February 21, 2020

Mining Public Data to Enhance Forest Assessment

David Diaz

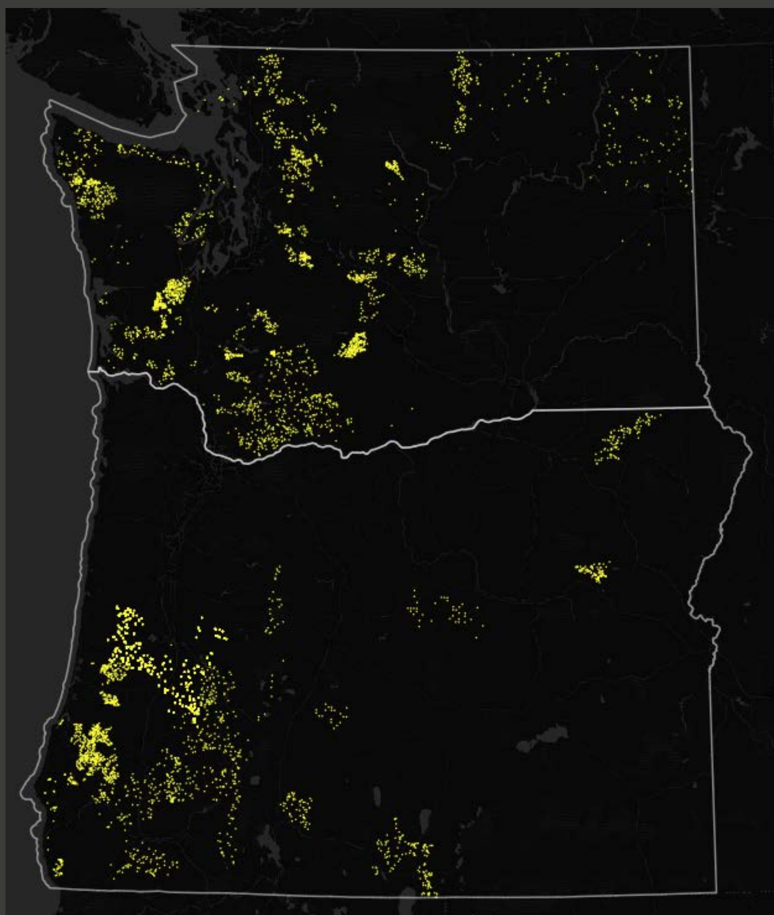
Ecotrust

MOTIVATION

Technology transfer to facilitate forest stewardship planning

Clackamas Tree School, 2016

Plot-sized pixels are increasingly used for building models to generate wall-to-wall forest maps



CSIRO PUBLISHING

www.publish.csiro.au/journals/ijwf *International Journal of Wildland Fire* **2009**, *18*, 235–249

LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment

Matthew G. Rollins

US Geological Survey, Center for Earth Resources Observation and Science (EROS), Sioux Falls, SD 57198, USA. Email: mrollins@usgs.gov

Predictive mapping of forest composition and structure with direct gradient analysis and nearest- neighbor imputation in coastal Oregon, U.S.A.

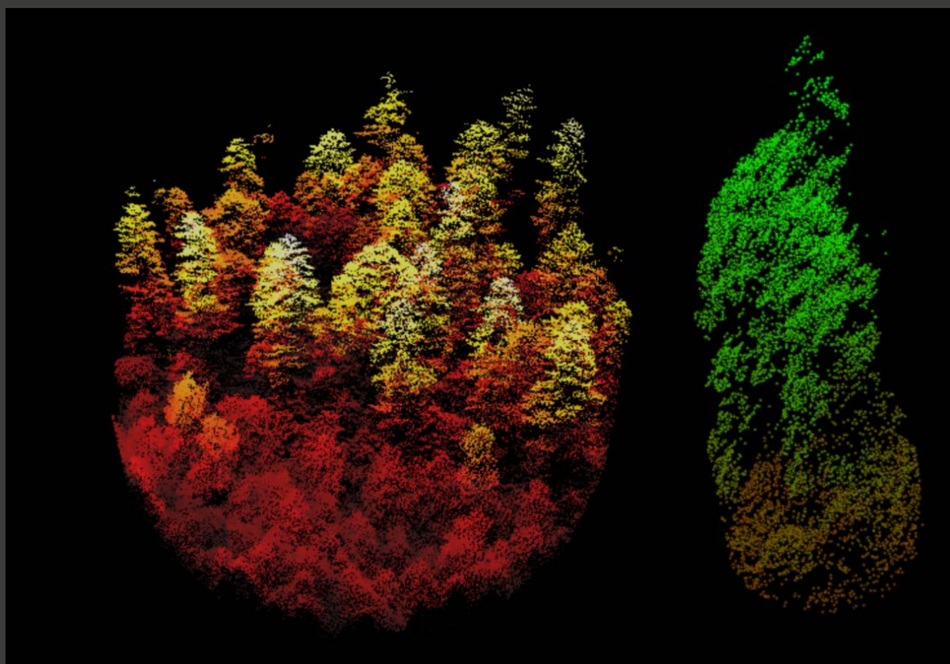
Janet L Ohmann and , Matthew J Gregory

Canadian Journal of Forest Research, 2002, 32(4): 725-741, https://doi.org/10.1139/x02-011

▶▶ Fixed-radius plots collected on public lands with precise location estimates.

