FUNDAMENTAL RESEARCH

For. Sci. 60(2):253–269
http://dx.doi.org/10.5849/forsci.12-101

forest management

Imputing Forest Structure Attributes from Stand Inventory and Remotely Sensed Data in Western Oregon, USA

Andrew T. Hudak, A. Tod Haren, Nicholas L. Crookston, Robert J. Liebermann, and Janet L. Ohmann

Imputation is commonly used to assign reference stand observations to target stands based on covariate relationships to remotely sensed data to assign inventory attributes across the entire landscape. However, most remotely sensed data are collected at higher resolution than the stand inventory data often used by operational foresters. Our primary goal was to compare various aggregation strategies for modeling and mapping forest attributes summarized from stand inventory data, using predictor variables derived from either light detection and ranging (LiDAR) or Landsat and a US Geological Survey (USGS) digital terrain model (DTM). We found that LiDAR metrics produced more accurate models than models using Landsat/USGS-DTM predictors. Calculating stand-level means of all predictors or all responses proved most accurate for developing imputation models or validating imputed maps, respectively. Developing models or validating maps at the unaggregated scale of individual stand subplots proved to be very inaccurate, presumably due to poor geolocation accuracies. However, using a sample of pixels within stands proved only slightly less accurate than using all available pixels. Furthermore, bootstrap tests of similarity between imputations and observations showed no evidence of bias regardless of aggregation strategy. We conclude that pixels sampled from within stands provide sufficient information for modeling or validating stand attributes of interest to foresters.

Keywords: gradient nearest neighbor (GNN), Landsat, LandTrendr, LiDAR, most similar neighbor (MSN), *k* nearest neighbor (*k*-NN), random forest (RF)

Operational forest managers and planners are constrained to use the forest inventory data available to them, collected on the ground and remotely. Most inventory data are in the form of traditional stand inventories, which provide a reliable measure of forest stocks at the stand level. Not all stands need to be measured on the ground to provide a useful landscape-level inventory (Moeur and Stage 1995, McRoberts 2008, Hudak et al. 2012). Remotely sensed data can be used to predict stand attributes in unmeasured stands. The predictive model associates the stand structure attributes of interest, or response variables, to predictor variables derived from the remotely sensed data. The remotely sensed data most often used come from Landsat, which is free and available globally, or similar types of multispectral satellite imagery. Increasingly, light detection and ranging (LiDAR) survey data are used where available, because LiDAR's increased sensitivity to canopy structure variation usually allows for more accurate predictions than is possible from Landsat, albeit at a cost for collecting the LiDAR data.

Forestry applications using LiDAR are moving from the research to operational realms, but many operational foresters remain deterred by the usually high initial costs of collecting LiDAR data and the specialized processing tools and expertise required (Hummel et al. 2011). Even if foresters were to universally adopt LiDAR-based forest inventory methods, however, there would remain a huge amount of underutilized legacy data in the form of traditional stand inventories. Stand inventory data may lack the spatial detail afforded by LiDAR but nevertheless accurately represent forest conditions measured on the ground. Forest planners summarize stand inventory data and consider management alternatives using the Forest Vegetation Simulator (FVS) (Dixon 2002, Crookston and Dixon 2005) or similar growth-and-yield models. Such models are broadly used to fill in the temporal intervals between stand inventories,

Manuscript received August 15, 2012; accepted October 14, 2013; published online November 28, 2013.

Affiliations: *Andrew T. Hudak (ahudak@fs.fed.us), USDA Forest Service, Rocky Mountain Research Station, Moscow, ID. A. Tod Haren (tharen@odf.state.or.us), Oregon Department of Forestry, Salem, OR. Nicholas L. Crookston (ncrookston@fs.fed.us), USDA Forest Service, Rocky Mountain Research Station. Robert J. Liebermann (rjliebermann@fs.fed.us), USDA Forest Service, Rocky Mountain Research Station, Moscow, ID. Janet L. Ohmann (janet.ohmann@orst.edu), USDA Forest Service, Pacific Northwest Research Station (retired), Corvallis, OR.*

