference in the VV- and VH-polarized backscatter) (Santoro et al., 2024b).

3.4 ALOS-2 PALSAR-2 mosaics

The ALOS-2 mission started on May 24, 2014, and carries an L-band SAR (PALSAR-2 instrument) with slightly improved performance than its predecessor, ALOS-1 PALSAR-1. ALOS-2 PALSAR-2 operates in several high-resolution (e.g., Fine Beam, FB) and a moderate resolution ScanSAR mode (WB) with resolutions of the order of 25 and 50 m, respectively. Each year global and repeated acquisitions are scheduled during seasons that are known to maximize the information content of the backscattered signal with respect to land surface properties. In both FB and WB mode, PALSAR-2 acquires data in single polarization (HH) and dual polarization (HH and HV, VV and VH over Japan), covering swaths of approximately 70 km and 250 km, respectively. While the acquisition plan foresees at least one global coverage per year at fine, i.e., 25 m, resolution in FB dual-polarization mode, multiple acquisitions may be available per year from WB mode, albeit only in the tropics.

L-band backscatter was identified as a crucial observable to complement Sentinel-1 backscatter time series for improving the retrieval of GSV and AGB particularly in high biomass forests. Because of the data policy applied by JAXA to ALOS-1 and ALOS-2 data, only a limited number of images can generally be obtained free of charge, which hinders large scale application. Large scale coverages of ALOS-2 PALSAR-2 data could only be obtained so far in the form of yearly backscatter mosaics (2015-2021) for the FB mode (Shimada and Ohtaki, 2010; Shimada et al., 2014) and per-cycle mosaics (46 days) for the WB mode. While the FB mosaics are publicly available, the ScanSAR mosaics are available only to a restricted research community (i.e., the Kyoto and Carbon (K&C) Initiative). The FB mode mosaics generally present almost complete global coverage (subset of the year 2020 mosaic for Europe is shown in Figure 6). The number of available observations from the mosaics is one at each location, therefore limiting the performance of the GSV and AGB retrieval.

Figure 6. ALOS-2 PALSAR-2 HV-pol mosaic produced by JAXA from FB data acquired in 2020.

For the second phase of the FCM project, JAXA granted exclusive access to all ALOS-2 FBD images acquired over Europe in 2017, 2020, 2021, and 2023. The ALOS-2 data were provided in the form of ca. 300 km long strips in range-doppler geometry and processed to radiometric terrain-corrected level by the FCM project team. Pre-processing was done with the commercial software package by GAMMA Remote Sensing and comprised:

1. multi-looking to obtain pixels with ca. 20 x 20 m^2 ground pixel posting,
2. topographic correction accounting for varying pixel scattering areas dependent on topography as with Frey et al. (2013) to produce “terrain-flattened” γ^0 backscatter intensity images,
3. geocoding and orthorectification based on the Copernicus 1-arcsecond Digital Elevation Model (DEM) to the target UTM map grid with 20 x 20 m^2 pixel size.

Figure 7 illustrates the number of images available for Europe for the four selected years. The best coverage was generally available for Northern Europe. In Southern and Central Europe, the number of available observations varied between two and four per year.

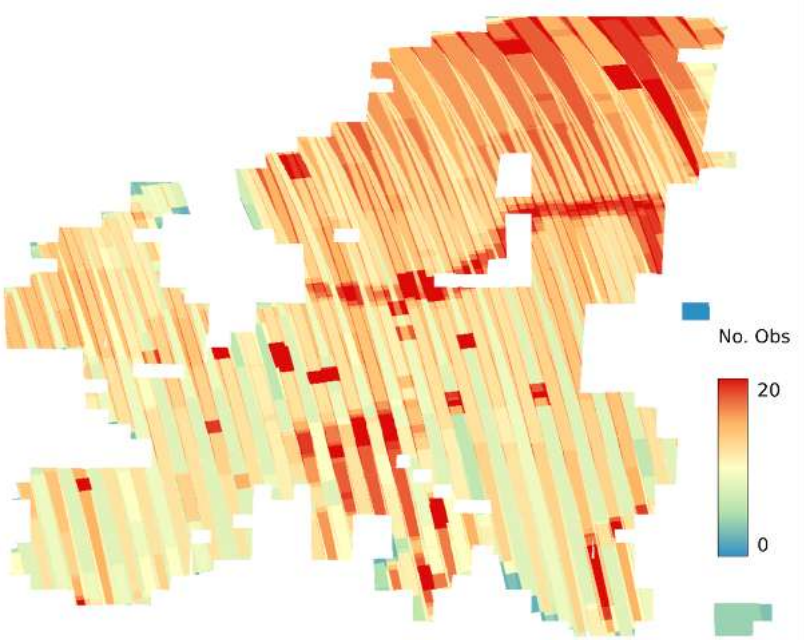

Figure 7. Coverage of ALOS-2 PALSAR-2 Fine-Beam data provided by JAXA for the year 2017, 2020, 2021, 2023.

3.5 TanDEM-X images

The German TanDEM-X mission flies the two satellites TerraSAR-X and TanDEM-X in a close orbit formation establishing a bistatic interferometer in space. The primary mission goal was generation of a global DEM (Krieger et al., 2007). The SAR data, jointly acquired by both satellites, are operationally processed by the Integrated TanDEM-X Processor. Processing comprises bistatic synchronization and focusing, filtering, co-registration, phase unwrapping and geocoding (Breit et al., 2012). Outputs are an individual scene-based (50 km × 30 km) DEM and a co-registered phase preserving single look slant range complex SAR images (CoSSC).

TanDEM-X data were previously demonstrated as a useful source of information for predicting the vertical structure of forests. While normally multi-polarizations measurements – polarimetric interferometric SAR (Pol-InSAR) signatures are most useful for forest structure assessment, also single-pol InSAR measurements in presence of external ground DEM (Praks et al., 2012, Krieger et al., 2014), and to considerable extent the magnitude of InSAR coherence (Olesk et al., 2016) can be used to estimate relationships between TanDEM-X observables and forest variables.

In FCM, primarily single-pol (HH) CoSSC images acquired during one close to baseline year were used in forest variable prediction and producing forest attribute maps. Tentative coverage over the Catalonia test site is shown in Figure 8.

Figure 8. TanDEM-X data coverage over Catalonia.

3.6 Spaceborne LiDAR data

LiDAR observations are closely related to vegetation structural features, thus being well-suited for direct prediction of forest variables related to the biomass. The density of global spaceborne LiDAR observations from the ICESat (2003-2009), ICESat-2 (2018-ongoing) and GEDI (2019-ongoing) missions is, however, still too coarse to allow for wall-to-wall prediction of forest variables. Spaceborne LiDAR observations are, therefore, considered here in the process of calibrating SAR models rather than used as predictors of biomass.

Between 2003 and 2009 the Ice Cloud and Elevation Satellite (ICESat) Geoscience Laser Altimeter System (GLAS) instrument collected information related to the vertical structure of forests in ca. 65 m large footprints collected every 170 m along track. The distance between tracks was of the order of 10s of km and increased towards the equator. The GLA14 product (version 34), which provides altimetry data for land surfaces only to which geodetic, was used to estimate canopy density (*CD*) calculated as the ratio of energy received from the canopy (returns above the ground peak) to the total energy received and the height (*h*) as the distance between the ground peak and signal beginning (*RH100*). Forest height was computed following the approaches in Simard et al. (2011) and Los et al. (2012), which calculated RH100 globally and defined a set of filters to discard footprints affected by topography and various noise sources in the waveforms (Santoro et al., 2021a).

Unlike the GLAS sensor, the Advanced Topographic Laser Altimeter System (ATLAS) onboard the ICESat-2 satellite, uses photon counting to retrieve elevation. With a frequency of 10,000 pulses per second, ATLAS achieves a much denser portrait of the surface compared to the 40 pulses used by GLAS. The measurement technique is, however, strongly affected by the power recorded by the instrument. ATLAS splits the laser into three

pairs of beams approximately 3.3 km apart. Each pair consists of a strong and weak energy beam (4:1 ratio). For vegetation studies, it is advised to flag measurements corresponding to weak beams because of the partly undetected vegetation layering in the returned signals. The ATL08 product (Neuenschwander and Pitts, 2019) contains geophysical parameters related to vegetation and terrain heights (in particular, the top-of-canopy height) but no metric of canopy density. The parameters are provided with a 100 m step size along the flight direction. Currently version 6 of the product is available from the National Snow and Ice Data Center (NSIDC) in the form of strips of photons collected along one orbit. To obtain segments from the original photon data, the original files are reformatted with the pysl4land Tool, a set of Python tools to process spaceborne LiDAR (GEDI and ICESat2) for land (pySL4Land) applications[3]. Herewith, the original photons are grouped into segments of 100 m length and 25 m width. Variables related to canopy height and corresponding quality flags are extracted.

Like GLAS, the Global Ecosystem Dynamics Investigation (GEDI) instrument (Dubayah et al., 2020) is a full waveform LiDAR. GEDI is installed on the International Space Station (ISS) and, therefore, obtains data for land masses between +/-52° latitude. The size of the footprint is smaller than for ICESat GLAS (25 m vs. 70 m diameter) and the density of observations is greater. GEDI acquires data for 8 parallel tracks, separated by about 600 m across track. Along each track, footprint centres are separated by 60 m. The distance between adjacent orbital tracks was about 1 km until January 2020 after which it increased to 70 km. From the waveform data, several height metrics, including canopy height (defined as H100) and canopy density are obtained. These level 2A (height metrics) and 2B (canopy density) data are provided at the level of individual footprints. Version 2 is currently distributed. Data from individual orbital files are reformatted with the pysl4land Tool and relevant variables related to canopy density and height are calculated.

[3] https://github.com/remotesensinginfo/pysl4land

4. Statistical approaches

4.1 Overview of statistical approaches

One of the overarching ideas of the FCM concept is that all output products are delivered with information on the uncertainty of the products. Typically, all EO based forest variable maps produced with the FCM tools are accompanied with two different types of uncertainty information. Firstly, error metrics (see section 4.2.) are calculated with reference field plots (whenever available). These metrics provide information on the overall level of uncertainty of the output products. Secondly, pixel-wise uncertainty layers are provided with most output products. The methods to calculate the pixel-wise uncertainty layers vary depending on the predictor model (see section 4.3) but all of the layers provide the standard deviation of the predictions. These layers enable users to analyse the expected level of uncertainty on pixel level and the spatial variation of the pixel level error within the area of interest.

In addition to the provision of plot and pixel level uncertainty information, two statistical frameworks to use EO-based forest maps in operational set-up have been demonstrated during the FCM project. These include the model-assisted estimation and two-step sampling approaches. The model-assisted estimation combines remote sensing and field data, improving the accuracy and enabling estimation for smaller geographic area than would be possible using only field plots. The two-step sampling approach, on the other hand, is a generic approach that can be implemented in highly varying purposes to efficiently utilize available datasets (e.g. including wall-to-wall maps, very high resolution imagery and field samples). In the FCM project the approach was demonstrated in the Colombian use case.

In this chapter, the statistical approaches used to provide the uncertainty information and to support the use of the outputs products in operational setting are described.

4.2 Output product error metrics

Whenever reference field plots or other suitable source of reference data are available, error metrics are provided with the products produced using the FCM tools. Standard set of error metrics provided with the products produced with the Probability, k-NN and UNet tools include the following:

1. *Root Mean Squared Error (RMSE)*, which quantifies the difference between predicted values and reference values. Lower RMSE values indicate more precise predictions. Conversely, high values indicate more error. Note that the RMSE includes also the prediction bias (see below). The RMSE is calculated as

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}} \qquad (4.2.1)$$

   where y_i represents the reference values, $\hat{y}_i$ represents the predicted values, i=1…..*n* indexes the observations, and *n* is the number of reference observations.

2. *Prediction bias* (Bias), which provides the difference between the mean of the predictions and the mean of the reference observations. This is an important error metric for forest monitoring as it tells about the usability of the predictor for large

area monitoring. Pixel level errors typically balance each other out for larger areas, but the bias reveals the level of systematic error in the predictions. The bias is calculated as

$$Bias = \frac{\sum_i (y_i - \hat{y}_i)}{n} \quad (4.2.2)$$

where y_i represents the reference values, $\hat{y}_i$ represents the predicted values, i=1…..*n* indexes the observations, and *n* is the number of reference observations.

3. *Coefficient of determination (R^2)*, which quantifies the proportion of the variation in the target variable that is predictable from the independent variables (used in the prediction). High R^2 values indicates a good fit of the predictor model. The R^2 is calculated as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (4.2.3)$$

where y_i represents the reference values, $\hat{y}_i$ represents the predicted values, i=1…..*n* indexes the observations, and *n* is the number of observations in the validation database.

The RMSE and bias values are also provided as values relative to the mean. The absolute value of the metric is compared to the mean value of the variable in the reference plots, thereby deriving relative metrics, denoted as RMSE% and Bias%. The relative values allow easier comparison of the error metrics between different areas of interest and between different variables.

Typically, the error metrics are calculated using an independent set of reference plots, and a validation set that has been extracted from the reference data before model training. The FCM tools also provide a crossvalidation tool to calculate error metrics for the k-NN predictions. This allows all available plots to be used for the mapmaking. This may have significant effect on the accuracy, particularly in areas where the number of available field plots is already limited. Using the crossvalidation tool also ensures that the split of the reference data does not affect the error metrics.

4.3 Standard deviation layers

4.3.1 k-NN

In addition to the prediction layer, the FCM k-NN tool outputs also a standard deviation layer for each target variable. These standard deviation layers have been calculated as the standard deviation of the k neighbours used to derive the prediction.

The standard deviation $\hat{s}(p)$ is calculated as

$$\hat{s}(p) = \sqrt{\frac{\sum_i (y_i(p) - \hat{y}(p))^2}{k}} \quad (4.3.1.1)$$

where $\hat{y}(p)$ is the prediction for pixel p and $y_i(p)$ are the values of the k neighbours used to calculate the prediction.

The standard deviation layers provide users with information on the level of uncertainty of the output products on pixel level. They also enable analysis of the spatial variation of the pixel level errors within the area of interest. It is important to note, however, that the pixel level errors are expected to balance out when cumulated over larger areas. Therefore, the prediction bias (see section 4.2) provides more useful information on the expected level of error for larger interest areas.

4.3.2 UNet

For UNet-model tool, epistemic uncertainty of UNet maps on a pixel-level is calculated as a model ensemble standard deviation. Here, the idea is that variability in predictions from multiple models trained on different data splits (folds) quantifies uncertainty stemming from data variability and model generalization.

The approach proceeds as follows. Given an overall training dataset D, we divide it into m folds/subsets $\{D_1, D_2, \dots, D_m\}$ and train m UNet models with the same architecture in such a way that n-th UNet model is trained on the whole dataset D leaving out D_n, with the process repeated for all $n = 1, \dots, m$ folds. Then during inference, for each mapping unit (pixel) p of the mapping area, an ensemble mean prediction and ensemble standard deviation are computed:

$$\bar{y}(p) = \frac{1}{m}\sum_{i=1}^{m} \hat{y}_i(p) \qquad (4.3.2.1)$$

$$\bar{s}(p) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i(p) - \bar{y}(p)\right)^2} \qquad (4.3.2.2)$$

Ensemble mean is reported as final UNet prediction map, and ensemble standard deviation reports the model-related uncertainty. As it considers model-related variance, aleatoric uncertainty is not handled and produced estimate can be considered optimistic compared to other uncertainty estimates. This uncertainty measure however does not require any independent ground truthing and can be reported also for maps produced by “blindly” applying the model.

When independent set-aside ground-truth data are available, conventional map-level accuracy metrics can be additionally computed, such as RMSE and systematic deviation (bias). Then, ensemble-derived epistemic uncertainty $s_{ep} = \bar{s}(p)$ and validation data derived bias $bias_{val}$ can be integrated into overall uncertainty reported on a pixel level:

$$s_{total} = \sqrt{s_{ep}^2 + bias_{val}^2} \qquad (4.3.2.3)$$

4.3.3 BIOMASAR

The uncertainty of a GSV prediction from a measurement of the backscatter is obtained by perturbing the measurement and the prediction model parameters (σ^0_{gr}, σ^0_{veg}, q, a and b) with individual uncertainties. The uncertainty of the prediction is then defined as the standard deviation of the vector of perturbed GSVs obtained by repeating the perturbation N times. The variance of the GSV prediction from Eq. (5.5.1.6) is then the sum of a variance component and a covariance component that accounts for the temporal correlation of prediction errors at a given pixel.

$$\delta(V_{mt})^2 = \sum_{i=1}^{N} w_i^2\, \delta(V_i)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i\, w_j\, Cov(V_i, V_j) \qquad (4.3.3.1)$$

where

$$Cov(V_i, V_j) = \delta V_i \delta V_j r_{ij} \qquad (4.3.3.2)$$

In Eq. (4.3.3.1), the variance component is modelled as a linear combination of the variances associated with the individual stem volume estimates $\delta(V_i)^2$, where w_i^2 is the weight introduced in Eq. (5.5.1.5) (Santoro et al., 2015). The covariance component is expressed in a similar manner where individual error co-variances are weighted. The error covariance in Eq. 4.3.3.2 is obtained from the pairwise standard deviation of GSV estimates from image *i* and image *j* and the corresponding correlation of errors, r_{ij}. To estimate the correlation of errors, a reference dataset is needed such as extensive plot inventory measurements or a LiDAR map of GSV with known accuracy

The variance of the stem biomass prediction is obtained with Eq. (4.3.3.3) and accounts for the variance of the stem volume prediction from Eq. (4.3.3.1) and of the wood density. The variance of the wood density, $\delta(WD)^2$, is modelled using the second order polynomials developed for Eurasian boreal forests (Thurner et al., 2014)

$$\delta(\delta B)^2 = (WD)^2 \cdot \delta(V_{mt})^2 + (V_{mt})^2 \cdot \delta(WD)^2 \qquad (4.3.3.3)$$

The variance of the total biomass density is then obtained by adding the variances of the individual biomass component. The variance of the branch, foliage and root biomass is modelled as in Thurner et al. (2014)

$$\delta(TB)^2 = \delta(SB)^2 + \delta(BB)^2 + \delta(FB)^2 + \delta(RB)^2 \qquad (4.3.3.4)$$

4.4 Model-assisted estimation

The aim of using model-assisted approaches is to improve forest variable estimates, typically over some larger area, that are i) only based on a probability sample, so-called Direct estimates, and ii) only based on remote sensing products, so-called Pixel-counting or Synthetic estimates. The model-assisted estimation is a general approach that can be applied to support integration of EO based products (such as the ones produced with the FCM tools) into existing forest monitoring schemes based on probability sampling. The details of the application vary case-by-case, but the general concept remains the same. As

an example of the procedure and algorithms, we describe here the way model-assisted estimation was applied in the Norway use case during the FCM project.

In the context of this use case, the probability sample consists of National Forest Inventory (NFI) field sample plots. It can, however, also be any other type of reliable reference measurements for the variable of interest, for example taken from aerial images or drone acquisitions, as long as they satisfy the requirements of a probability sample. In the context of the FCM project, the remote sensing products are maps of forest attributes including volume and biomass. But again, the EO based products can be any kind of variable of interest mapped wall-to-wall by remote sensing or other approaches. For simplicity, we will describe the methodology in the remainder of the text only for the context of this project with timber volume as the variable of interest, and Norwegian NFI data as the probability sample. The aim is to improve the mean volume estimate (over all land uses) for the productive low-land stratum of the NFI in Norway south of Nordland County as the area of interest.

The improvement of model-assisted approaches is achieved by combining the direct NFI-based estimate of the mean volume with the remote sensing-based pixel-counting estimate. The following steps are needed for a model-assisted estimate:

1) Determine the difference between volume (m^3/ha) observed at each sample plot and the mapped (predicted) volume (m^3/ha) at the same location. The difference is calculated as observed minus predicted. Due to this difference, the applied model-assisted estimator in our case is referred to as the Difference estimator.
2) Calculate the mean of the differences. This mean is a correction factor for the systematic error or bias in the remote-sensing based pixel-counting estimator. If, for example, the map more commonly predicts lower values than observed, then the correction factor will be positive and indicates a systematic underestimation by the map.
3) Determine the pixel-counting estimate by calculating the mean over all wall-to-wall map pixels covering the area of interest.
4) Obtain the model-assisted difference estimate by adding the pixel-counting estimate and the correction factor.
5) Calculate the variance and standard error of the differences. This is the design-based standard error of the correction factor and the difference estimator itself.

For determining the improvement of using the map in addition to the NFI field data, the direct estimate, including the variance and standard error is calculated. If the map is even slightly correlated with the field data, then the variance of the differences is smaller than the variance of the observations and the difference estimator is more precise than the direct estimator. If the map is of low quality, maybe because it exhibits artifacts that result in outliers, also the opposite may be the case. An intuitive measure of the improvement is the Relative Efficiency (RE), which is the ratio of the variance of the direct estimator (nominator) and the variance of the difference estimator (denominator). If RE is larger than 1, the difference estimator is more efficient than the direct estimator. The RE can be seen as a factor by which the number of field samples would need to be multiplied to achieve the same accuracy with the direct estimator as with the difference estimator (assuming that these additional observations would not be used in the difference estimator).

The difference estimator provides improvements in two ways: 1) It results in higher precision of the direct NFI-based estimate (if the map has ok quality), and 2) it provides a reliable uncertainty estimate for the remote-sensing based map. There are, however, also a few requirements which include that the reference sample and map data 1) are assumed to be fully independent, 2) have temporal agreement, and 3) agree in resolution. In practise, these requirements are seldomly fully met.

Using equations, the steps for a model-assisted estimate are:

The difference d is given by

$$d_i = y_i - \hat{y}_i \quad (4.4.1)$$

with $i \ldots n$ indexing the sample plots and n = number of observations, y are observed values and $\hat{y}$ are mapped values extracted or otherwise obtained for sample plots.

The correction factor C is given by

$$\hat{C} = \sum_i d_i / n \quad (4.4.2)$$

The pixel counting estimate is given by the mean of all pixels in the AOI

$$\mu = \sum_i \hat{y}_i / N \quad (4.4.3)$$

where N is the total number of pixels.

The difference estimator is given by

$$\widehat{Y^{Diff}} = \mu + \hat{C} \quad (4.4.4)$$

The variance of $\widehat{Y^{Diff}}$ and $\hat{C}$ is given by

$$V\left(\widehat{Y^{Diff}}\right) = V\left(\hat{C}\right) = 1/n \sum_i d_i^2 / (n-1) \quad (4.4.5)$$

with the standard error

$$SE = \sqrt{V(\cdot)} \quad (4.4.6)$$

In addition, the direct estimate is given as

$$\widehat{Y^{Dir}} = \sum_i y_i / n \quad (4.4.7)$$

with variance

$$V\left(\widehat{Y^{Dir}}\right) = 1/n \sum_i (y_i - y_{mean})^2 / (n-1) \quad (4.4.8)$$

and

$$RE = V\left(\widehat{Y^{Dir}}\right) / V\left(\widehat{Y^{Diff}}\right) \quad (4.4.9)$$

4.5 Two-step sampling approach

Sometimes there are multiple different types of datasets available for a given interest area, including e.g. LiDAR or very high resolution remote sensing data, medium resolution satellite data and a possibility for field plot measurements. In these kinds of situations, statistical approaches can be defined to enable efficient use of the datasets and derivation of rigorous uncertainty information. As an example of a statical framework, we present a 'two-step' sampling approach created for the Colombian use case demonstration. The phrase 'two-step' refers to two separate samplings, one being a field measurement campaign and the other being a sampling for visual interpretation on Planet data. These two steps are then combined in the estimates of biomass.