Agency	Plot Size	Plots
BLM	1/8 ac (41.6 ft radius)	1,860
WA DNR	1/10 ac (37.2 ft radius)	3,500
USFS	1/4 ac (58.9 ft radius)	800

Correlate plots with remote-sensing data, then train models to fill in the gaps



185 ft radius
1 ha plot

41.6 ft radius
1/8 ac plot

ELSEVIER

Remote Sensing of Environment
Volume 112, Issue 5, 15 May 2008, Pages 2232-2245

Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data

Andrew T. Hudak [a], Nicholas L. Crookston [a], Jeffrey S. Evans [a], David E. Hall [a], Michael J. Falkowski [b]

Relatively modest errors in locating plots and trees can substantially affect predictive models.

Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass

G.W. Frazer [a,*], S. Magnussen [a,1], M.A. Wulder [a,2], K.O. Niemann [b,3]

OPERATIONAL LABELS

Simple, reproducible, forest typing

United States
Department
of Agriculture

Forest Service

Forest
Management
Service Center

Fort Collins, CO

April 2013

A Compendium of NFS Regional Vegetation Classification Algorithms

Don Vandendriesche

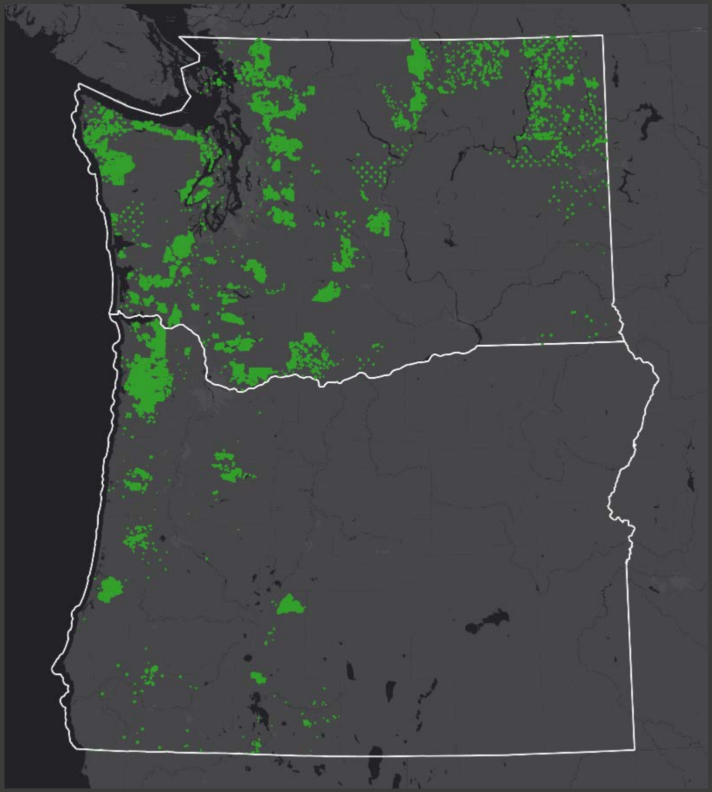
EXAMPLE

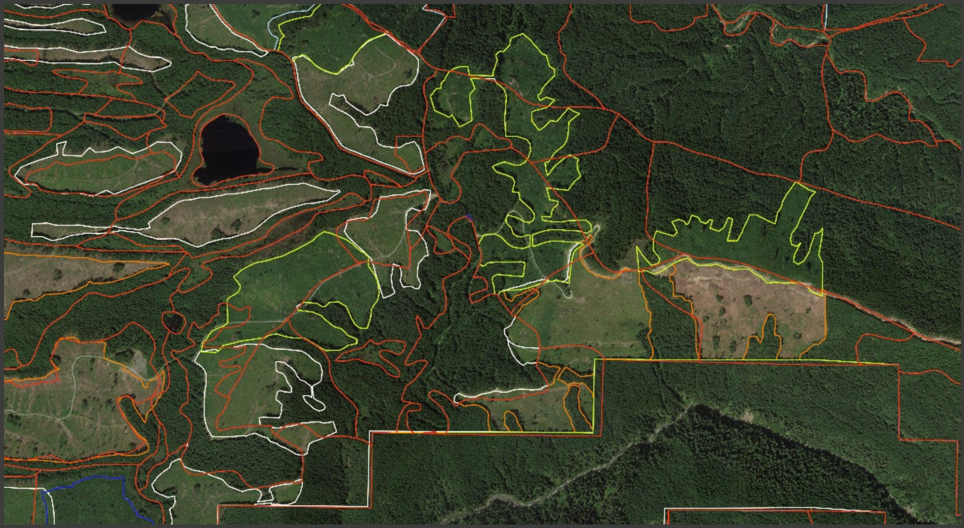
- Douglas-fir / western hemlock / mix
- Medium Tree (10-15” QMD)
- Moderate Cover (40-60%)
- Single Story

DOMINANCE TYPE	
One-species dominance	
Two-species dominance	
Three-species dominance	
Mixed-species dominance	
Non-vegetated	
SIZE CLASS	
0 – Nonstocked	0 - 1" qmd
1 – Seed/Sap	1 - 5" qmd
2 – Small Tree	5 - 10" qmd
3 – Medium Tree	10 - 15" qmd
4 – Large Tree	15 - 20" qmd
5 – Very Large Tree	20"+ qmd
CANOPY COVER	
0 – Sparse	0 - 10%
1 – Open	10 - 40%
2 – Moderate	40 - 70%
3 – Closed	70%+
CANOPY LAYERS	
0 – Nonstocked	
1 – Single Story	
2 – Multiple Story	