Acknowledgments: *We thank Dave Enck for assistance in interpreting the SLI protocol, Emmor Nile for obtaining 2007–2009 LiDAR data, Jim Muckenhoupt for providing 2010 LiDAR data, Justin Braaten and Robert Kennedy for providing LandTrendr data, and Jeff Hamann for productive conversations about this project. Funding for this research was provided by the Oregon Department of Forestry through Collective Agreement 11-CO-11221633-136.*

whereas imputation models are commonly used to fill in the spatial gaps between inventoried stands. It is incumbent on the forestry research community to develop modeling approaches that make good use of existing stand inventory data, growth-and-yield projections, and stand imputation methods, even as newer technologies (such as LiDAR) and tools (such as growth projection and imputation modeling software) become available.

A traditional stand inventory is composed of a number of subplots; the number is conditioned on the variability in forest conditions sampled to represent the stand. More often than not, stand inventory subplot locations are not recorded in the field with the same accuracy or precision as forest inventory plot locations designed for LiDAR-based model calibration and validation. Plots established to calibrate or validate LiDAR are by design fixed-radius plots that allow for more accurate tallying of forest attributes within a defined area, whereas stand inventory subplots may use variable-radius sampling, fixed-radius plots, or a combination of these and other sampling methods. Fixed-radius plots typically take longer to characterize than variable-radius plots, so not as many fixed-radius plots can be installed across a forested landscape for the same cost. Hummel et al. (2011) found that the accuracy and cost of LiDAR-based inventories using fixed-radius plots for training data were comparable to those of traditional stand inventories using variable-radius subplots. This cost-benefit analysis did not include other benefits of LiDAR: the availability of LiDAR over a larger area than was characterized in the stand units sampled; the higher spatial resolution of pixel-based maps derived from LiDAR compared with polygon-based stand maps; and the added benefits of LiDAR data for other applications besides forestry, such as road planning and hydrology.

Whereas modelers may have little choice in what type of data are available (LiDAR or Landsat, plot-level or stand-level), they do have more flexibility in choosing a modeling method. Regression models have been broadly used but are typically limited to univariate responses, which break associations between the independently predicted stand attributes of interest (Tomppo et al. 2008). Imputation methods are often preferred for operational use because they preserve associations between multiple stand attributes of interest in the sample units (e.g., stands) provided that $k = 1$, with k representing the number of nearest neighbors (Moeur and Stage 1995, McRoberts 2008). Nine nearest neighbor (NN) selection techniques are available in the yaImpute package (Crookston and Finley 2008) of the R statistical software suite (R Core Team 2012). Using $k = 1$, Hudak et al. (2008, 2009) found that the method based on random forest (RF) classification trees was most accurate for k-NN imputation of coniferous forest structure attributes from LiDAR metrics. Two alternative methods in package yaImpute that have been widely and successfully applied for k-NN imputation are most similar neighbor (MSN) (Moeur and Stage 1995), which uses canonical correlation analysis and gradient nearest neighbor (GNN) (Ohmann and Gregory 2002), which uses canonical correspondence analysis. MSN has often been the chosen method to impute stand structure attributes from reference stands to target stands (Moeur and Stage 1995, Temesgen et al. 2003); GNN is used for the same purpose but also to impute species-level or plant community responses along regional environmental gradients (Ohmann and Gregory 2002, Ohmann et al. 2011, Wilson et al. 2012). We view all three as alternative k-NN methods in that they use alternative methods of computing the distance between reference observations and target observations. In the simplest case where $k = 1$, they also can be referred to as three-neighbor selection methods.