Here we present the main concepts and considerations that were taken into account when defining the framework. The aim is that this description serves as a guideline for future users defining similar frameworks for their use cases. Note that in our example case the field measurement campaign was conducted first. The order of the samplings could have been also different. This would affect also the optimal statistical procedures.

4.5.1. Design and analysis of field campaign sampling

Field reference data is crucial for accurately calibration of estimates derived from EO data. Over the study area in Colombia, several land cover (LC) classes were identified with EO data (Sentinel-2, PALSAR-2), presumably containing varying amounts of biomass. For the sake of this example, these LC classes are called 'Primary 1', 'Primary 2', 'Regrowth 1' and 'Inundated'. Accurate biomass estimates were needed for these four LC classes for the study area; the other land cover classes (presumed to have no or very little biomass) could be given biomass estimates based on them.

At the top level, the sampling design of the field campaign was elementary: in the four LC classes we made simple random sampling (SRS) designs, independently of each other. In the practical level, there were several important details that needed to be considered. The locations had to be accessible in practise (within 6 km from rivers) and there had to be spare locations in case the field crew found some locations inaccessible during the field campaign. The total sample size was limited due to time and resources limitations. When divided into four different LC classes, this meant small subsamples and large uncertainties. Furthermore, it was possible that the true LC class of a selected location was different from the LC class of the map that was used to select the location. These issues meant that there was an inevitable danger of selection bias in the process that finally produced the data.

The field measurement campaign was carefully documented, and we were able to assure that the sampling design was followed without any such compromises that were not considered beforehand. This is important for the statistical validity of the results. In this example case, there were finally 46 randomly selected locations with known biomass and with known inclusion probabilities. Therefore, we were confident that the effect of the selection bias is small.

After the field campaign data was processed, some exploratory data analyses were performed to see possible outliers and whether the sampling distribution was normal. Box-Whisker plots (Figure 9) are useful for obtaining an overview of data.

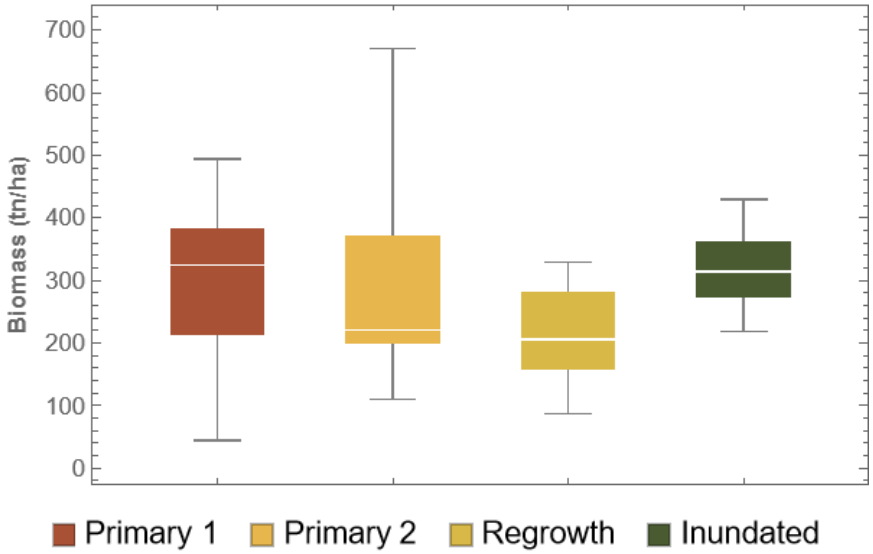

Figure 9. Box-Whisker plots of Colombia use case field measurement campaign.

For most of the classes the distributions look as expected, with relatively normal distributions, although with rather high variance in some LC classes. However, in the case of 'Primary 2' sample, we immediately noticed different characteristics compared to the other samples. To further investigate the distribution of plots within each class, we used the Normal-Quantile plots. The Normal-Quantile plots are plots of points,

$$\left(\Phi^{-1}\left(\frac{i}{n+1}\right),\ x_{(i)}\right),\ i = 1, \ldots, n \qquad (4.5.1.1)$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution and $x_{(i)}$ is the i :th order statistic. Generally, the sampling distributions can be assumed to follow the normal distribution, and the Normal-Quantile plot is a tool to visually detect deviations from this assumption.

Figure 10 presents the Normal-Quantile plots for each LC class measured in the Colombia field measurement campaign. While the other classes show close to normal distribution, the 'Primary 2' sample clearly deviates from the normal distribution. This can be seen e.g. from the large difference between the mean and median (the two horizontal lines). Moreover, the 'Primary 2' sample may be bimodal, with a rather large gap in observations between 200 and 300 t/ha. Investigation for the reasons of possible bimodality revealed some problems in the correspondence between the LC class 'Primary 2' and real-world

'Primary 2' class. The real-world class included two different types of forests, which had shown similar EO characteristics, but have different levels of biomass.

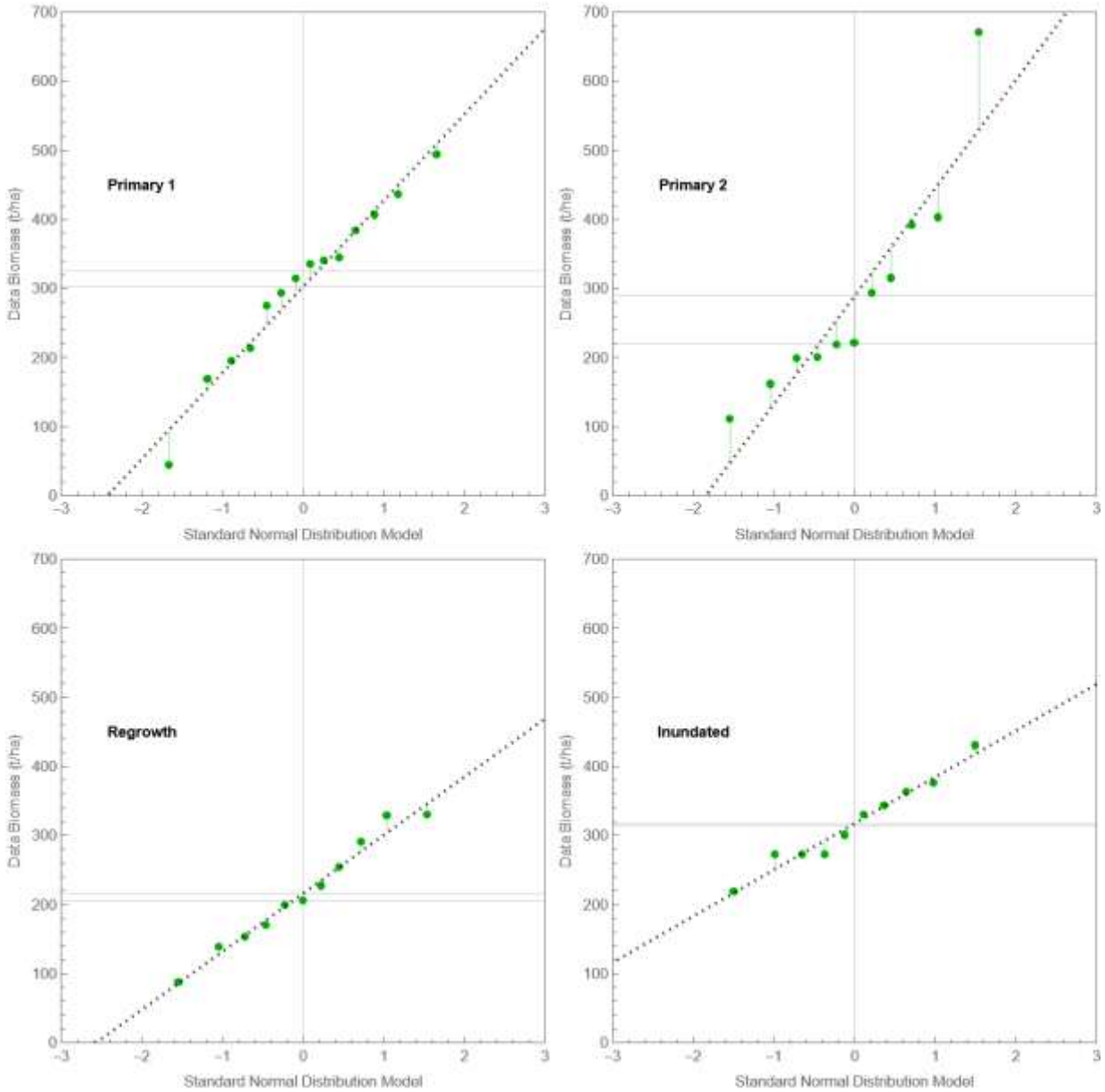

Figure 10. Normal-Quantile Plots of Colombia use case field measurement campaign.

When the normal distribution is a plausible model for a sample, it means that the two parameters, mean and variance, are sufficient to describe the data. The effect of possible outliers can be studied by computing robust estimates of the sample mean and comparing them against the sample mean. Such robust estimates include Trimmed mean, Winsorized mean, and median.

Finally, the confidence intervals around the mean can be computed in various ways, including the t-distribution based model and bootstrapping. We preferred bootstrapping in this example case since it provides meaningful results also when the sampling distribution deviates from the normal distribution.

4.5.2. Design and analysis of sampling for visual interpretation

The visual image interpretation with 5 m resolution NICFI Planet mosaic was conducted to improve understanding on the LC class characteristics and distribution. Altogether 1554 visual sample plots (100 x 100 m) were evaluated, recording the LC class distribution. Within this information, it was possible to finetune LC class distribution in the interest area and define biomass estimates for the classes where no field sampling was conducted, thereby improving the biomass estimates derived from the LC map for particular interest areas.

The sampling design for visual interpretation (Figure 11) was a stratified design with three strata. In two of the strata, indexed by h, $h \in \{1,3\}$, it was also a two-stage design. In stratum $h = 2$, a simple random sampling design was performed. In strata $h \in \{1,3\}$, a grid was first randomly selected and then a simple random sampling was performed from the grid.

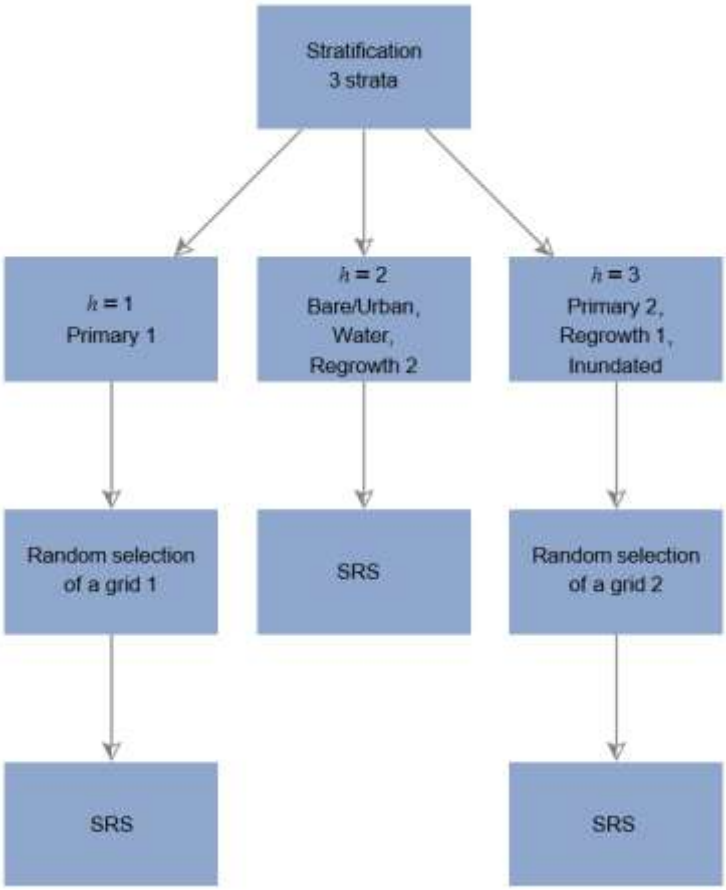

Figure 11. Stratified design for the visual interpretation (SRS stands for stratified random sampling).

Table 4 below collects our notation. The population unit in the visual interpretation sampling is a square of size 1 ha. In the cases of the first-stage selection of a grid of population units, the possible grids are disjoint, and the union of all grids cover the whole area.

Table 4. Statistical notions used.

Notation	Type	Explanation
y	Generic name	Land Cover class y
h	Design variable	Indicates the stratum in question: $h \in \{1, 2, 3\}$
m	Design parameter	The number of all possible grids
N_h	Design parameter	The size of the stratum h population
W_h	Design parameter	Stratum weight
n_h	Design parameter	The sample size in stratum h
k	Random variable	Index of the randomly selected grid in the two–stage designs, $k \in \{1, \ldots, m\}$
y_{hi}	Random variable	The amount of y in the randomly selected i:th population unit of stratum h
y_{hki}	Random variable	The amount of y in the randomly selected i:th unit of the k:th grid of stratum h
$N_h(m, k)$	Random variable	The size of the randomly chosen grid k

The sizes of the grids vary; therefore, the usual sample mean of the two-stage sampling in strata h, $h \in \{1,3\}$, is biased. The following formula defines an unbiased estimate:

$$\overline{y}_h^k = \frac{mN_h(m,k)}{N_h}\overline{y}_h \qquad (4.5.2.1)$$

where $\overline{y}_h$ is the ordinary sample mean. That is, the ordinary sample mean must be multiplied by a factor, which depends on the design parameters and the chosen grid. The variance of the unbiased estimator is

$$Var(\overline{y}_h^k) = \frac{m}{N_h^2}\left(\sum_{j=1}^{m} N_h(m,j)(N_h(m,j) - n_h)\right)Var(\overline{y}_h) + \frac{1}{N_h^2}\left(E\overline{y}_h^k\right)^2 E(mN_h(m,j) - N_h)^2 \qquad (4.5.2.2)$$

See Chapter 11.2 in (Cochran 1977) for a similar example. Finally, the variance of the stratified sample mean is computed by the formula

$$Var(\overline{y}_{st}) = W_1^2 Var(\overline{y}_1^{k_1}) + W_2^2 Var(\overline{y}_2) + W_3^2 Var(\overline{y}_3^{k_3}) \qquad (4.5.2.3)$$

Above k_1 and k_3 refer to selected grids in strata $h = 1$ and $h = 3$ respectively. Recall that sampling in different strata is performed independently of each other, including the first-stage selection of a grid.

With this approach we were able to provide improved biomass estimates with confidence intervals for each of the LC classes. These estimates can be used to calculate benchmark

biomass estimate for the area of interest and any desired sub-areas. The estimates can also be used to evaluate biomass changes together with the benchmark map and activity data. For all of these outputs, confidence intervals can be provided.

4.6 European wide biomass map accuracy assessment

An independent accuracy assessment for the European wide biomass map was conducted in line with the new CEOS LPV protocol for biomass from space calibration and validation. The new CEOS protocol contains a dedicated section about using existing in-situ data as reference for the validation of larger area biomass maps, assuming they are properly screened, processed and harmonized, to allow for comparison with large area biomass map predictions. The validation procedures were mostly developed as part of the CCI-Biomass project (CCI Biomass 2020) and have been slightly adapted to the FCM case.

The accuracy assessment of a European wide map required an effort to include a large number of different reference data sources covering all different geographical regions and forest types across Europe. Thus, we relied on AGB reference data that were not specifically produced for validation purposes but that were rather collected within the context of National Forest Inventories (NFI) and other efforts at local to regional scale. This had consequences, i.e. that we could not rely on a design based sample. Also, the sampling frames were different as the biomass map concerns mean forest biomass density discretised in spatial grid cells (including non-forested area) while the inventories employ non-uniformly sized and typically small plots within forested areas. Thus, specific care had to be taken for the map-plot comparison. The assessments were performed at the map pixel level, as well as spatially aggregated over larger pixel blocks.

It is important to realize that the reference data were also estimates and therefore affected by errors that should be taken into account when using them in the map-plot comparisons (Réjou-Méchain et al. 2017, Réjou-Méchain et al. 2019). We deliberately did not specify the biomass variable of interest, as the retrieval will target growing stock volume (GSV, m^3/ha) and convert this to above-ground biomass (AGB, Mg/ha) and below-ground biomass (BGB, Mg/ha). In principle, European NFIs report all variables in their field data while research inventory plots maintained by scientific investigators do mostly register AGB only.

An extensive dataset of forest in-situ data across Europe was acquired for the purpose of the validation. Plots included in the database underwent a series of quality checks (see below). *In situ* forest data were not used for calibration of the European wide map to guarantee full independence from the production process and because the project's biomass map processing chain did not rely on such calibration procedure. The dataset was part of the *AGBref* database, a global collection of forest biomass reference dataset (Araza et al. under preparation).

The following *in situ* data selection criteria were used for product validation. *In situ* data needed:

- A proper citable reference source and metadata to assess the procedures and quality of biomass prediction.
- Precise coordinates (4-6 decimals for coordinates in decimal degrees).
- A census date within ten years from the reference year of the map products to avoid temporal inconsistency with the assessed maps.

- Measurements of all trees of diameter ≥ 10 cm (or less) were included in the estimates.
- Sites that were not deforested between the year of the inventory and the reference year of the biomass map. This assessment was based on the forest loss layers of the Hansen dataset (Hansen et al., 2013).

For map product validation, the response design encompassed different steps leading to the assessment of differences between map and plot values and the data harmonization procedure is pictured in Figure 3. The plots used in our comparison may have been surveyed at a different time than the map to be assessed, they typically differ in spatial support (i.e., the area covered by individual plots) from the map pixels and they measure different spatial entities (average biomass over a pixel area versus forest biomass within a forest plot). Therefore, data harmonization was needed prior to the analysis of differences.

Differences between the inventory date of inventory plots and the reference year of the map were harmonized using updated IPCC growth rates (IPCC 2019). For dealing with the distinct sampling populations in terms of both different spatial support and the inclusion of non-forested areas within map pixels we multiplied the temporally adjusted plot measurements by forest fraction. This forest fraction was computed by putting a 10% threshold on a tree cover product (or any other available forest map provided by the user). This was undertaken both at pixel level and over larger aggregated blocks. In the rare case of more than one inventory plot occurring within a pixel, the average of the adjusted biomass per plot was used. The correction for forest fraction was applied only to plots with an area below 1 ha.

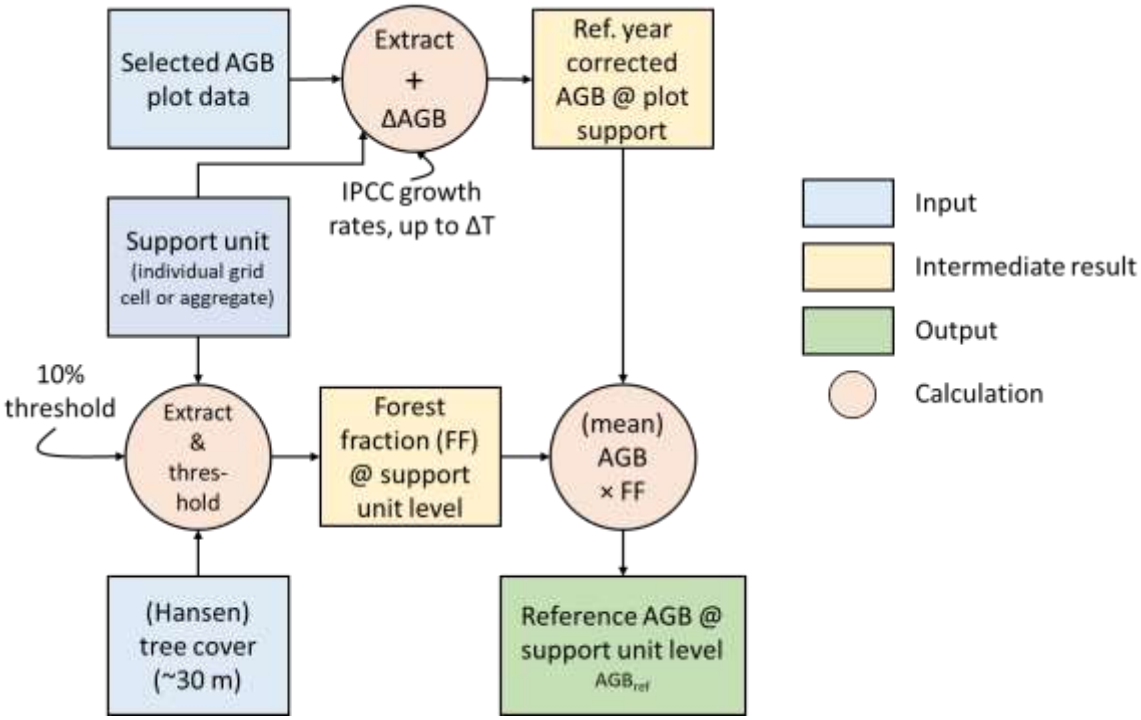

Figure 12. Overview of reference data harmonization steps for the European wide map.
The flowchart refers to AGB as the forest variable of interest.

As straightforward way of analysing map vs. plot differences and account for the expected differences in the accuracy of plots in different size categories, we introduced a weight to reflect the “quality” of the plot data in the accuracy analysis i.e., plot biomass estimates

were aggregated with inverse-variance weights so that the resulting reference value matched the spatial support of the map (*Plot2Map* approach, Figure 12). The accuracy reporting was done for different biomass ranges.

Two temporally matched *AGBref* subsets were prepared: the 2020 *AGBref* subset was used to validate the 2020 map, whereas the 2015 subset was paired with the 2017 map. This strategy limited growth-rate mismatch while guaranteeing an adequate sample size for error calculation. For validation at 10 km grid level, only those 10 km grid cells that contained more than five field plots were selected, in accordance with the "minimum-plots" quality flag (Araza et al. 2022). The filtered *AGBref* locations formed a dense corridor across continental Europe, with highest concentrations in France, Germany, Poland and the Czech Republic as well as across the Baltic–Scandinavian belt Figure 13).

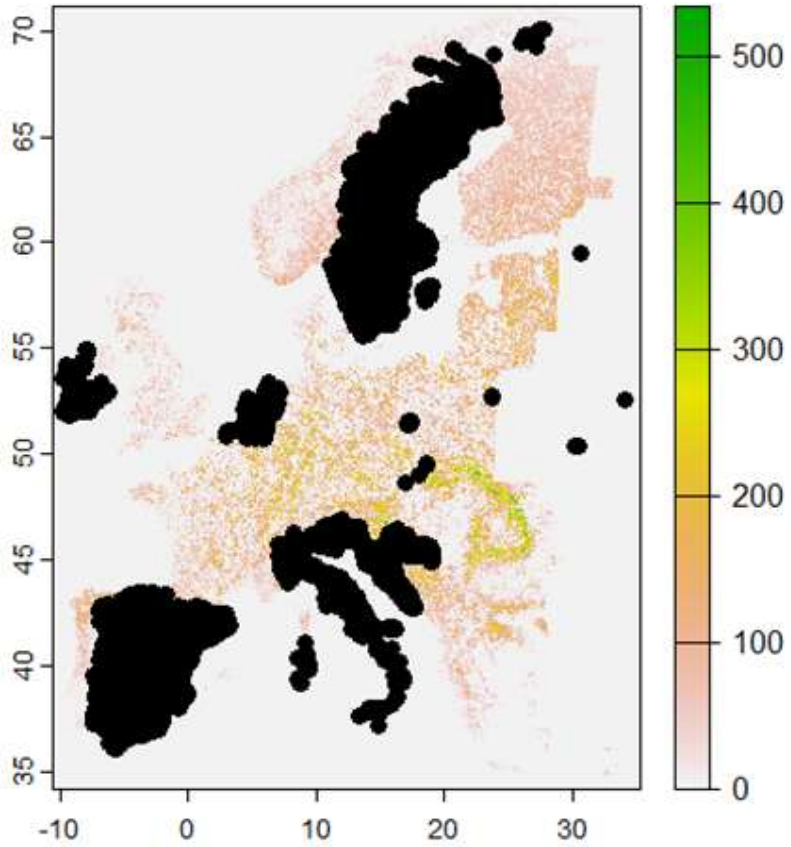

Figure 13. Overview of the locations of AGBref plots used and compared with the European wide map.

