EGU2020: Sharing Geoscience Online

Present and future global vegetation dynamics and carbon stocks from observations and models (8 May 2020)

Utility of hyperspectral remote sensing data in estimating forest structure variables in boreal forests of Finland

Eelis Halme, Petri Pellikka and Matti Mõttus

Introduction

Background

- Estimating Finnish boreal forest variables using machine learning regression algorithms
  - Gaussian process regression (GPR)
  - Support vector regression (SVR)
  - Performance and accuracy?
- Hyperspectral and multispectral remote sensing data
  - AISA imager (128 bands, res. 0.7 m)
  - Sentinel-2 (10 bands, res. 10 m)
  - Additional benefit from higher spectral resolution?
- New forest data from the Finnish Forest Centre
  - Usability of the new stand-level data?

Study objectives

- The specific objectives of the study were:

1) to investigate the estimation accuracy of forest variables in Finnish boreal forest from stand-level data using kernel-based regression algorithms;

2) to study the suitability of newly available Finnish Forest Centre stand-level data for training the kernel-based regression methods for forest variable retrieval;

3) to assess the additional value of hyperspectral remote sensing data, compared with multispectral optical satellite remote sensing data, in estimating forest variables of Finnish boreal forest.

Material

Remote sensing data

- Hyperspectral AISA
  - 128 bands, 400–1000 nm, spectral res. 4.7 nm, spatial res. 0.7 m
  - Airborne flight campaign
- Multispectral Sentinel-2 (S2)
  - 10 bands in VNIR, spatial res. 10 m
  - Level-2A product
- Images from June 2017
- Study site in the southern boreal forest zone

Hyperspectral AISA image that is composed of nine separate flight lines. The covered area is about 3000 ha. Hyytiälä forestry field station (61°50'44"N, 24°17'10"E) in Juupajoki, Finland, is marked with a red dot.

Reference data

- Forest data from the Finnish Forest Centre
  - Stand-level
  - Simulated to the end of 2017
  - Open data (CC BY 4.0)
- Set of independent in situ measurements
  - Plot-level
  - Measured in summer 2013

Variables of interest
• Mean height (m) • Basal area (m^2/ha) • Leaf area index • Stem biomass (t/ha) • Main tree species   • *Basal area and LAI weighted*

Main tree species

- Categorical value
  - How to estimate using regression algorithms?

  → Variables quantifying basal area and LAI per species within each stand

  → Predictions for the main tree species were calculated from the estimated species-specific values

- Two different versions based on species dominance
  1. Using all stands
  2. Using stands with 75% species dominance present

