Species Prediction
Using Single Photon LiDAR
Results from the Algonquin Park Forest

PREPARED FOR:
KTTD

PREPARED BY:
Margaret Penner
Murray Woods

December 2022

Table of Contents

1 Introduction 1

2 Predicting percent conifer 3

2.1 Objective 3

2.2 Methods 3

2.3 Results 6

3 Predicting Species and forest type 7

3.1 Objective 7

3.2 Data 7

3.3 Results 8

4 Summary 13

5 Next Steps .... **Error! Bookmark not defined.**

6 Literature cited 14

1 Introduction

Recent projects in Ontario have used Single Photon LiDAR (SPL) to derive Forest Resource Inventories (FRIs). SPL, along with fixed area field plots, was used to estimate attributes using an area-based approach (ABA). The results were raster maps of quantitative attributes including heights, volumes and basal area. Polygon boundaries and the associated species composition and age were extracted from existing, historic inventories (T1 inventories) and combined with the SPL inventories (T2 inventories) to produce a current polygon inventory. The T1 species composition was from interpretation of aerial photographs that were at least 10 years old and dependent on the polygon delineation from that time. A new attribute in the T1 inventories was a VERT code – a classification of the vertical distribution of crowns in the forest canopy (Figure 1). If the polygon had two layers, a species composition and age were interpreted for each layer.

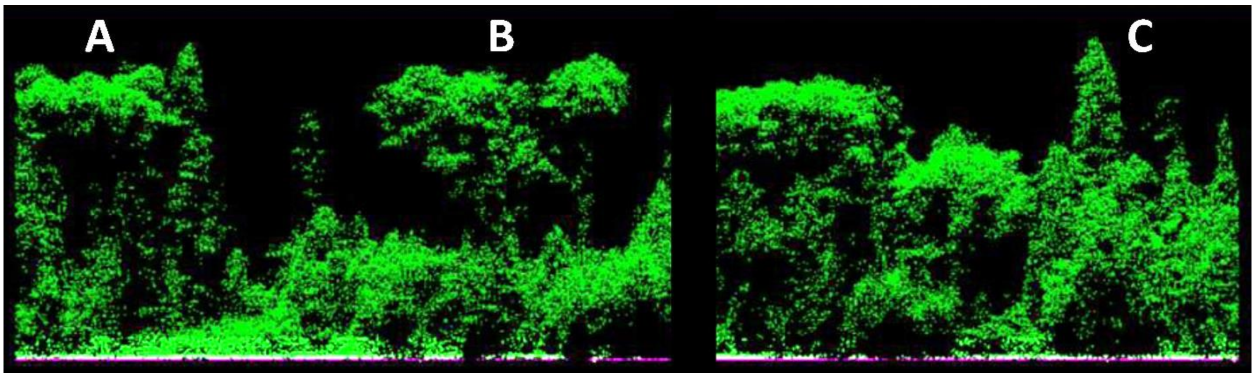

Figure 1. A profile view of a single photon LiDAR (SPL) point cloud is given. Vertical structure can be an important forest inventory attribute. It can range from a single canopy layer (A), two-story (B) to complex (C).

A previous project looked at the using SPL to predict vertical structure - whether a pixel was a single canopy later or two canopy layers. Predictions at the pixel level can then be aggregated to the polygon level to provide inventory estimates by layer.

For the Algonquin Park Forest (APF), approximately 40% of the forest was classified through Photo Interpretation (PI) as Two Tiered (TT) (either Single with Veterans (SV), Two-tiered with a dominant overstory (TO), Two-tiered with a dominant understory (TU), Two-tiered with a dominant overstory and veterans (MO) or two-tiered with a dominant understory and veterans (MU)) (Figure 2a). Approximately half of the TT polygons were Hardwood selection and another 25% were pine shelterwood. Of the over 10,000 hardwood selection polygons that were TO, 85 polygons had an understory species composition that included more than one species. This may reflect a relatively pure understory but also likely is a result of it being very difficult to see and/or time consuming to assess the understory.

In the Dog River – Matawin Forest (DRM), approximately 20% is classified as TT with the majority of that being classified as SV (Figure 2b). Very little area was assessed as TO or TU. This may be the result of clearcutting being the primary silvicultural system in the boreal. It may also have resulted from some confusion around the VERT photo interpretation definitions and assigning polygons with more than 3 species to VERT = CX by default.