The Oregon Department of Forestry (ODF) possesses stand-level inventory data across a sizeable portion of state forest lands in Tillamook District, northwestern Oregon, USA (Figure 1). ODF currently uses FVS for operational forest planning and seeks to incorporate k-NN imputation methods into its forest inventory system for the purpose of improving the accuracy, confidence, and utility of its forest inventory estimates. The goal of this study is to test alternative methods of using ODF's available forest inventory, LiDAR, Landsat, and US Geological Survey (USGS) elevation data sets. Our first objective is to compare the GNN, MSN, and RF methods of k-NN imputation and choose the most accurate method with which to proceed using LiDAR or Landsat/USGS-digital terrain model (DTM) predictors. Based on the findings of Hudak et al. (2008), we hypothesize that RF would be the most accurate neighbor selection method using either LiDAR-derived metrics or Landsat/USGS-DTM-derived variables. Our second objective is to compare the accuracies of imputation methods formulated at the level of inventory stands, inventory subplots, and a hybrid approach of "stand samples." We hypothesize that the stand-level approach will be most accurate. Our third objective is to compare three validation strategies for imputed map pixels, again at the stand, subplot, and stand sample levels. We hypothesize that aggregating all imputed pixels within reference stands will be most accurate.

Methods

Study Area

The study area (Figure 1) encompasses ODF's Tillamook District, about 102,000 ha located in northwest Oregon west of the Oregon Coast Range crest (45.06−45.81° N and 123.43−123.96° W). The area is within the Coast Range, Volcanics ecoregion (Thorson et al. 2003). Predominant soils are gravelly loam of igneous origin (Soil Survey Staff 2013). The landscape is dominated by steep, dissected slopes commonly exceeding 80%. Elevations range from near sea level to >1,000 m (Oregon State Service Center for GIS 1997). The climate is characterized by cool wet winters and relatively dry warm summers. Average daily temperature was 5° C in the winter to 15° C in the summer for the period between 1981 and 2010. For the same time period, average annual precipitation was about 200 cm at lower elevations to nearly 500 cm at higher elevations. Most precipitation occurs as rain in the winter months with some snow accumulating at the upper elevations (Wang et al. 2012).

The western edge of the study area is within the Sitka spruce (*Picea sitchensis*) vegetation zone, whereas the majority of the site is within the western hemlock (*Tsuga heterophylla*) vegetation zone (Franklin and Dyrness 1973). A series of intense fires between 1933 and 1951 burned through much of the study area. Salvage logging and rehabilitation efforts including tree planting and aerial seeding occurred through the 1970s. Consequently, the bulk of the forest is relatively young with varying levels of stocking. Predominant conifer species include Douglas-fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*), Sitka spruce (*Picea sitchensis*), western redcedar (*Thuja plicata*), and others occurring less frequently. Typical hardwood tree species include red alder (*Alnus rubra*) and bigleaf maple (*Acer macrophyllum*). Hardwoods occur throughout the forest but are concentrated along riparian zones and within large patches in low-lying areas (ODF 2009, 2010).

