Forest Carbon and Climate Program
Department of Forestry
MICHIGAN STATE UNIVERSITY

Report: Predictors of Family Forest Harvest Behavior in US Forest Service Ecoregion 212

Kylie Clay (MSU FCCP)
Chad Papa (MSU FCCP)
Hunter Stanke (MSU FCCP)

Predictors of Family Forest Harvest Likelihood and Intensity in USFS Ecoregion 212

Kylie Clay[*], Chad Papa[*], Hunter Stanke[*]

[*] Michigan State University, Forest Carbon and Climate Program
Contact: kclay@msu.edu

Michigan State University, Forest Carbon and Climate Program (FCCP) has conducted plot-level statistical analysis with the following objectives:

1) identifying covariates that best predict harvest likelihood (HL) and harvest intensity (HI) on non-industrial private forestland in northern Michigan, Minnesota, and Wisconsin (Ecoregion 212) for the three forest type groups (FTGs) of interest (Aspen/ Birch [AB]; Maple/ Beech/Birch [MBB]; White/Red/Jack Pine [Pine]);
2) identifying appropriate subregions for donor pool selection; and
3) identifying tiers of plot-level carbon potential according to key indicators (to inform potential caps or funding tiers).

Table of Contents

Figures

Tables

Statistical Methods

To conduct the plot-level analysis key predictors of HL and HI, we developed a random forest (RF) model. In this section, we detail briefly what a random forest model is and why this approach was selected.

Machine learning is a widely used technique to automate both supervised and unsupervised classifications in order to identify patterns within datasets. Specifically, RF models, a type of machine learning algorithm and an extension of classification and regression trees (CART) techniques, are a suite of non-parametric models that utilize decision trees to classify datasets. RF models split observations in a pairwise hierarchical manner based on an algorithm-generated basic rule that minimizes within-group variation and maximizes between-group variation (Breiman, 2001). This enables rapid classification and estimation of importance for dependent variables (Ziegler and Konig, 2014). RF models have grown in popularity due to ease of parameter tuning (i.e., an analyst needs only to determine input variables, number of trees to generate, and the number of variables to sample at each decision step) and model insensitivity to variable magnitudes and distribution (i.e., models do not require data rescaling) (Wager et al., 2014).

RF offers advantages over other parametric approaches (such as generalized linear models or logistic regression models), including handling residual noise for predictions and probability estimates for multi-category dependent variables (Ziegler and Konig, 2014). RF models can be prone to overfitting, since models inherently reduce variance and mean square error through complex model building processes that can generate many trees. However, bootstrapping samplers and bootstrap aggregation inherent to RF model techniques minimize overfitting; additionally, straight-forward checks of model results can limit bias and increase validity (Ziegler and Konig (2014). RF model estimates characterize error, strength, and correlation and can also be used to measure variable importance (Breiman, 2001), including for high-dimensional problems involving many features (Ziegler and Konig, 2014).

Data Description

Here, we provide a description of the input used in the statistical analyses.

We derive all input data (i.e., predictors and response) from:

1. The US Department of Agriculture, Forest Inventory and Analysis: https://www.fia.fs.usda.gov/ (using the rFIA R package: https://rfia.netlify.app/)
2. US Census: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html
3. Mill location data [provided by state DNRs]:
    - Michigan:https://midnr.maps.arcgis.com/apps/webappviewer/index.html?id=a75cedbbefc547dca4e1c340d65ee3cb)

- Minnesota: https://webapps15.dnr.state.mn.us/timber_producers/companies
- Wisconsin: https://dnr.wisconsin.gov/topic/forestbusinesses/industries

We consider only plots encompassing privately-owned forestland in our analyses (including tribal lands). See `src/summarizeVariables.R` for the procedures used to summarize condition-level FIA data.

We exclude any plot not meeting the following conditions:

- Plot falls exclusively on private forestland
- Single condition is present, and its attributes are constant through time (e.g., has always been recorded as a red pine plantation)
- Trees present at least one plot visit (e.g., post-clearcut is considered non-treed forestland)
- Annual-to-Annual plot, i.e., same plot design used at all visits

Input data are stored in `results/plot_vars.csv`. Variable definitions are as follows:

Dependent Variables (Harvest Indicators)

Harvest intensity:

- `REMV_NETVOL_ACRE`: (numeric) average annual net merchantable volume (cu.ft.) per acre harvested during the remeasurement interval. We compute HI for all remeasured plots (most have been remeasured multiple times) in terms of a percentage of net merchantable volume removed that can be attributed to tree harvesting across all plot visits (i.e., sum across remeasurements).

Harvest (binary):

- `HARVESTED`: (factor/ binary) binary code indicating if tree harvesting occurred on the plot between the remeasurement interval (`HARESTED=1` when harvesting occurred, and `HARVESTED=0` otherwise)

Independent Variables (Predictors/ Co-Variates)

- `FORTYPCD`: (factor) code for forest types
- `SITECLCD`: (factor) code for site productivity classes
- `STDORGCD`: (factor) binary code indicating clear evidence of artificial regeneration (i.e., plantation status)
- `PHYSCLCD`: (factor) code for physiographic classes
- `ECOSECCD`: (factor) code for ecoregion
- `STATECD`: (factor) code for state
- `RDDISTCD`: (factor) code for straight-line distance to nearest improved road
- `SLOPE`: (numeric) slope of condition (%)
- `ASPECT`: (numeric) aspect of condition (degrees)
- `ELEV`: (numeric) elevation of condition
- `PREV_BAA`: (numeric) live tree basal area per acre at initial measurement
- `PREV_QMD`: (numeric) live tree quadratic mean diameter at initial measurement

- `PREV_NETVOL_ACRE`: (numeric) net merchantable volume at initial measurement
- `GSSTK`: (numeric) initial stocking of growing stock (absolute value 0-167%)
- `prop.forest`: (numeric) proportion of landscape within 10km of fuzzed plot locations that is classified as forestland (derived from [National Land Cover Database 2016: https://www.mrlc.gov/national-land-cover-database-nlcd-2016]).
- `dist.to.mill`: (numeric) distance to nearest mill, calculated using fuzzed and swapped plot coordinates and mill coordinates.
- `n.mills.50km`: (numeric) number of mills within a 50km radius
- `pop.current`: (numeric) 2019 county population [US Census data]
- `pop.growth`: (numeric) county population growth 2011-2019 [US Census data]
- `prop.small.private`: (numeric) proportion of forestland within 1km of fuzzed plot location that is classified as private (family/small owner) ownership (derived from [Sass et al, 2020: https://www.nrs.fs.fed.us/pubs/61623]).

Please see the FIA Database Documentation: (https://www.fia.fs.usda.gov/library/database-documentation/current/ver90/FIADB%20User%20Guide%20P2_9-0-1_final.pdf) for definitions associated with forest type, site productivity, stand origin, and physiographic class codes.

Covariate Importance Results

To assess variable importance associate with the HL and HI models, we calculated the loss in predictive accuracy associated with the removal of each variable. To calculate model predictive accuracy, we used a 5-fold cross validation technique to evaluate out-of-sample performance (that is, we systematically and sequentially removed a portion of the plots and tested the ability of the model to predict results on those plots). **Figure 1** and **Figure 2** visualize the results of both the HL and HI analyses, respectively.

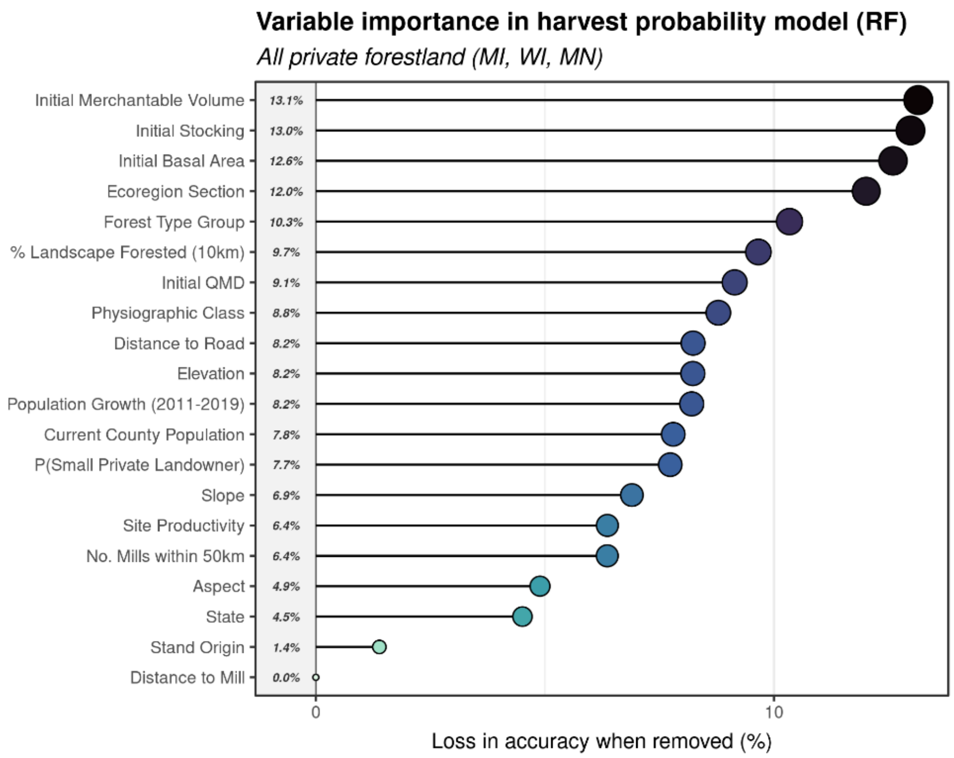


Figure 1. Variable importance in harvest probability model.

