VTT

Forest inventory using EO data

ESA RepreSent user workshop
4 May 2023

Jukka Miettinen, Oleg Antropov and Tuomas Häme

Forest inventory using Earth Observation data

- **Forest inventories** provide detailed information on the state of the forest and its changes.
- Information is needed from sample plot level to forest compartment level and other defined small areas, and for large area monitoring (provincial to global level).
- **Variables of interest**: traditional (forest area, height, species, diameter, growing stock volume…) and increasingly expanding (biomass, carbon, biodiversity, ecosystem services…).
- **Earth Observation data allows:**
  - Monitoring in remote or hardly accessible areas
  - Wall-to-wall maps with increased information on spatial distribution
  - High temporal frequency
  - Estimation for small areas when the plot sample size does not allow direct estimation;

Image source: Google Earth; Forest information: © Metsäkeskus and Finnish National Land Survey, 2015

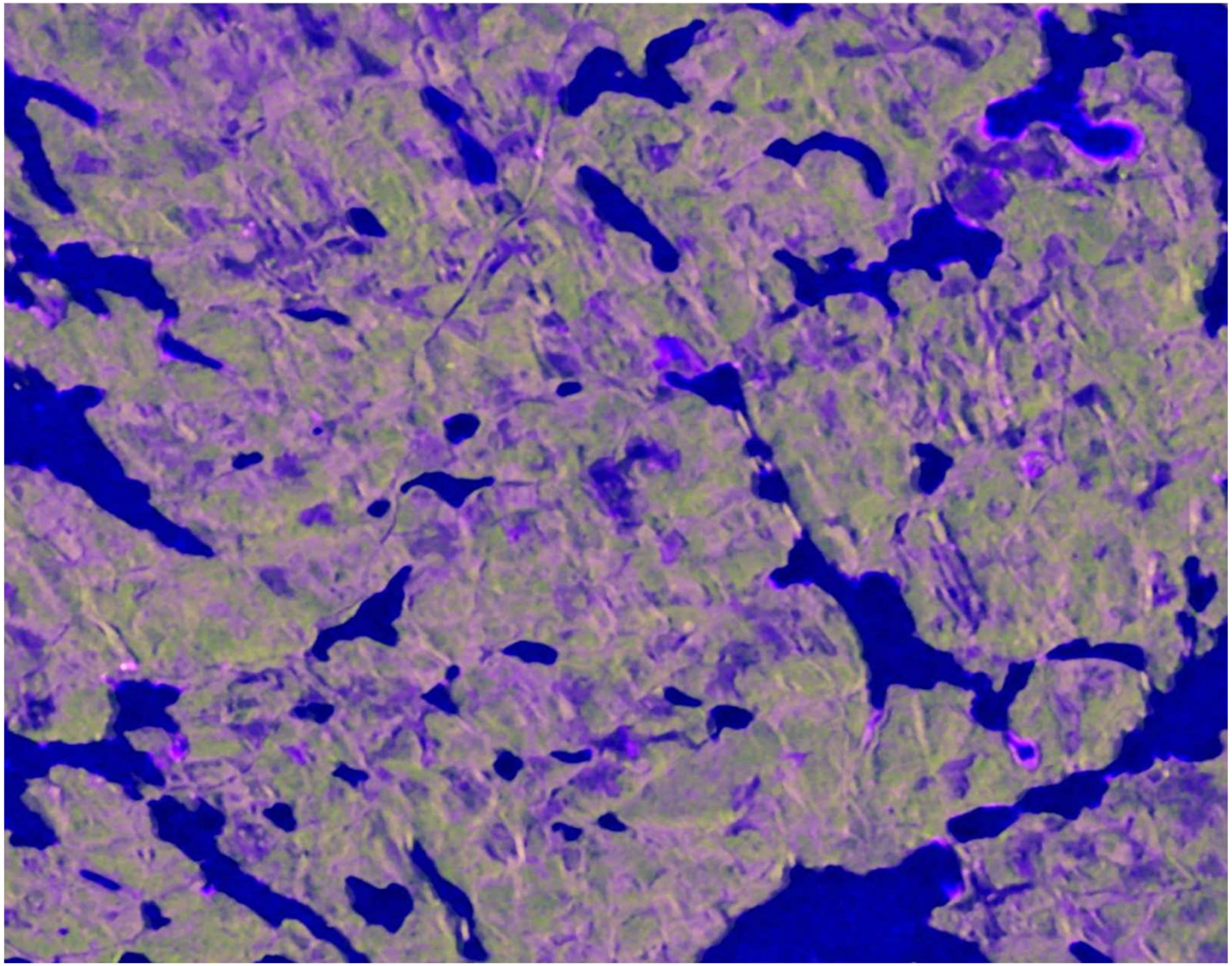
Fit for purpose – remote sensing data

Data type	Frequency	Spatial coverage	Spatial detail	Cost
Satellite (10-30 m optical)	***	***	*	*
Satellite (10-30 m radar)	***	***	*	*
Satellite (< 1m)	**	**	**	**
Aerial images	*	**	***	***
Aerial LiDAR	*	**	***	***
Drone images	*	*	***	***