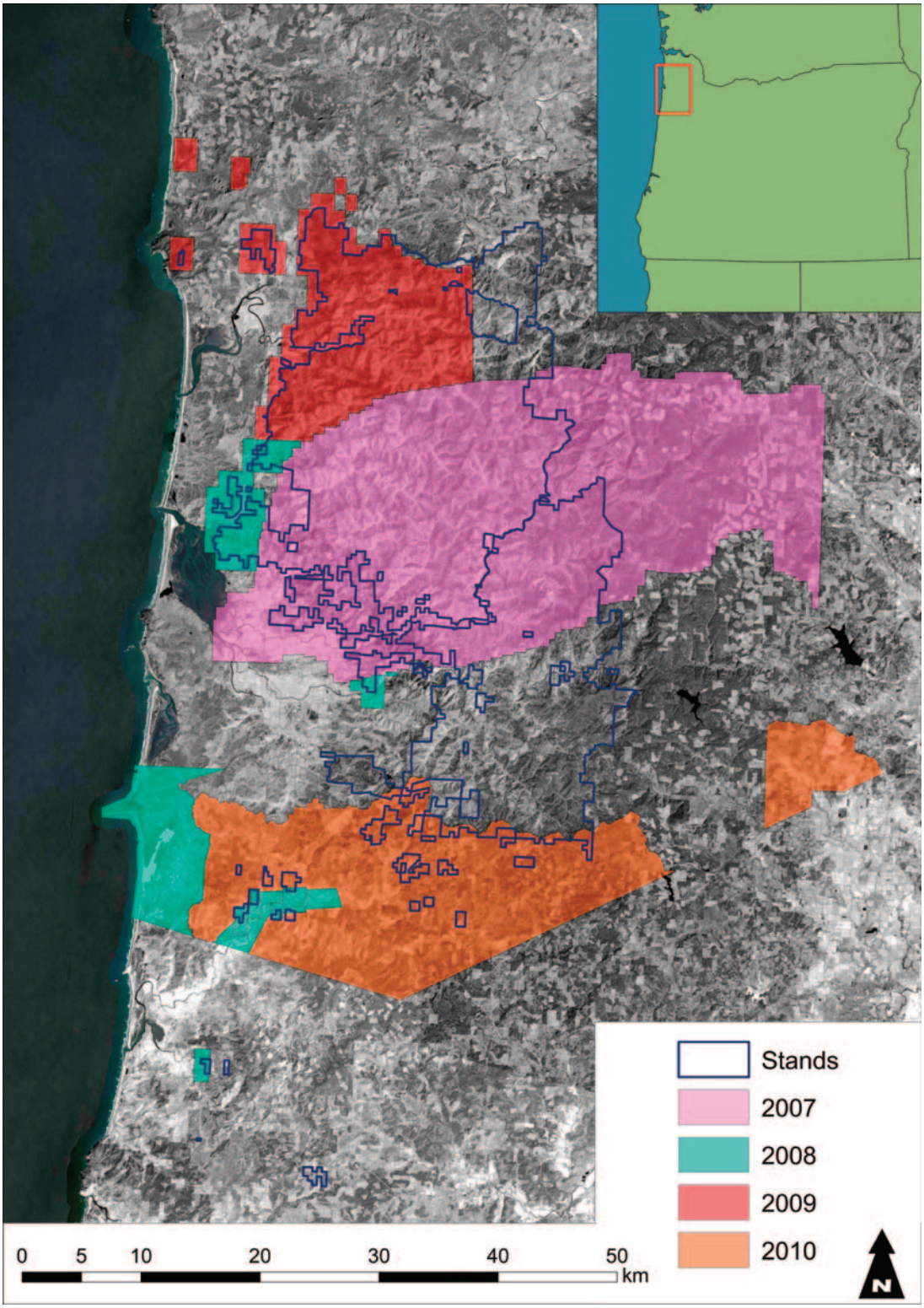

Figure 1. Location of the Tillamook study area in northwestern Oregon, USA. The southern, straightline edge of the color-shaded area is the southern boundary of Landsat TM Path/Row 47/28 containing 2007–2010 Landsat time series images processed with LandTrendr. LiDAR data south of this boundary were not included in the analysis. The background image is a 2006 multiscene LandTrendr-derived tasseled cap brightness image of broader extent.

Inventory Data

ODF's stand level inventory (SLI) was originally designed to support stratified double sampling (ODF 2004). Stands of similar vegetation and topographic characteristics were delineated and assigned a stratum identifier according to predominant tree species, size class, and estimated stocking. A line plot cruise was used to collect vegetation estimates for a random sample of stands within each stratum (Figure 2). A combination of variable and fixed-radius plots, centered on the subplot location, were used to sample vegetation data. Large live trees and snags (defined as standing

