A South Carolina Lidar Forest Modeling Analysis

4/7/2021, OLI Meeting

Jacob Strunk, Bob McGaughey

USFS VMaRS Team

Study Site

- Savannah River Site in South Carolina
- Department of Energy facility but non-facility lands are managed by the Forest Service
- Total area is ~200,000 acres with ~170,000 acres of forested land
- Lands managed for wood products and habitat for various species
- Some areas are restricted or require special precautions for entry
- Earlier lidar project in 2009-2010

Project Objectives

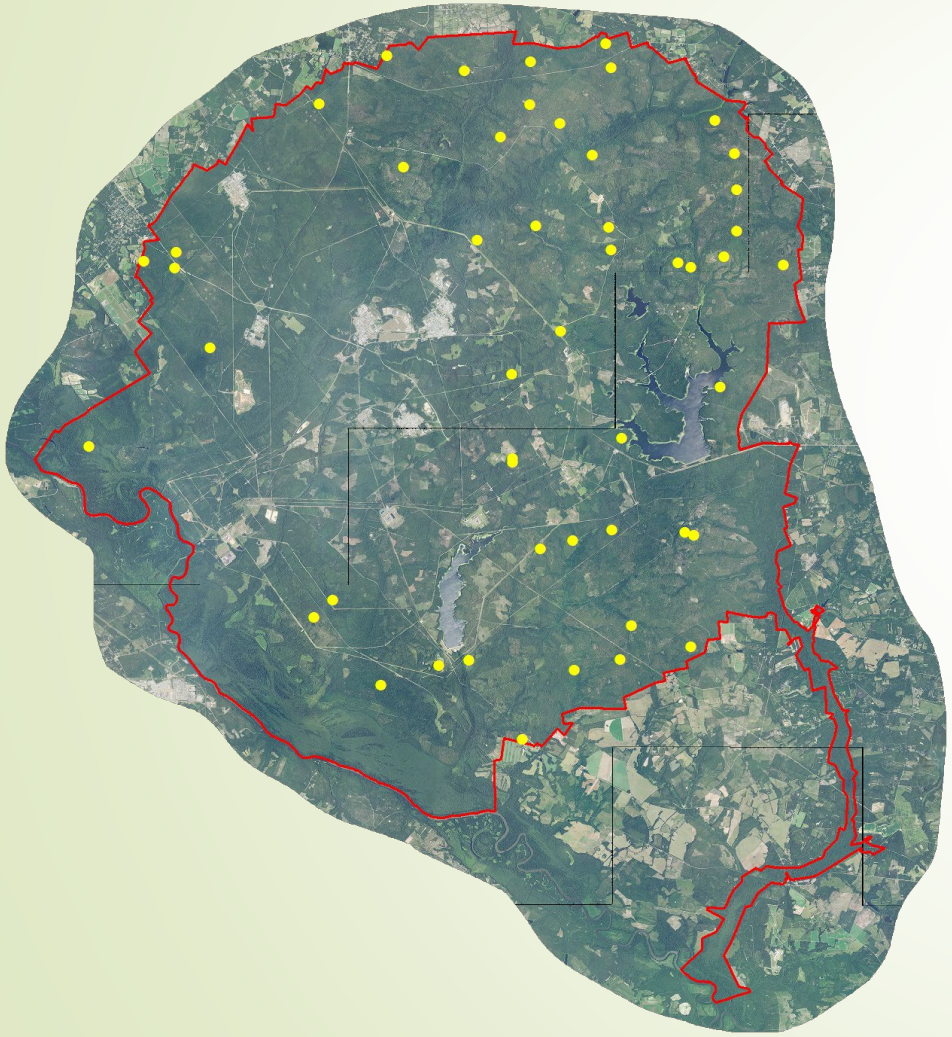
1. Build operational inventory

2. Research Investigation

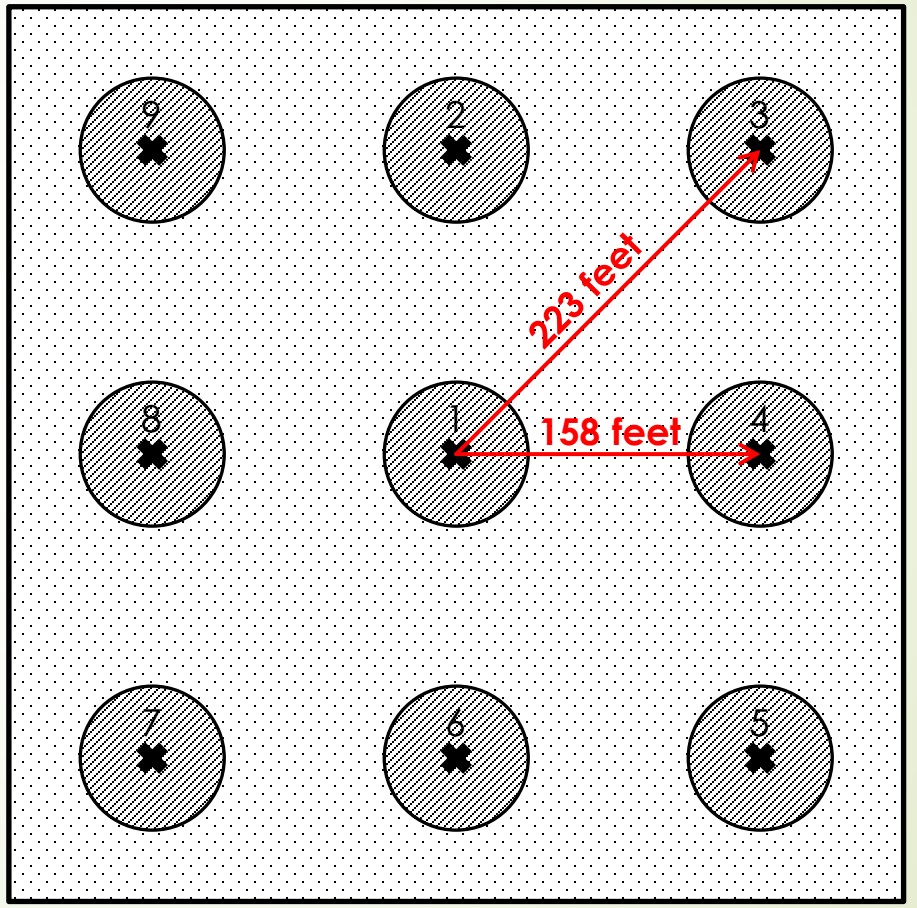
Sample Design

- 548 Plot Fixed Radius Plots
- 50 Validation Stands
- Lidar (leaf off)

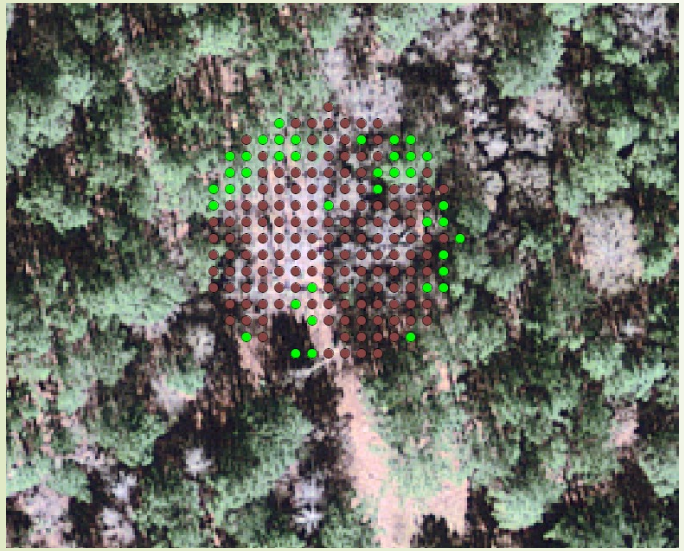
50 Validation-stands (9 x 0.1 acre plots)



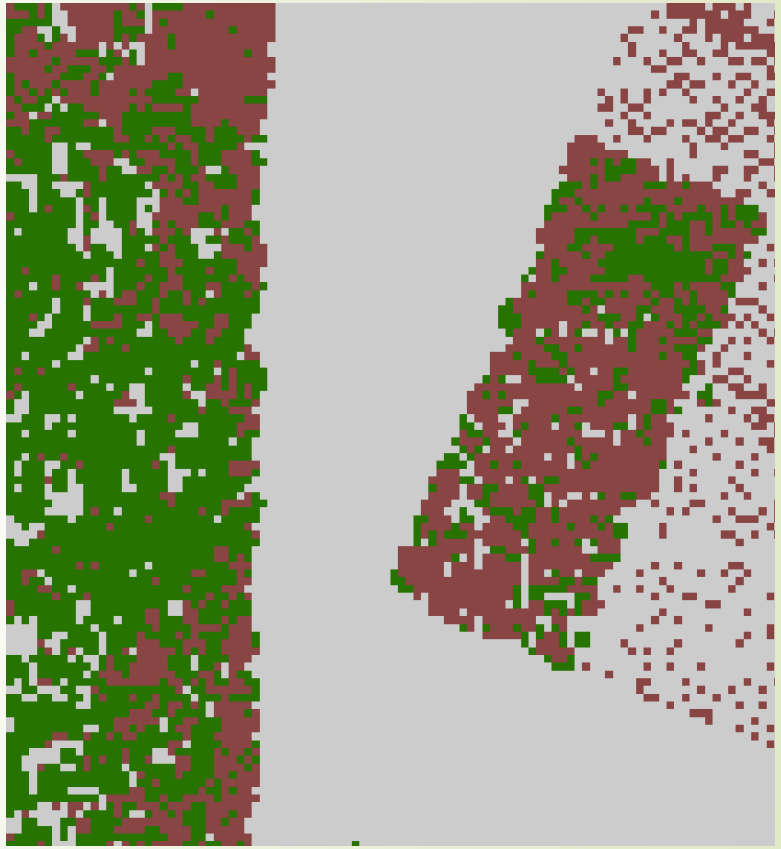
Validation-stand (5 acres)



Hardwood Index



Mixed hardwood/conifer



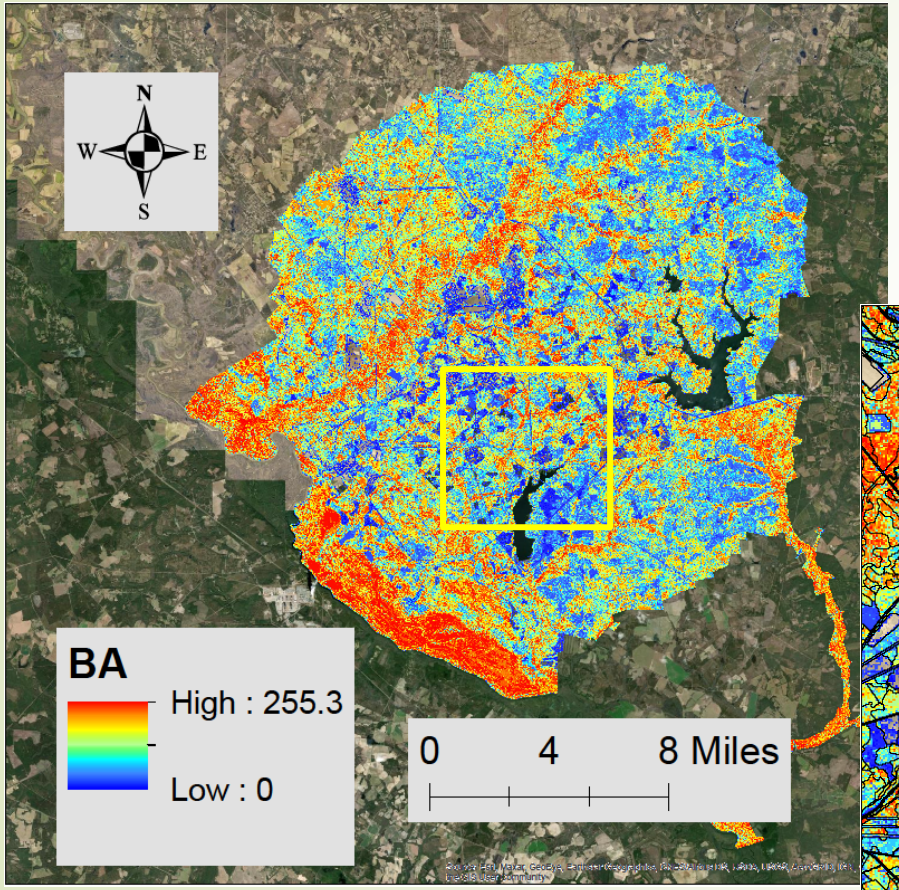
Random Forest (RF) model on homogeneous plots (2m)
-height and intensity metrics

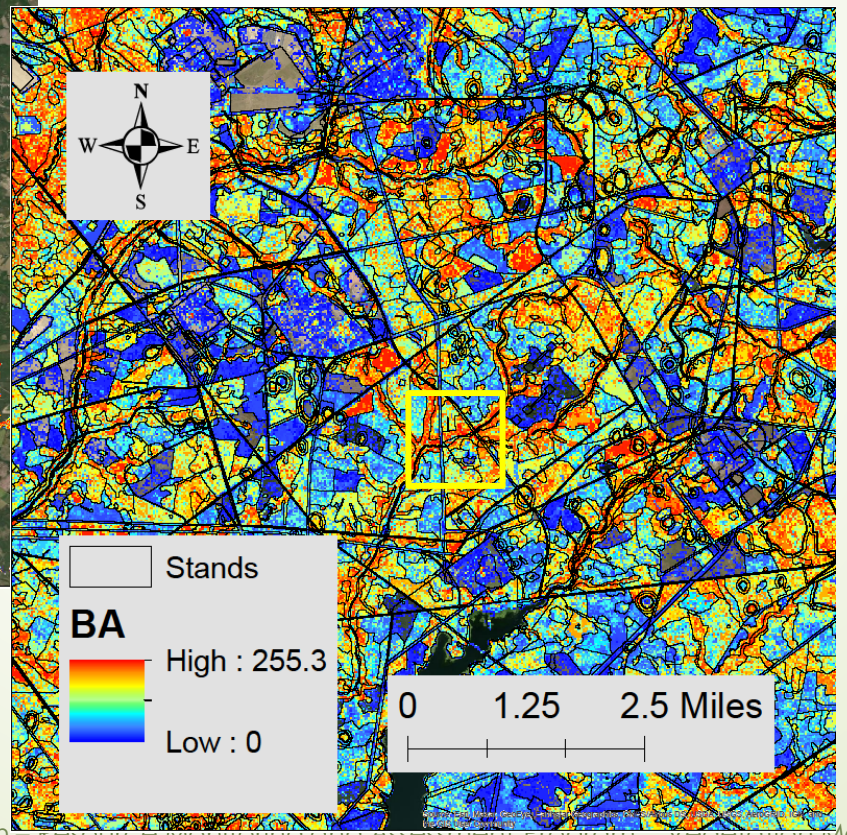
Operational Products

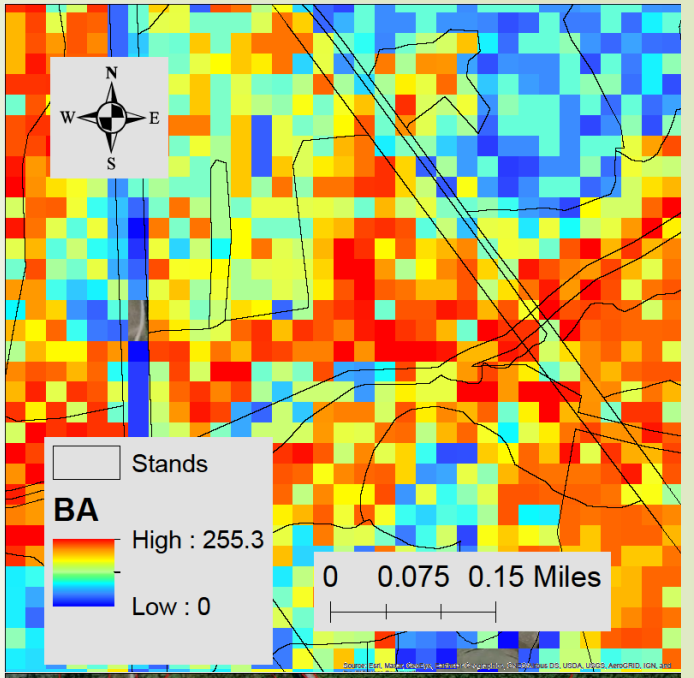
- Raster layers
- Vector layer (many attributes)
- Site-wide Tree List

