Forest Carbon Monitoring

ESA Forest Carbon Monitoring: *Evaluation of various Earth Observation datasets and methods*

POLINSAR & BIOMASS 2023 Workshop

Oleg Antropov, *Jukka Miettinen*, *Tuomas Häme*, *Lauri Seitsonen*, VTT, Finland

Ronald E. McRoberts, University of Minnesota, USA

Maurizio Santoro, *Oliver Cartus*, Gamma Remote Sensing, Switzerland

Shaojia Ge, Nanjing University of S&T, China

Matteo Pardini, *Kostas Papathanassiou*, *Irena Hajnsek*, DLR, Germany

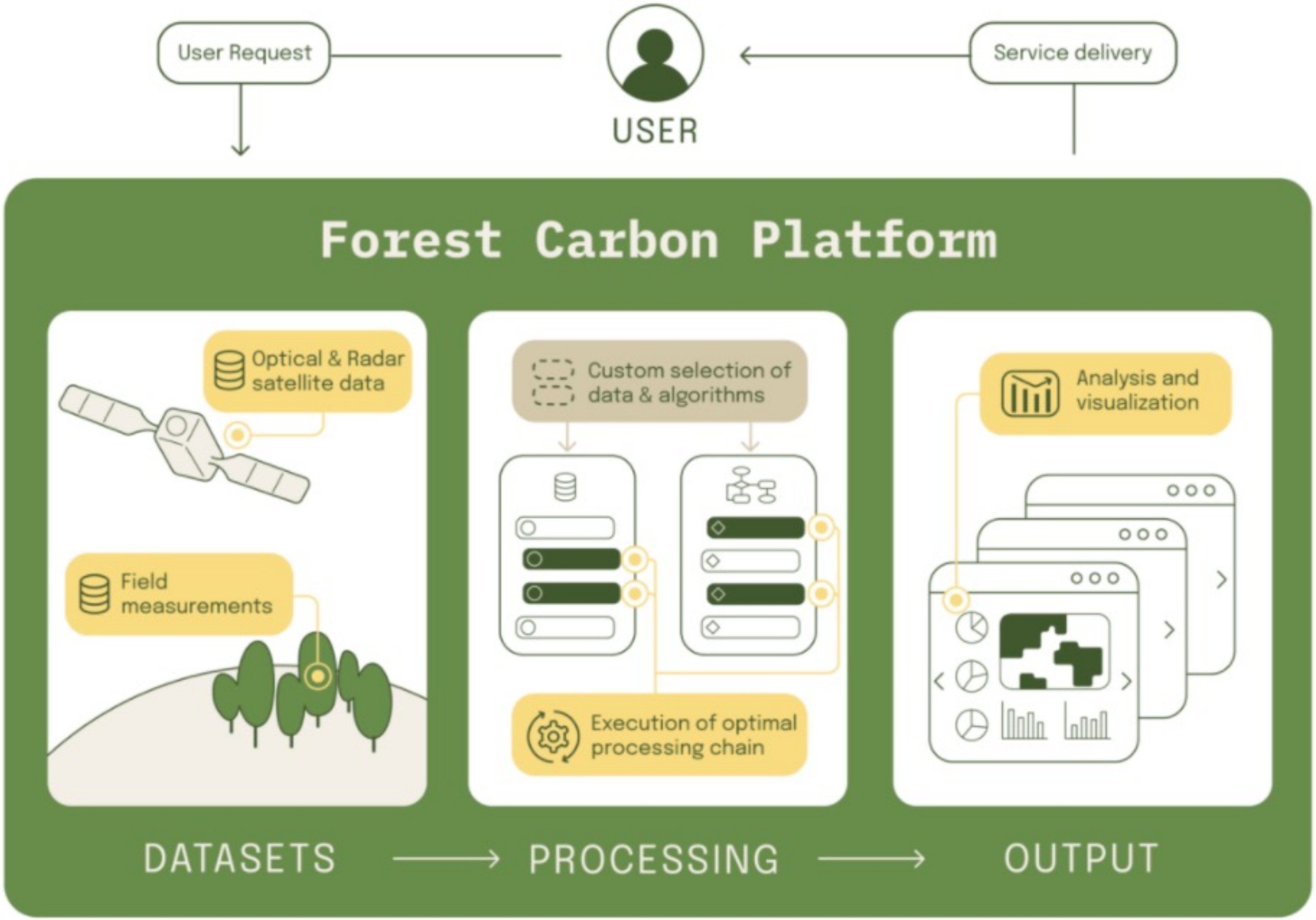
Outline

- Forest Carbon Monitoring project: overview
- EO based forest inventory and carbon estimation pathways
- Data & Methods evaluation
    - Study sites and modelling approaches
    - Selection of results on forest variable prediction and feature selection
    - General conclusions on role of sensor data and methods
    - Deep learning model transfer example

Objective

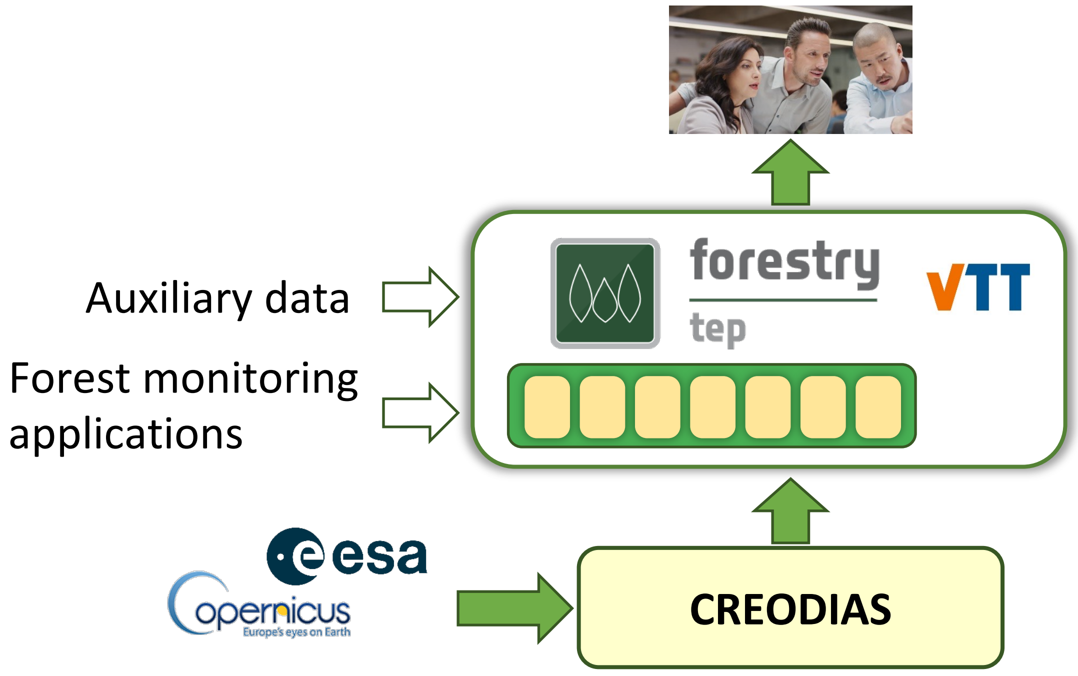
Key features

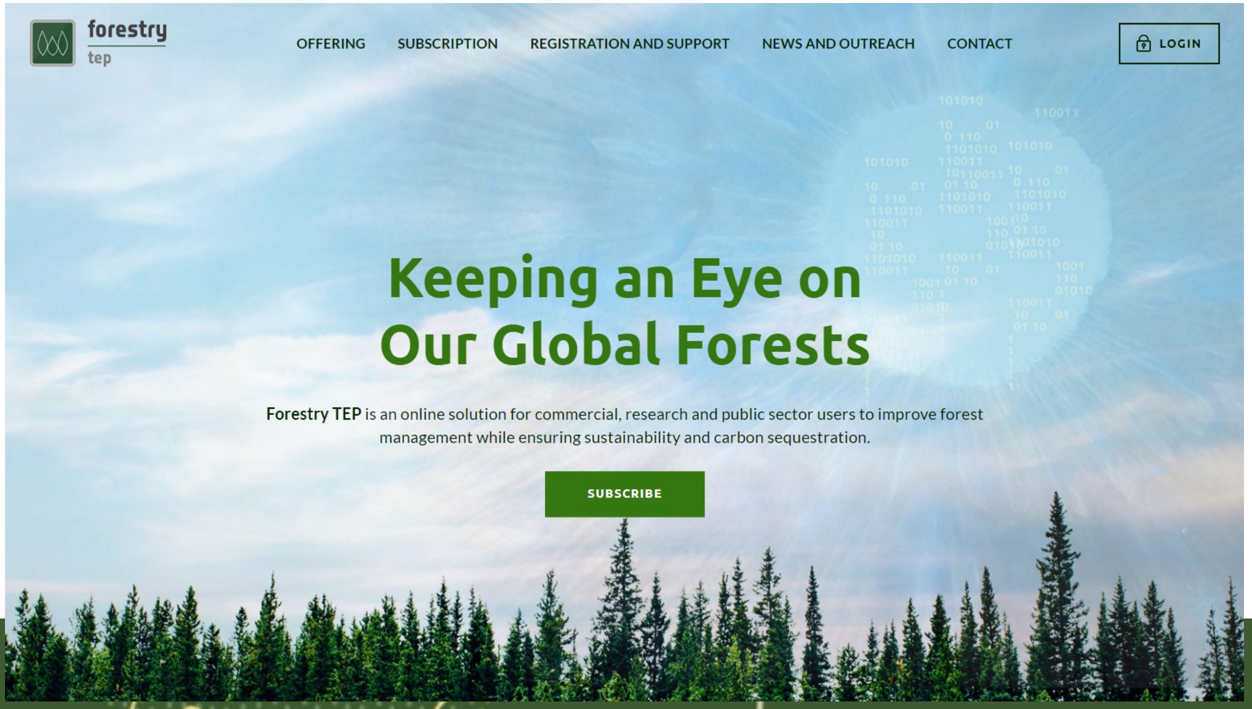
1. Close integration of in-situ and remotely sensed data.
2. Process-based forest ecosystem carbon modelling integrated into the system.
3. Flexibility to user needs ranging from private company area monitoring to continental analyses.