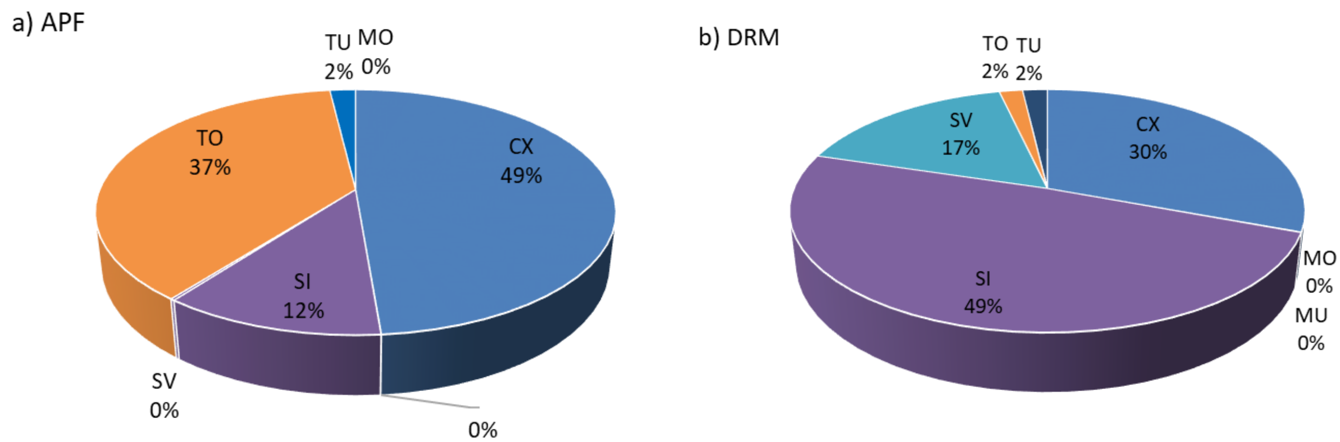

Figure 2. The proportion of productive forest area in the various vertical structure classes (VERT) is given for the Algonquin Park Forest (a – APF) and the Dog River – Matawin Forest (b – DRM). SI = single –storied, CX = complex, SV = single-storied with residual veterans, TO = two-tiered with a dominant overstory, TU = two-tiered with a dominant understory, MO = two-tiered with a dominant overstory and veterans, MU = two-tiered with a dominant understory and veterans.

The T2 species composition is taken from the T1 inventories. This may not be that useful for polygons that are TT in the T2 inventory.

- The polygon may not have been interpreted as a TT polygon in T1 and there is only one species composition (for all layers combined). This is particularly true in the boreal.
- It is difficult for photo interpreters to see the understory and the understory species composition is often a single species.
- The T1 inventory is at least 10 years old and the T1 species composition may not reflect the T2 conditions.

ABA approaches to predicting species composition from LiDAR have been investigated (van Ewijk et al 2014, Wilson et al. 2012, Donoghue et al. 2007). Less work has focused on the species composition of the understory. More work has been undertaken using Individual tree crown (ITC) approaches (e.g., Prieur et al 2022).

SPL may not be the best technology for this purpose. The laser beamlets have less energy than linear LiDAR and canopy penetration is less than conventional LiDAR (Figure 3), particularly for leaf-on conditions for hardwoods and dense canopy conditions for conifers.

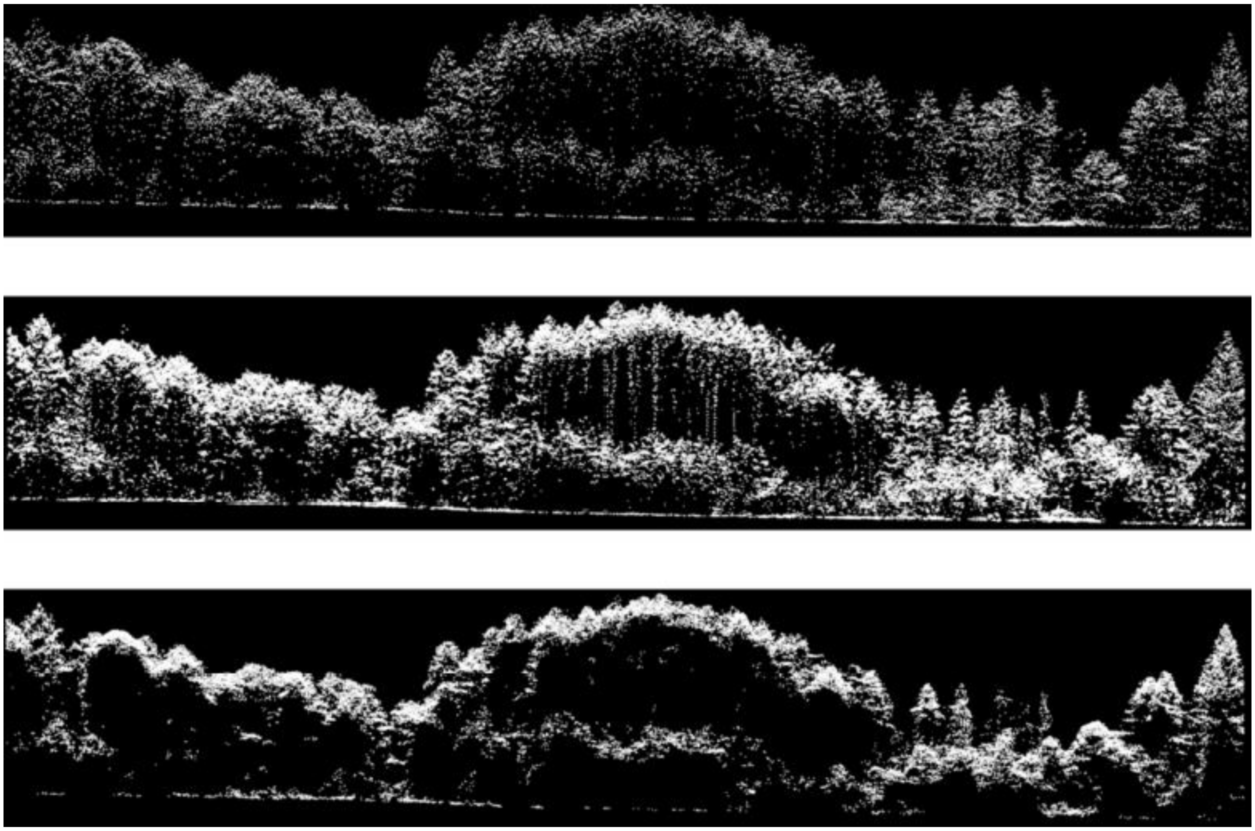

Figure 3. Transect extracted from the three lidar datasets: (**top**) monospectral ALS (ALS12); (**middle**) multispectral ALS (MSL16), with the three channels combined; and (**bottom**) photon-counting lidar (SPL18). Taken from Prieur et al. (2022).

2 Predicting percent conifer

2.1 Objective

The objective of this project was to investigate the use of Single Photon LiDAR (SPL) to predict % conifer by layer using an ABA.

2.2 Methods

Procedure

1. The vertical structure (VERT) of each ground plot was predicted using adimensional predictors (Table 1).
2. The CDHt was predicted using the normal LiDAR predictors and all plots.