TRADITIONAL INVENTORY PLOTS

Public domain Stand-Level Inventory

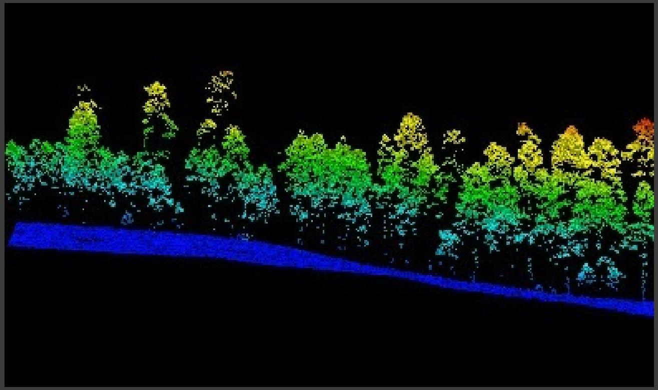




- WA DNR: 369,000 stand polygons over six different years (2004-2017)
- ODF: 15,000 current stand polygons

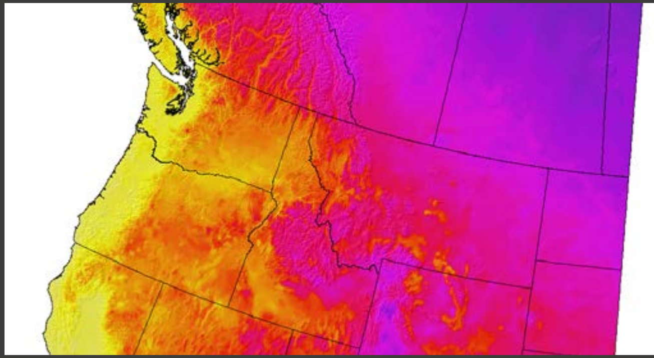
450,000 plot measurements have occurred in these stands, containing tree-level observations (species, diameter, height, etc.)

GENERATING FEATURES



LIDAR

Publicly-available lidar point clouds covering several million acres are being processed into 20+ rasters characterizing terrain and canopy. 0.5-1m resolution surface and intensity rasters, 10m resolution canopy metrics.



CLIMATE

Down-scaled monthly, seasonal, and annual climatic data and derived metrics relevant for vegetation modeling extracted from Climate WNA (Wang 2012).

IMAGERY

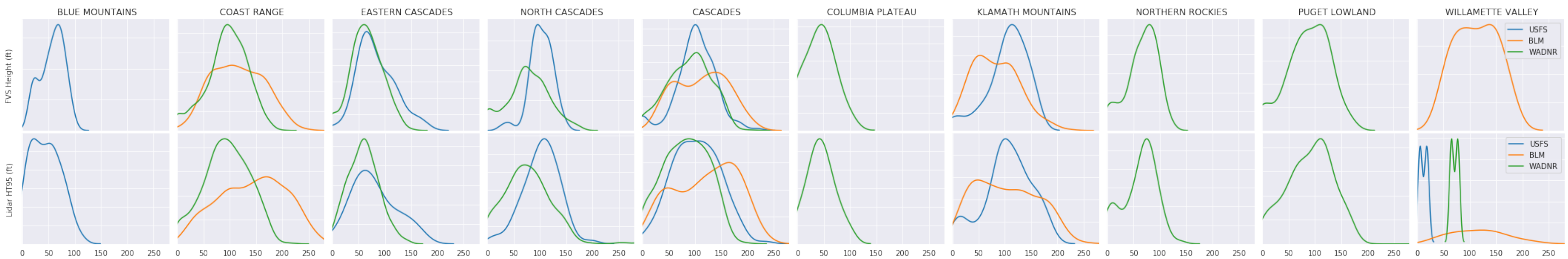
Aerial (NAIP) and satellite imagery extracted using Google Earth Engine. Time series of several images per year collected from SENTINEL and LANDSAT to facilitate species identification. Histograms of values within stand polygons also extracted.

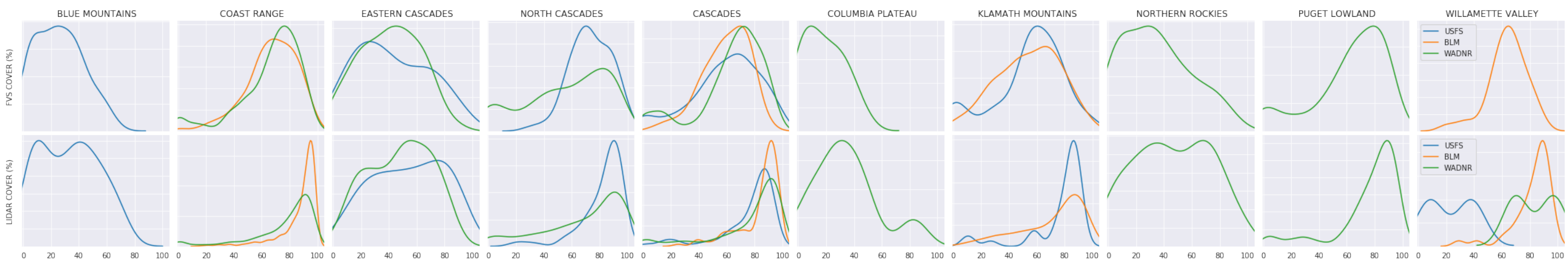
(we'll settle for airplanes)