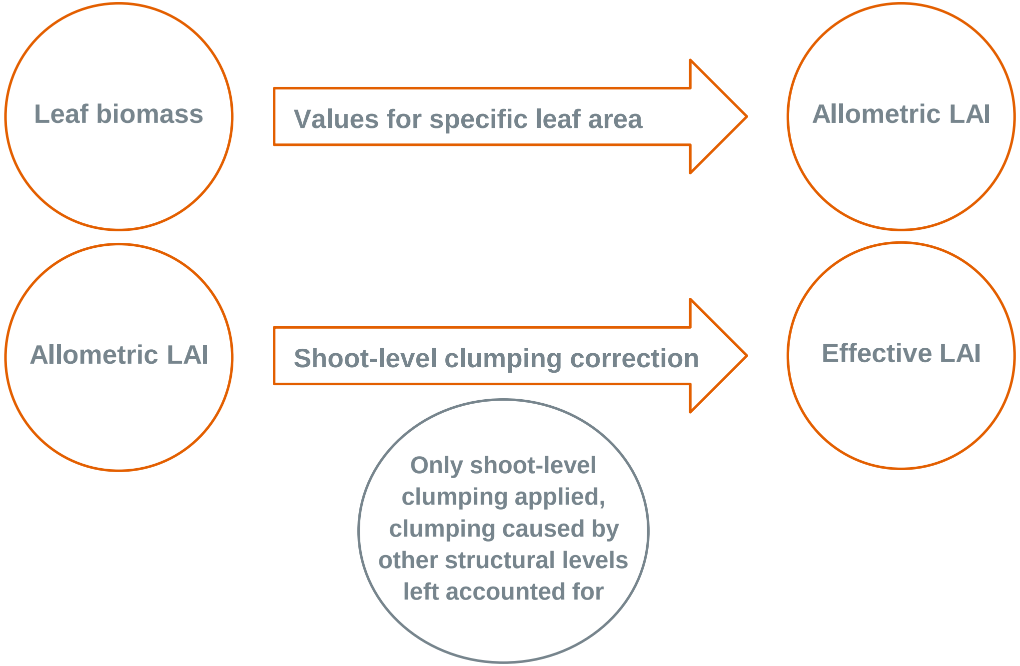
Species-specific variables in a stand

- *Pine* basal area (m^2/ha)
- *Spruce* basal area (m^2/ha)
- *Broadleaved tree* basal area (m^2/ha)

- *Pine* LAI
- *Spruce* LAI
- *Broadleaved tree* LAI

Stand leaf area index (LAI)

- Effective LAI was computed from stand leaf biomass using allometry and species-specific values from literature



Methods

Algorithms

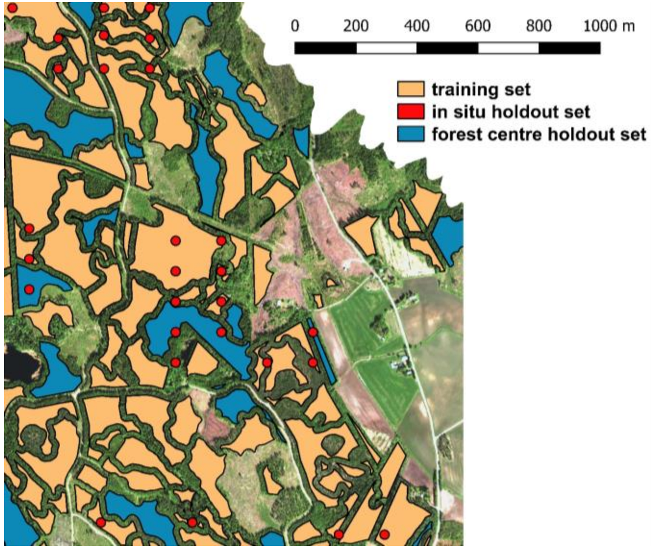
- Implementation in scikit-learn
- Gaussian process regression (GPR)
  - Kernel and probabilistic approach for solving regression task
  - Our kernel included a radial basis function (RBF) and a white noise kernel
- Support vector regression (SVR)
  - Kernel function and ε-insensitive error function for solving regression task
  - We used RBF kernel
- 5-fold cross validation in hyperparameter tuning

`sklearn.gaussian_process.GaussianProcessRegressor`

`sklearn.svm.SVR`

Datasets

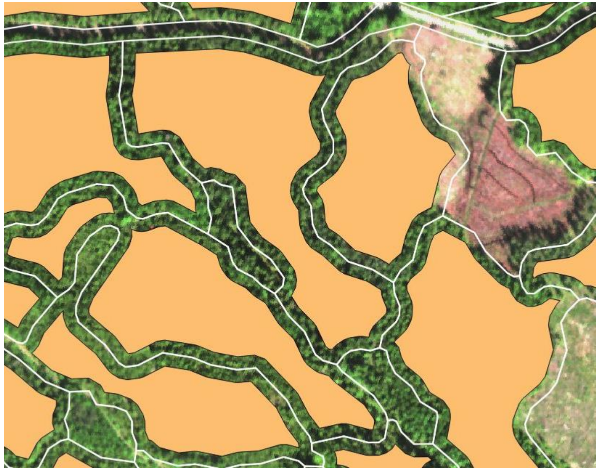
- Training set
  - Stand-level data (Forest Centre)
  - 625 stands
- Two holdout sets
  - Stand-level (Forest Centre)
  - Plot-level (in situ)
  - 120 geometries



A small subset of the stands and in situ plots. The stands were downsized with 10 m buffer and the plots have a diameter of 30 m.

