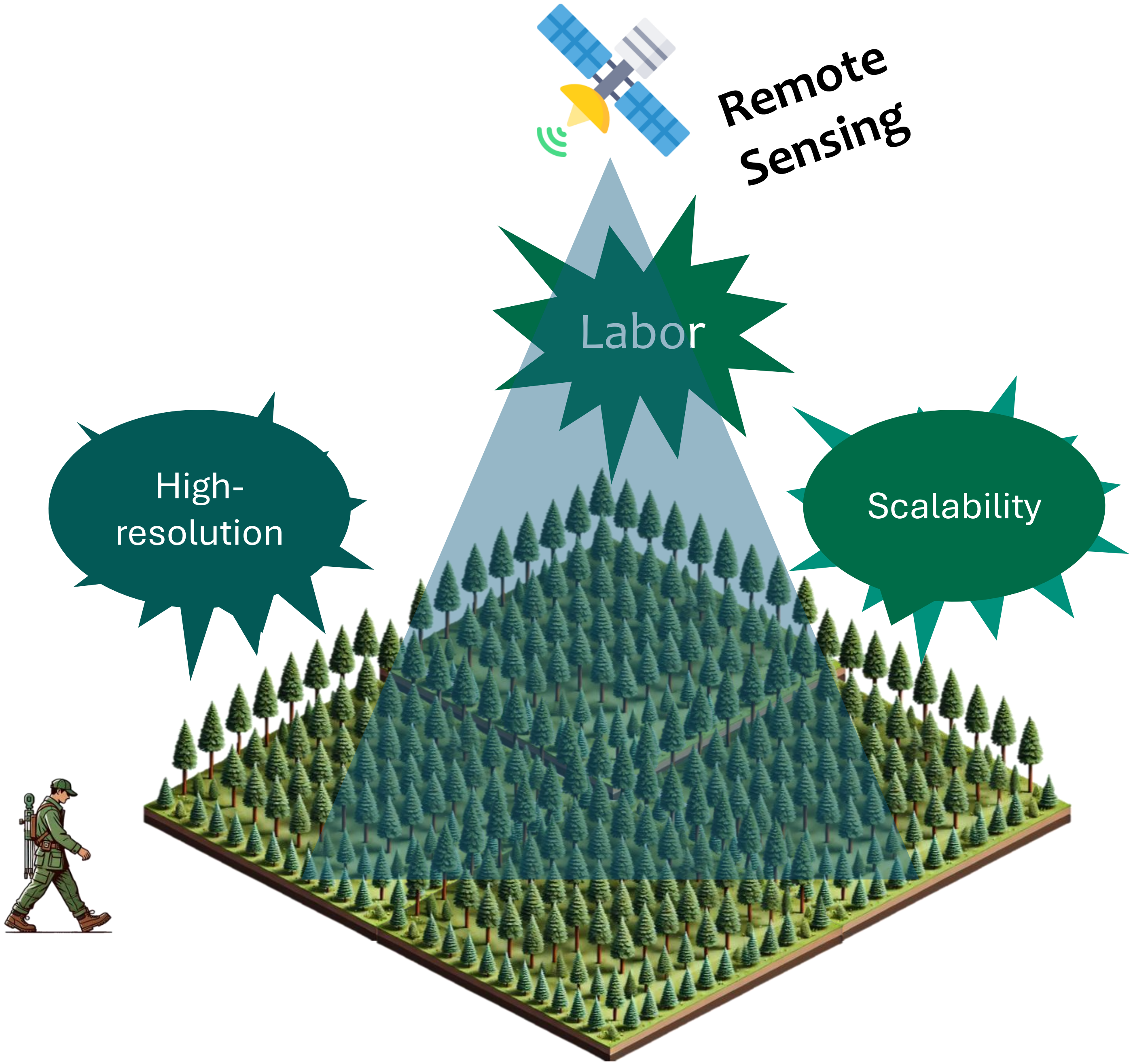
*FREC Seminar 2025*

# Remote Sensing Meets AI: Predicting Yield in *Pinus taeda* (L.) Plantations with Opensource Time-Series Satellite Data

Gunjan Barua
*Ph.D. Candidate*

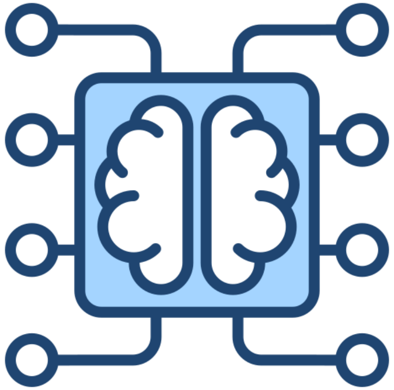
VIRGINIA TECH


Remote Sensing
Labor
High-resolution
Scalability

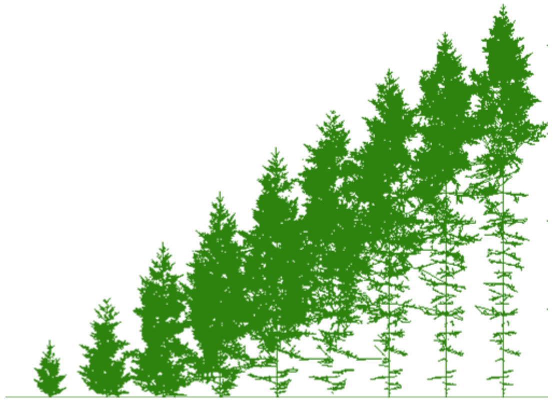
## Research Questions

- Can open-source satellite data (SAR & Multispectral) 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?

Satellite Data



Artificial Intelligence



Yield Modeling

# Study Area




- 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)