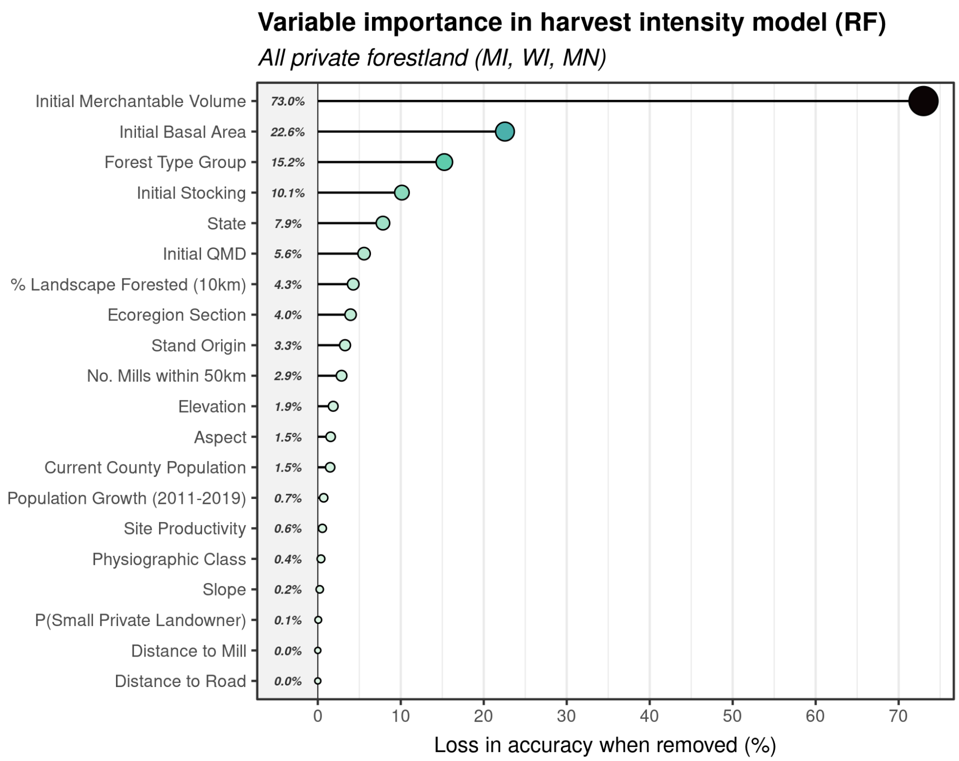


Figure 2. Variable importance in harvest intensity model (RF).

Geographic Variations in Predicted Harvest Likelihood and Harvest Intensity

Should the geographic variation prove useful for strategic landowner recruitment, we used the random forest model to predict HL and HI across all FIA donor plots (ecoregion 212). **Figure 3** and **Figure 4** visualize these results.

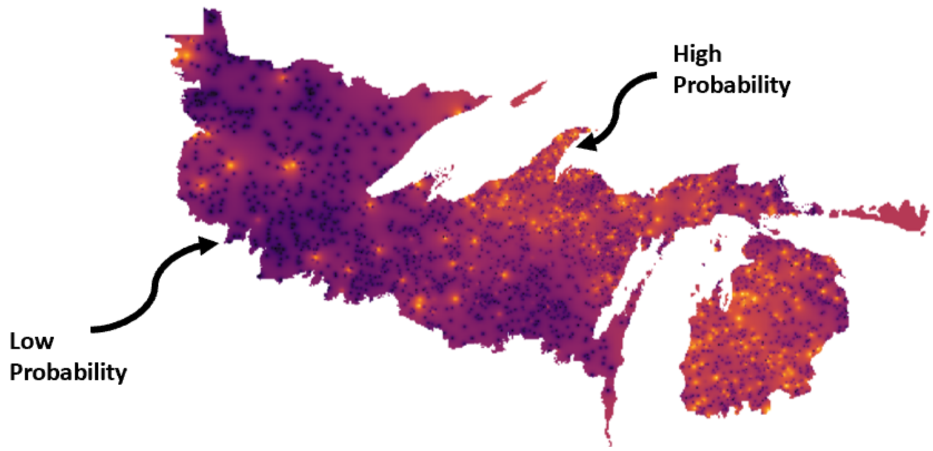


Figure 3. Predicted Harvest Likelihood.

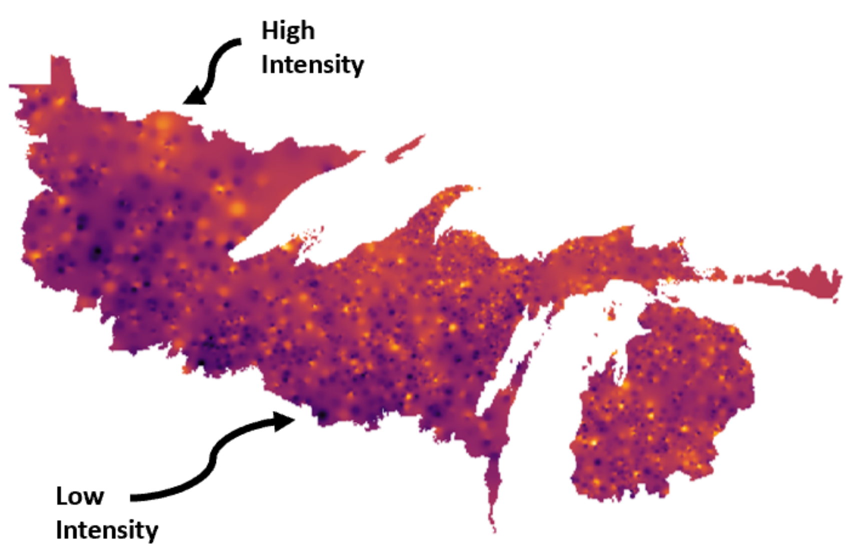


Figure 4. Predicted Harvest Intensity.

Subregions for Donor Pool Selection

To create more appropriate and refined donor pools, we binned discrete groups of USFS Ecological Sections (Cleland et al. 2007) based on trends in forest condition and harvest behavior.

Per Figure 5, we initially identified three distinct groups based on the prevalence and management behavior of the three FTGs of interest: an AB dominated group largely centered in Minnesota, a MBB dominated group largely centered in Wisconsin, and an AB group largely centered in Michigan. We did not consider grouping Minnesota and Michigan ecological sections despite their similar FTG makeup due to differing common management regimes (for aspen, in particular).

Ultimately, we opted to merge the Minnesota and Wisconsin groups as not doing so would have left too few plots of the minority FTG in each of the respective sub-regions. Specifically, both MBB in MN and AB in WI would have had insufficient donor plots from which to draw within their sub-region. Per the approved FFCP methodology, in the case of insufficient donor plots within the subregion, the donor pool moves to the larger ecoregion (in the case of the Northwoods, this was all of ecoregion 212); this was seen as less preferred given MI's distinct aspen growth and management practices. Figure 6 visualizes the two sub-region groupings that were determined most appropriate.

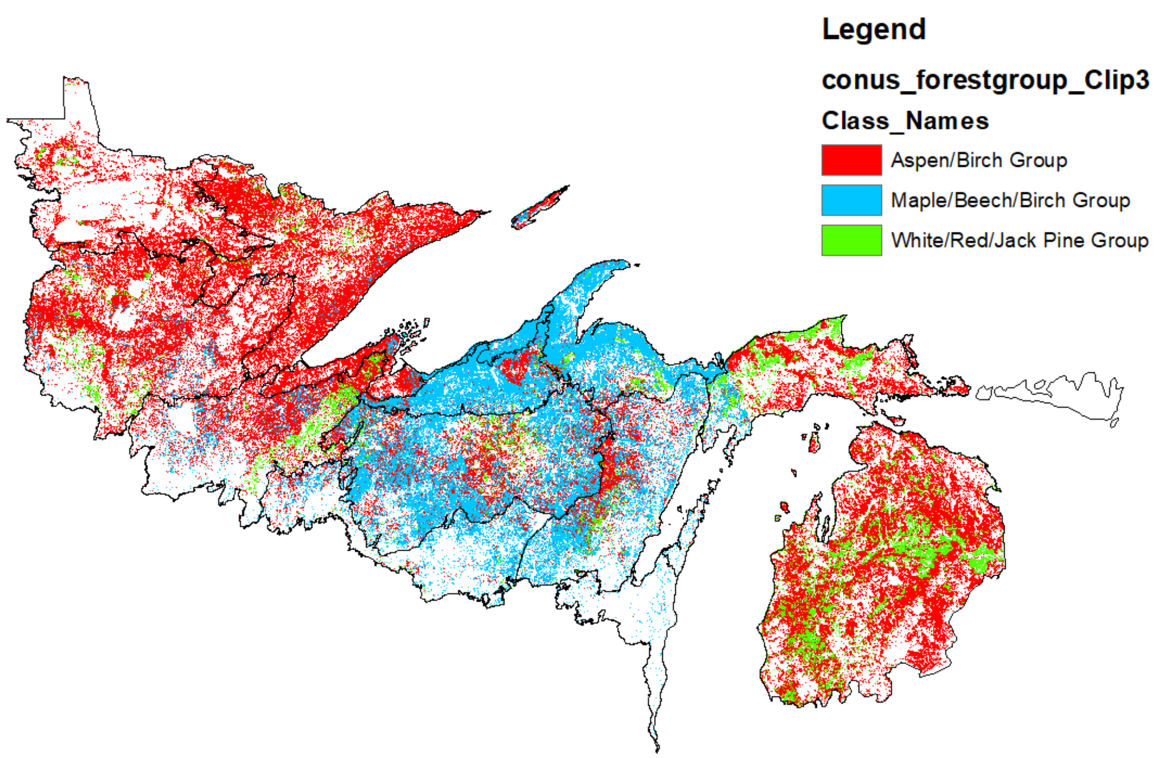


Figure 5. Three forest type groups of interest in USFS ecoregion 212.
Source: USFS (2005)—LANDSAT and MODIS

