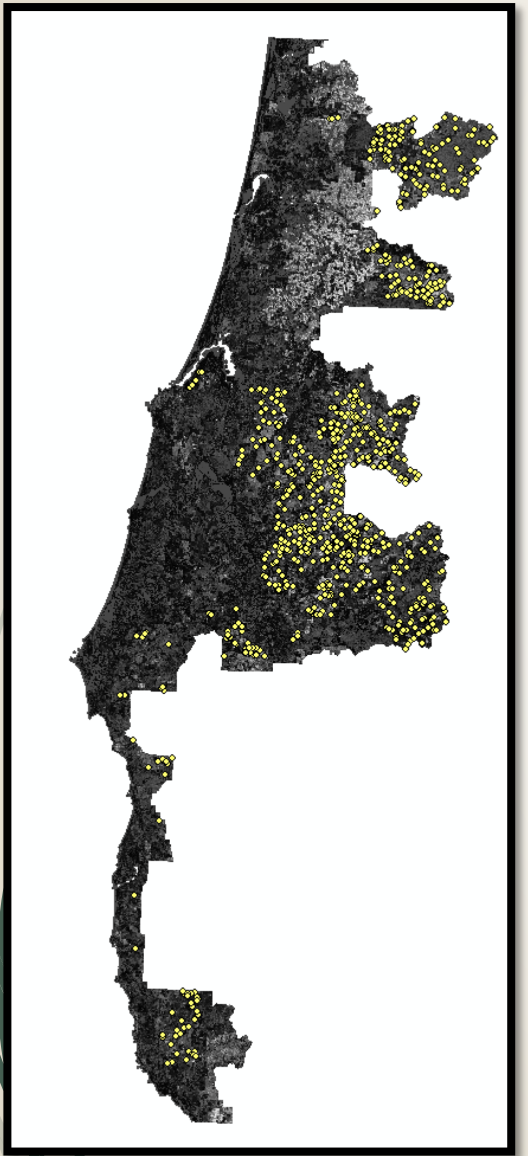
Species-composition Modeling Using Lidar and Spectral Indices

Francisco Mauro, Research Associate
Forest Engineering, Resources and Management
College of Forestry, Oregon State University

Objective:

Test the limits for species identification using easily accessible remote sensing dataset

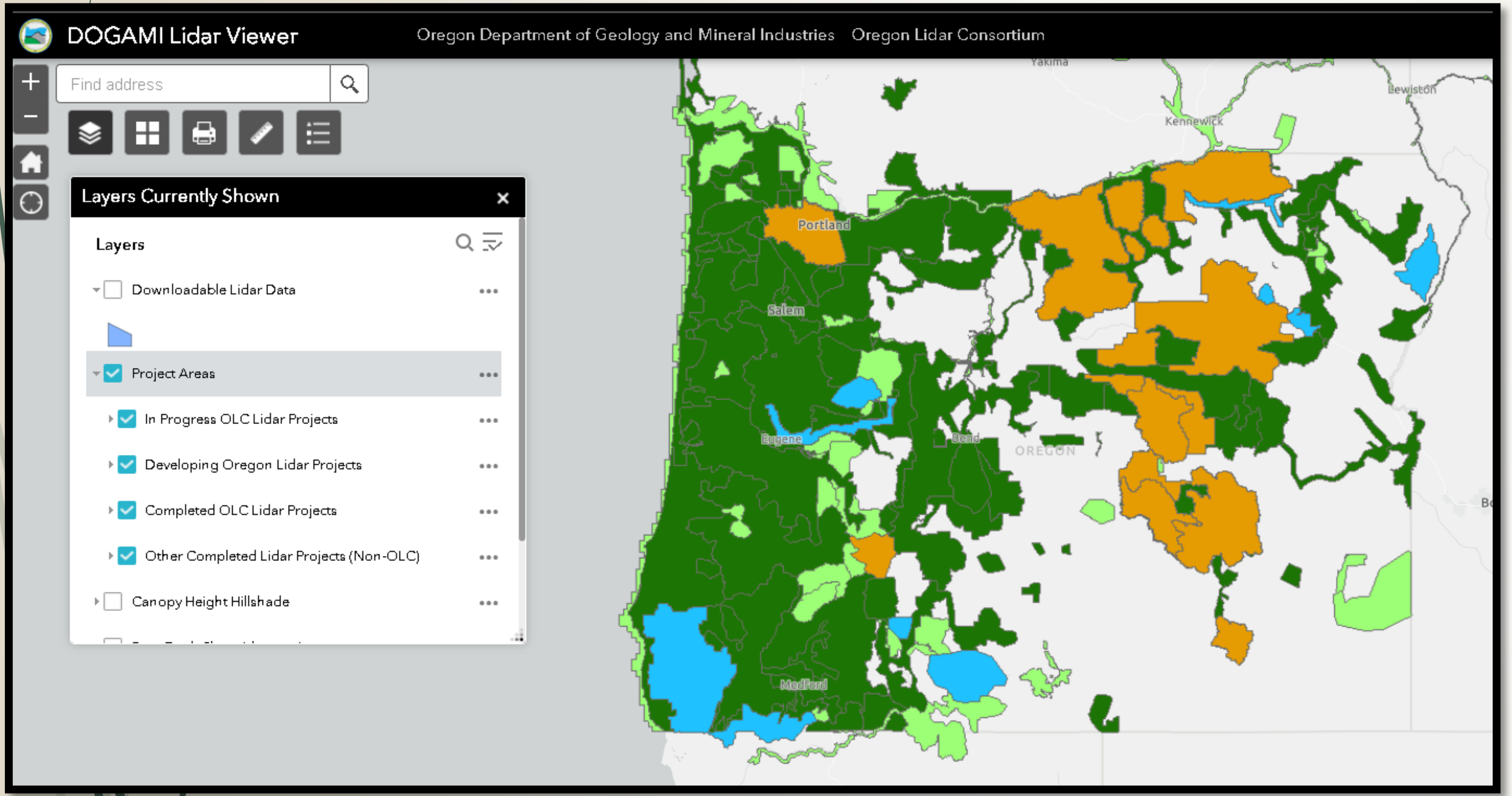
Study area



- Coos Bay lidar acquisition (2009)
- 899 ground fixed radius with GPS coordinates plots of 1/10 ac
  - **Positive:** High quality ground data
  - **Positive:** Pre-existing models for several structural attributes to complement the species classification
  - **Negative:** High dominance of Douglas fir in the area

Sources of auxiliary information

Lidar predictors (Structure)