## Satellite Data (From Google Earth Engine)

Sentinel 1 (SAR Data)

Sentinel 2 (Multispectral Data)

**Time-series Data: January 2016 – December 2017**

**Target Variable:** Plot-level total yield in January 2018

## 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 | $VV * VH$ |

## Satellite Data (Contd.)

| | |
|---|---|
| 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) | $(VV * VH) / VV$ |

## Satellite Data (Contd.)

- 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 | $\frac{NIR - Red}{NIR + Red + 0.16}$ |

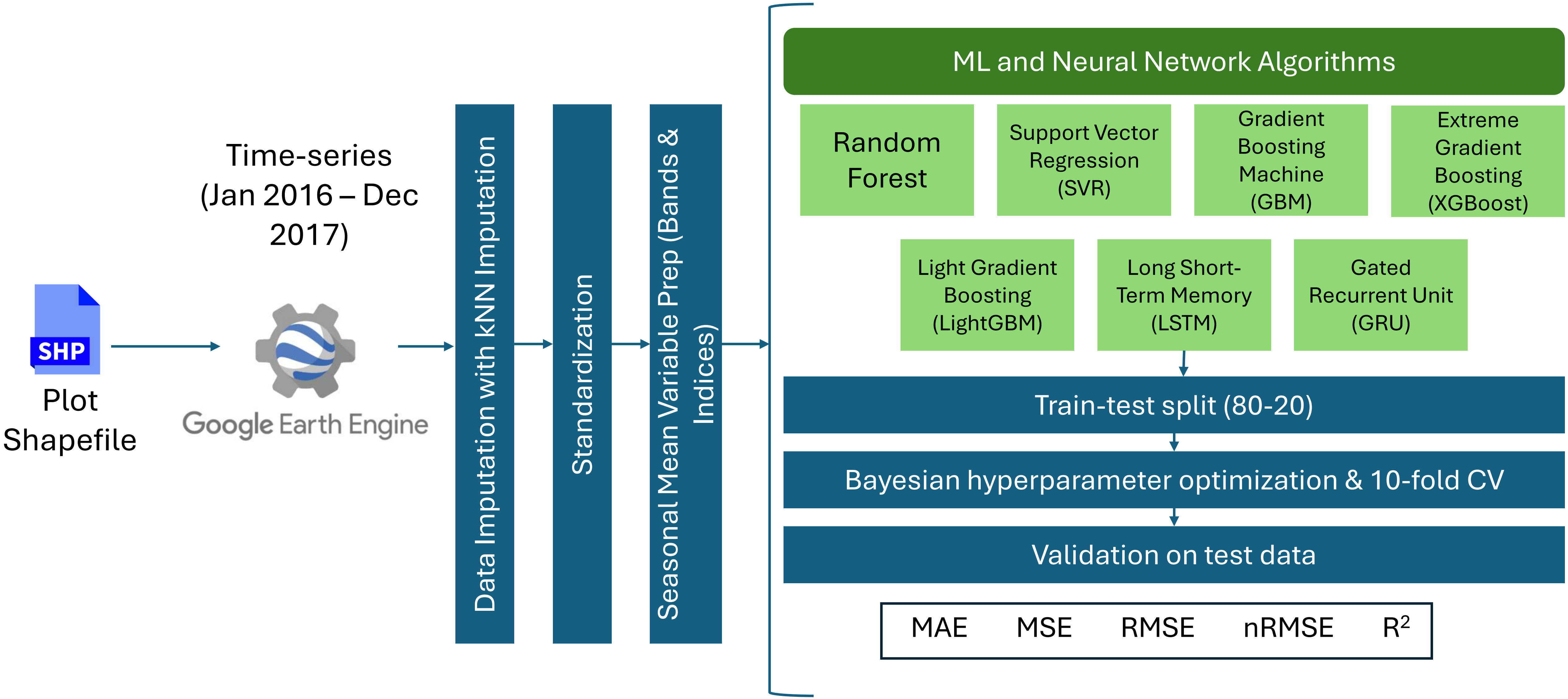
# Satellite Data (Contd.)

