Indicators from remote sensing data using machine learning

Janne Mäyrä, SYKE
Andras Balazs, LUKE
25.01.2021

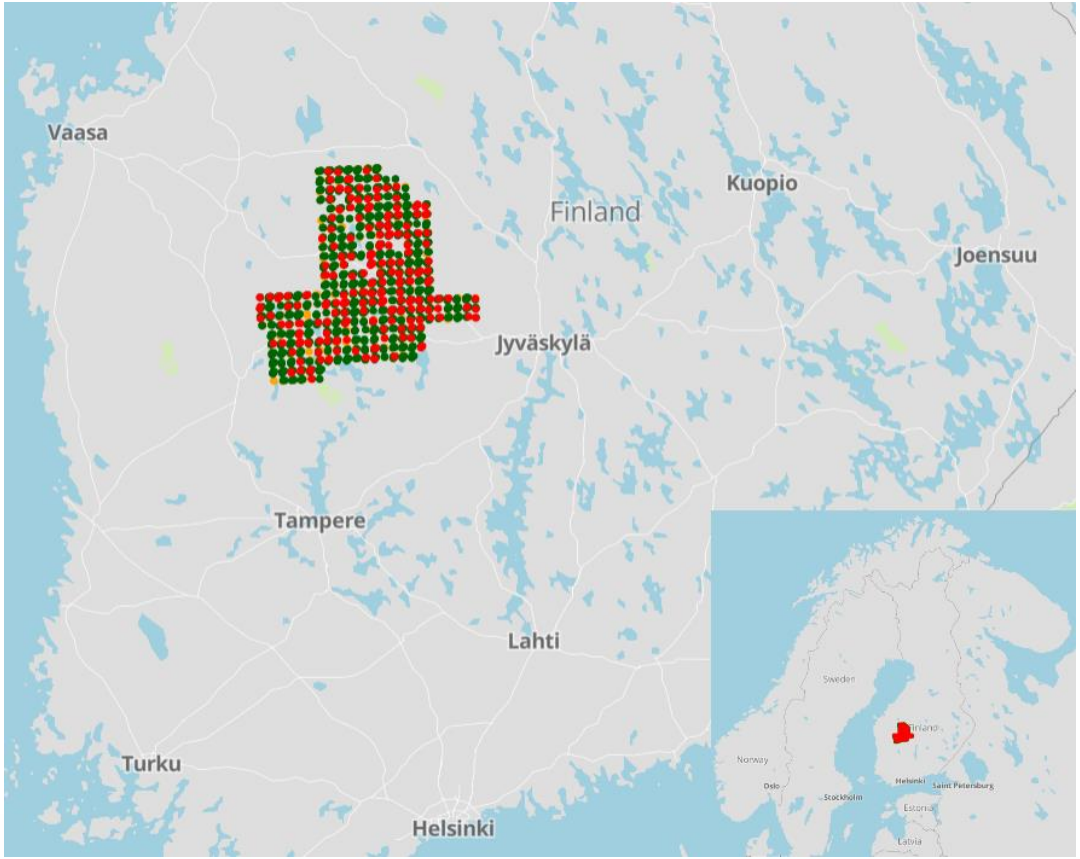
SYKE

Data

- Remote sensing data
  - Aerial false-color images with high ground resolution (0.3m)
  - Airborne laser scanning data with average point density of 1.66 pts/m²
- Ground reference data
  - Around 1500 circular field plots with 9m radius over an area of ca. 5800 km²
  - The reference data includes total and species-wise total growing stock (m³/ha), mean diameter at breast height (cm) and mean height (m)

SYKE

Ground reference data



SYKE

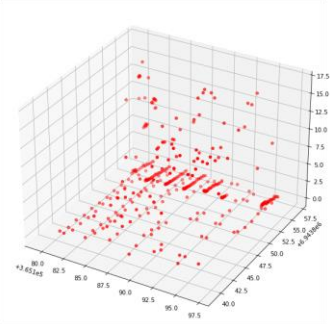
Methods

- k-nearest neighbors combined with genetic algorithm
  - Similar method currently used in Finnish MS-NFI
- Traditional machine learning methods, such as Random Forest (RF) and Artificial Neural Networks (ANN)
- Modern deep learning methods, especially Convolutional Neural Networks
  - CNNs are nowadays the method of choice in different computer vision tasks

SYKE

Data processing for k-NN and traditional machine learning methods

- Traditional methods are not able to process raw data
- From aerial imagery, we extracted optical features as well as textural features (around 100 features total)
- From LiDAR point clouds, around 70 point cloud level metrics were computed