The rest of the analysis uses those plots predicted to be two-storied (VERT = SV or TT)

3. The reference top of the lower layer (ref_top_LL) was estimated using kernel smoothing.
4. The reference relative top of the lower layer was calculated as rel_ref_top_LL = ref_top_LL/CDHt.

5. The relative top of the lower layer was predicted using adimensional predictors. The predicted top of the lower layer (pred_top_LL) was calculated as the predicted relative top of the lower layer and the predicted CDHt.
6. The reference % conifer by layer (ref_pct_con_UL and ref_pct_con_LL) were calculated from the field data (live trees with Dbh ≥ 7.1 cm) using basal area.
7. The % conifer by layer was predicted using adimensional SPL predictors. For the upper layer, the SPL predictors were calculated using the entire point cloud. For the lower layer, the predicted top of the lower layer was rounded up to the nearest 2m (e.g., pred_top_LL = 6.5m was rounded up to 8m). The usual SPL predictors were calculated from returns below the threshold.

There were some plots with no trees below the ref_top_LL or below the pred_top_LL and it was not possible to calculate the % conifer for the lower layer. For these plots, the % conifer of the lower layer was set to the % conifer of the upper layer.

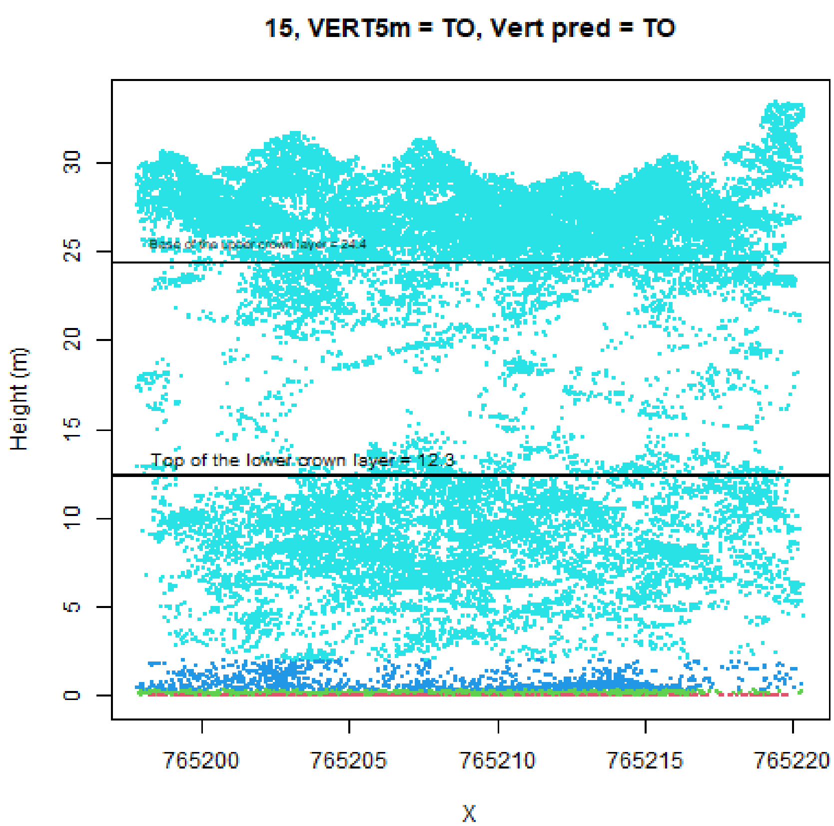

Figure 4. An example of a calibration plot predicted to be two-tiered. The % conifer for the upper layer was predicted using the entire LiDAR point cloud. The predicted height of the lower layer (pred_top_LL) was 15.55m. This was rounded up to 16m and the % conifer for the lower layer was predicted using the LiDAR point cloud below 16m.

Table 1. The LiDAR predictors are described.

Attribute	Description	Standard	adimensional
zmean	mean height of z	Y	
zentropy	entropy of height distribution (z)	Y	Y
pzabovezmean	percentage of returns above zmean	Y	Y
zq5	height of the 5th percentile of z	Y	
zq10	height of the 10th percentile of z	Y	
zq15	height of the 15th percentile of z	Y	
⋮			
zq90	height of the 90th percentile of z	Y	
zq95	height of the9 5th percentile of z	Y	
zq99	height of the 99th percentile of z	Y	
zpcum1	Height of the 1st decile	Y	Y
zpcum2	Height of the 2nd decile	Y	Y
⋮			
zpcum9	Height of the 9th decile	Y	Y
zsd95	standard deviation of z trimmed to 95%	Y	
zskew95	skewness of z trimmed to 95%	Y	
zkurt95	kurtosis of z trimmed to 95%	Y	
vdr	Vertical Distribution Ratio (max-median)/max	Y	
cv	coefficient of variation of z returns	Y	
vci_1m	vegetation complexity index - 1m bins	Y	Y
dns_2m	canopy cover % above 2m (number of all returns above 2m / number of all returns) * 100	Y	
dns_4m	canopy cover % above 4m	Y	
⋮			
dns_28m	canopy cover % above 28m	Y	
dns_30m	canopy cover % above 30m	Y	
vegden_0_2	Percent vegetation returns between 0 and 2m	Y	
vegden_2_4	Percent vegetation returns between 2 and 4m	Y	
⋮			
vegden_26_28	Percent vegetation returns between 26 and 28m	Y	
vegden_28_30	Percent vegetation returns between 28 and 30m	Y	
L1	L1 moment of vegetation points	Y	
L2	L2 moment of vegetation points	Y	
L3	L3 moment of vegetation points	Y	
L4	L4 moment of vegetation points	Y	
Lskew	L Skewness of vegetation points	Y	
Lkurt	L Kurtosis of vegetation points	Y	
Lcoefvar	L Coefficient of Variation of vegetation points	Y	
lpi	Lidar penetration index - count of returns between (-0.15 - 0.15)/all points (-.15 to 30m) * 100 [Uses Class 2,3,4,5]	Y	Y
ri_pts	rumple index based on Lidar points - 1m DSM	Y	Y
zsd95_norm	zsd95/zmean		Y
zskew95_norm	zskew95/zmean		Y
zkurt95_norm	zkurt95/zmean		Y
L2_norm	L2/L1		Y
L3_norm	L3/L1		Y
L4_norm	L4/L1		Y
Lskew_norm	Lskew/L1		Y
Lkurt_norm	Lkurt/L1		Y

2.3 Results

The results were disappointing. The % conifer in the upper layer was predicted using adimensional predictors. The random forest model explained 26% of the variance.