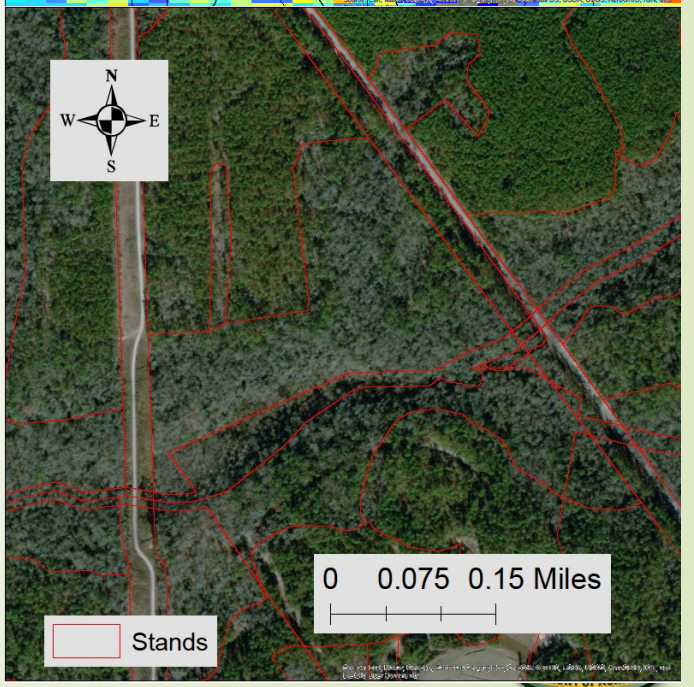
- FPS Database (https://fbrinstitute.org/)
  - Evaluate management alternatives
  - Demonstrate sustainable yield

Raster Products

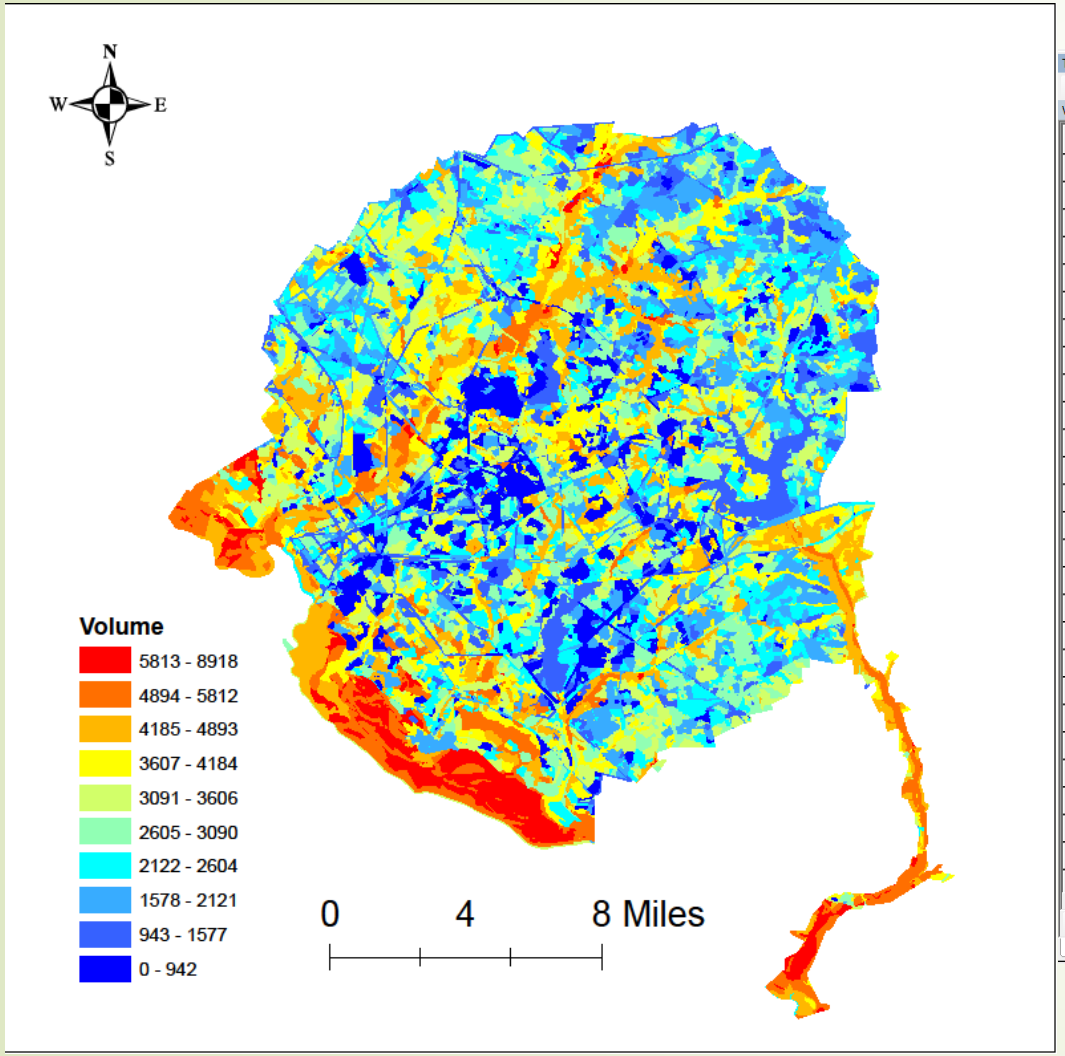








Vector layers



Table

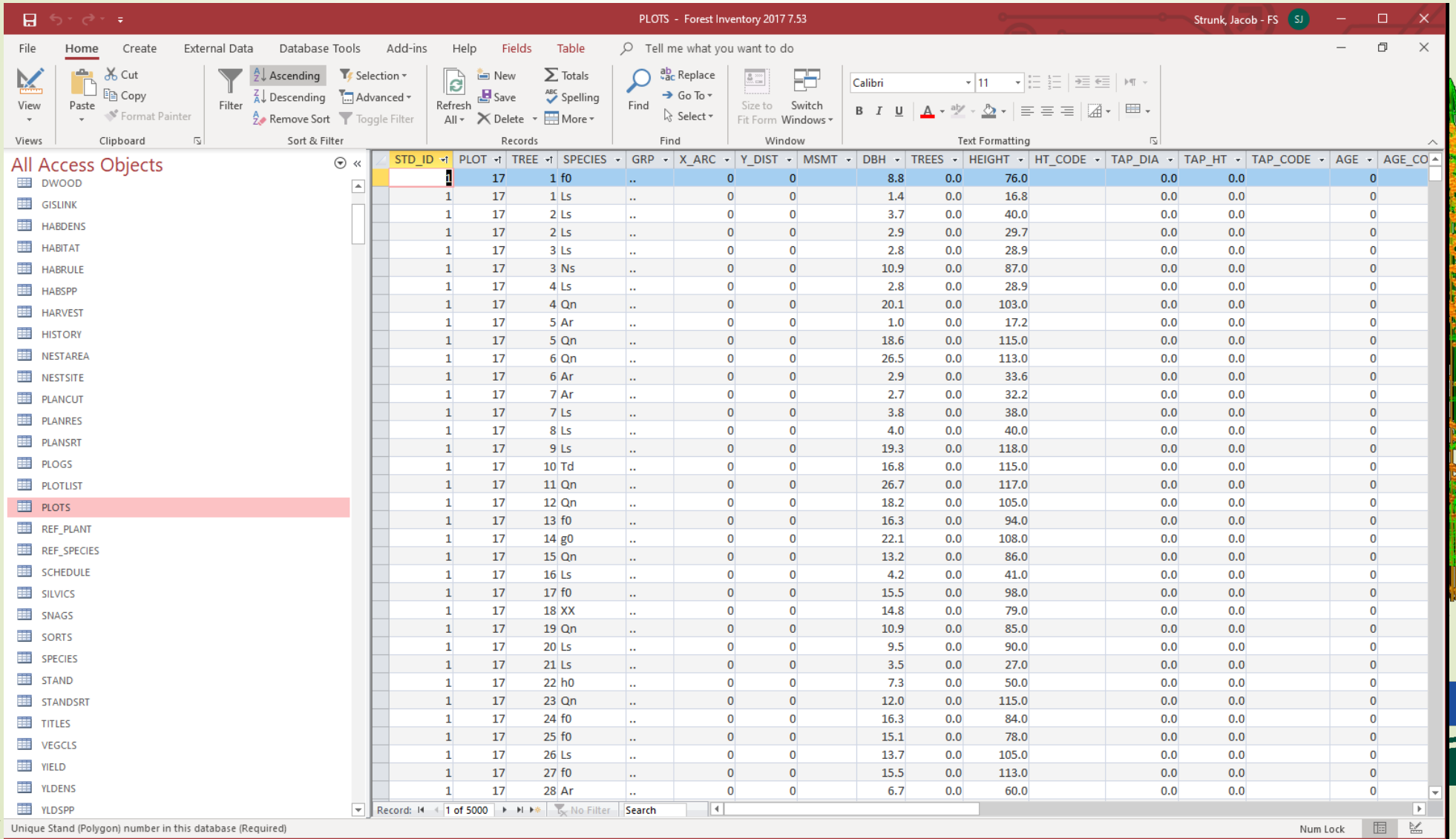
Volume

ACRES	YEAR_	STAND_A	ba_hs_l_	ba_hw_	ba_sw_l	vol_hs_l	vol_hw_l_	vol_sw_l	bbm_hs_l_tp	bbm_hw_l_tp	bbm_sw
72.85	1990	29	62.63	6.64	56	1836.78	136.68	1700.1	28.8	2.47	
8.19	1990	29	64.76	23.06	41.7	1740.78	523.39	1217.39	28.33	9.18	
20.71	1990	29	70.26	22.84	47.42	1985.12	542.62	1442.5	31.69	9.26	
6.54	1990	29	67.04	7.32	59.71	1806.08	137.89	1668.19	28.61	2.59	
21	1990	29	77.64	7.91	69.73	1930.03	135.24	1794.79	30.9	2.69	
53.33	1990	29	110.8	31.95	78.85	2756.63	706.07	2050.55	45.01	12.9	
47.9	1990	29	79.31	20.84	58.47	2148.69	421.64	1727.05	34.65	7.87	
47.26	1990	29	104.32	32.61	71.7	2946.42	737.55	2208.87	47.74	13.27	
39.02	1990	29	126.33	12.07	114.27	3512.01	238.78	3273.23	55.82	4.51	
21.92	1990	29	88.62	24.11	64.52	2094.54	475.93	1618.6	34.43	8.98	
50.29	1990	29	83.58	9.97	73.61	2166.38	188.65	1977.73	34.52	3.57	
0.09	1990	29	66.1	13.3	52.8	1302.9	211.2	1091.7	21.6	3.9	
13.22	1990	29	102.7	5.26	97.43	2605.87	79.28	2526.58	41.34	1.62	
40.75	1990	29	107.19	11.18	96	2792.31	209.03	2583.28	44.43	3.98	
19.1	1990	29	118.74	14.8	103.95	3392.5	307.43	3085.08	54.02	5.79	
17.27	1990	29	72.78	49.06	23.71	1656.29	1111.36	544.94	27.32	18.73	
47.46	1990	29	55.55	10.19	45.36	1513.79	211.85	1301.95	24.26	3.82	
1.52	1990	29	86.54	60.27	26.29	2017.97	1304.27	713.71	34	22.93	
33	1990	29	97.62	16.33	81.29	2453.84	318.27	2135.58	39.44	5.88	
1.62	1990	29	44.14	14.44	29.71	1003.73	260.23	743.49	17.01	4.81	
13.36	1990	29	76.77	10.38	66.4	1861.01	164.16	1696.86	29.92	3.34	
46.9	1990	29	135.01	9.04	125.97	3916.96	174.62	3742.34	61.99	3.3	
3.39	1990	29	116.83	104.76	12.08	2895.59	2599.28	296.31	44.72	40.18	
20.9	1990	29	81.05	59.3	21.75	2099.69	1474.41	625.28	35.07	25.19	
61.38	1990	29	138.38	20.76	117.63	4009.68	470.05	3539.63	63.97	8.58	
3.98	1990	29	124.82	91.08	33.72	3694.51	2564.67	1129.84	60.02	42.52	
57.05	1990	29	140.33	11.62	128.71	4037.48	251.84	3785.63	64.08	4.63	

(0 out of 8746 Selected)

Volume

Tree List (FPS Database)



Research Question

- Model Type (PS, OLS, kNN, RF)
- Stand vs Plot
- No. Predictors
- Spp groups (vol x hw, sw)
- DBH (vol x spp grp x 2-inch bins)

