SEGMENTED OR POOLED MODELS?

Machine learning estimation of stem volume from Lithuanian NFI, LiDAR and Sentinel-2 data

VYTAUTAS MAGNUS UNIVERSITY
Faculty of Informatics

AUTHORS: Arnas Matusevičius · Monika Papartė · Gabrielė Kasputytė · Gintautas Mozgeris · Tomas Krilavičius

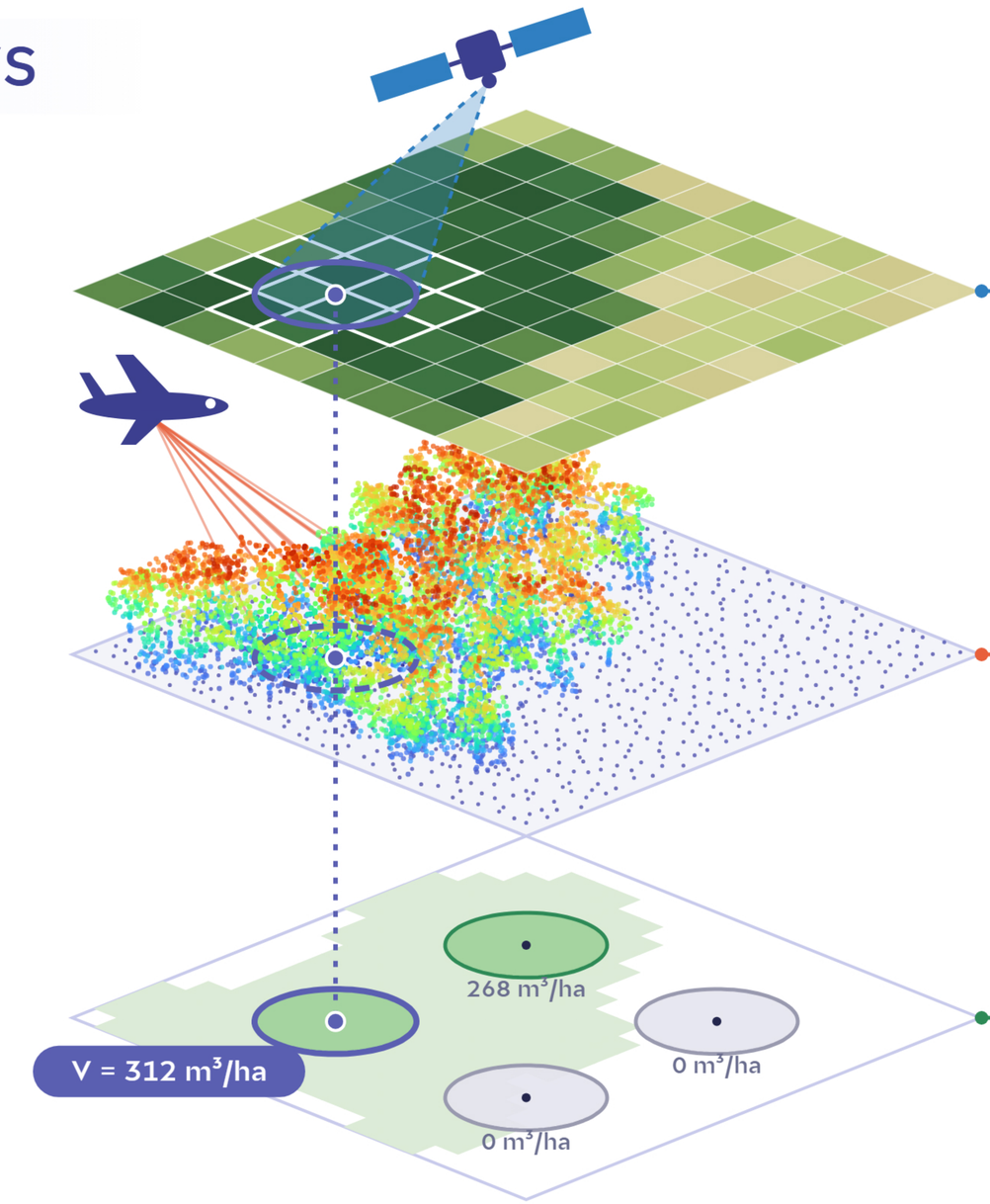
arnas.matusevicius@vdu.lt

forest 4.0

1 ONE PLOT, THREE VIEWS

Lithuania is about one-third forest. The **National Forest Inventory (NFI)** measures stem volume precisely, but only on sample plots. **Airborne laser scanning (ALS)** and **Sentinel-2** cover the whole country.

Can machine learning carry plot volume onto every hectare, and should forests be **split into segments** to do it?