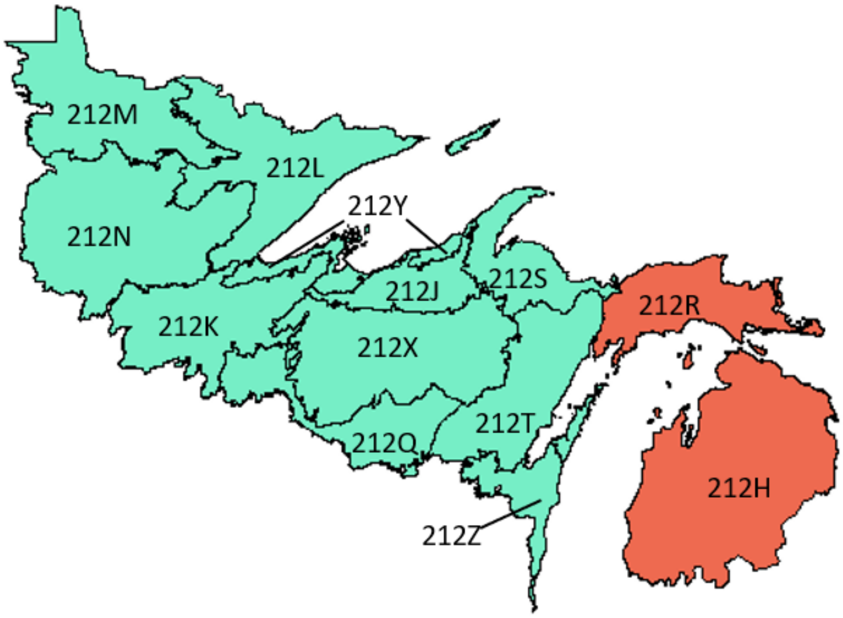


Figure 6. Selected ECOSUBCD Groupings

Tiers (to Inform Caps and/or Payment Structures)

With an aim of informing potential 1) landowner and donor pool caps (e.g., cutoffs to program participation) and 2) tiered payment structures, we systematically assessed how caps on each covariate and groups of covariate caps influenced HL and HI. The objective was to identify covariates whose caps might most efficiently increase predicted HL and HI (i.e., without unduly reducing the donor pool).

Our process was to first determine initial, or "100%", cap selections for each of the covariates (below or above which plots would be deemed ineligible), and then to experiment with the impact on HL and HI as those caps became systematically stricter and looser. We determined the initial cap selections for each covariate based on partial dependence plots (PDP), histograms, and knowledge about the FTG-specific harvesting practices (see Appendix I for all PDPs and histograms). The PDPs help visualize predicted HL at different levels of the covariate of interest, while the histograms help visualize the effect different caps have on the donor pool (n). Combing these sources of information, we identified data-driven, theoretically relevant caps for each of the covariates.

Once initial caps were selected, we experimented with variations in cap groupings to maximize both impact and efficiency (i.e., we assessed the impact on n, HI, and HL by applying different mixes of covariates' caps). Because many of the covariate of interest were highly correlated, there was no value added (in terms of n, HL, or HI) in including all variable caps. Through a series of systematic assessments, we identified the Quadratic Mean Diameter (QMD) and Merchantable Volume caps as driving impact on n, HL, and HI.

The tables and figures below demonstrate both what those caps represent (in each covariate's unit of analysis) and their impact on n, HL, and HI for each FTG in the two subregions. To facilitate table and figure comprehension: when caps are set at 0%, no cap has been applied to the donor pool; 100% caps represent the initially determined cap. As intended, all subregions and FTGs reveal the same general trajectory, albeit at different rates, of increasing HL and HI and decreasing n as caps increase in intensity. This suggests that landowners with higher QMD and Merchantable Volume levels stand to have more of a carbon impact on the landscape, though excluding landowners below identified levels would have a negative impact on the number of eligible donor and participant plots.

Aspen-Birch

Table 1. Quadratic Mean Diameter (QMD) and Merchantable Volume Caps Impacts on Donor Pool (n), Harvest Likelihood (HL), and Harvest Intensity (HI) for Aspen-Birch

Aspen-Birch							
		Caps and Results by Cap Intensity					
	Variable Cap Used	***0%***	***25%***	***50%***	***75%***	***100%***	***125%***
	QMD (inches)	0	0	0	0	0	0
	Merchantable Vol (cu.ft/ ac)	0	250	625	750	1000	1250
	Result Indicator						
Group 1 (MN; WI)	*n*	450	272	181	112	67	36
	Predicted Harvest Likelihood	0.024	0.034	0.039	0.052	0.066	0.084
	Predicted Harvest Intensity	54.781	75.908	99.765	134.039	147.735	194.009
Group 2 (MI)	*n*	198	135	92	58	24	17
	Predicted Harvest Likelihood	0.019	0.025	0.030	0.028	0.046	0.045
	Predicted Harvest Intensity	72.957	81.519	98.342	113.396	189.205	221.554

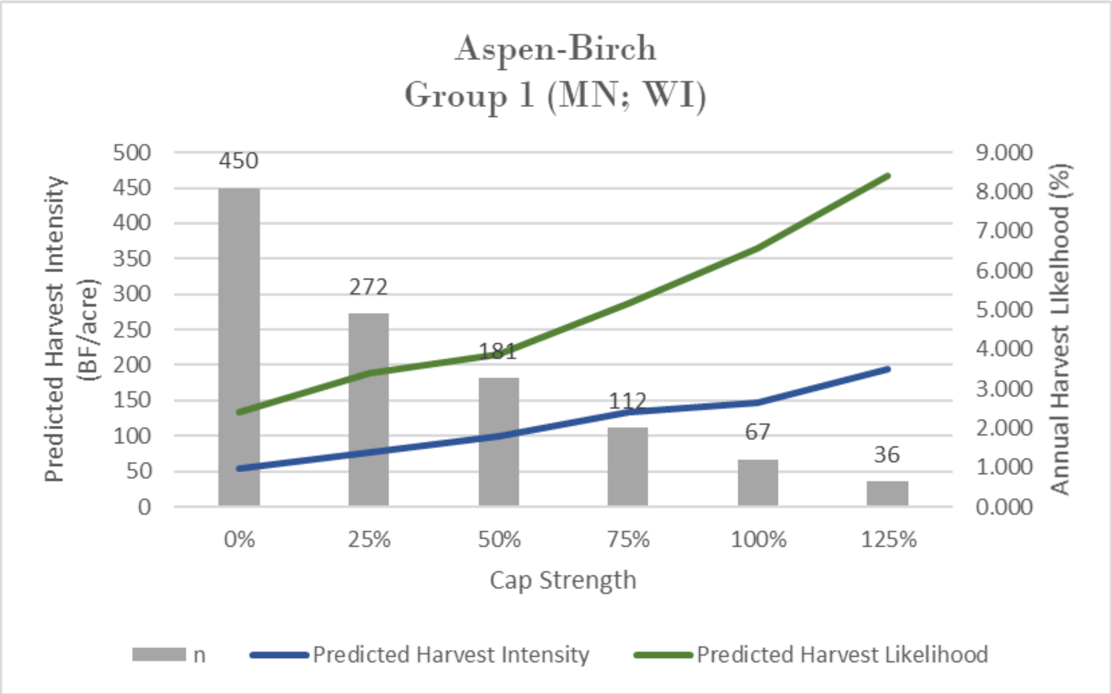


Figure 7. Predicted harvest intensity for Aspen-Birch, Group 1 (MN, WI).

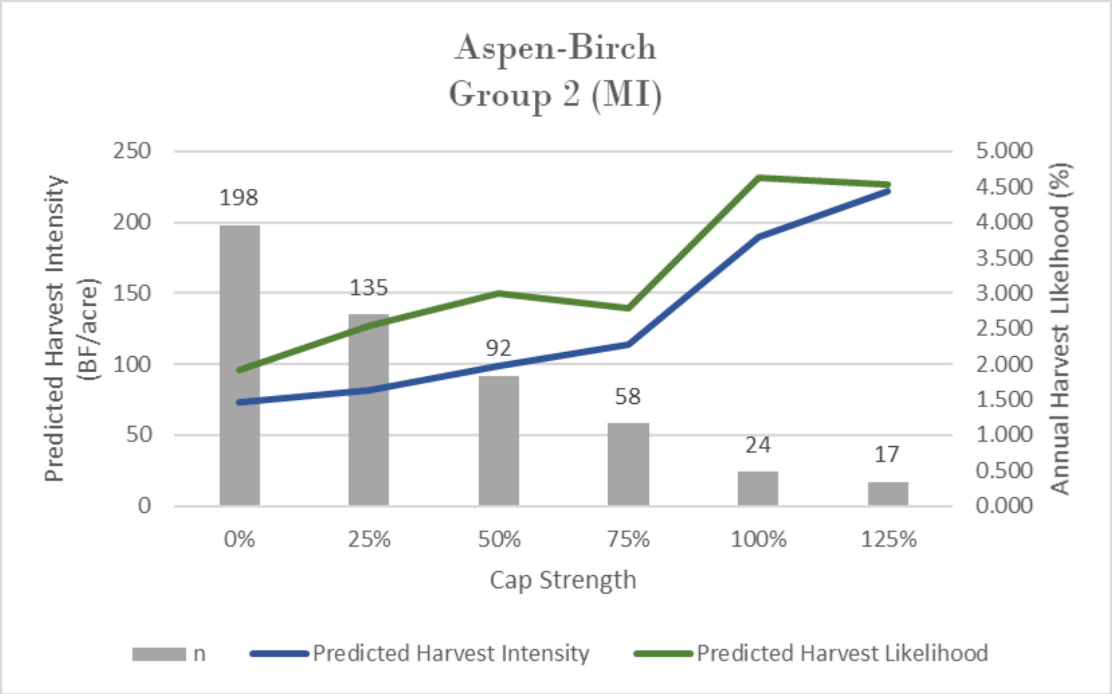


Figure 8. Predicted harvest intensity for Aspen-Birch, Group 2 (MI).