- **Growth projection**
- *DBH prediction methods?*
- **Midstory inference**

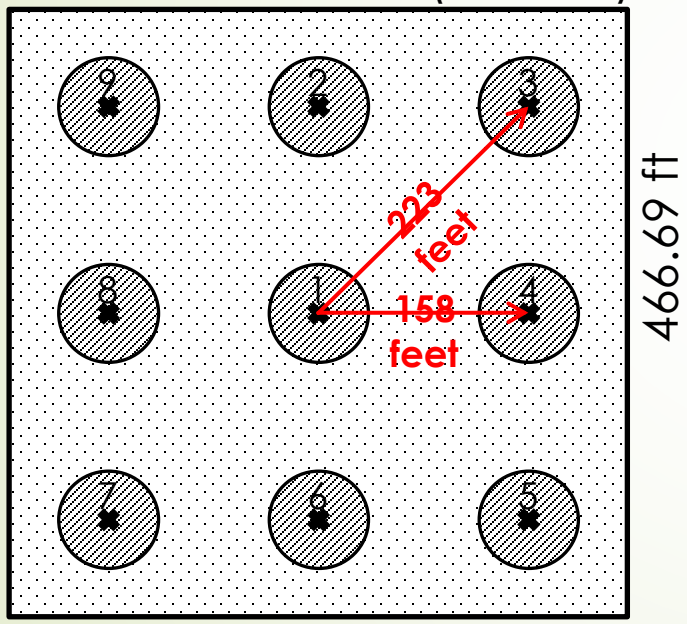
Findings

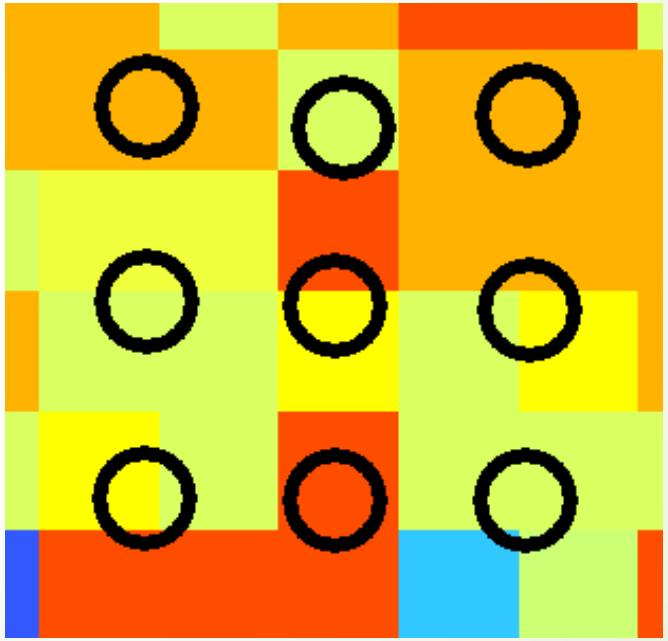
- R^2 – proportion of variance explained
- **Stand level**
- Plot level

- Attributes
  - ba = basal area
  - bbm: biomass
  - cbh: crown base height
  - den: density, trees per acre
  - lor: Lorey's height
  - qmd: quadratic mean diameter
  - t40: top 40, top 40 trees per acre
  - vol: board foot volume / acre

Stand level Inference

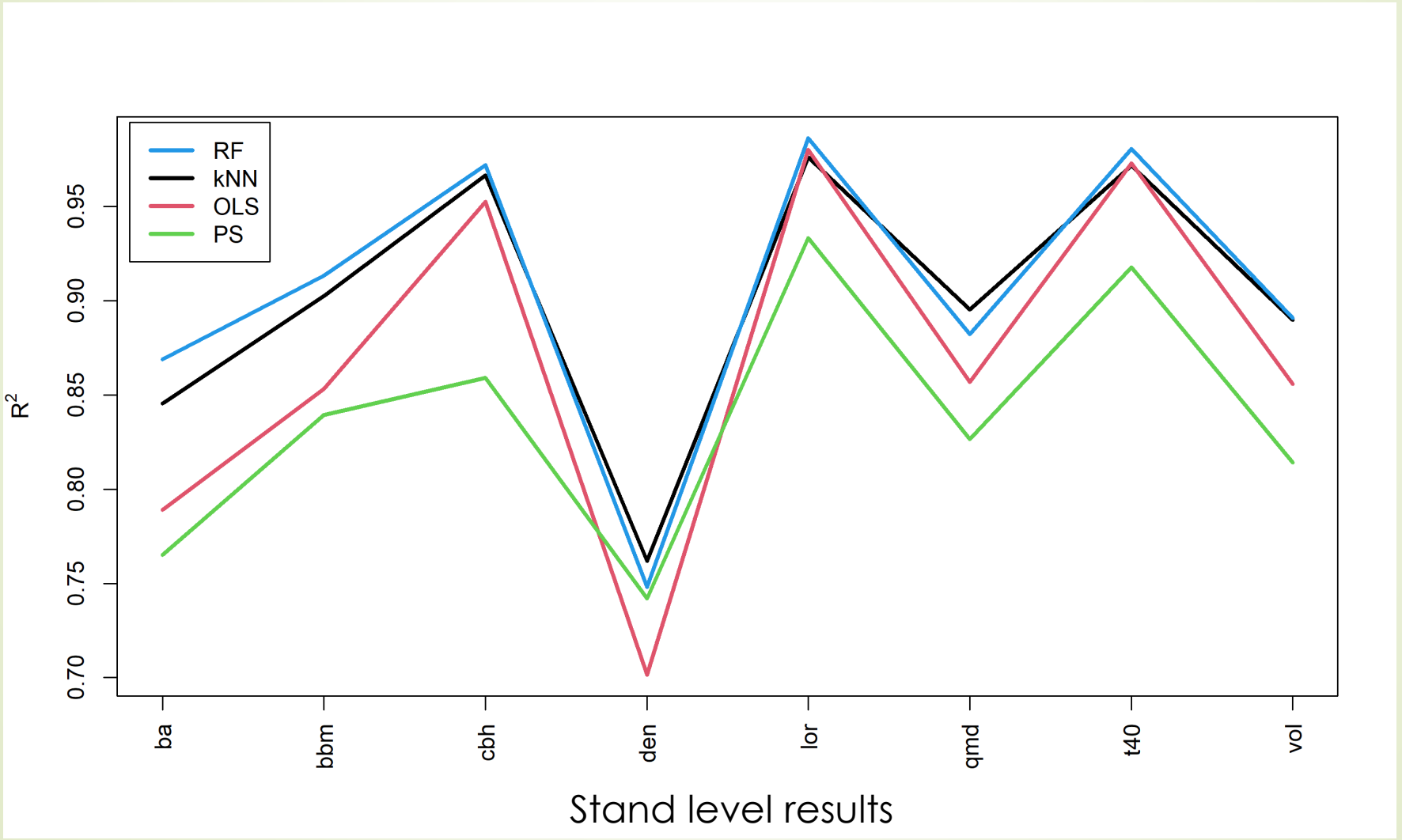
Validation-stand (5 acres)



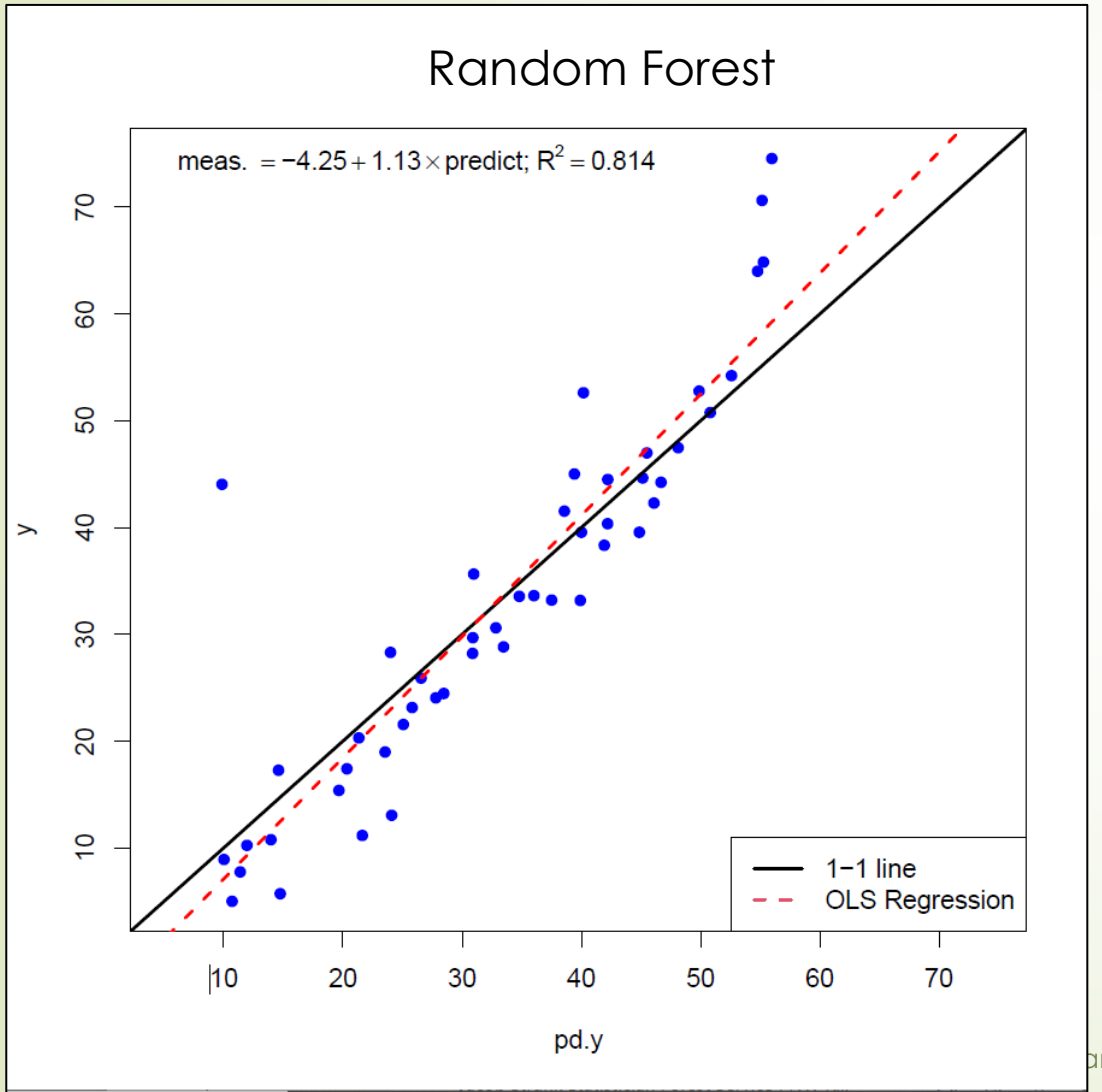


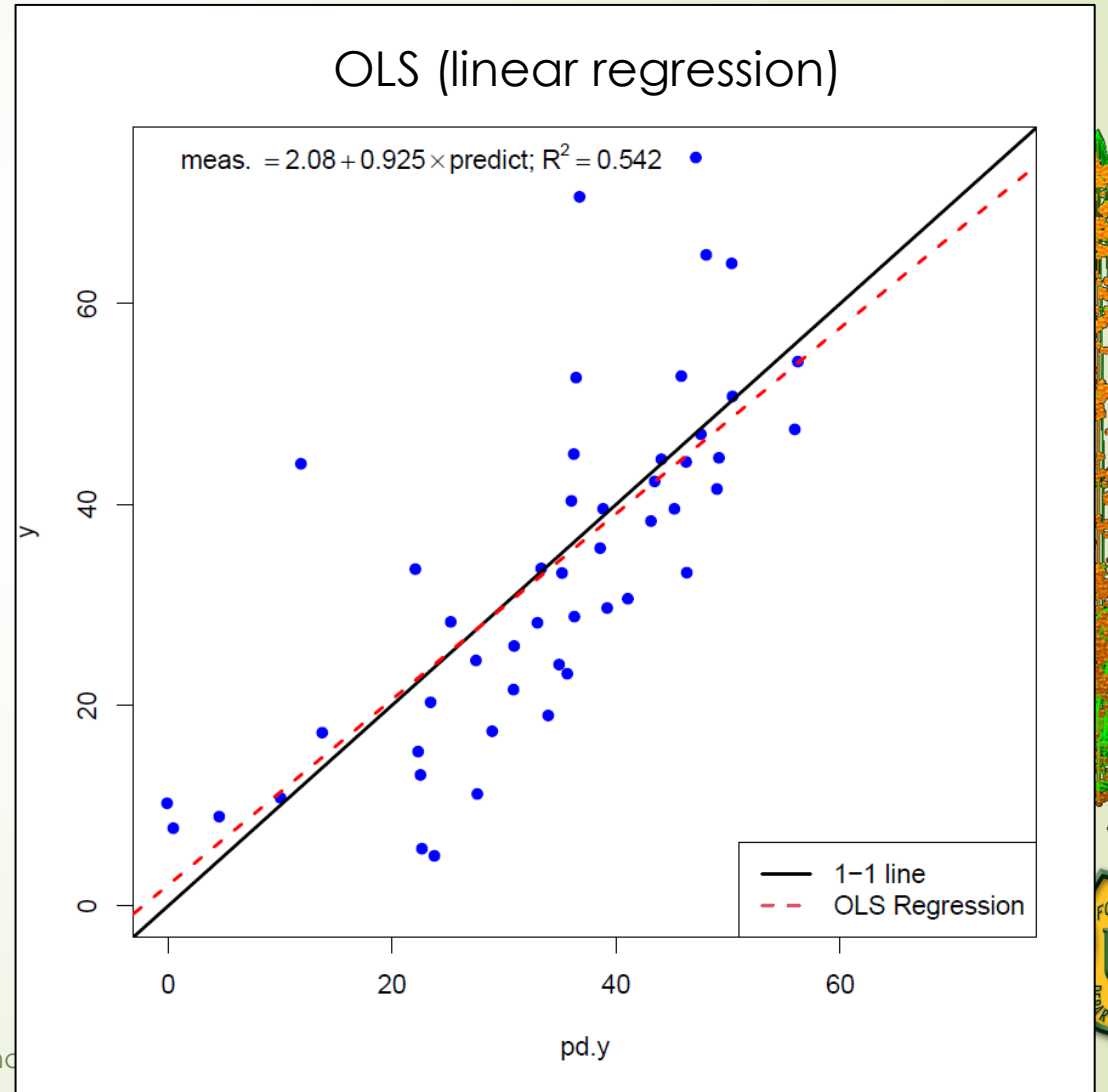
Real Plot locations (irregular)

Model Type (RF and kNN best)