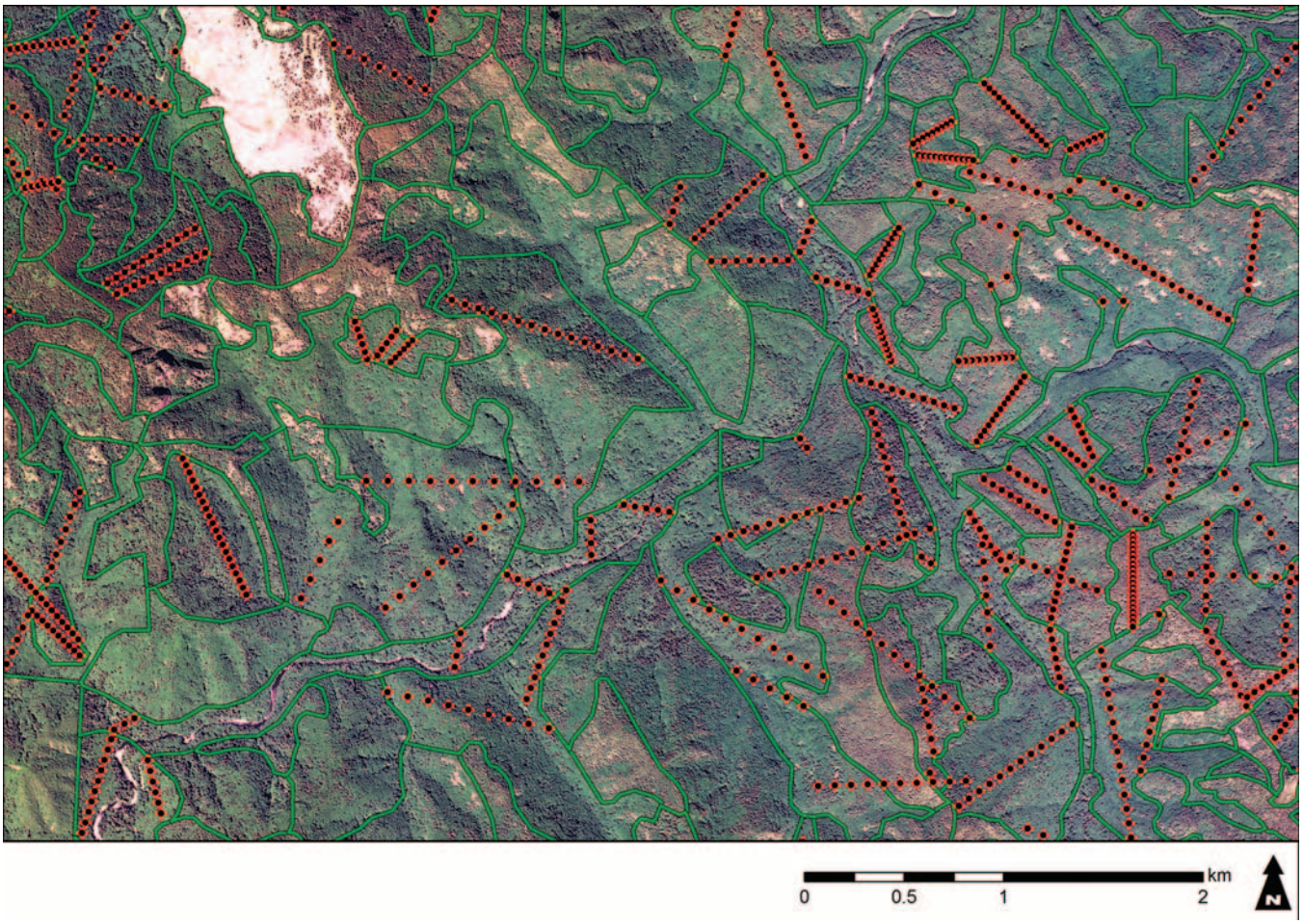

Figure 2. Close-up view of a portion of the Tillamook study area showing estimated subplot locations in reference stands. The background image is a true-color 2011 image from the National Agriculture Imagery Program.

dead trees ≥1.37 m dbh), with dbh >13.7 cm, were measured within a variable-radius plot. The basal area factor was held constant within a stand and was chosen before initiation of a cruise so that approximately 8–10 live trees would be selected per subplot. Small trees (6.1 cm < dbh ≤ 13.7 cm) were sampled within a 5.2-m fixed radius subplot. Seedlings and saplings (dbh ≤6.1 cm) were sampled within a 2.4-m fixed-radius subplot. Nontree vegetation was sampled with a 7.3-m fixed-radius subplot. Finally, a 30.5-m planar transect, originating at the subplot center location and extending in a random direction, was used to sample downed wood; downed wood pieces were tallied if >91 cm in length and >15.2 cm in diameter on either end. In some stands, snags were sampled within a fixed-radius plot, and the small tree and sapling plot radius varied. These deviations occurred on ~5% of the selected stands.