Maple/Beech/Birch

Table 2. Quadratic Mean Diameter (QMD) and Merchantable Volume Caps Impacts on Donor Pool (n), Harvest Likelihood (HL), and Harvest Intensity (HI) for Maple/Beech/Birch

Maple/ Beech/ Birch									
		Caps and Results by Cap Intensity							
	Variable Cap Used	**0%**	**25%**	**50%**	**75%**	**100%**	**112.5%**	**125%**	**137.5%**
	QMD (inches)	0	1.75	3.5	5.25	7	7.875	8.75	9.625
	Merchantable Vol (cu.ft/ ac)	0	125	250	375	500	562.5	625	687.5
	Result Indicator								
Group 1 (MN; WI)	*n*	776	759	720	515	228	*140*	91	*51*
	Predicted Harvest Likelihood	0.041	0.041	0.043	0.046	0.054	*0.055*	0.066	*0.075*
	Predicted Harvest Intensity	58.661	60.082	60.575	67.576	75.017	*77.364*	76.850	*85.521*
Group 2 (MI)	*n*	346	339	315	199	97	*62*	34	*20*
	Predicted Harvest Likelihood	0.043	0.045	0.045	0.051	0.063	*0.062*	0.070	*0.071*
	Predicted Harvest Intensity	72.597	72.857	75.355	86.496	103.996	*111.366*	132.238	*133.882*

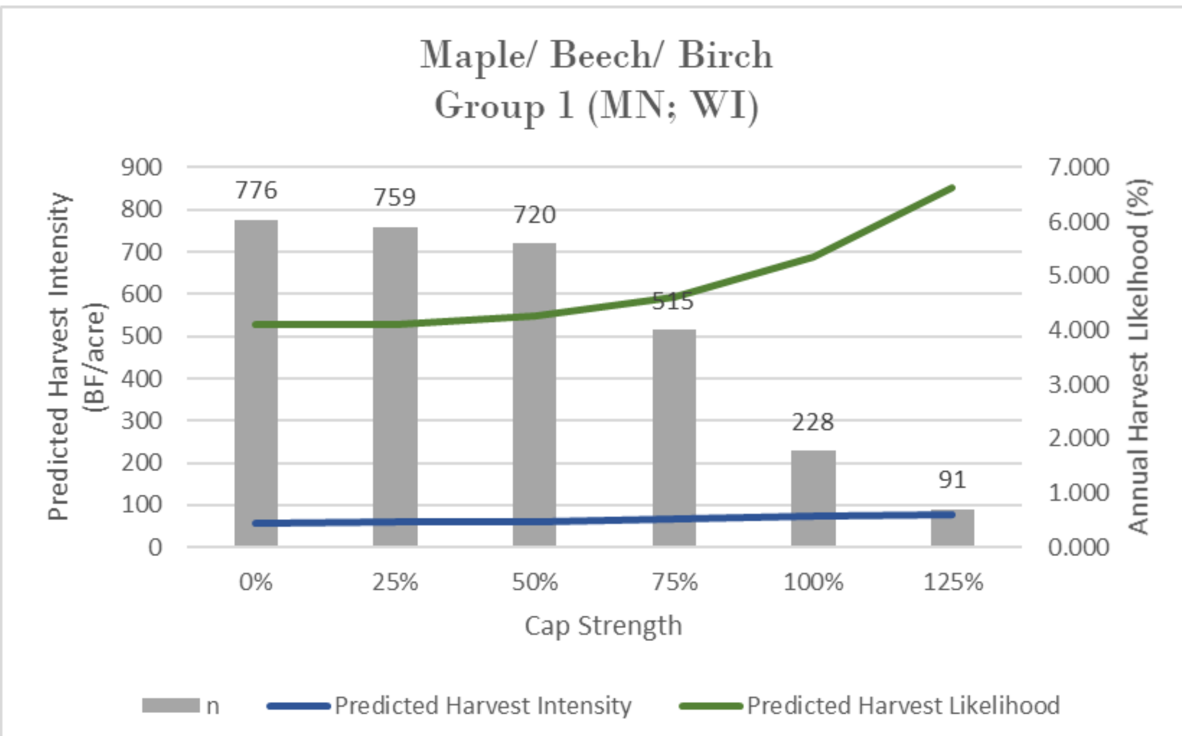


Figure 9. Predicted harvest intensity for Maple/Beech/Birch, Group 1 (MN, WI).

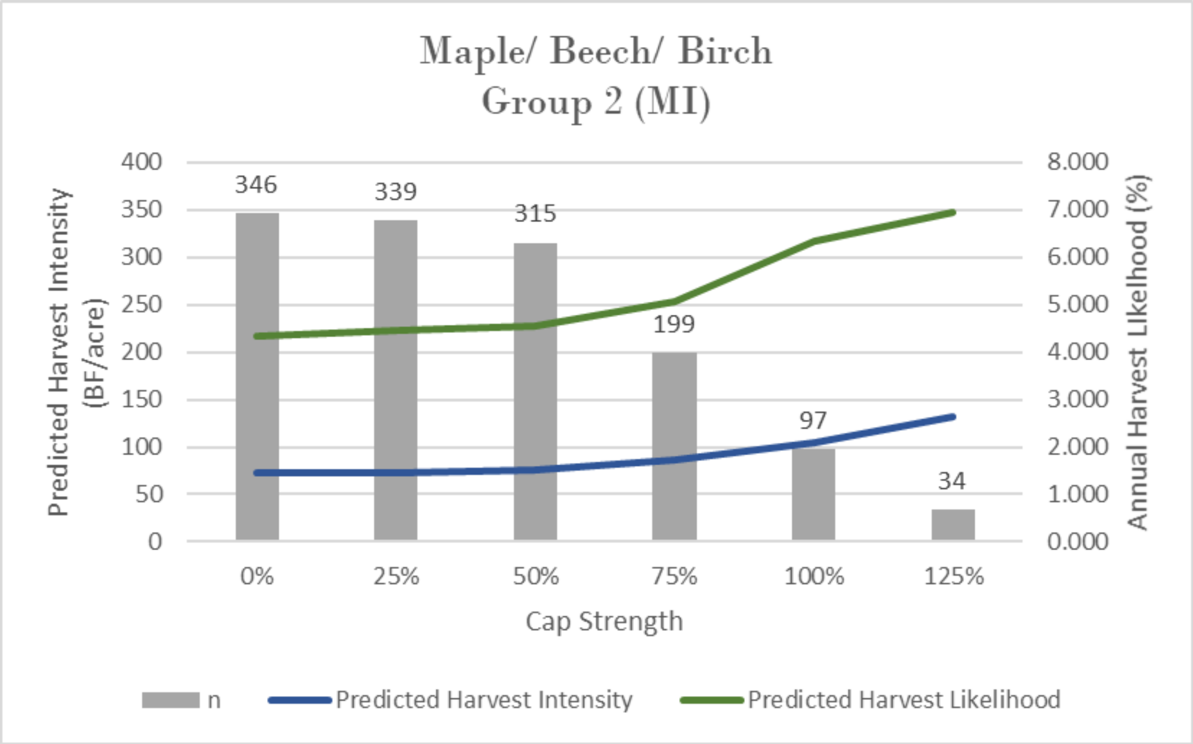


Figure 10. Predicted harvest intensity for Maple/Beech/Birch, Group 2 (MI).