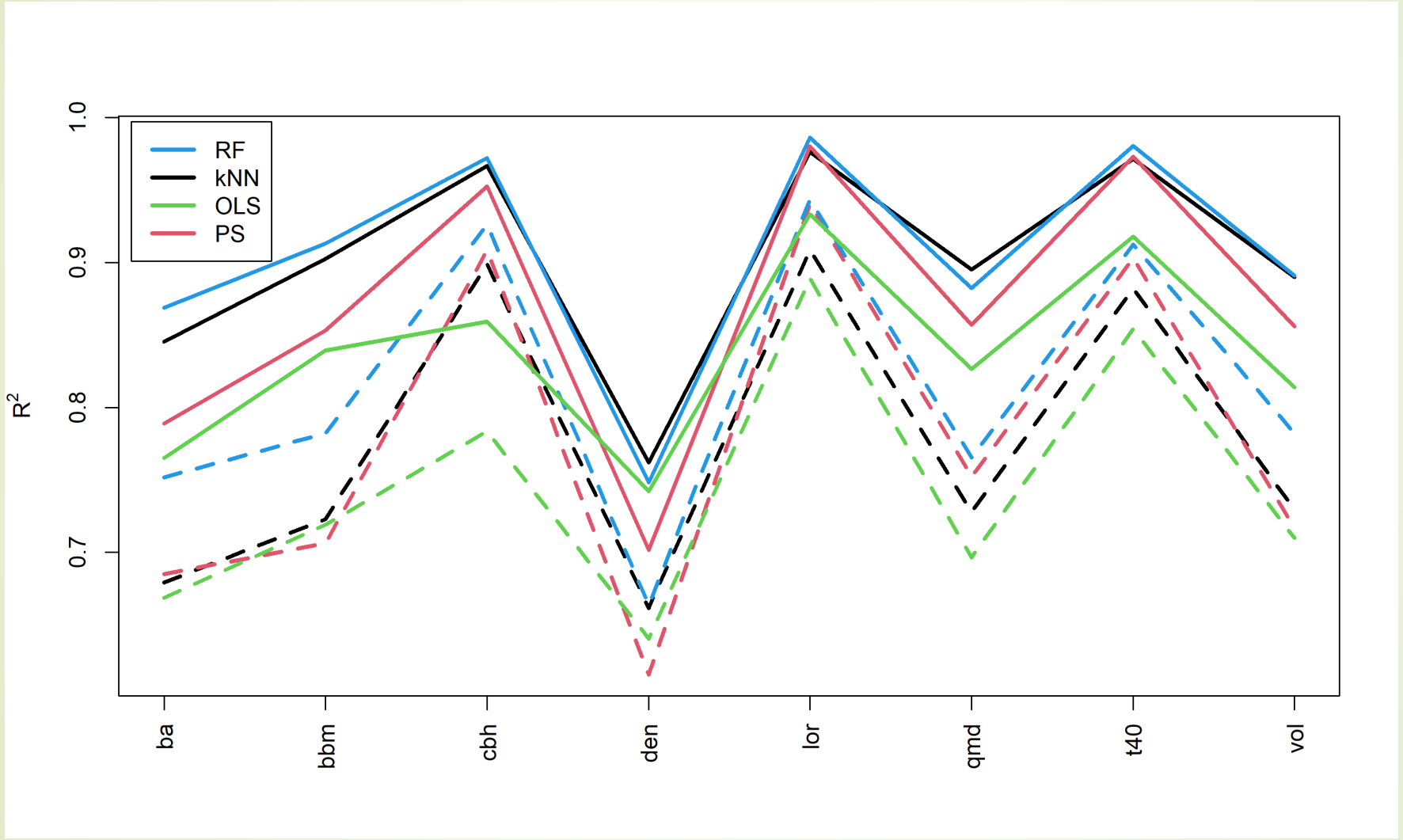


Example Visual Diagnostics (Biomass, SW)

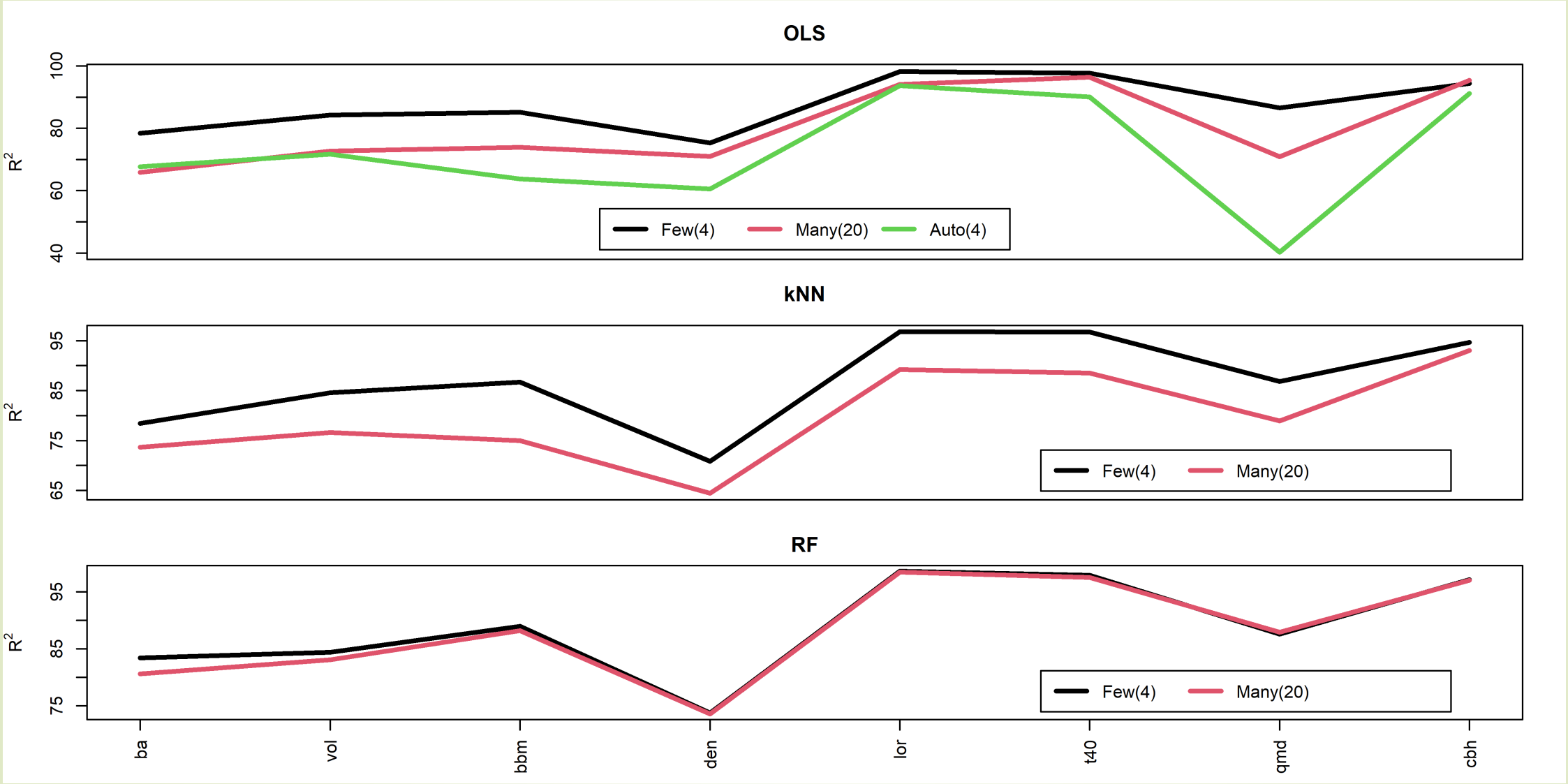




Stand (solid) always better than **Plot** for same model type



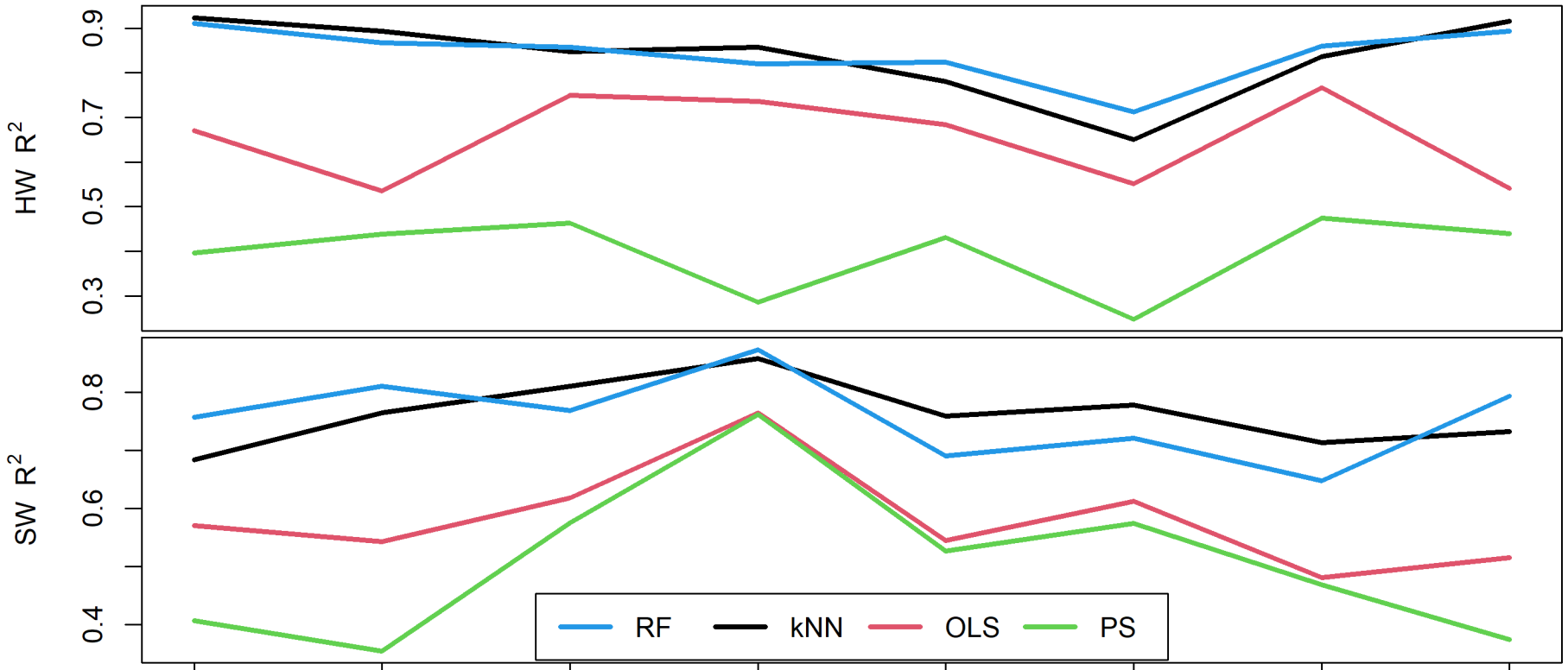
No. Predictors



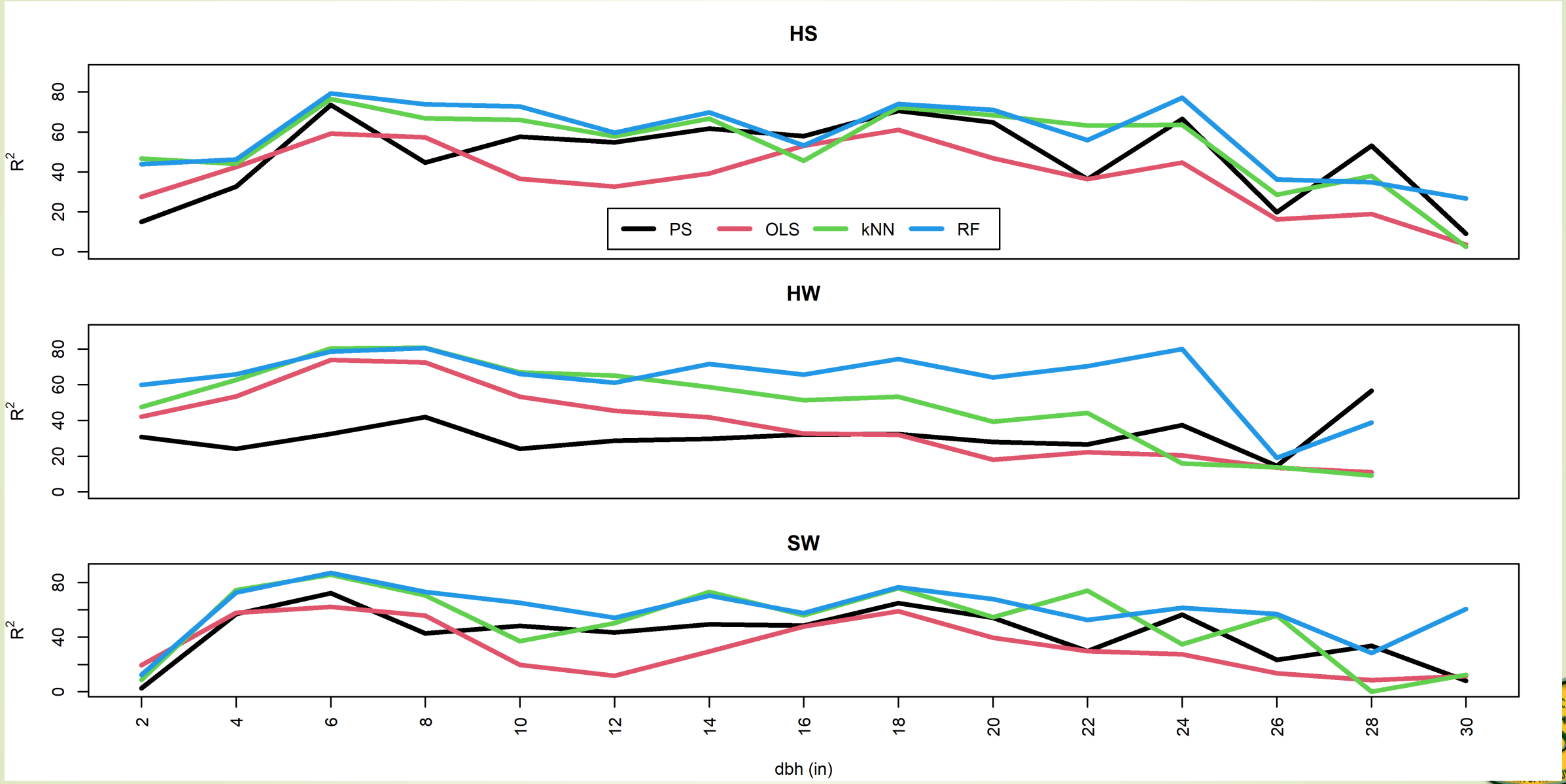
Few=Elev.P20,Elev.P95,Profile.area,HWProportion

Many=Elev.minimum,Elev.mean,Elev.stddev,Elev.skewness,Elev.L1,Elev.P05,Elev.P10,Elev.P20,Elev.P40,Elev.P50,Elev.P60,Elev.P70,Elev.P80,Elev.P90,Elev.P99,Canopy.relief.ratio,Percentage.first.returns.above.2.00,X.All.returns.above.2.00.....Total.first.returns....100,HWProportion,Profile.area