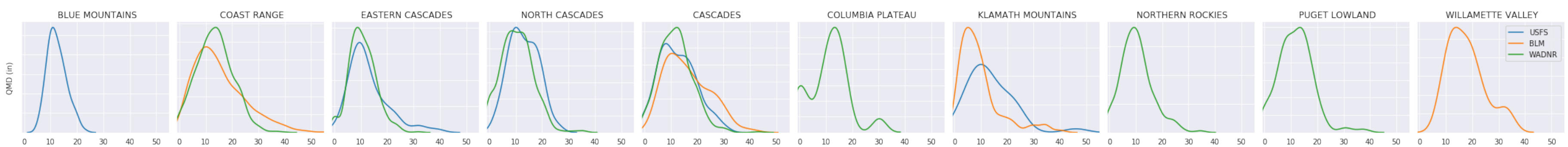
Working with data across agencies, ecoregions, and lidar acquisitions

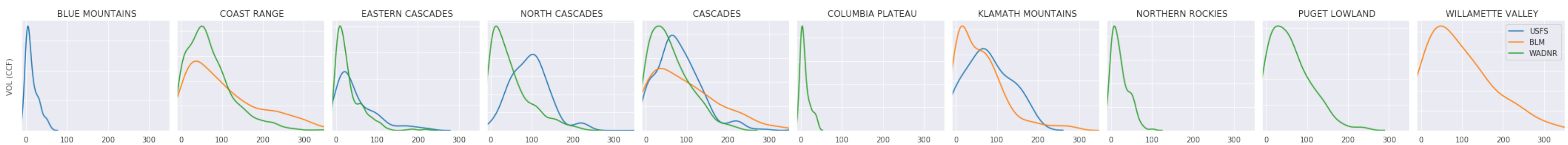
ECOREGION	# PLOTS	# LIDAR ACQS	# LIDAR YRS	# AGENCIES
Blue Mountains	104	2	2	1
Cascades	579	21	9	3
Coast Range	1,511	18	9	2
Columbia Plateau	8	4	4	1
Eastern Cascades	202	11	6	2
Klamath Mountains	253	7	5	2
North Cascades	405	18	6	2
Northern Rockies	71	4	3	1
Puget Lowland	110	14	7	1
Willamette Valley	41	2	2	2
TOTAL	**3,284**	**70**	**11**	**3**