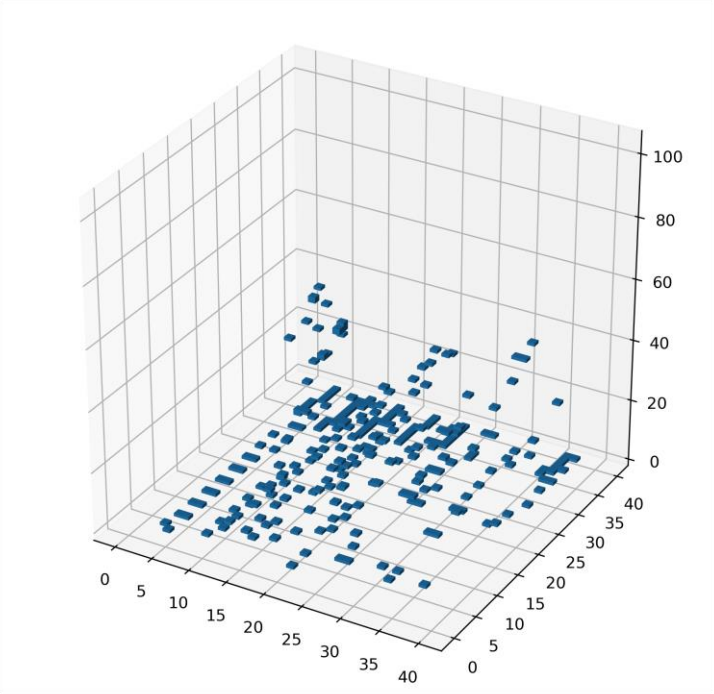


zmax=16.785, zmean=2.63,…,green_max=0.76, red_max=0.85…

SYKE

Data processing for deep learning methods

- The main advantage of deep learning methods is that they are able to extract features from the raw data
- Aerial images can be fed to CNN without any processing
- LiDAR point clouds, however, need to be **voxelized** before feeding to the models

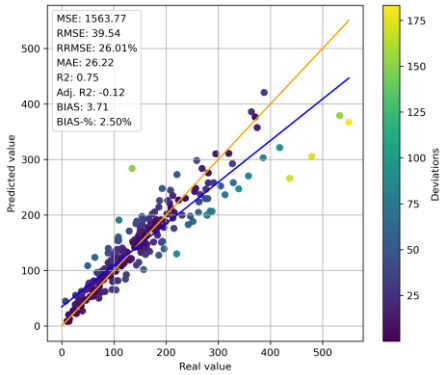


SYKE

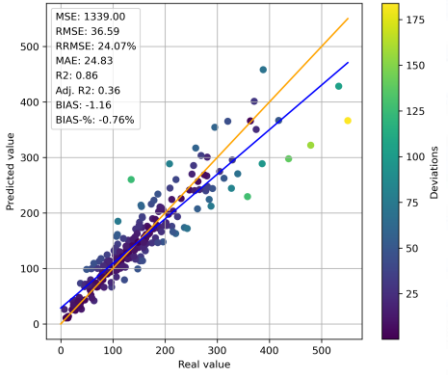
Total volume (m³)

- CNN was able to outperform other methods by using only voxelized LiDAR data
- For larger total volumes, all methods tend to underestimate total volume