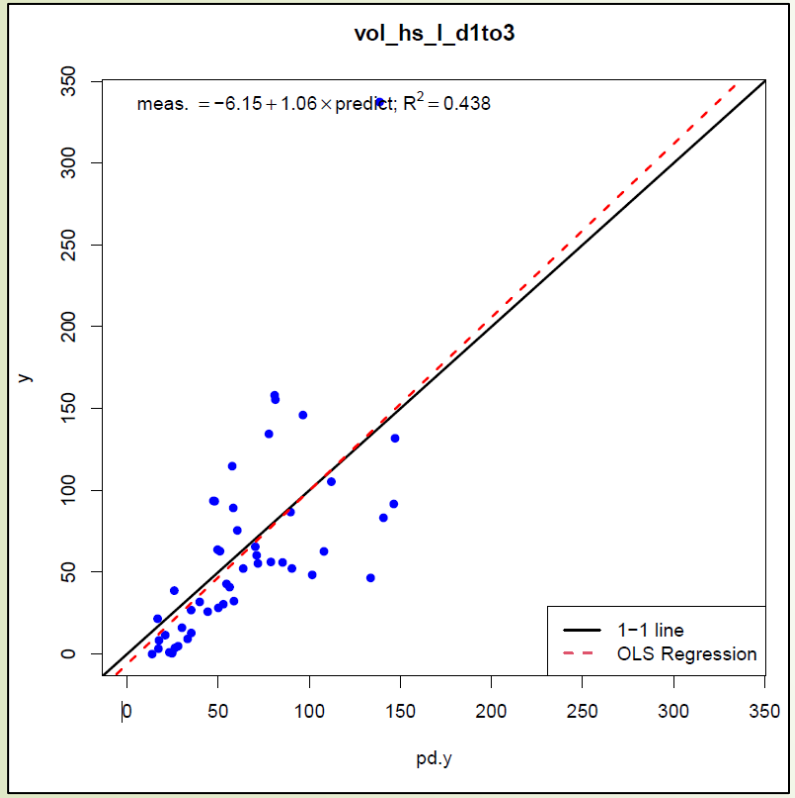
Spp Group

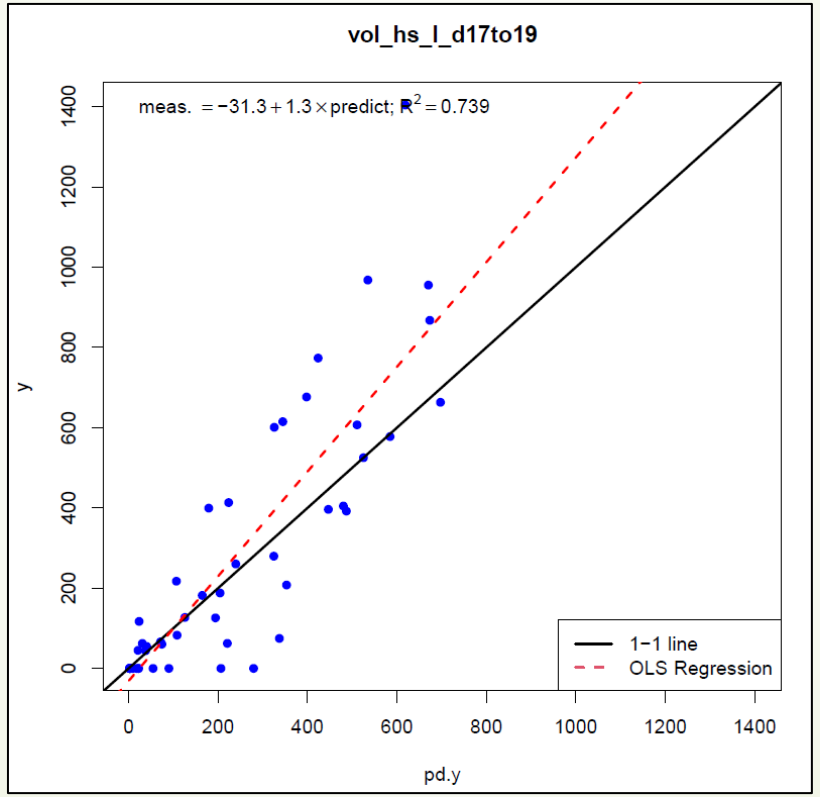

HW R^2
0.3
0.5
0.7
0.9
SW R^2
0.4
0.6
0.8
RF
kNN
OLS
PS
ba
bbm
cbh
den
lor
qmd
t40
vol

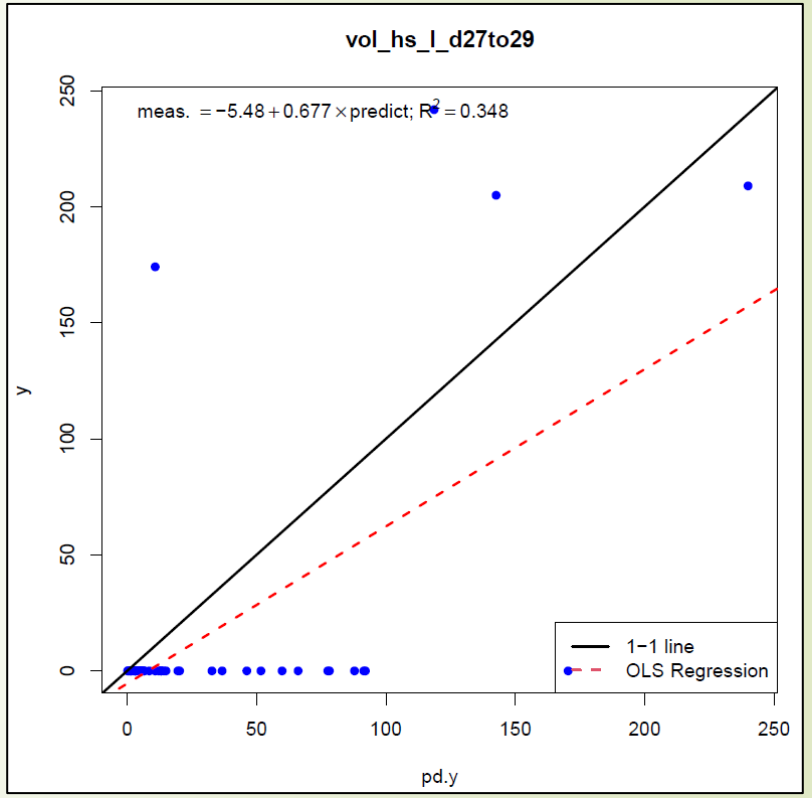
Vol x DBH x Spp Group



RF Results


vol_hs_l_d1to3
meas. = −6.15 + 1.06 × predict; R² = 0.438
1−1 line
OLS Regression
y
pd.y
0
50
100
150
200
250
300
350


vol_hs_l_d17to19
meas. = −31.3 + 1.3 × predict; R² = 0.739
1−1 line
OLS Regression
y
pd.y
0
200
400
600
800
1000
1200
1400


vol_hs_l_d27to29
meas. = −5.48 + 0.677 × predict; R² = 0.348
1−1 line
OLS Regression
y
pd.y
0
50
100
150
200
250

Manuscript nearly complete

- More nuance …

Table ??. Prediction performances ($R^2 \times 100\%$) for forest attributes for all species (hs), hardwood species (hw), and softwood (sw) species

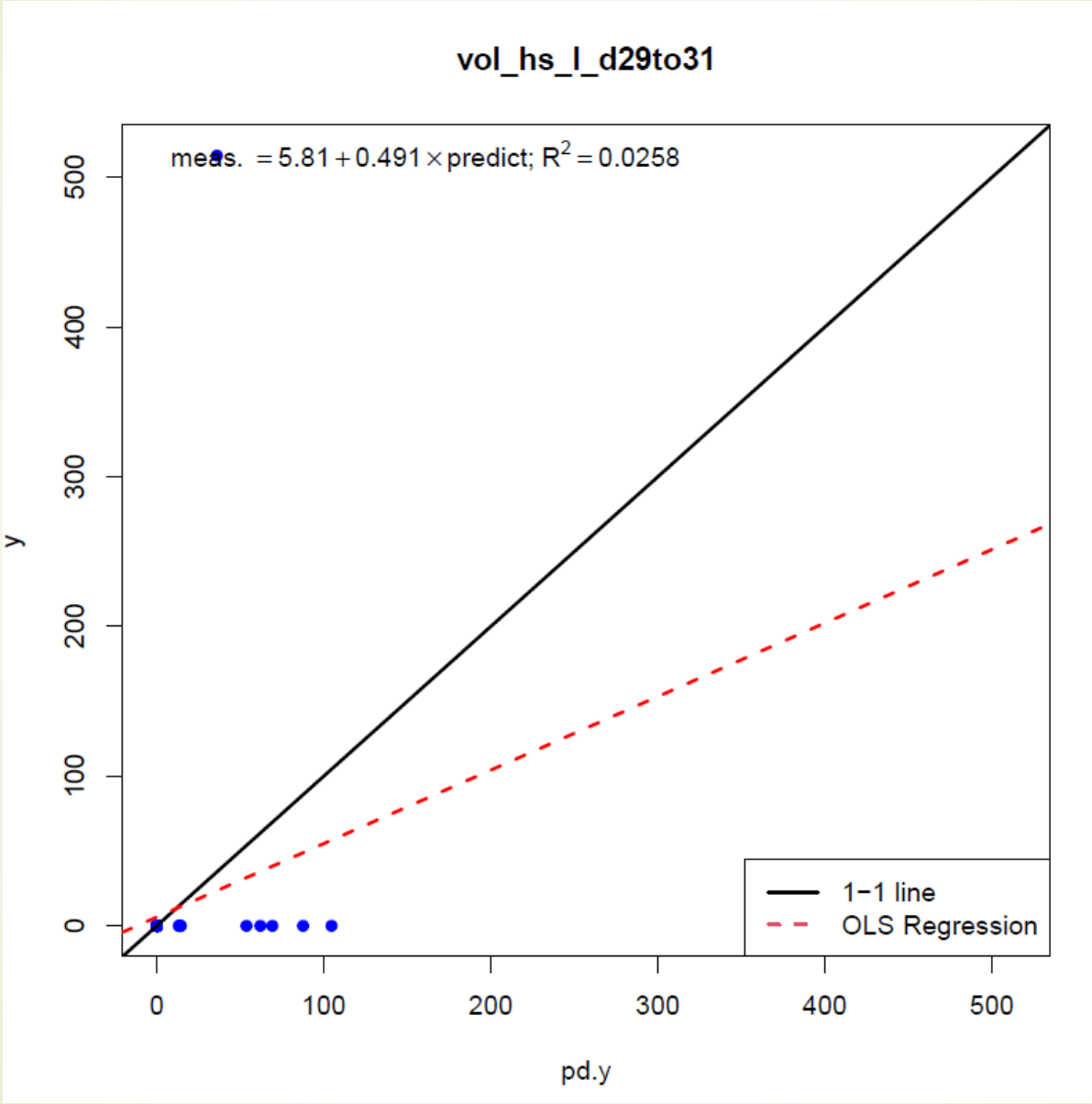
		Plot				Stand			
attribute	group	PS	OLS	kNN	RF	PS	OLS	kNN	RF
ba	hs	67	68	68	75	77	79	85	87
ba	hw	47	54	72	73	40	67	92	91
ba	sw	29	55	61	66	41	57	68	76
bbm	hs	72	71	72	77	84	85	90	90
bbm	hw	48	42	66	68	44	54	89	87
bbm	sw	30	53	68	71	35	54	77	83
cbh	hs	78	91	90	93	86	95	97	97
cbh	hw	54	60	65	67	46	75	85	85
cbh	sw	46	52	64	63	58	62	81	76
den	hs	64	62	66	65	74	70	76	74
den	hw	44	67	74	74	29	74	86	83
den	sw	52	67	76	77	76	76	86	87
lor	hs	89	94	91	95	93	98	98	99
lor	hw	51	53	58	64	43	68	78	82
lor	sw	41	43	54	53	53	54	76	68
qmd	hs	70	75	73	76	83	86	90	88
qmd	hw	40	34	41	47	25	55	65	71
qmd	sw	34	37	38	44	57	61	78	72
t40	hs	85	90	88	91	92	97	97	98
t40	hw	56	63	67	70	48	77	84	86
t40	sw	36	39	52	49	47	48	71	63
vol	hs	71	72	73	78	81	86	89	88
vol	hw	48	42	68	71	44	54	92	89
vol	sw	30	52	66	69	37	51	73	80
median R^2	hs	72	74	73	78	84	86	90	89
median R^2	hw	48	54	66	69	44	68	86	86
median R^2	sw	35	52	62	64	50	56	76	76

Conclusions

- Stand exceeds plot performance
- kNN, RF performed similarly
- RF is most robust (vol etc., dbh, spp)
- kNN is most convenient (tree list), but sensitive to tuning
- PS fared poorly ☹
- OLS was mediocre
- Few targeted predictors > many predictors > Automated VS
- HW Index yielded good Vol x HW performance
- Diameter predictions with RF pretty good

Questions?

kNN


vol_hs_l_d29to31
meas. = 5.81 + 0.491 × predict; R² = 0.0258
y
pd.y
0
100
200
300
400
500
1−1 line
OLS Regression

Hardwood proportion

- Use Random Forest (RF) classification model
  - L-moment coefficient of variation of return heights
  - Canopy cover (all returns > 2m / total 1st returns)
  - Intensity P50 (median)
  - Intensity P10
- Classification happens using 2m cells, proportion is computed for plots and 30m cells (area-wide)
- Final classification model: overall classification accuracy: 94%
  - 70% of 2m cells used to build classifier
  - 30% used to compute overall classification accuracy