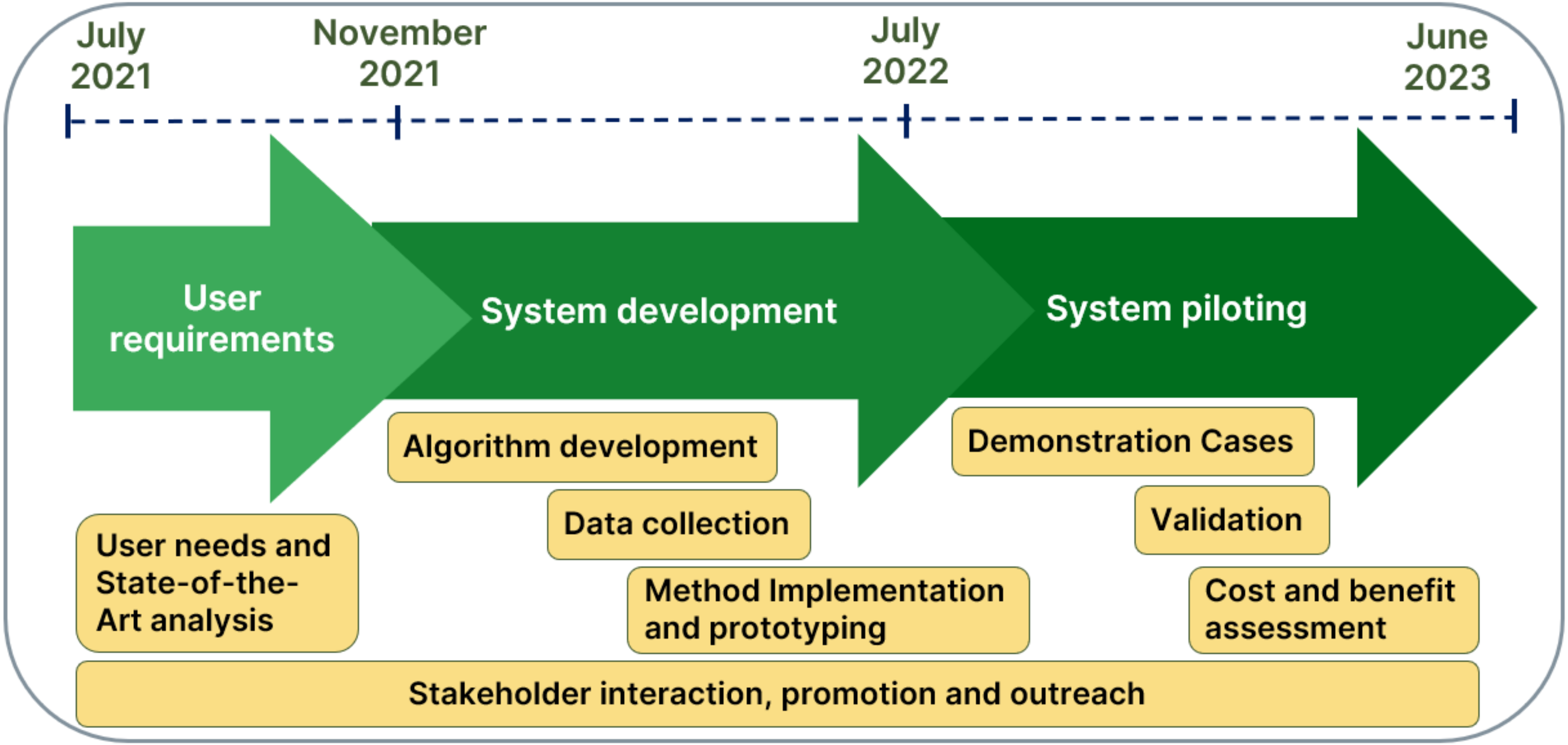
Forestry TEP

- **The platform demonstrations were implemented on Forestry TEP**
- **Ways to use Forestry TEP**
  - Use available applications that combine EO data and your own input datasets
  - Develop your own processing scripts
  - Share or license applications
  - Access or share output products
- **Two modes of usage**
  - Online web user interface
  - REST API for interconnecting between systems
- **All information available at: https://f-tep.com**





Project flow



Demonstration products

Forest structure variable products

- Stem density
- Height
- Diameter
- Basal area
- Growing stock volume
- Species proportions
- Site type

Biomass and growth products

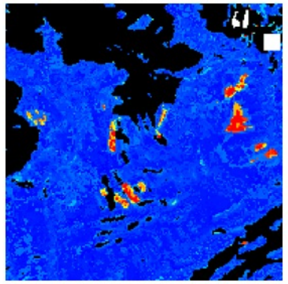
- Above ground biomass
- Below ground biomass
- Stem volume increment

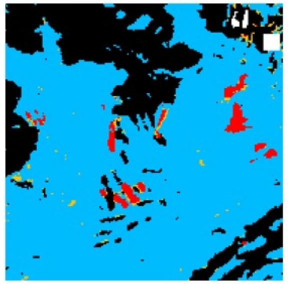
Change products

- Change magnitude
- Change type
- Biomass decrease mask

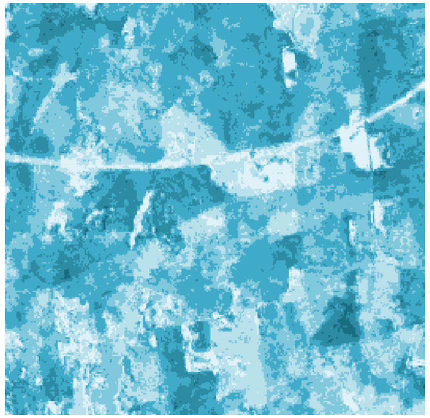

Sentinel-2 2020

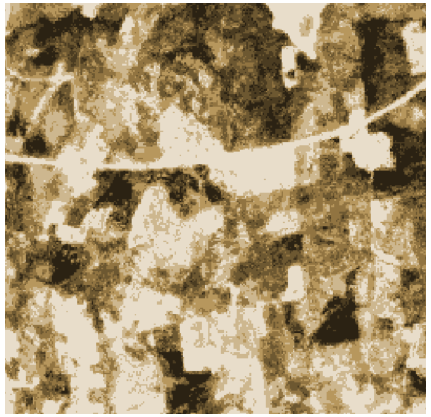

Sentinel-2 2021

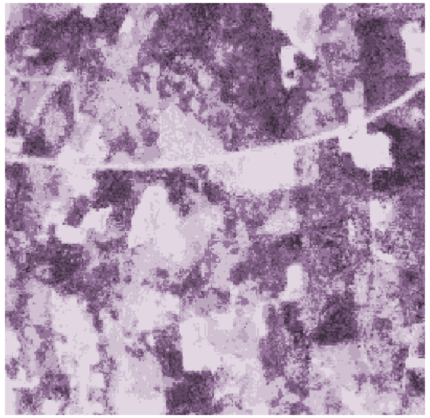

Change magnitude


Biomass decrease mask

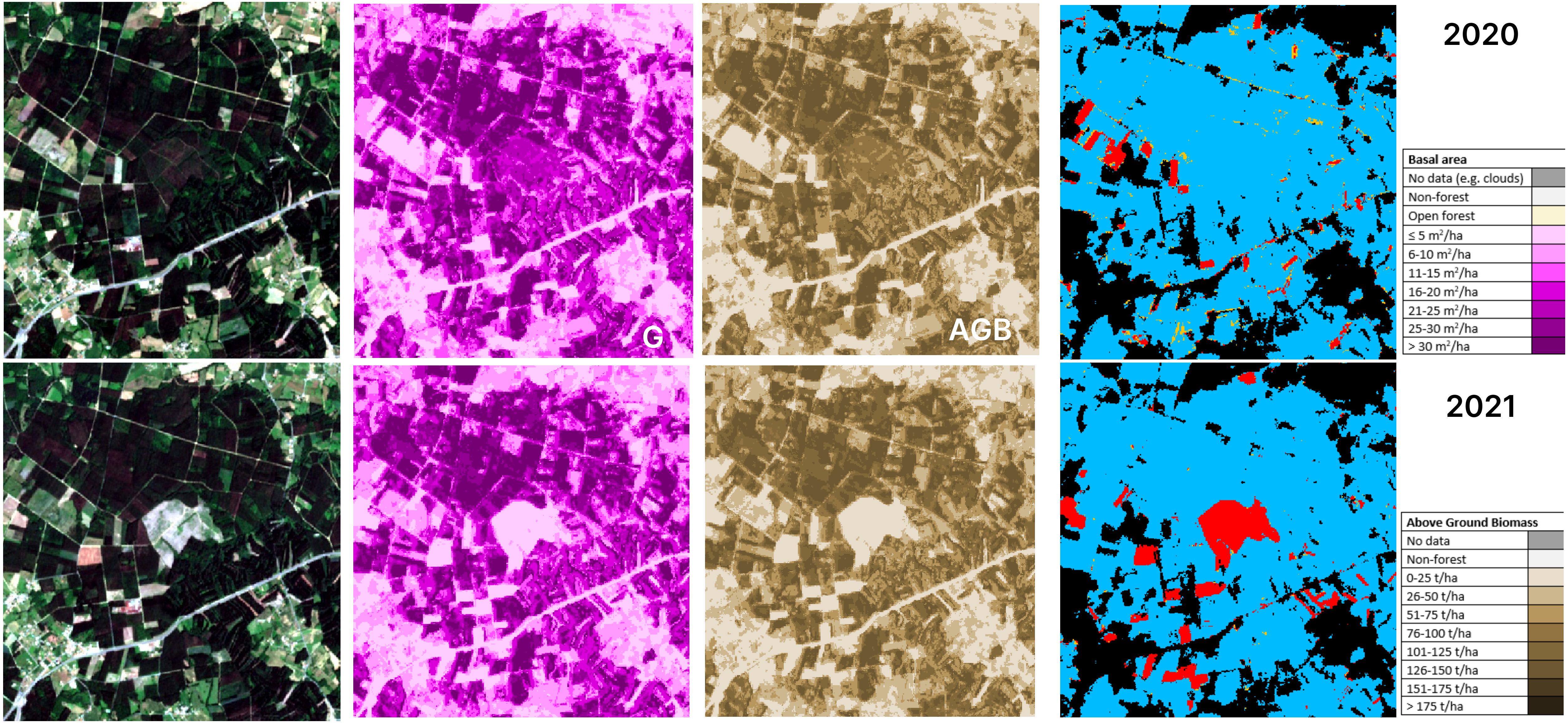

Sentinel-2


Height


Growing Stock Volume

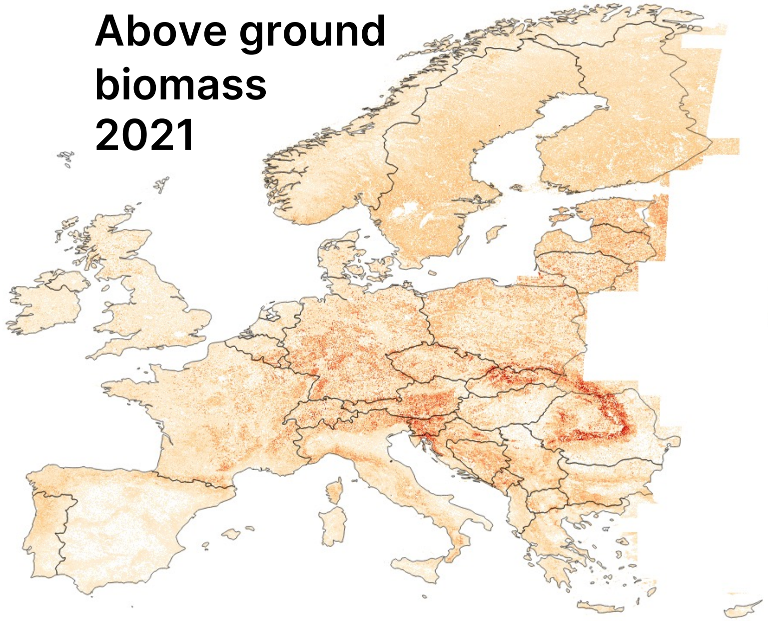

Below Ground Biomass

Example of local level products - Galicia

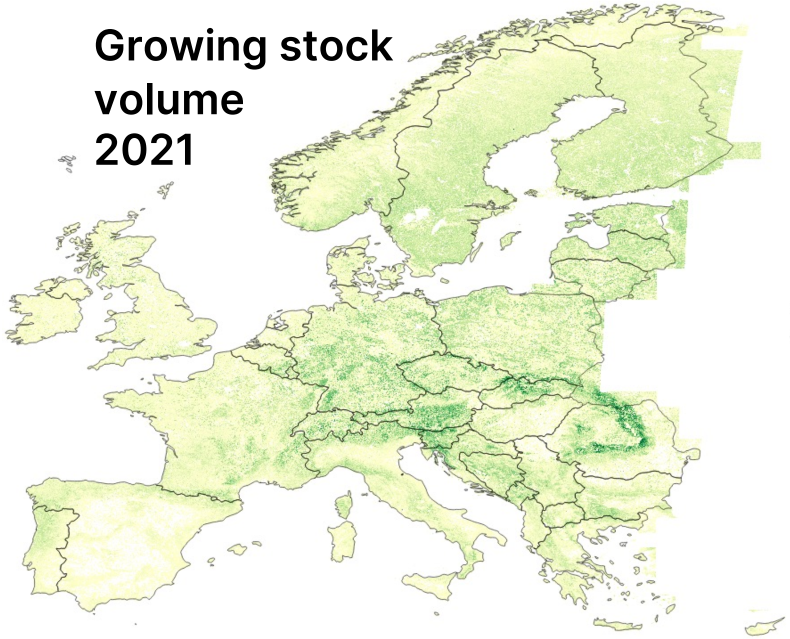


Examples of products - European wide mapping (I)