White/Red/Jack Pine

Table 3. Quadratic Mean Diameter (QMD) and Merchantable Volume Caps Impacts on Donor Pool (n), Harvest Likelihood (HL), and Harvest Intensity (HI) for White/Red/Jack Pine

White/ Red/ Jack Pine							
		Caps and Results by Cap Intensity					
	Variable Cap Used	***0%***	***25%***	***50%***	***75%***	***100%***	***125%***
	QMD (inches)	0	0	0	0	0	0
	Merchantable Vol (cu.ft/ ac)	0	125	250	375	500	625
	Result Indicator						
Group 1 (MN; WI)	*n*	90	85	74	60	49	41
	Predicted Harvest Likelihood	0.025	0.027	0.028	0.034	0.035	0.034
	Predicted Harvest Intensity	56.896	60.323	63.049	67.871	77.147	81.807
Group 2 (MI)	*n*	77	68	60	52	41	29
	Predicted Harvest Likelihood	0.024	0.027	0.031	0.035	0.044	0.051
	Predicted Harvest Intensity	75.804	79.358	81.554	88.357	90.886	106.400

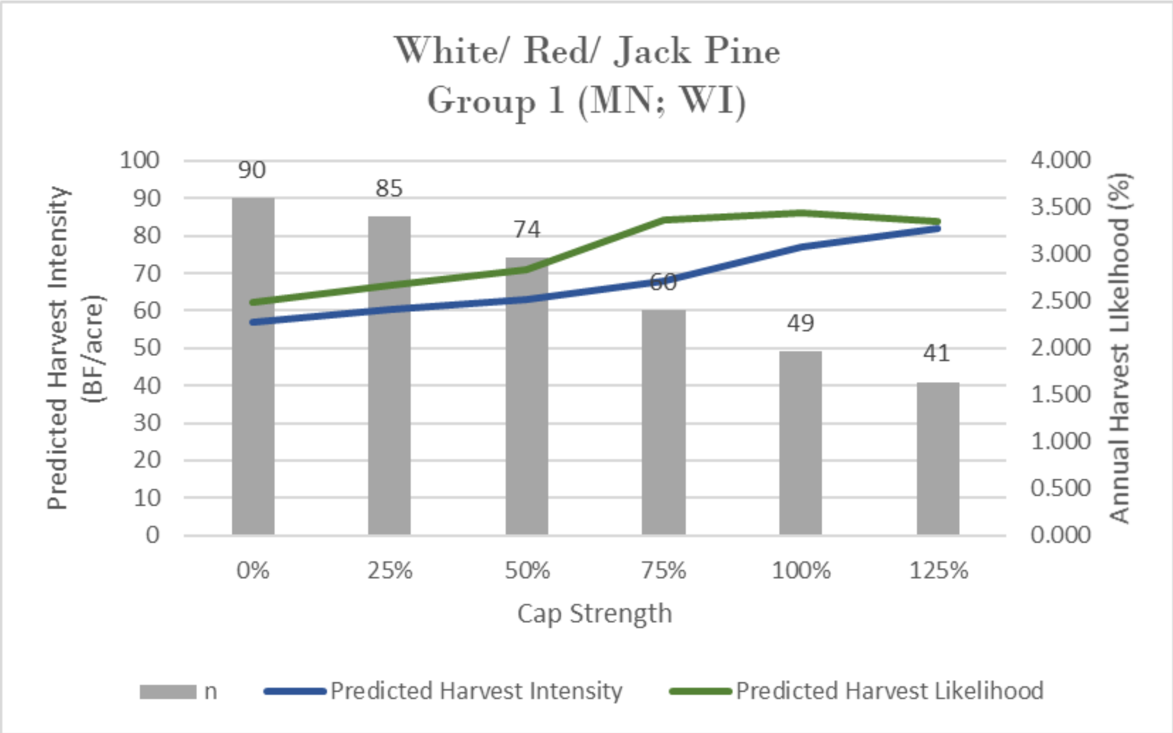


Figure 11. Predicted harvest intensity for White/Red/Jack Pine, Group 1 (MN, WI).

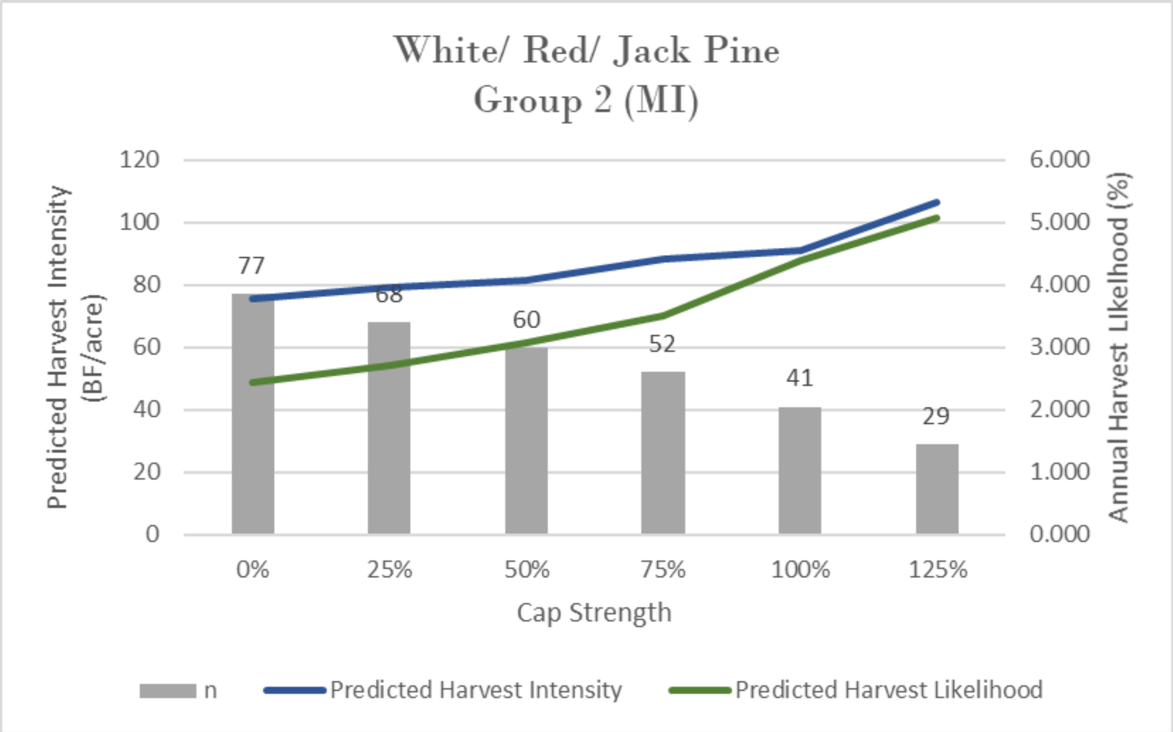


Figure 12. Predicted harvest intensity for White/Red/Jack Pine, Group 2 (MI).

References

Breiman, L. (2001). Random Forests. *Machine Learning*. 45:5-32.

Cleland, D.T.; Freeouf, J.A.; Keys, J.E.; Nowacki, G.J.; Carpenter, C.A.; and McNab, W.H. (2007). Ecological Subregions: Sections and Subsections for the conterminous United States. Gen. Tech. Report WO-76D [Map on CD-ROM] (A.M. Sloan, cartographer). Washington, DC: U.S. Department of Agriculture, Forest Service, presentation scale 1:3,500,000; colored. https://www.fs.usda.gov/research/treesearch/48672

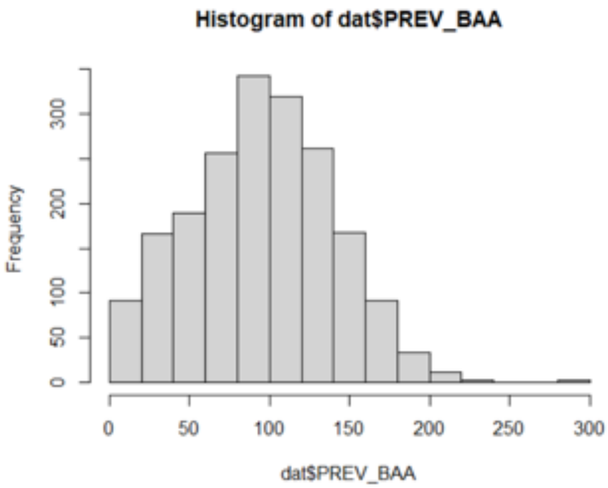
Sass, Emma M.; Butler, Brett J.; Markowski-Lindsay, Marla. (2020). Distribution of forest ownerships across the conterminous United States, 2017. Res. Map NRS-11. Madison, WI: U.S. Department of Agriculture, Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-RMAP-11. [Scale 1: 10,000,000, 1: 80,000,000].

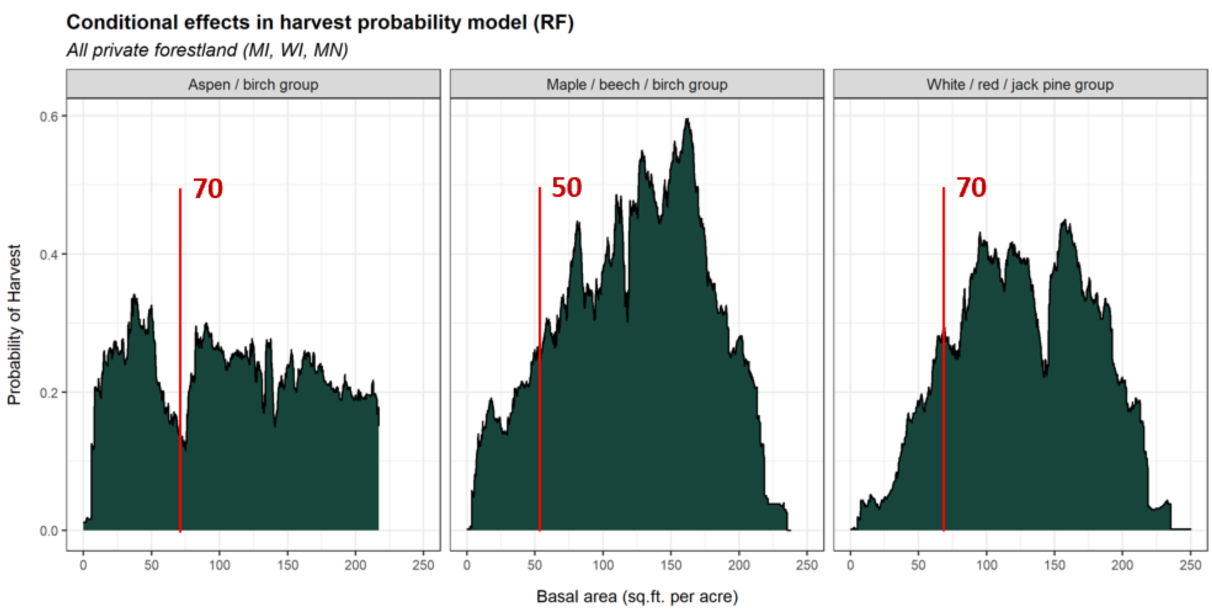
Wager, S., T. Hasie, B. Efron. (2014). Confidence Intervals fro Random Forests: the Jackknife and the infinitesimal Jackknife. *Journal of Machine Learning Research*. 15:1625-1651.