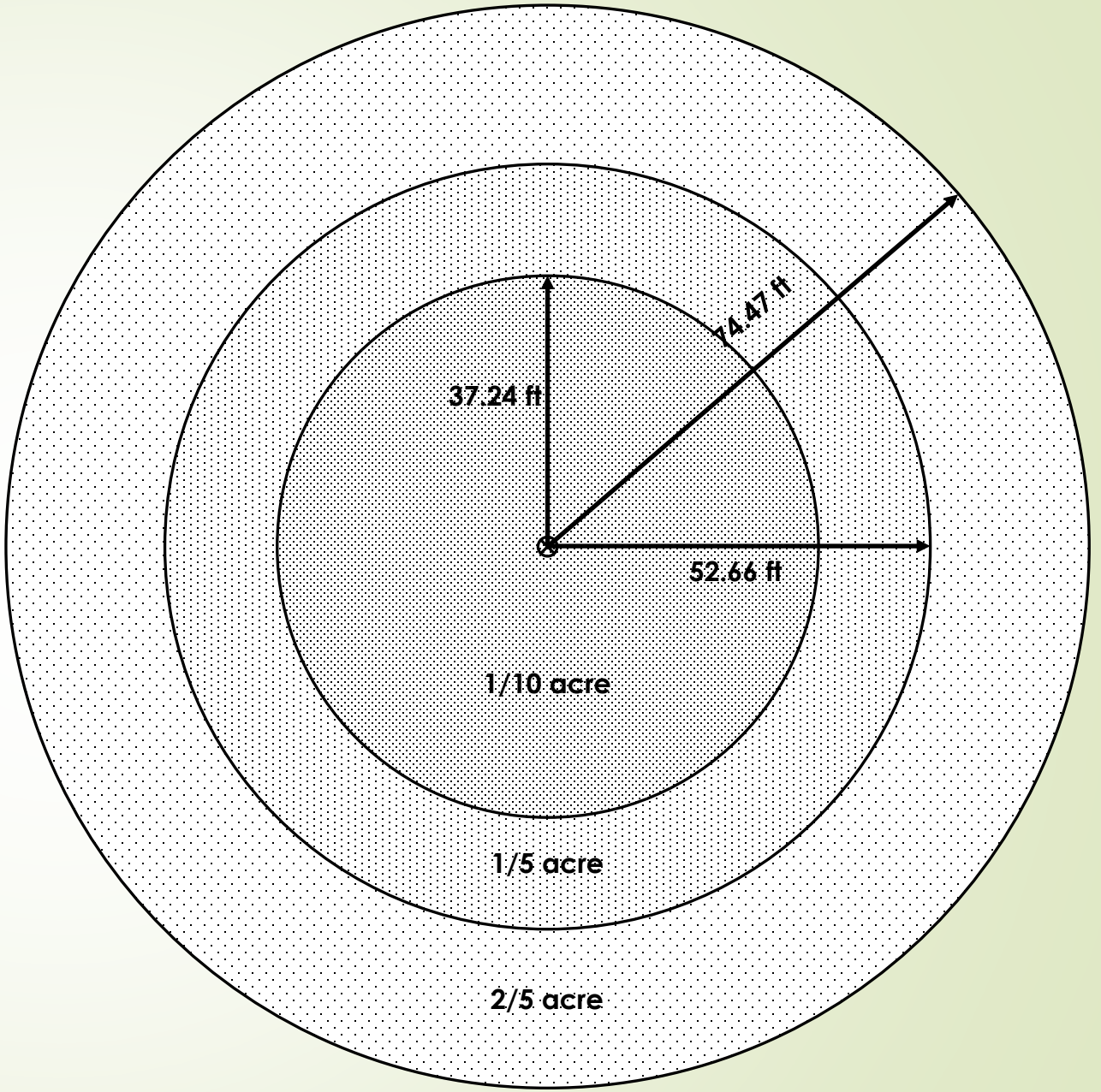
Lidar data

- Acquisition completed March 2018
- 368 miles^2
- 40° FOV, 50% overlap
- ~8+ pulses/m^2 (aggregate)

Lidar System Acquisition Parameters	
Item	**Parameter**
System	Leica ALS-70 HP
Nominal Pulse Spacing (m)	0.7
Nominal Pulse Density (pls/m²)	4.4
Nominal Flight Height (AGL meters)	1500
Nominal Flight Speed (kts)	130
Pass Heading (degree)	142
Sensor Scan Angle (degree)	40
Scan Frequency (Hz)	47.3
Pulse Rate of Scanner (kHz)	339.2
Line Spacing (m)	902
Pulse Duration of Scanner (ns)	4
Pulse Width of Scanner (m)	0.33
Central Wavelength of Sensor Laser (nm)	1064
Sensor Operated with Multiple Pulses	Yes
Beam Divergence (mrad)	0.22
Nominal Swath Width (m)	1,150
Nominal Swath Overlap (%)	50
Scan Pattern	Triangle

Plot design

Trees >= 3” DBH on 1/10-acre plot	Plot Size (acre)	# plots
< 9	2/5	217
< 19	1/5	218
19+	1/10	113

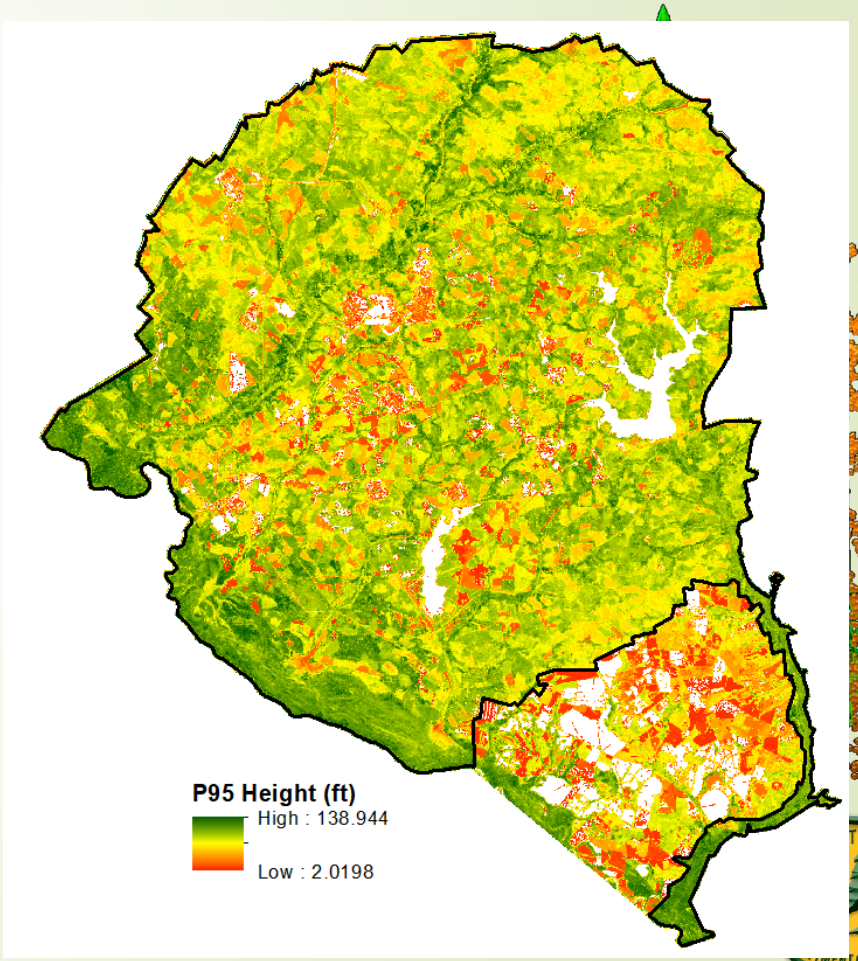


GNSS survey for all plots

- Javad Triumph2 receivers on tripods
- 15+ minute occupations with 1 second epochs
- Post-processed
- Anticipated accuracy better than 1m horizontal
- Measured accuracy
  - 10 large rocks that could be identified in point cloud
  - Collect for 15+ minutes
  - Digitize high point on rock in 2009 point cloud
  - Ave difference: 0.52m; StdDev:0.27m; min: 0.14m; min: 1.10m

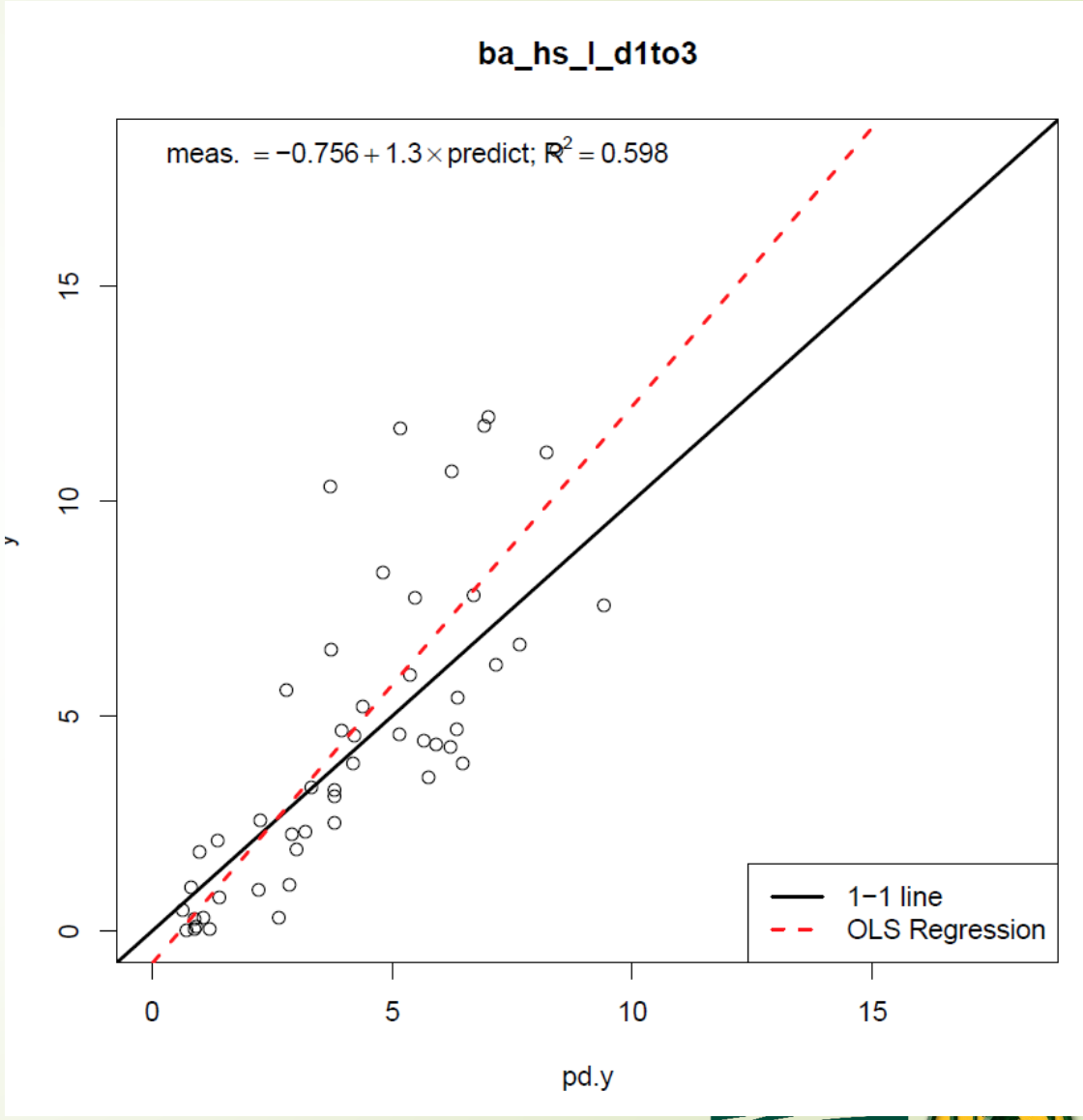
Lidar metrics

- Clipped point data using plot location and size
- Computed full set of metrics using height & intensity for all plots (grid & mini-stand)
- Also computed metrics over acquisition area for 30m cells

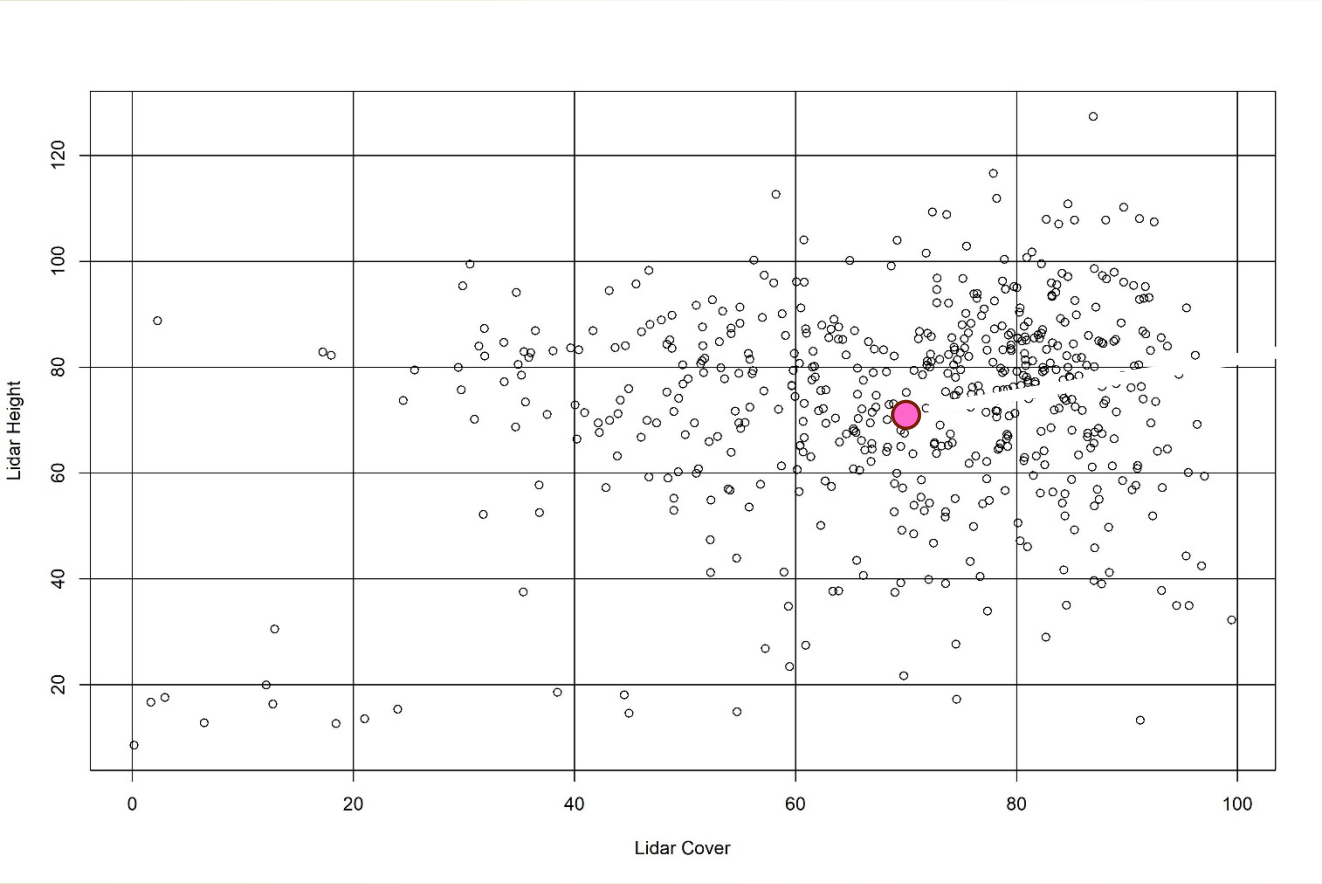


Analysis Overview

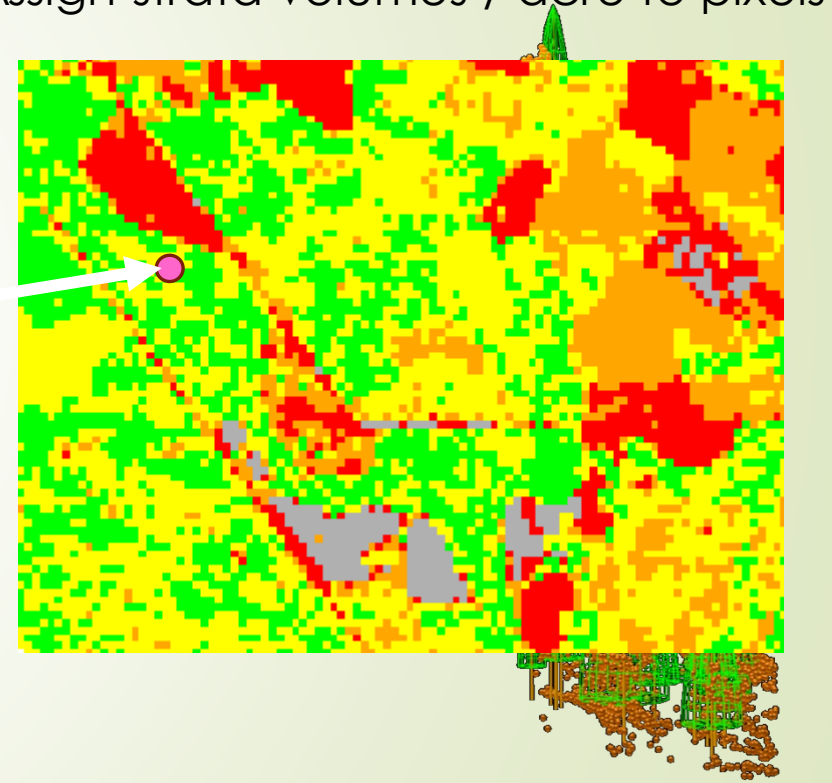
- Compare: Post Stratification (PS), OLS, kNN
- **Validate with mini-stands**
  - Measured versus predicted
- Suite of forest metrics
  - BA
  - BA 1 to 3 inches, 3-5 inches …
  - Vol
  - Biomass
  - …