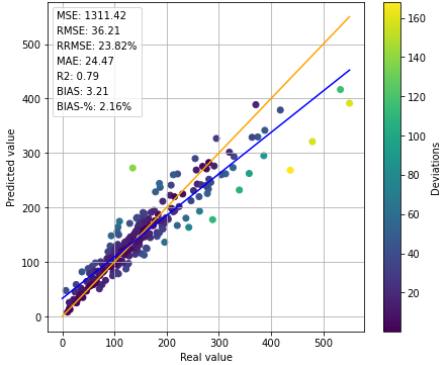
RF



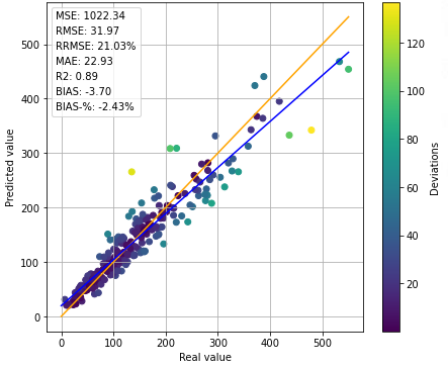
ANN



k-NN



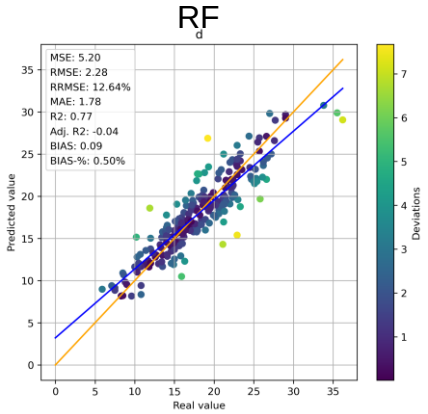
CNN

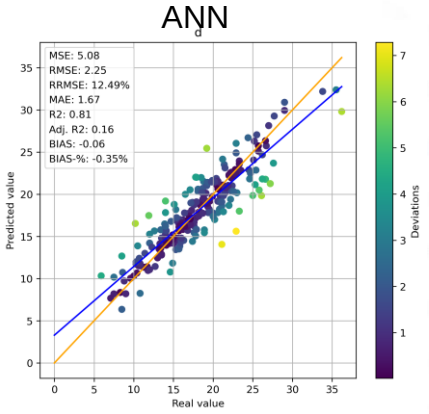


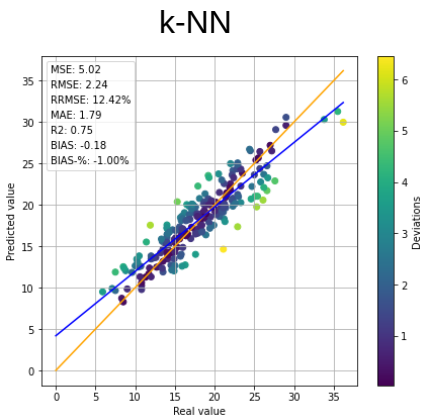
SYKE

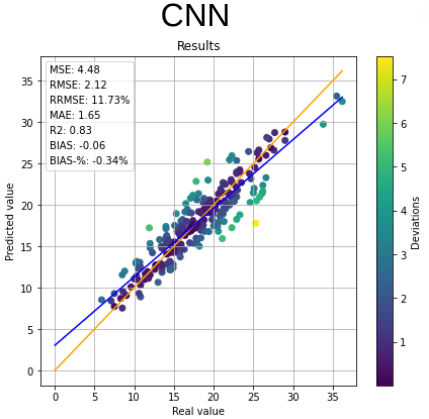
DBH (cm)

- All tested methods achieve almost equal results, with only minor differences between best and worst results







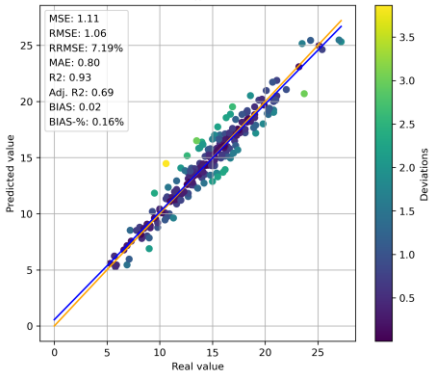


SYKE

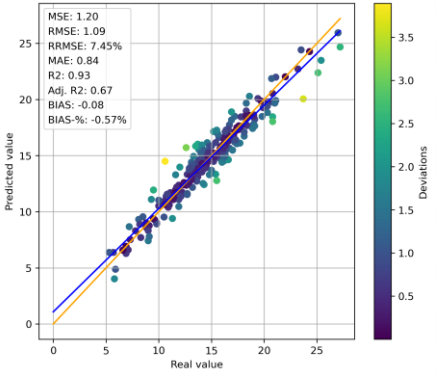
Average height (m)

- Again, no major differences between methods
- However, average height can be modelled from LiDAR data without any complex model