Ziegler, A., Inke R. Konig. (2014). Mining data with random forests: current options for real-world applications. *WIREs Data Mining Knowl Discov*. 4:55-63

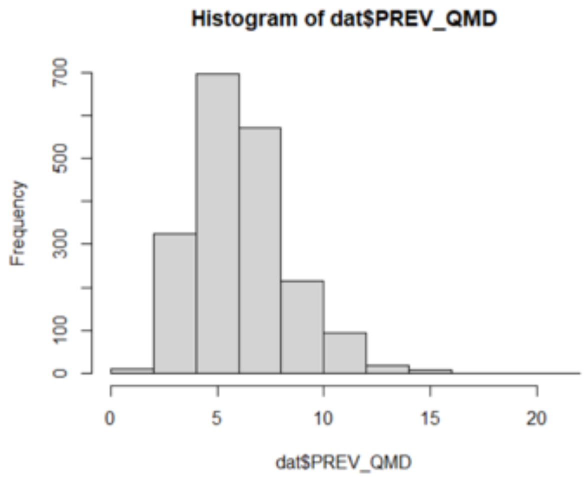
Appendix I: Histograms and Partial Dependence Plots for Initial Cap Selection

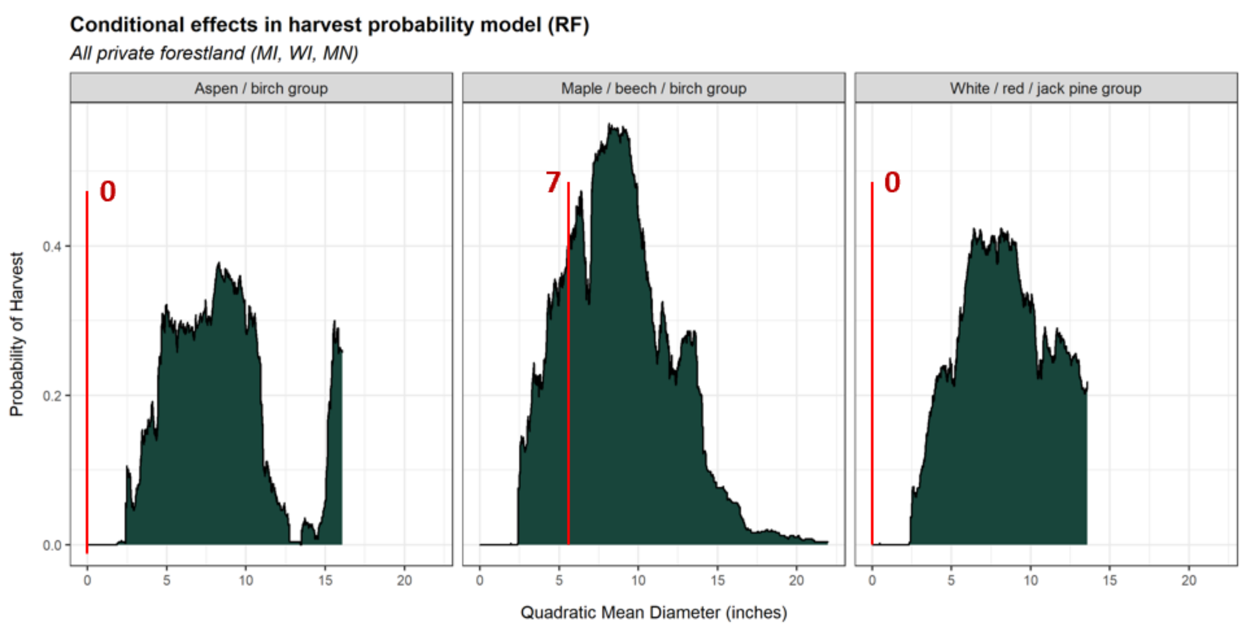
Basal Area



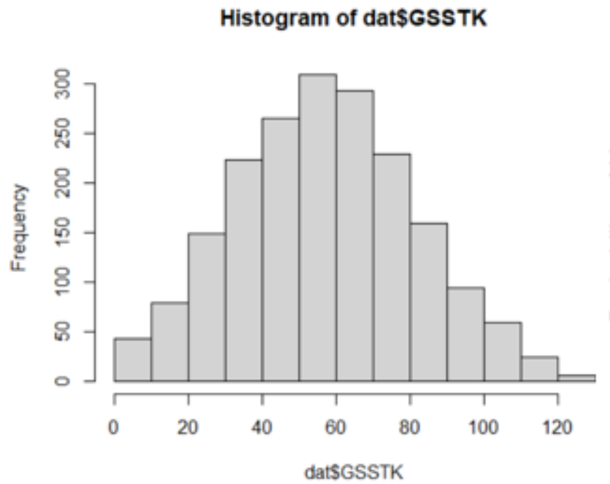