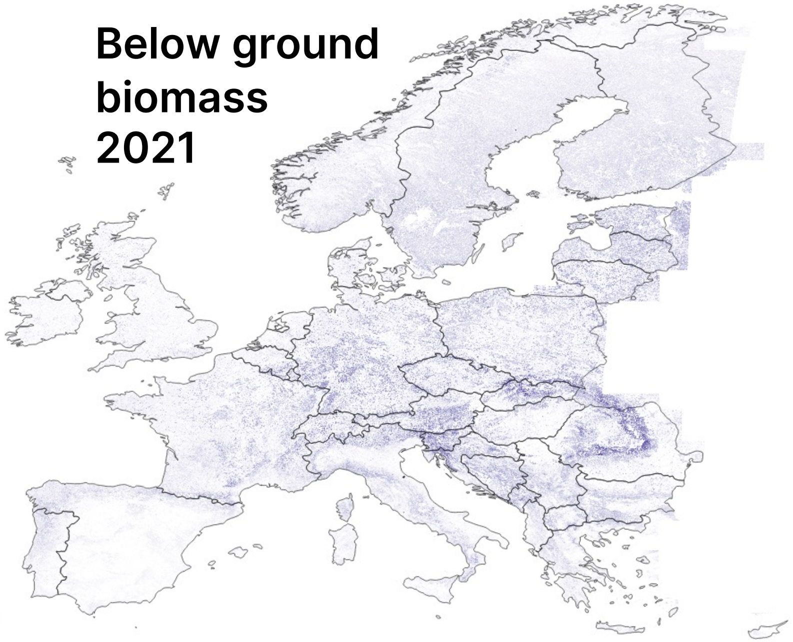
Above ground biomass 2021

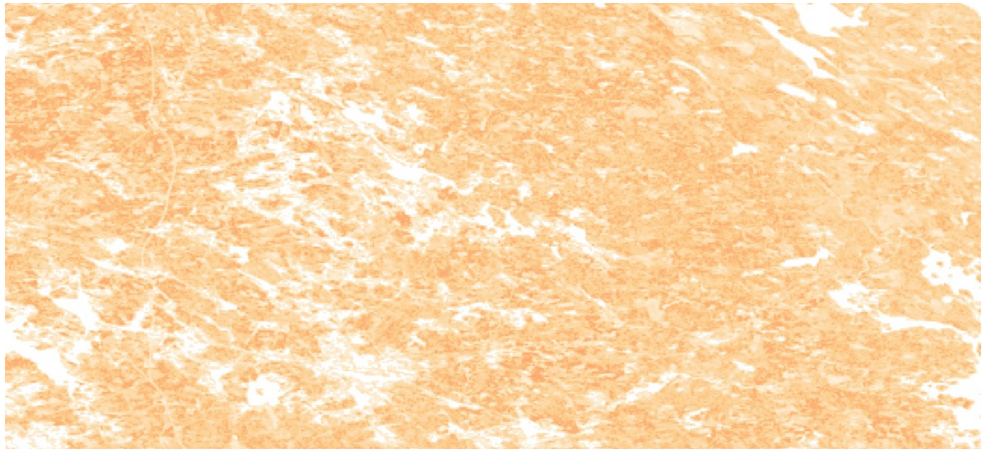


Growing stock volume 2021



Below ground biomass 2021

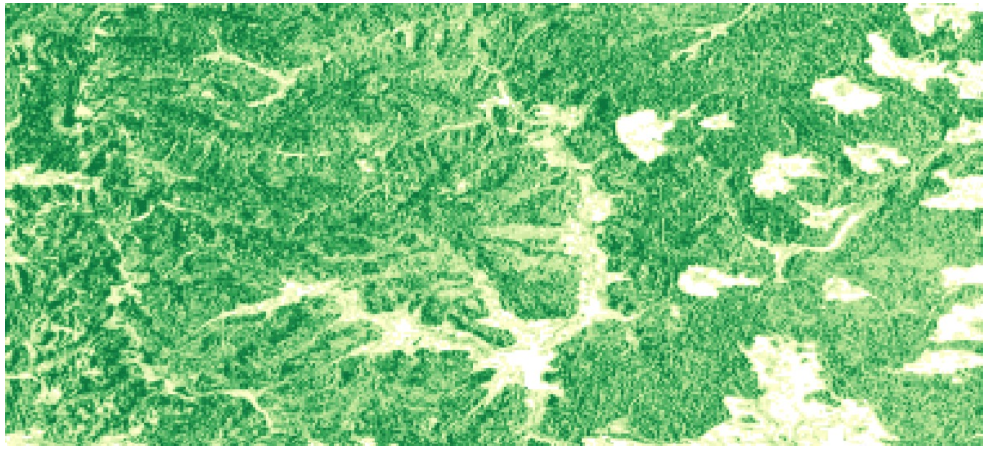




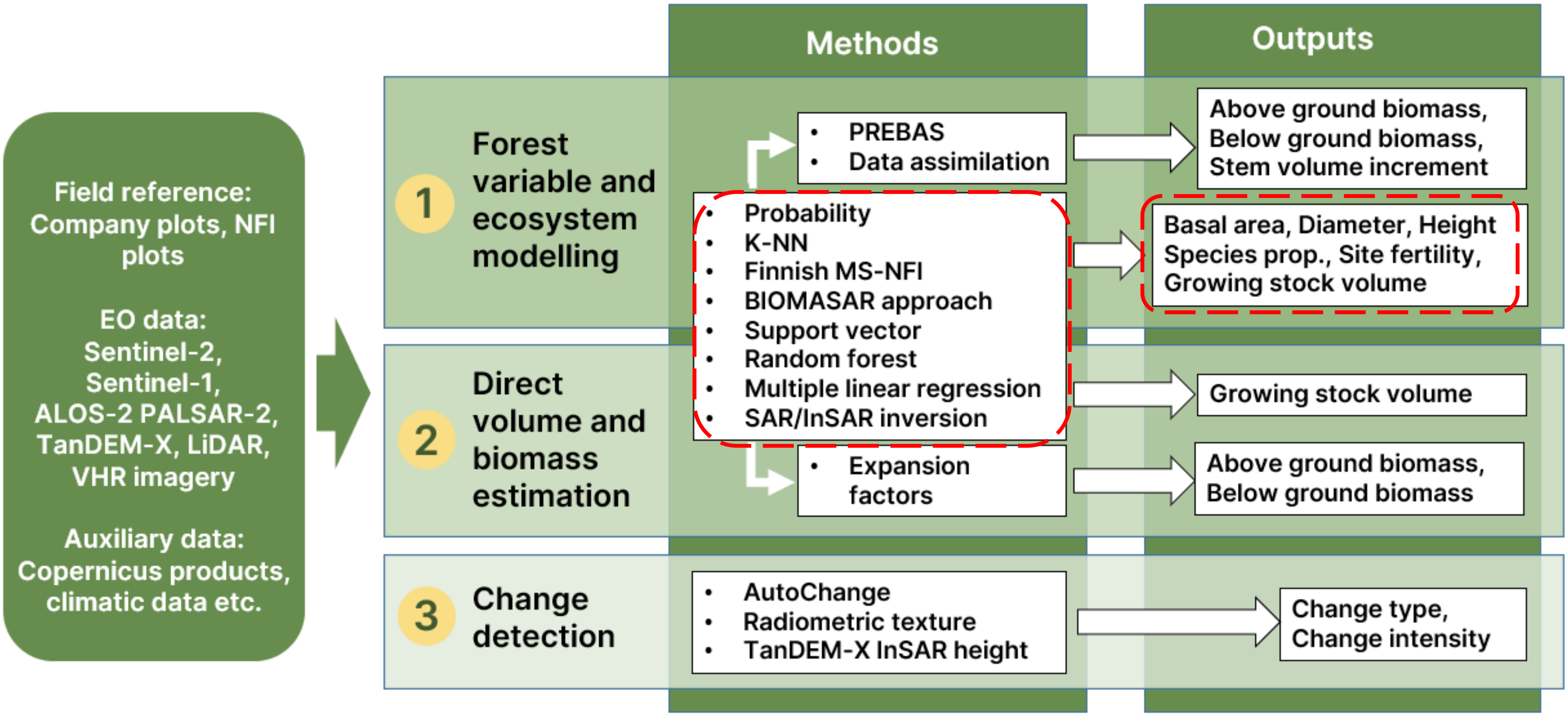
20 X 30 km subset in Finland

20 X 30 km subset in Germany

2021



Three main pathways in the algorithm evaluation



Forest variable prediction intercomparison

Testing and benchmarking forest inventory mapping approaches over Test sites using local computing facilities

↓

Implementing several selected approaches (EO data + methods) in cloud environment (Forestry TEP)

↓

Methodology demonstrations with estimations and uncertainty assessment over Demonstration sites

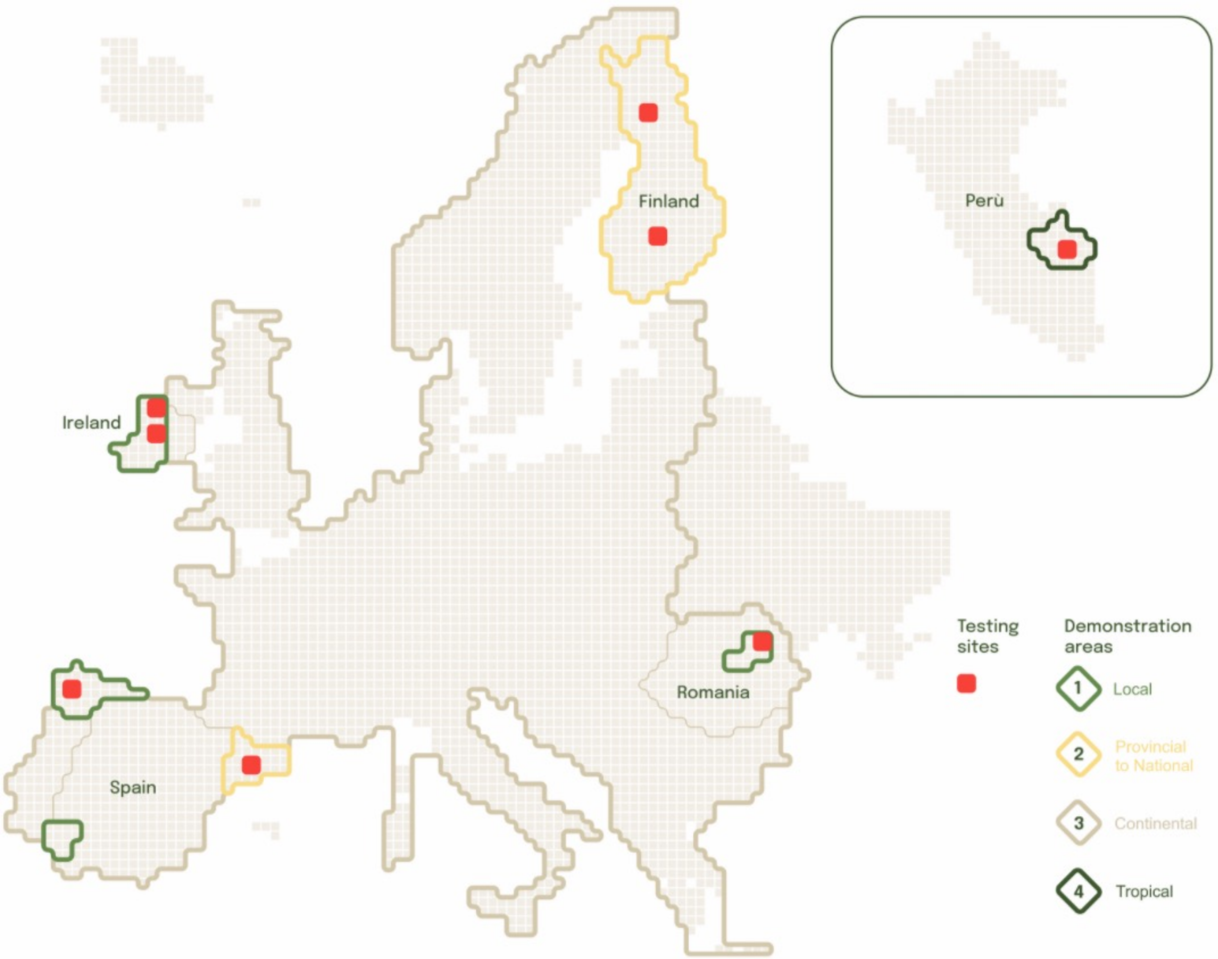
Method intercomparison rationale

Data & methods intercomparison summary

Variable	Algorithms to be tested	Satellite data types	
		Satellite optical	Satellite radar
Basal area, Diameter, Tree species	Probability	x	x
	k-NN	x	x
Tree height	Probability	x	x
	k-NN	x	x
	SAR/InSAR model inversion	x	x
	Regression (SVR, RF, MLR)	x	x
Site fertility	Probability	x	
	k-NN	x	
	Classification (SVM, RF)	x	x
Growing stock volume	Probability	x	x
	k-NN	x	x
	SAR/InSAR model inversion		x
	Regression (SVR, RF, MLR)	x	x
	BIOMASAR approach		x
Above-ground biomass	PREBAS		
	BIOMASAR approach		x
Below-ground biomass	PREBAS		
	BIOMASAR approach		
Change detection	AutoChange	x	
	TanDEM-X InSAR height		x
	Sentinel-1 radiometric contrast		x