Applying regressions

- Target values
  - Forest variables of interest
  - One regression per one variable
- Feature values
  - Zonal mean reflectances
  - Reflectances averaged within stands or plots
  - Different values for AISA and S2 images
- Accuracy assessment
  - Root-mean-square error (RMSE)
  - Relative RMSE and bias
  - Coefficient of determination (R^2)



Original stands (white) were downsized with 10-meter buffer (orange) before calculating the zonal mean reflectances to avoid spectral mixing.

Results

Stand-level evaluation

- Good accuracies, differences small between algorithms and datasets
- GPR generally more accurate, however, SVR was faster

Table: Relative and absolute root-mean-square errors of **the most accurate forest variable estimations** for SVR and GPR algorithms on **stand-level**. AISA and Sentinel-2 correspond to the remote sensing images that were used to compute the stand zonal mean reflectances to the used datasets. Also, relative bias and coefficient of determination (R^2) are given.

Forest variable	Algorithm & dataset	RMSE%	RMSE	Bias%	R^2
mean height	GPR & AISA	15 %	2.60 m	-2 %	0.66
basal area	SVR & AISA	17 %	3.84 m^2/ha	-2 %	0.68
LAI	GPR & AISA	20 %	0.66	0 %	0.83
stem biomass	GPR & Sentinel-2	28 %	21.46 t/ha	-2 %	0.63

Plot-level evaluation

- Less accurate estimations
- In situ plots lacked similarity with the stand-level training data
  - Relative biases were high

Table: Relative and absolute root-mean-square errors of **the most accurate forest variable estimations** for SVR and GPR algorithms on **plot-level**. AISA and Sentinel-2 correspond to the remote sensing images that were used to compute the zonal mean reflectances to the used datasets. Also, relative bias and coefficient of determination (R^2) are given.

Forest variable	Algorithm & dataset	RMSE%	RMSE	Bias%	R^2
mean height	GPR & Sentinel-2	37 %	6.14 m	11 %	0.21
basal area	GPR & AISA	45 %	8.66 m^2/ha	25 %	0.12
LAI	GPR & AISA	65 %	1.65	45 %	0.33

Main tree species

- Overall accuracies on stand-level rather high

	AISA		Sentinel-2	
Main tree species	**SVR (%)**	**GPR (%)**	**SVR (%)**	**GPR (%)**
Stand-level				
Species dominance ignored				
based on basal area	90	**93**	88	87
based on LAI	88	87	83	84
Over 75% dominance present				
based on basal area	94	98	88	90
based on LAI	**100**	98	93	95
Plot-level				
based on basal area	**71**	69	68	68

Discussion and Conclusions

Meeting study objectives

- Higher spectral resolution can have a positive influence on accuracy
  - Especially with variables related to species-specific information (e.g, LAI)
  - Spatial resolution has smaller effect
- Estimation accuracies on stand-level were surprisingly good
  - Especially for mean height and basal area
  - These variables are known to have good accuracy also in the original forest data
- The new stand-level forest data seemed suitable for forest variable retrieval
  - High potential in further remote sensing applications
  - E.g., developing or parameterizing global vegetation models

Future improvements

- Plot-level data remains difficult to upscale to stand-level with reasonable accuracy
  - Model transferability between different scales of data difficult also for machine learning algorithms
- Further research needed
  - How to scale plot-level data to stand-level?
  - How to use plot- and stand-level data interchangeably in forestry applications?

beyond the obvious

Eelis Halme
eelis.halme@vtt.fi

www.vtt.fi