Blue Ridge SAF Chapter Meeting 2025

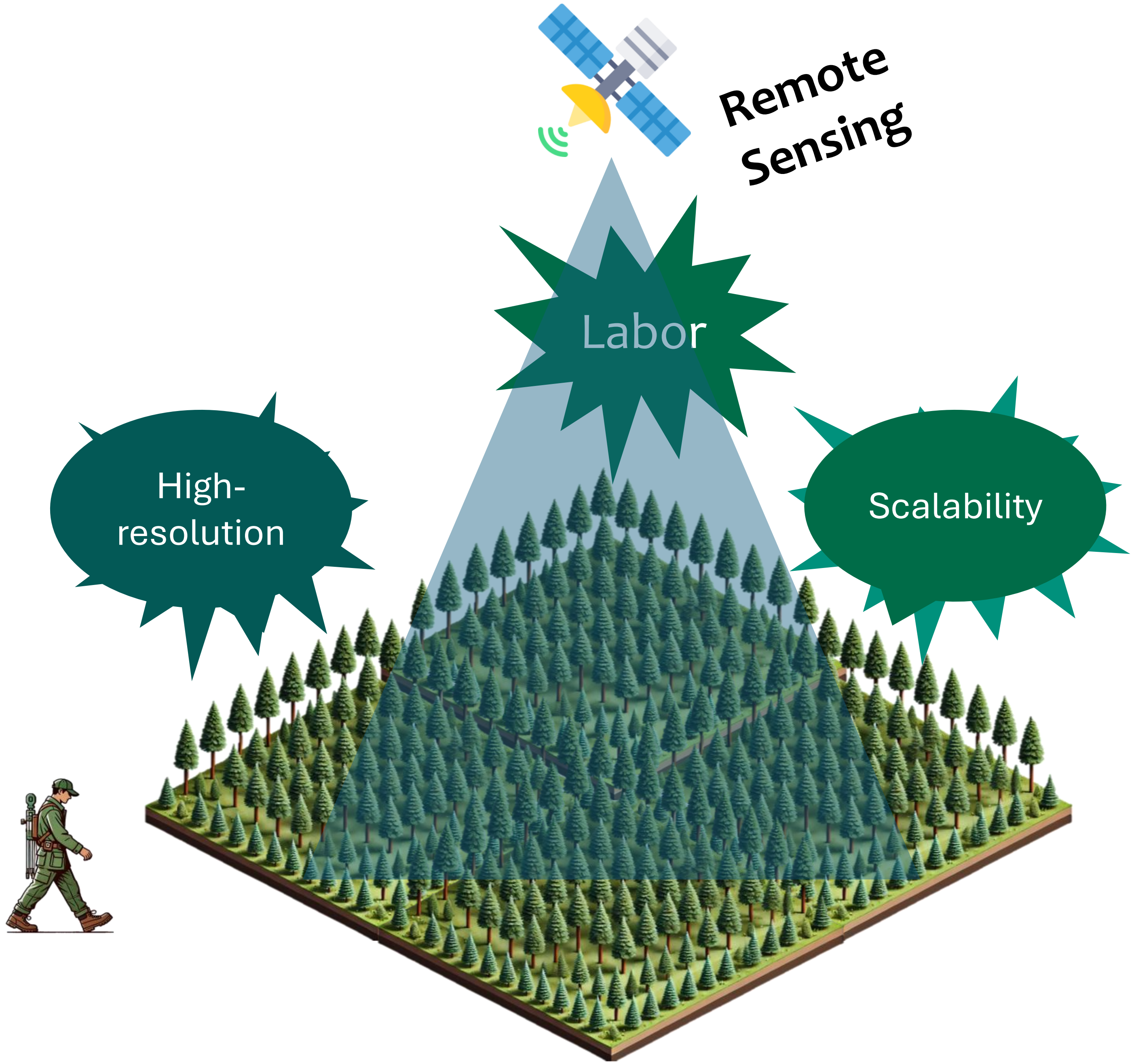
Remote Sensing in AI: Predicting yield of *Pinus taeda* (L.) plantations with LiDAR and Satellite Data