Testing sites

Site	Country	Forest types	Climate zone	Topography
1	Finland	Semi-natural coniferous and broadleaf	Arctic	Hilly
2	Finland	Semi-natural coniferous and broadleaf	Boreal	Gently undulating
3	Ireland	Mainly coniferous plantations, some broadleaf	Atlantic	Gently undulating
4	Romania	Semi-natural coniferous and broadleaf	Temperate/ Continental	Hilly
5	Spain	Eucalypt plantations, some natural forests	Atlantic	Hilly
6	Spain	Semi-natural coniferous and broadleaf	Mediterranean	Hilly to Mountainous
7	Peru	Amazonian evergreen	Tropical	Gently undulating



Forest variable prediction intercomparison

Rationale: Benchmarking presently available satellite image datasets and suitable classification/prediction methodologies to identify

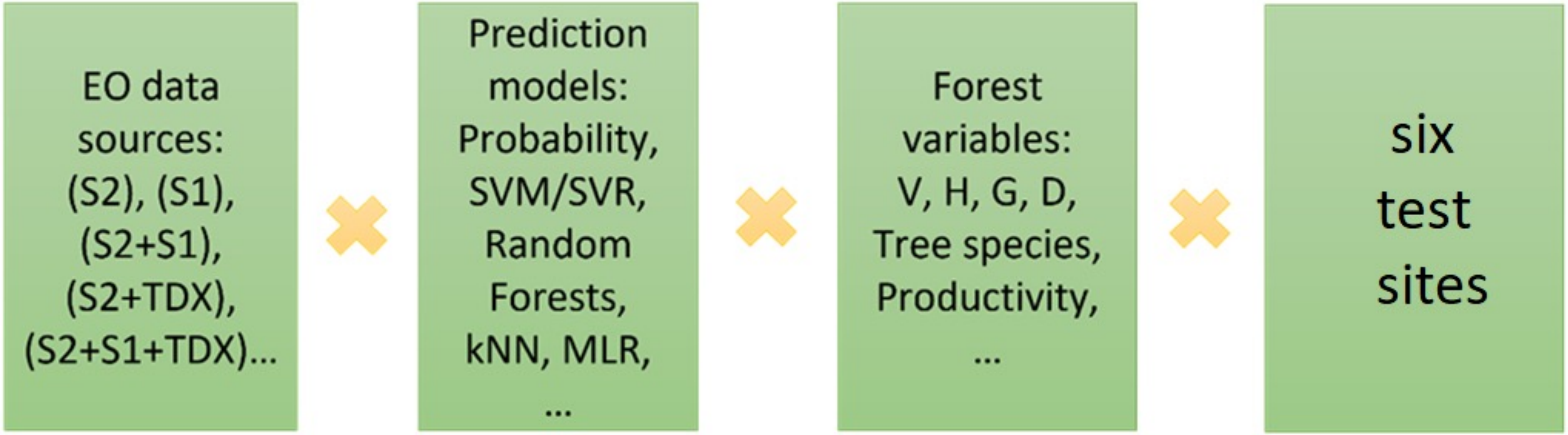


Illustration of studied combinations in the method intercomparison

Forest inventory with Earth Observation data

- **Forest inventories** provide detailed information about the current state of the forest and its change.
- Information can be reported on sample unit level (plots), on forest compartment level, other small-area or large-area level.
- **Forest variables**: forest tree height, canopy closure, tree species, growing stock volume, diameter at breast height, basal area.
- Data sources “traditionally” used in connection with forest inventories: aerial images, field survey, ALS data

Use of satellite Earth Observation data as auxiliary data along with plot-level data:

- allows to increase precision of estimation compared to using only forest plots;
- enables estimation for small areas when the plot sample size does not allow direct estimation;
- allows producing estimates in remote or hardly accessible areas;
- enables producing wall-to-wall maps with reference information key for model training and uncertainty quantification.

Image source: Google Earth, forest information: Metsäkeskus and National Land Survey of Finland, 2015

Sensor image data

- **Optical multispectral images**
- **Synthetic aperture radar images**
  - **Multitemporal / time-series**
  - **Multipolarization**
  - **Interferometric**
- **Various combinations of SAR and optical images**

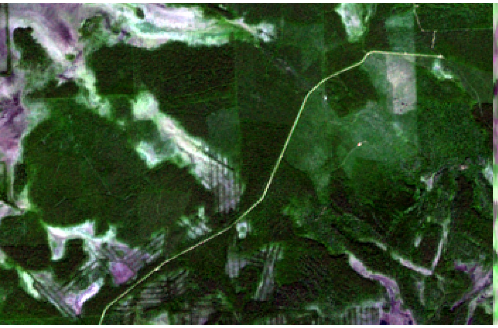
$$I\,(m,x,y) = F\,[\ldots, \{\text{ target properties }\}, \ldots]$$

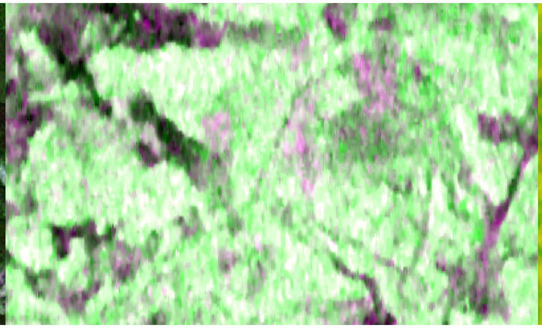
$$\{\text{ target properties }\} = F^{-1}\,[\,I\,(x,y)\,]$$

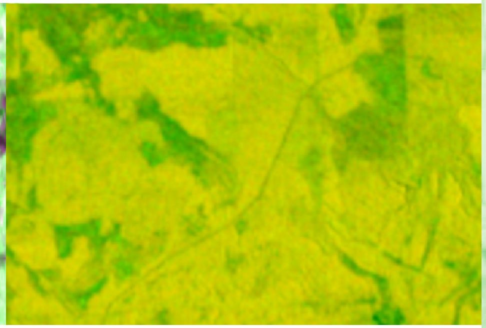
{ target properties} : {volumetric water content, roughness, orientation, vertical structure, density, spatial structure}

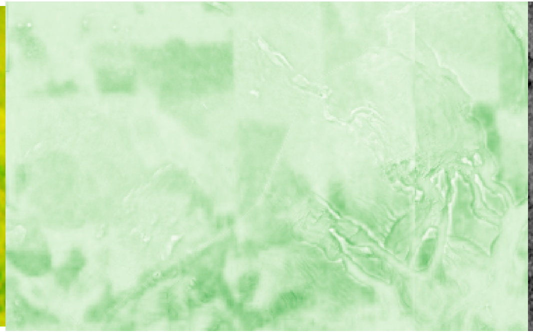
{ forest variables} : {growing stock volume, height, DBH, tree species, …}

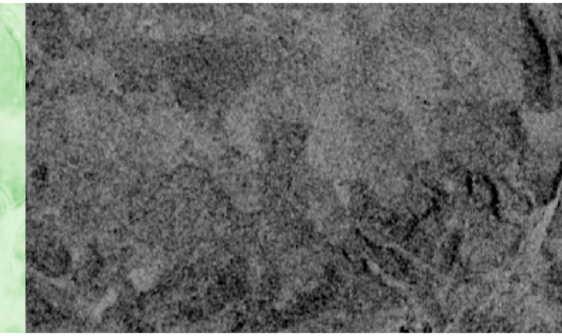
$$\{\text{ forest variables }\} = Z\,[\,I\,(m,x,y)\,]$$


Sentinel-2


ALOS-2 PALSAR-2


Sentinel-1


TanDEM-X InSAR CHM


TanDEM-X coherence

Modeling principles

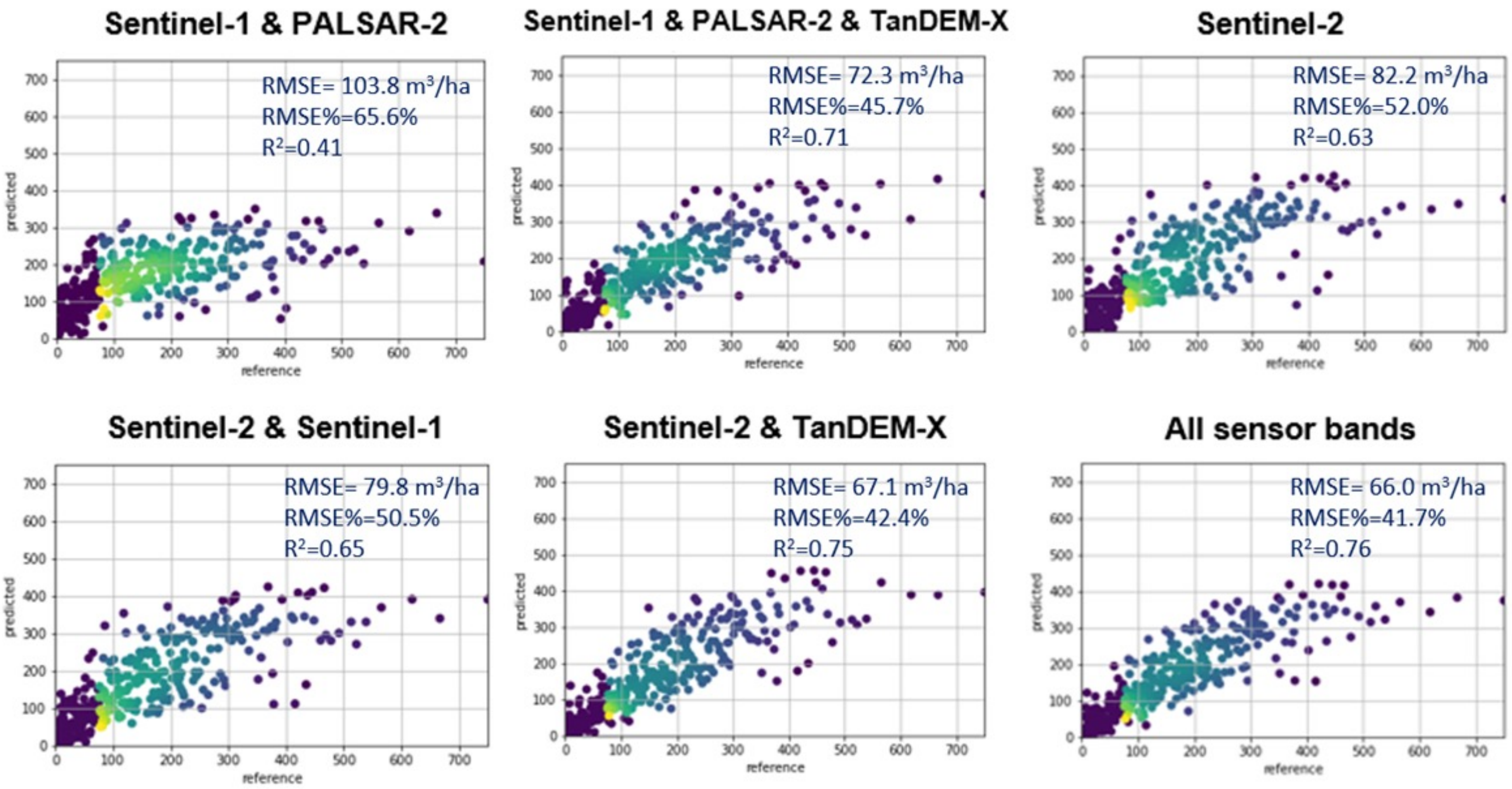
Models describing relationship between forest variables and RS observables:

- physics-based and semi-empirical (motivated by wavelength, resolution, env conditions), reference data used for “calibration”
    - normally suitable for a given sensor/wavelength
      (e.g., InSAR coherence models for vegetation, WCM vegetation)
- Statistical parametric models (partly overlaps with earlier), model fitting is used, reference data are used for teaching models
    - often don’t care about “nature” of EO data
- Non-parametric approaches - completely dependent on reference data
    - normally don’t care about “nature” of EO data
- Semi-supervised approaches - utilize EO data even when reference data are missing