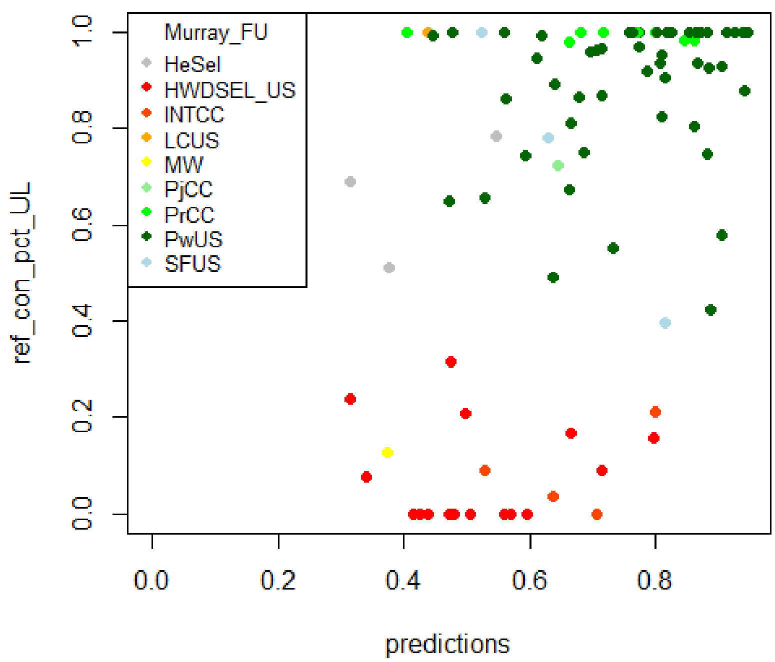

Figure 5. The predictions of the fraction conifer of the upper layer are given by forest unit.

The results for the lower layer were even worse. The % variance explained was negative indicating that many of the predictors were uncorrelated with % conifer.

Table 2. The % variance and the top predictor are given for predicting the conifer percent of the upper layer (con_pct_UL) and lower layer (con_pct_LL) using adimensional and standard predictors.

Dependent variable	Predictors	mean of squared residuals	% variance explained	Top three predictors (%incMSE)		
con_pct_UL	adimensional	0.10256	26.24	kurt95_norm	ri_pts	zpcum9
con_pct_UL	standard	0.08614	38.05	vegden_16_18	vegden_20_22	vegden_16_20
con_pct_LL	adimensional	0.14884	-0.43	zpcum2	zpcum3	kurt95_norm
con_pct_LL	standard	0.14997	-1.19	pzabovemean	zq55	zq80

3 Predicting Species and forest type

3.1 Objective

Rather than predict the conifer percent, the prediction of the dominant species of a pixel was investigated. The direct prediction of forest type (conifer, tolerant hardwood, intolerant hardwood) at the pixel level was also investigated.

3.2 Data

The species training dataset was created as follows.

- Pure species polygons (i.e. Mh = Hard Maple) or pure forest-type polygons (i.e., MhBe = Tolerant Hwd) were extracted from the APF T2 inventory (see Table 3). Hardwood cells with conifer species and Conifer cells with hardwood species were avoided. Note the species composition in the T2 inventory is largely derived from photo interpretation for the T1 inventory.
- Within the pure species polygons, pixels that appeared to be homogenous in terms of species composition were extracted.
- A range of crown closure was targeted during grid cell selection
- Each pixel was assigned a forest type and species string.
- Multiple pixels may be in the same polygon

Type = Conifer, Tolerant Hardwood, or Intolerant Hardwood.

Spec = The FRI PI stand species. If it is a single species like "Mh" that means the stand was interpreted as Mh100. When there are 2 or more species listed, those species, and possibly more, were in the stand.

Table 3. The sample sizes are given by species and forest type.

Species	Type	N
Bf	Conifer	19
Cw	Conifer	52
He	Conifer	2
Pj	Conifer	80
Pr	Conifer	132
PrPlant	Conifer	129
PrPlant_Thin	Conifer	321
Pw	Conifer	355
Sw	Conifer	19
Subtotal		1109
Mh	Tolerant Hardwood	302
MhBe	Tolerant Hardwood	187
MhBeBy	Tolerant Hardwood	139
MhByBe	Tolerant Hardwood	100
MhByMr	Tolerant Hardwood	26
MhMrBe	Tolerant Hardwood	90
MhMrBy	Tolerant Hardwood	200
MhMrByBe	Tolerant Hardwood	54
Or	Tolerant Hardwood	144
Subtotal		1242

Bw	Intolerant Hardwood	170
Po	Intolerant Hardwood	281
PoBw	Intolerant Hardwood	99
Subtotal		550

The SPL intensity attributes are given in Table 4. According to Prieur et al. (2022), no intensity normalisation is required and it can be considered an adimensional predictor.

Table 4. The intensity predictors are given.