Sentinel-2 optical satellite

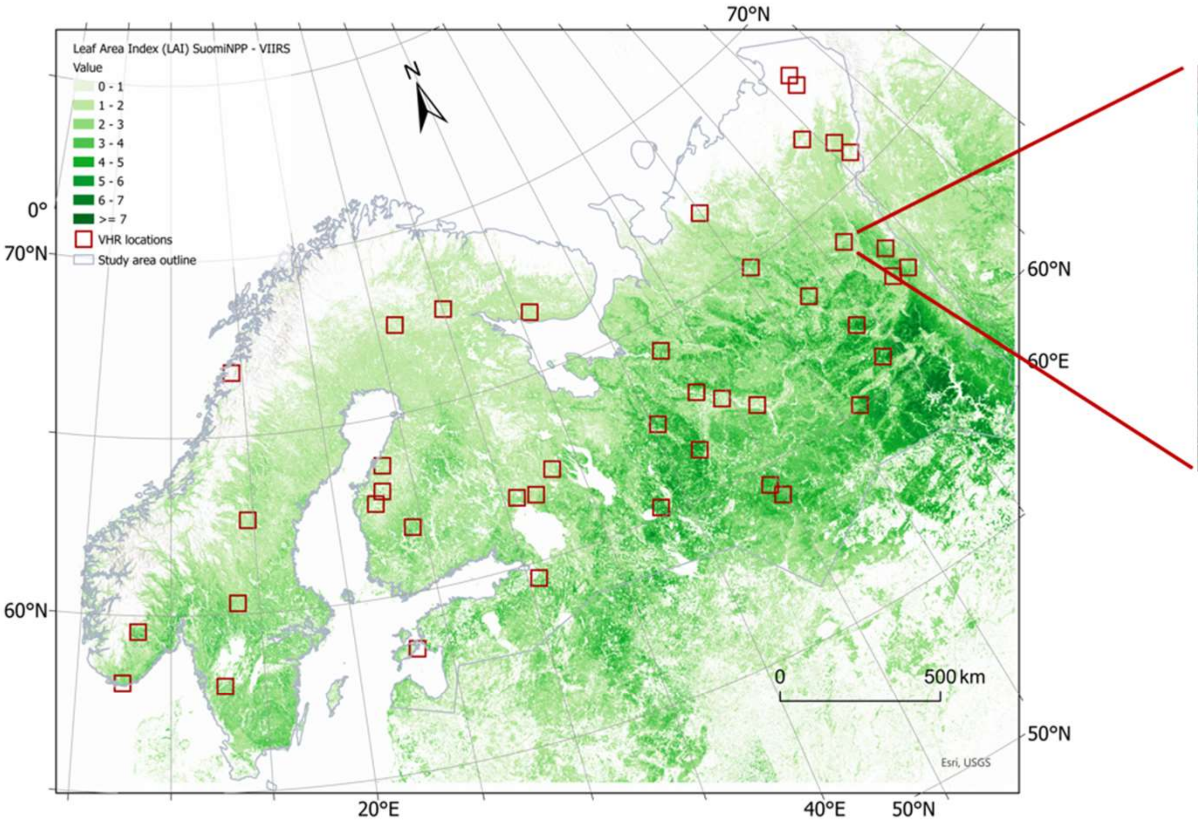
RGB:
NIR
Red
Green
Blue

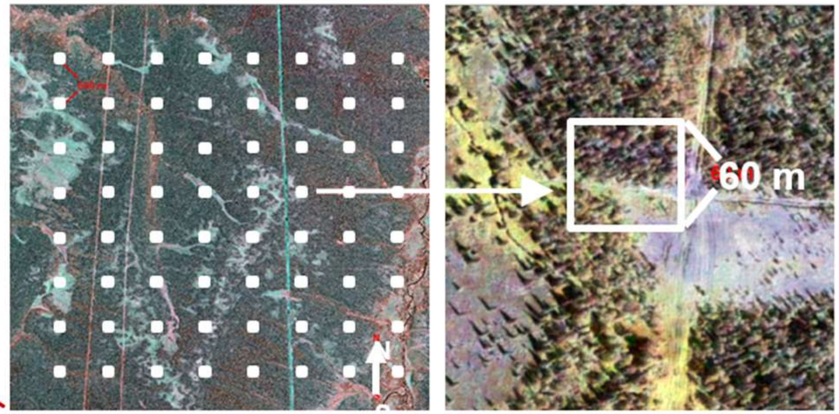
Sentinel-1 C-band SAR satellite

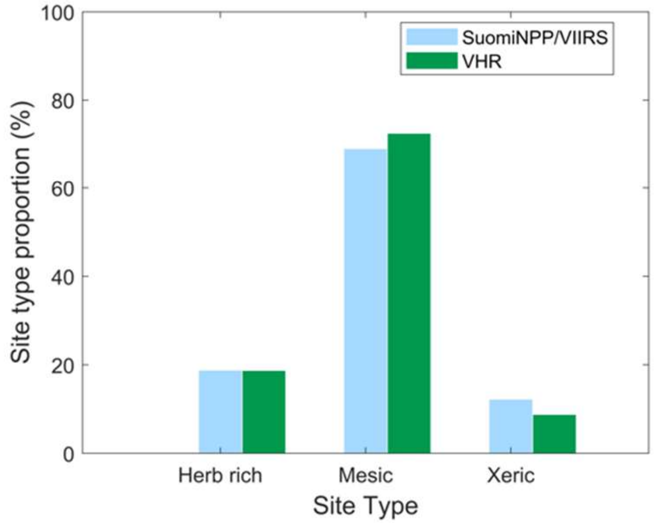


RGB:
VV
VH
VV-VH difference
© ESA

Combining satellite datasets







Häme, T. Astola, H., Kilpi, J., Rauste, Y., Sirro , L., Mutanen, T., Parmes, E., Rasinmäki, J., Imangholiloo, M. (upcoming) Forest Area and Structural Variable Estimation in Boreal Forest Using Suomi NPP VIIRS Data and a Sample from VHR Imagery. *Remote Sensing* (to be submitted)

Deriving information on forest variables from EO data

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

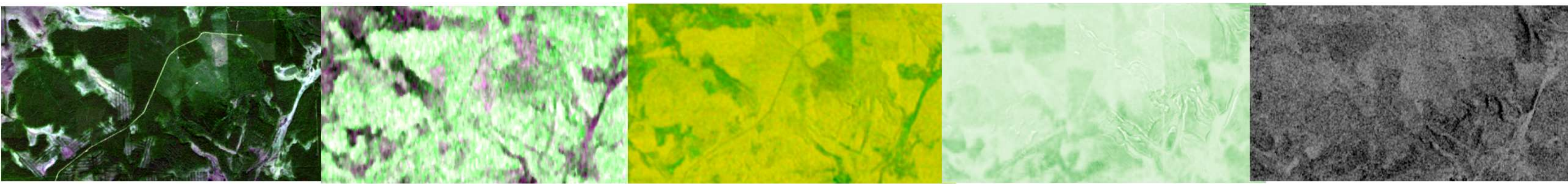
I (m,x,y) = F [..., {target properties}, ...]

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

{target properties} : {spectral properties, water content, roughness, orientation, density, vertical & spatial structure...}

{forest variables} : {height, DBH, species, basal area, growing stock volume, biomass...}

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



Sentinel-2 © ESA | ALOS-2 PALSAR-2 © JAXA and METI, 2015 | Sentinel-1 © ESA | TanDEM-X InSAR CHM | TanDEM-X coherence © DLR

Traditional approaches for prediction/classification

VTT

- **Physics-based and semi-empirical**
  - Reference data used for "calibration"
  - Normally suitable for a given sensor/wavelength
  - WCM (water cloud model), RVoG (random volume over ground)…
- **Statistical parametric models**
  - Model fitting is used, reference data for teaching
  - Often don't care about "nature" of EO data
  - MLR (multiple linear regression)…
- **Non-parametric approaches**
  - Completely dependent on reference data