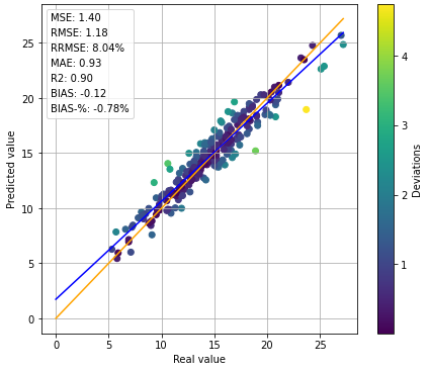
RF



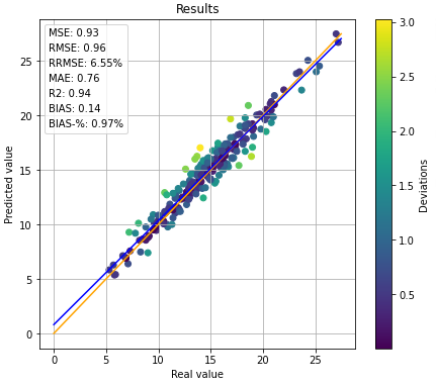
ANN



k-NN



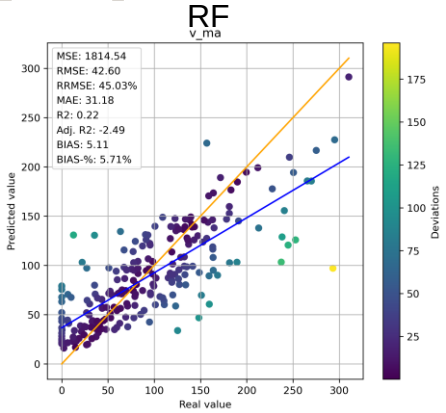
CNN

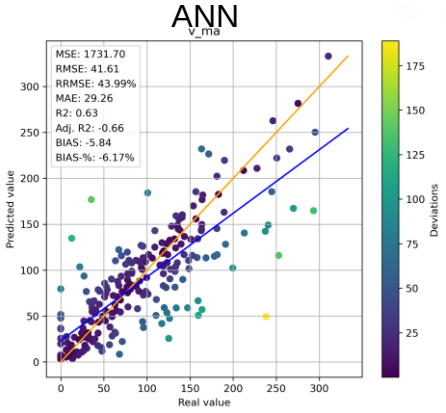


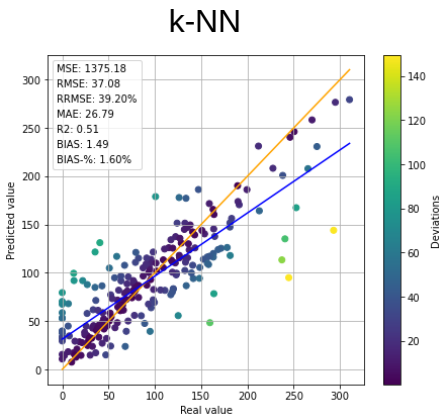
SYKE

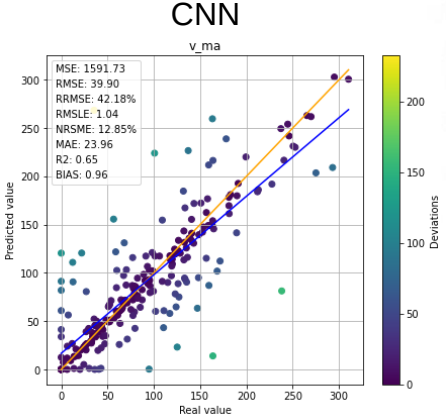
Volume of pine (m³)

- Predictions for CNN were acquired with two-step process
  - Proportions of species from aerial images
  - Species-wise volume from predicted total volume







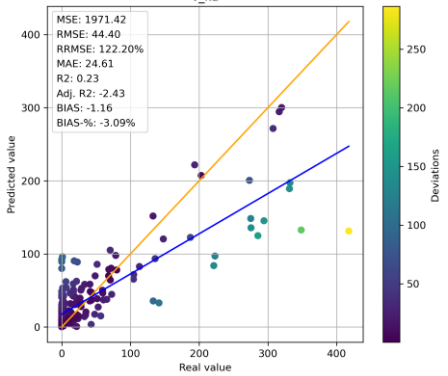


SYKE

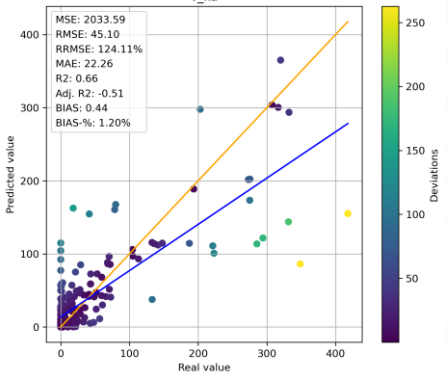
Volume of spruce (m³)

- Predictions for CNN were acquired with two-step process
  - Proportions of species from aerial images
  - Species-wise volume from predicted total volume