5. Algorithm descriptions

5.1 Overview of the section

In this section, we describe the underlying algorithms of the FCM tools used to predict continuous forest structural variables and changes. EO-based features derived from the data sources described in chapter 3 (or similar sources) are typically used as predictor variables, while in some cases also Airborne LiDAR Scanning (ALS) based features can be used. Some of the algorithms are applicable for predicting one forest attribute at a time, while several approaches can produce predictions of multiple forest variables simultaneously. Depending on sensor considerations, some tools are suitable or optimal for predicting only selected sets of forest variables. A good example is the better suitability of optical satellite data to predict forest tree species composition, while radar backscatter at longer wavelength should be a better candidate for GSV prediction.

The algorithms described in this chapter include:

1. Probability, a forest classification and prediction algorithm.
2. k-NN, a non-parametric algorithm widely used in forest monitoring.
3. UNet, a popular convolutional neural network recently introduced for pixel-level forest mapping regression task (forest variable prediction from EO images)
4. BIOMASAR, a physical approach for forest growing stock volume and biomass prediction.
5. Autochange, an image-to-image change detection algorithm.
6. PREBAS, a process-based ecosystem model for prediction and forecasting of biomass and carbon fluxes.
7. Data assimilation, an approach to combine information from several input sources to enable consistent temporal monitoring of forest areas.

For potential user of the FCM tools, it is important to understand the potential and limitations of the available monitoring algorithms. This chapter provides the description of the algorithms and observation of the performance of the algorithms achieved in the FCM use case demonstrations.

5.2 Probability

5.2.1 Algorithm description

The Probability forest classification and prediction approach (Häme et al. 2001) approach includes three phases: 1) Proba Cluster, 2) Proba Model and 3) Proba Estimates. The overall workflow of the forest structural variable prediction is illustrated in Figure 14. The process is started with the Proba Cluster module, performing image clustering of the input images using k-means clustering and maximum likelihood classification. After the clustering, the Proba Model module is used to associate the field measurements with the clusters. Both spectral statistics and forest variable values are needed for each cluster whose number is a parameter value. The forest variable values for each cluster can be computed as an arithmetic mean or median of all the field measurements belonging to this cluster.

Based on earlier experience (Häme et al. 2001; Sirro et al. 2018; Miettinen et al. 2021), median value is recommended to derive cluster values for most variables, while mean value of the sample plots falling into a given cluster is used for proportional variables (such as tree species proportions). The median approach is less affected by potential outlier plots, but the average approach produces more reasonable predictions for the proportional variables (ensuring them summing up to 100%).

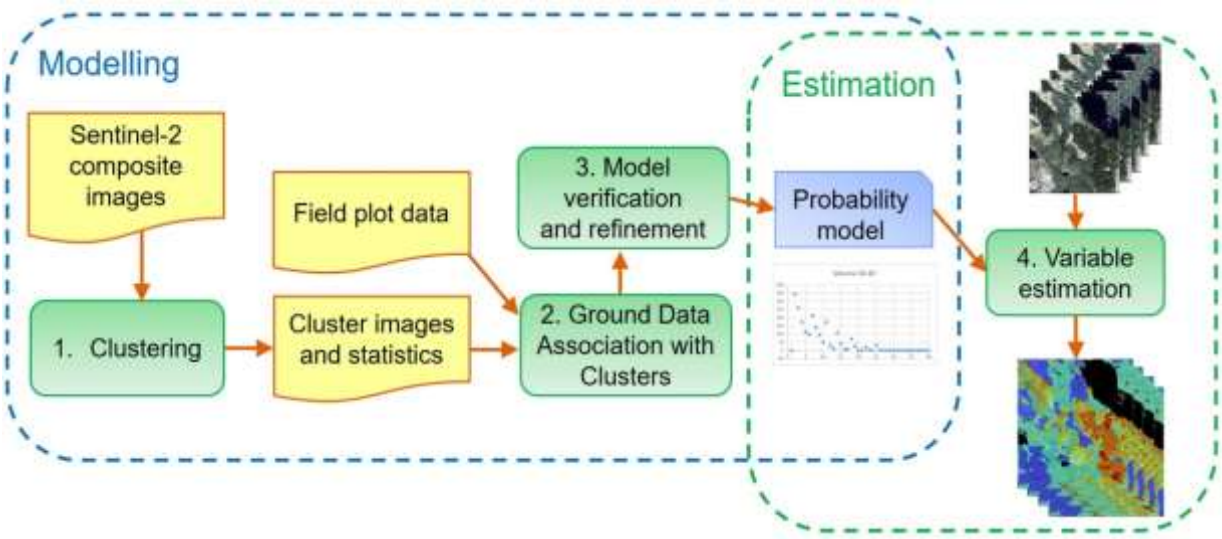

Figure 14. Overall workflow of the forest structural variable prediction using the Probability approach.

The resulting model can be analysed by comparing manually cluster spectral distribution in the spectral coordinate system and via visual analysis of a satellite image. This manual phase allows modification of the model. For instance, clusters representing non-forest may lack field observations and consequently reference data values, but their structural variable values can be manually set to zero.

Finally, the Proba Estimates is used to compute a forest-variable prediction for each image pixel. A multivariate normal distribution for each cluster is characterized using its mean vector and covariance matrix. A cluster membership probability for a spectral vector x is computed for five spectrally closest clusters and these probabilities are scaled to sum up to 1. These cluster membership probabilities are used as weights when deriving a final prediction for a given pixel as a weighted sum of reference data values for five spectrally closest clusters (Häme et al., 2001), calculated as:

$$f(x) = \sum_{c=1}^{N} P(c|x)\, f_c \qquad (5.2.1.1)$$

where *f(x)* is the target variable value for spectral vector *x, P(c|x)* the probability for spectral vector *x* belonging to cluster *c, f_c* the target variable value for cluster *c* and *N* the number of clusters.

5.2.2 Performance

Overall, the Probability method provided comparable results to the k-NN algorithm in sites where both methods were applied and compared. However, the greatest benefit of the Probability method is that it can be applied to areas with very limited field reference data available (e.g. less than 50 plots). In these situations, it may be impossible to apply the k-NN method at all, or the output may be of very low quality. The Probability method, on the other hand, allows manual investigation and modification of the model, which makes it feasible to use it in areas with limited field reference data availability.

The Probability method was used in five FCM use case demonstrations: Galicia, Ireland, Extremadura, Styria and Peru. In addition, it was tested in several test sites during the main project. As with all the prediction models, the Probability error metrics varied strongly between demonstration sites, depending on the availability and type of field reference data, the used EO datasets and EO image quality. Table 5 provides two examples of the error levels in two sites, one in Ireland and one in Finland. In both of these examples, a combination of Sentinel-1 and Sentinel-2 data was used. The Ireland results can be considered exceptionally good, while the Finland results provide a more typical level of expected accuracy.

Table 5. Examples of the error levels of output products produced with the Probability method.

		G (m^2)	D (cm)	H (m)	GSV (m^3/ha)	N (N/ha)	Spruce% (%)	Pine% (%)	Larch% (%)	BL% (%)
Ireland	RMSE	6.87	3.04	3.08	54.47	520.83	25.36	5.57	36.17	22.00
	RMSE %	20.13	18.99	21.16	28.4	30.24	48.06	60.88	162.86	139.33
	Bias	-1.00	-0.53	-0.55	-7.85	50.94	6.04	1.93	-16.96	9.04
	Bias %	-2.93	-3.31	-3.76	-4.09	2.96	11.45	21.07	-76.35	57.28
Finland	RMSE	6.87	6.23	5.04	85.88		28.01	30.48		26.87
	RMSE %	39.02	37.1	33.37	54.46		87.5	72.27		108.99
	Bias	0.65	-0.16	0.17	8.42		0.63	1.13		-2.1
	Bias %	3.71	-0.95	1.1	5.34		1.98	2.68		-8.51

*D = diameter, G = basal area, H = height, GSV = growing stock volume, N = density/number of trees, Spruce% = proportion of spruce, Pine% = proportion of pine, Larch% = proportion of larch and BL% = proportion of broadleaf

Figure 15 illustrates the effects of EO dataset combinations in forest variable prediction with the Probability method. The clear improvement in volume prediction with the inclusion of TanDEM-X can be seen as a narrower point cloud. The RMSE and RMSE% for this case were 73.72 m^3/ha and 46.75%, with the relative bias of 2.18%. In comparison, the corresponding values using Sentinel-1 and Sentinel-2 were 85.88 m^3/ha and 54.46%, with the relative bias of 5.34% (Table 5).

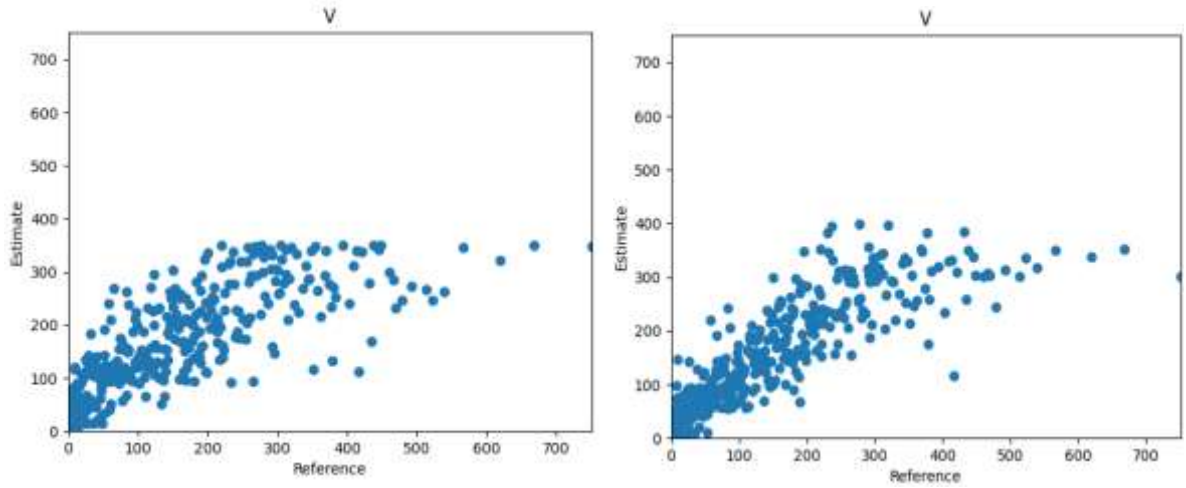

Figure 15. Scatter plots of GSV prediction with Probability using Sentinel-1 + Sentinel-2 (left) and Sentinel-2 + TanDEM-X (right).

Overall, the experiences with the Probability method gained over the course of the FCM project highlight the significance of the available datasets (both reference data as well as EO data). The available dataset combinations largely define the level of uncertainty that can be reached in a given site. In addition to that, there are naturally variations between sites due to ecological and environmental conditions. All this leads to fact that it is very difficult to define an expected range of error levels for a given site before tests with the available datasets combinations have been conducted.

5.3 k-NN

5.3.1 Algorithm description

The k-Nearest Neighbour method (k-NN; Alt, 2001) is a popular non-parametric and distribution-free algorithm for forest monitoring. It has been widely used to predict numerous forest structural variables in different parts of the world (Chirici et al. 2016). When abundant field reference dataset is available (preferably over 100 field sample plots), it provides a fast and efficient tool for conducting forest mapping and monitoring in the target area. As a multivariate method it allows predicting several target variables simultaneously, thus ensuring also their relationships.

In the k-NN algorithm, the predictions for the target variable values (such as GSV or AGB) are obtained as linear combinations of the attribute values in a set of k units selected from a reference set of units with known values (Figure 16). The choice of these units is determined by a distance-metric defined on the auxiliary variable space. The k reference units with the smallest distances to the target unit in the auxiliary space are selected. Simultaneous prediction of all forest variables distinguishes k-NN from most other prediction approaches. It is a non-parametric estimator since predictions can be made without estimating any parameters, as well as distribution-free prediction approach because predictions can be made without any prior distributional assumptions.

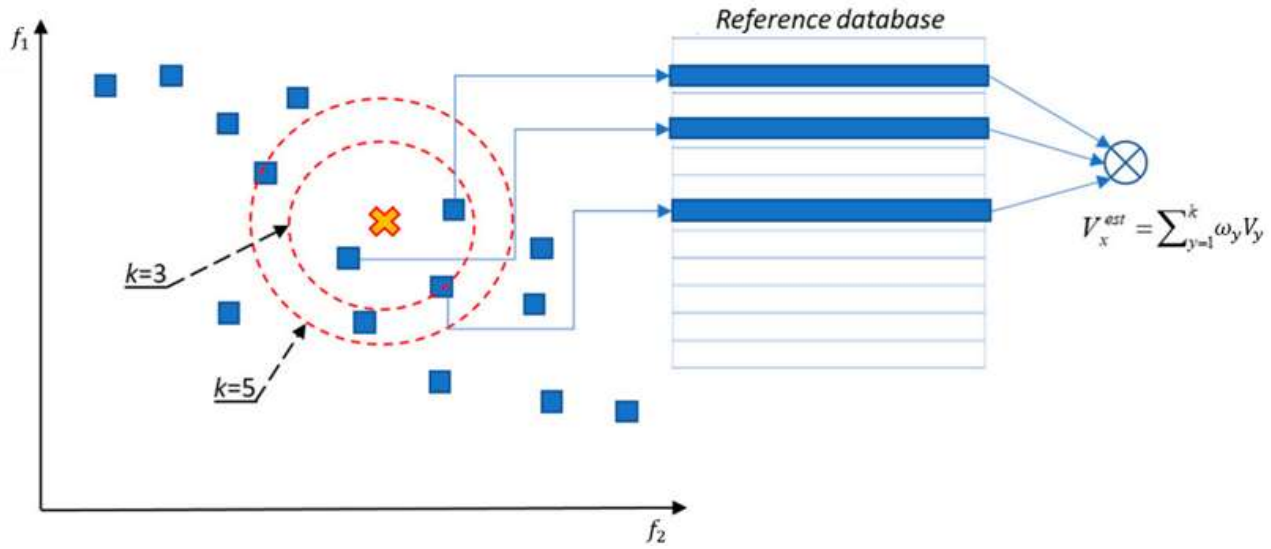

Figure 16. Schematic representation of k-NN method for predicting continuous variables (Antropov et al. 2017).

The considered reference and prediction units can be forest plots, pixels or stands. The k-NN predicted vector $\hat{y}_p$ for pixel p is calculated as

$$\hat{y}_p = \sum_{i \in I} w_{i,p}\, y_i, \qquad (5.3.1.1)$$

where y_i is the vector of observations for the i-th contributing unit in the reference set, I is the subset of contributing units that are nearest with respect to the distance metric, and $w_{i,p}$ is the weight of *i*-th contributing unit calculated as

$$w_{i,p} = \begin{cases} \dfrac{d_{i,p}^{-t}}{\sum_{i \in I} d_{i,p}^{-t}} & i \in I \\ 0 & \text{otherwise} \end{cases}, \qquad (5.3.1.2)$$

where $t \in [0,2]$. Common choices are $t = 0$, which weights all reference set plots equally thus making the prediction a simple averaged vector, and $t = 1$ or $t = 2$ which weight units inversely to their feature space distance or distance squared from pixel p. Popular selections for the distance metric are Mahalanobis distance (Kendall & Stewart, 1968) or weighted Euclidean distance,

$$d_{i,p} = \sqrt{\sum_{l=1}^{L} v_l \left(x_{i,l} - x_{p,l}\right)^2}, \qquad (5.3.1.3)$$

where $d_{i,p}$ denotes the distance in feature space between pixels i and p, and l indexes the features; and vector v_l consists of weights associated with l individual features.

5.3.2 Performance

Overall, the k-NN algorithm was found to work reliably and consistently in different types of conditions and with a variety of variables, as long as sufficient number of reference field plots are available. It is not recommended to use the k-NN approach with less than 100 reference field plots, unless it has been verified that the plots provide a representative sample of the entire target population and range of EO data spectral values. As already discussed above, the performance compared to the Probability method is rather similar. The benefit of k-NN is the fast and easy implementation. But with small numbers of reference field plots the Probablity method is a safer option.

The k-NN method was used in three FCM use case demonstrations: Romania, Catalonia and Norway. In addition, it was used as the benchmark algorithm for making dataset comparisons in the testing sites during the main project. The experiences gathered from these tests and demonstrations provided valuable information on the usability of the algorithm and the level accuracy that can be reached in various ecological and environmental conditions, and with a wide range of EO dataset combinations.

Table 6 provides two examples of the error levels in output products produced with the k-NN algorithm, one in Romania and one in Catalonia. In both of these examples, a combination of Sentinel-1 and Sentinel-2 data was used. The two sites have rather similar levels of uncertainty, with relative RMSE ranging typically between 30% and around 50% percent depending on the variable, while the bias can be expected to be typically less than 4%. Both of these two examples fall into the general level of uncertainty observed in the cases where the k-NN method has been applied.

Table 6. Examples of the error levels of output products produced with the k-NN method.

		G (m^2)	D (cm)	H (m)	GSV (m^3/ha)	Con% (%)	BL% (%)	AGB (t/ha)
Romania	RMSE	15,31	11,11	54,03	220,48	17,16	17,10	
	RMSE %	33,6	30,7	22,2	43,9	31,4	37,7	
	Bias	-1,46	-1,16	-1,76	-7,84	-1,80	0,96	
	Bias %	-3,2	-3,2	-0,7	-1,6	-3,3	2,1	
Catalonia	RMSE	8,05	6,21	33,69	51,77	28,24	28,23	42,19
	RMSE %	40,2	32,9	39,5	51,1	84,6	42,4	46,8
	Bias	-0,26	-0,27	3,66	-0,57	1,87	-1,77	0,22
	Bias %	-1,3	-1,4	4,3	-0,6	5,6	-2,7	0,2

*D = diameter, G = basal area, H = height, GSV = growing stock volume, Con% = proportion of conifers, BL% = proportion of broadleaf and AGB = above ground biomass

The scatter plots presented in Figure 17 reveal generally rather good agreement with the reference and predicted forest structural variable values. However, two typical tendencies are worth noting. Firstly, the low values are on average overestimated. This can be seen exceptionally clearly in the Romanian basal area predictions, as a sharp rise from zero. Secondly, all of the scatter plots show a tendency of saturation at higher values of the variables. For example, the Catalonia basal area predictions seem to saturate at a level of around 30 m^2/ha, although some values in the reference data are over 40 m^2/ha. Together, these two tendencies lead to “averaging” effect of the predictions, meaning that the predictions tend to gravitate towards the mean of the reference data. Although this does

not affect the average statistics over the interest areas, it is very important to understand the effects of the averaging e.g. in modelling context. The output map products typically show higher proportion of middle-range forests and underestimate the proportion of low and high values.

It is also important to highlight that the averaging tendency is a common feature in all EO based forest monitoring, not restricted only to k-NN algorithm. This is particularly true for traditional machine learning algorithms. New deep learning approaches, such as the UNet method presented in Section 5.4 seem to have high potential in reducing the saturation effects.

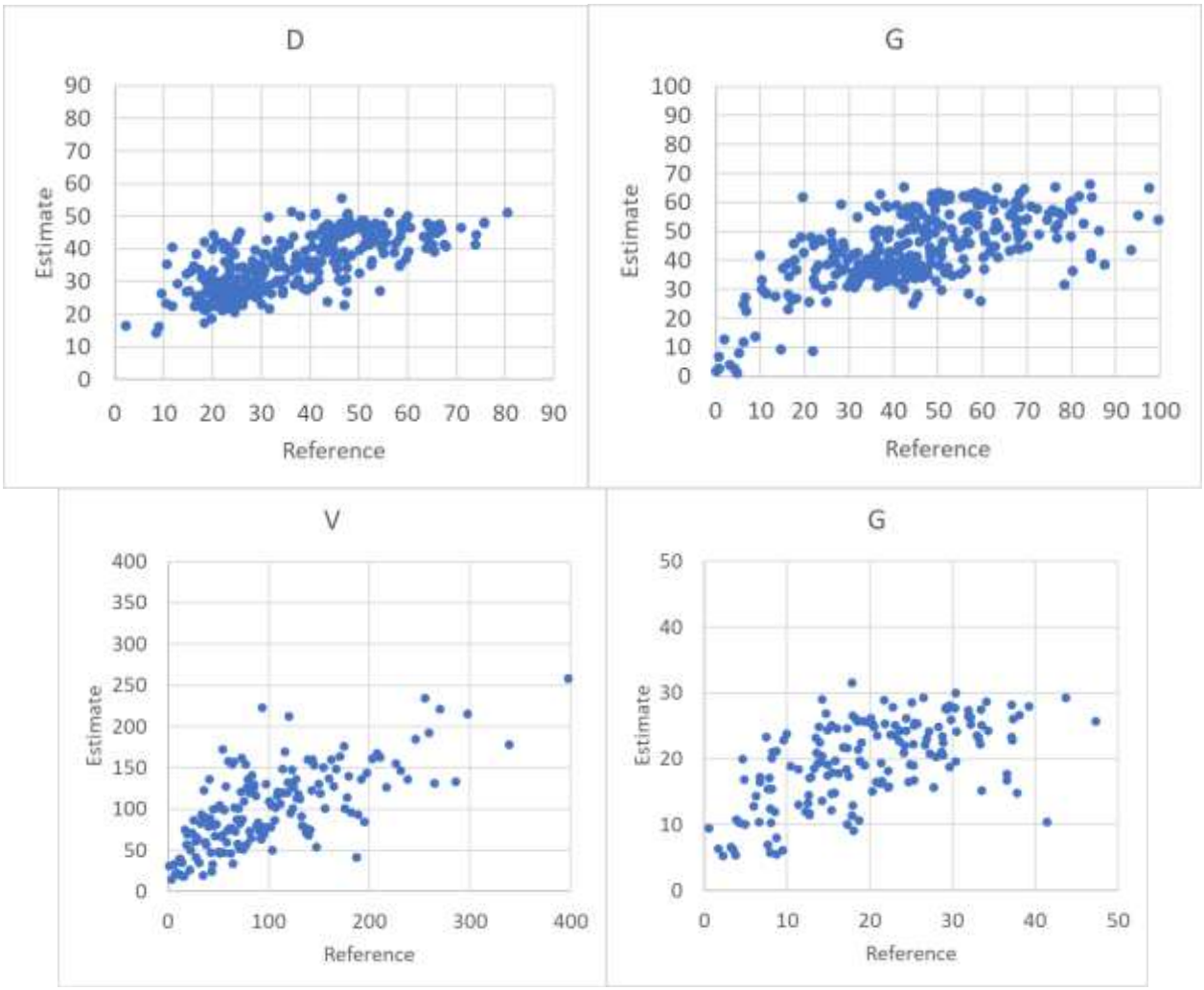

Figure 17. Scatter plots of diameter (D) and basal area (G) from Romania (top row) and growing stock volume (V) and basal area (G) from Catalonia (bottom row). All produced with k-NN method.