Gunjan Barua

Ph.D. Candidate

Forest Resources & Environmental Conservation

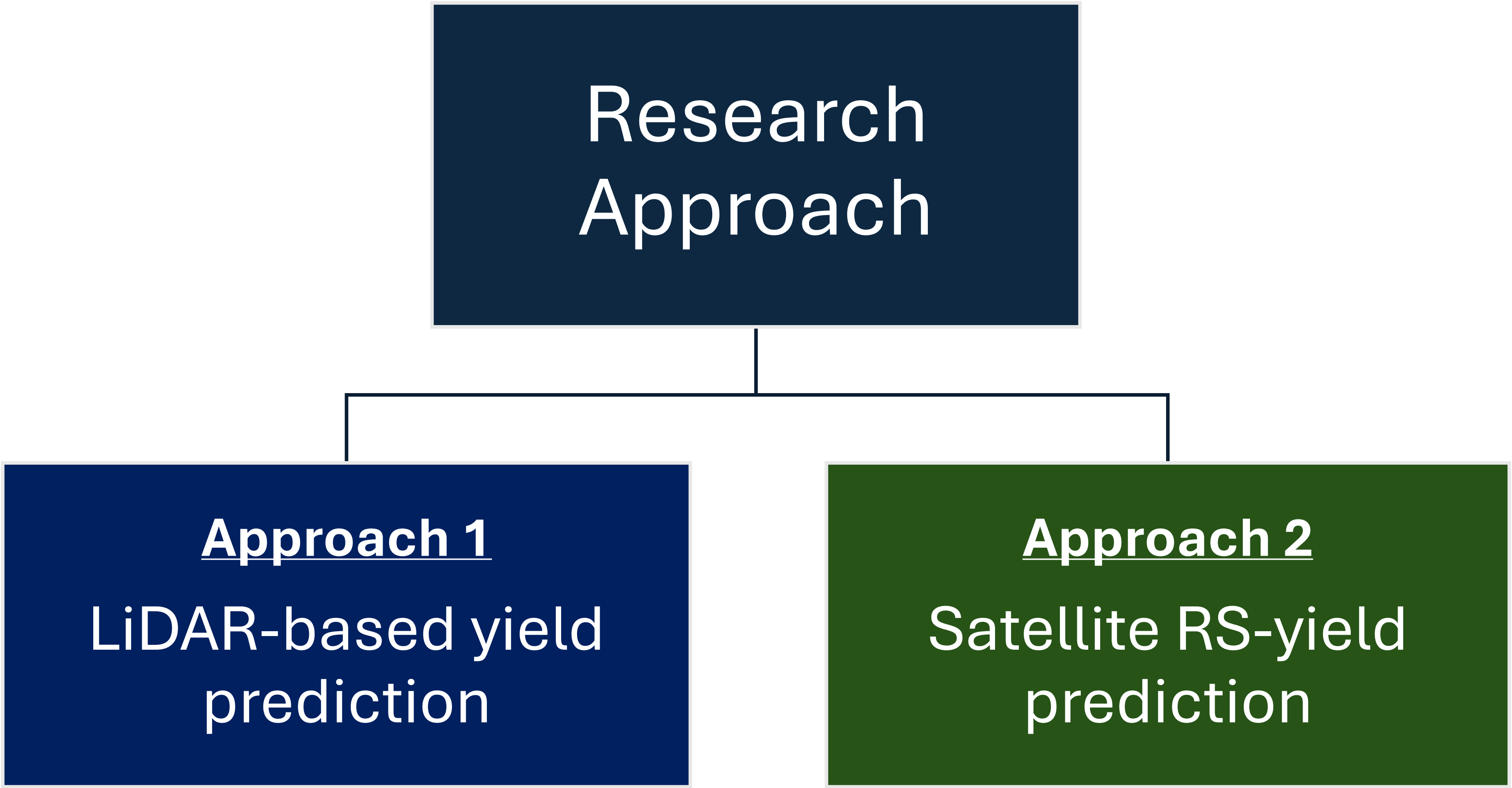
VIRGINIA TECH

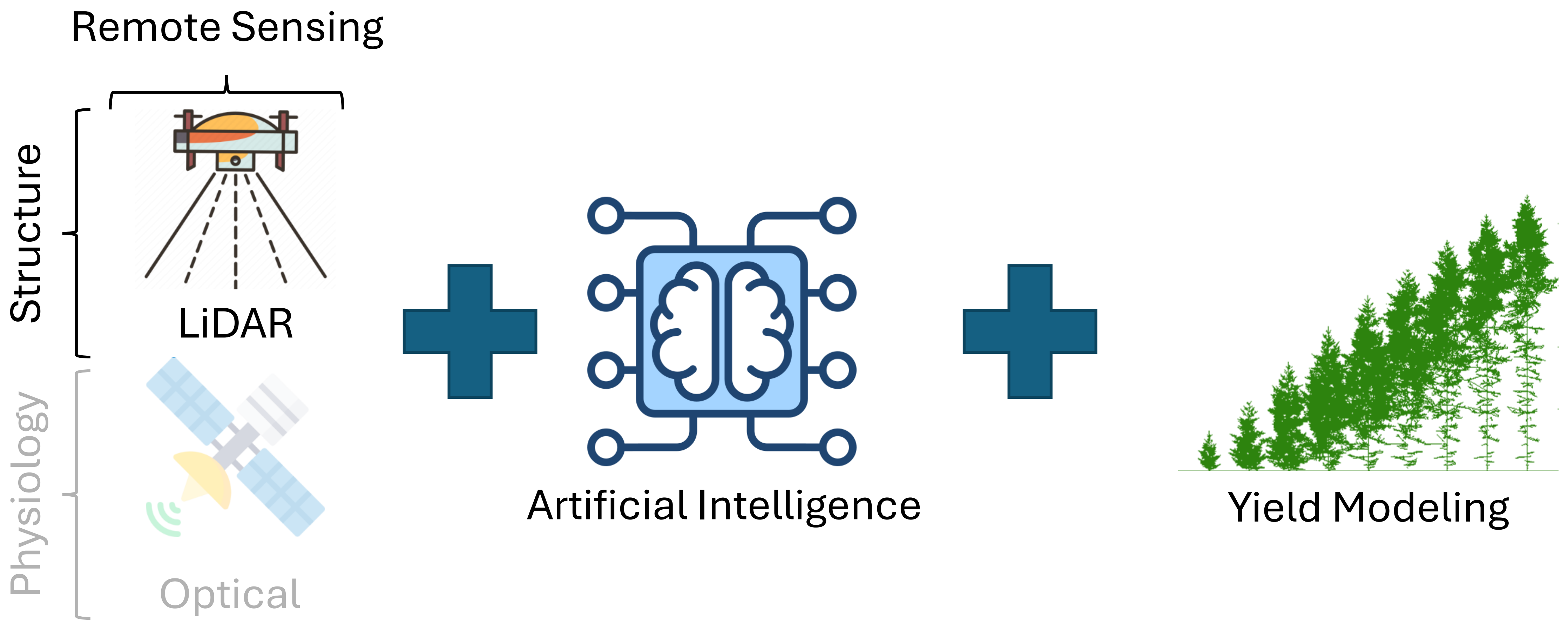


Remote Sensing
Labor
High-resolution
Scalability

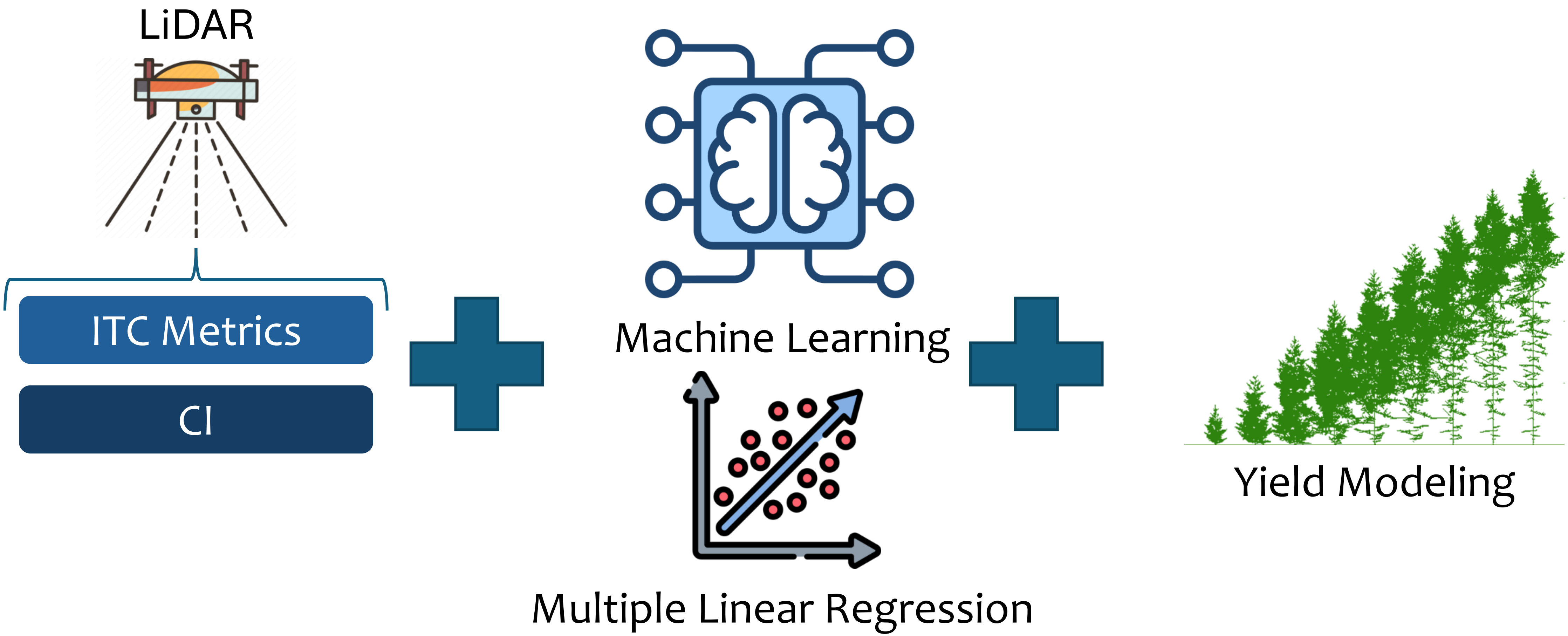
Research Questions

- Can LiDAR alone or multispectral satellite data alone be used for predicting the yield of *Pinus taeda* (L.) plantations?
- If machine learning and/or neural network models are used, which methods provide the best results?
- Which variables are important for predicting the yield?
- Do planting density/thinning conditions play any role in the accuracy of the prediction results?


Research Approach
Approach 1
LiDAR-based yield prediction
Approach 2
Satellite RS-yield prediction

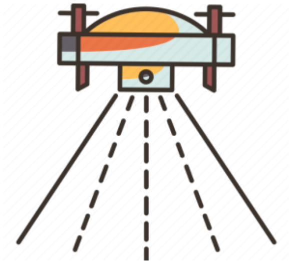

Remote Sensing
Structure
LiDAR
Physiology
Optical
Artificial Intelligence
Yield Modeling