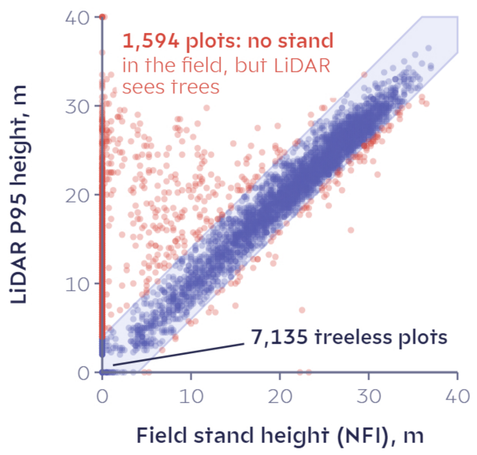


HOW IT FITS TOGETHER

1. **NFI plots** give measured stem volume, m^3/ha
2. **ALS + Sentinel-2** describe every plot with 150 features
3. **Machine learning** learns the link and maps volume everywhere

1 in 3 plots has trees, matching Lithuania's ~33% forest cover. The other plots have zero volume.

2 DATA PREPARATION



16,332 table rows

▼ − 3,375 no field volume

12,957 NFI plots

▼ − 2,080
LiDAR and field height differ > 4 m

10,877 plots used

± 4 m rule adapted from Björnberg et al. 2026

EFFECT ON ACCURACY

32% → 26%

rRMSE, forested plots: before → after

Forecasting the volume of these removed plots led to an error of **86%** on them, as the difference between LiDAR and NFI calculations differed greatly.

3 SEGMENTED OR POOLED?

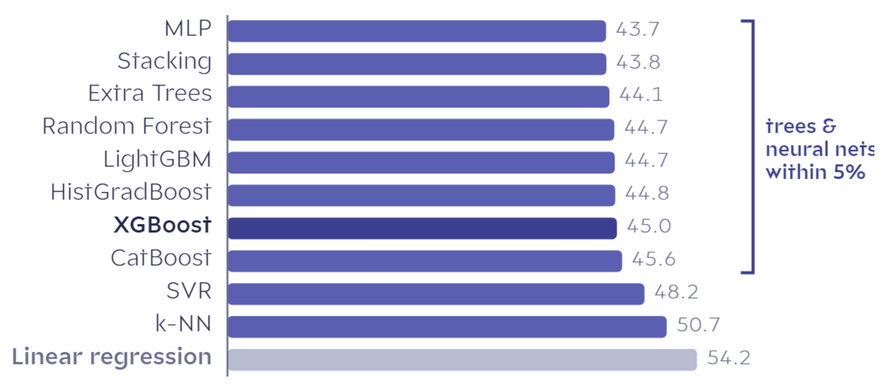
Split the plots into segments with one model each, or pool them into one model?



4 ALGORITHMS & ACCURACY

Linear regression vs 10 ML methods

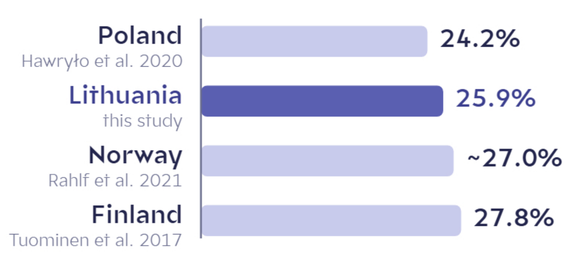
RMSE on all plots, m^3/ha · 50× repeated hold-out



Forested plots vs other NFIs

relative RMSE · CV grouped by plot clusters

−15% RMSE, XGBoost vs linear regression



5 CONCLUSIONS

POOLED.

One model for all plots was as accurate as segmented ones, so far.

Clean reference data matters

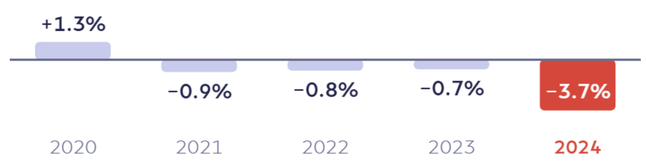
Dropping the 16% of plots where LiDAR and field disagree cuts the error from 32% to 26%.

Algorithm choice matters little

Tree-based and neural models are within 5% of each other.

Time is the real problem

Field and LiDAR years differ by up to 5 years: plots measured in 2024 come out too low. Bias by held-out NFI year, % of mean volume:



6 NEXT STEPS

- Species as a predictor in one pooled model
- Growth-adjust plots to the LiDAR year
- Forecast between 5-year NFI visits

FUNDING

- This research has received funding from Horizon Europe Framework Programme (HORIZON), call Teaming for Excellence (HORIZON-WIDERA-2022-ACCESS-01-two-stage) - Creation of the centre of excellence in smart forestry "Forest 4.0" No. 101059985".
- This research has been co-funded by the European Union under the project "FOREST 4.0 - Center of Excellence for the development of a sustainable forest bioeconomy", No. 10-042-P-0002.

Funded by the European Union

Co-funded by the European Union