| | | |
|---|---|---|
| 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 | $\frac{NIR - RedEdge1}{NIR + Red}$ |

# Satellite Data (Contd.)

| | | |
|---|---|---|
| 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 | $-0.283 \times Green - 0.660 \times Red + 0.577 \times RedEdge2 + 0.388 \times Water\ Vapor$ |

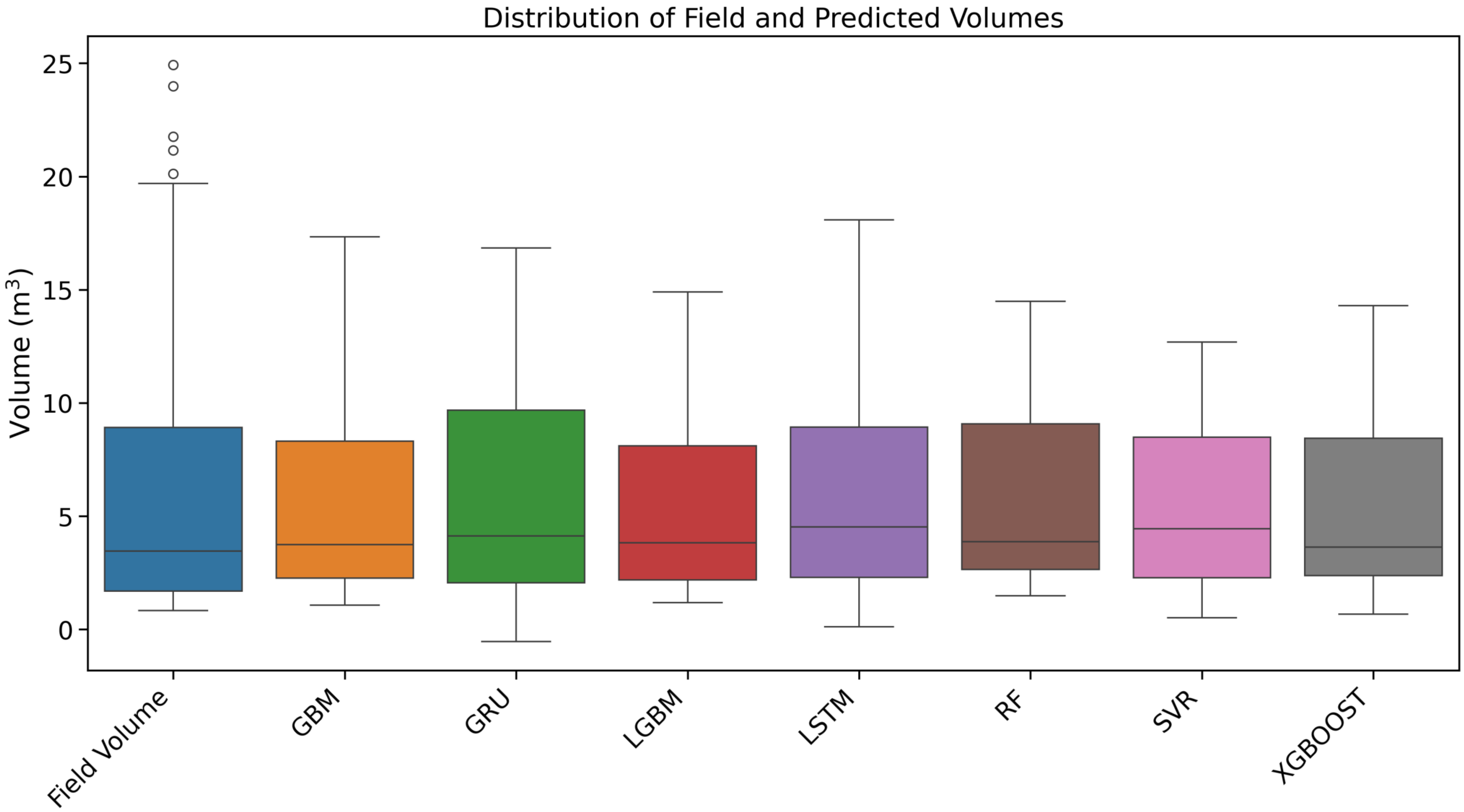
# Approach 2: Satellite Remote Sensing