Approach 1: LiDAR



Approach 1: LiDAR

LiDAR



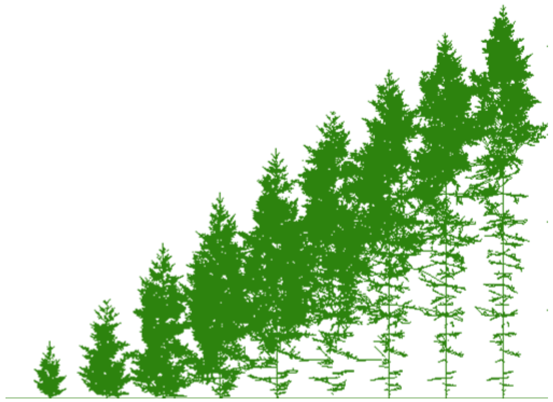
ITC Metrics

CI

Random Forest

Support Vector Regression

Multiple Linear Regression

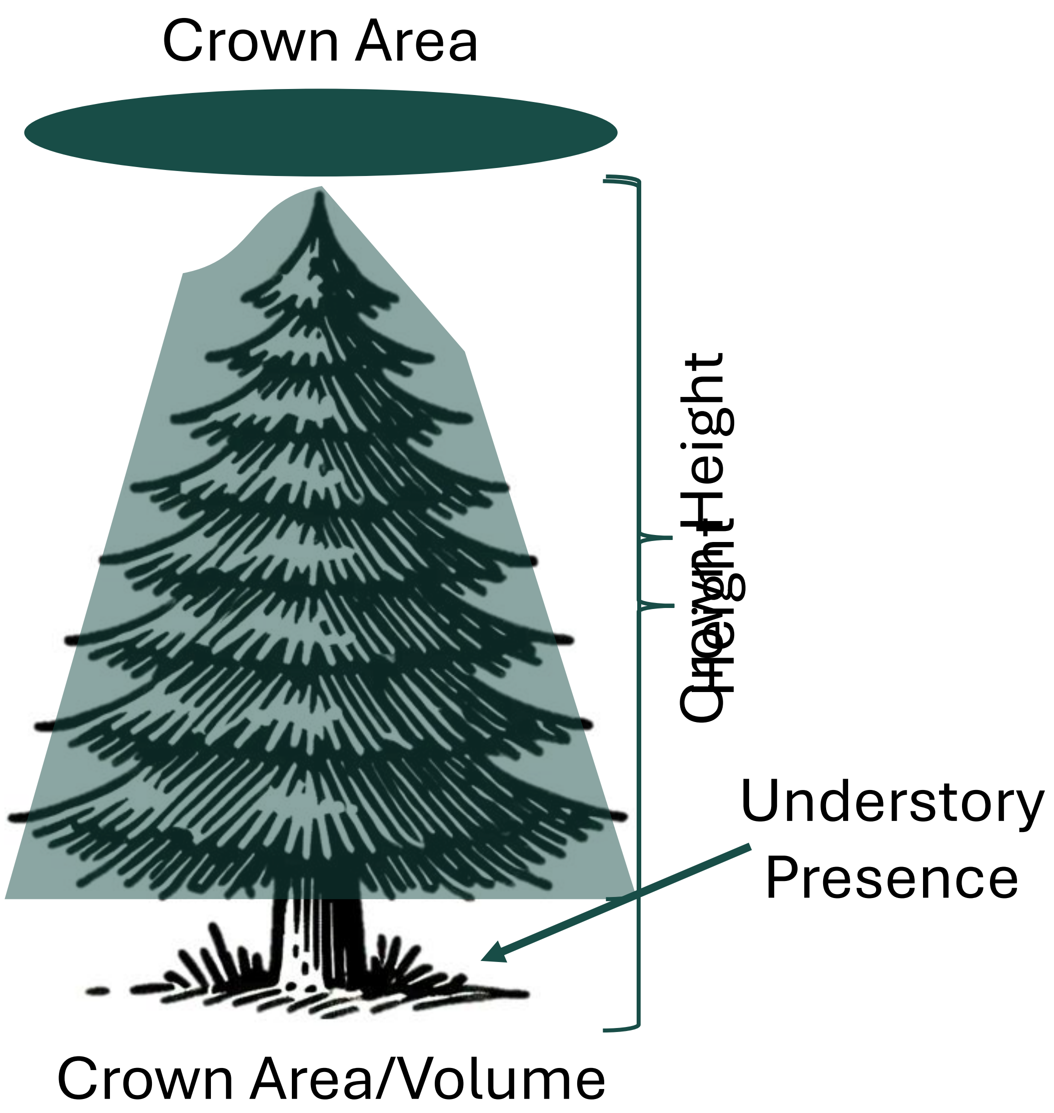


Yield Modeling

Approach 1: LiDAR

Individual Tree Crown Metrics

Three-dimensional characteristics of individual tree crowns obtained

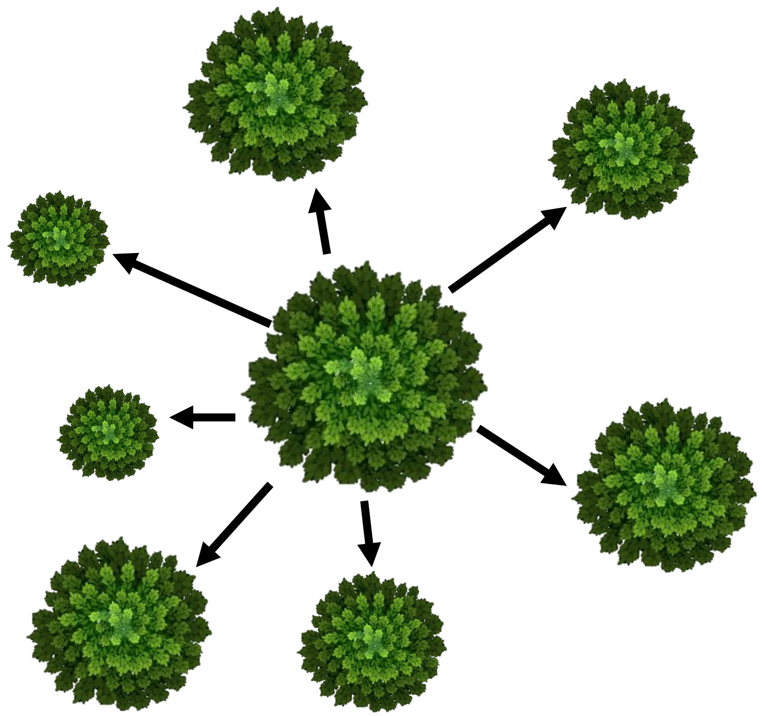


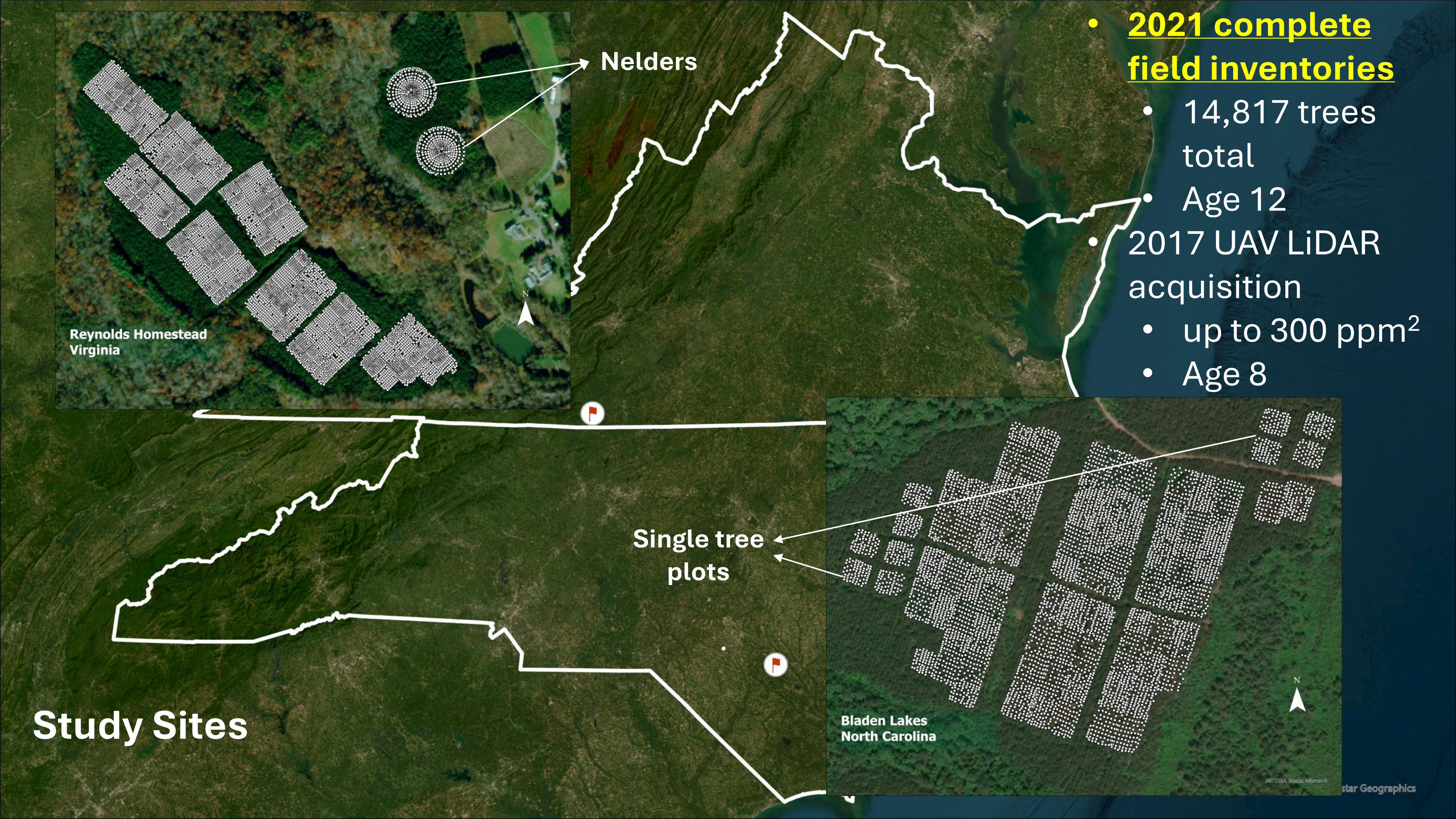
Approach 1: LiDAR

Distance-dependent Competition Indices

A quantitative measure used to assess the competitive interactions among trees in a forest stand

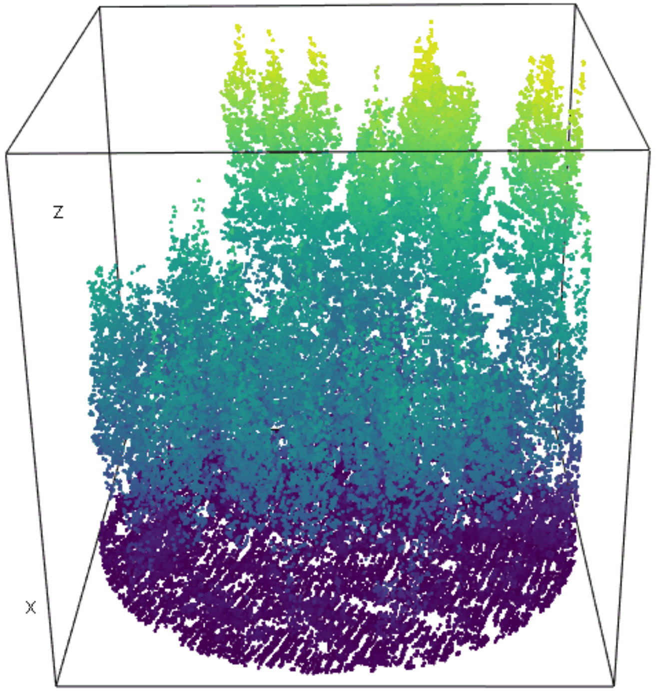
Important for understanding how tree growth, survival, and yield are influenced by neighboring trees.





Study Sites
Nelders
Reynolds Homestead
Virginia
Single tree
plots
Bladen Lakes
North Carolina
N
N
• 2021 complete field inventories
• 14,817 trees total
• Age 12
• 2017 UAV LiDAR acquisition
• up to 300 ppm²
• Age 8

Approach 1: LiDAR



Individual tree crown delineation and variable preparation

44 Variables

Individual Tree Crown (ITC) metrics

Competition Indices (CI)

Data collected in 2017

Field data from 2021

Approach 1: LiDAR

List of Variables (n =44)

Lidar generated ITC predictors:

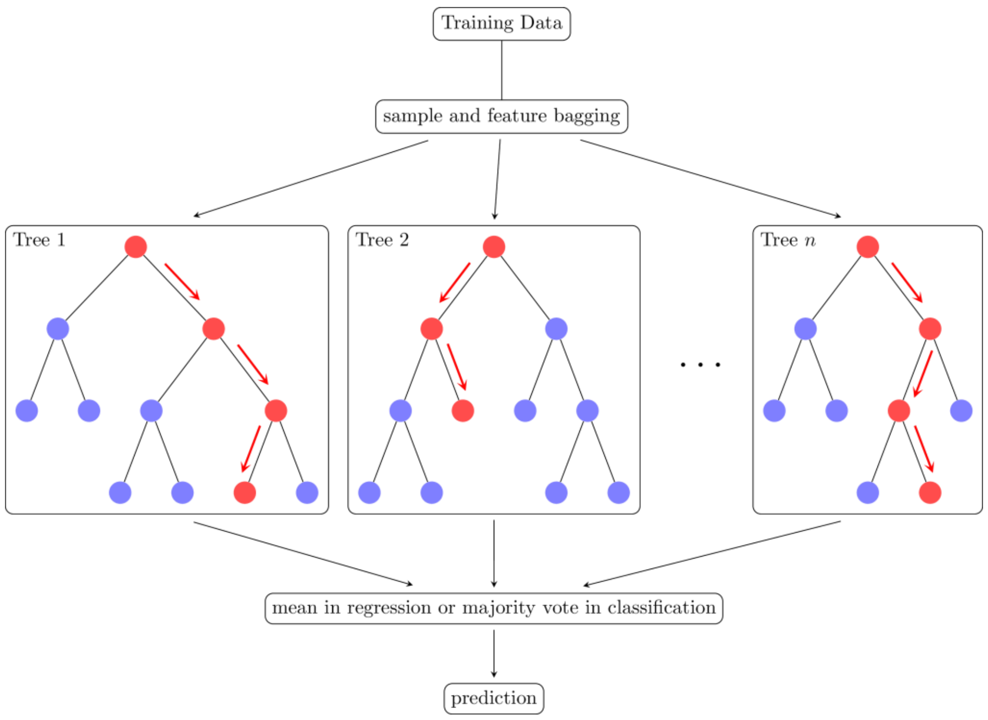
- Height
- Crown Diameter
- Crown Area
- Height to Live Crown (HTLC)
- Single-tree LAI
- Crown Volume (top 10%, 20%, 30%, 40%, 50%)
- Crown Surface Area (top 10%, 20%, 30%, 40%, 50%)
- Understory present (1 or 0)
- Understory Proportion (Canopy Returns/Understory Returns)
- Number of neighboring trees

Lidar generated CI predictors:

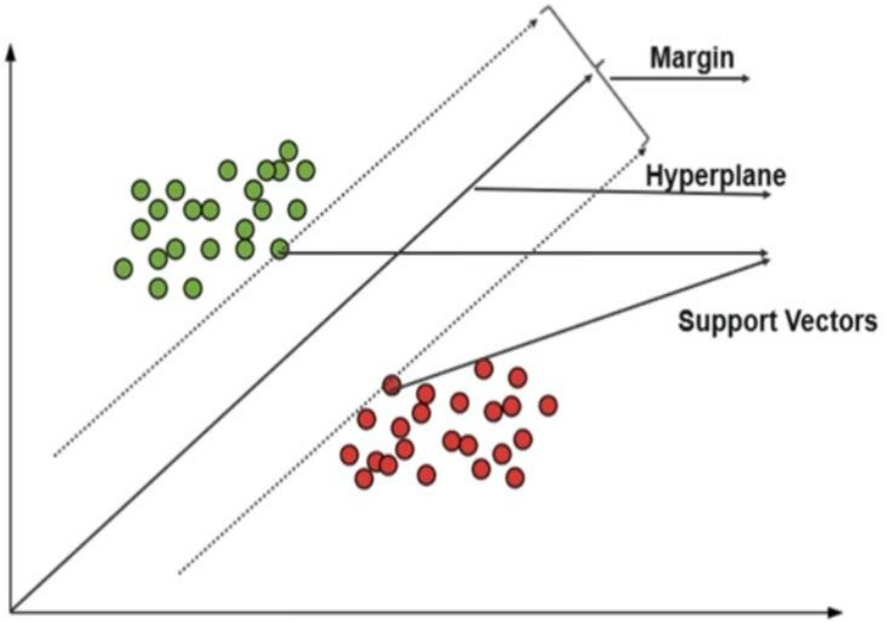
- CI from the crown area (convex-hull)
- CI from the crown area (alpha-hull)
- CI using tree height
- CI using maximum crown diameter
- CI using LAI
- CI using height to the live crown
- CI using understory presence
- CI using understory proportion
- CI using crown volume (top 10%, 20%, 30%, 40%, 50%)
- CI using crown surface area (top 10%, 20%, 30%, 40%, 50%)
- Number of neighbors
- Silva competition index value