Elevation metrics

+

Return proportions

+

Latitude & Longitude

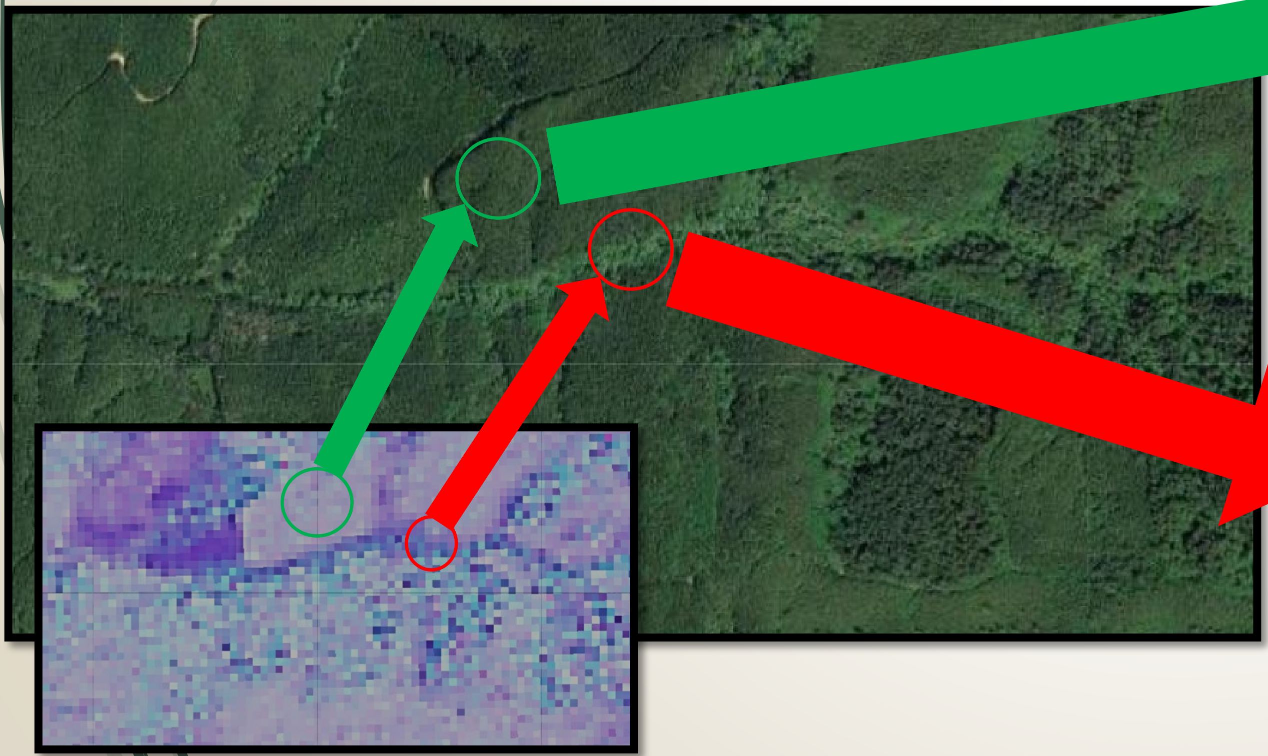
---

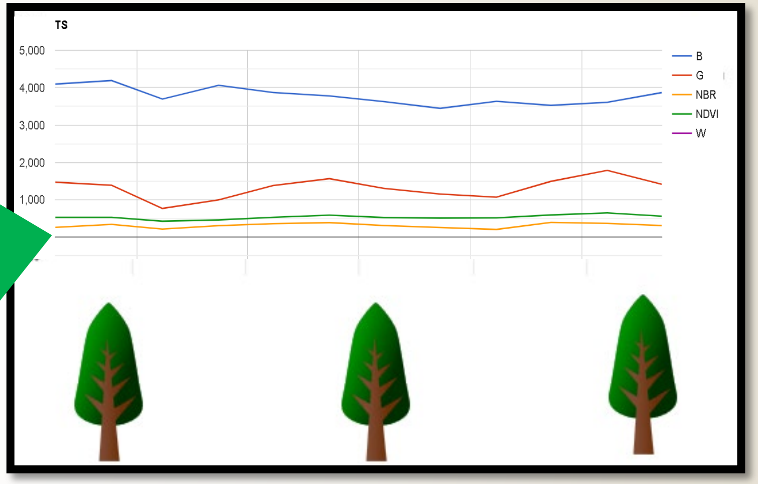
90 Lidar metrics

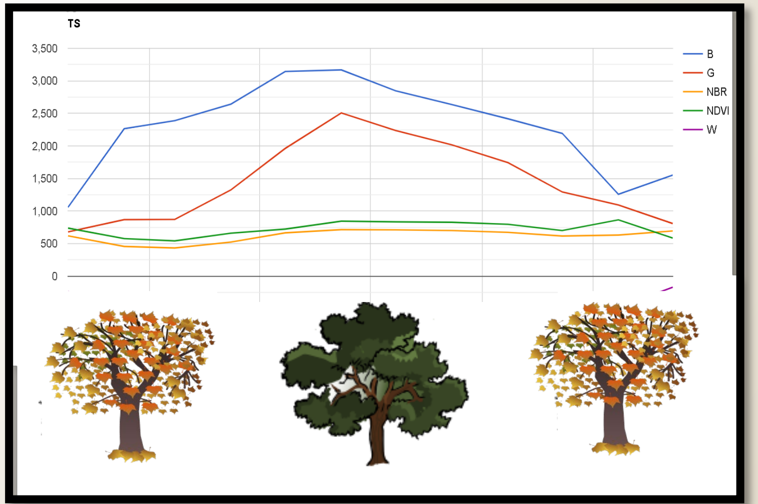
Landsat & Sentinel 2 predictors (Phenology) Google earth engine

Monthly means, max, min and SD for 5 spectral indexes

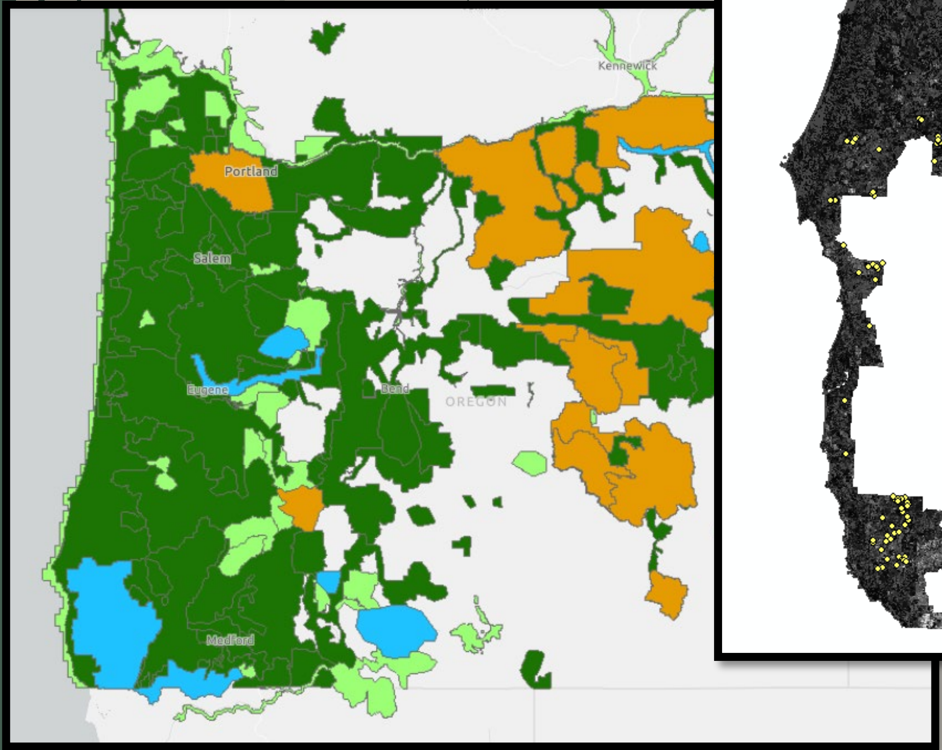
Conifers→Stable spectral signatures
Broad-leaved→Varying spectral signatures (spring vs winter)