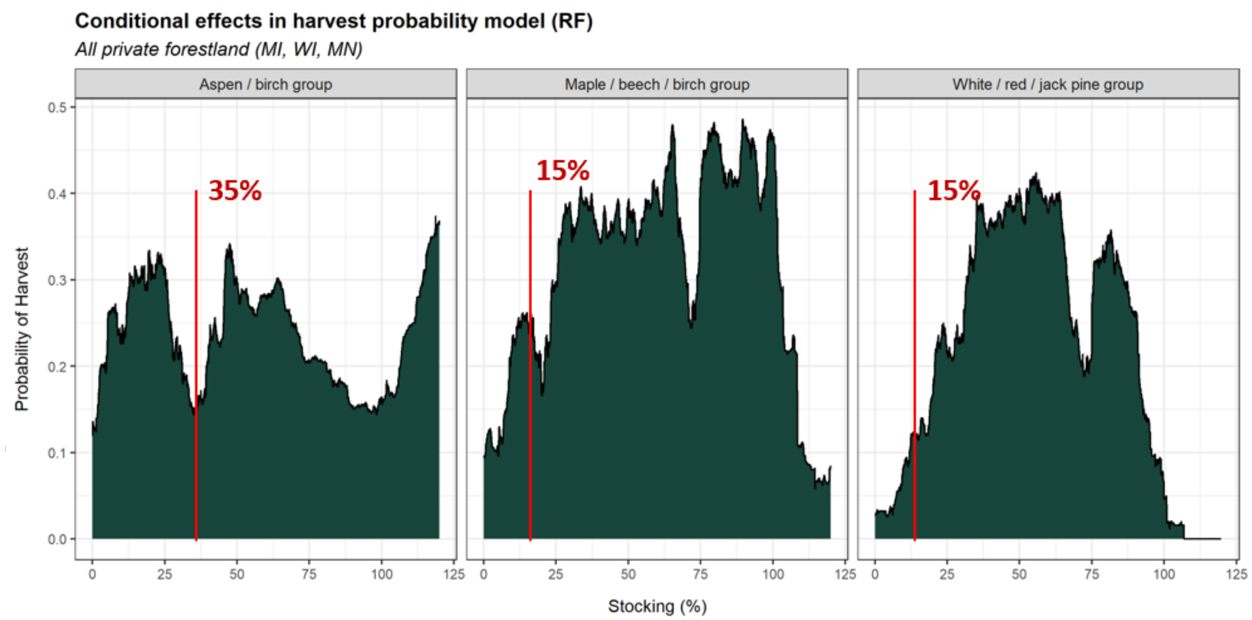
Quadratic Mean Diameter



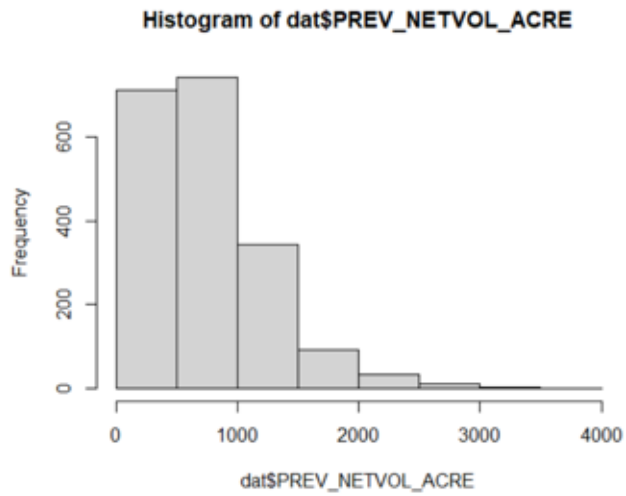


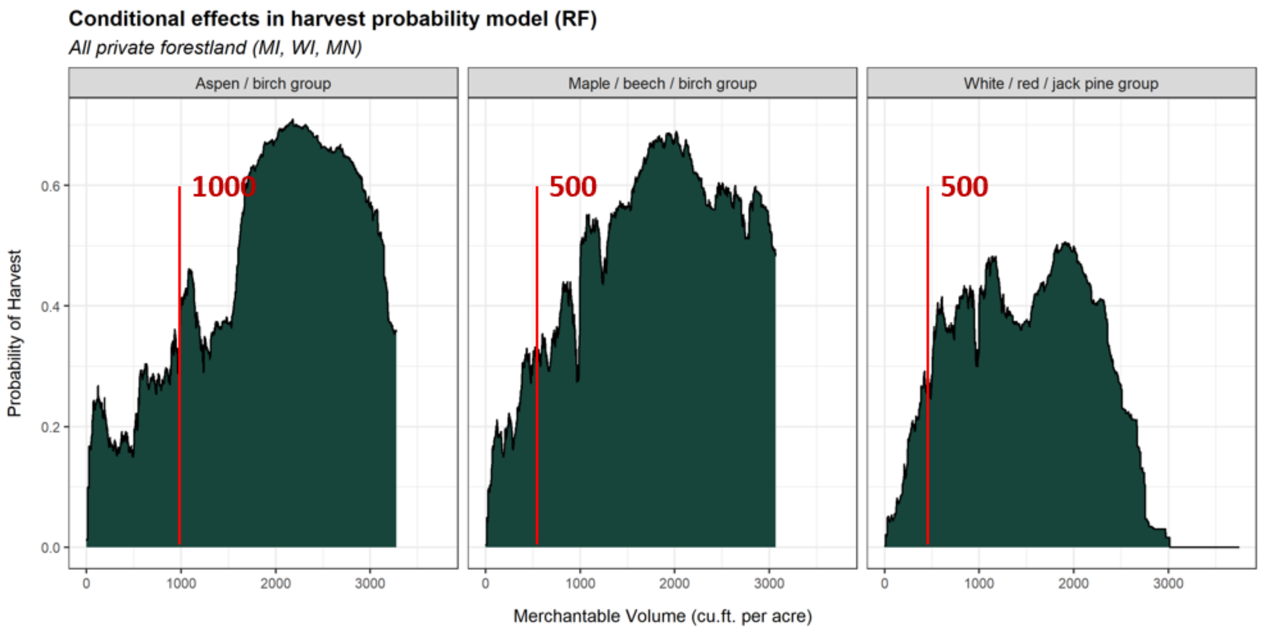
Stocking Percent



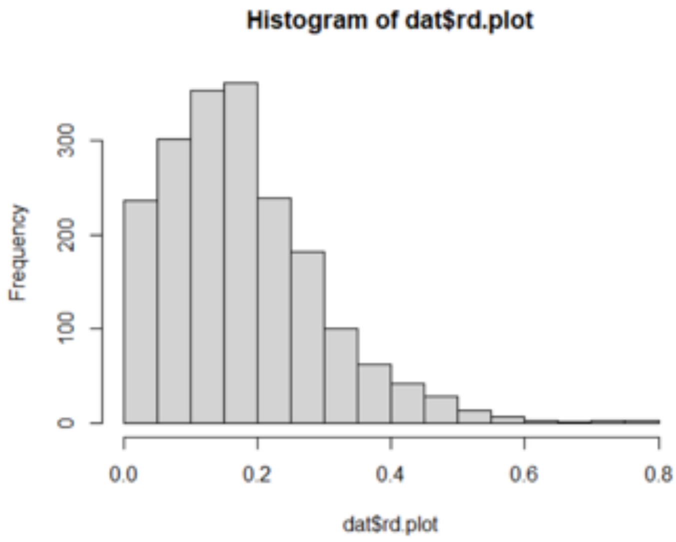


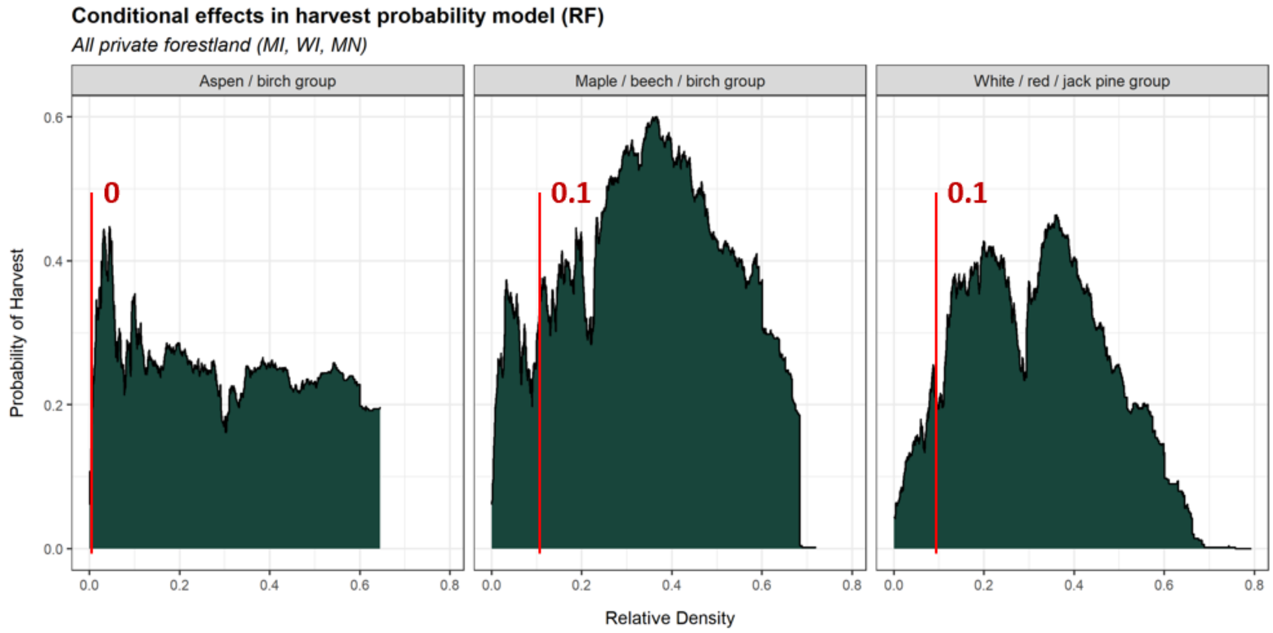
Merchantable Volume



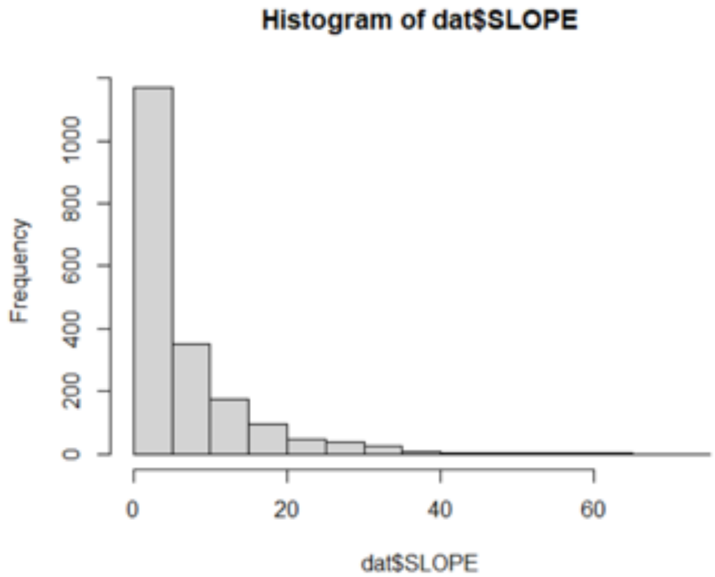


Relative Density (> 5 in diameter trees)