Approach 1: LiDAR

Random Forest



Support Vector Machine



Input Data → 80% Training → Validation → Stratified 20% Test

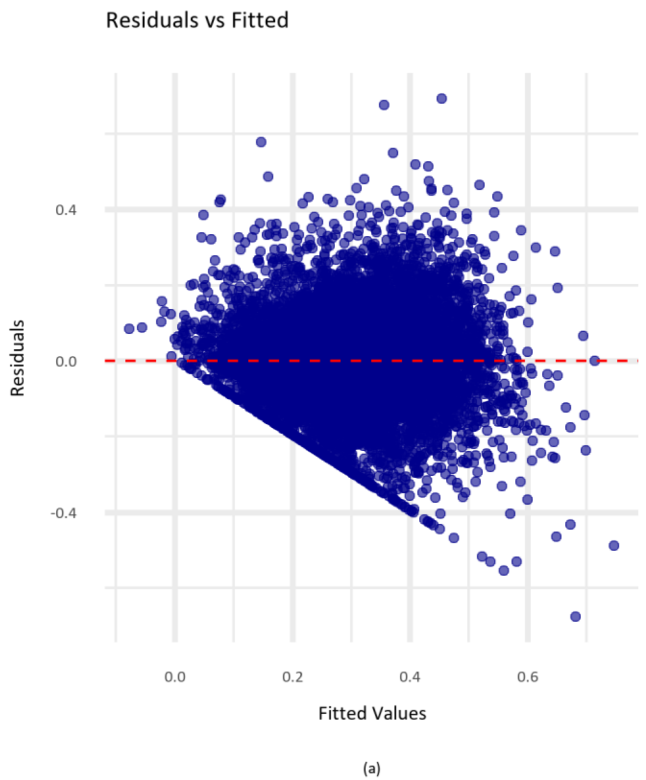
Validation Metrics

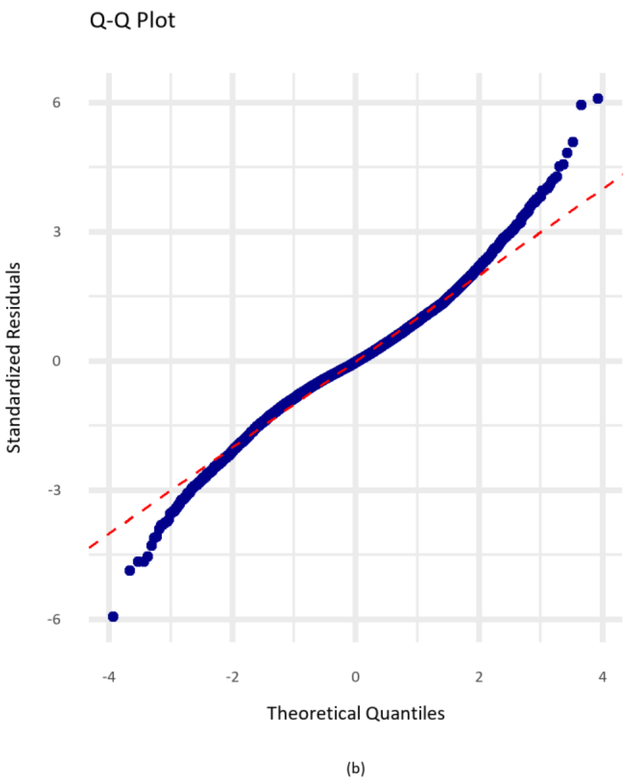
nRMSE	MAE
RMSE	MSE
R^2	

Approach 1: LiDAR

Multiple Linear Regression failed to meet the homoscedasticity assumption

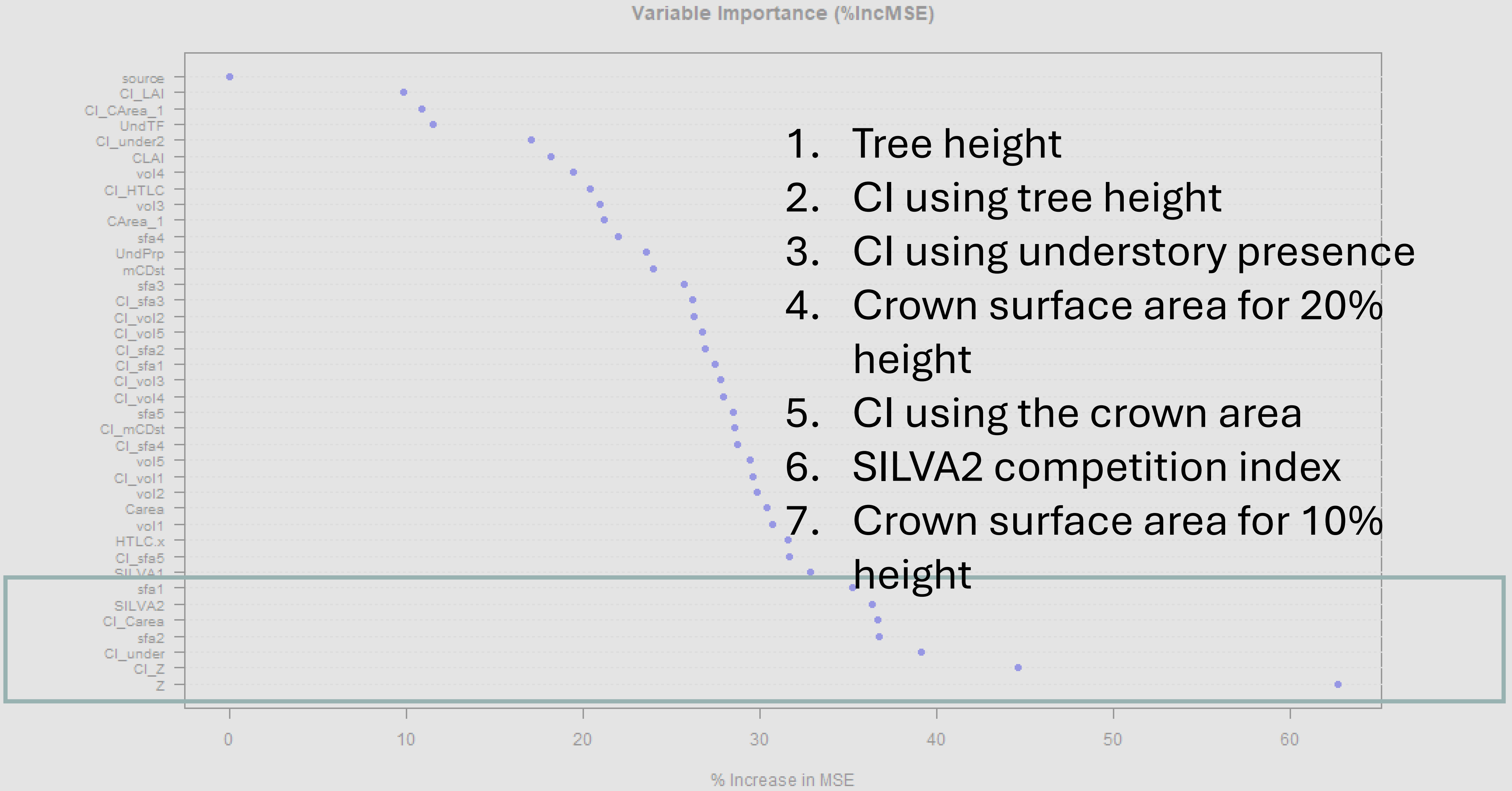
Studentized Breusch-Pagan Test indicates heteroscedasticity (BP = 465.8, df = 7, p < 0.001)





Approach 1: LiDAR

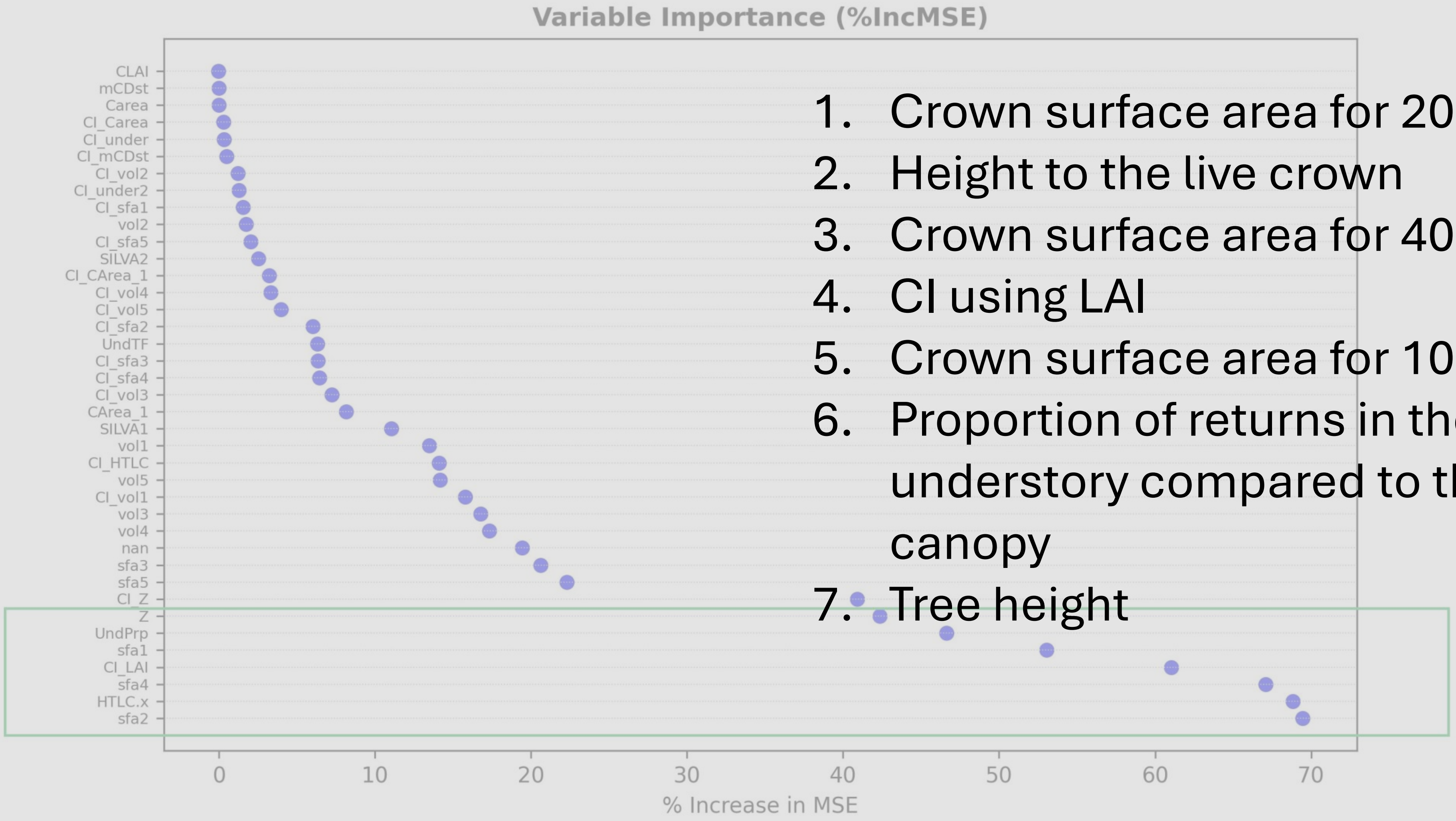
Important Predictors for Random Forest



1. Tree height
2. CI using tree height
3. CI using understory presence
4. Crown surface area for 20% height
5. CI using the crown area
6. SILVA2 competition index
7. Crown surface area for 10% height