Attribute	Description
Imean	Mean of intensity *new*
Isd	Standard deviation of intensity *new*
Iskew	Skewness of intensity *new*
ikurtosis	Kurtosis of intensity *new*

However, banding in the intensity attributes (an example is given in (Figure 6) was clear and the intensity attributes were not considered further.

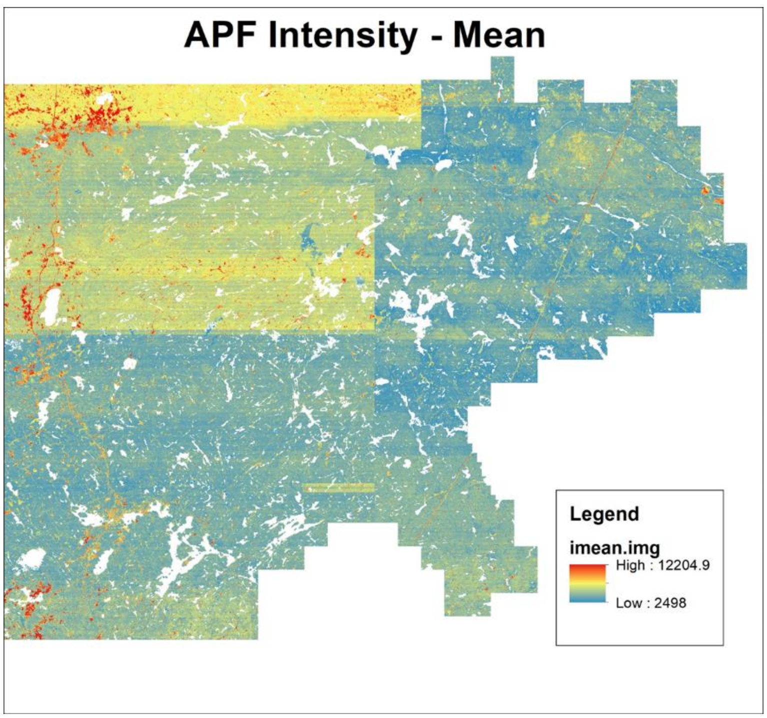

Figure 6. The mean intensity is given. There is distinct banding that is not the result of forest attributes.

3.3 Results

The top two adimensional predictors of forest type are ri_pts and lpi (Figure 7).

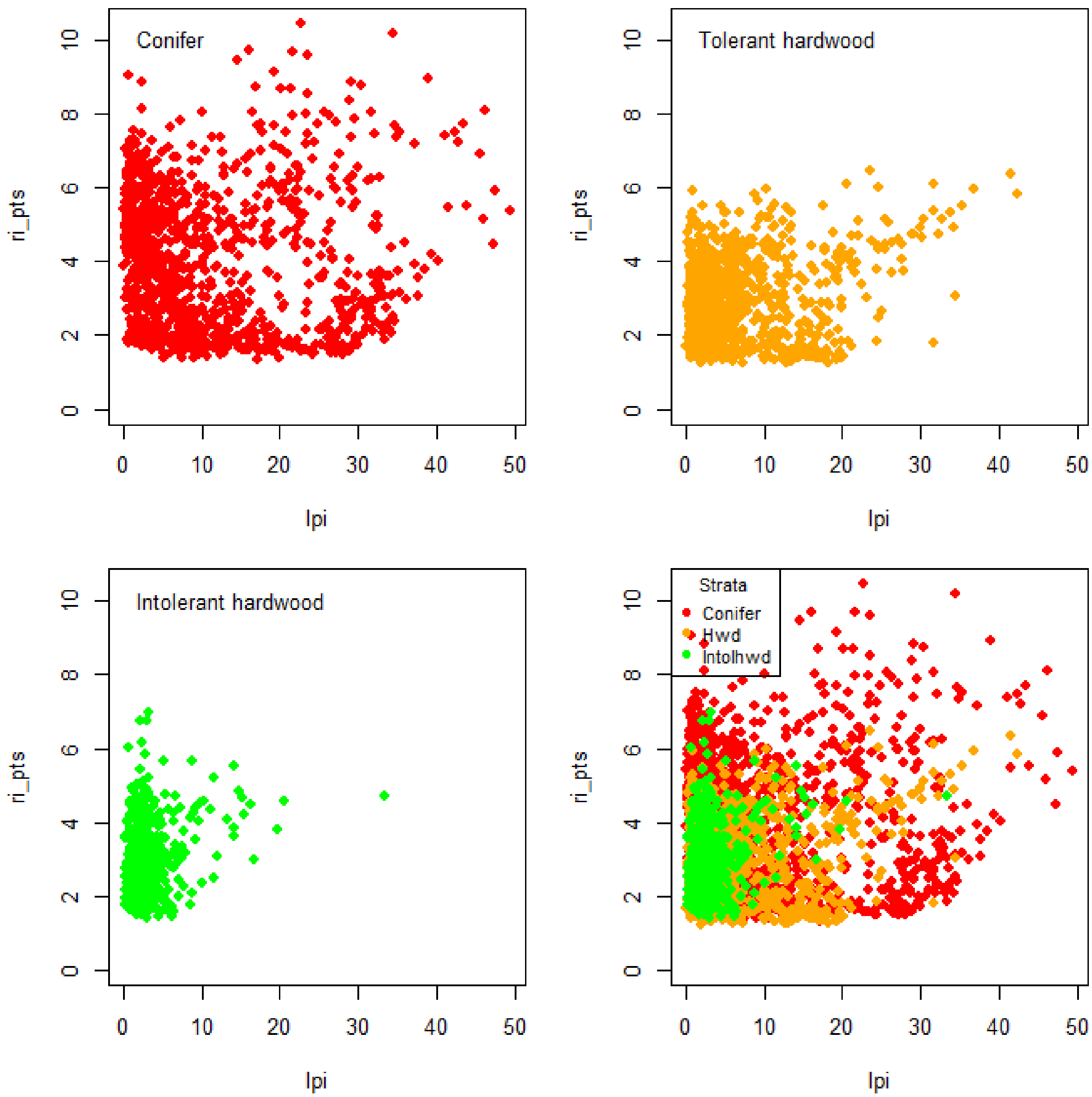

Figure 7. The data are plotted by rumple index (ri_pts) and the LiDAR penetration Index (lpi). Graphs are given by forest type and then all forest types combined. In predicting forest type, ri_pts and lpi were top predictors (excluding ri_pts and lpi resulted in the biggest mean decreases in accuracy).