Background: Post-Stratification



Assign strata volumes / acre to pixels



Background: kNN Imputation

- Prediction
  - Get P90 for some pixel (target)

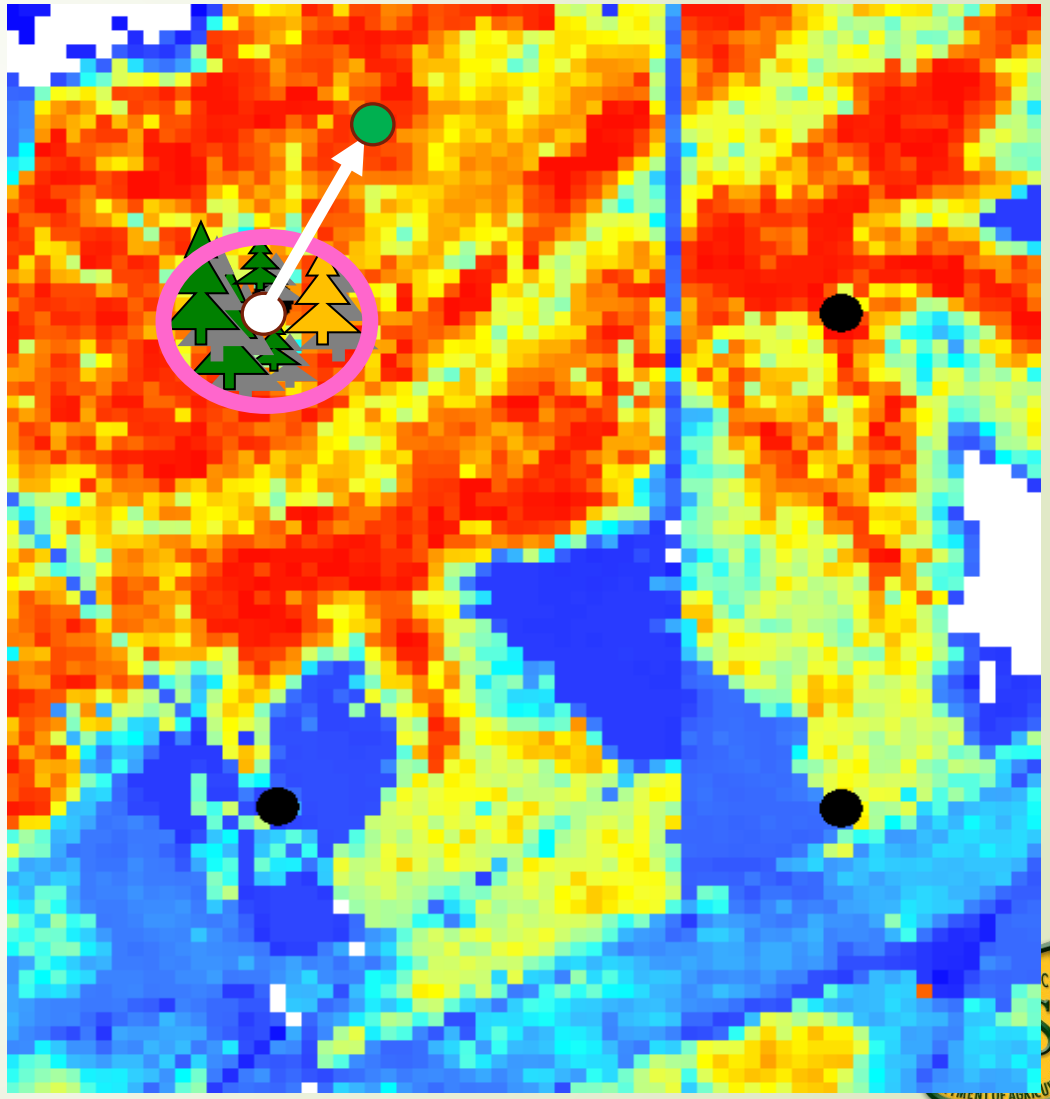
  - 74'
  - Find a field plot with similar P90 (donor)

  - 78'

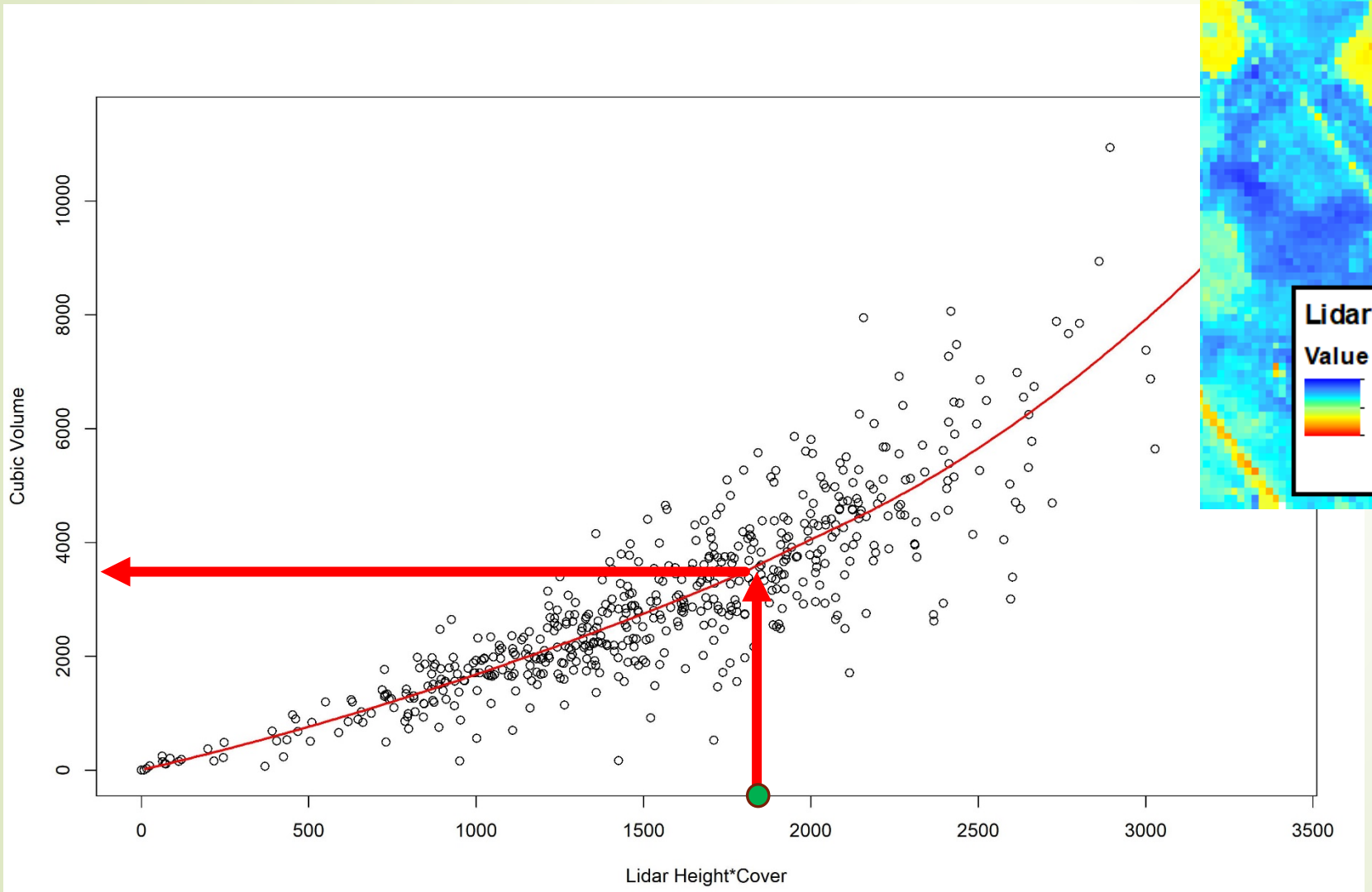
  - Grab Volume from donor

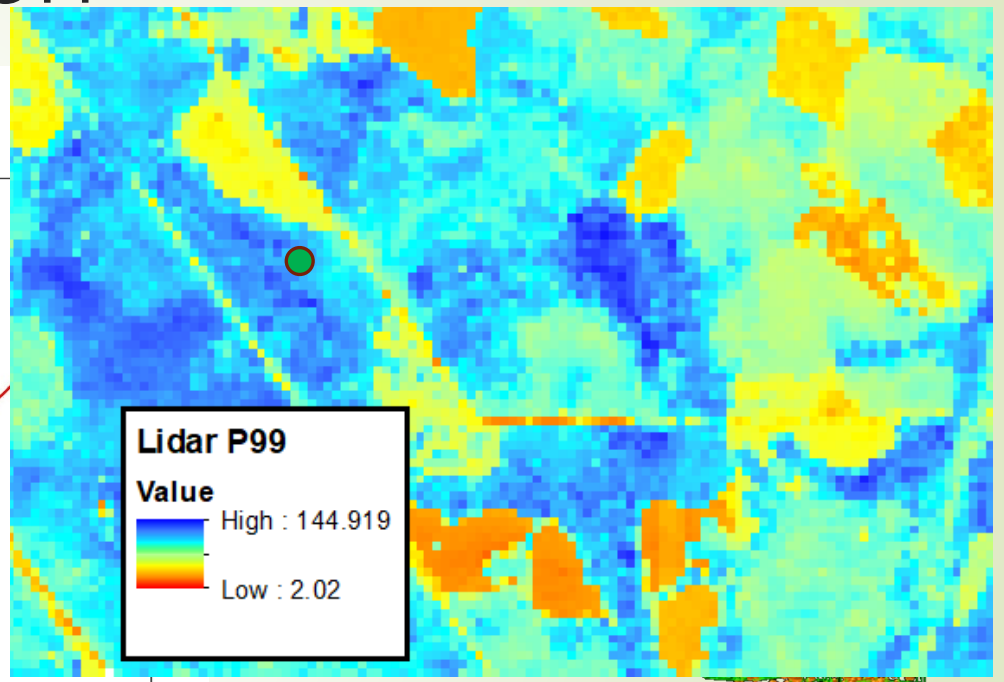
  - 6000 ft^3/acre
  - Give to target

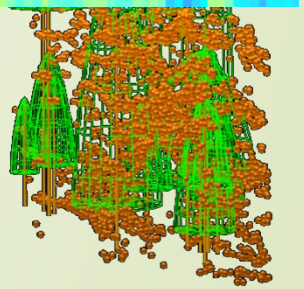
  -> 6000 ft^3/acre



Background: OLS Regression







FBRI FPS Integration

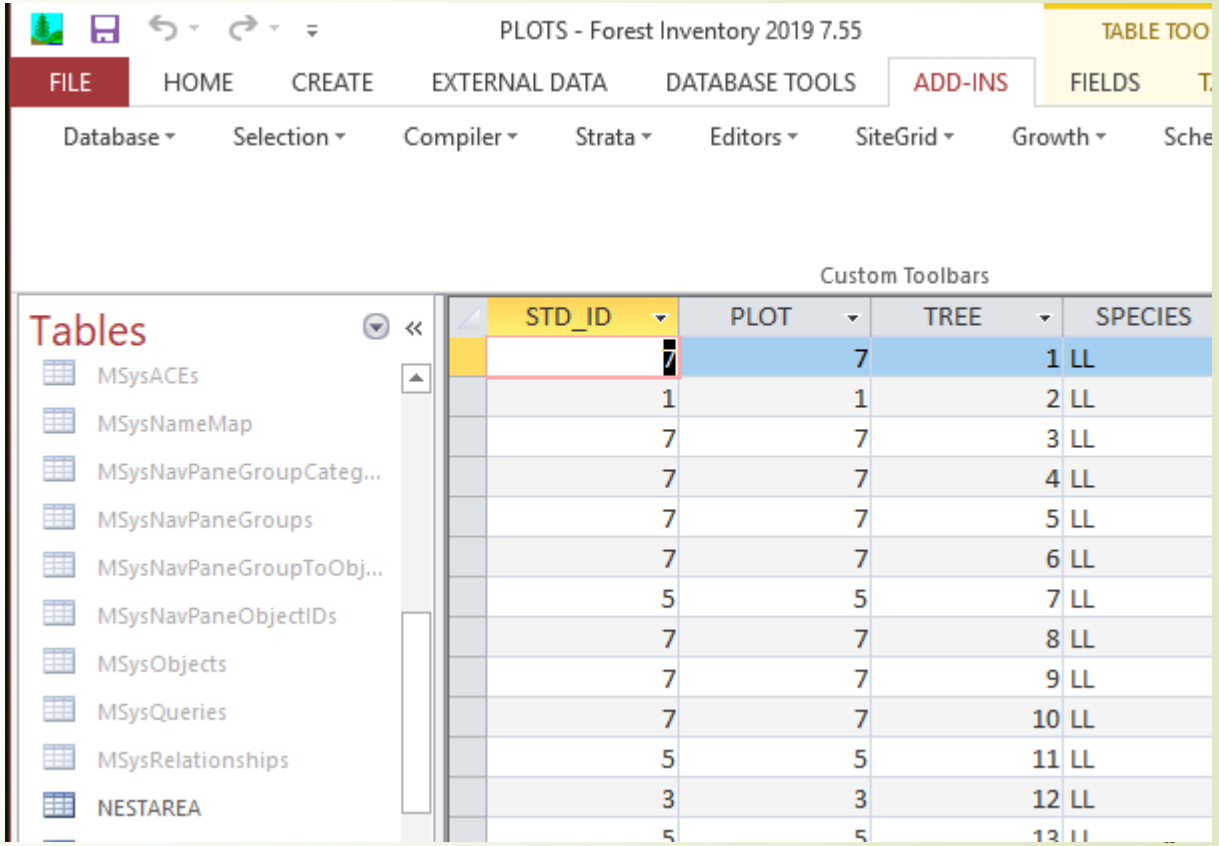
Important to Land manager

In progress:

- Impute tree lists to entire ownership
- Add to FPS
- Compare mini-stand TL projections with predicted TL projections

Current hurdles:

- species crosswalk
- link sqlite database (too big for Access)



The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.)