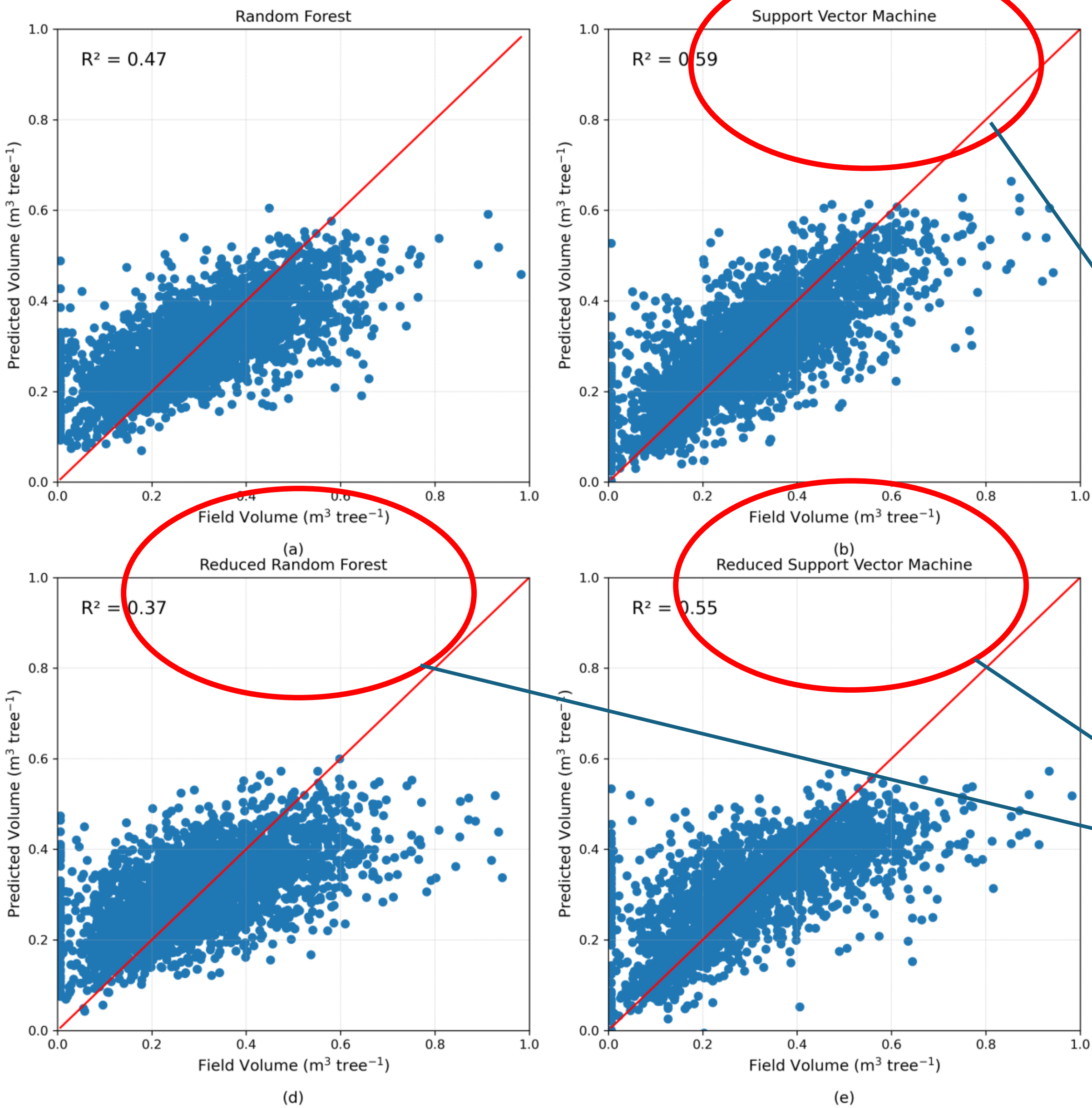
Approach 1: LiDAR

Important Predictors for Support Vector Machine



1. Crown surface area for 20% height
2. Height to the live crown
3. Crown surface area for 40% height
4. CI using LAI
5. Crown surface area for 10% height
6. Proportion of returns in the understory compared to the canopy
7. Tree height