Methodologies for prediction/classification

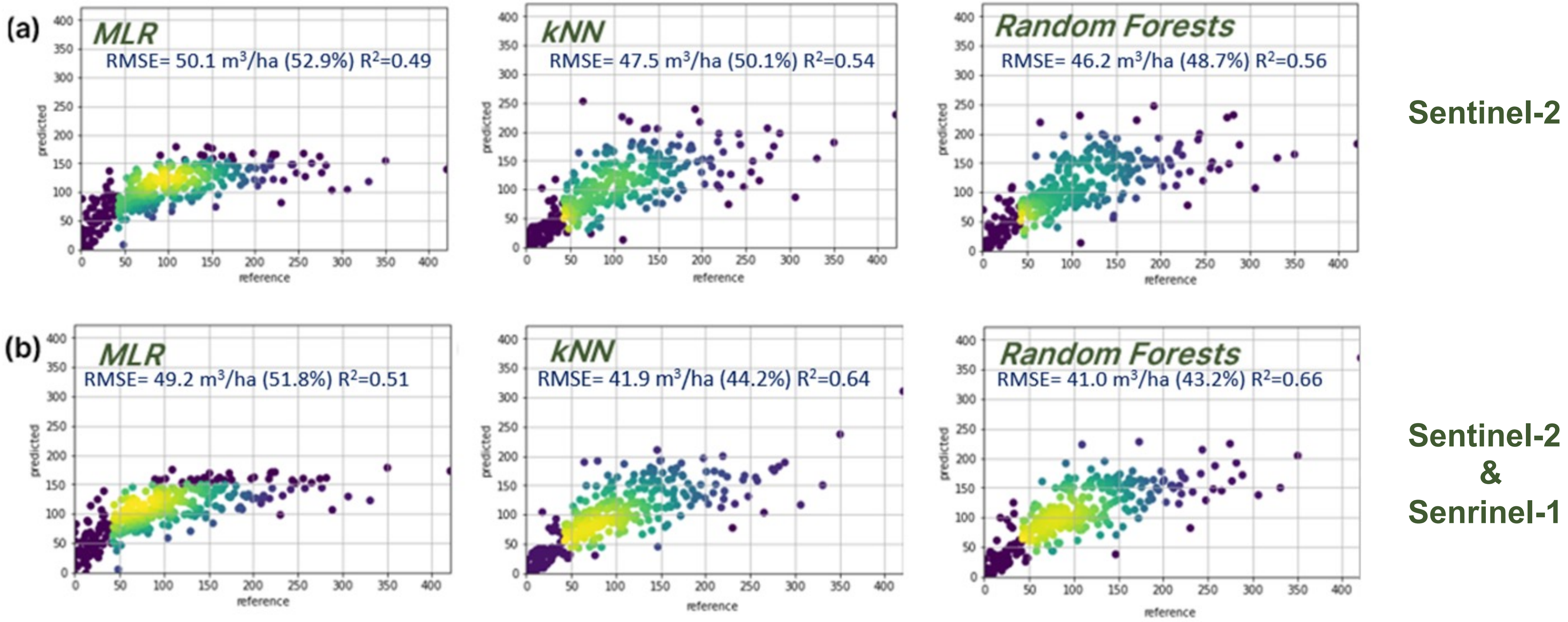
- **Parametric, semi-empirical and physics-based models:**
  - WCM (water cloud model) derived
  - RVoG (random volume over ground) derived
- **Statistical parametric methodologies:**
  - MLR
- **Machine learning non-parametric methods:**
  - k-NN,
  - support vector regression,
  - random forests
- **Semisupervised non-parametric methods:**
  - Probability

EO data intercomparison: SAR and optical images



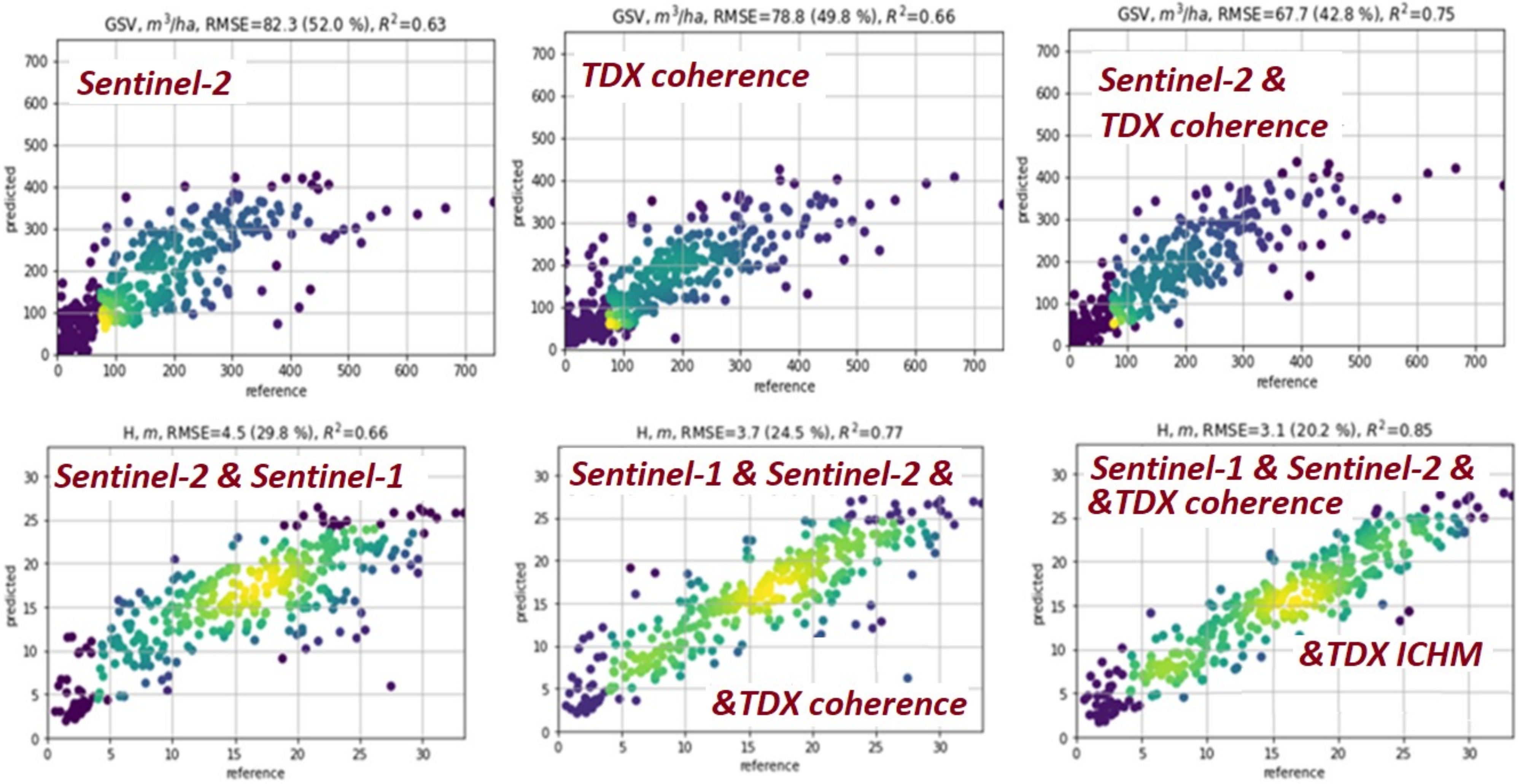
Southern Finland site GSV predictions with kNN using various combinations of EO images

EO data intercomparison: SAR and optical images



GSV prediction over Northern Finland site using various methods and EO data combinations:
(a) Sentinel-2; (b) Sentinel-2 & Sentinel-1

EO data intercomparison: role of vertical structure



Southern Finland site forest variable predictions using various EO imagery with the k-NN method: top row – growing stock volume, bottom row – forest tree height.

Role of the TanDEM-X dataset was important with all methods and many forest variables, least with forest tree species proportions and site index.

Feature selection

Height (H)	Basal area (G)	GSV (V)	Diameter at breast height (D)	PINE proportion	SPRUCE proportion	Broadleaf trees (BL) proportion

Rank	H	G	V	D	PINE	SPRUCE	BL
1	TDX-Coh	S2-5	TDX-CHM	TDX-Coh	S2-1	S2-6	S2-4
2	TDX-CHM	TDX-CHM	S2-6	S2-4	S2-4	S1-VH	S2-2
3	S2-5	P2-HV	TDX-Coh	TDX-CHM	TDX-CHM	S1-VV	S2-3
4	S2-7	S2-4	P2-HV	S1-VV	S1-VH	S2-5	TDX-CHM
5	S2-1	S2-1	S2-5	P2-JD	S1-VV	S2-1	S2-6
6	S2-2	S2-2	S2-4	S1-VH	S2-2	S2-2	S2-5
7	P2-HH	S2-6	P2-HH	P2-HH	TDX-Coh	S2-7	S2-7
8	S2-4	P2-HH	S1-VH	P2-HV	P2-HV	P2-HH	P2-HH
9	P2-HV	P2-JD	S2-2	S2-5	S2-5	P2-HV	P2-HV
10	S2-3	S1-VV	S2-1	S2-7	P2-HH	P2-JD	S2-1
11	P2-JD	S1-VH	S2-7	S2-1	S2-6	TDX-CHM	S1-VV
12	S1-VH	S2-3	S2-3	S2-2	S2-7	TDX-Coh	S1-VH
13	S2-6	S2-7	P2-JD	S2-3	P2-JD	S2-3	TDX-Coh
14	S1-VV	TDX-Coh	S1-VV	S2-6	S2-3	S2-4	P2-JD

Sequential feature selection of EO data features over Northern Finland site

S2-1	Sentinel-2 band 2 Blue
S2-2	Sentinel-2 band 3 green
S2-3	Sentinel-2 band 4 Red
S2-4	Sentinel-2 band 8 NIR
S2-5	Sentinel-2 band 5 VegRE
S2-6	Sentinel-2 band 11 SWIR
S2-7	Sentinel-2 band 12 SWIR
S1-VH	Sentinel-1 VH-pol
S2-VV	Sentinel-1 VV-pol
P2-HH	ALOS-2 PALSAR-2 HH-pol
P2-HV	ALOS-2 PALSAR-2 HV-pol
P2-JD	ALOS-2 PALSAR-2 day index
TDX-Coh	TanDEM-X coherence magnitude
TDX-CHM	TanDEM-X InSAR CHM

Feature selection

Height (H)	Basal area (G)	GSV (V)	Diameter at breast height (D)	PINE proportion	SPRUCE proportion	Broadleaf trees (BL) proportion

- Lasso feature selection, alpha=1.0:

	S2-1	S2-2	S2-3	S2-4	S2-5	S2-6	S2-7	S1-VH	S1-VV	P2-HH	P2-HV	P2-JD	TDX-CHM	TDX-Coh
H	0,0	0,0	0,0	0,3	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,2	2,4
G	0,0	0,0	0,0	0,8	3,1	1,3	0,0	0,0	0,0	0,2	1,7	0,0	2,7	0,0
V	0,0	2,5	0,0	15,1	36,2	0,0	7,0	0,0	1,1	4,1	10,8	0,0	59,9	9,5
D	0,0	0,0	0,0	2,1	1,3	0,0	0,0	0,0	0,0	0,1	0,0	0,0	2,1	2,9
PINE	1,6	16,9	0,4	20,9	0,0	0,0	0,0	26,7	25,0	0,0	1,4	0,0	5,7	0,0
SPRUCE	2,5	0,0	0,0	0,0	0,0	12,9	0,0	27,6	25,4	1,3	0,0	0,0	0,0	0,0
BL	0,0	15,5	0,0	22,0	6,4	14,6	0,0	0,6	0,0	2,4	0,0	0,0	3,0	1,9

- Random forest ranking:

	S2-1	S2-2	S2-3	S2-4	S2-5	S2-6	S2-7	S1-VH	S1-VV	P2-HH	P2-HV	P2-JD	TDX-CHM	TDX-Coh
H	0.014	0.089	0.019	0.11	0.093	0.052	0.042	0.018	0.015	0.023	0.04	0.0012	0.26	0.22
G	0.021	0.12	0.027	0.08	0.15	0.17	0.12	0.02	0.013	0.024	0.1	0.0023	0.1	0.046
V	0.018	0.078	0.02	0.13	0.15	0.13	0.096	0.011	0.011	0.022	0.06	0.0014	0.19	0.087
D	0.019	0.046	0.016	0.16	0.09	0.079	0.055	0.017	0.014	0.036	0.036	0.0012	0.23	0.2
PINE	0.084	0.074	0.15	0.22	0.059	0.067	0.039	0.058	0.11	0.033	0.032	0.0022	0.038	0.032
SPRUCE	0.051	0.079	0.086	0.11	0.098	0.14	0.13	0.07	0.095	0.038	0.022	0.0031	0.035	0.042
BL	0.051	0.073	0.068	0.27	0.053	0.11	0.064	0.029	0.041	0.084	0.052	0.0032	0.055	0.043

- Mutual information ranking:

	S2-1	S2-2	S2-3	S2-4	S2-5	S2-6	S2-7	S1-VH	S1-VV	P2-HH	P2-HV	P2-JD	TDX-CHM	TDX-Coh
H	0.2	0.5	0.3	0.5	0.6	0.4	0.4	0.3	0.1	0.2	0.3	0.0	1.0	0.7
G	0.4	0.8	0.4	0.6	0.9	1.0	0.8	0.3	0.2	0.4	0.8	0.0	0.6	0.4
V	0.5	0.8	0.4	0.7	1.0	0.9	0.7	0.1	0.2	0.5	0.6	0.0	0.9	0.6
D	0.3	0.6	0.3	0.7	0.8	0.7	0.5	0.3	0.3	0.4	0.5	0.0	1.0	0.8
PINE	0.8	0.3	0.7	1.0	0.6	0.2	0.3	0.0	0.7	0.1	0.0	0.2	0.1	0.0
SPRUCE	0.7	0.7	0.7	0.6	0.7	1.0	0.8	0.1	0.3	0.3	0.2	0.0	0.4	0.0
BL	0.0	0.1	0.2	1.0	0.3	0.6	0.4	0.0	0.0	0.6	0.4	0.0	0.2	0.0

S2-1	Sentinel-2 band 2 Blue
S2-2	Sentinel-2 band 3 green
S2-3	Sentinel-2 band 4 Red
S2-4	Sentinel-2 band 8 NIR
S2-5	Sentinel-2 band 5 VegRE
S2-6	Sentinel-2 band 11 SWIR
S2-7	Sentinel-2 band 12 SWIR
S1-VH	Sentinel-1 VH-pol
S2-VV	Sentinel-1 VV-pol
P2-HH	ALOS-2 PALSAR-2 HH-pol
P2-HV	ALOS-2 PALSAR-2 HV-pol
P2-JD	ALOS-2 PALSAR-2 day index
TDX-Coh	TanDEM-X coherence magnitude
TDX-CHM	TanDEM-X InSAR CHM

Forest variable prediction results: Methods

- Over majority of test sites, **MLR** proved to be a robust prediction method in the sense that increasing number of independent variables improved prediction accuracy.
- Basic **InSAR/SAR models** often required supervision/**fine-tuning** to achieve accuracy levels similar to other studied approaches. However, they seem robust when lacking reference data.
- **kNN** and **Probability** approach have demonstrated similar performance levels and were suitable for **multivariate prediction** of forest attributes.
- **Nonparametric methods** (e.g., kNN) often favoured smaller dimensionality of feature space and appear very **sensitive to non-representative** data.
- **RF** was somewhat superior to **SVR** (aside from site index), with both approaches yielding the **best** possible **predictions after finetuning** their hyperparameters.
- RF & SVR demonstrated the best possible predictions for several forest variables.
- **Visual assessment** of produced maps can **affect final ranking**

Forest variable prediction results: EO datasets

- Sentinel-2 or combined **Sentinel-2 & Sentinel-1** was the most important data combination for predicting **tree species** proportions.
- For **other structural variables**, most centrally GSV and forest height, the best predictions were provided by **combining radar and optical datasets**, with a key role of Sentinel-2 and TanDEM-X datasets.
- From "all forest variables" perspective it is worthwhile to say that Sentinel-2 was the single best dataset, followed by TanDEM-X in case it was available.
- For practically all sites, **combining Sentinel-1 with Sentinel-2** improved prediction accuracy by a small **margin of 2-4 percentage units**, indicating it is useful to combine the two Copernicus datasets.
- For several studied prediction methods and test sites, using all data bands simultaneously provided the best performance.
- With non-parametric approaches, such as kNN and Probability method, **excluding "noisy" bands improved the prediction** in several cases. Use of feature weighting in prediction can be useful to overcome the issue.

Conclusions on data and method combinations (I)

Datasets	Accuracy*	Considerations
Sentinel-1 only	50-80%	• Time series required • Limited accuracy • All weather capability
Sentinel-2 only	20-60%	• Required for species • Best single dataset • Inter-image variation
Sentinel-2 + Sentinel-1 or PALSAR2	20-60%	• Minor improvement to Sentinel-2 alone
Sentinel-2 + Sentinel-1 + TanDEM-X coherence	20-40%	• Great improvement for Height and GSV • Limited availability

* Typical plot level accuracy variation between variables and sites. RMSE percent of the mean.

Conclusions on data and method combinations (II)

- Recommended "decision tree" for selecting methods:

Field data for training available? — yes → Sufficient* number of field plots? — yes → Multiple output variables ? — yes → *k-NN*

Field data for training available? — no → *BIOMASAR*

Growing stock vol.
Above ground bio.
Below ground bio.

20-100 m res.

Sufficient* number of field plots? — no → *Probability*

Basal area
Diameter
Height
Growing stock vol.
Species
Site

PREBAS

Above ground bio.
Below ground bio.
Stem volume incr.

10 m res.

Multiple output variables ? — no → *Many options*

Individual interest variables (e.g., forest height) can be predicted with a method like Random Forests.

10 m res.

k-NN

Basal area
Diameter
Height
Growing stock vol.
Species
Site

PREBAS

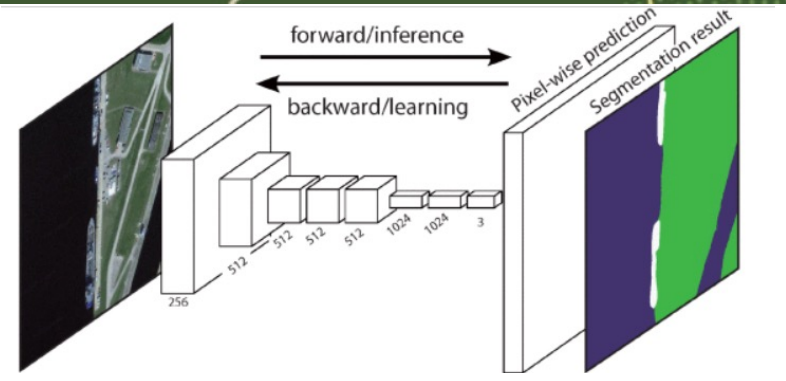
Above ground bio.
Below ground bio.
Stem volume incr.

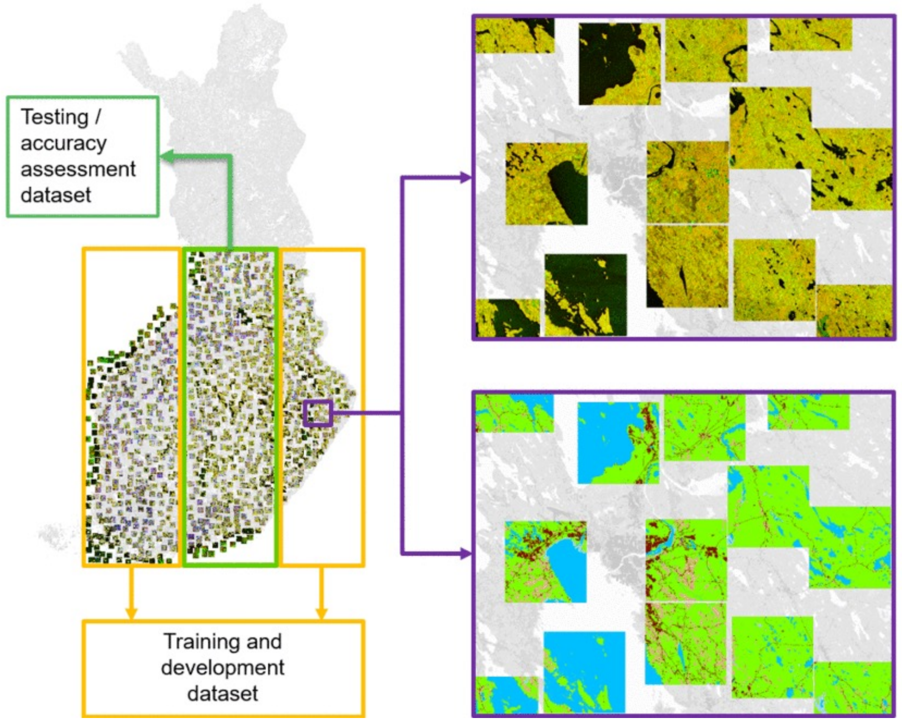
10 m res.

*Typically at least 100 plots

New Horizons: Deep Learning

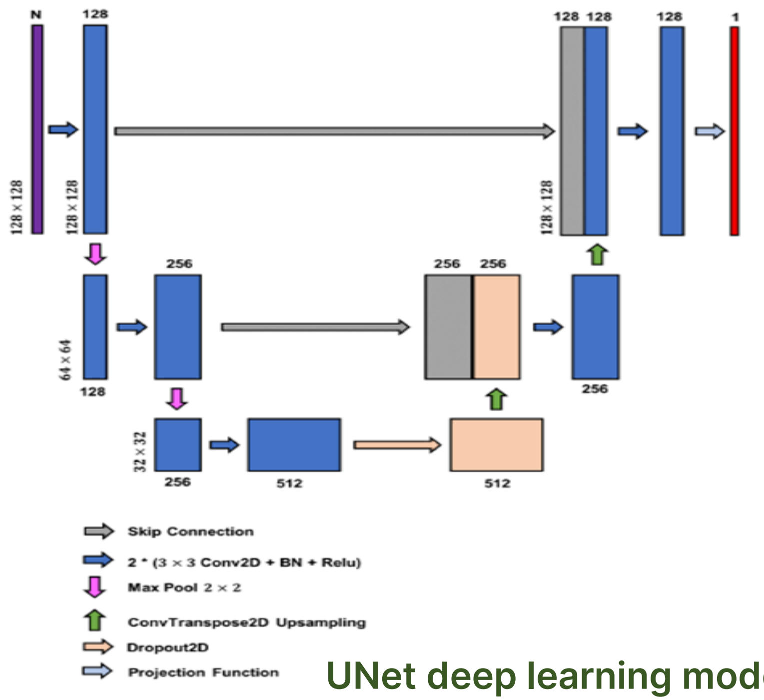
- Capable of automatically extracting spatial textural and temporal dependencies vs ”hand-engineered features”
- Require high quality and extensive reference data labels, that is fully segmented labels
- Already quite popular in semantic segmentation tasks with EO data, such as land cover mapping
- Semi-supervised learning scenarios already demonstrated
- Possible domain adaptation or model transfer
- Several ”pilot” studies in forest variable prediction using EO data



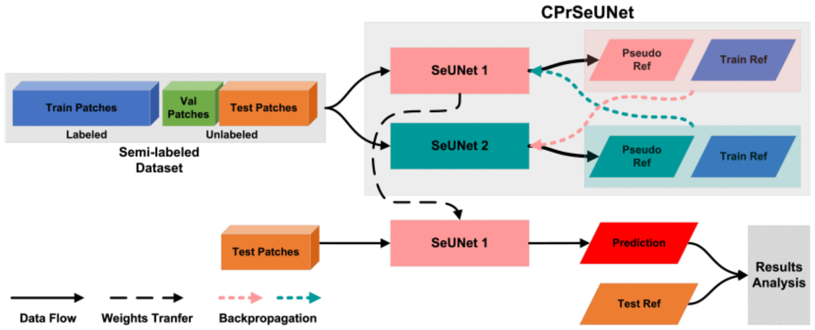


Šćepanović et al., Wide-Area Land Cover Mapping With Sentinel-1 Imagery Using Deep Learning Semantic Segmentation Models," in *IEEE JSTARS*, 2021