TRANSFER LEARNING
continued



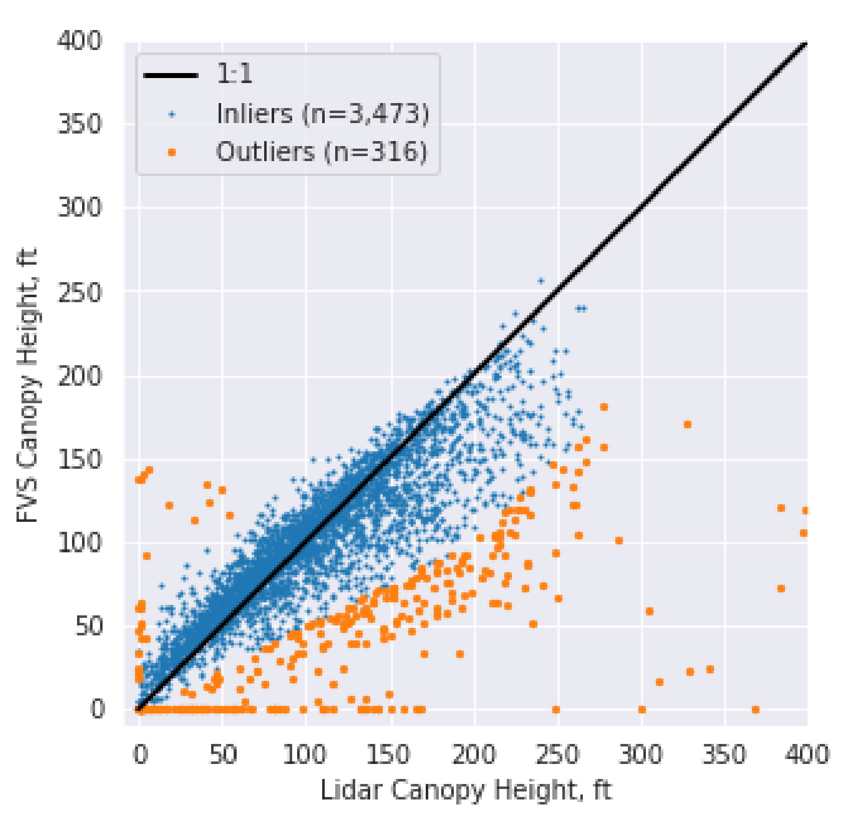






OUTLIER REMOVAL

Addressing time lags and co-registration error



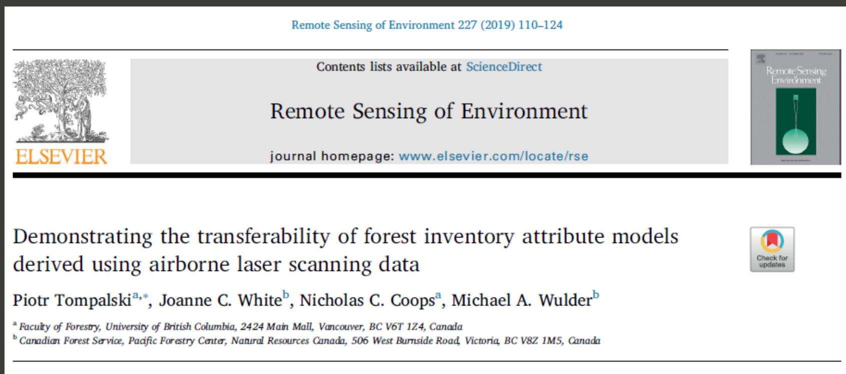
FVS simulation was used to grow field-measured trees forward on an annual basis.

For each year where a plot had available lidar data, the closest-matching year of satellite imagery and FVS data were chosen. Differences of 5+ yrs between FVS and lidar data were discarded (n=829).

Lidar clips with fewer than 4 returns/m2 were then discarded (n=103)

▶▶ FVS Top Height and Lidar HT95 were then compared to exclude plots with substantial differences (n=316)

Transferring Point Cloud Models in Forest Inventory


Remote Sensing of Environment 227 (2019) 110–124

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data

Piotr Tompalski[a,*], Joanne C. White[b], Nicholas C. Coops[a], Michael A. Wulder[b]

[a] Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

[b] Canadian Forest Service, Pacific Forestry Center, Natural Resources Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada

Learning from 239 plots from
3 study areas in coastal BC

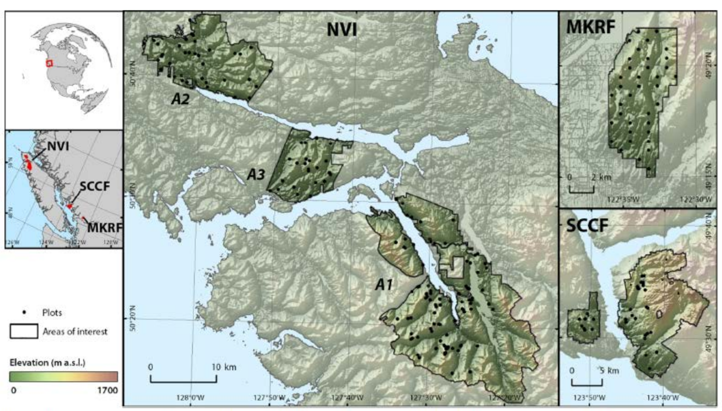


Fig. 1. Outline of the study areas used in the presented research. NVI – North Vancouver Island and sub-areas A1, A2, and A3; MKRF – Malcolm Knapp Research Forest; SCCF – Sunshine Coast Community Forest. The distance between NVI and SCCF was approximately 270 km, and between NVI and MKRF approximately 350 km.

▶▶ "Scenario 1" investigates how well models trained to predict TOPHT, QMD, and VOL in one area extrapolate to another.

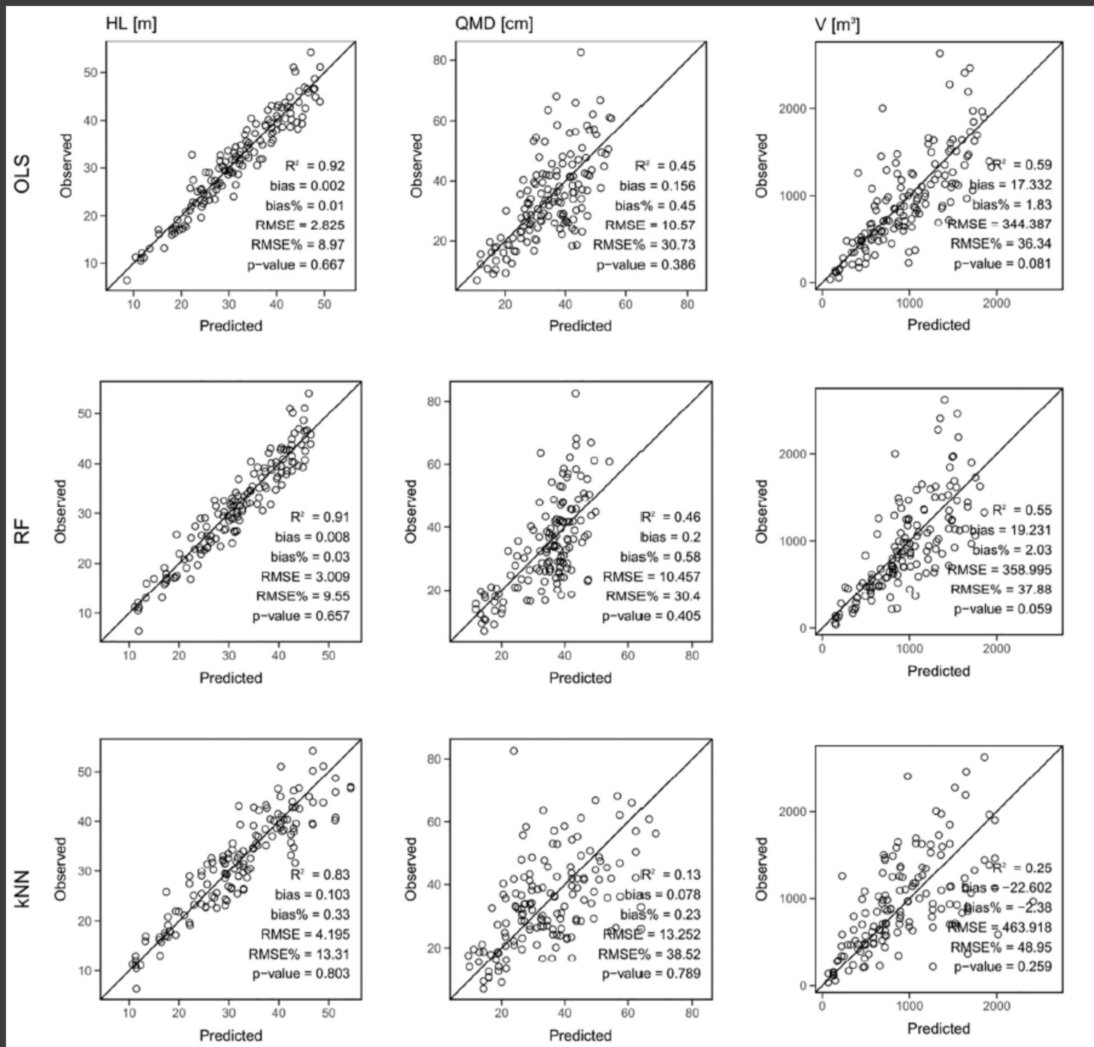
Benchmarks provided by "global" models trained on all plots. "Regional" models are trained on plots from a single study area or sub-area, and tested on another one.