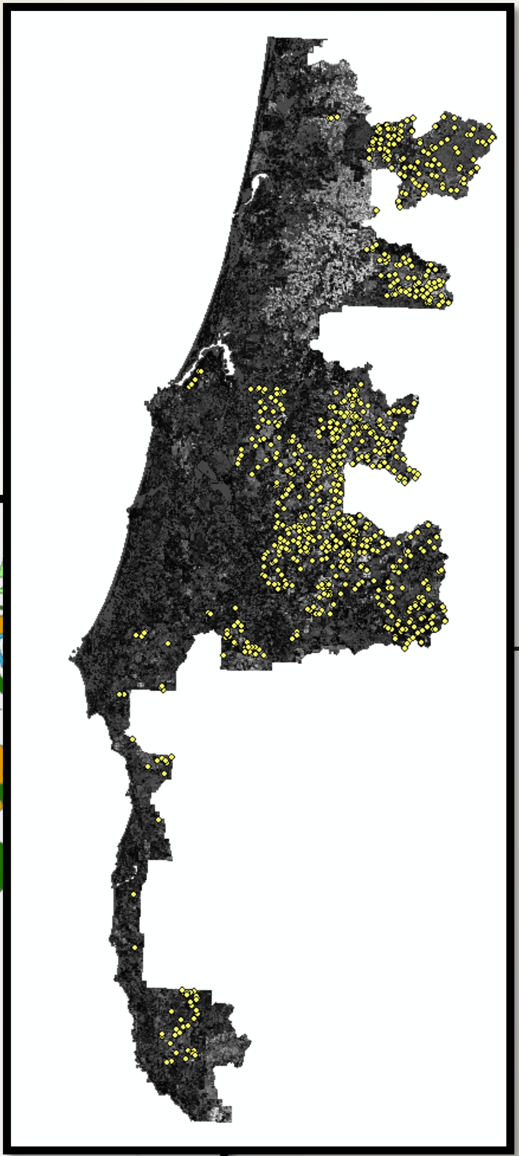






Lidar intensity metrics



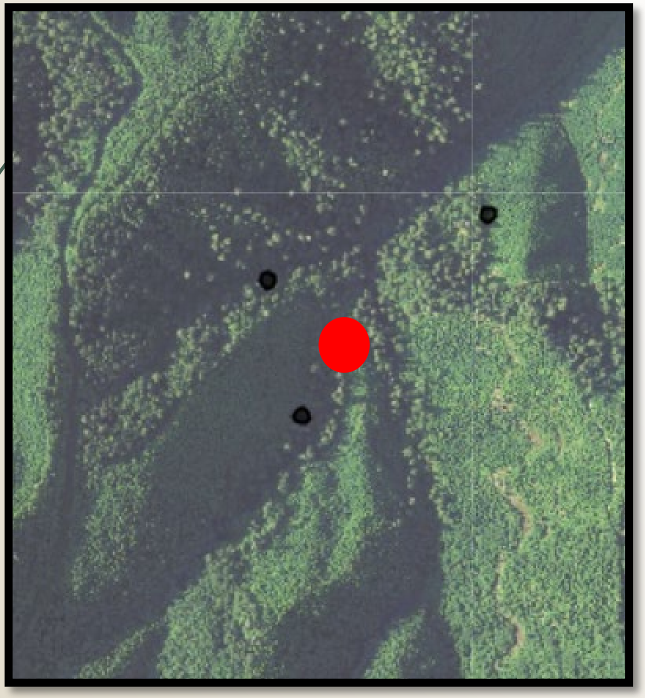


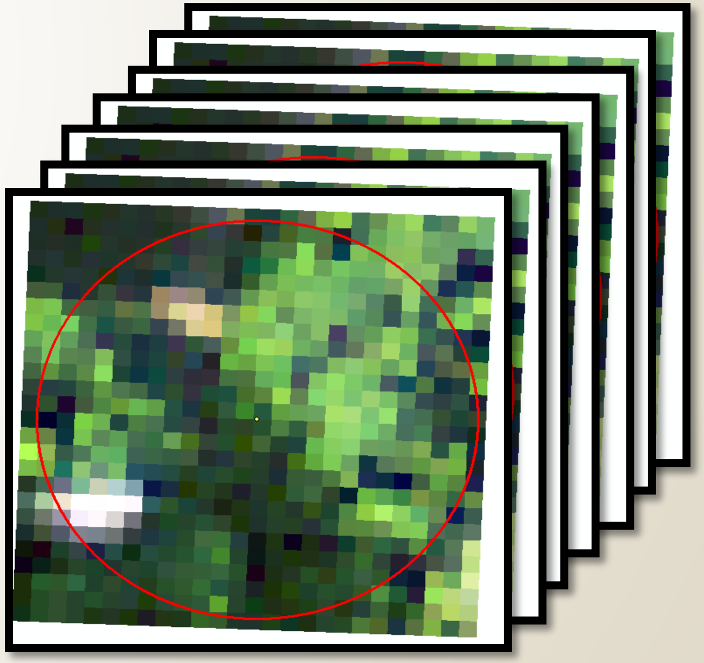
Intensity metrics

30 Lidar metrics

NAIP imagery from Google earth engine

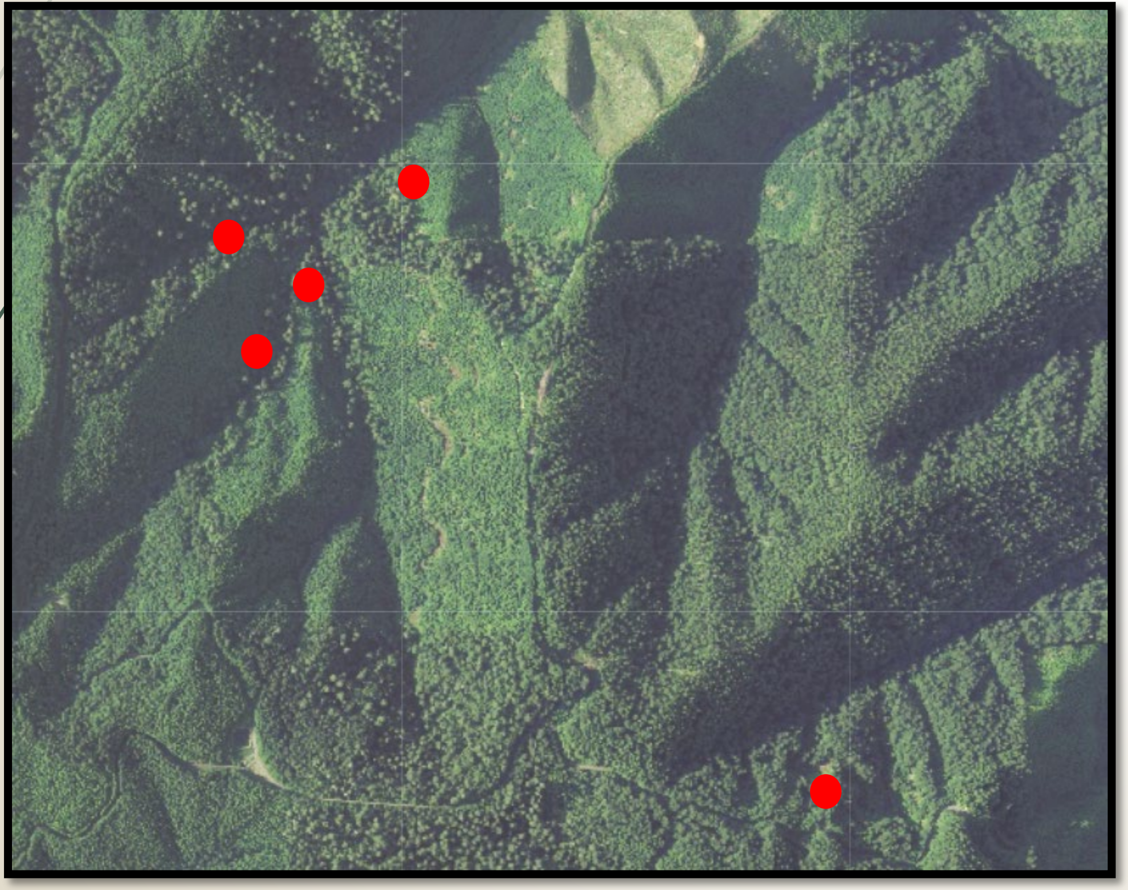
Image “chips” (26 x 26 pixels) for each plot
4 Bands per chip: R, G, B, NIR, 4 band ratios NDVI, NDWI, bare earth mask & shadows mask





NAIP imagery from Google earth engine

Year 2009: 1 m resolution
Visual inspection, relief shadows and tiling effects
Discarded