The following is a summary of data collection procedures fully described in the SLI Field Guide (ODF 2004). All data were recorded with electronic data recorders using custom software developed in-house by ODF staff. The collection software included routines for error checks and selection of subsample candidate trees consistent with the SLI Field Guide. Species, dbh, and visible damage were recorded for each tree and snag on a subplot. Height was recorded for all trees <10 cm dbh. Large tree height was estimated from a local regression curve fit to a subsample of at least three trees for each species and across the dbh range of a stand. Total height, percent live crown, and crown class were recorded for the height sample trees. Stand-level site index was estimated from a subsample of at least three dominant or codominant trees of the most common species. Site sample trees were measured for total height and an increment core was taken to estimate age at breast height. Species, length, small end diameter, diameter at intercept, and large end diameter were recorded for each downed wood piece. Decay class and wildlife use codes (Johnson 1998, ODF 2004) were recorded for all snags and downed wood pieces. Species, percent cover, and average height were recorded for nontree vegetation. Slope, aspect, elevation, plant association, slope position, and other condition codes were recorded for each subplot. All sample tallies were expanded to per acre values by the data collection software before being imported into the inventory database.

Sampling intensity was determined before data collection according to perceived uniformity within the stand, resulting in 9–25 subplots per stand. Stand boundaries were updated after harvest operations and other disturbances. Subplots were removed from the inventory data if they no longer represented the stand. The remaining subplots were assumed to represent the unaltered portion of the stand and were retained in the inventory database to support forest operation planning. Of 5,872 stands on state forest lands in Tillamook District (Figure 1), 2002–2010 inventory data were available in 1,122 (19% of the stands), 2007–2010 LiDAR data were available in 4,334 (74% of the stands), and 2007–2010 Landsat and USGS-DTM data were available in all stands. Details about

how these data sets were processed are described in the following sections.

Growth Projection

The Pacific Northwest Coast variant (Keyser 2008) of the FVS (Dixon 2002) was used to estimate stand and subplot conditions for the years of the LiDAR collection. FVS projections were made at both the stand and stand subplot levels. Subplot site variables (elevation, slope, aspect, and site index) were used to initialize the FVS projections. Averages of the subplot site values were computed for the stand-level projections. Previous work by ODF (2006) found that the species default values in FVS for maximum stand density index (SDI) and maximum dbh for red alder, resulted in unrealistic growth and mortality estimates. Therefore, the following species level modifiers were added to the FVS projections to maintain consistency with previous ODF growth-and-yield projects: a maximum SDI of 720 was used for western hemlock, 300 for red alder, and 600 for all other species; a maximum dbh of 61 cm was also specified for red alder. A projection of each stand and subplot was made for each of the LiDAR collection years with the growth cycle length set as the difference between the inventory collection year and the LiDAR collection year. Only stands sampled during or before a given LiDAR collection year could be projected. It is recognized that some bias may be incurred by using growth cycles shorter than the FVS standard of 10 years. However, no assessment or correction of this bias was performed for this study. Regeneration and ingrowth were not included in the FVS projections.

Downed wood samples collected with SLI were converted to tons per acre for input into FVS. Cubic foot volume by size class was computed for each downed wood piece using a conical frustum volume formula for all downed wood pieces. Weight was calculated by multiplying the computed volume times the specific gravity (Miles and Smith 2009) of the sample species. The specific gravity of Douglas-fir was used for samples of unknown species. No adjustment was made for decay class. Downed wood estimates from SLI were limited to size classes >15.2 cm, and FVS defaults were assumed for the smaller size classes.