Three types of models are fit:

1. Linear (OLS)
2. kNN Regression
3. Random Forest Regression

Model performance quantified using R^2, as well as absolute and relative bias and RMSE.

Performance of Global Models



In English, please:

TOP HEIGHT (ft)

model	*RMSE*	
OLS	9.3	9%
RF	9.9	10%
kNN	13.8	13%

QMD (in)

model	*RMSE*	
OLS	4.2	31%
RF	4.1	30%
kNN	5.2	39%

VOL (CCF/ac)

model	*RMSE*	
OLS	49	36%
RF	51	38%
kNN	66	49%

Transfer between study areas

Table 4. Average differences in bias% and RMSE% for the transferred models.

Attribute	Method	Scenario 1	
		Δbias%	ΔRMSE%
HL	OLS	0.67	−0.14
	RF	6.32	1.51
	kNN	1.17	2.94
QMD	OLS	9.69	−1.33
	RF	4.49	0.75
	kNN	12.80	23.55
V	OLS	12.29	4.98
	RF	5.03	−2.69
	kNN	8.88	0.23

Positive values (increase in bias% or RMSE%) indicate a decrease in prediction accuracy.

- **Average bias always increased with global models.** Few models saw improved precision (RMSE).
- Height transfer is most robust. Transfer learning for other variables cautioned, particularly to other forest types.
- RF was more robust for QMD and VOL transfer while OLS was most robust for Height transfer.

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Top Height

Tompalski et al. (2019)

TOP HEIGHT (ft)		
model	*RMSE*	
OLS	9.3	9%
RF	9.9	10%
kNN	13.8	13%

RMSE

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	11.9	11.9	11.9	12.4	10.6	11.2	11.9	104
Coast Range	16.7	16.6	16.9	17.3	15.9	15.8	16.8	1,580
North Cascades	16.3	15.5	15.5	12.7	12.8	14.0	15.0	443
Cascades	14.5	14.2	15.4	15.0	13.8	13.7	13.5	613
Klamath Mountains	15.9	16.2	17.6	15.7	13.8	14.0	15.1	275
Eastern Cascades	13.8	13.4	14.6	11.7	9.9	9.8	11.4	203
Northern Rockies	10.5	10.3	10.6	11.5	10.1	10.7	11.2	83
Puget Lowland	7.2	7.1	8.2	8.0	6.8	6.2	8.2	122
Willamette Valley	7.2	6.9	8.0	12.8	9.6	8.9	11.7	41
All	15.5	15.3	15.8	15.3	14.1	14.2	15.0	3,473

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Top Height

Tompalski et al. (2019)

TOP HEIGHT (ft)

model	*RMSE*	
OLS	9.3	9%
RF	9.9	10%
kNN	13.8	13%

RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	22%	22%	22%	23%	20%	21%	22%	104
Coast Range	15%	15%	15%	15%	14%	14%	15%	1,580
North Cascades	18%	18%	18%	14%	14%	16%	17%	443
Cascades	15%	14%	16%	15%	14%	14%	14%	613
Klamath Mountains	16%	17%	18%	16%	14%	15%	16%	275
Eastern Cascades	19%	18%	20%	16%	14%	14%	16%	203
Northern Rockies	15%	15%	15%	17%	15%	16%	16%	83
Puget Lowland	7%	7%	8%	8%	7%	6%	8%	122
Willamette Valley	6%	6%	6%	10%	8%	7%	10%	41
All	16%	15%	16%	15%	14%	14%	15%	3,473

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Top Height

Tompalski et al. (2019)