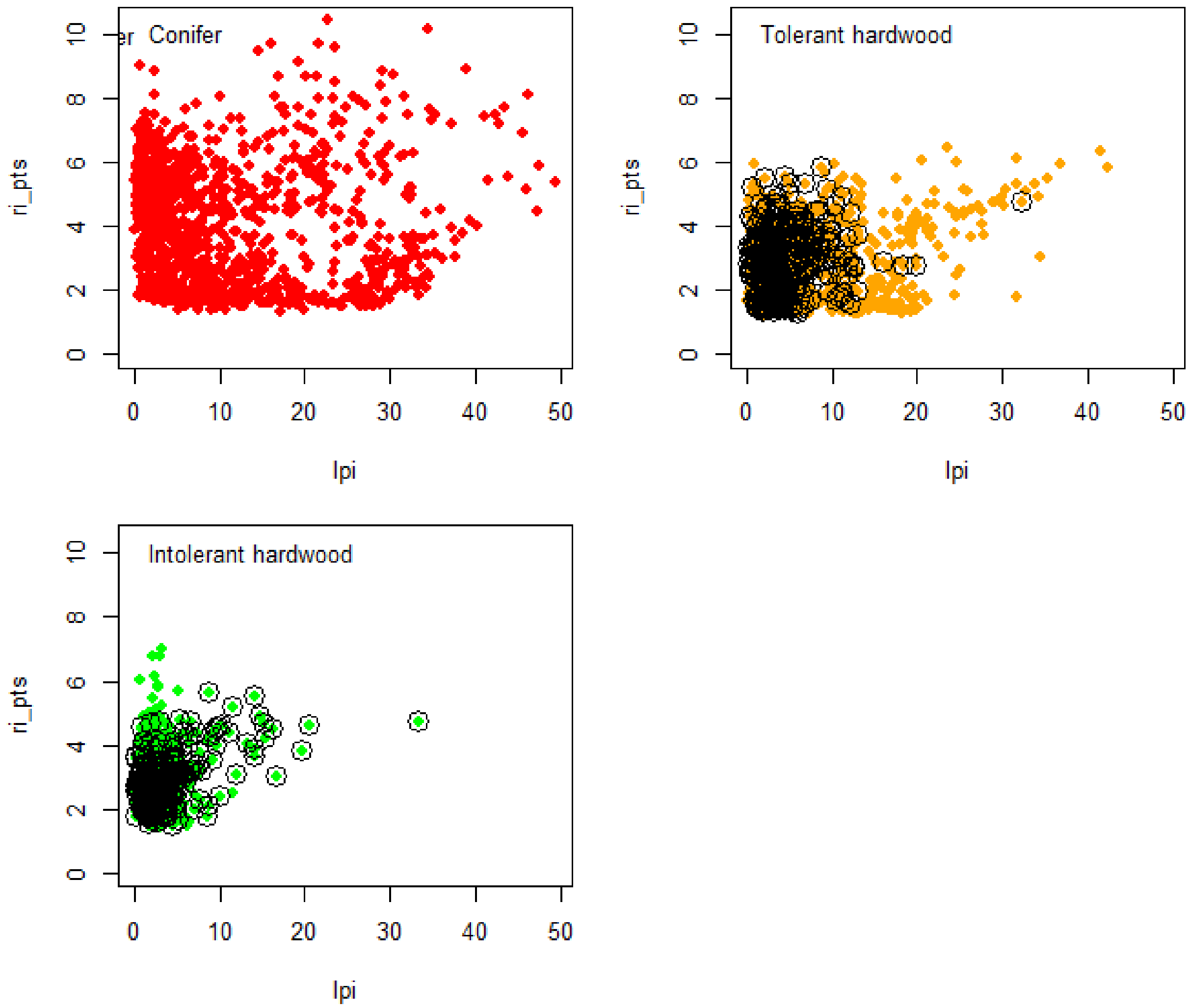

Figure 8. The same as Figure 7 except the single species samples for Tolerant hardwood and Intolerant Hardwood are circled. All the conifer samples are single species.

These were predicted using adimensional predictors. The intolerant hardwoods had the poorest prediction agreement.

Table 5. The confusion matrix for predicting forest type is given.

	Predicted			
Actual	Conifer	Tol Hwd	Intol Hwd	Agreement
Conifer	1035	54	20	93%
Tol Hwd	70	1095	77	88%
Intol Hwd	46	200	304	55%
Agreement	90%	81%	76%	84%

If tolerant hardwood and intolerant hardwoods are combined (post-prediction), the agreement increases.

Table 6. The intolerant and tolerant hardwood classes in the previous table are combined.