FVS was used to summarize projected stand and subplot data, and the Fire and Fuels Extension (FFE) was used to generate estimates of snags, downed wood, and other forest biomass pools and their carbon loads (Rebain 2010). The nine "standard" FVS stand summary variables with broad utility in forestry were selected a priori as the response variables to impute with the exception of board feet (Table 1). Because board feet is not the same as merchantable volume and does not have an easy equivalent in metric units, total carbon from the FFE carbon report was selected instead as a more useful response variable with broad ecological utility. Only trees of >5.1-cm dbh were included in the FVS summaries of trees per hectare (TPH), basal area (BA), and quadratic mean diameter (QMD).

The 2010 inventory included the largest number of reference stands (n = 1,122), as newly inventoried stands were added to the geodatabase every year beginning in 2002. There were 856 reference stands with coincident LiDAR data collected in either 2007, 2008, 2009, or 2010. Matching the inventory growth year to the LiDAR collection year resulted in 309 reference stands matched in 2007, 59 stands matched in 2008, 241 matched in 2009, and 247 in 2010. That left 266 reference stands situated completely outside the extent of the 2007–2010 LiDAR surveys; 2010 stand summaries were matched to just the Landsat/USGS-DTM data in these cases. Some 2010 reference stands that were new to the inventory (n = 168) were matched to LiDAR (and Landsat) data collected in one of the earlier years.

Table 1. Nine forest structure attributes chosen as response variables for the study, summarized with FVS at the stand level (n = 1,122).

Response variable	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
TPH (trees/ha)	0.0	368.3	516.6	591.3	714.0	5,116.9
BA (m^2/ha)	0.0	32.5	40.4	40.7	48.0	100.8
SDI	0.0	261.0	310.0	319.7	370.8	937.0
CCF	4.0	194.0	237.0	236.1	279.0	500.0
Ht (m)	0.6	26.5	30.0	29.7	33.8	53.3
QMD (cm)	0.0	26.9	31.9	32.4	36.9	68.9
TVol (m^3/ha)	0.0	308.8	405.1	428.9	508.1	1,635.4
MVol (m^3/ha)	0.0	243.4	328.4	352.5	426.8	1,454.0
TCarb (tonnes/ha)	44.9	210.7	259.6	262.1	306.6	686.0

TPH, trees per ha; BA, basal area; SDI, stand density index; CCF, crown competition factor; Ht, height; QMD, quadratic mean diameter; TVol, total volume; MVol, merchantable volume; TCarb, total carbon.

Remotely Sensed Data

LiDAR

ODF has been collecting LiDAR data to support various agency projects and objectives in earnest since 2007. LiDAR acquisition has typically occurred opportunistically in cooperation with adjacent landowners or other federal and state agency projects. LiDAR data within the study area were collected under three separate cooperative agreements: first in 2007, second in 2008–2009, and third in 2010 (Table 2). There is a degree of overlap in the acquisition dates between the 2008 and 2009 deliveries; data within the 2009 delivery were wholly acquired during 2009, whereas portions of the 2008 delivery were acquired during both years (Watershed Sciences 2009).

LiDAR data were delivered as classified binary (.las) files projected in Oregon State Plane units of feet and the 1983 North America Datum (NAD83). Each .las file represented a square tile on the landscape. The large number of tiles dictated that they be subdivided into five batches, with the 2007 LiDAR tiles (the largest coverage) divided into two batches, whereas the 2008, 2009, and 2010 LiDAR tiles were processed as separate batches. The lastile utility of LAStools (Isenburg 2012) was used to retile the .las tiles into 1 × 1-km tiles with a 6.1-m overlap to preclude edge artifacts on subsequent interpolation of a seamless DTM from the classified ground returns, using the GridSurfaceCreate utility of FUSION (McGaughey 2012). Canopy height, intensity, density, and topographic metrics (Table 3) were calculated at 10 × 10-m resolution using the FUSION LTK batch processor. The output ASCII text files were converted to image rasters. The outer 30 m of all rasters was trimmed to eliminate edge artifacts in some metrics along the edges of the four delivery areas (Figure 1), before mosaicking of the rasterized metrics derived from the four LiDAR surveys across the entirety of the study area.