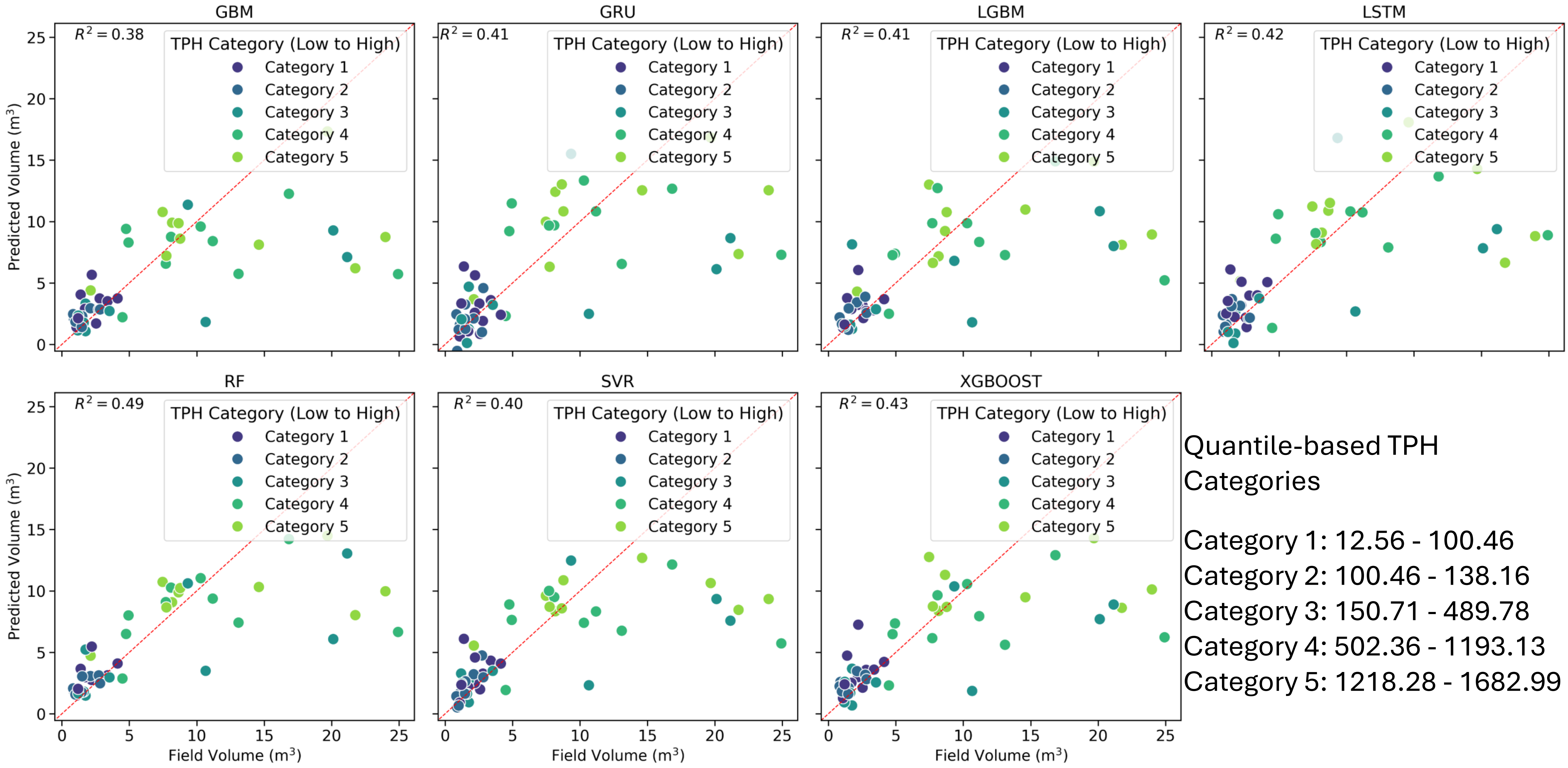
# Results

| | | 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 |

# Results (Contd.)



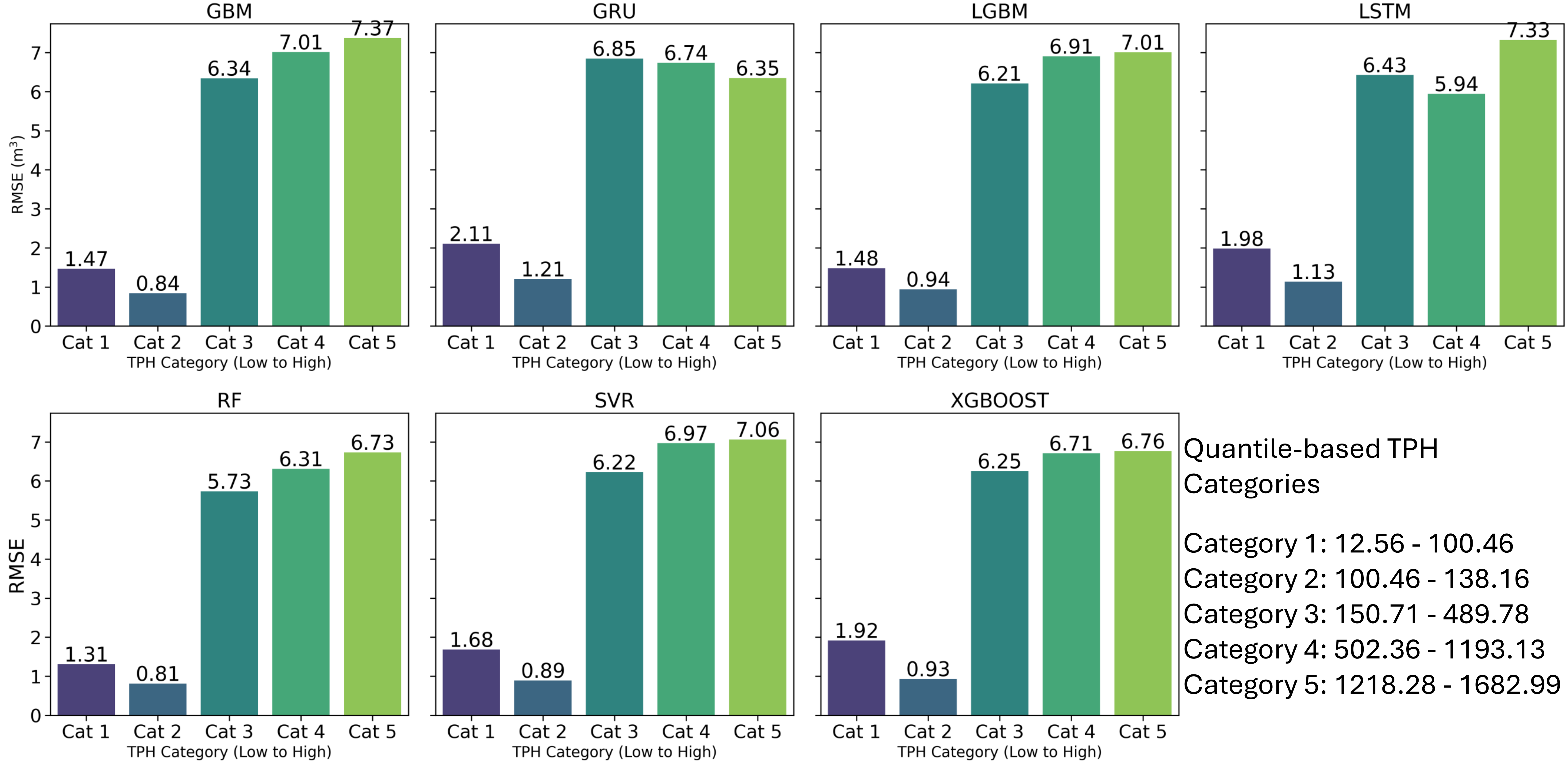
# Results (Contd.)