The extensive k-NN testing allowed also evaluation of the effects of the EO data combinations. The tests highlight the importance of finding optimal dataset combinations (Figure 18). Particularly the availability of datasets strongly related to the height of the canopy (such as TanDEM-X coherence or canopy height models) greatly improve the prediction accuracy. The findings of Teijido et al. (2025) support the finding that the effects of the dataset combinations are clearly larger than the effects of the methods used in the prediction. This is important to keep in mind when selecting the most suitable algorithm to use in a particular case. It is more important to first gather the EO and reference datasets and then only choose the algorithm that best suit prediction with the available datasets.

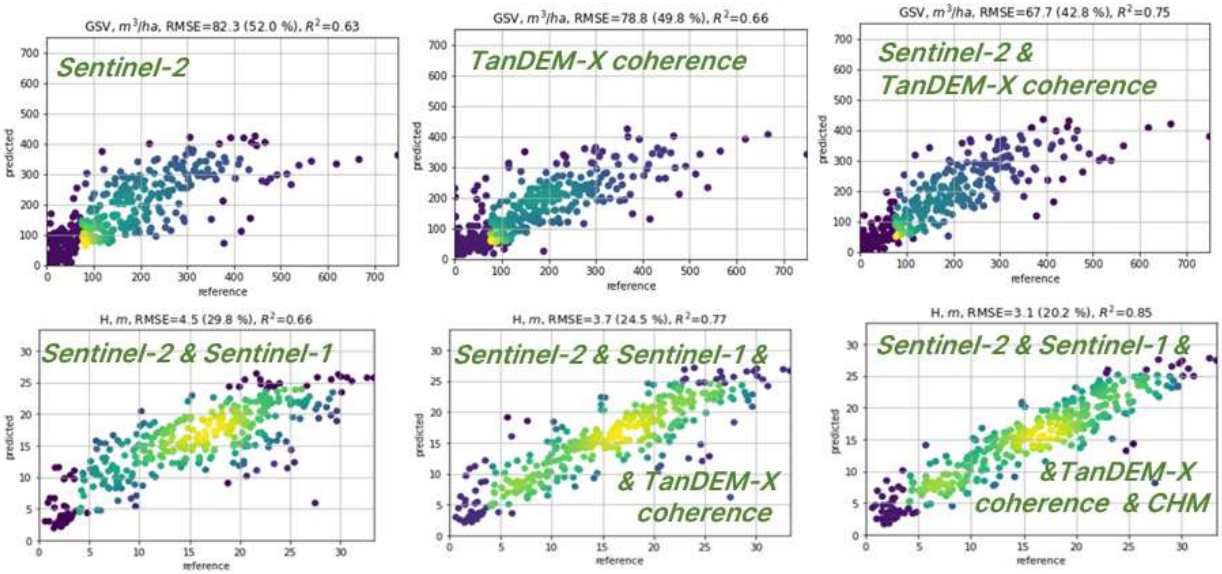

Figure 18. Growing stock volume (top) and height (bottom) prediction with various EO data using the k-NN method.

5.4 UNet

5.4.1 Algorithm description

Since recently, deep learning approaches popular in computer vision tasks and consistently beating conventional machine learning methodologies across various benchmark datasets, got attention in EO based forest mapping. We selected UNet, a variant of fully convolutional network, as a baseline deep learning model for forest variable prediction from multisource SAR and optical images. The UNet model was originally proposed for biomedical image segmentation and is presently often used in various semantic segmentation tasks.

The basic UNet (also known as Vanilla UNet) uses convolutional network to extract features (Ronneberger et al. 2015). Unlike basic CNN (Krizhevsky et al. 2012), the fully convolutional and skip-connection structures allow UNet to extract deeper features of input data, maintain good fusion ability at all levels, while keeping the feature map size unchanged, suggesting it an excellent choice for pixel-level classification (semantic segmentation) and regression tasks.

The overall structure of UNet is symmetric, similar to encoder–decoder, shown in Figure 19 below. The encoder is responsible for feature extraction, and the decoder restores the feature map to the original size. Each box in the UNet indicates a feature map, where the corresponding size is denoted near the boxes. The blue arrow indicates a double-convolution structure as a core unit of the model, composed by cascading a two-dimensional convolution, batch-normalization and ReLu activation. The two-dimensional convolution captures features at current level and an activation layer projects the obtained feature map to a nonlinear feature space.

A 2×2 pooling downscales the original feature map to half of its spatial size, expanding the receptive field for the subsequent double-convolution. As the model goes deeper, the larger receptive field means more global information of the input data can be captured.

In decoder, the green arrow indicates the upsampling operation to restore the size of feature maps. As the pooling operation discarded some details, applying skip-connection, represented by grey arrows, the shallow feature maps are concatenated to deep features recovered from upsampling. The final arrow represents a 1×1 convolution projection function, which maps the last feature map to the target space. The 1×1 convolution kernel size preserves the spatial size and enables pixel-level prediction.

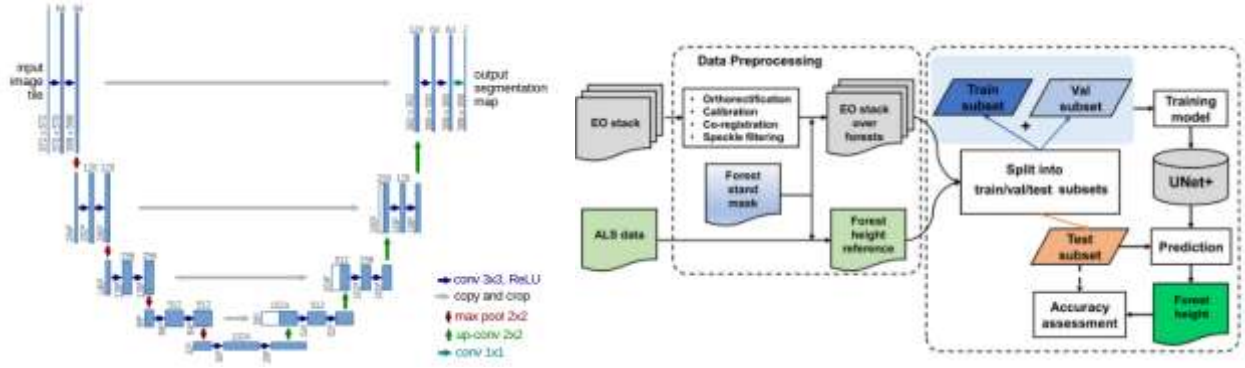

Figure 19. Basic UNet model structure after Ronneberger et al (2015) and the overall UNet model pretraining pipeline for EO based forest variable prediction.

The model can be effectively trained using spatially explicit representations. In scenario when forest plots are available, training from scratch typically leads to unsatisfactory results not better than with conventional pixel-based methods as spatial representations are not effectively learned. In this situation, and effective approach is pretraining a general model using fully segmented labels, followed by model finetuning with forest plots. The model can be further enforced with attention mechanisms (Ge e al. 2022) or used as a part of semi-supervised contrastive regression approaches (Ge et al. 2023b). The UNet algorithm has also shown good results in transfer learning tasks with forest plot data (Ge et al. 2023).

5.4.2 Performance

In the FCM project, the UNet algorithm was used to produce the demonstration products in the Catalonian and Norwegian use case areas. In addition, the algorithms were tested in several testing sites, including training and finetuning the models with different types of datasets. Two main approaches were evaluated: 1) training and application of the model in target area and 2) geographic or temporal transfer of the model to target area or year. The findings on the performance of these two types of applications are highlighted in this section to illustrate the level accuracy that can typically be reached with the model in different situations.

Table 7 provides an example of error level reached in the Norwegian demonstration area, with a model trained using ALS based wall-to-wall forest variable layers and multi-source EO data (Sentinel-2, Sentinel-1 and PALSAR-2 mosaic). Comparison to k-NN results from the same use case reveal consistently better RMSE and R^2 metrics of the UNet based results. Particularly the large difference in the R^2 values indicate larger explaining power of the UNet algorithm.

Table 7. Examples of the error levels of output products produced with the UNet method, compared to corresponding k-NN results.

	UNet				k-NN			
	D	G	H	V	D	G	H	V
RMSE	5.50	8.56	3.33	79.92	5.57	9.44	3.81	91.27
RMSE %	35.3	51.9	28.4	69.4	36.0	58.2	32.6	80.8
Bias	0.84	1.27	0.56	12.49	0.05	0.06	0.11	0.77
Bias %	5.4	7.7	4.8	10.9	0.3	0.4	0.9	0.7
R^2	0.32	0.64	0.61	0.63	0.13	0.5	0.42	0.45

*D = diameter, G = basal area, H = height, V = growing stock volume

However, the UNet based results also have significant bias in the Norwegian use case, compared to the nearly unbiased k-NN results (Table 7). Depending on the use case, this may or may not have significant effect on the usability of the method. For example, if the maps are used as input to model-assisted estimation, the bias will be corrected in the estimation phase. The bias is also visible in the density scatter plots, which otherwise show high agreement between the predicted and the reference values (Figure 20).

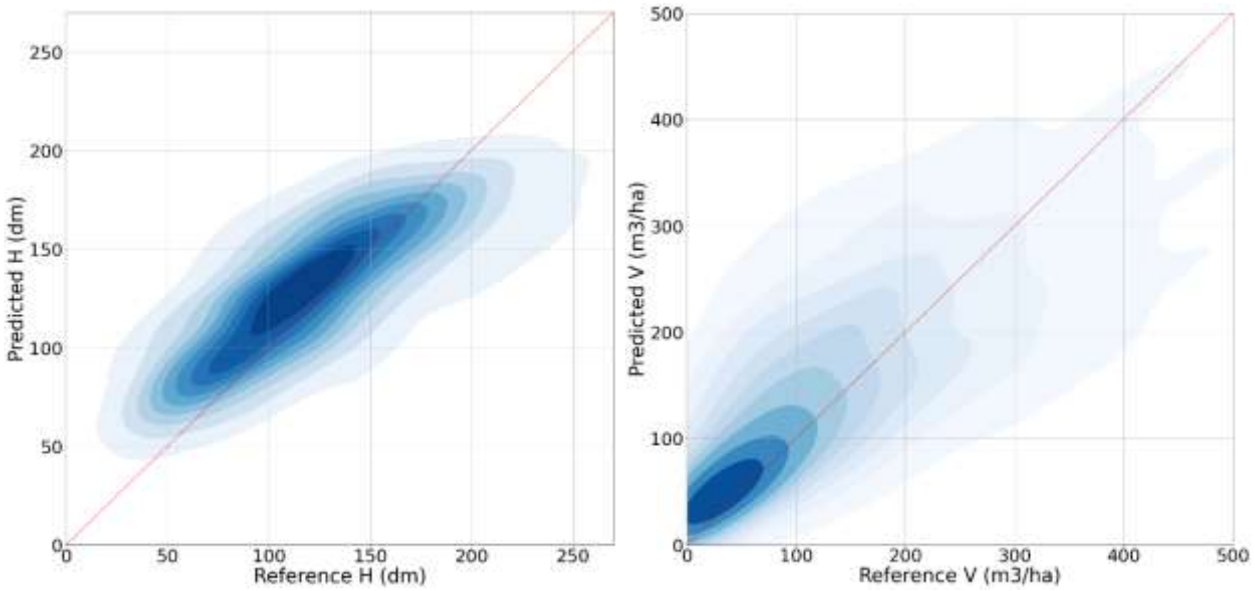

Figure 20. Height (left) and growing stock volume (right) density scatter plots in the Norway demonstration use case for UNet algorithm.

For operational application of the UNet models, a model transfer (either geographic or temporal) is often required. This is because suitable training data may not be available in the target area. In this case, a model trained in similar ecological conditions can be transferred to the target area by finetuning it with a small number of plots. Model transfer does not necessarily affect negatively the accuracy of the results. To illustrate the behaviour of models before and after fine-tuning, Table 8 presents height prediction results with a UNet model that was trained in Finland with ALS based wall-to-wall forest variable maps and applied in Norway without (“blind”) and with fine-tuning (“fine-tuned”). The results were compared to k-NN and a UNet model trained with Norwegian ALS based wall-to-wall forest variable maps (“SR16”).

Table 8. Examples of the error levels for height prediction with various UNet model training options and the benchmark k-NN method. See text above for more details.

	Height			
	k-NN	UNet		
		Blind	Fine-tuned	SR16
RMSE	36.05	46.73	31.42	31.38
RMSE %	31.7	41.1	27.6	27.6
Bias	3.66	0.31	8.59	7.22
Bias %	3.2	0.3	7.5	6.3
R^2	0.44	0.06	0.58	0.58

It can be seen that the model finetuning brought the results to the same level that could be reached with the model trained with local reference data. Figure 21 illustrates how model transfer changes the shape of the prediction scatterplots, compared to “blind” application and a model trained with local reference data.

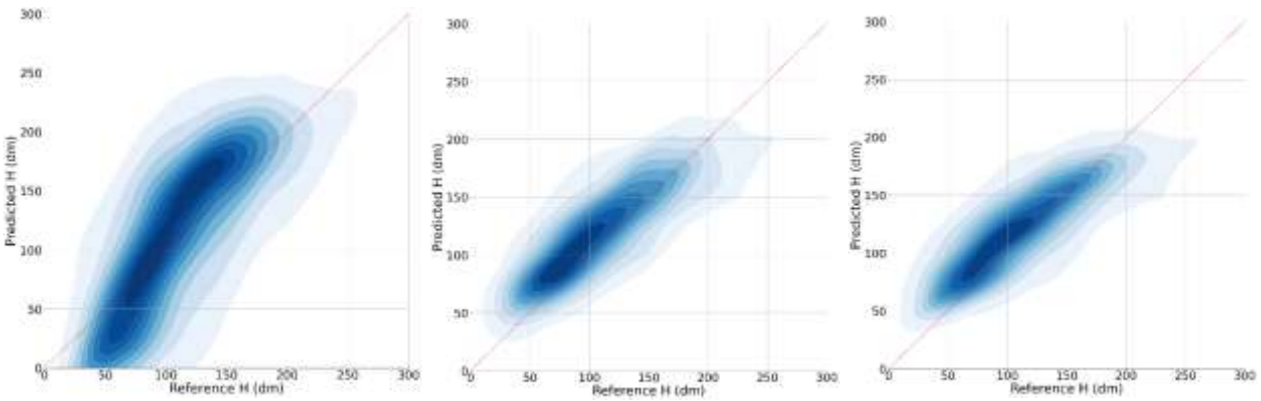

Figure 21. Density scatter plots of height prediction in Norway with “blind” application of Finnish model (left), fine-tuned Finnish model (centre) and Norwegian model (right).

Overall, comparisons between k-NN and UNet based forest structure maps revealed consistently better error metrics for the UNet models. Furthermore, the maps produced with UNet models resulted in a more natural-looking distribution of forest variable values, with clearer distinction of adjacent forest stands (Figure 22). It also enabled prediction of higher volume and biomass, reducing the saturation effects observed in volume and biomass predictions. The improved high-volume prediction was particularly clear in stand-level accuracy assessment conducted in the Norwegian use case demonstration (Figure 23). It is important to realise that all of the reference stands had volume above 120 m^3/ha. There is relatively good agreement in the UNet maps up to around 400 m^3/ha, while saturation effects start to be very clear in the kNN maps already from around 250 m^3/ha onwards.

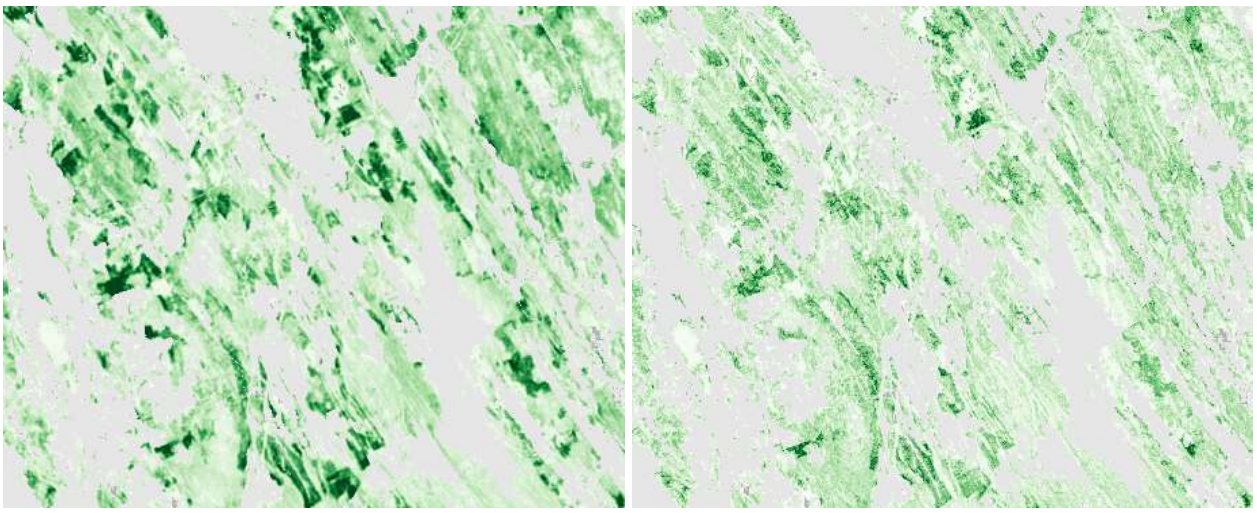

Figure 22. Volume maps produced with UNet (left) and k-NN (right). Grey indicates non-forest area. Volume (green) range 0-600 m^3/ha.

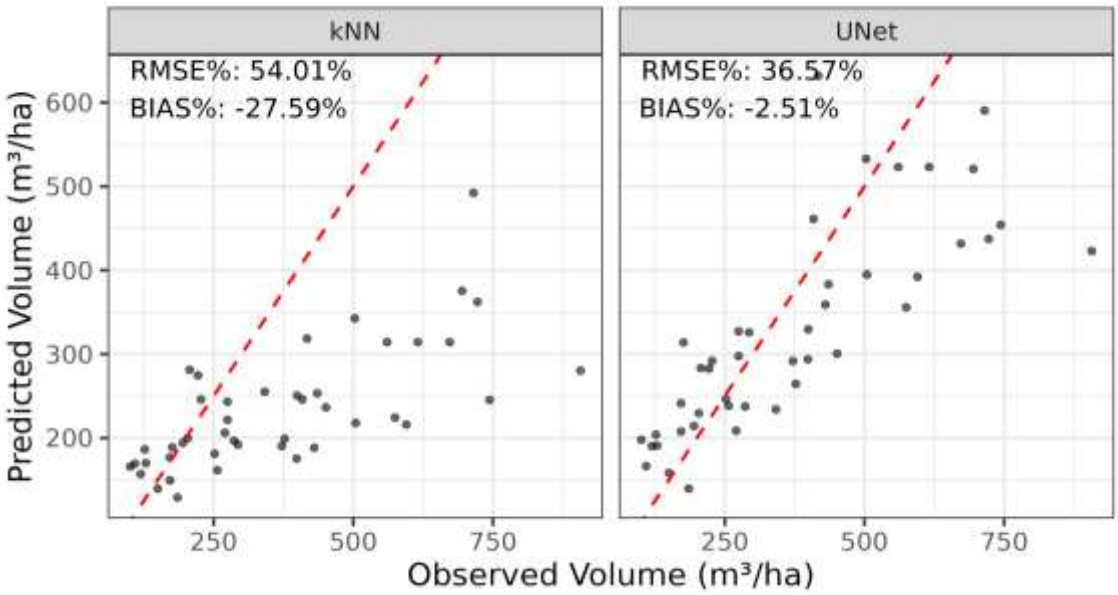

Figure 23. Observed vs. predicted mean volume per stand for kNN and UNet-based maps.

A potentially negative aspect of the UNet algorithm is that it is developed and run individually for different variables. This may, in theory, lead to discrepancy between the forest structural variables. This is also the case with model transfer, as the models for each forest structure variable are fine-tuned separately. The model transfer to the target area is at its best a very effective method to utilise existing UNet model in the area of interest, without the need of extensive field measurement campaigns. However, it is recommended that the representativeness of the plot data used in the fine-tuning, as well as the consistency of the results between forest structural variables is evaluated carefully.

5.5 BIOMASAR

5.5.1 Algorithm description

Prediction of forest biomass density, defined either in the form of structural parameters such as Growing Stock Volume (GSV, unit m^3 ha^{-1}) or in the form of organic mass such as Above Ground Biomass (AGB, unit Mg ha^{-1}), requires observations of both the horizontal (i.e., tree density) and vertical (i.e., height) properties of a forest. To predict the mass, in addition, tree form factors and wood densities are needed. Remotely sensed data from space do not offer such a variety of observations. Therefore, biomass can only be inferred by means of mathematical models, which are tailored to adapt to remotely sensing data available with the aid of reference biomass data from ground surveys. This aspect becomes even more crucial when available measurements have limited sensitivity to biomass, which is the case for satellite missions with imaging instruments currently in operation.

The unavailability of spatially dense datasets of reference biomass measurements for most regions of the world implies that model-based biomass mapping from satellite remotely sensed data of large areas, e.g., continents, requires strong generalizations of local model training. Eventually, this results in strongly biased estimators of biomass (Mitchard et al., 2014, Avitabile et al., 2016). An approach that can overcome such limitations is a self-calibrating method. Rather than training the prediction model with biomass measurements, the model parameters are predicted by deriving statistical parameters of the satellite observations (SAR backscatter in our case) for given forest conditions. Herewith, the biomass prediction model linking biomass to satellite data makes use of auxiliary remotely sensed datasets (e.g., canopy density, LiDAR-based metrics, land cover) together with statistics predicted from forest inventory data (Santoro et al., 2011, Cartus et al., 2012). Despite several approximations in how the model is trained, the performance of such calibration methods, referred to as BIOMASAR, was found comparable to the performance achieved when the models were trained with in situ measurements (Santoro et al., 2011). This suggested the application of BIOMASAR for large-scale mapping of biomass using spaceborne SAR backscatter observations (Santoro et al., 2015). More recently, such methods have been implemented to generate the global datasets of GSV and AGB provided in the framework of ESA's GlobBiomass project (Santoro et al., 2021a) and CCI+ Biomass project (Santoro et al., 2024a).