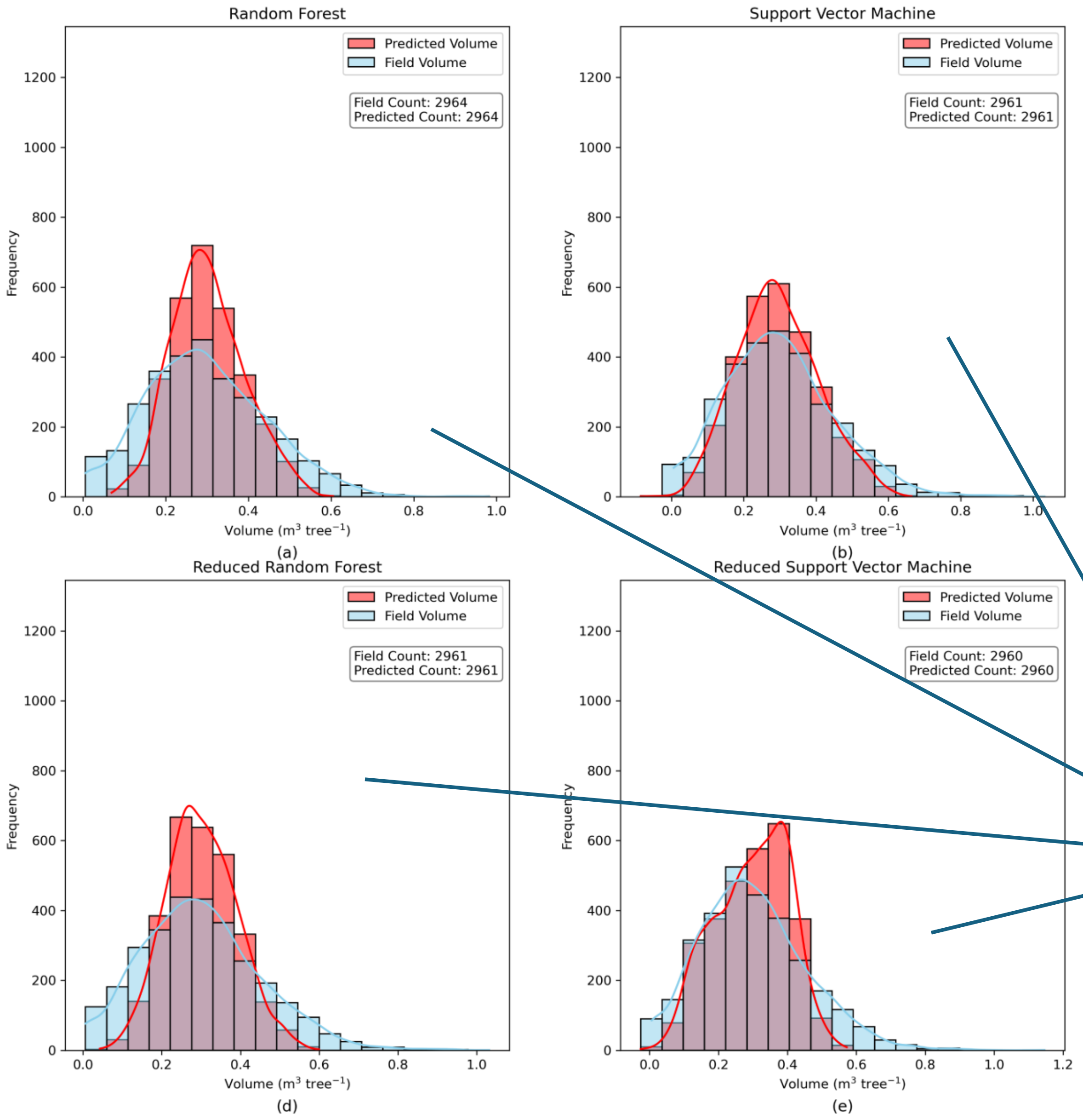
Approach 1: LiDAR



Best predicts individual yield

Feature reduction affects model performance

Approach 1: LiDAR



Tend to track the overall shape decently

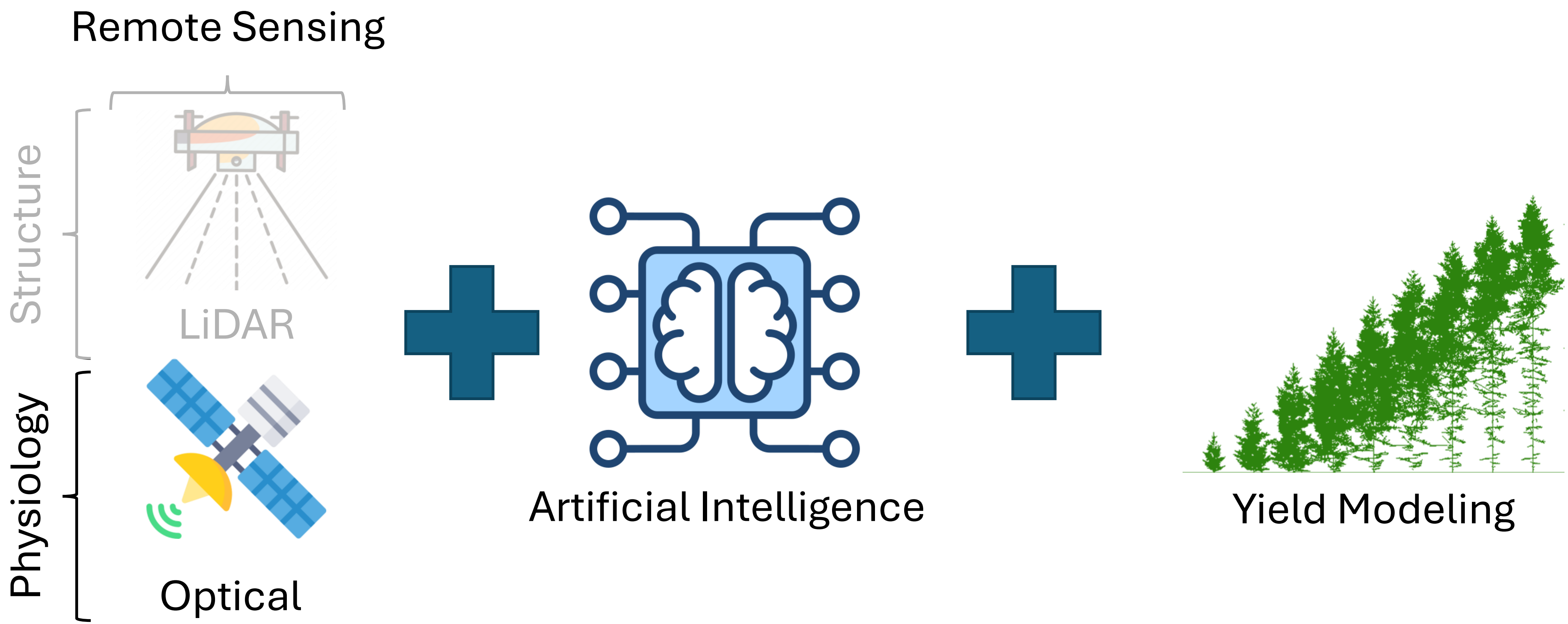
Approach 1: LiDAR

	Random Forest Model (RF_{full})	Support Vector Machine Model (SVM_{full})	Reduced RF Model (Top 7 Variables)	Reduced SVM Model (Top 7 Variables)
MAE (m^3 $tree^{-1}$)	0.08	0.07	0.09	0.07
MSE (m^3 $tree^{-1}$)	0.01	0.01	0.01	0.01
RMSE (m^3 $tree^{-1}$)	0.11	0.1	0.12	0.1
nRMSE (%)	10.86	9.59	11.88	9.14
Accuracy (%)	89.14	90.41	88.12	90.86
R^2	0.48	0.59	0.37	0.55
Predicted values (m^3)	898.48	873.33	892.98	880.42
Field values (m^3)	884.99	886.66	886.66	888.46
Difference (%)	1.53	-1.5	0.71	-0.9

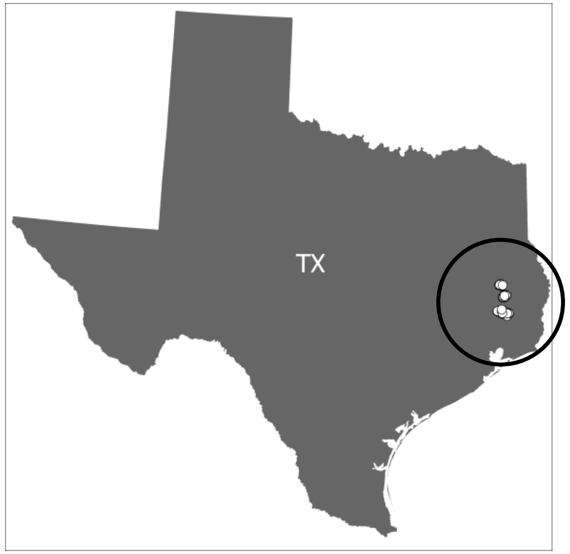
Approach 1: LiDAR

- Slightly significant differences in mean absolute error (MAE) across models ($F_{3, 64}$ = 2.95, p = 0.04)
- Planting density caused significant differences in accuracy across the models ($F_{9, 11830}$ = 12.06, p < 0.001)
- Overall, both RF and SVM show promise for predicting yield over a 4-year period
- Individual tree-level yields when aggregated at the stand-level, provide a good estimate of timber volume compared to field estimates
- Manuscript submitted to *Remote Sensing of Environment*

Approach 2: Satellite Remote Sensing



Approach 2: Satellite Remote Sensing



- 258 plots
- 3 states: Georgia, Florida and Texas
- Data collected within December 2017 – February 2018
- Contains typical field information (mean height, quadratic mean diameter, number of trees, total volume)

Approach 2: Satellite Remote Sensing

Satellite Data (From Google Earth Engine)

- Sentinel 1
  - C-band SAR data
  - 2 bands: VV and VH
  - January 2016 – December 2017 (24 Months)

VV_VH_Ratio	VV/VH
VH_VV_Ratio	VH/VV
VV_VH_Difference	$VV - VH$
VH_VV_Difference	$VH - VV$
Normalized_VVVH	$(VV - VH)/(VV + VH)$
Sum_VVVH	$VV + VH$
Product_VVVH	$VV * VH$

Approach 2: Satellite Remote Sensing

Satellite Data (From Google Earth Engine)

DPDD (Dual-polarization difference descriptor)	$(VV + VH) / \sqrt{2}$
Gamma_nought_VH	$VH/\cos(angle * (\pi)/180)$
Gamma_nought_VV	$VV/\cos(angle * (\pi)/180)$
RVI (Radar Vegetation Index)	$(4 * VH)/ (VV + VH)$
VDDPI (Volume density dual-polarization index)	$(VV * VH) / VV$

Approach 2: Satellite Remote Sensing

Satellite Data (From Google Earth Engine)

- Sentinel 2
  - Level-1C (Top of Atmosphere)
  - Multispectral bands
  - 13 bands: (B1 to B12)
  - January 2016 – December 2017 (24 Months)

NDVI	Normalized Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$
EVI	Enhanced Vegetation Index	$2.5 * \frac{NIR - Red}{NIR + 6 \times Red - 7.5 \times Blue + 1}$
SAVI	Soil-adjusted Vegetation Index	$\frac{NIR - Red}{NIR + Red + 0.5} \times 1 + 0.5$
MSAVI2	Modified Soil-adjusted Vegetation Index 2	$0.5 \times (2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)})$
OSAVI	Optimized Soil-Adjusted Vegetation Index	$\frac{NIR - Red}{NIR + Red + 0.16}$

Approach 2: Satellite Remote Sensing

Satellite Data (From Google Earth Engine)

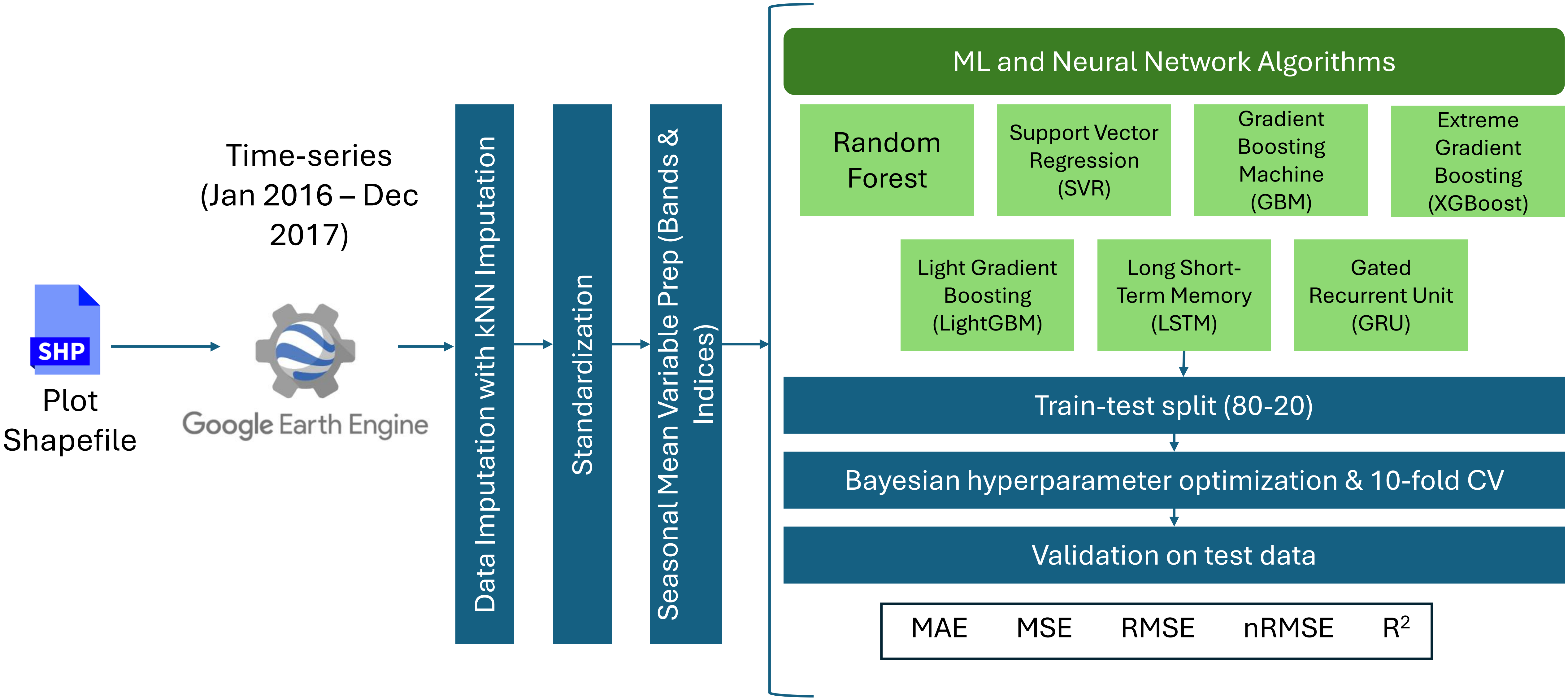
NDWI_McFeeters	Normalized Difference Water Index (McFeeters version)	$\frac{Green - NIR}{Green + NIR}$
MSI	Moisture Stress Index	$\frac{SWIR1}{NIR}$
NDRE	Normalized Difference Red Edge Index	$\frac{NIR - RedEdge1}{NIR + RedEdge1}$
SIPI3	Structure Insensitive Pigment Index	$\frac{NIR - Blue}{NIR - Red}$
SR	Simple Ratio Index	$\frac{NIR}{Red}$
DVI	Difference Vegetation Index	$NIR - Red$
REIP	Red Edge Inflection Point	$700 + 40 \times \left(\frac{\frac{Red + RedEdge3}{2} - RedEdge1}{RedEdge2 - RedEdge1}\right)$
LCI	Leaf Chlorophyll Index	$\frac{NIR - RedEdge1}{NIR + Red}$