Dimensionality reduction of auxiliary information

Lidar Elevation metrics

Lidar Return proportions

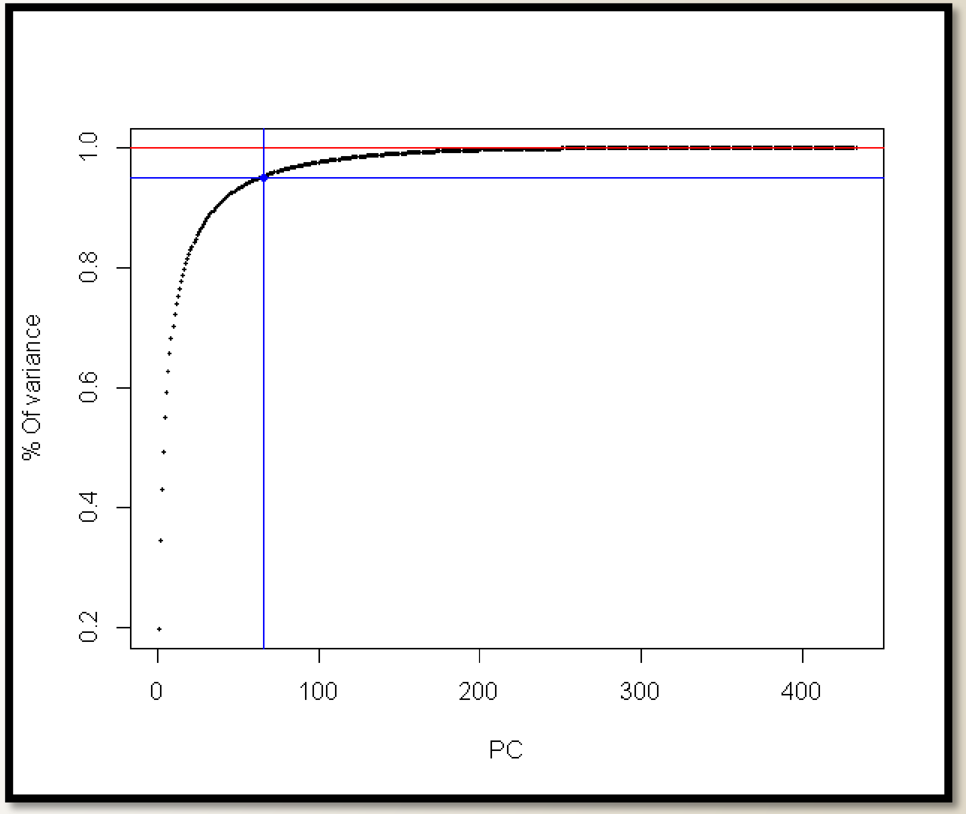
Latitude & Longitude

Lidar Intensity metrics

Landsat indexes

Sentinel 2 Indexes

PCA



66Principal components

Ground data

Ground data

- Tree measurements of DBH, Species tree height→Volume
- Plots contain several trees → Systematic way to define categories for species classification
- Plot sizes (1/10 ac)←→ Stand size (20-50 ac) (200-500 modeling units)

Categories based on dominant species

Dominance defined based on Volume proportion

Dominant species

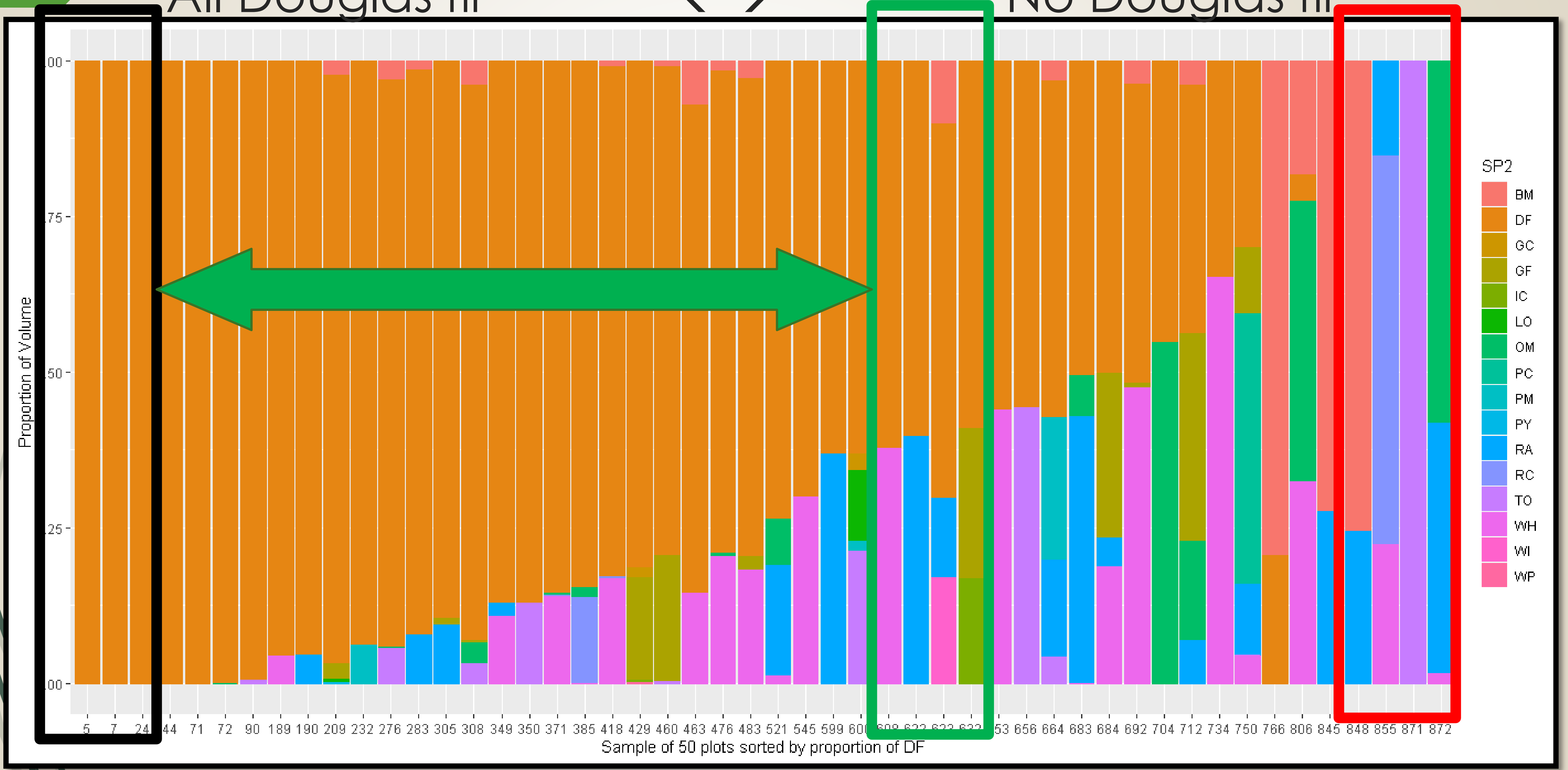
Species	Plots	% Plots
DF	704	78.31%
WH	56	6.23%
RA	29	3.23%
BM	28	3.11%
RC	27	3.00%
GF	16	1.78%
PC	11	1.22%
TO	10	1.11%
OM	10	1.11%
KP	2	0.22%
SS	2	0.22%
LO	1	0.11%
GC	1	0.11%
PP	1	0.11%
WO	1	0.11%

Direct reclassification

Species	Plots	% Plots
DF	704	78.31%
WH	56	6.23%
RA	29	3.23%
BM	28	3.11%
RC	27	3.00%
GF	16	1.78%
PC	11	1.22%
TO	10	1.11%
OM	10	1.11%
Other	8	0.89%