The BIOMASAR retrieval approach can be summarized as follows:

- Input: SAR backscatter (e.g., from Sentinel-1 or ALOS-2 PALSAR-2)
- Retrieval model: Water Cloud Model integrated with structural functions (allometries) to predict GSV
- Model training: self-calibration
- Feature selection: weighted average of GSV predictions from individual SAR backscatter observations

Prediction of GSV from the SAR backscatter images is based on the method proposed in Santoro et al. (2021b). The relationship between the SAR backscatter and GSV is expressed with the physically-based Water Cloud Model with gaps (Askne et al., 1997) in Eq. (5.5.1.1). This model described the backscattered intensity from a forest as a function

of the backscatter from the forest floor through gaps in the canopy, the backscatter from the forest floor attenuated by the canopy and the backscatter from the canopy.

$$\sigma^0_{for} = (1-\eta)\sigma^0_{gr} + \eta\sigma^0_{gr}T_{tree} + \eta\sigma^0_{veg}(1-T_{tree}) \quad (5.5.1.1)$$

The model parameters σ^0_{gr} and σ^0_{veg} represent the backscattering coefficient of the ground and vegetation layer, respectively. T_{tree} represents the two-way tree transmissivity and is expressed as with α being the two-way attenuation per meter through a tree canopy and h being the depth of the attenuating layer, which is assumed to correspond to the canopy height.

The model expresses the forest backscatter as a function of η and h, i.e., canopy density and height. To establish a dependency upon GSV, the two variables are replaced with forest structural models relating canopy density to canopy height in Eq. (5.5.1.2) (Santoro et al., 2024b), and h to GSV in Eq. (5.5.1.3) (Santoro et al., 2024b). The estimation of the model parameters q, a, b, relies on spaceborne LiDAR and statistics of GSV from NFIs (Santoro et al., 2024b).

$$\eta = 1 - e^{-qh} \quad (5.5.1.2)$$

$$V = \mathrm{a} \cdot h^b \quad (5.5.1.3)$$

In this way, the retrieval model expressed the SAR backscatter as a function of GSV only:

$$\sigma^0_{for} = \sigma^0_{gr}\left(e^{-q(aV)^b} + e^{-\alpha(aV)^b} - e^{-(q+\alpha)(aV)^b}\right) + \sigma^0_{veg}\left(1 - e^{-q(aV)^b} - e^{-\alpha(aV)^b} + e^{-(q+\alpha)(aV)^b}\right) \quad (5.5.1.4)$$

To predict stem volume from a measurement of the SAR backscatter, the model parameters σ^0_{gr}, σ^0_{veg} and α need to be computed first. The prediction is implemented in the form of a model self-calibration not requiring *in situ* data as part of a training set. Self-calibration means that the backscatter of pixels in correspondence of areas with small andlarge canopy densities (e.g., based on the Global Forest Change dataset by Hansen et al., 2013) are extracted and the median backscatter value for each class is calculated. The value for small canopy densities is associated with σ^0_{gr}. The value for large canopy densities is associated with “dense forests” and, therefore, needs to be compensated for residual ground contribution to obtain the value representative of the backscatter from the canopy only, i.e., σ^0_{veg}. The self-calibration is undertaken with a sliding window approach and separately for each image to adapt the predictions of the model parameters to the local conditions of the forest at the time of image acquisition. For a detailed description of the implementation in this project, it is referred to Santoro et al. (2024b).

The key feature of the BIOMASAR approach is the combination of GSV predictions from multiple SAR backscatter observations because it improves the accuracy of the prediction compared to a single-image prediction regardless of the SAR dataset (Santoro et al., 2011; Cartus et al., 2012). Individual predictions of GSV, V_i, are combined into a final value, V_{mt}, with a weighted average (Kurvonen et al., 1999).

$$V_{mt} = \frac{\sum_{i=1}^{N} w_i V_i}{\sum_{i=1}^{N} w_i} \quad (5.5.1.5)$$

Each weight $w_i = (\sigma^0_{veg,i} - \sigma^0_{gr,i})$ is defined as the difference between the predictions of the model backscatter coefficients for the specific image, *i*. This approach indeed favors predictions corresponding to images acquired under conditions that maximize the sensitivity of the backscatter to stem volume (Santoro et al., 2011; Cartus et al., 2012).

An additional step is pursued if several SAR datasets are available (e.g., Sentinel-1 and ALOS-2 PALSAR-2). In this case, each set of SAR observations is piped into a specific BIOMASAR module to exploit frequency-specific strengths and to reduce the impact of systematic weaknesses of each dataset on the final predictions. The dataset-specific predictions of biomass from Eq. (5.5.1.5) are eventually merged with the aim of reducing biases and uncertainties. The merging consists of a weighted average of e.g., the C- and L-band GSV predictions; the procedure is outlined in Santoro et al. (2024b). When SAR images are available for several years, yearly estimates of GSV can be generated. Merging of the C- and L-band datasets is then implemented on a year-to-year basis, i.e., the weights are defined for each year. To harmonize the computation, the estimation of the weights relies on a cost function that is minimized across years following the procedure described in Santoro et al. (2024a).

In principle, the BIOMASAR approach can be implemented to predict AGB instead of GSV by plugging in an allometry that relates canopy height to AGB as currently undertaken in the CCI Biomass project, where AGB is the forest variable of interest (Santoro et al., 2024a). The reason for pursuing a prediction of GSV is that for the European forest landscape, GSV is the primary forest variable of interest. AGB, BGB and carbon-related variables can then be predicted from GSV by simple scaling. Here, we introduce two approaches.

Stem volume can be converted to stem biomass, *SB*, with an estimate of the wood density, *WD* (unit: g/cm^3) in Eq. (5.5.1.6). SB can then be used to predict the biomass density in branches, *BB*, foliage, *FB*, and roots, *RB*, (Thurner et al., 2014) to obtain an estimate of total biomass with Eq. (5.5.1.7). In other words, total biomass represents the sum of the above- and below-ground biomass.

$$\mathrm{SB} = \mathrm{V} \cdot \mathrm{WD} \quad (5.5.1.6)$$

$$TB = AGB + BGB = SB + BB + FB + RB \quad (5.5.1.7)$$

For the wood density, we propose to use average values per leaf type reported by Thurner et al. (2014) because they are based on extensive datasets from European forests. The biomass of branches, foliage and roots is modelled as a function of stem biomass with a power-law function (Thurner et al., 2014). Again, we shall use the coefficient estimates proposed by Thurner et al. (2014) for broadleaves and conifers because they are based on extensive measurements from European forests. For the stratification of the landscape by leaf type (broadleaves and conifers, i.e., either needleleaf deciduous or needleleaf evergreen), we shall use a dataset contemporary to the SAR observations and with similar spatial resolution.

AGB can also be estimated from GSV with a simple Biomass Conversion and Expansion Factor (BCEF). The BCEF is defined as the product of wood density and the stem-to-total

biomass expansion factor representing the proportion of above-ground biomass to the stem biomass (Santoro et al., 2021a).

$$AGB = \mathrm{WD} \cdot \mathrm{BEF} \cdot \mathrm{GSV} = \mathrm{BCEF} \cdot \mathrm{GSV} \quad (5.5.1.8)$$

A global raster dataset of BCEF estimates was generated from extensive measurements of wood density and biomass proportions (Santoro et al., 2021a). The dataset was found to be accurate across almost the entire range of BCEFs worldwide. The major limitation of this dataset is the limited spatial resolution (1 km), which hinders reproducing small-scale spatial patterns of species composition. For this reason, the approach with the BCEF is only introduced to benchmark the approach proposed by Thurner et al. (2014).

5.5.2 Performance

5.5.2.1 BIOMASAR model calibration

The BIOMASAR approach relies on a model relating the forest backscatter to canopy density and height, and on two allometries that relate canopy density, canopy height and growing stock volume, to allow for a direct prediction of the latter from SAR backscatter measurements.

For the allometry that relates canopy density and height (Eq. 5.5.1.2) we produced estimates of the model parameter *q* from ICESat GLAS metrics of the two variables on a 1° tiling basis as currently implemented in the CCI Biomass algorithm (version 6). We also tested the set of estimates of the model parameter obtained with the same LiDAR dataset but different stratification algorithm (Kay et al., 2021). These estimates were characterized by a larger variability of values and caused frequent over- or underestimation of the modelled backscatter at the test sites, thus being deemed as less reliable. Compared to the previous version of this document, our new set of estimates replaces the dataset therein presented and based on Santoro et al. (2022).

For the allometry that relates canopy height and GSV (Eq. 5.5.1.3), we first compared the results derived purely from forest inventory plot data with those obtained from ICESat-2 LiDAR (height, RH98 metric) and inventory statistics (GSV) (Figure 24). The relationship between the two variables was established at the level of provincial averages because of the weak correlation at the level of individual inventory plots. This comparison could be undertaken in six European countries for which plot data from national forest inventories are freely available. The strong similarity of the ensemble allometry (Figure 24) reinforces the use of provincial statistics, published by most NFIs in Europe, and averages from ICESat-2 LiDAR canopy heights. Figure 24, however, shows different relationships depending on the country. To accommodate for the spatial variability of the association between canopy height and GSV, we stratified the ICESat-2 and provincial GSV statistics by ecoregion (Dinerstein et al., 2017), leading us to four allometric functions that describe the relationship between canopy height and GSV across the European forest landscape (Figure 25).

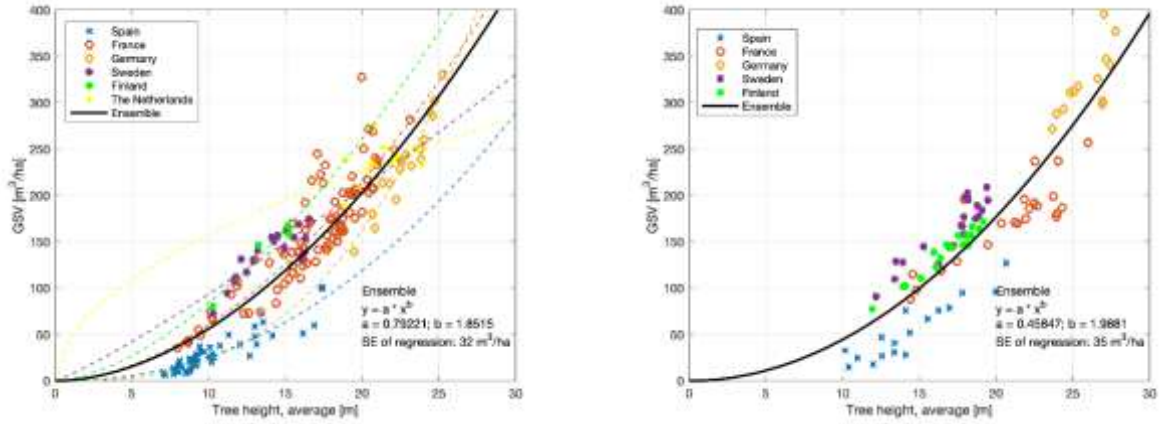

Figure 24. Observations of average tree height and GSV from NFI data for administrative units from six countries and corresponding model fits (dashed curves) using Eq. 5.5.1.3 (left panel). The black solid curve represents the model fit to the ensemble of all observations. For the ensemble, the plot shows the coefficients and the SE of the regression. In the right panel, we show observations of average ICESat-2 canopy height and GSV from NFI data for administrative units from those six countries. The black solid curve represents the fit of Eq. 5.5.1.3 to the ensemble of all observations.

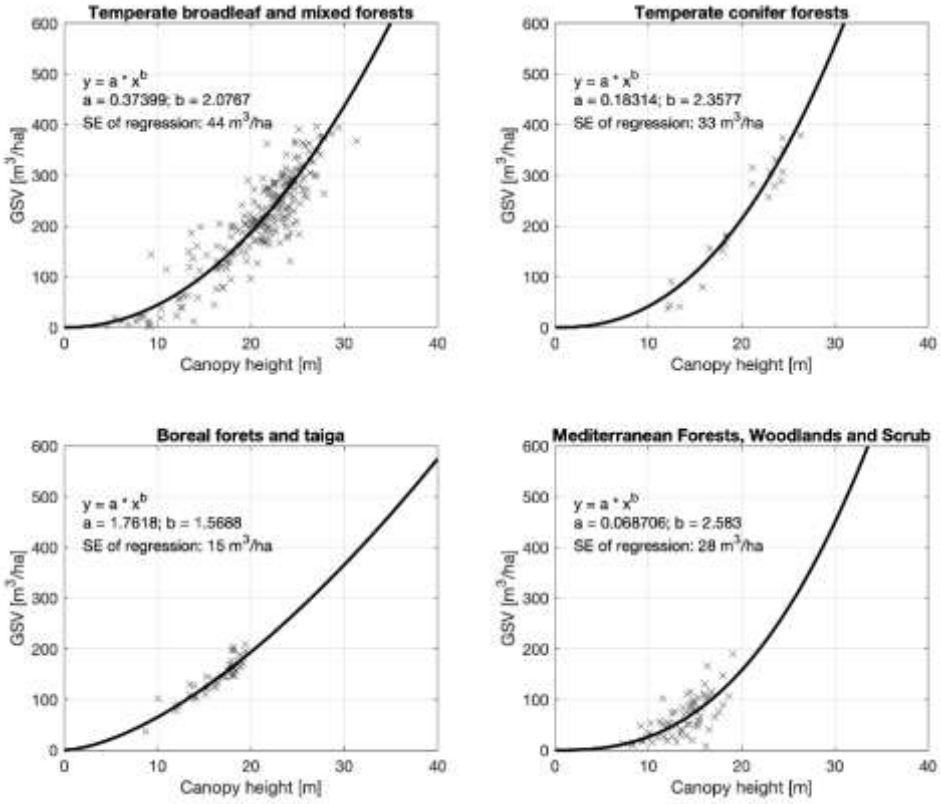

Figure 25. Measurements of canopy height from ICESat-2 data averaged at the level of NFI units and corresponding GSV value published by the NFIs stratified by ecoregion (Dinerstein et al., 2017). In each panel, the curve represents the fit of Eq. 5.5.1.3 to the measurements. Coefficients of Eq. 5.5.1.3 and the standard error of the regression are visualized in the upper left corner of each panel.

Validation of the BIOMASAR approach was undertaken at the sites of Catalonia, Finland N, Finland S and Romania for which extensive observations of GSV from field measurements were available. To validate the self-calibration approach, we compared the modelled backscatter from Eq. 5.5.1.4 with the same modelled backscatter obtained with a

least squares regression to the training dataset at each site and for each date of the Sentinel-1 and ALOS-2 datasets. In Figure 26, the scatter plots show the estimates of the WCM parameters σ^0_{gr} and σ^0_{veg} for the sites of Catalonia and Finland N and the Sentinel-1 datasets. These were the only sites characterized by moderate to high correlation between backscatter and GSV observations. For the Finnish site, we observe strong agreement between estimates whereas for Catalonia, there appears to be a systematic offset depending on polarization and parameter. We explain the discrepancy as a consequence of the fitting procedure based on inventory data, which does often not generate a realistic estimate. This result is relevant in the overall context of deeming reliable such model fits based on ground reference data.

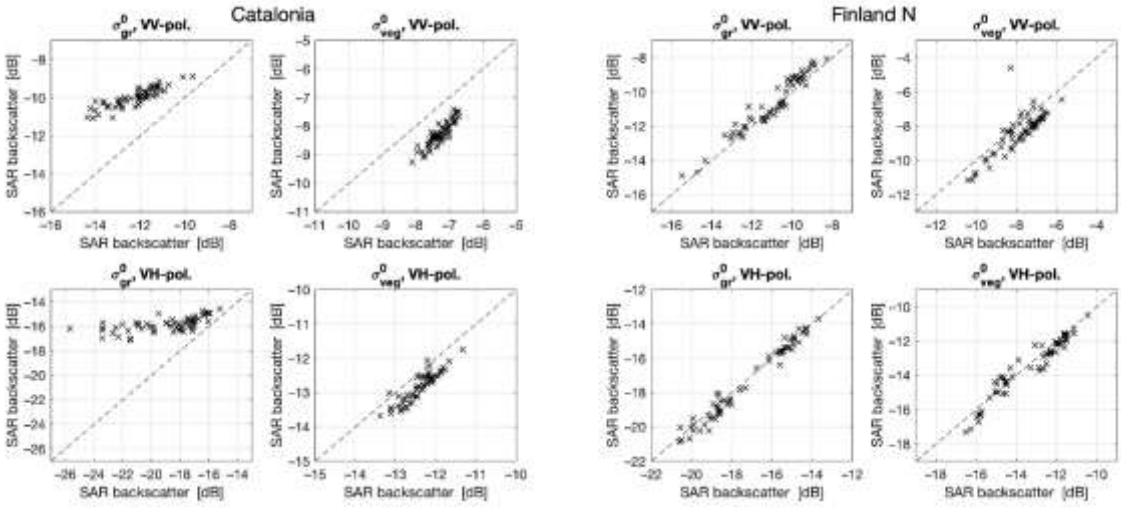

Figure 26. Scatter plots comparing the estimates of σ^0_{gr} and σ^0_{veg} from the regression fit to the observations (x axis) and from the self-calibration (y axis) for VV- and VH-polarized Sentinel-1 images over the test site of Catalonia and Finland N (Santoro et al., 2024b).
The dashed line represents the identity line.

The same analysis was undertaken for the time series of ALOS-2 observations, here represented by the backscatter in the yearly mosaics produced by JAXA (2015-2020). The individual ALOS-2 PALSAR-2 used for the pan-European map were not available at the time of this study. The estimates of σ^0_{gr} and σ^0_{veg} from the two model fitting procedures show an overall strong similarity except for the HV-polarized dataset over Catalonia where the external calibration generated slightly higher values than the traditional regression based on ground reference data. We have already identified such an issue with the Sentinel-1 data and explain the discrepancy because of the potentially unrealistic values associated to σ^0_{gr} and σ^0_{veg} which were due to the low correlation between backscatter and GSV. Our analysis demonstrated the importance of retrieving GSV from multiple L-band observations.

The accuracy of the GSV estimates obtained with the BIOMASAR approach and with the traditional model training procedure based on reference GSV values is reported with respect to a set of inventory plots that were not touched during the model development and calibration phase. The metrics are calculated at plot level. GSV class medians are also displayed in the figures to visualize the tendency. The GSV estimates were obtained for each SAR dataset and then merged to obtain a final value that may reduce sensor-specific over- or under-estimations.

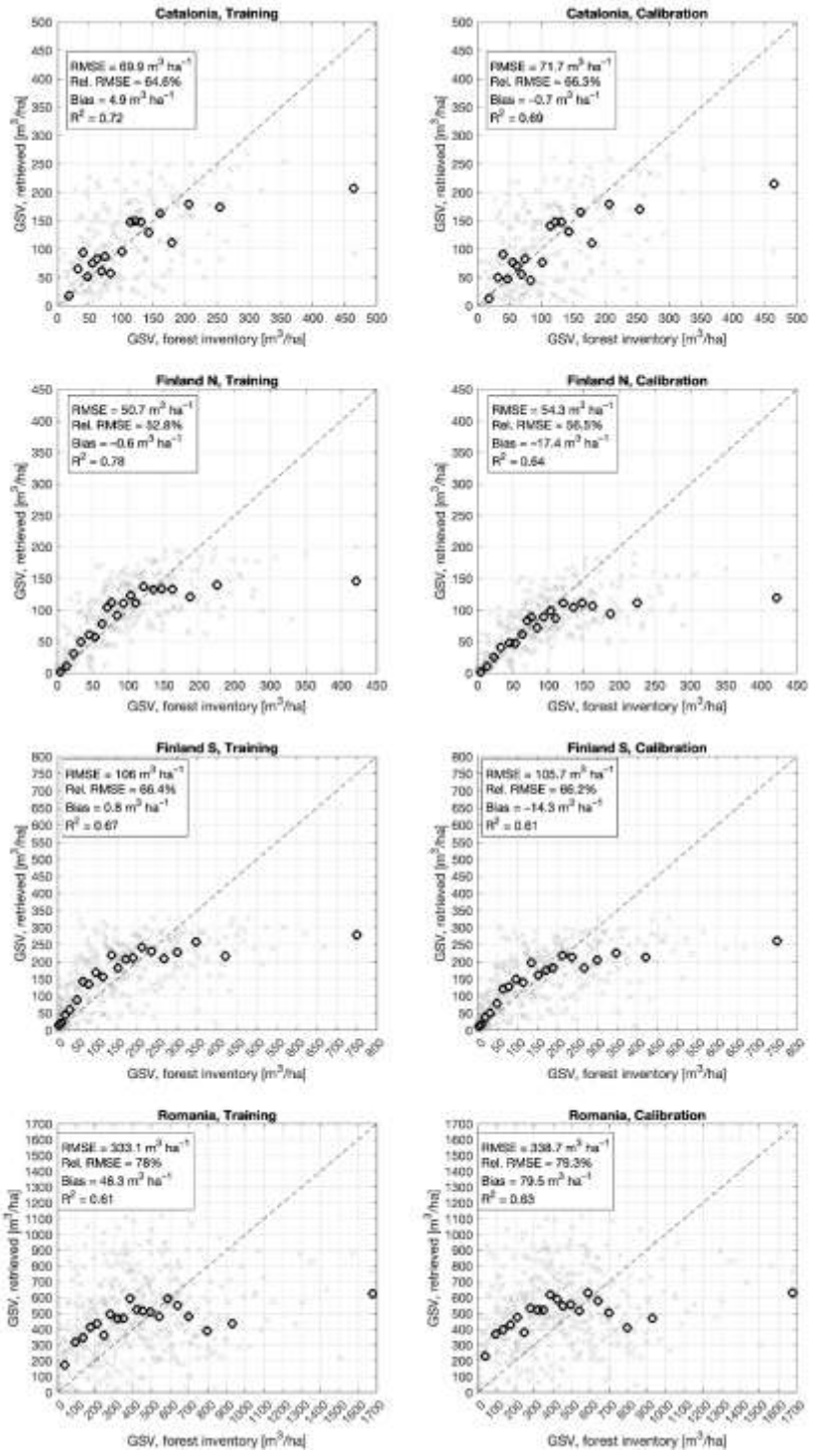

Figure 27. Scatter plots comparing the merged estimates of GSV from the WCM trained with ground refence data ("Training", left panels) and from the WCM calibrated with the BIOMASAR algorithm ("Calibrated", right panel) to the GSV from the inventory data for the test site of Catalonia.
Crosses illustrate the comparison at the plot level. Circles represent the median value of the estimated GSV for a given range of GSV from the inventory. The dashed line represents the identity line.

The merging weights differed depending on the test site. The weighting factor for the L-band GSV estimates ranged between nearly 0 (Finland N), 0.23 (Catalonia), 0.54 (Romania) and 0.9 (Finland S). In general, the results obtained with BIOMASAR were on average reliable and only slightly poorer compared to those obtained by fitting the WCM to the reference measurements of GSV (Figure 27). For Catalonia, the weighting was geared towards the C-band dataset, which was correct given the poorer performance of the

retrieval with L-band. For Finland N, the weighting excluded the L-band estimates although the retrieval results were of higher quality. This is a shortcoming of how the normalization procedure for the weight was implemented for this analysis. The set of w_s values were obtained for the four sites only and Finland N was characterized by the lowest of the values resulting in a normalized value of 0. For Finland S, the L-band estimates were preferred, which agrees with the superior performance of the retrieval compared to C-band. Finally, for Romania, the weights were of similar magnitude; however, given the lack of correlation between GSV and backscatter, these results are not relevant. Indeed, using the same model but different calibration approaches led to completely different retrieval results, a consequence of the insensitivity of the backscatter to GSV.