Approach 2: Satellite Remote Sensing

Satellite Data (From Google Earth Engine)

NDII	Normalized Difference Infrared Index	$\frac{NIR - SWIR1}{NIR + SWIR1}$
NDLI	Normalized Difference Lignin Index	$\frac{\log(SWIR1) - \log(SWIR2)}{\log(SWIR1) + \log(SWIR2)}$
NMDI	Normalized Multi-band Drought Index	$\frac{NIR - (SWIR1 - SWIR2)}{NIR + (SWIR1 - SWIR2)}$
GNDVI	Green Normalized Difference Vegetation Index	$\frac{NIR - Green}{NIR + Green}$
CVI	Chlorophyll vegetation index	$9 \times \frac{RedEdge1}{Green^2}$
GLI	Green Leaf Index	$\frac{2Green - Red - Blue}{2Green + Red + Blue}$
TC_Brightness	Tasselled Cap – Brightness	$0.3037 \times Blue + 0.2793 \times Green + 0.4743 \times Red + 0.5585 \times NIR + 0.5082 \times Cirrus + 0.1863 \times SWIR2$
TC_GVI	Tasselled Cap Green Vegetation Index	$-0.283 \times Green - 0.660 \times Red + 0.577 \times RedEdge2 + 0.388 \times Water\,Vapor$

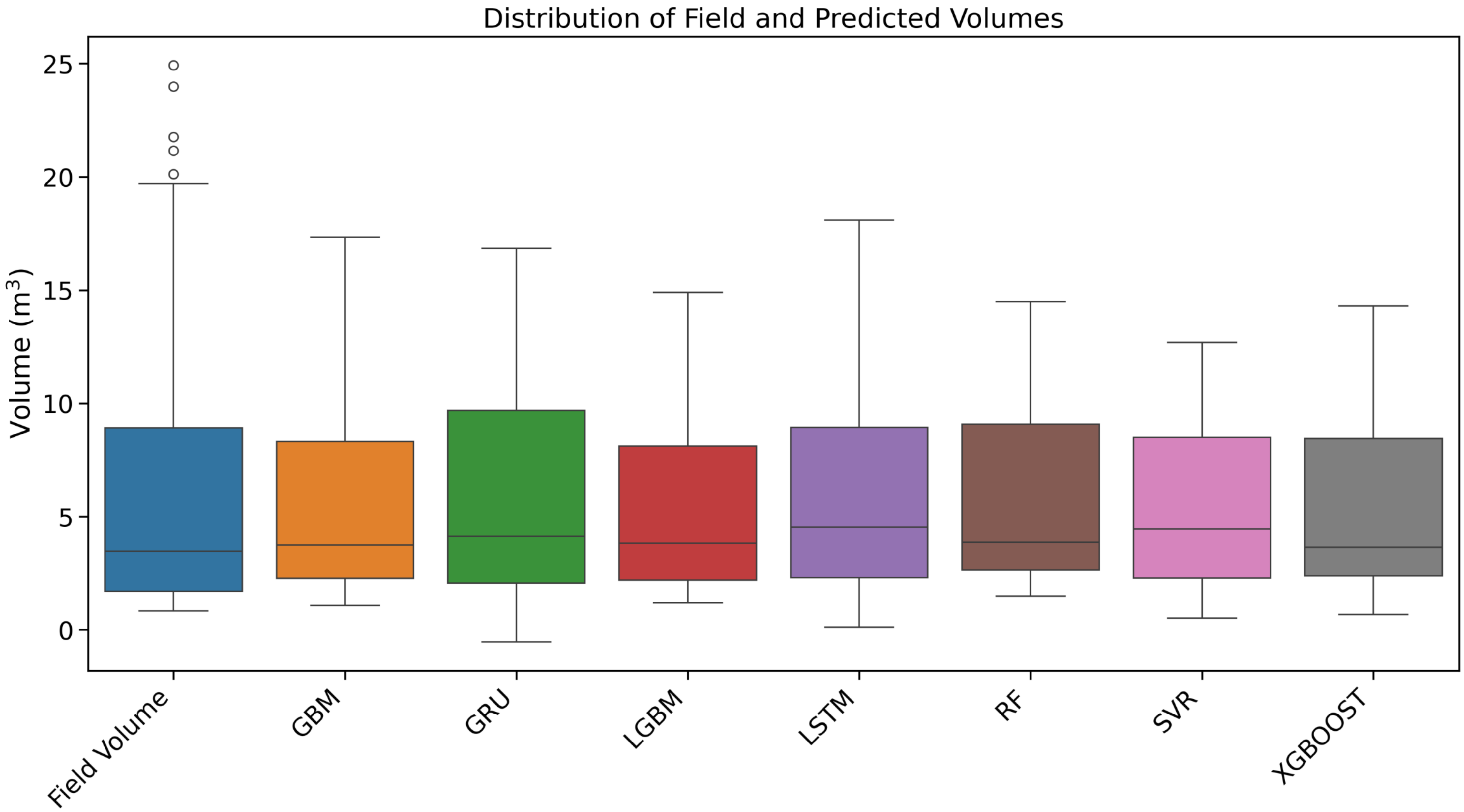
Approach 2: Satellite Remote Sensing



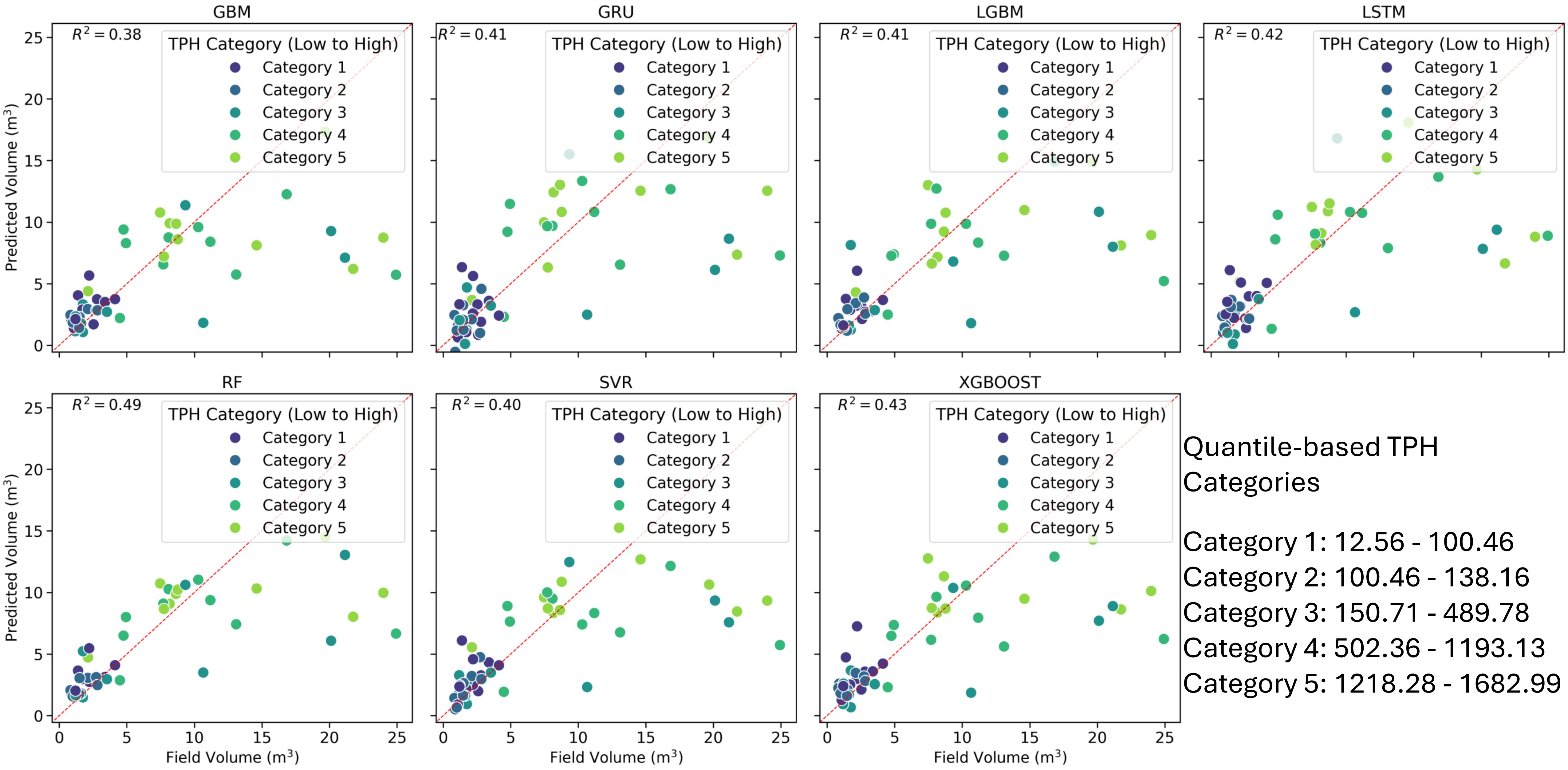
Approach 2: Satellite Remote Sensing

		1st		2nd		3rd	
Metrics	SVR	RF	GBM	XGBoost	LightGBM	LSTM	GRU
MAE (m^3)	3.02	2.70	2.97	2.94	2.99	3.18	3.30
MSE (m^3)	27.13	23.20	28.23	25.92	26.67	26.35	26.79
RMSE (m^3)	5.21	4.82	5.31	5.09	5.16	5.13	5.17
nRMSE (%)	21.62	19.99	22.05	21.13	21.43	21.31	21.48
R^2	0.40	0.49	0.38	0.43	0.41	0.42	0.41