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

# Results (Contd.)




Quantile-based TPH Categories

Category 1: 12.56 - 100.46

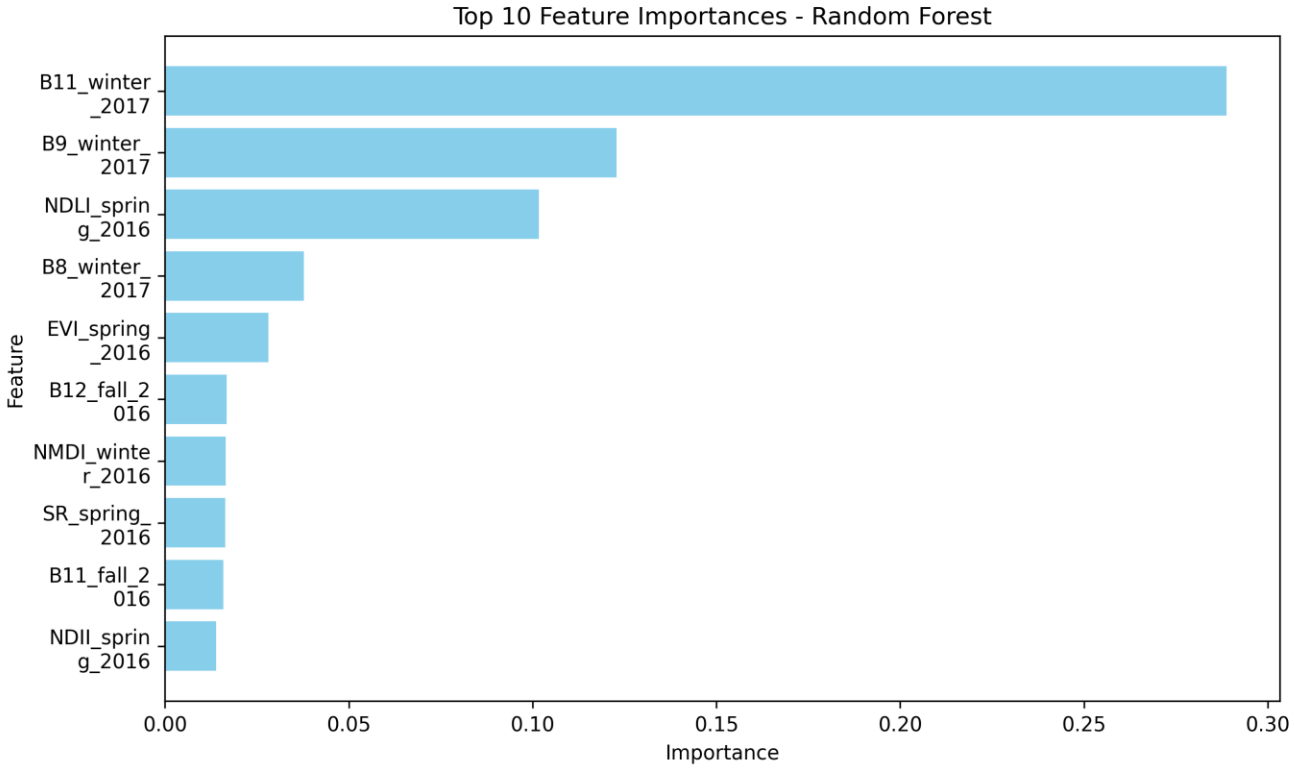
Category 2: 100.46 - 138.16

Category 3: 150.71 - 489.78