Tree List Imputation

- kNN pretty convenient
- Pull in actual tree records
- Previous study was promising

forests

MDPI

Article

An Examination of Diameter Density Prediction with k-NN and Airborne Lidar

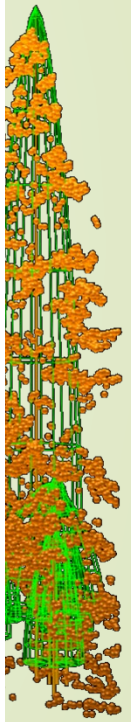
Jacob L. Strunk [1,*], Peter J. Gould [2], Petteri Packalen [3], Krishna P. Poudel [4], Hans-Erik Andersen [1] and Hailemariam Temesgen [4]

1 USDA Forest Service Pacific Northwest Research Station, University of Washington, P.O. Box 352100, Seattle, WA 98195-2100, USA; handersen@fs.fed.us
2 Washington State Department of Natural Resources, P.O. Box 47000, 1111 Washington Street SE, Olympia, WA 98504-7000, USA; Peter.Gould@dnr.wa.gov
3 Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland; petteri.packalen@uef.fi
4 Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, Peavy 204, OR 97331-5706, USA; Krishna.Poudel@oregonstate.edu (K.P.P.); hailemariam.temesgen@oregonstated.edu (H.T.)
* Correspondence: jstrunk@fs.fed.us; Tel.: +1-(541)-737-4457; Fax: +1-(541)-737-4316

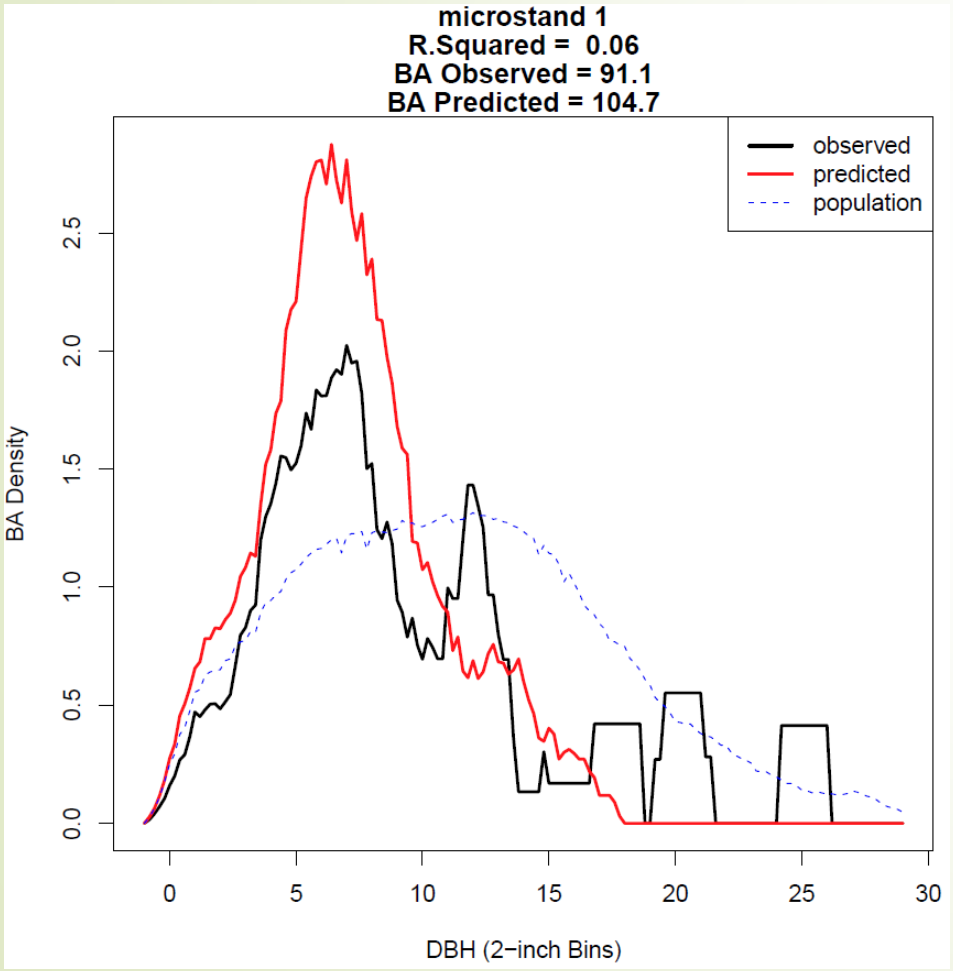
Received: 29 September 2017; Accepted: 10 November 2017; Published: 16 November 2017

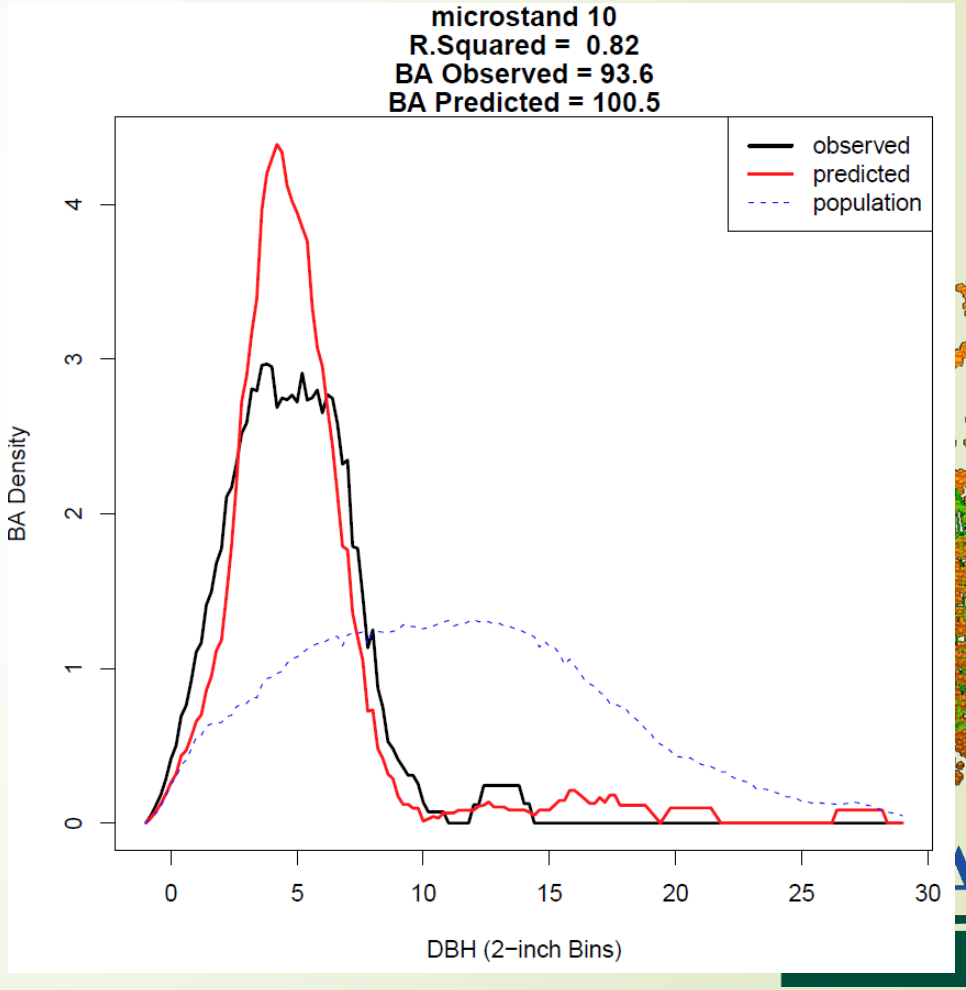
Abstract: While lidar-based forest inventory methods have been widely demonstrated, performances of methods to predict tree diameters with airborne lidar (lidar) are not well understood. One cause for this is that the performance metrics typically used in studies for prediction of diameters can be difficult to interpret, and may not support comparative inferences between sampling designs and study areas. To help with this problem we propose two indices and use them to evaluate a variety of lidar and k nearest neighbor (k-NN) strategies for prediction of tree diameter distributions. The indices are based on the coefficient of determination (R^2), and root mean square deviation (RMSD). Both of the indices are highly interpretable, and the RMSD-based index facilitates comparisons with alternative (non-lidar) inventory strategies, and with projects in other regions. K-NN diameter distribution prediction strategies were examined using auxiliary lidar for 190 training plots distribute across the 800 km^2 Savannah River Site in South Carolina, USA. We evaluate the performance of k-NN with respect to distance metrics, number of neighbors, predictor sets, and response sets. K-NN and lidar explained 80% of variability in diameters, and Mahalanobis distance with k = 3 neighbors performed best according to a number of criteria.

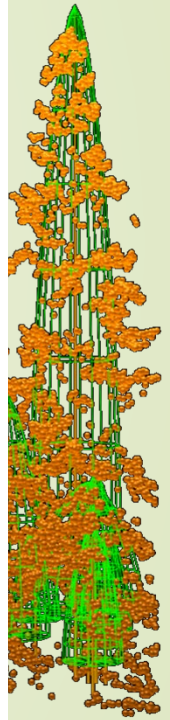
Keywords: forest inventory; dbh; diameter distribution; performance criteria



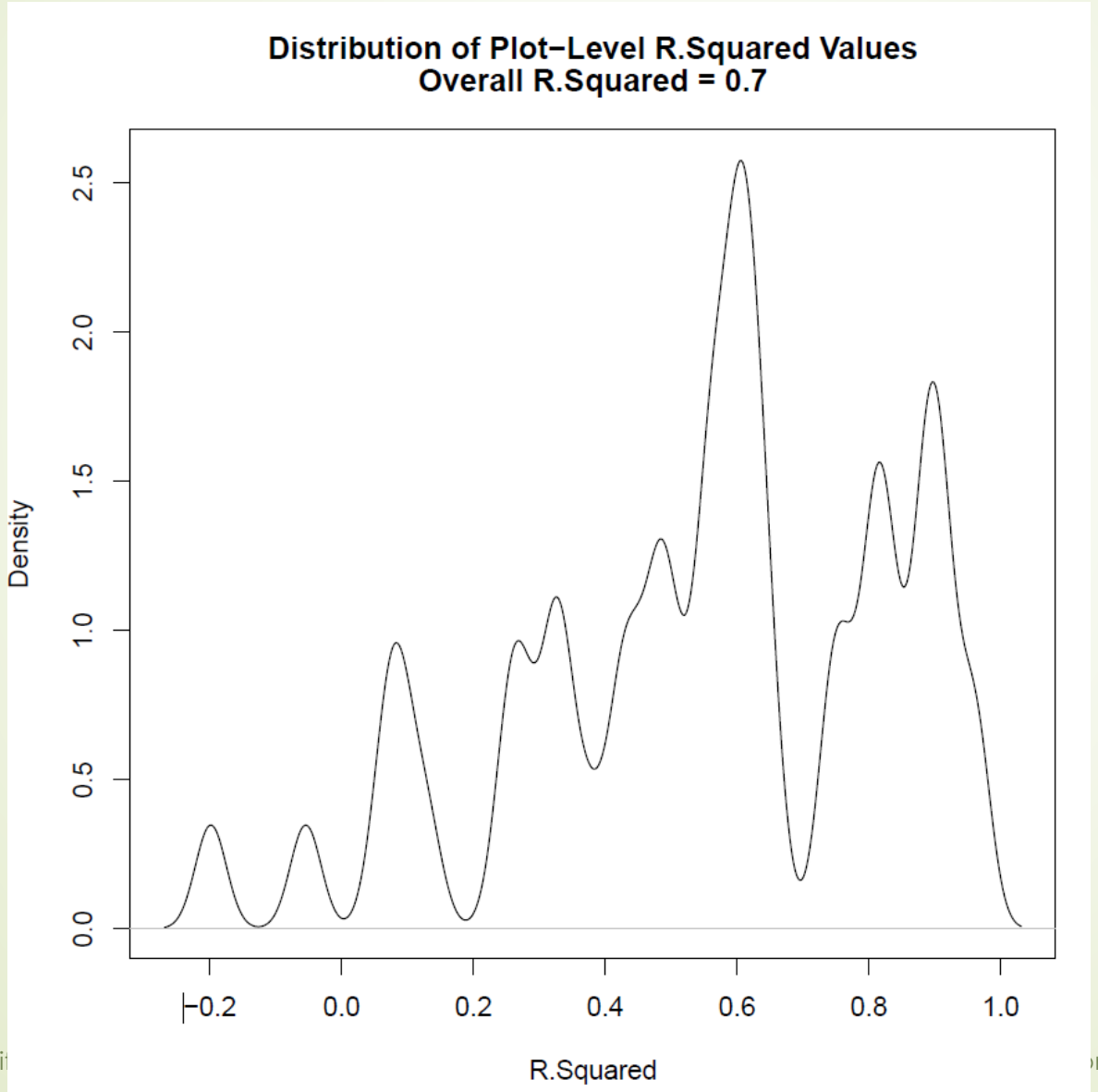
kNN Tree Lists: Example Stands







Tree List Performance: Overall



HW / SW Classification

- Identify plots with only hardwood or conifer species for training
  - Used total basal area on plot: 39 hardwood plots, 25 conifer plots
- Compute metrics (height and intensity) using 2m cells
- Isolate 1st returns within 2m of the canopy surface and compute another full set of metrics
- Use Random Forest (RF) with all metrics and then use variable importance scores to select 4 variables for the classification model
  - L-moment coefficient of variation of return heights
  - Canopy cover (all returns > 2m / total 1st returns)
  - Intensity P50 (median)
  - Intensity P10

HW / SW Classification (cont)

- Run RF again with only the 4 variables (split data for training and validation: 70% and 30%)
- Final classification model: overall classification accuracy: 94%
- Use the classification results for 2m cells to compute the proportion of the plot occupied by hardwood species
- Repeat the process using 2m metrics over entire acquisition area to populate 30m cells with the hardwood proportion

Plot protocol

- Trees < 3" DBH (1/10-acre plots)
  - Tally live trees by DBH class (1-2" & 2-3") and species
- Trees ≥ 3" DBH (live and dead) (plot size varies)
  - Species, DBH, total height, crown class