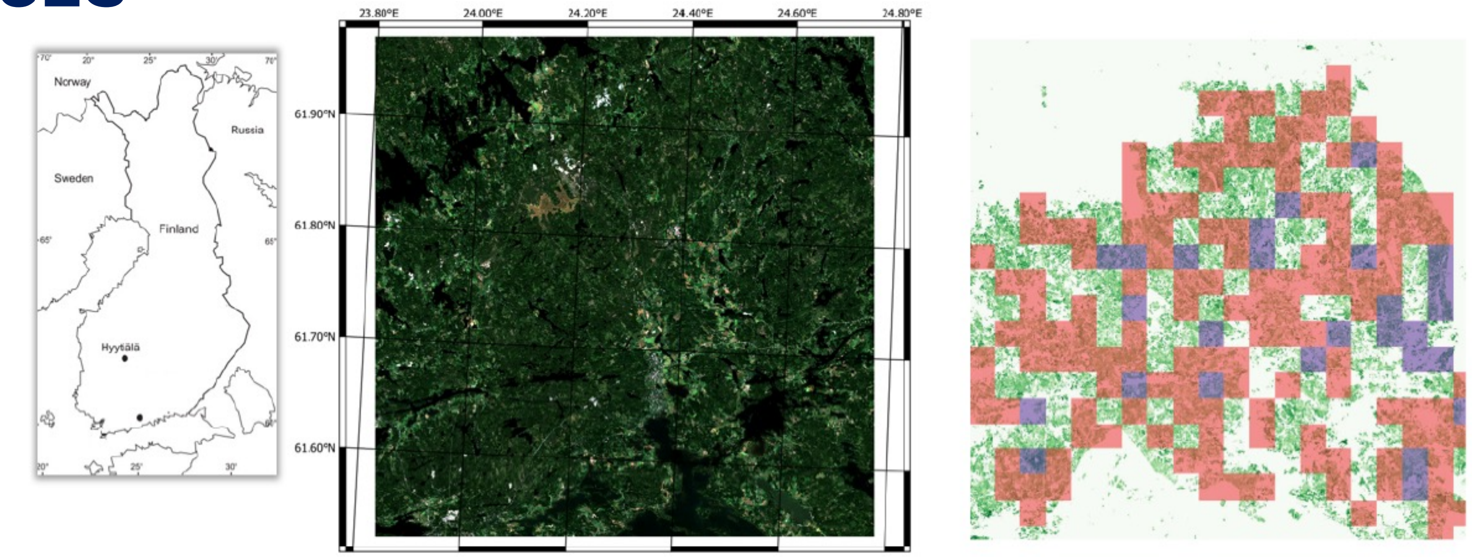
UNet based improved models



UNet deep learning model



Improved CPrSeUNet model



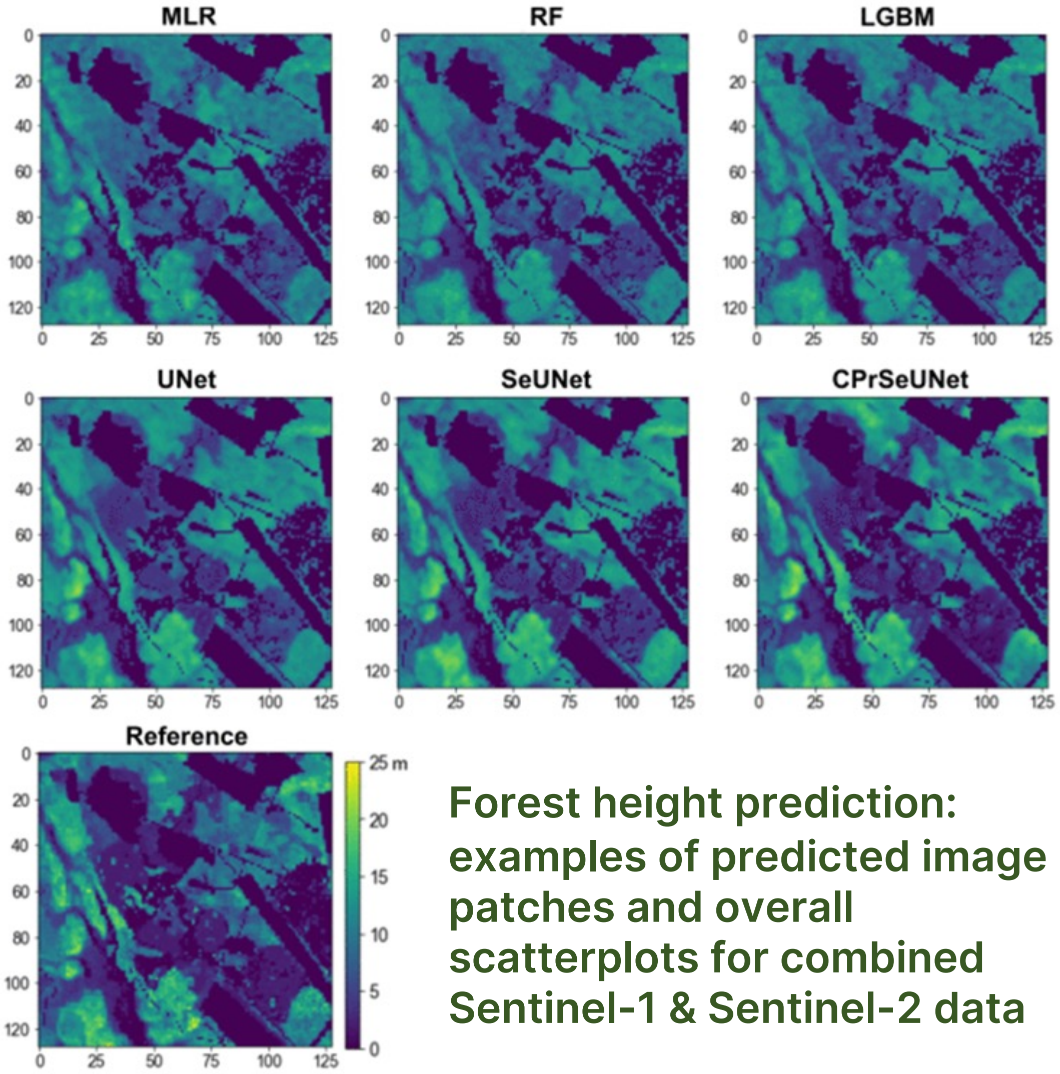
Study site location and division into training(red), validation (blue) and testing sets, 50x50 km2 size

Key points:

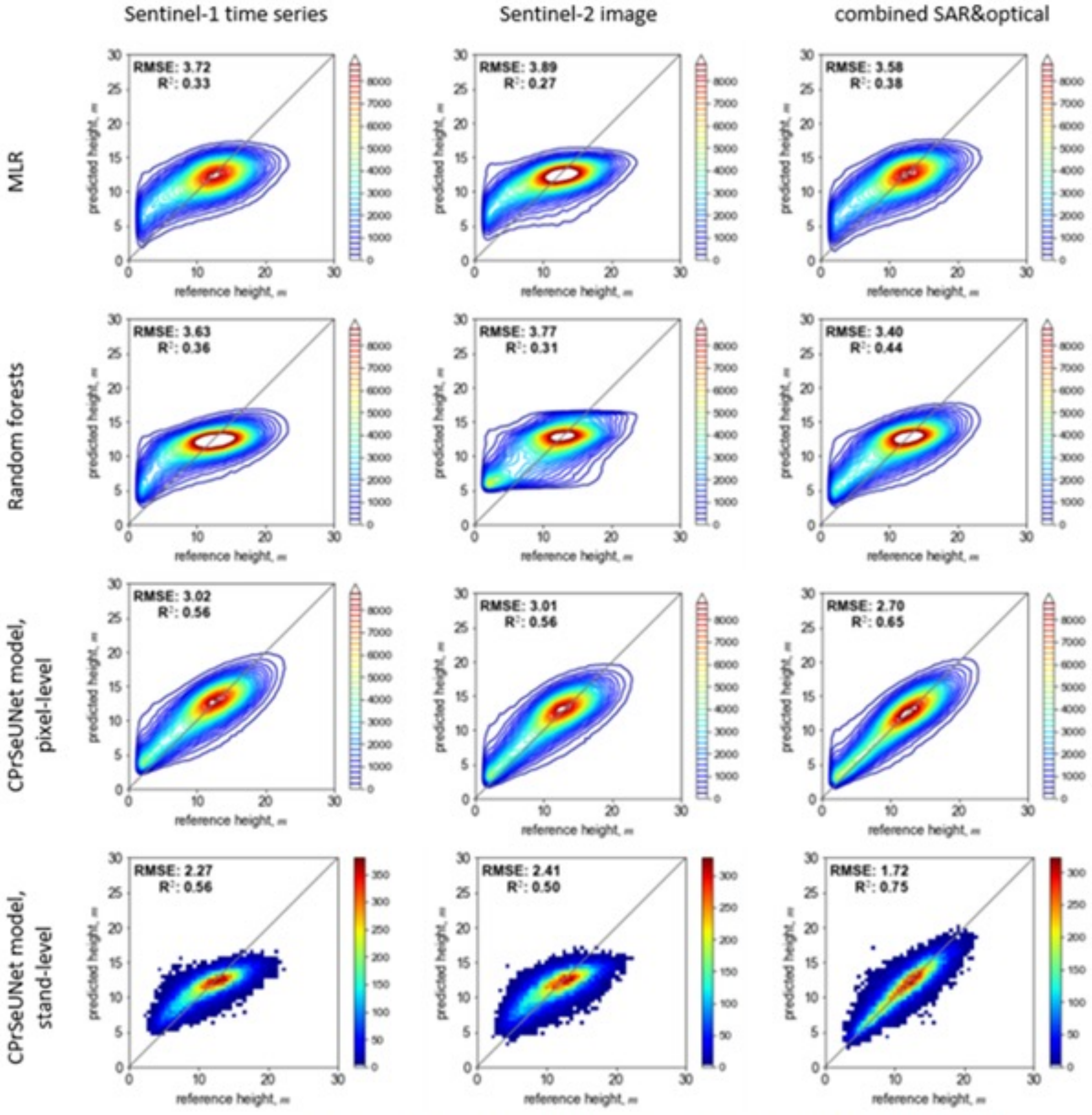
- Target variable – forest tree height, reference data – airborne laser measurements, predictor variables – features from several EO datasets (radar channels, optical bands)
- Comparison with machine learning approaches MLR, SVR, RF
- Testing separately Sentinel-1 images (frozen/nonfrozen), Sentinel-1 time series (27 datatakes), "good" Sentinel-2 image, SAR and optical combined

Ge, Antropov et al., "Improved Semisupervised UNet Deep Learning Model for Forest Height Mapping With Satellite SAR and Optical Data," IEEE JSTARS, vol. 15, pp. 5776-5787, 2022.