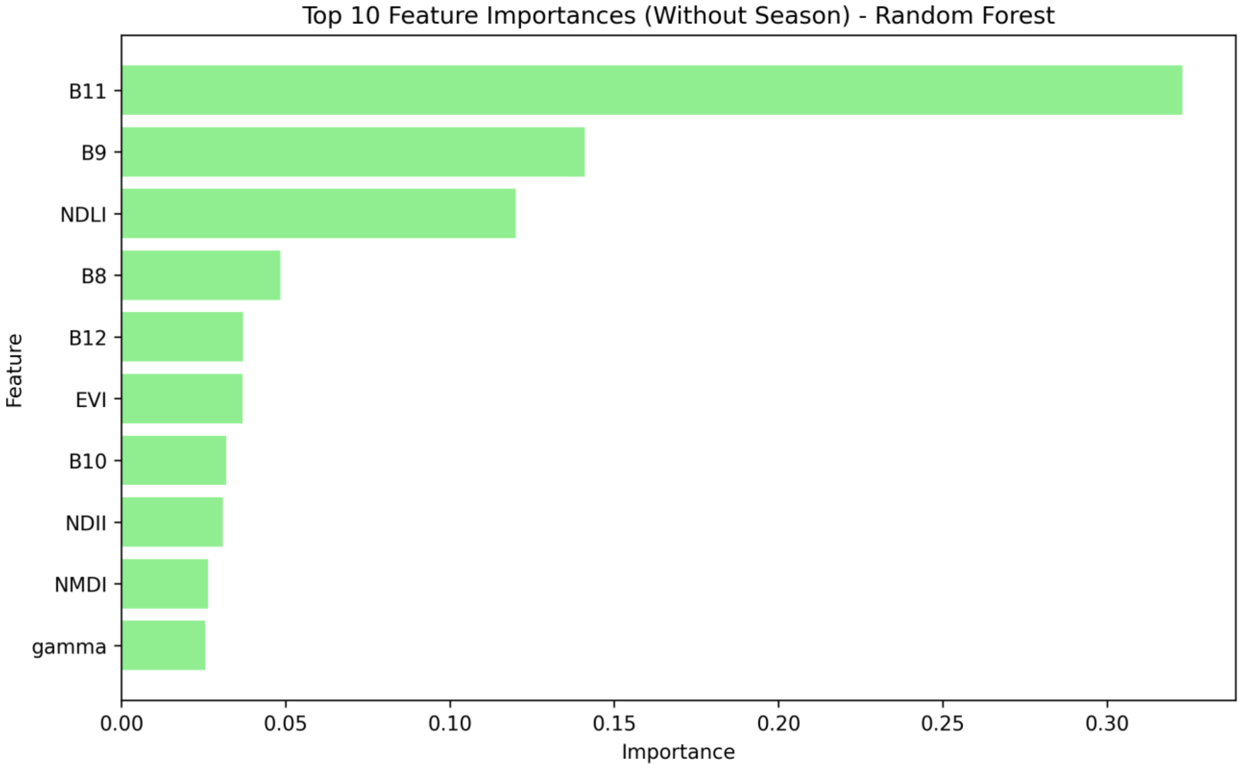
Category 4: 502.36 - 1193.13

Category 5: 1218.28 - 1682.99

# Results (Contd.)




Gini importance (Mean decrease in impurity)



# Conclusion

- 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