TOP HEIGHT	
model	*ΔRMSE%*
OLS	-0,1%
RF	+1,5%
kNN	+2,9%

Change in RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	+1%	+1%	0%	+1%	1%	+2%	+2%	104
Coast Range	0%	0%	0%	-1%	-1%	0%	0%	1,580
North Cascades	0%	+2%	+2%	-1%	0%	+2%	-1%	443
Cascades	+1%	+1%	+2%	+1%	+2%	+2%	0%	613
Klamath Mountains	+1%	+1%	+2%	+1%	+1%	+1%	0%	275
Eastern Cascades	+2%	+2%	+4%	-3%	-4%	-3%	-10%	203
Northern Rockies	0%	0%	+1%	-3%	-5%	-3%	+2%	83
Puget Lowland	-2%	-1%	0%	-4%	-2%	-5%	-3%	122
Willamette Valley	-1%	0%	0%	-3%	-5%	-5%	+2%	41

Positive values indicate a decrease in prediction accuracy.

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Top Height

Tompalski et al. (2019)

TOP HEIGHT	
model	*Δbias%*
OLS	+0.6%
RF	+6.3%
kNN	+1.2%

Change in Bias%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	+2%	2%	-11%	-3%	-3%	-3%	+4%	104
Coast Range	+1%	0%	+2%	0%	+1%	0%	+1%	1,580
North Cascades	+2%	+1%	-1%	+2%	+2%	+2%	0%	443
Cascades	-3%	-3%	-4%	-3%	-3%	-3%	-3%	613
Klamath Mountains	+4%	+4%	+4%	+3%	+4%	+4%	2%	275
Eastern Cascades	0%	0%	-3%	0%	0%	-1%	-2%	203
Northern Rockies	-1%	0%	-3%	-1%	-2%	-1%	0%	83
Puget Lowland	0%	0%	-1%	-1%	0%	0%	-1%	122
Willamette Valley	0%	0%	0%	-2%	+2%	+2%	-4%	41

Positive values indicate a decrease in prediction accuracy.

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Quadratic Mean Diameter

Tompalski et al. (2019)

QMD (in)		
model	*RMSE*	
OLS	4.2	31%
RF	4.1	30%
kNN	5.2	39%

RMSE

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	2.1	2.1	2.2	1.8	1.9	1.8	1.7	104
Coast Range	5.5	5.5	5.4	5.7	5.4	5.4	5.6	1,580
North Cascades	3.5	3.5	3.7	3.4	3.5	3.5	3.6	443
Cascades	3.7	3.7	3.9	3.6	3.3	3.4	3.6	613
Klamath Mountains	7.2	7.2	7.3	6.5	6.7	6.8	7.2	275
Eastern Cascades	3.7	3.7	3.7	3.7	3.7	3.8	3.5	203
Northern Rockies	2.9	2.9	2.7	2.9	2.8	2.6	2.3	83
Puget Lowland	2.7	2.7	2.9	3.4	2.9	2.8	3.3	122
Willamette Valley	6.4	6.3	6.1	8.5	6.6	6.2	6.2	41
All	5.0	5.0	5.0	5.0	4.8	4.8	4.9	3,473

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Quadratic Mean Diameter

Tompalski et al. (2019)

QMD (in)		
model	*RMSE*	
OLS	4.2	31%
RF	4.1	30%
kNN	5.2	39%

RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	20%	20%	21%	17%	18%	17%	16%	104
Coast Range	36%	36%	35%	37%	35%	36%	36%	1,580
North Cascades	29%	29%	31%	28%	28%	28%	29%	443
Cascades	30%	30%	31%	28%	26%	27%	29%	613
Klamath Mountains	67%	67%	67%	60%	62%	62%	66%	275
Eastern Cascades	32%	32%	32%	32%	32%	33%	30%	203
Northern Rockies	31%	31%	29%	31%	30%	29%	26%	83
Puget Lowland	22%	22%	24%	28%	24%	23%	27%	122
Willamette Valley	30%	29%	29%	40%	31%	30%	29%	41
All	37%	37%	37%	37%	35%	36%	36%	3,473

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Quadratic Mean Diameter

Tompalski et al. (2019)

QMD	
model	*ΔRMSE%*
OLS	-1.3%
RF	+0.8%
kNN	+23.6%

Change in RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	-1%	-1%	+1%	-8%	-8%	-8%	-7%	104
Coast Range	0%	0%	-3%	0%	0%	+1%	+1%	1,580
North Cascades	-2%	-3%	-2%	-3%	+1%	0%	-1%	443
Cascades	-1%	-1%	0%	-2%	-2%	-3%	-3%	613
Klamath Mountains	+8%	+9%	+8%	-1%	+1%	+2%	+10%	275
Eastern Cascades	0%	0%	1%	-5%	-2%	-3%	-7%	203
Northern Rockies	+3%	+3%	1%	-3%	-7%	-11%	-5%	83
Puget Lowland	-7%	-8%	-7%	-19%	-5%	-10%	-7%	122
Willamette Valley	+7%	+6%	+7%	+7%	+2%	+8%	+5%	41

Positive values indicate a decrease in prediction accuracy.