  - Normally don't care about "nature" of EO data
  - kNN (k nearest neighbours), SVR (support vector regression), RF (random forests)…
- **Semi-supervised approaches**
  - Utilize EO data even when reference data are missing
  - Probability (VTT forest estimation tool)…

- Traditional approaches work well for many purposes
- Deep Learning (DL) methods gain popularity:
  - Effectively including spectral-spatial and sometimes temporal relationships in modeling process;
  - Learn useful features from the data, without need for manual feature engineering and selection;
  - Effectively handle large amounts of data
  - Often deliver better prediction/classification accuracies in regression and sematic segmentation tasks

Deep learning (DL) models

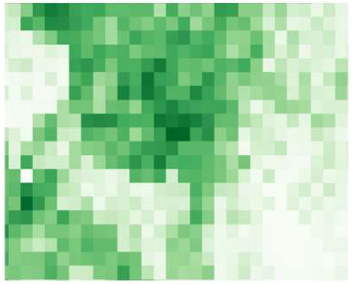
- require large amounts of labeled data for training, which can be time-consuming and expensive to obtain, and often require expert annotation.
- features learned by DL models can be difficult to interpret, thus hard to understand how the model is making its predictions.
- DL models can be sensitive to noise and artifacts in the data, which can lead to errors in the segmentation/regression results, e.g., atmospheric, radiometric artefacts or sensor noise.
- DL models can be computationally intensive, when processing large satellite images.

DL model training:

weakly-supervised labels:

input EO image

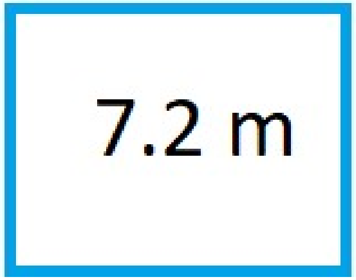
fully-segmented label, e.g., ALS-based data



partial label, e.g., forest stand data

single pixel label, e.g., forest plot data

image-level reference



Further, we will discuss unsupervised, self-supervised, and weakly-supervised approaches to handle such data

beyond
the obvious