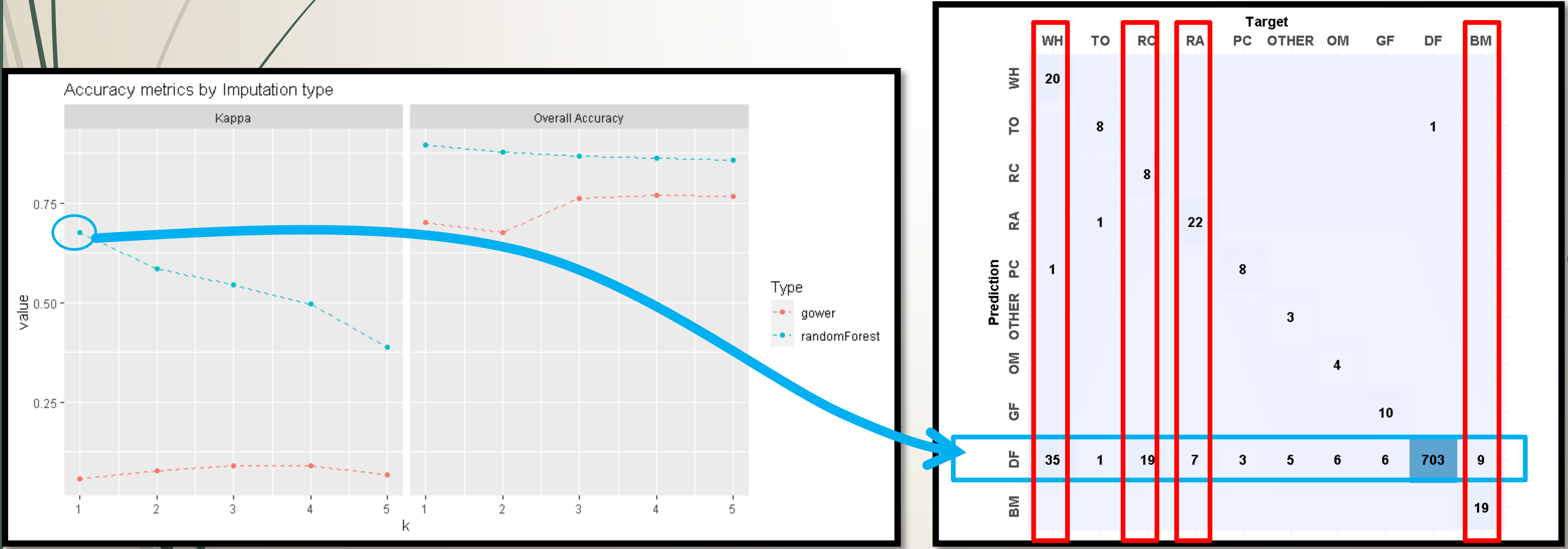
Ground data

All Douglas fir ←→ No Douglas fir



K-NN Imputation models:

- 1 to 5 Neighbors
- Gower distance and random forest distance
- Random forest distance with k = 1 best model
- Overall accuracies above 80% **but DF proportion in the area is 78%**
- **Kappa indexes 0.68 in the best case. (Poor-moderate)**

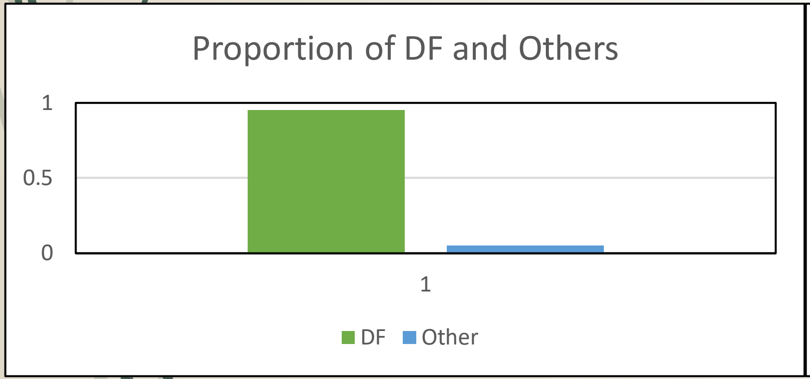


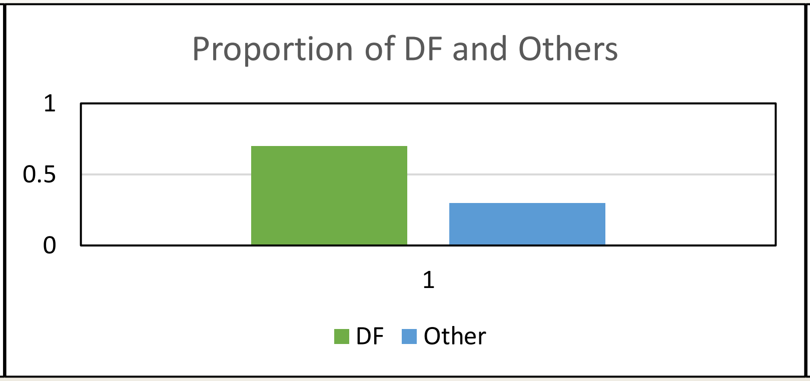
Douglas fir is so dominant that most species go to Douglass fir

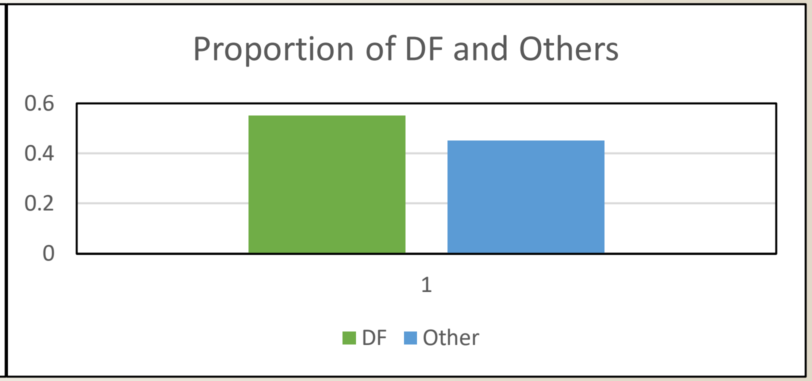
Definition of new categories:

- Defining categories based on the **dominant species very imbalanced dataset (~80% DF)**
- Also, with 1 dominant species→ Many unknowns









Redefining the problem

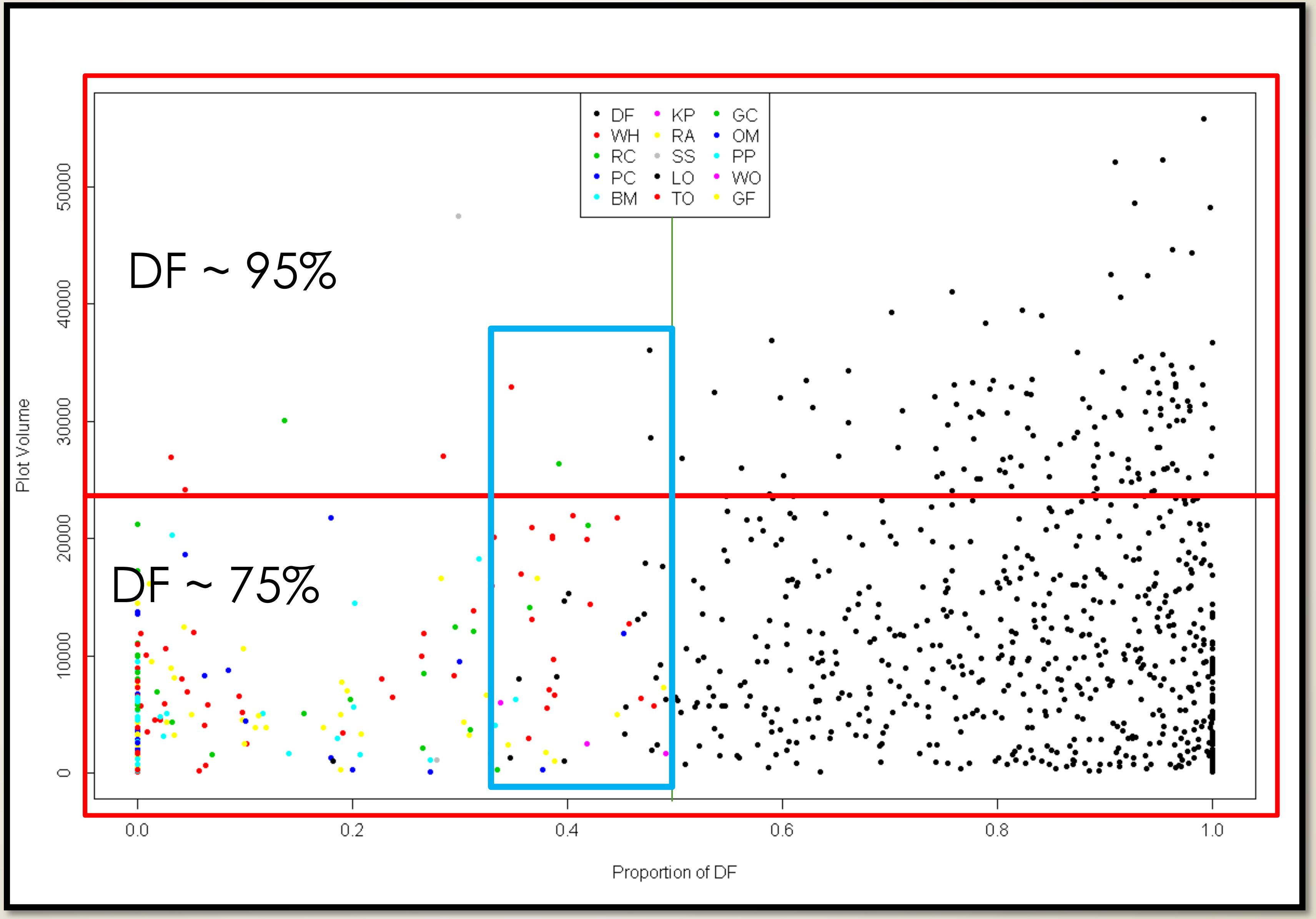
Definition of new categories:

We look at different ways of defining new categories for the classification problem.