Unet+ based models

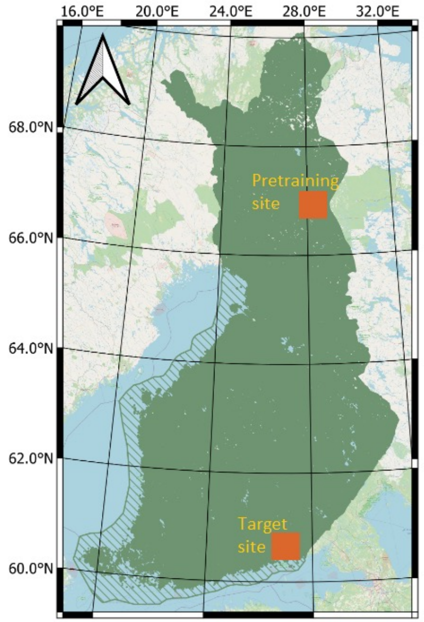


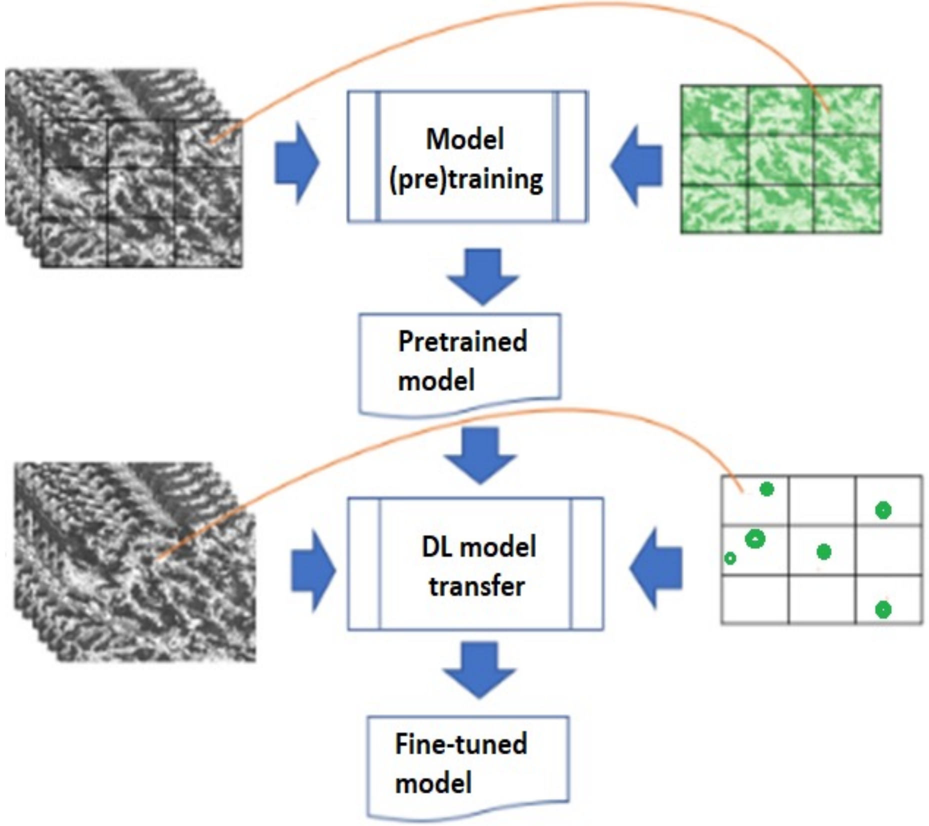
Forest height prediction: examples of predicted image patches and overall scatterplots for combined Sentinel-1 & Sentinel-2 data

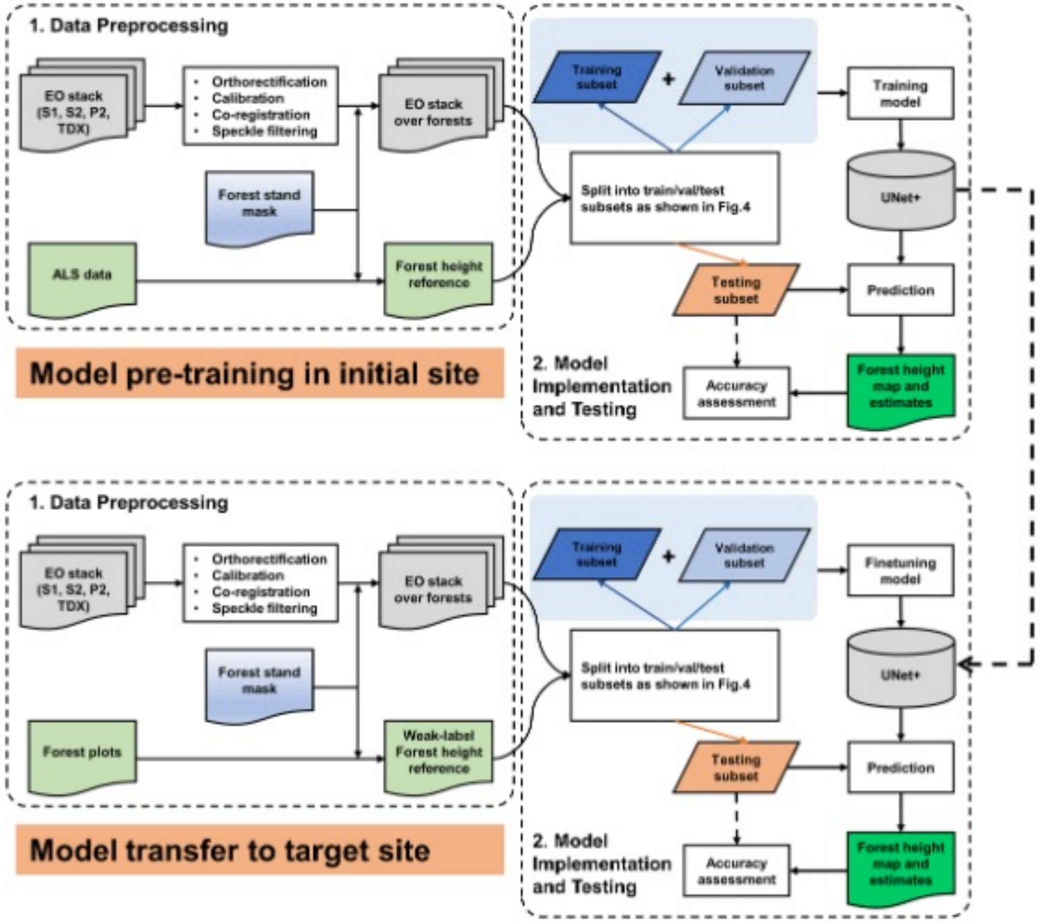


Forest height prediction performance (Ge et al. 2022)

Unet+ model transfer



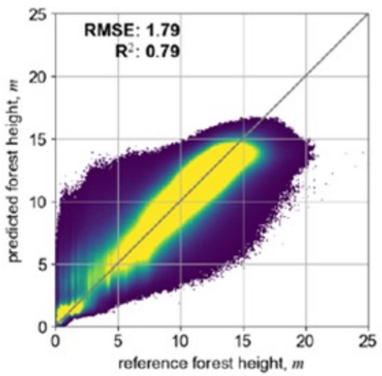


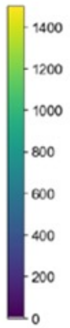


Ge, Antropov, Häme, Miettinen et al. Deep learning models with transfer learning in boreal forest mapping using multi-source satellite SAR/InSAR and optical images, *submitted, 2023.*

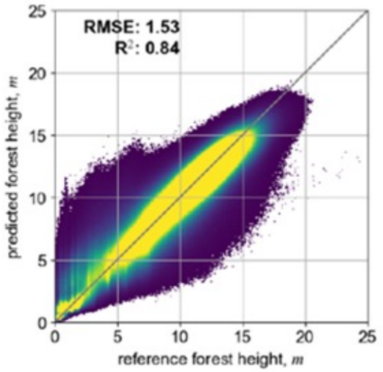
Unet+ model transfer

Pretraining with ALS data



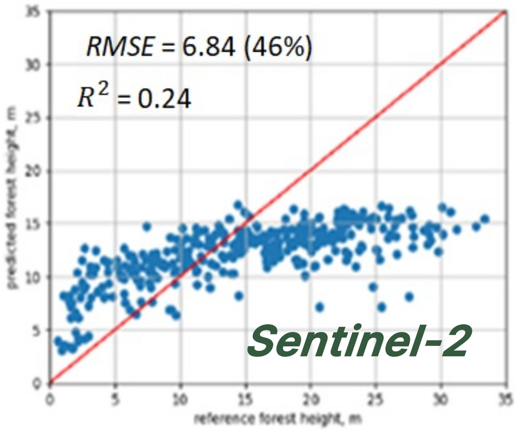


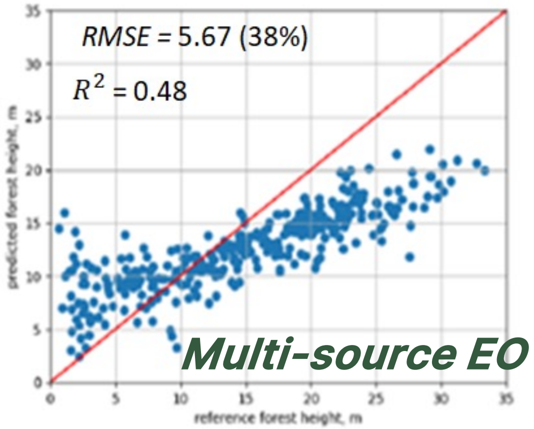
Sentinel-2



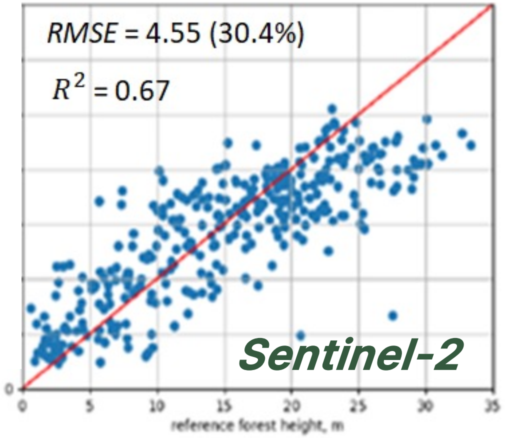
Multi-source EO

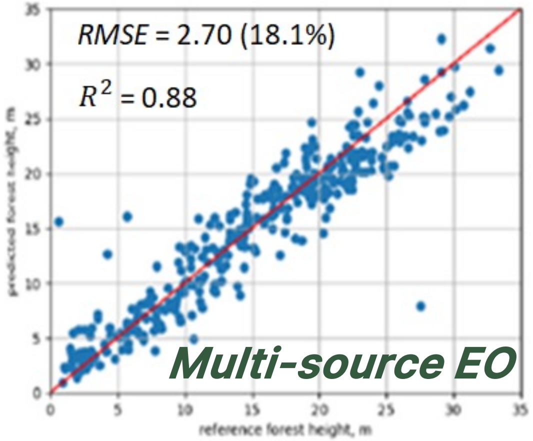
Non-finetuned model applied over target site



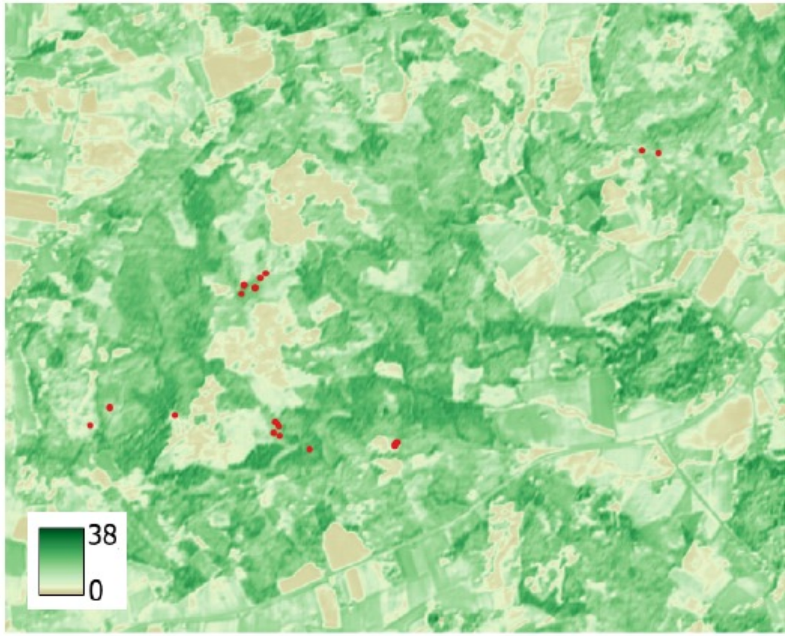


After fine-tuning with forest plots:





Multi-source EO: Sentinel-2 & Sentinel-1 & ALOS-2 PALSAR-2 &TanDEM-X



forest height map

Antropov, Ge, Häme, Miettinen et al. Deep learning models with transfer learning in boreal forest mapping using multi-source satellite SAR/InSAR and optical images, *submitted, 2023.*

Conclusions

- Processing chains developed and tested for wide area forest variable mapping using high resolution optical and radar satellite images
- Optimal pathways suggested using currently available satellite and reference data
- Optimal sensor combinations identified, SAR + optical combination recommended
- Important role of TanDEM-X and vertical structure
- Deep learning model potential with Copernicus and multi-source EO datasets

Forest Carbon
Monitoring

Thank you!

More information at:
https://www.forestcarbonplatform.org