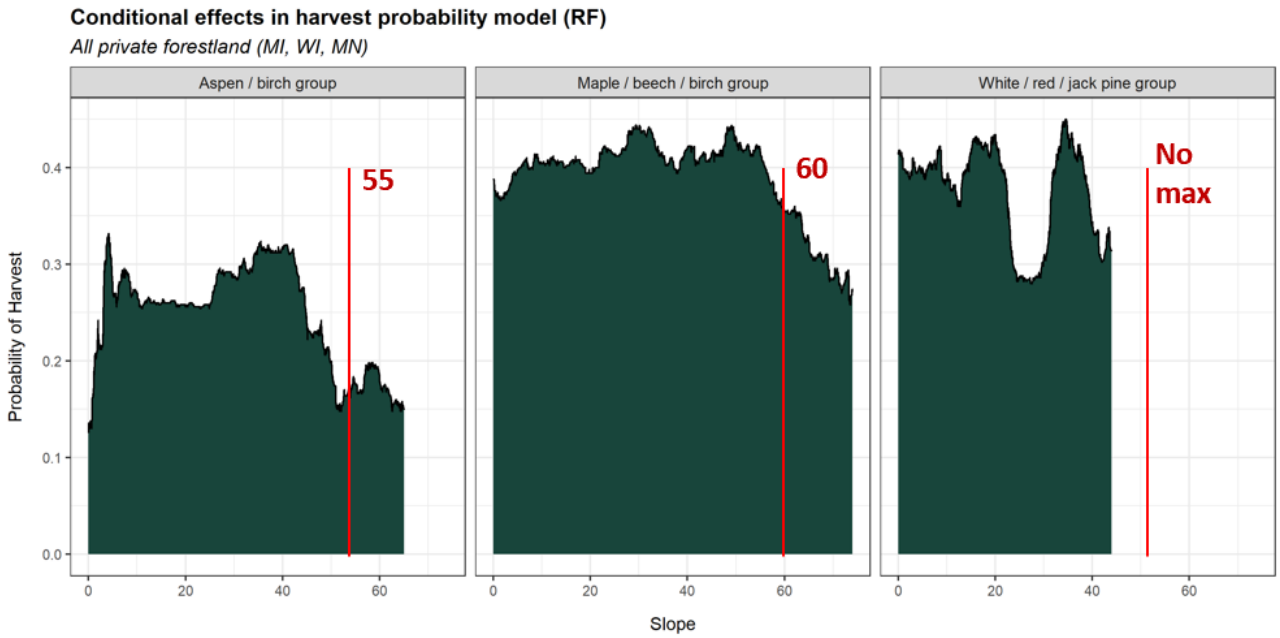




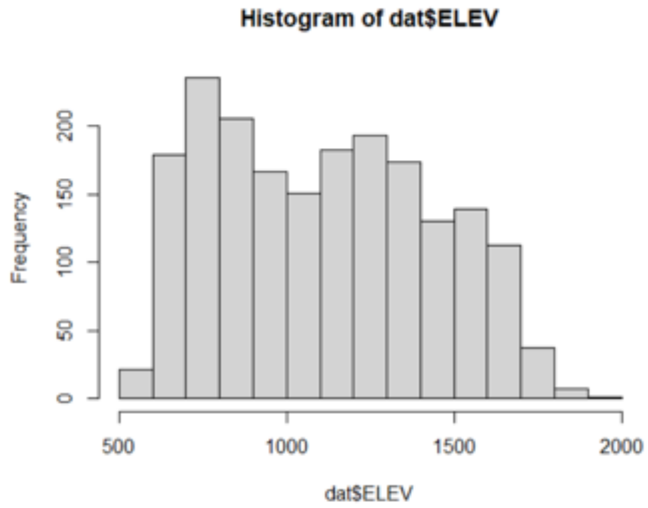
Slope

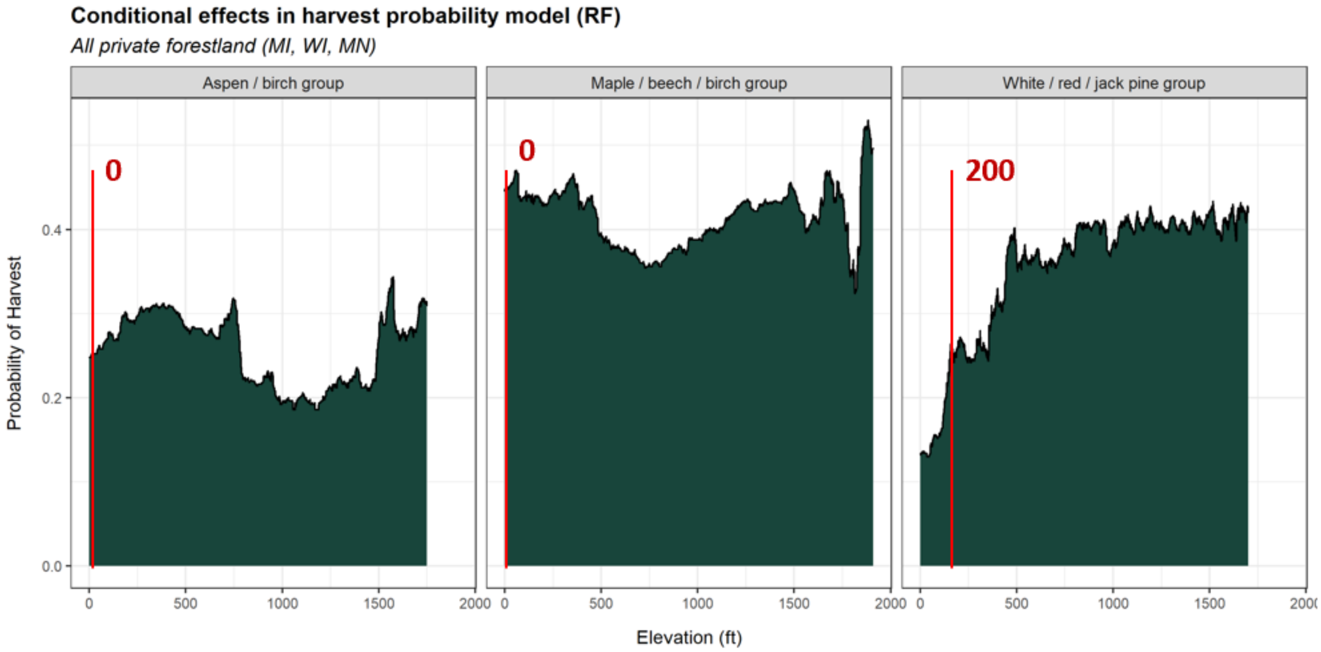

Histogram of dat$SLOPE
Frequency
0
200
400
600
800
1000
0
20
40
60
dat$SLOPE

.

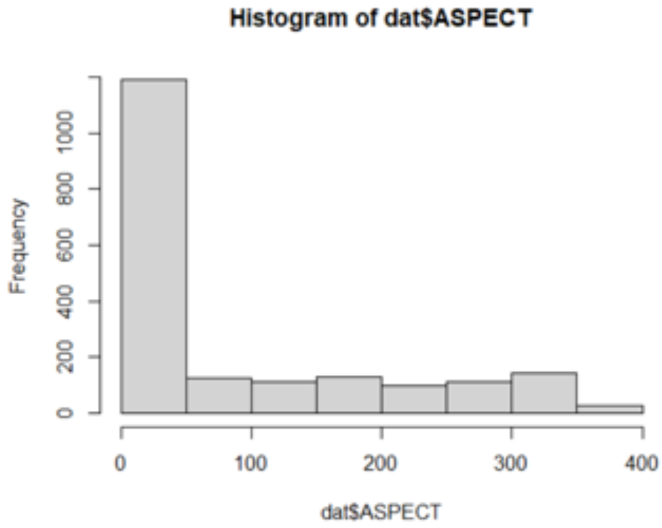

Conditional effects in harvest probability model (RF)
All private forestland (MI, WI, MN)
Aspen / birch group
Maple / beech / birch group
White / red / jack pine group
Probability of Harvest
0.0
0.1
0.2
0.3
0.4
55
60
No
max
0
20
40
60
Slope

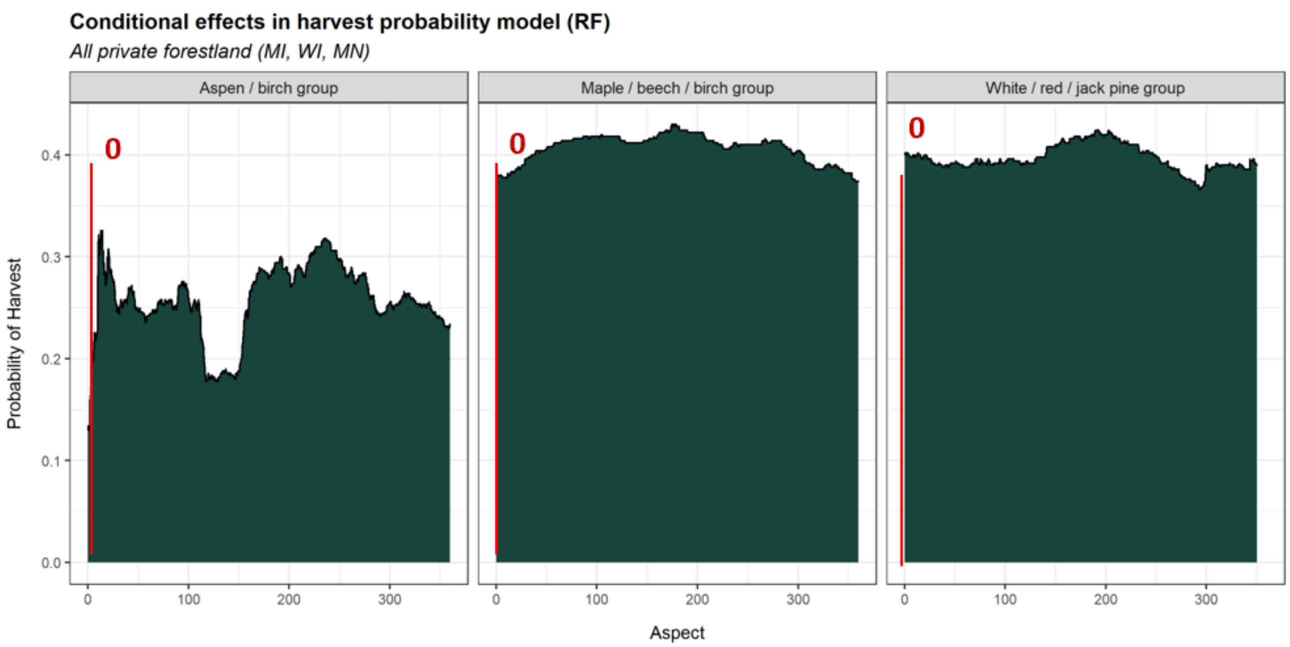
Elevation


Histogram of dat$ELEV
Frequency
0
50
100
150
200
500
1000
1500
2000
dat$ELEV


Conditional effects in harvest probability model (RF)
All private forestland (MI, WI, MN)
Aspen / birch group
Maple / beech / birch group
White / red / jack pine group
0
0
200
Probability of Harvest
0.0
0.2
0.4
Elevation (ft)

Aspect


Histogram of dat$ASPECT
Frequency
0
200
400
600
800
1000
0
100
200
300
400
dat$ASPECT


Conditional effects in harvest probability model (RF)
All private forestland (MI, WI, MN)
Aspen / birch group
Maple / beech / birch group
White / red / jack pine group
0
0
0
Probability of Harvest
0.0
0.1
0.2
0.3
0.4
0
100
200
300
Aspect