Approach 2: Satellite Remote Sensing



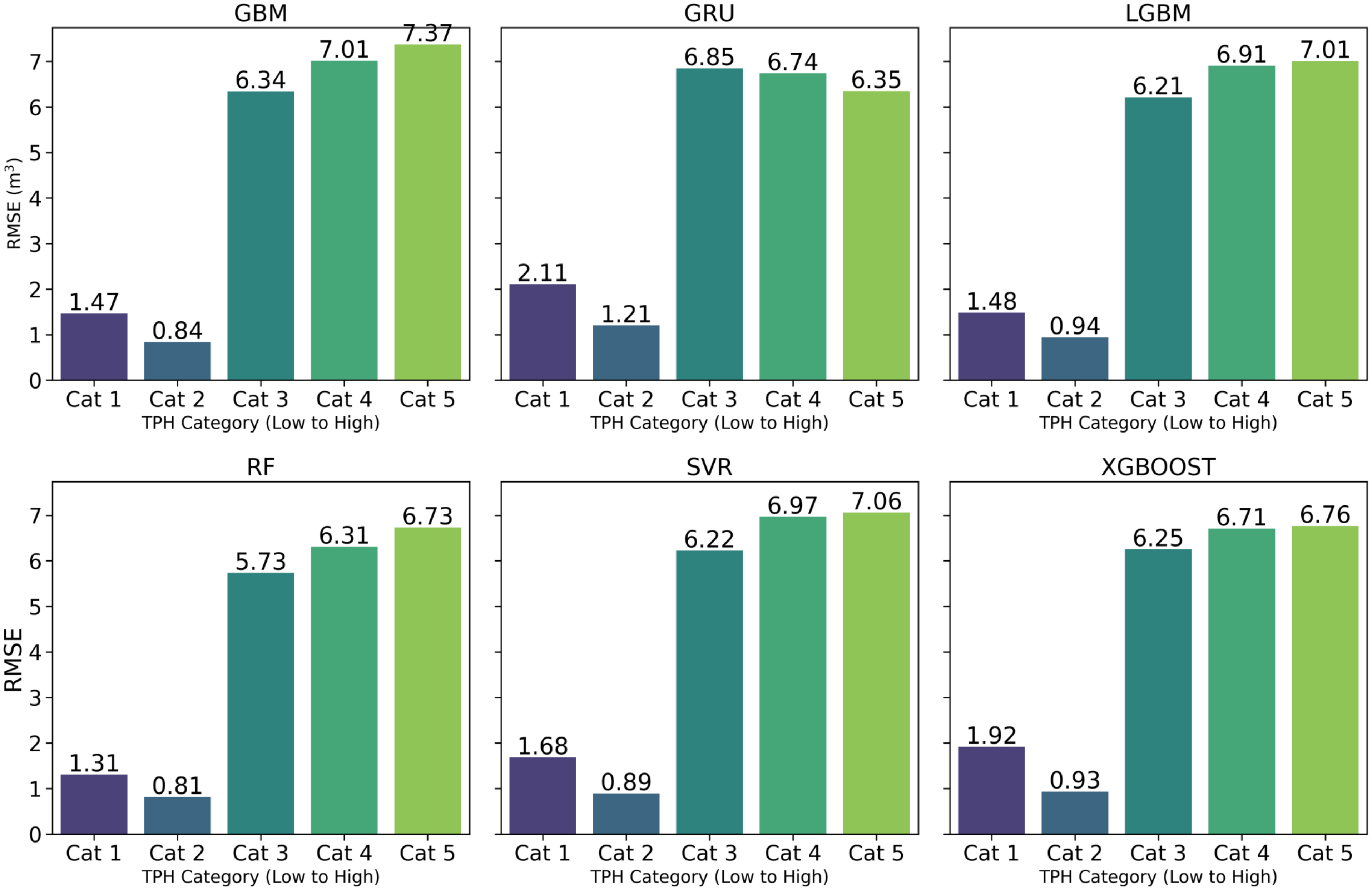
Approach 2: Satellite Remote Sensing



Quantile-based TPH Categories

Category 1: 12.56 - 100.46
Category 2: 100.46 - 138.16
Category 3: 150.71 - 489.78
Category 4: 502.36 - 1193.13
Category 5: 1218.28 - 1682.99

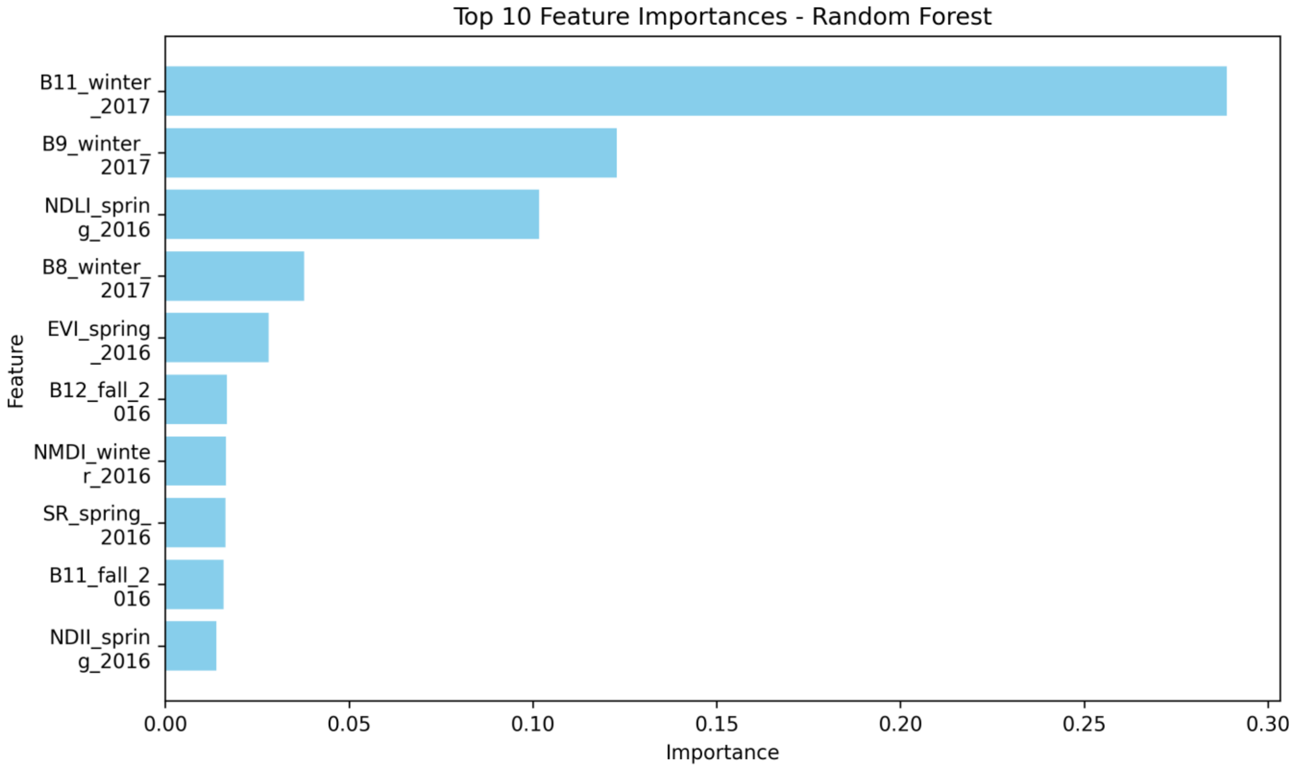
Approach 2: Satellite Remote Sensing



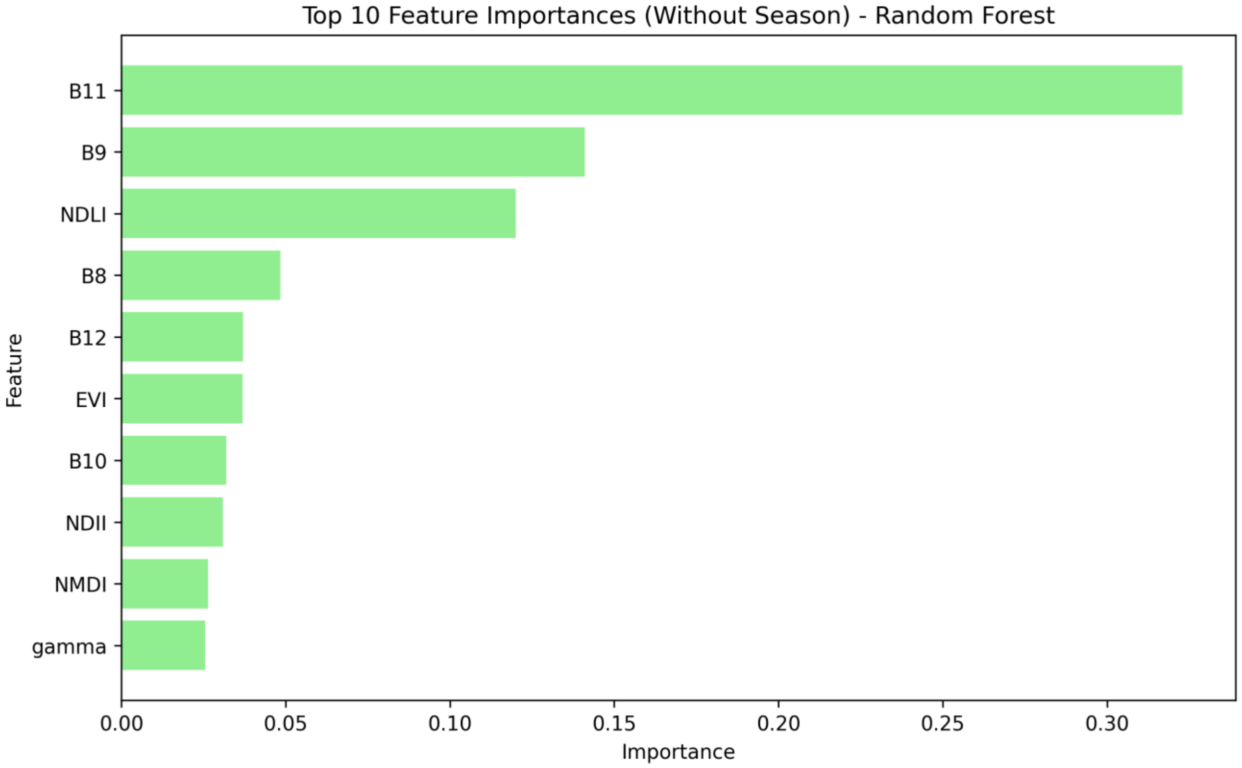
Quantile-based TPH Categories

Category 1: 12.56 - 100.46
Category 2: 100.46 - 138.16
Category 3: 150.71 - 489.78
Category 4: 502.36 - 1193.13
Category 5: 1218.28 - 1682.99

Approach 2: Satellite Remote Sensing



Gini importance (Mean decrease in impurity)



Approach 2: Satellite Remote Sensing

- RF model best predicts the plot-level yield of Loblolly Pine using the time-series bands and indices, followed by XGBoost and LSTM
- The prediction errors are typically lower at lower planting densities
- In the RF model, the planting density significantly affects the accuracy ($F_{4,64} = 16.64$, $p < 0.001$)
- Shortwave infrared band 1 (SWIR 1 – band 11) and water vapor band (band 9) were deemed the most important bands in the RF model. Further investigation is required
- Further investigation required to assess the impact of thinning status on the yield prediction

Thank you!

I am open to your questions and comments

Acknowledgment

FOREST PRODUCTIVITY COOPERATIVE