	Predicted		
Actual	Conifer	Hardwood	Agreement
Conifer	1033	74	93%
Hardwood	116	1676	94%
Agreement	90%	96%	93%

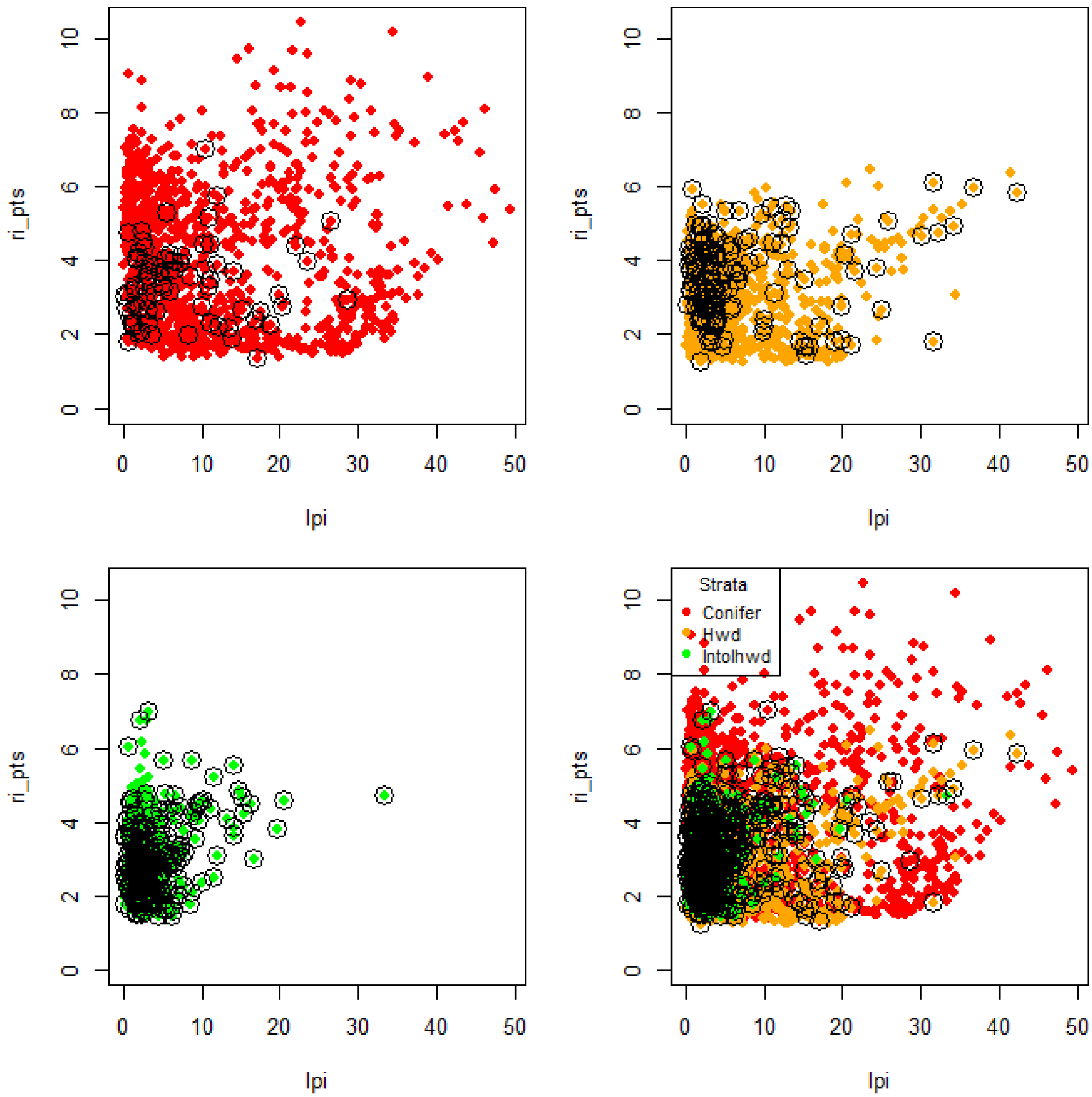

Figure 9. Same as the previous figure except the misclassified points are circled.

The forest type with the highest rate of misclassification is the intolerant hardwood type with white birch having the highest rate of misclassification.

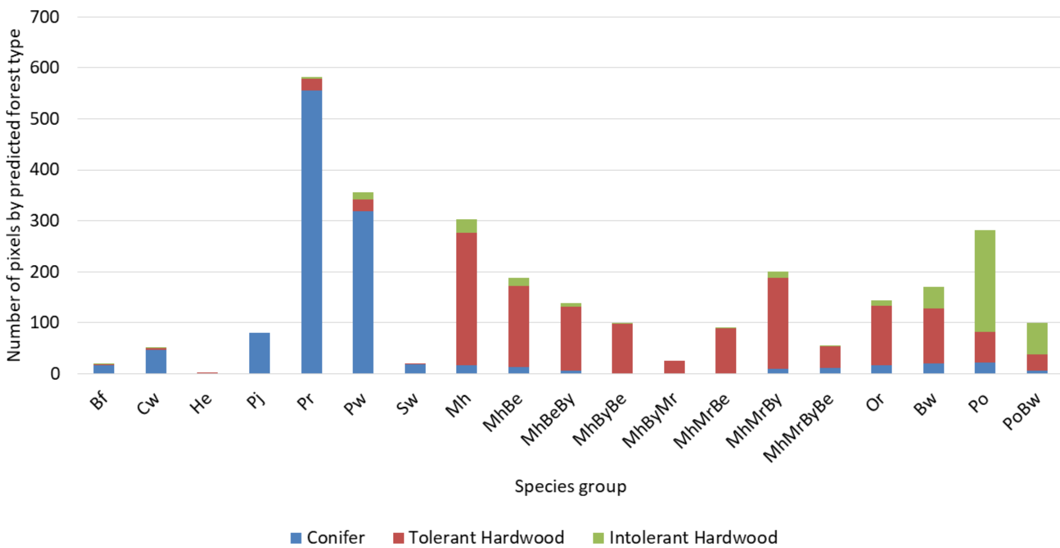

Figure 10. The calibration pixels are given by actual species group and predicted forest type.

There appears to be some predictive ability at the species group level as well.