1. Proportions based on different Volume thresholds
2. Clusters of species proportions
3. Combinations based on first 2 dominant species

Exploratory analysis

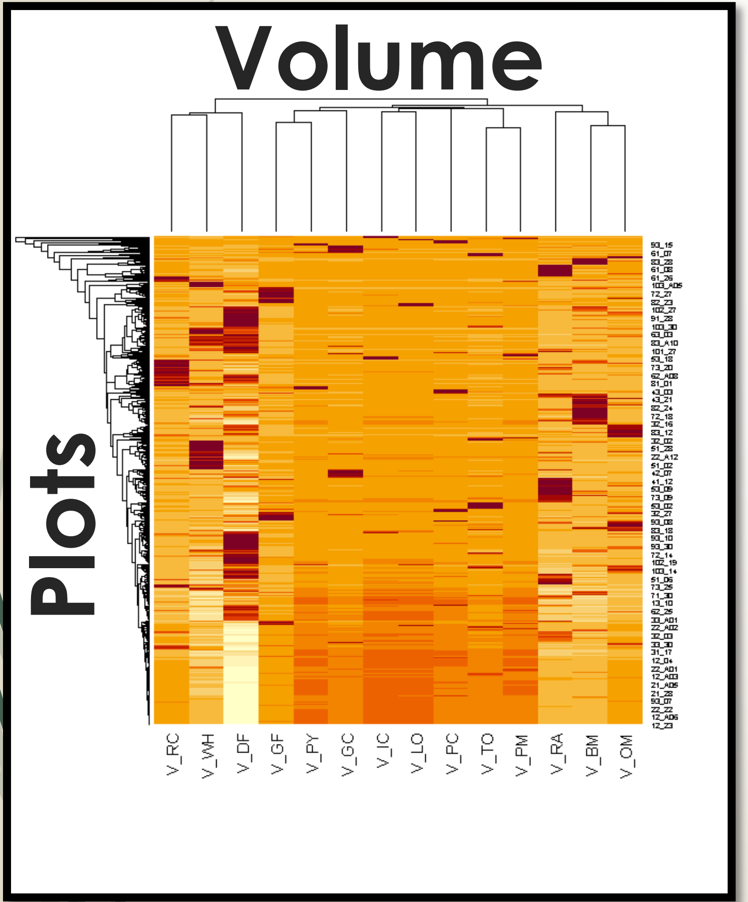
More mixing for plots with low or medium Volume

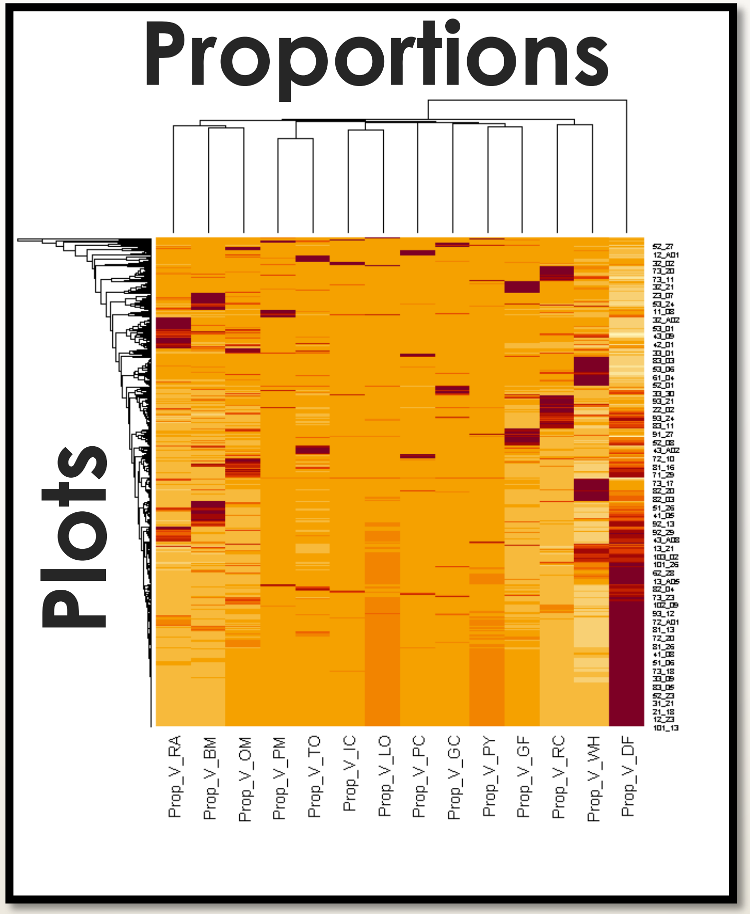


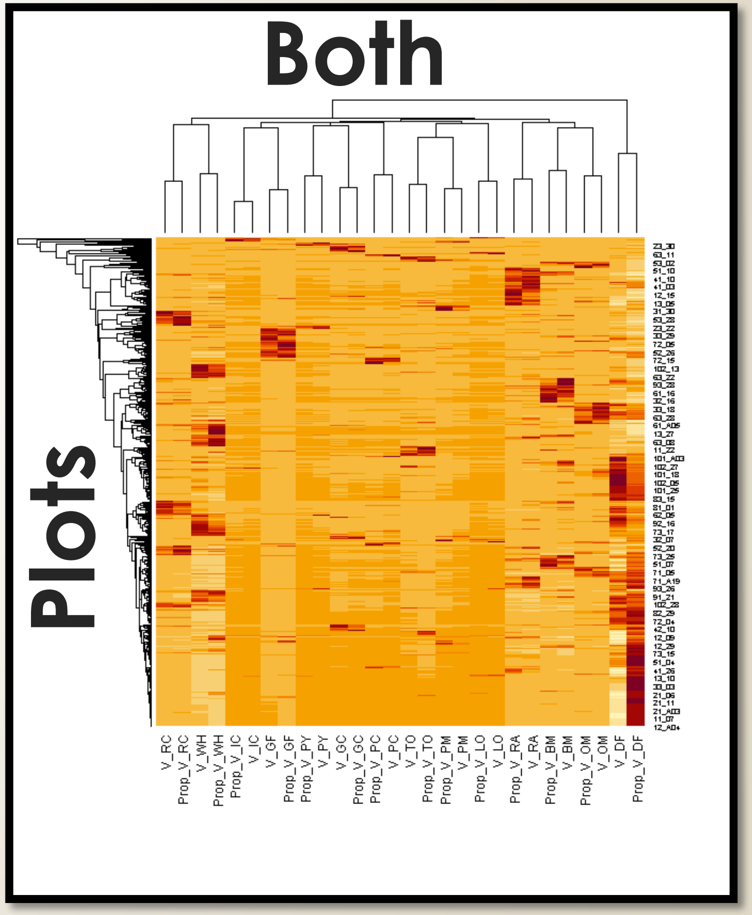
Exploratory analysis

Cluster analysis to identify “groups of associated species”

Based on Volume by species, proportion of Volume by species, both







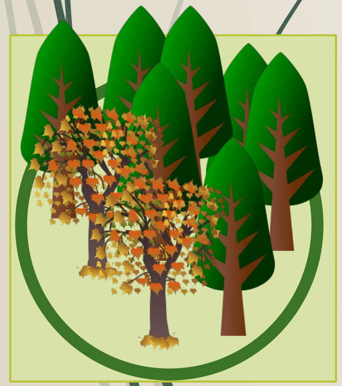
Categories based on dominant species
Dominance defined based on Volume proportion