The uncertainties of the SAR and LiDAR measurements and of the individual model parameters quantified by their standard deviations were propagated to obtain an estimate of the GSV's uncertainty for a given SAR observation. For each band, the GSV uncertainty was then expressed as the weighted average of the individual uncertainties and accounted for the temporal correlation of the errors. Finally, the uncertainty of the merged GSV was computed as the weighted average of the band-specific uncertainties. We visualize the uncertainty of the GSV as a function of the estimated GSV in Figure 28 for Catalonia, Finland N and S. The results for Romania are omitted because of the overall very high uncertainty (> 100% of the estimate). The uncertainty at the pixel level was considerable and differed between sites. We attribute this to the sensitivity of the backscatter to GSV, which was more pronounced in Finland N than elsewhere, and to the temporal correlation of errors which was remarkable at all sites.

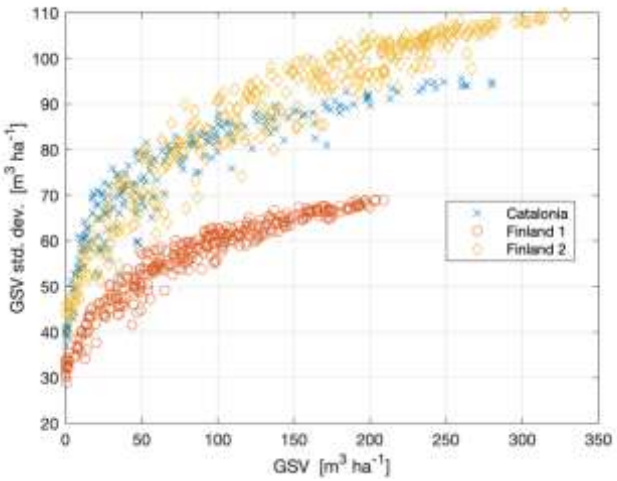

Figure 28. Standard deviation of GSV estimates as a function of the estimated GSV at plot level for the sites of Catalonia, Finland 1 and 2.

BIOMASAR requires a set of auxiliary parameters such as canopy density, maximum GSV etc. In the development phase, the values were mostly a constant and reflected the local conditions of each site. For the implementation on Forestry TEP, where the mass processing was run, the auxiliary information consisted of raster datasets derived from EO data expressing per-pixel values of e.g. canopy density, canopy height, maximum biomass, land cover type etc. Indeed, none of these parameters can be detailed at the spatial resolution of the pan-European dataset (20 m) with in situ measurements. The accuracy of each of those datasets has therefore an impact on the accuracy of the pan-European data product.

Figure 29 shows the comparison of plot-based and map-based values of GSV for the same four sites as above. Compared to the development phase, the results have somewhat lower accuracy. For the two Finnish sites, the layer of maximum GSV based on ICESat-2 heights constrained the retrieval to an interval of GSV values smaller than the real one, causing underestimation at high GSV levels. The somewhat bended relationship between plot- and map-based values was due to imperfections in the allometry relating canopy height and canopy density. This was based on ICESat-1 data which were acquired primarily under leaf-off conditions, thus biasing such model. For Catalonia, the errors were explained as a consequence of inaccurate maximum GSV estimates. The results for Romania were caused by imperfect calibration of the Water Cloud Model when implemented on Forestry TEP. The errors were introduced when selecting “ground” pixels, many of these having been neglected as a consequence of the strong filtering of land cover types implemented in the Forestry TEP processing.

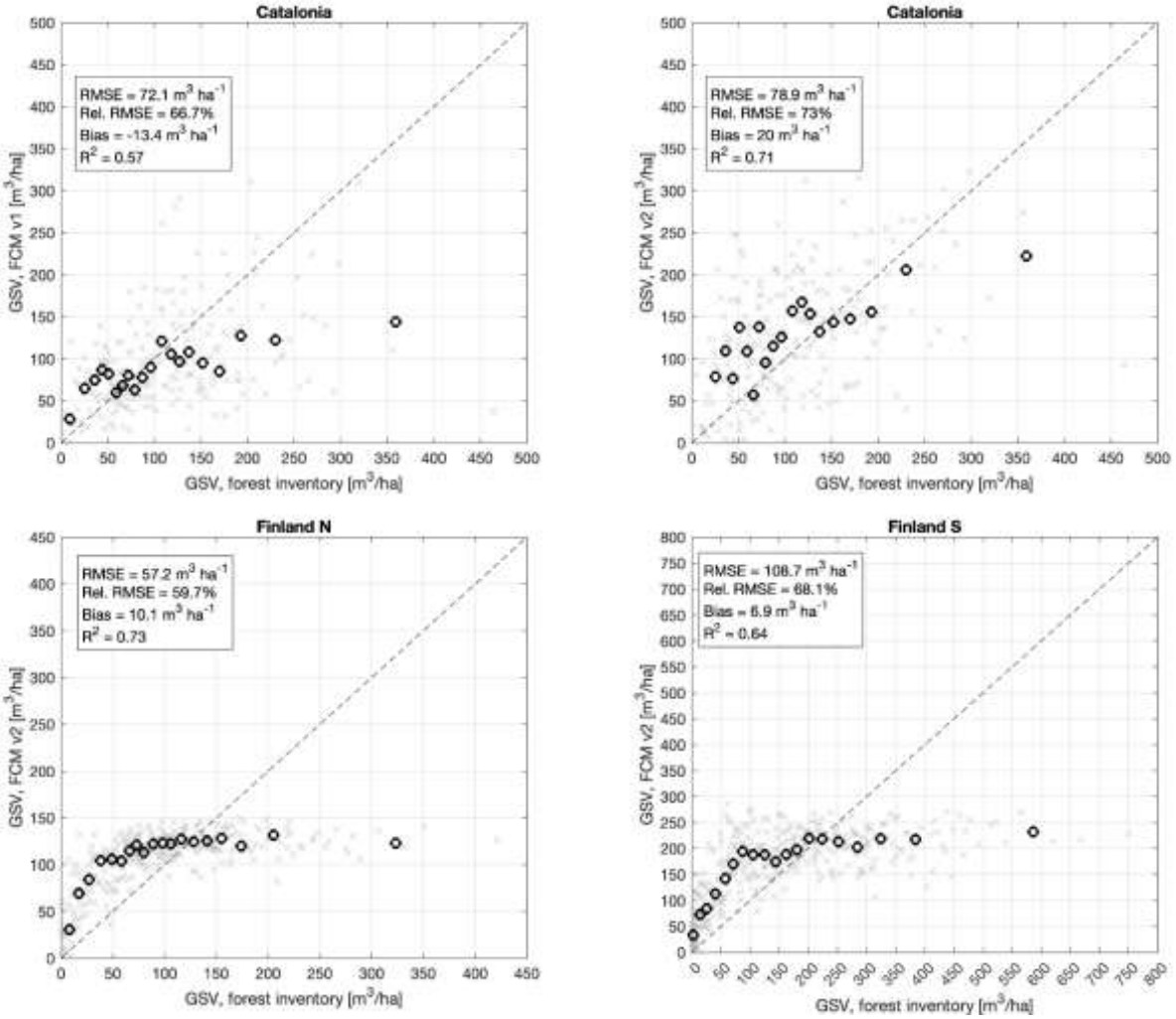

Figure 29. Scatter plots comparing the map-based estimates of GSV from the BIOMASAR approach implemented on FTEP for the pan-European processing with ground refence data.
Crosses illustrate the comparison at the plot level. Circles represent the median value of the estimated GSV for a given range of GSV from the inventory. The dashed line represents the identity line.

The results at plot level indicate that the pan-European processing captured the spatial distribution of GSV (and thereof of AGB and BGB) but had substantial issues at the spatial resolution of the maps. Aggregating the maps to coarse spatial resolution confirmed this indication. Figure 30 shows an example for the Romanian site where plot-based and map-

based average values of GSV are compared at different aggregation levels. Much of the variance affecting the full spatial resolution (Figure 29) disappeared after averaging as also confirmed by the R^2 value, which increased from 0.15 at full resolution to 0.76 at 1 km and close to 1 at coarse resolutions.

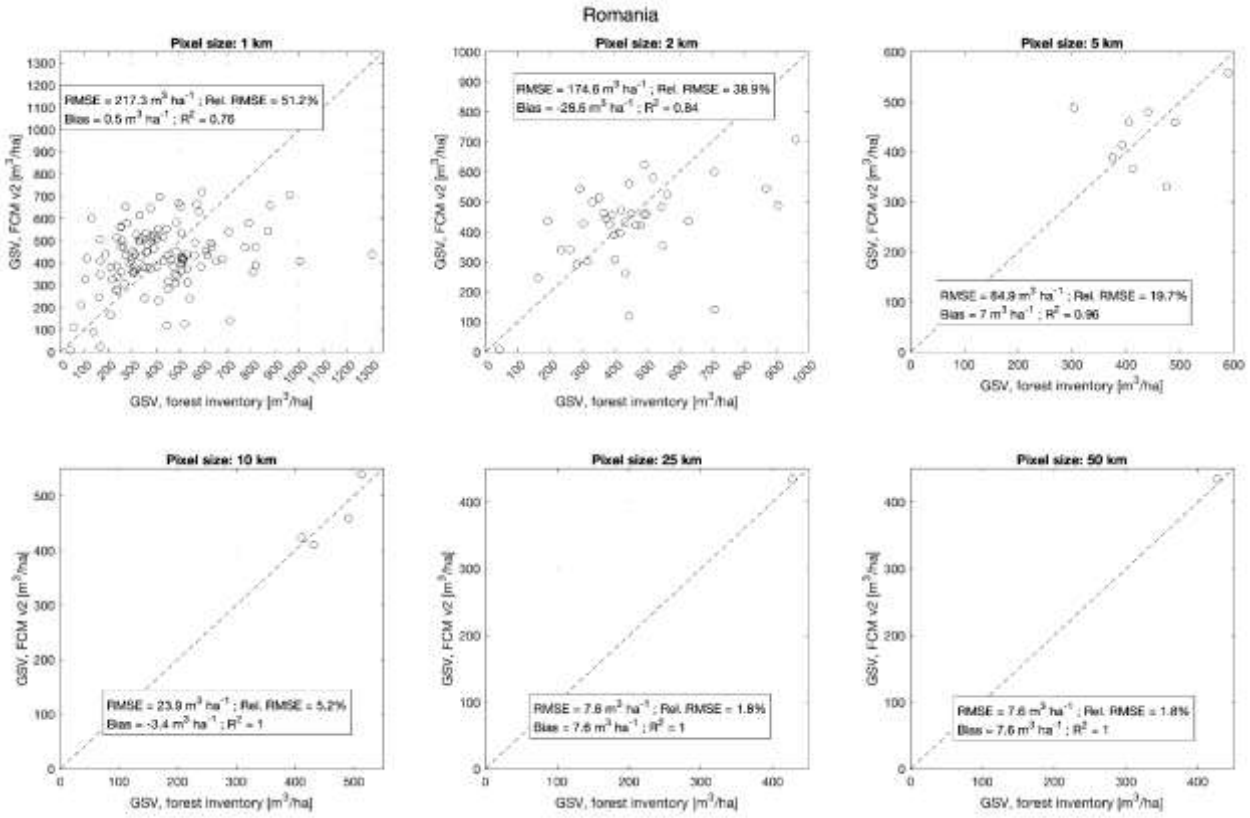

Figure 30. Scatter plots comparing average values of GSV from the pan-European map product and from the dataset of in situ measurements for the site of Romania at different levels of aggregation.

5.5.2.2 Output product accuracy

The previous section provided description of the model calibration results and evaluation of the performance of the model. In addition to this, the resulting European volume and biomass maps were validated by comparing them to published provincial level NFI statistics and by conducting an independent accuracy assessment with a set of NFI plots from countries across Europe (as described in Section 4.6).

The provincial level analysis was conducted by comparing averages from FCM map data values of GSV at the level of individual provinces with values reported by European NFIs in their periodic reports on forest resources (Figure 30). Even if the GSV statistics by the NFIs were not always coincident with the map-based values, this comparison is indicative of the quality of the FCM data product. For Nordic and Mediterranean countries, the agreement was strong. For some countries in Central Europe (Germany, Poland, Czech Republic, Hungary), the map underestimated GSV. Here, we identified an issue with the land cover definition used to select “ground” pixels when calibrating the Water Cloud Model. The processing masked out “pastures” according to the Copernicus High Resolution Layer, which reduced substantially the number of pixels on which the model could be calibrated and biased it. Strong topography caused underestimation in Switzerland. Strong land fragmentation and country-specific forest definitions explained the discrepancies in the United Kingdom, Ireland and the Netherlands.

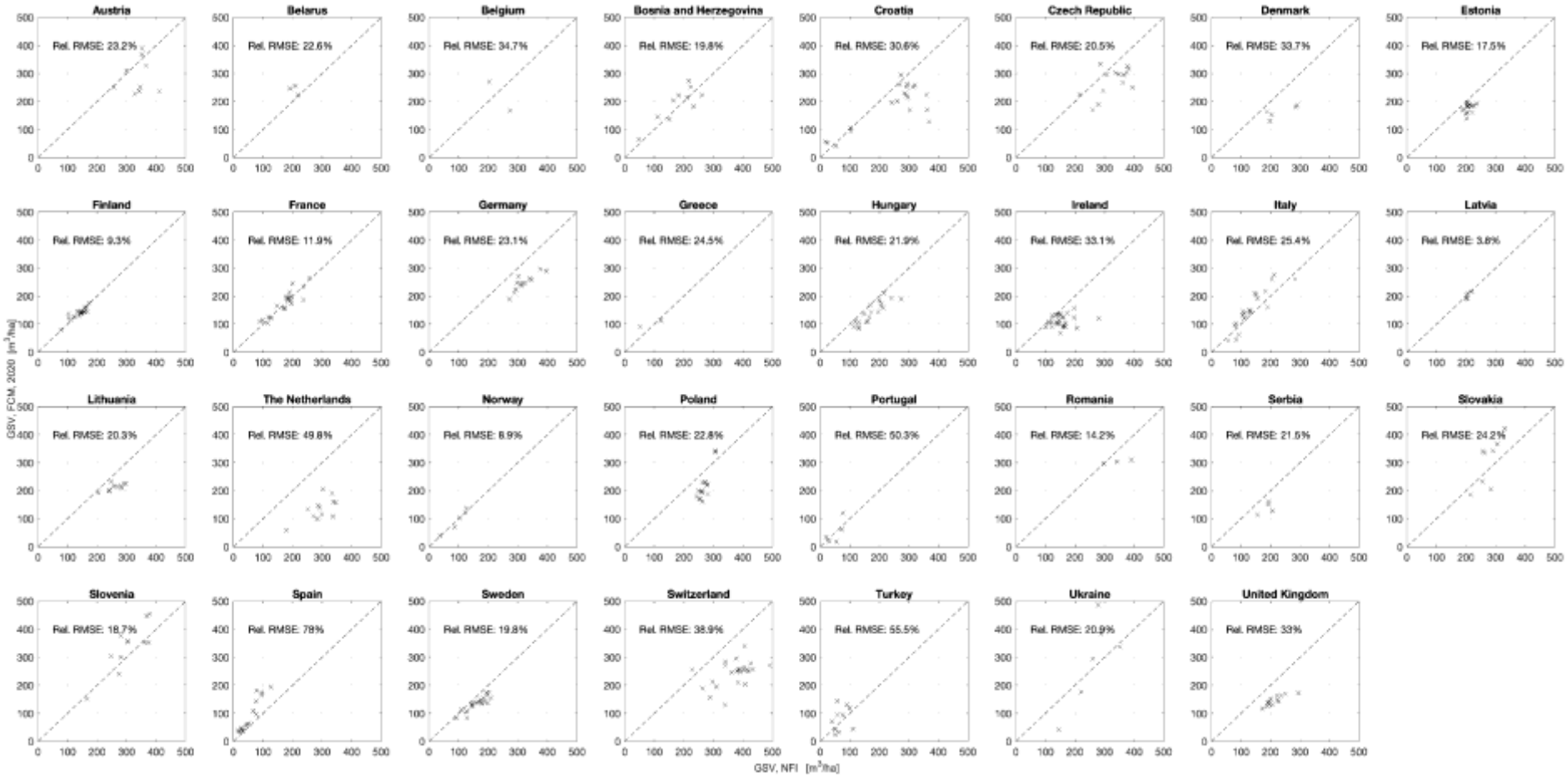

Figure 31. Comparison of GSV averages from this study with values published by European National Forest Inventories.

The independent accuracy assessment with a set of NFI plots was conducted in 10 km aggregates (as described in Section 4.6). Results of the analysis are presented in Figure 32. The two graphs show high consistency between the two annual maps. The patterns of the scatter plots are very similar in both years, indicating inter-annual consistency in the accuracy of maps.

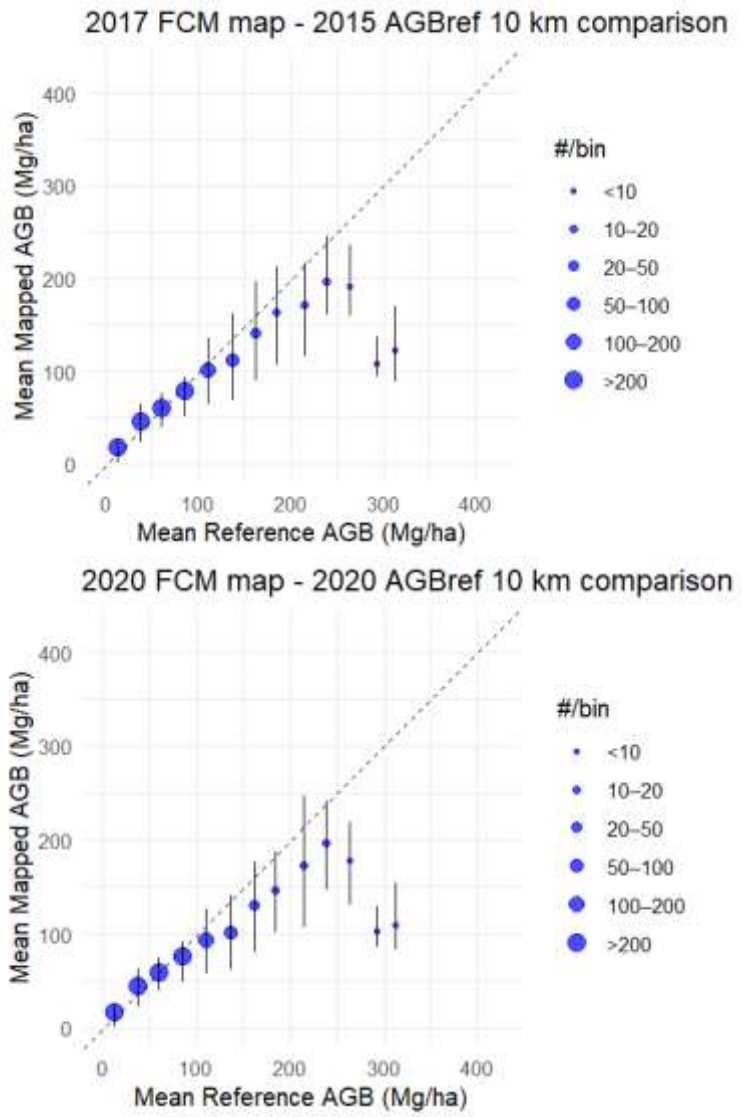

Figure 32. Validation results of the 2017 (left) and 2020 (right) European wide biomass maps.

The reference and mapped values have close agreement up to about 250 Mg/ha for both maps. Systematic underestimation emerges near 300 Mg/ha, driven mainly by high-volume forests in mountainous Croatia which could also reflect conservative masking rules applied during FCM backscatter calibration, as documented earlier in this ATBD. Moreover, the underestimation in high-biomass and temporal noise may be linked to sparse ALOS-2 coverage and remain as target for further improvement.

More detailed analysis of the independent product validation can be found in the product “Delivery note” made available with the maps.

5.6 Autochange

5.6.1 Algorithm description

The Autochange change detection approach (Häme et al. 2020) is a bi-temporal change detection method developed at the VTT Technical Research Centre of Finland. The method provides change intensity and type as outputs. It is assumed that the majority of the image area does not include changes of interest to the user.

The computational flow of the process is illustrated in Figure 33. To reduce time, the process uses a sample of pixels for clustering. The sample consists of mean spectral vectors of groups of pixels representing relatively homogeneous ground targets. The homogeneity is tested by computing the standard deviation vector (alternatively coefficient of variation) of each n x n pixel group in both images. A sample of m groups with lowest non-zero deviations is selected for the clustering. The zero deviation groups are rejected because they can represent a multiplied single pixel due to resampling. Use of the standard deviation as a criterion favours low-reflectance groups that typically represent mature forests. They are over-represented in the sample, which improves the performance of the algorithm in the detection of forest cuts.

Clustering is performed by *k*-means on the selected homogeneous pixel groups (later observations) of the pre-change image to compose *primary clusters* (image representing the state at the start of monitoring). The image spectral bands are standardized before clustering because the k-means algorithm uses the Euclidean distance and is thus sensitive to the dynamic range of a spectral band. The primary clusters are sorted by their 'biomass' index (BM), based on their reflectance of the red band, which has a relatively high correlation with the biomass (NRC 1970).

The primary clusters are used for two purposes: to acquire general information about the land cover before the changes, and to set the initial state for the second level clustering of the post-change image (image representing the state at the end of monitoring).

The observations of the post-change image are labelled with the cluster numbers of the primary clusters without performing a new clustering. The spectral intensity distribution of a primary cluster will consequently become heterogeneous if a change has occurred between the moments of acquisition of the two images. A secondary clustering to n sub-clusters is performed within each labelled primary cluster of the post-change image. This extracts the observations that represent change to specific secondary clusters.

The Euclidean distance between each secondary cluster and its post-change primary cluster is computed as the length of the spectral mean vector and represents the magnitude of change (CM). The CM gets a value of 100 if the Euclidean distance of a secondary cluster to its primary cluster of each spectral band is as large as the standard deviation of the CM band. The benefit of computing the change magnitude from the post-change image only is that the pre- and post-change images do not have to represent identical spectral bands or even do not have to be acquired using the same instrument.

The change type is computed using two spectral features: the biomass index (BM) and the Normalized Difference Vegetation Index (NDVI), producing four change types:

1. BM decrease & NDVI increase: can indicate, e.g. conifer tree removal with remaining broad-leaved trees and grasses

2. BM decrease & NDVI decrease: typical change due to clear-cuts
3. BM increase & NDVI increase: biomass growth
4. BM increase & NDVI decrease: biomass growth with associated decrease in broadleaved trees or shrubs and grasses. Such change can be associated with silvicultural operations on conifer regeneration sites, for instance.