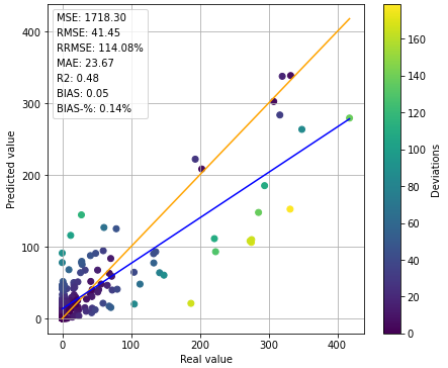
RF



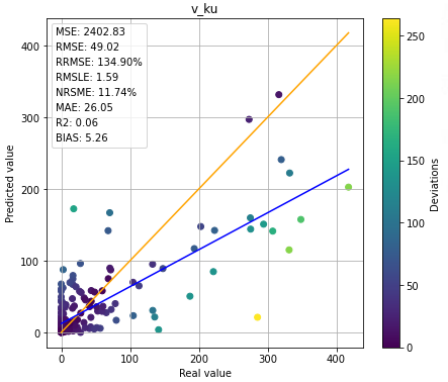
ANN



k-NN



CNN

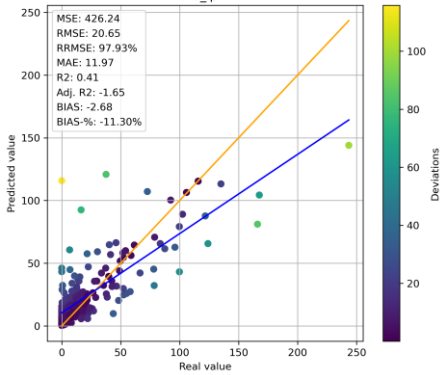


SYKE

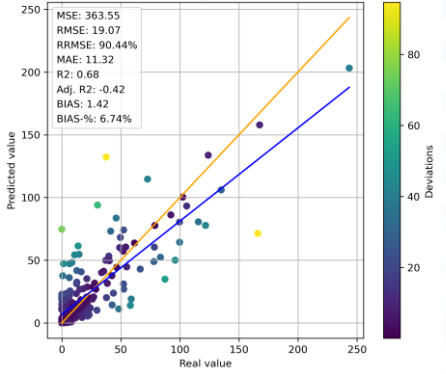
Volume of deciduous trees (m³)

- Predictions for CNN were acquired with two-step process
  - Proportions of species from aerial images
  - Species-wise volume from predicted total volume

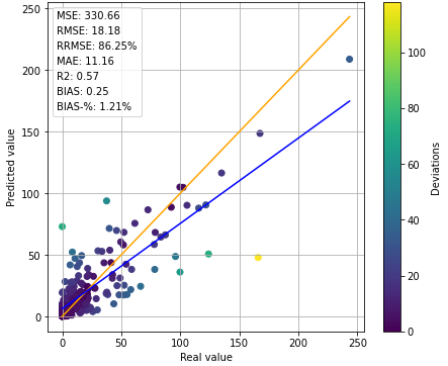
RF



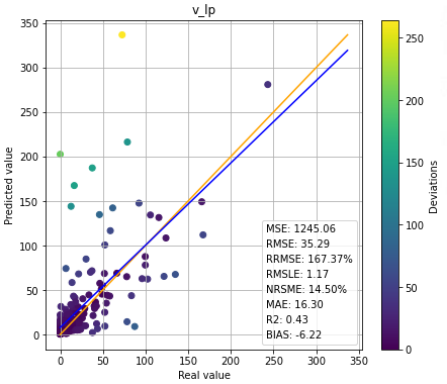
ANN



k-NN



CNN



SYKE

Recap of results (RMSE-%)

	RF	ANN	k-NN	CNN
Volume	26.01%	24.07%	23.82%	21.03%
DBH	12.64%	12.49%	12.42%	11.73%
Height	7.19%	7.45%	8.04%	6.55%
Pine	45.03%	43.99%	39.20%	42.18%
Spruce	122.20%	124.11%	114.08%	134.90%
Deciduous	97.93%	90.44%	86.25%	167.37%

SYKE

Conclusions

- For most of the variables, 3D-CNN achieved the best results
- For other attributes than species-wise volumes, utilizing only LiDAR data yielded similar or better results than aerial imagery or combination of both data sources
- Proportions of different species are the most difficult attributes to model
- Higher laser point density is expected to improve results of CNN methods

SYKE

From forest attributes to ecosystem indicators

- In order to acquire comparable ecosystem indicators estimated forest variables are scaled between 0 and 1
- This enables comparability not only between different areas within the same project but also between projects
- Forest variables useful as indicators e.g.
  - tree species composition
  - mean height as a proxy for age
- Vertical and horizontal structure of forest canopy is one of the key ecosystem indicators, but e.g. current laser data point density is not high enough for extracting those indicators

SYKE