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Quadratic Mean Diameter

Tompalski et al. (2019)

QMD	
model	*Δbias%*
OLS	+9.7%
RF	+4.5%
kNN	+12.8%

Change in Bias%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	-25%	-24%	-21%	-13%	-21%	-20%	-12%	104
Coast Range	-1%	0%	2%	-1%	0%	0%	1%	1,580
North Cascades	3%	2%	1%	7%	5%	4%	5%	443
Cascades	-2%	-2%	-3%	-2%	-1%	-2%	-3%	613
Klamath Mountains	30%	30%	29%	25%	23%	24%	32%	275
Eastern Cascades	-4%	-4%	-4%	-9%	-8%	-7%	-3%	203
Northern Rockies	2%	3%	8%	-9%	-5%	-4%	10%	83
Puget Lowland	-3%	-3%	-5%	-6%	2%	1%	1%	122
Willamette Valley	-13%	-12%	-14%	-13%	-9%	-16%	-10%	41

Positive values indicate a decrease in prediction accuracy.

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Cubic Volume

Tompalski et al. (2019)

VOL (CCF/ac)

model	*RMSE*	
OLS	49	36%
RF	51	38%
kNN	66	49%

RMSE

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	8	8	8	9	9	9	10	104
Coast Range	51	51	51	56	53	54	56	1,580
North Cascades	35	35	36	30	28	26	30	443
Cascades	28	28	28	29	28	28	30	613
Klamath Mountains	29	29	30	26	27	26	28	275
Eastern Cascades	15	15	15	16	16	16	15	203
Northern Rockies	6	7	7	10	7	6	6	83
Puget Lowland	19	19	20	22	20	20	23	122
Willamette Valley	43	40	46	62	54	49	55	41
All	40	40	40	42	40	40	43	3,473

MEANWHILE, SOUTH OF THE BORDER...

Performance of Global Models: Cubic Volume

Tompalski et al. (2019)

VOL (CCF/ac)		
model	*RMSE*	
OLS	49	36%
RF	51	38%
kNN	66	49%

RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	54%	59%	57%	62%	60%	66%	69%	104
Coast Range	49%	49%	48%	53%	50%	51%	53%	1,580
North Cascades	59%	59%	61%	51%	47%	45%	51%	443
Cascades	36%	36%	37%	38%	36%	36%	38%	613
Klamath Mountains	38%	38%	40%	34%	36%	34%	36%	275
Eastern Cascades	42%	41%	40%	44%	44%	45%	42%	203
Northern Rockies	22%	24%	26%	36%	23%	20%	22%	83
Puget Lowland	28%	28%	30%	32%	29%	30%	33%	122
Willamette Valley	35%	33%	37%	50%	44%	40%	45%	41
All	49%	49%	49%	52%	50%	49%	52%	3,473

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Cubic Volume

Tompalski et al. (2019)

VOL	
model	*ΔRMSE%*
OLS	+5.0%
RF	-2.7%
kNN	+0.2%

Change in RMSE%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	9%	13%	12%	14%	16%	18%	28%	104
Coast Range	-1%	-1%	-1%	-1%	-1%	0%	4%	1,580
North Cascades	5%	-1%	-2%	9%	1%	1%	-1%	443
Cascades	0%	0%	0%	-1%	-2%	-2%	3%	613
Klamath Mountains	-4%	-4%	-4%	-8%	-4%	-4%	-3%	275
Eastern Cascades	2%	2%	0%	0%	1%	3%	3%	203
Northern Rockies	-19%	-19%	-9%	-25%	-36%	-54%	-2%	83
Puget Lowland	-1%	0%	1%	-4%	0%	0%	1%	122
Willamette Valley	7%	8%	14%	-5%	-8%	-10%	12%	41

Positive values indicate a decrease in prediction accuracy.

MEANWHILE, SOUTH OF THE BORDER...

Transferability: Cubic Volume

Tompalski et al. (2019)

VOL	
model	*Δbias%*
OLS	+12.3%
RF	+5.0%
kNN	+8.9%

Change in Bias%

ECOREGION	ELASTICNET	LASSO	LASSO_5	KNN	RF	GB	SVM	n
Blue Mountains	5%	9%	6%	2%	4%	3%	13%	104
Coast Range	0%	0%	1%	0%	1%	-1%	-2%	1,580
North Cascades	-1%	-6%	-8%	0%	0%	0%	-3%	443
Cascades	-1%	0%	-1%	-1%	-1%	2%	1%	613
Klamath Mountains	1%	0%	1%	-1%	2%	4%	-5%	275
Eastern Cascades	-5%	-3%	-8%	3%	1%	3%	-1%	203
Northern Rockies	20%	22%	18%	24%	22%	26%	17%	83
Puget Lowland	1%	1%	-4%	-2%	2%	-5%	-9%	122
Willamette Valley	-6%	-6%	-9%	0%	4%	2%	-16%	41

Positive values indicate a decrease in prediction accuracy.

Food for thought

- Predictive ability for regional and global models were substantially different among ecoregions.
- Global models often did not substantially affect predictive accuracy and bias, though some ecoregions showed strong shifts in either direction.
- Some global models show substantially improved predictive accuracy and reduced bias compared to regional models.
- Comparable transferability for TOPHT, QMD, and VOL, though certain ecoregions buck this trend.
- Higher number and diversity of training examples intuitively support greater generalization, particularly for machine learning algorithms, contrasting with Tompalksi et al. (2019) recommendations.
- Mode for fitting and tuning algorithm hyperparameters using cross-validation may meaningfully contribute to better transferability as well.

Thank you.

Ecotrust