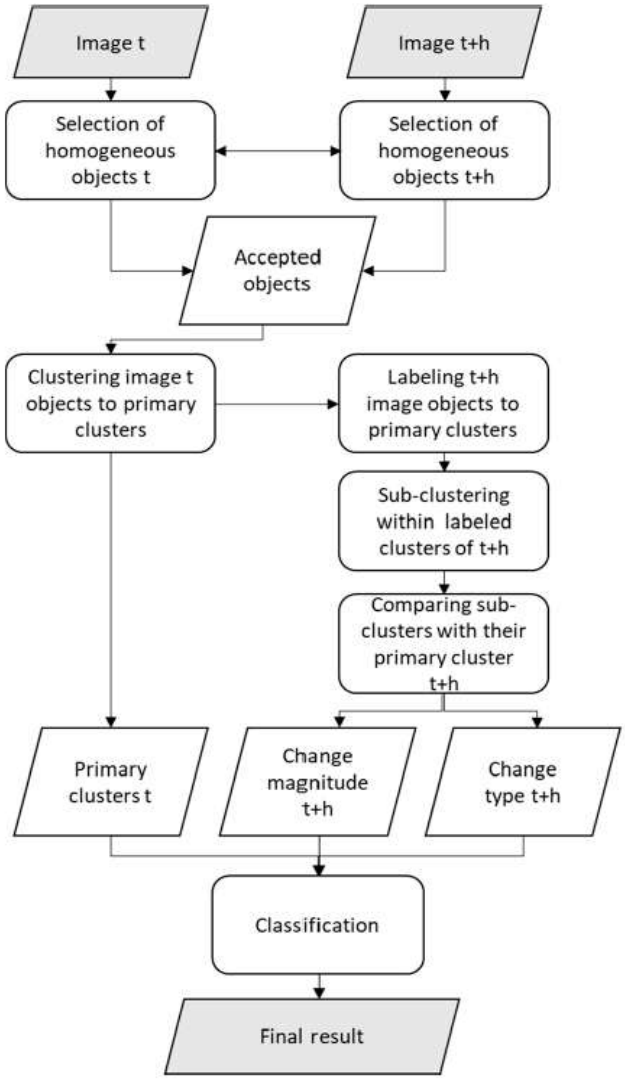

Figure 33. Flowchart of the Autochange change detection method.

The outputs from processing include:

1. Primary cluster mean intensity vector from the pre-change image.
2. Change magnitude, computed using the primary and secondary cluster mean intensity vectors from the post-change image.
3. Change type, computed using the primary and secondary cluster BM and NDVI means from the post-change image.

The final change classification is compiled by applying simple logical operations to the three-band output image. For instance, accepting only primary clusters from the pre-

change image that reflect mature forest, change types that indicate biomass decrease and selecting a threshold for change magnitude produces mapping of clear-cut areas. Alternatively, external forest masks can be used to identify changes in the forest area.

5.6.2 Performance

The use of the Autochange algorithm involves selection of a few key parameters including the 1) image bands to be used, 2) number of primary clusters, 3) number of secondary clusters and 4) sample size. The selection of the parameters directly affect the performance of the algorithm and may differ depending on the goals of the change detection (e.g. the type changes the user is most interested in). Optimal combination of the parameters need to be sought case by case. The following list discussed the main points of the parameter selection affecting the performance of the algorithm.

1. Different types of changes are reflected differently in the bands available in the EO datasets used. For example, in the FCM project it was important to find a *combination of image bands* that best reflect changes (particularly reduction) of biomass, tolerate variation in seasonal and atmospheric properties inside images and are applicable in different geographic regions.

2. *The primary clusters* form homogeneous land cover classes in the pre-change image. If a forest mask applied as a mask in change detection, all primary clusters represent forest types and the number of primary clusters can be reduced. The aim is to use optimal number of clusters to separate different forest types but also to avoid too large number of clusters to enable sufficient number of observations in each cluster.

3. *The secondary clustering* is performed within the primary clusters in the post-change image. The number of the secondary clusters affect the separation of different land cover types in the later image. The larger the number of clusters the smaller the number of observations in each cluster. Very large number of secondary clusters may result with clusters that have only few observations and affect negatively the results. Typically, when aiming at detecting only large magnitude changes like clear cuts, small number of secondary clusters can be used. When aiming at detecting also subtle changes the number of secondary clusters is often increased. Also, for example remaining traces of clouds in the later image affect the secondary clustering, which needs to be considered when selecting the number clusters.

4. It is important that the initial sample for clustering includes observations from the changed areas, although their proportion of the total area is small. Therefore, the parameter defining the *size* of the initial *sample* should ensure large enough number of observations for the clustering. However, too large sample would include observations that represent heterogenous land cover such as border areas between forest and open land.

As already mentioned above, the selection of the key parameters have a strong effect on the performance of the change detection. Therefore, only general indications of the suitability of the algorithm for a given change detection purpose can be given. As an example of the performance of the Autochange algorithm, we present here the parameters and example output products from the European-wide change detection 2020-2021 conducted during the main phase of the FCM project. The parameter set used in the detection is defined in Table 9. The input EO data were the Sentinel-2 composite images presented in Chapter 3.

Table 9. Set of parameters used for European wide application of Autochange.

Parameter	Explanation	Values
hmg_threshold	Number of observations	2000000
limits	Intensity limits for band 10 (quality band) of pre-change image	10 4001 10000
limits2	Intensity limits for band 10 (quality band) of post-change image	10 4001 10000
hmg_size	Sample box dimension of observations in pixels	2
cluster_count	Number of initial clusters	30
cluster_count2	Number of secondary clusters	5
bands	Bands used in the pre-change image	3 5 7 (B04 B08 B12)
bands2	Bands used in the post-change image	3 5 7 (B04 B08 B12)

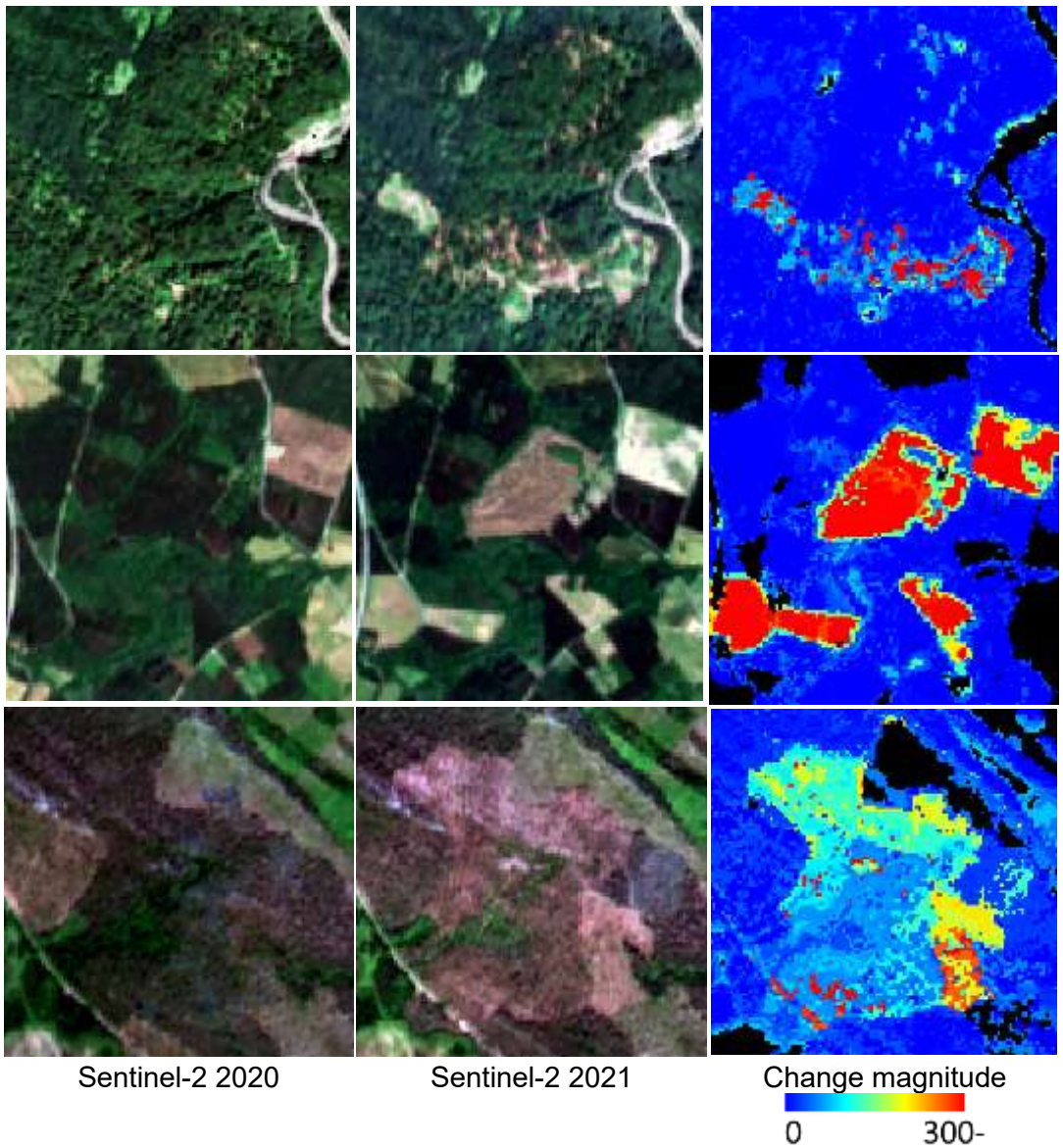

Figure 34. Sentinel-2 true colour composites 2020 and 2021 and corresponding change magnitude for the pixels whose change type indicates biomass decrease.
Results have been computed with the parameters used in the European wide demonstration (Table 9). Upper row in Romania, middle row in Ireland and bottom row in Sweden. Area 1 x 1 km.

Figure 34 shows examples of the change detection results in Romania, Ireland and Northern Sweden using the parameters defined in the Table 9. The selection of the parameters for the European wide mapping needed to be done conservatively, taking into account the variation within the continent. For smaller area analyses, parameters can be chosen more specifically to highlight particular types of changes in the particular conditions of the interest area. Nevertheless, even with the European wide parameters one the CM layers provide a range of different magnitudes, varying from the high values in the clearcut areas to lower values in areas where more subtle changes have taken place.

5.7 PREBAS

5.7.1 Algorithm description

PREBAS is a semi-empirical forest growth simulator (Mäkelä 1997; Valentine and Mäkelä 2005; Peltoniemi et al. 2015; Minunno et al. 2019), to predict forest carbon and water fluxes, current biomass and dimensional growth of even-aged forest stands. PREBAS consists of a daily-time-step module (PRELES) to predict photosynthesis, soil moisture and evapotranspiration, and an annual-time-step module (CROBAS) to allocate the assimilated carbon to respiration and structural growth of biomass components. Intended for large-scale applications in forestry, the model has modest input requirements and feasible runtimes. Parameterized with Bayesian calibration for three boreal species, the model has been demonstrated to perform adequately in country-wide applications (Minunno et al 2019; Holmberg et al. 2019; Minunno et al 2016).

PRELES (PREdict Light-use efficiency, Evapotranspiration and Soil water) predicts photosynthesis or gross primary production (GPP) and evapotranspiration using a light-use-efficiency (LUE) approach linked to soil moisture. PRELES was developed so as to run with standard weather data (Peltoniemi et al. 2015). It calculates photosynthesis using potential LUE and multiplicative modifying factors that depend on the environmental drivers. One of these is soil moisture, which is predicted in PRELES using a simple bucket model that takes precipitation as input and is depleted by evapotranspiration. Evapotranspiration is divided into transpiration by the canopy and evaporation from surfaces and ground (including the ground layer). A similar modelling approach is used as in photosynthesis, where modifying factors reduce the potential evapotranspiration. Transpiration also strongly depends on photosynthesis due to their link through stomatal control. PRELES therefore shows strong interlinkages between photosynthesis, evapotranspiration and soil water (Tian et al. 2020).

CROBAS is an individual tree growth model that can be applied for different stand configurations, climates and sites. Stand configurations are derived from the structural forest variables, and the weather effects are incorporated through impacts on photosynthesis, respiration and tissue longevity. In the FCM project, we use the climate-dependent potential photosynthetic production of a stand, quantified by PRELES, to derive the geographic variation of the other relevant metabolic parameters, following the procedure proposed by Mäkelä et al. (2016). Edaphic site characteristics are described in terms of an aggregated soil fertility parameter that regulates below-ground allocation of carbon.

Total tree growth in CROBAS equals annual net photosynthetic production. Respiration is divided into growth and maintenance components, where maintenance is assumed

proportional to live biomass and growth respiration is a proportion of growth. Total growth is allocated annually to the biomass components that comprise foliage, fine roots and three sapwood fractions, stems, branches and coarse roots. Carbon allocation between wood and foliage is based on the pipe model (Shinozaki et al. 1964) with dynamic crown rise, and allocation between fine roots and foliage assumes that fine-root to foliage ratio depends on nutrient availability, quantified in terms of site type (Valentine et al. 2013).

The PREBAS model is integrated into the forest carbon monitoring platform in such a manner that the forest variable inputs are derived from the forest variable maps produced by the methods described above and the PREBAS model is used to output above ground and below ground biomass. It can also be used to provide a wide range of carbon flux products and forecasts as well.

5.7.2 Performance

As an example of typical level of performance of the PREBAS, we here present calibration results of the model in the Norwegian use case demonstration area. PREBAS model requires daily meteorological inputs for photosynthetically active radiation (PAR), air temperature (TAir), precipitation (P), vapor pressure deficit (VPD), and ambient CO2 concentration. This input was prepared from E-OBS data, an ensemble dataset available at a 0.1-degree grid. The data on global radiation, daily mean temperature, daily precipitation sum, and daily mean relative humidity from E-OBS were used. The PAR was calculated from global radiation, and VPD was calculated using relative humidity and temperature (Allen et al., 1998).

The model was calibrated using Norwegian National Forest Inventory (NFI) data. The dataset consisted of multi-temporal forest stand structure data from 1913 stands. These stands consisted of pine, spruce, and broadleaved trees, with 36% of the stands being mono-specific, 44% with two of the three species, and the remaining 20% being the mixed stands of pine, spruce, and broadleaved trees. The NFI data included repeated measurements of height, diameter at breast height, basal area, height of crown base, volume, and biomass. The NFI data was divided into calibration and validation sets (20% of the stands).

The model was initialized using the first available NFI measurements from 2001. The repeated measurements were assimilated into the model to calibrate the model parameters. Bayesian calibration is a statistical approach used to refine model parameters by integrating prior knowledge with observational data. The prior parameter values were defined from previous calibrations of PREBAS model components (PRELES - Minunno et al., 2016 and CROBAS - Minunno et al., 2019). The parameter ranges were defined based on expert knowledge. The likelihood functions (Sivia et al., 2006) for outputs were defined which were used to update the prior parameters. The posterior distribution was numerically sampled using the Differential Evolution Markov Chain Monte-Carlo algorithm (ter Braak & Vrugt, 2008). The Gelman-Rubin diagnostic (Gelman & Rubin, 1992) was used to test the parameter convergence. When the parameter convergence was achieved, the calibrated set of parameters were used for model simulations for calibration and validation stands. The model performance was evaluated using R^2 and RMSE values, then the mean square errors were compared.

The observed vs. simulated values of forest structural variables (Figure 35) showed improvements in the R^2 and RMSE values when using the calibrated parameters in both the calibration and validation stands.

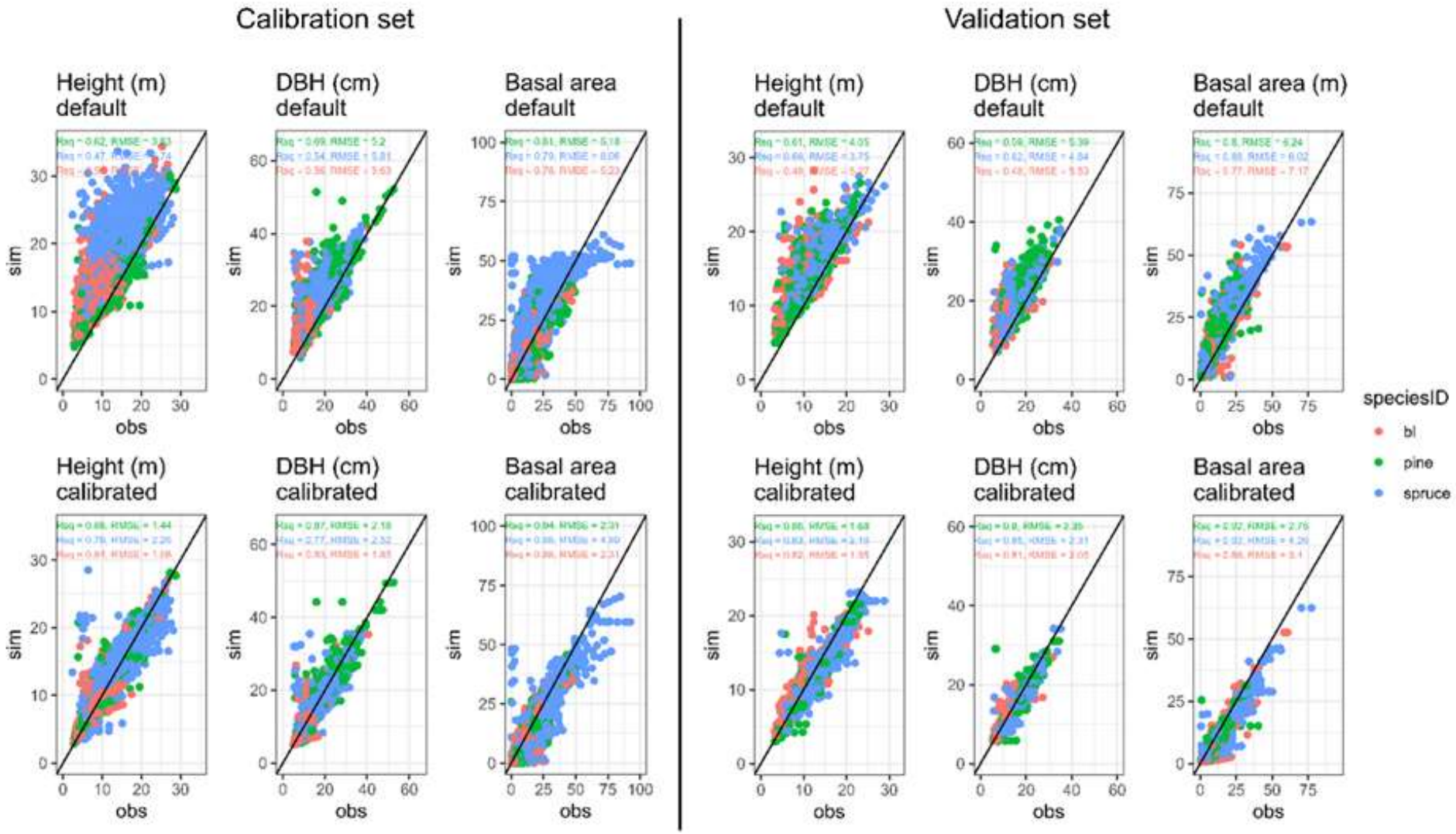

Figure 35. Observed vs. simulated values using the default and the calibrated parameter sets in PREBAS model.

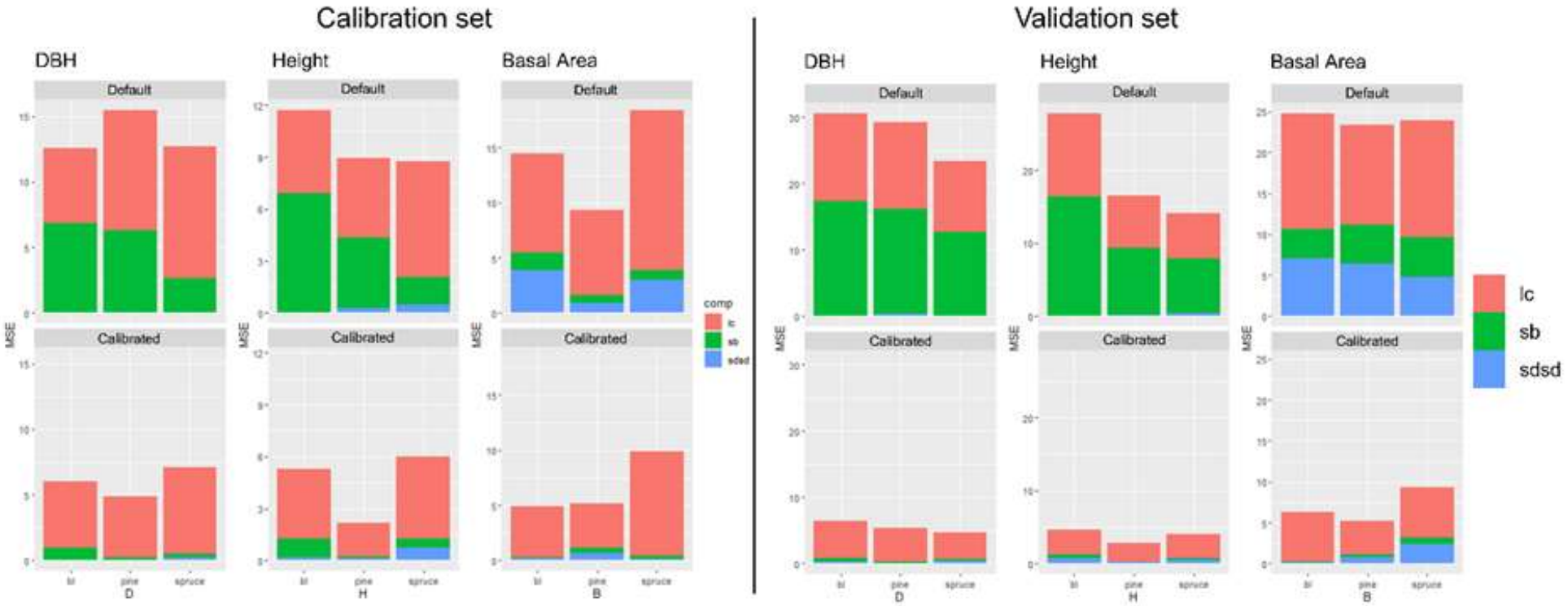

Figure 36. Mean square error and its components calculated from PREBAS simulations using default and calibrated parameters.

The mean square errors were decomposed into three components, representing the average deviation of the simulations from the data (i.e., bias error, sb), if the variability in the data was well-captured by the simulations (i.e., variance error, calculated as the square difference between standard deviations (sdsd)) and the ability of the model to reproduce the pattern of the fluctuations among the data (i.e., phase shift error representing the lack of correlation (lc)) (Kobayashi & Salam, 2000). The simulations using the calibrated model parameters showed overall reductions in MSE (Figure 36). The reduction varied between 50% and 80% of the MSE measured prior to calibration.

5.8 Data assimilation

One of the problematic issues of EO based forest monitoring is the potential inconsistency of the time series of output products on pixel and stand level. For any certification purpose or regulatory monitoring, it is essential that the time series of the results will provide logical and consistent trends corresponding to the actual changes in the target area. To improve the temporal consistency of the predictions, the FCM concept provides the data assimilation (DA) approach. The DA is based on an ensemble Kalman filter that allows to integrate all the available information in the DA framework. By combining the PREBAS process-based ecosystem model predictions and the EO-based predictions with the DA, the time series of output maps provide a more consistent time-series for the users.

The high-level flowchart of the implementation of the DA as demonstrated in the Norway use case is provided in Figure 37. The forest model PREBAS is initialized with the forest structural variables estimated with EO data of the first year (2017). PREBAS is run until the next available EO dataset (2019), at which point the PREBAS and EO-based predictions of the forest structural variables are combined by means of the DA framework to derive a new prediction. The new prediction is used to initialise PREBAS and this cycle is repeated for 2021 and 2023.