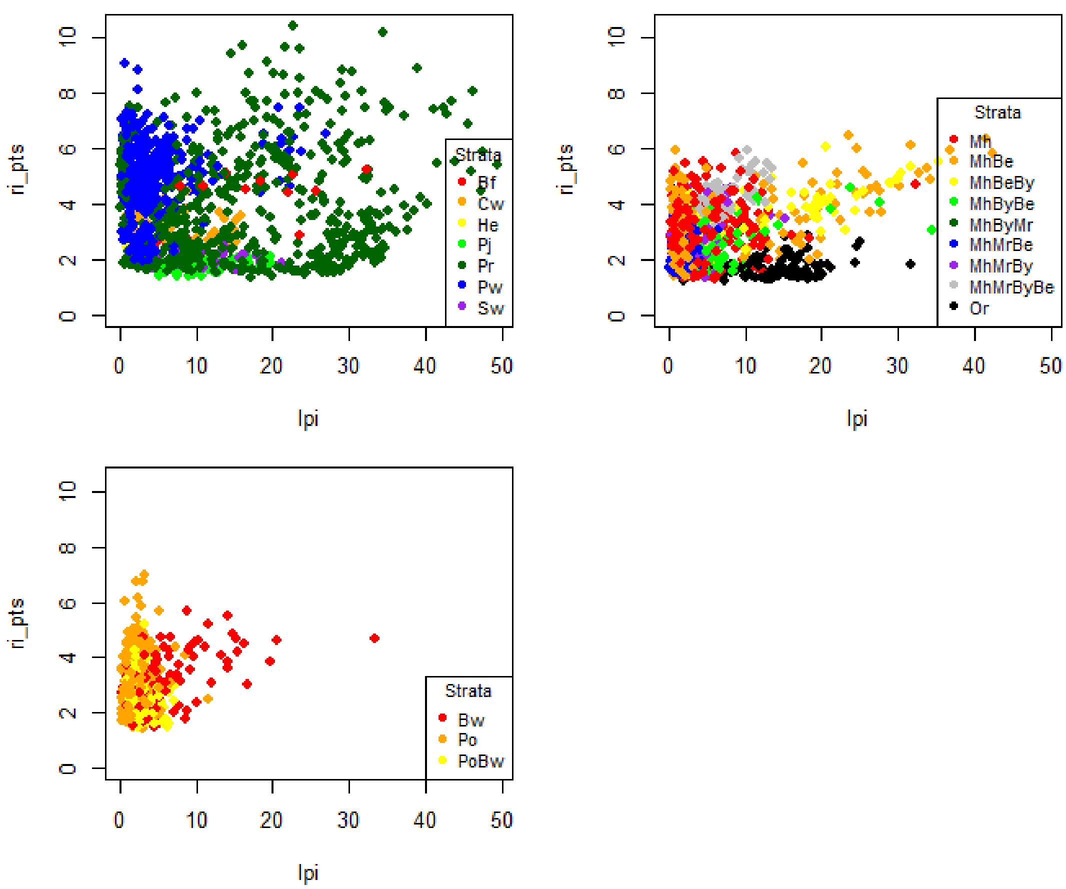

Figure 11. The Conifer (top left), tolerant hardwood (top right) and intolerant hardwood (bottom left) observations are plotted by species group.

The results are not as good when applied to the original LiDAR calibration plots (71% compared to 84% agreement with forest type). The original LiDAR calibration plots generally represent more mixed conditions in terms of species composition. In addition, their species composition comes from field measurement rather than photo interpretation.

Table 7. The confusion matrix for predicting forest type is given for the LiDAR plots.

	Predicted			
Actual	Conifer	Tol Hwd	Intol Hwd	Agreement
Conifer	84	31	15	65%
Tol Hwd	5	71	6	87%
Intol Hwd	2	6	3	27%
Agreement	92%	66%	13%	71%

If tolerant hardwood and intolerant hardwoods are combined (post-prediction), the agreement increases.

Table 8. The intolerant and tolerant hardwood classes in the previous table are combined.

	Predicted		
Actual	Conifer	Hardwood	Agreement
Conifer	84	46	65%
Hardwood	7	86	92%
Agreement	92%	65%	76%

4 Summary

The results of predicting % conifer by layer at the pixel level using LiDAR attributes were not promising.

The results of predicting forest type (conifer, tolerant hardwood or intolerant hardwood) using LiDAR attributes at the pixel level are promising. The quality of the training data (based on production photo interpretation of an older (T1) inventory) limit any further investigation.

5 Recommendations

The results of predicting forest type (and species) of the overstory at the pixel level are promising. The next step is to improve the quality of the training data. The training data in this project used the species composition from the T1 inventory which was a production inventory and based on 2007-2010 Photography. Also, selection of candidate training grid cells was performed in 2D by an untrained photo interpreter. A better training set would be to:

- Have experience interpreters use the 2019 imagery and identify relatively pure conditions at a grid cell scale. Ideally, the training dataset would have approximately equal sample sizes by species or species group. In addition, for pine and tolerant hardwoods it would be good to have representation of management (recent vs cutting in the past) also equally represented. Consideration should also be given to sampling across gradients of canopy cover.

The 223 SPL inventory calibration plots can be used as an independent validation dataset.

6 Literature cited

Donoghue, D., Watt, P., Cox, N., Wilson, J. 2007. Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. Rem. Sensing of Environ. 110:509-522.

Prieur, J.-F.; St-Onge, B.; Fournier, R.A.; Woods, M.E.; Rana, P.; Kneeshaw, D. A Comparison of Three Airborne Laser Scanner Types for Species Identification of Individual Trees. Sensors **2022**, 22, 35. ttps://doi.org/10.3390/s22010035.

Van Ewijk, K., Randin, C., Treitz, P., Scott, N. 2014. Predicting fine-scale species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. Rem. Sensing of Environ. 150: 120-131.

Wilson, T., Lister, A., Riemann, R. 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. For. Ecol and Mgmt. 271: 182-198.