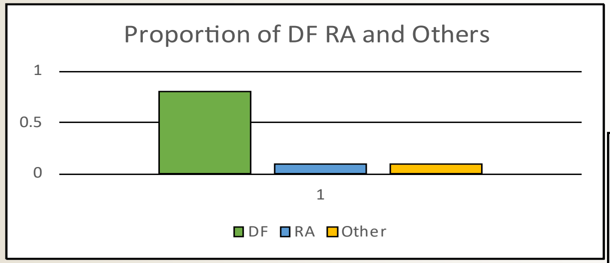
- One class with was pure Douglas fir
- One class with combinations involving minor species
- Douglas fir & Hemlock 200

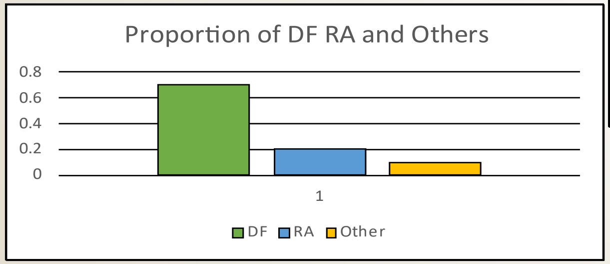
2 Species comb	# of plots	%plots
BM:DF	89	9.90%
BM:RA	11	1.22%
DF:GC	13	1.45%
DF:GF	54	6.01%
DF:OM	61	6.79%
DF:PC	21	2.34%
DF:PM	17	1.89%
DF:Pure	79	8.79%
DF:RA	88	9.79%
DF:RC	84	9.34%
DF:TO	51	5.67%
DF:WH	201	22.36%
Other	102	11.35%
RA:WH	13	1.45%
RC:WH	15	1.67%

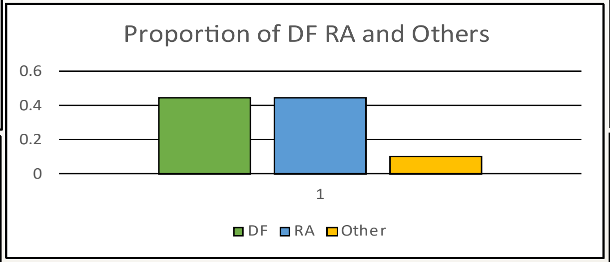
Definition of new categories:

- Defining categories as combinations based on the **first 2 dominant species** provided more balanced groups
- It makes the problem slightly different
- As with the dominant species. We don't know "HOW dominant" the species are

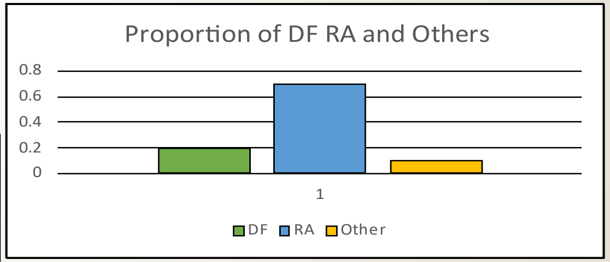


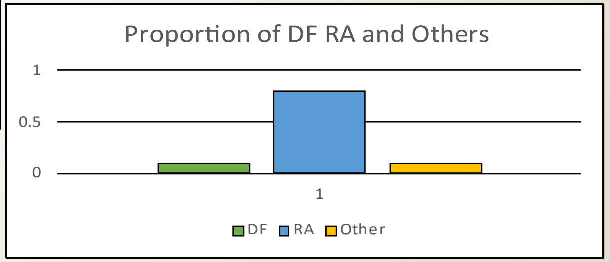






All plots are DF:RA





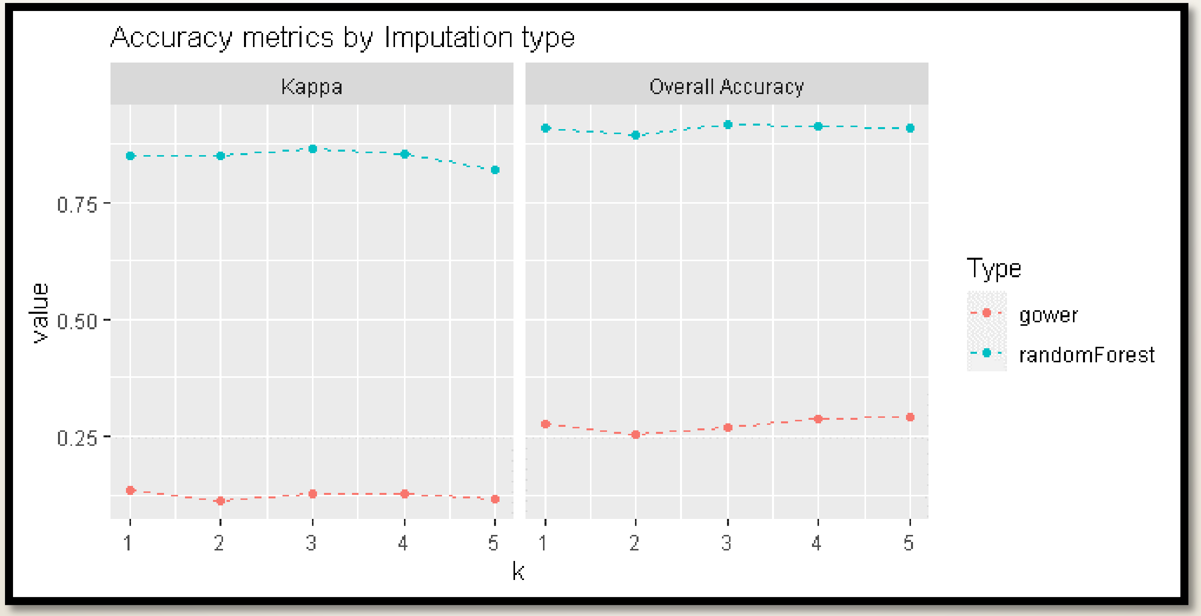
Random forest models for classification:

- Kappa = 0.13
- 4 categories completely missed
- Very few assignments conif+conif Hardw+Hardw

Prediction \ Target	RC:WH	RA:WH	Other	DF:WH	DF:TO	DF:RC	DF:RA	DF:Pure	DF:PM	DF:PC	DF:OM	DF:GF	DF:GC	BM:RA	BM:DF
RC:WH															
RA:WH															
Other	1	7	31	7	6	3	10	7	8	5	2	18	2	3	4
DF:WH	13	5	31	167	10	65	32	32	2	10	34	16	7	2	35
DF:TO		1	10	1	20		1	5	2	2	5	1	2		
DF:RC			2	5	1	4	3	1		1	1	3			1
DF:RA			7	2	3	1	19	5		1	2	2	1	2	3
DF:Pure			2	3	7	2	3	14	2	1	5	3	1		1
DF:PM								1							
DF:PC											1				
DF:OM			1	5	4	1		4		1	9	1		1	3
DF:GF	1		10	2		4	3	2	2		1	10			2
DF:GC															
BM:RA															
BM:DF			8	9		4	17	8	1		1			3	40

K-NN Imputation models:

- Overall accuracies above 80%
- **Ongoing process**
- **Kappa indexes 0.86 in the best case. (Improved the performance)**
- Part of the success →**More balanced classes**