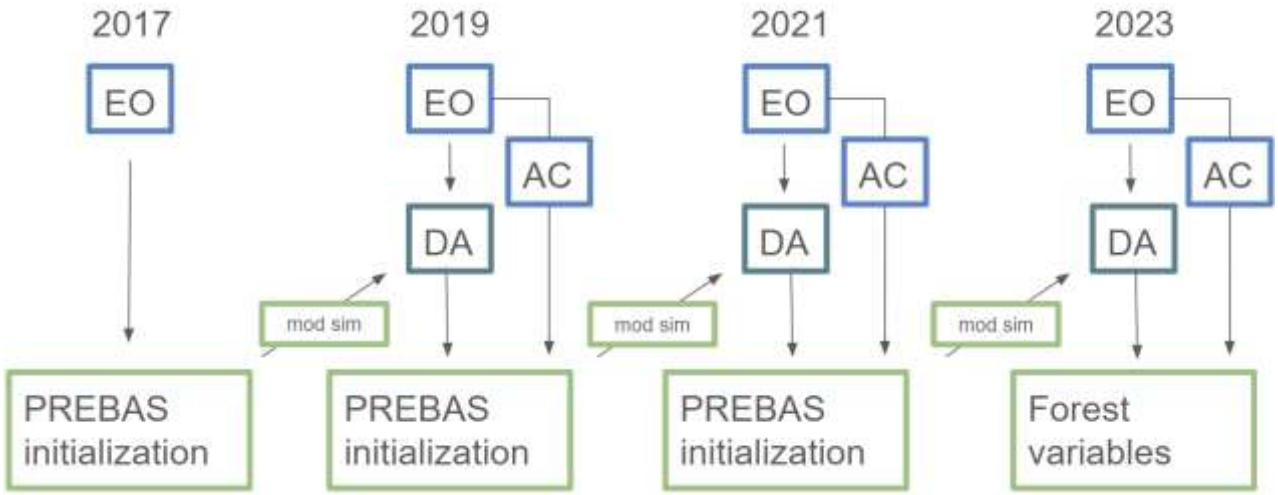

Figure 37. High level illustration of the data assimilation flow chart.

Earth observation predictions of forest structural variables (EO) are used to initialize the forest model PREBAS in case of the first measurements or when changes of forest cover are detected (AC). Otherwise EO are combined with forest model predictions (mod sim) to update the status of forest structural variables (DA).

In DA the different sources of information, in this case forest model predictions and satellite-based predictions, are integrated accounting for their relative uncertainty. Model predictive uncertainty is quantified using Montecarlo simulations, whereas for EO based predictions the pixel based uncertainty estimates are used. Autochange detections are integrated in the analyses. For pixels where a change in forest structure is detected, the DA step is skipped and PREBAS is initialized directly with the most recent EO-based predictions. For the pixels covered by clouds, data assimilation is bypassed, and model predictions are utilized to forecast forest growth until the next available S2 data.

The data assimilation framework consists of five steps (Figure 38). The description here is derived mainly from Minunno et al. (2025). Steps 2 to 5 are implemented at pixel level:

1. Emulator calibration
2. Monte Carlo simulations for the uncertainty quantification of initial state
3. Forecast step
4. Data assimilation
5. Map production

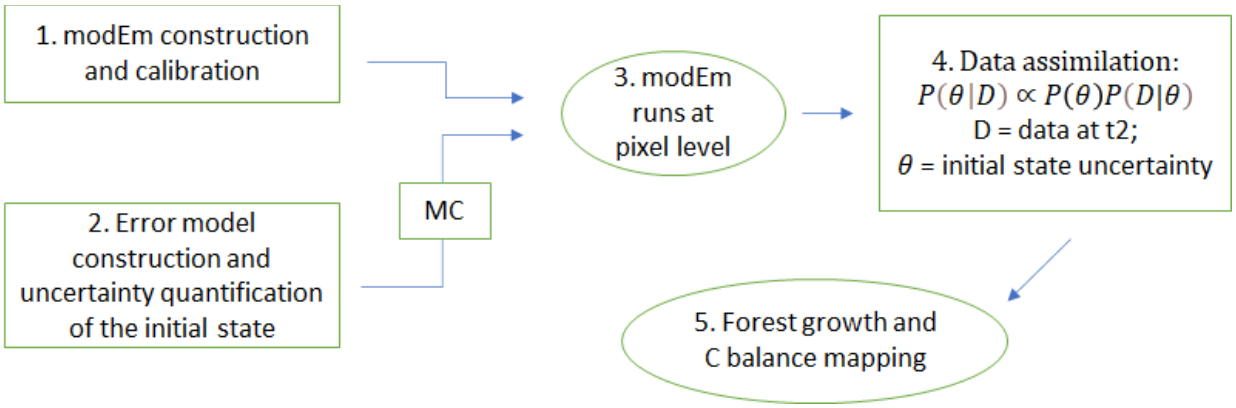

Figure 38. Flowchart for the data assimilation framework of forest structural variables. P(θ|D): posterior probability distribution of forest structural variables integrating data between EO-based and modelled predictions; P(θ): prior distribution, given by the initial state uncertainty of Sentinel-2 predictions for the first year propagated to t2 by means of model predictions; P(D|θ): likelihood function calculated using forest structural variables calculated by model predictions and Sentinel-2 predictions for t2.

1. In step 1, the emulators are fitted using PREBAS outputs. A large number of pixels (e.g. around 20 000 for one Sentinel-2 image) are randomly extracted. PREBAS is initialized with EO-based predictions and the outputs (e.g. B, D, H, species coverage) are modelled for t2 to fit the emulators.
2. In step 2, for each pixel, 1000 samples of interest variables (e.g. B, D, H and species coverage) are drawn, using a multivariate normal distribution fitted on the basis of EO-based predictions and field measurements.
3. In step 3, the emulator is run 1000 times for each pixel using the inputs generated in step 2. By means of the emulator runs the forest structural variables are computed at t2 with the associated uncertainty.

4. In step 4, the emulator forecasts are combined with satellite-based predictions at t2 using the Bayesian approach. The forecasts are used to construct the prior distribution for the forest structural variables and the new predictions at t2 are encoded in the likelihood defined in the accuracy assessment analysis. The posterior is calculated using the Kalman filter as an analytical solution to compute the moments of forest structural variables.
5. In step 5, maps of carbon balance and forest growth and their relative uncertainties are generated using the maximum a posteriori (MAP) predictions and their relative uncertainty expressed by standard deviation.

The idea is that the data assimilation approach allows creation and updating of consistent time series of forest variable and carbon flux predictions. The monitoring system could be be continuously updating with increasing accuracy, as new observations (either EO-based on from other sources) are obtained.

6. Conclusion

The extensive testing and use case demonstrations conducted during the project emphasize the effects of the conditions and goals of the use case on the selection of the most suitable tool. None of the tools can be said to be the best option overall, but the most optimal tool (or a combination of tools) need to be chosen case-by-case taking into account the available datasets, the characteristics of the area of interest and the objectives of the use case.

Empirical methods with field data measurements from the area of interest provide more flexibility to meet varying user needs (e.g., related to the required forest variables) and maximizing the potential of available datasets. Typically, highest accuracies could be reached in the test areas with approaches combining local field reference data with a fusion of optical and radar satellite data.

On the other hand, a method like BIOMASAR enables production of large area forest volume and biomass maps with consistent methodology and without field data from the area of interest. The reliance on radar data time series also decreases impact of weather variations in the results. With a range of methods from semi-automated empirical to fully automated physical methods, the FCM toolbox can respond to different user requests with varying interest variables, data availability and the size of geographic interest area.

The availability of the EO and reference datasets for the interest area is the most important factor effecting the level of accuracy that can be reached in the mapping. The selection of EO datasets used in a particular use case depends both on the physical availability of datasets as well as the willingness of the user to purchase EO data in addition to freely available datasets. The algorithm to be used in the mapping should be selected based on the available EO and reference data, as well as the objectives of the use case.

Although the selection of optimal algorithm to be used depends on several practical and scientific aspects and has to be evaluated case-by-case, some general guidance can be given to users:

1. *Availability of field reference data:* The type and availability of field reference data is one of the most important deciding factors when choosing the most suitable algorithm. If there is no field data available from the user or other sources to train the model for the area of interest, there are two options: to apply blindly a model trained in another area or to use the BIOMASAR approach. The choice depends largely on the variables of interest and the required spatial and temporal detail, as well as the size of the interest area. The former choice is feasible if reference data or models from similar ecological conditions are available, and particularly if a small number of field plots can be measured for model finetuning purposes. The latter approach is particularly suitable for large area mapping up to continental and global levels. Four years of growing stock volume and biomass maps (2017, 2020, 2021 and 2023) of European wide biomass maps are available on the FCM product portal.

2. *Amount of available reference data:* A second important aspect is the amount of available reference data. In the case of extensive wall-to-wall reference data layers or high number of field plots, the choice of algorithm can be based on practical preferences or comparative tests. If the number of plots is limited (<100), and/or its representativeness cannot be guaranteed, the Probability method or UNet model transfers are the most recommended options. The benefit of the semi-automated

Probability approach is that it allows visual evaluation and manual fine-tuning of the model. This becomes a vital asset in situations where the field is not representative and needs to be supplemented by visual evaluation of the interest area in combination with the unsupervised clustering (the first step of the Probability process). Another option is the transfer of UNet model into the target area by finetuning the model with the limited number of reference data.

3. *Number of interest variables:* The benefit of the k-NN and Probability algorithms is that they can produce multivariate predictions, meaning that multiple variables are predicted in one process using the same reference data, thereby retaining the relationships between the variables (such as diameter, basal area, height and biomass). Other algorithms, such as the UNet, produce predictions for single variables. This may in some cases result in unrealistic combinations of forest variable predictions, which in turn affects the process-based ecosystem modelling. The user needs to consider the main goals use case and the trade-offs between the single and multivariate algorithms.

It is also important to note that no specific number of plots can be used as the threshold for “sufficient” number of plots. It strongly depends on the variability of the forest area and the range of different characteristics of the area included in the plots. Typically, at least 100 plots are required for reliable k-NN implementation and preferably at least 50 for UNet model finetuning. But the characteristics and representability of the field dataset needs to be evaluated case-by-case.

Overall, the selection of the optimal algorithm (or an existing product) is often based mainly on practical aspects on reference data and EO data availability. The FCM tools offer a range of options to choose from. This document has provided the scientific basis for the tools with examples of the typical levels of uncertainty that can be reached. In addition, key issues to be considered in the selection of the algorithm have been discussed. Further information and examples of use cases can also be found in the Forest Carbon Monitoring website. We recommend anyone interested in using the FCM tools to contact the FCM team with very low threshold through the website (https://www.forestcarbonplatform.org/) or through the Forestry TEP platform (https://f-tep.com/). We will be happy to advice on the options bilaterally considering the details of each individual use case.

References

Allen, RG, Pereira, LS, Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome , 300 (9), D05109.

Alt, H. (2001) The Nearest Neighbor. In: Alt, H. (eds) Computational Discrete Mathematics. Lecture Notes in Computer Science, vol 2122. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45506-X_2

Antropov, O., Rauste, Y., Häme, T., & Praks, J. (2017). Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. Remote Sensing, 9.

Araza, A., de Bruin, S., Herold, M., Quegan, S., Labriere, N., Rodriguez-Veiga, P., Avitabile, V., Santoro, M., Mitchard, E.T.A., Ryan, C.M., Phillips, O.L., Willcock, S., Verbeeck, H., Carreiras J., Hein, L., Schelhaas, M.-J., Pacheco-Pascagaza, A.M., da Conceição Bispo, P., Vaglio Laurin, G., Vieilledent, G., Slik, F., Wijaya, A., Lewis, S.L., Morel, A., Liang, J., Sukhdeo, H., Schepaschenko, D., Cavlovic, J., Gilani, H. and Lucas, R. (2022) A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. Remote Sensing of Environment 272: 112917.

Askne, J., Dammert, P.B.G., Ulander, L.M.H., Smith, G., 1997. C-band repeat-pass interferometric SAR observations of the forest. IEEE Transactions on Geoscience and Remote Sensing 35, 25–35.

Astola, H., Häme, T., Sirro, L., Molinier, M. and 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.

Attema, E. P. W., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. Radio Science, 13 , 357-364.

Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., Ashton, P.S., Banin, L., Bayol, N., Berry, N.J., Boeckx, P., de Jong, B.H.J., DeVries, B., Girardin, C.A.J., Kearsley, E., Lindsell, J.A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E.T.A., Nagy, L., Qie, L., Quinones, M.J., Ryan, C.M., Ferry, S.J.W., Sunderland, T., Laurin, G.V., Gatti, R.C., Valentini, R., Verbeeck, H., Wijaya, A., Willcock, S., 2016. An integrated pan-tropical biomass map using multiple reference datasets. Global Change Biology 22, 1406–1420.

Cartus et al. 2012. Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. Remote Sensing of Environment 124, 466–478. https://doi.org/10.1016/j.rse.2012.05.029

Cartus, O., Santoro, M., Wegmuller, U., & Rommen, B. (2019). Benchmarking the retrieval of biomass in boreal forests using p-band sar backscatter with multi-temporal c- and l-band observations. Remote Sensing, 11.

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

Cochran, W.G. (1977). Sampling techniques (3rd ed.). New York: John Wiley & Sons.

Dinerstein, E., Olson, D., Joshi, A., et al. 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. BioScience 67, 534–545. https://doi.org/10.1093/biosci/bix014

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Science of Remote Sensing 100002.

Frey, O., Santoro, M., Werner, C.L., Wegmuller, U., 2013. DEM-Based SAR Pixel-Area Estimation for Enhanced Geocoding Refinement and Radiometric Normalization. IEEE Geosci. Remote Sensing Lett. 10, 48–52.

Ge, S., Antropov, O., Häme, T., McRoberts, R.E. & Miettinen, J. (2023a) Deep Learning Model Transfer in Forest Mapping Using Multi-Source Satellite SAR and Optical Images. Remote Sensing 15: 5152. doi: 10.3390/rs15215152

Ge, S., Gu, H., Su, W., Lönnqvist, A., & Antropov, O. (2023b) A Novel Semisupervised Contrastive Regression Framework for Forest Inventory Mapping with Multisensor Satellite Data. *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023

Ge, S., Gu, H., Su, W., Praks, J. & Antropov, O., Improved Semisupervised UNet Deep Learning Model for Forest Height Mapping with Satellite SAR and Optical Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 5776-5787, 2022.

Häme, T., Sirro, L., Kilpi, J., Seitsonen, L., Andersson, K. and Melkas, T. (2020) A hierarchical clustering method for land cover change detection and identification. Remote Sensing 12: 1751.

Häme, T., Stenberg, P., Andersson, K., Rauste, Y., Kennedy, P., Folving, S., & Sarkeala, J. 2001. AVHRR-based forest proportion map of the Pan-European area. Remote Sensing of Environment, 77(1), 76-91.

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., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21-st century forest cover change. Science 342, 850–853.

Holmberg, M., T. Aalto, A. Akujärvi, A.N. Arslan, I. Bergstrom, K. Bottcher, I. Lahtinen. 2019. “Ecosystem Services Related to Carbon Cycling - Modeling Present and Future Impacts in Boreal Forests.” *Frontiers in plant science* 10: 343-351. doi: 10.3389/fpls.2019.00343.

Kay, H., Santoro, M., Cartus, O., Bunting, P., Lucas, R., 2021. Exploring the Relationship between Forest Canopy Height and Canopy Density from Spaceborne LiDAR Observations. Remote Sensing 13, 4961.

Kobayashi, K., & Salam, M. U. (2000). Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. *Agronomy Journal*, *92*(2), 345–352.

Kurvonen, L., Pulliainen, J., & Hallikainen, M. (1999). Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 SAR images. IEEE Transactions on Geoscience and Remote Sensing, 37 , 198{205.

Los, S.O., Rosette, J.A.B., Kljun, N., North, P.R.J., Chasmer, L., Suárez, J.C., Hopkinson, C., Hill, R.A., Van Gorsel, E., Mahoney, C., Berni, J.A.J., 2012. Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. Geoscientific Model Development 5, 413–432.

Mäkelä, A. 1997. “A carbon balance model of growth and self-pruning in trees based on structural relationships.” *Forest Science* 43: 7-24. doi: 10.1093/forestscience/43.1.7.

Mäkelä, A., M. Pulkkinen, and H. Mäkinen. 2016. “Bridging empirical and carbon-balance based forest site productivity - significance of below-ground allocation.” *Forest Ecology and Management* 372: 64-77. doi: 10.1016/j.foreco.2016.03.059.

Mäkisara et al. 2019. The multi-source national forest inventory of Finland methods and results 2015. Technical Report 8/2019, Natural Resources Institute Finland (Luke). Natural resources and bioeconomy studies. http://urn.fi/URN:ISBN:978-952-326-711-4

Miettinen, J., Carlier, S., Häme, L., Mäkelä, A., Minunno, F., Penttilä, J., Pisl, J., Rasinmäki, J., Rauste, Y., Seitsonen, L, Tian X. and Häme, T. (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: 9492-514. doi: 10.1080/01431161.2021.1998715

Minunno, F., Miettinen, J., Tian, X., Häme, T., Holder, J., Koivu, K. and Mäkelä, A. (2025) Data assimilation of forest status using Sentinel-2 data and a process-based model. Agricultural and Forest Meteorology 363: 110436. DOI: 10.1016/j.agrformet.2025.110436

Minunno, F., M. Peltoniemi, S. Härkönen, T. Kalliokoski, H. Mäkinen, and A. Mäkelä. 2019. “A. Bayesian calibration of a carbon balance model PREBAS using data from permanent growth experiments and national forest inventory.” *Forest Ecology and Management* 440: 208-257.

Minunno, F., M. Peltoniemi, S. Launiainen, M. Aurela, I. Mammarella, A. Lindroth, A. Lohela, M. Minkkinen, and A. Mäkelä. 2016. “Calibration and validation of a semi-empirical flux ecosystem model for coniferous forests in the Boreal region.” *Ecological Modelling* 341: 37-52. doi: 10.1016/j.ecolmodel.2016.09.020.

Mitchard et al. 2014. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. Global Ecology and Biogeography 23, 935–946. https://doi.org/10.1111/geb.12168

Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. Remote Sensing of Environment 221, 247–259.

NRC (1970). National Research Council. *Remote Sensing with Special Reference to Agriculture and Forestry*; The National. Academies Press: Cambridge, MA, USA. ISBN 978-0-309-01723-7.

Olesk, A.; Praks, J.; Antropov, O.; Zalite, K.; Arumäe, T.; Voormansik, K. Interferometric SAR coherence models for characterization of hemiboreal forests using TanDEM-X data. Remote Sens. 2016, 8, 700.

Peltoniemi, M.S., T. Markkanen, S. Härkönen, P. Muukkonen, T. Thum, T. Aalto, and A. Mäkelä. 2015. “Consistent estimates of gross primary production of Finnish forests --- comparison of two process models.” *Boreal Environment Research* 20: 196–212.

Praks, J.; Antropov, O.; Hallikainen, M.T. LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X-and L-band. IEEE Trans. Geosci. Remote Sens. 2012, 50, 3831–3843

Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham.

Santoro et al. 2015. Forest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from Envisat ASAR. Remote Sens. Env. 168, 316–334. https://doi.org/10.1016/j.rse.2015.07.005

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmuller, U., & Wiesmann, A. (2011). Retrieval of growing stock volume in boreal forest using hyper-temporal series of envisat asar scansar backscatter measurements. Remote Sensing of Environment, 115 ,1040 490-507.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabilie, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., Cazzolla Gatti, R., da Conceição Bispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E.T.A., Morel, A., Pacheco Pascagaza, A.M., Ryan, C.M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A., Willcock, S., 2021a. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth System Science Data 13, 3927–3950.

Santoro, M., Cartus, O., Fransson, J.E.S., 2021b. Integration of allometric equations in the water cloud model towards an improved retrieval of forest stem volume with L-band SAR data in Sweden. Remote Sensing of Environment 253, 112235.

Santoro, M., Cartus, O., Wegmüller, U., Besnard, S., Carvalhais, N., Araza, A., Herold, M., Liang, J., Cavlovic, J., Engdahl, M.E., 2022. Global estimation of above-ground biomass from spaceborne C-band scatterometer observations aided by LiDAR metrics of vegetation structure. Remote Sensing of Environment 279, 113114.

Santoro, M., Cartus, O., Quegan, S., Kay, H., Lucas, R.M., Araza, A., Herold, M., Labrière, N., Chave, J., Rosenqvist, Å., Tadono, T., Kobayashi, K., Kellndorfer, J., Avitabile, V., Brown, H., Carreiras, J., Campbell, M.J., Cavlovic, J., Bispo, P.D.C., Gilani, H., Khan, M.L., Kumar, A., Lewis, S.L., Liang, J., Mitchard, E.T.A., Pacheco-Pascagaza, A.M., Phillips, O.L., Ryan, C.M., Saikia, P., Schepaschenko, D., Sukhdeo, H., Verbeeck, H., Vieilledent, G., Wijaya, A., Willcock, S., Seifert, F.M., 2024a. Design and performance of the Climate Change Initiative Biomass global retrieval algorithm. Science of Remote Sensing 10, 100169. https://doi.org/10.1016/j.srs.2024.100169

Santoro, M., Cartus, O., Antropov, O., Miettinen, J. (2024b) The Estimation of the Growing Stock Volume of Forests with Synthetic Aperture Radar-Based Models: A Comparison of Model-Fitting Methods. Remote Sensing 16: 4079. doi: 10.3390/rs16214079

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007-2010). Remote Sensing of Environment 155, 13–31.

Shimada, M., Ohtaki, T., 2010. Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 3, 637–656.

Shinozaki, K., K. Yoda, K. Hozumi, and T. Kira. 1964. “A quantitative analysis of plant form—the pipe model theory. I. Basic analyses.” *Japanese Journal of Ecology* 14: 97–105. doi: 10.18960/seitai.14.3_97.

Simard, M., Pinto, N., Fisher, J.B., Baccini, A., 2011. Mapping forest canopy height globally with spaceborne LiDAR. Journal of Geophysical Research 116.

Sirro, L., Häme, T., Rauste, Y., Kilpi, J., Hämäläinen, J., Gunia, K., De Jong, B., Paz Pellat, F. 2018. Potential of Different Optical and SAR Data in Forest and Land Cover Classification to Support REDD+ MRV. Remote Sensing 10: 942.

Sivia, D., Skilling, J., Sivia, D., & Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial* (Second Edition, Second Edition). Oxford University Press.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14 , 199-222.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S.R., Schmullius, C., 2014. Carbon stock and density of northern boreal and temperate forests. Global Ecology and Biogeography 23, 297–310.

Teijido-Murias, I., Antropov, O., López-Sánchez, C.A., Barrio-Anta, M. and Miettinen, J. (2025) Forest Height and Volume Mapping in Northern Spain with Multi-Source Earth Observation Data: Method and Data Comparison. Forests 16: 563. doi: 10.3390/f16040563

ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435–446. https://doi.org/10.1007/s11222-008-9104-9 GFOI (2014). Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative. Group on Earth Observations, Geneva, Switzerland.

Tian, X.L., F. Minunno, T.J. Cao, M. Peltoniemi, T. Kalliokoski, and A. Mäkelä. 2020. “Extending the range of applicability of the semi-empirical ecosystem flux model PRELES for varying forest types and climate.” *Global Change Biology* 26: 2923-2943. doi: 10.1111/gcb.14992.

Valentine, H.T. and A. Mäkelä. 2005. “Bridging process-based and empirical approaches to modeling tree growth.” *Tree Physiology* 25: 769–779. doi: 10.1093/treephys/25.7.769.

Valentine, H.T., R.L. Amateis, H. Gove, and A. Mäkelä. 2013. “Crown rise and crown-length dynamics: application to loblolly pine.” *Forestry* 86: 371-375. doi: 10.1093/forestry/cpt007.