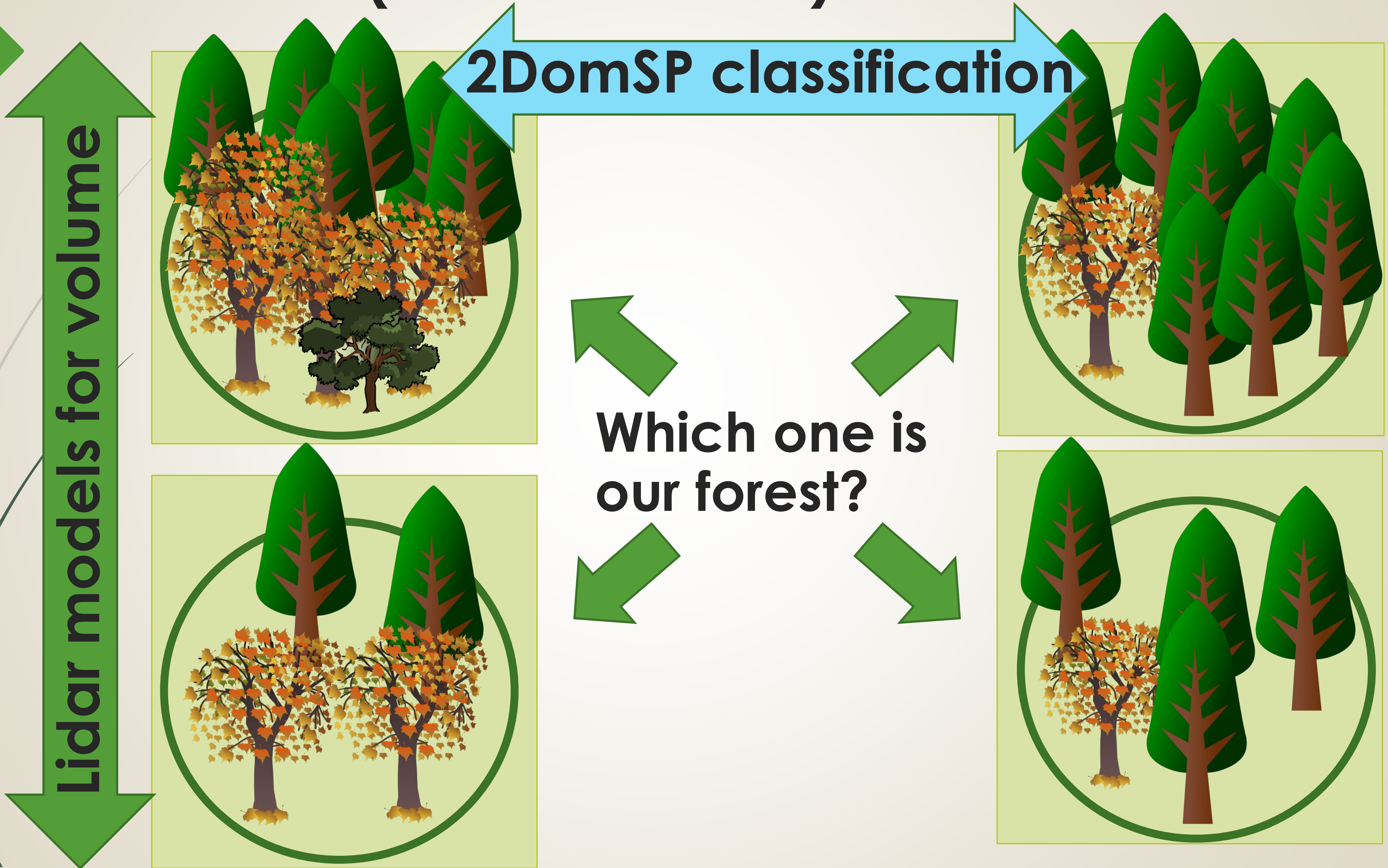


K-NN Imputation models:

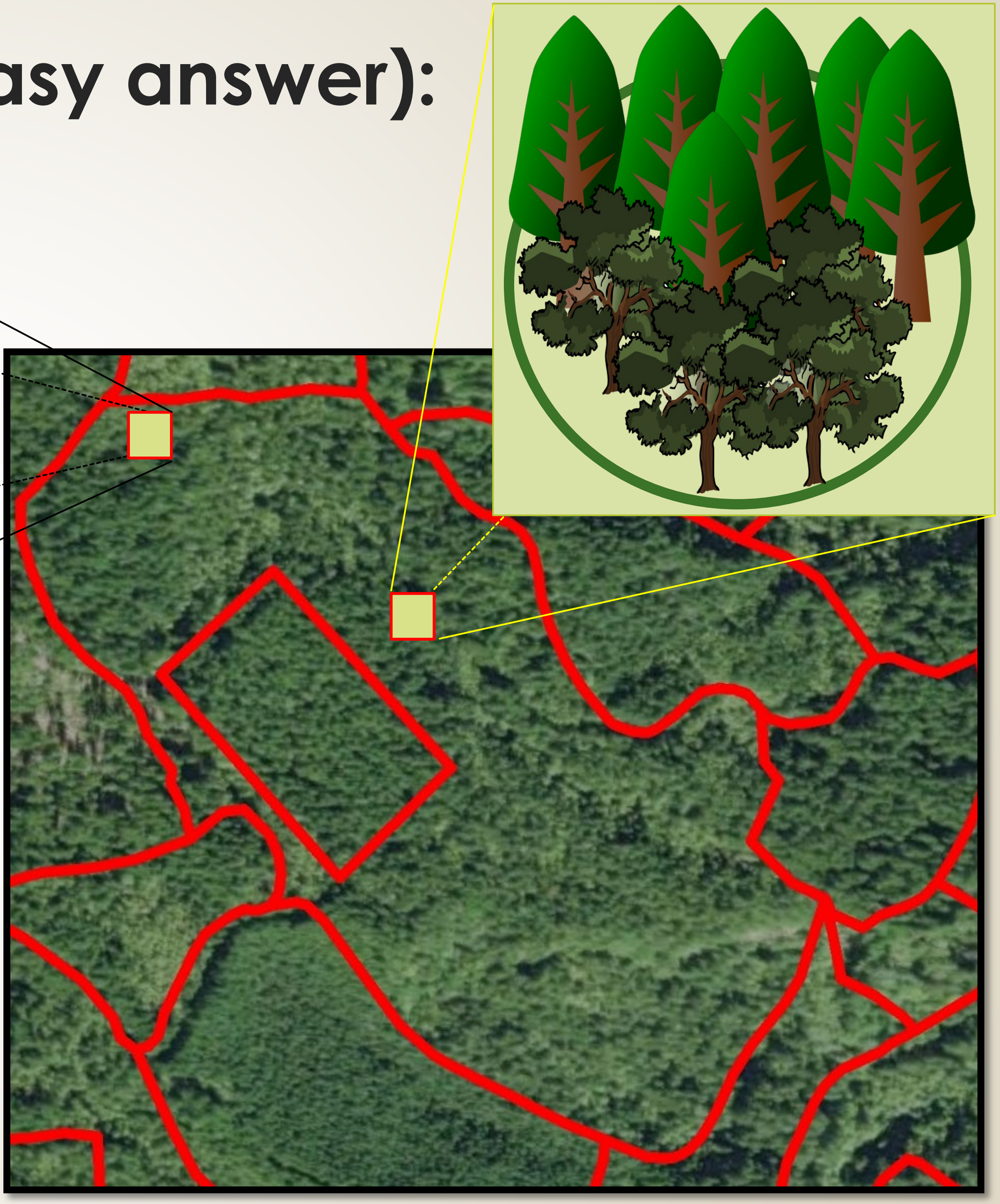
Target

Prediction	RC:WH	RA:WH	Other	DF:WH	DF:TO	DF:RC	DF:RA	DF:Pure	DF:PM	DF:PC	DF:OM	DF:GF	DF:GC	BM:RA	BM:DF
RC:WH	10														
RA:WH		4													
Other	1	3	90						2	1				2	1
DF:WH	3	6	5	198		3	2	2			4	1	1		5
DF:TO			1		51		1		2		2		1		1
DF:RC				1		79							1		1
DF:RA			1				85								
DF:Pure			2					76			1				
DF:PM									12						
DF:PC										20					
DF:OM				1				1			51	1		1	
DF:GF	1		1			1					1	51			
DF:GC													9		
BM:RA														8	
BM:DF			2	1		1			1		2	1	1		81

Question (with an answer):



Question (without an easy answer):



Aggregation to stand values. 1 Stand can be 200-500 pixels

More important→ How to validate the aggregation

Conclusions on the classification:

- Imputation worked better than Random Forest
- Making 2 species categories help breaking the imbalance of the dataset
- Improvements are necessary.
  - **Positive:** We do better than classifying at random. Improvements **god to moderate**.
  - **Negative:** Classifications provide little information about the species mixing.

2 Questions: How much detail we need and at what level. Stands, plots-pixels? How to validate at stand level.

Thank you for your attention