Landsat/USGS-DTM

Landsat image data (30 m) and a 10-m USGS-DTM were available across the entire study area (Figure 1). Landsat Thematic Mapper (TM) time series images in Path/Row 47/28 were obtained

Table 2. LiDAR collection parameters and quality estimates reported by the vendor.

Collection year	2007	2008	2009	2010
Customer	Puget Sound LiDAR Consortium	Oregon LiDAR Consortium	Oregon LiDAR Consortium	Oregon Department of Geology and Mineral Industries
Acquisition dates	Apr. 27, 2007–May 11, 2007	Oct. 23, 2008–June 15, 2009	Apr. 16, 2009–June 15,2009	Jan. 3, 2010–July 14, 2010
Sensor	Leica ALS50 phase II	Leica ALS60	Leica ALS60	Leica ALS60 and ALS50 phase II
Altitude (m AGL)	900	900	900	900 and 1,300
Scan angle (°)	±14	±15	±15	±14
Flight line side-lap (%)	50	60	60	50
Pulse rate (kHz)	>105	93–99	93–99	>105
First return density (points/m^2)	7.71	8.61	8.61	9.17
Ground return density (points/m^2)	0.71	0.96	0.96	1.06
Average absolute accuracy (m RMSE)	0.03	0.03	0.03	0.03

AGL, above ground level; RMSE, root mean square error.

from the USGS open Landsat archive and already had subpixel geolocation accuracy, so no further geometric processing was required. Atmospheric effects were mostly removed using the cosine-theta (COST) correction of Chavez (1996) as implemented by Schroeder et al. (2006) to a single reference image in the time series. Cloud-free pixels in the other images in the time series were then normalized to the COST image using multivariate alteration detection and calibration (MADCAL) algorithms from Canty et al. (2004). After normalization of the time series, the Landsat tasseled cap transformation as defined by Crist (1985) for reflectance data was applied, producing the brightness, greenness, and wetness tasseled cap indices for all images in the time series. Further Landsat image preprocessing details are available in Kennedy et al. (2010).

The LandTrendr image time series analysis tool is designed to quantify multiple disturbances of varying duration and magnitude (Kennedy et al. 2010). Fits of the three tasseled cap bands at an annual timestep minimize phenology and similarly transient sources of intra-annual spectral variation, resulting in a more accurate characterization of the scene than would be captured by any individual image from the time series for a given year (Kennedy et al. 2010). Which of the four sets of 2007–2010 fitted tasseled cap bands to use for a given stand was conditioned on the collection year of the 2007–2010 LiDAR, so that the Landsat tasseled cap indices used were matched to the collection year of the LiDAR where available or otherwise the 2010 Landsat tasseled cap indices. Topographic metrics were calculated from the USGS-DTM using an ERDAS Imagine add-on tool developed at the USFS Remote Sensing Applications Center (Ruefenacht 2014).

The 10-m resolution USGS-DTM topographic metrics were resampled to 30 m to match the spatial resolution of the tasseled cap indices from LandTrendr (Table 3). Single pixel values at estimated subplot locations and zonal means of pixels within stand polygons were extracted in ArcGIS.

Predictor Variable Selection

The RF method available in the yaImpute R imputation package builds a set of classification trees (a forest) for each response variable (Crookston and Finley 2008) by calling the RF of Breiman (2001) as implemented by Liaw and Wiener (2002) in the randomForest package of R. The RF method is currently unique among the neighbor selection methods (including GNN and MSN) available in yaImpute in that it takes advantage of the bootstraping process in randomForest, thus helping development of more rigorous models for two reasons. First, randomForest at each iteration extracts a 33% out-of-bag random sample of the observations with replacement, to compare against predictions. Second, randomForest at each iteration randomly samples predictor variables from the pool of candidate variables to assess the effect on model error rate. Together, these features are highly effective for preventing model overfit (Breiman 2001, Liaw and Wiener 2002). Variable importance values are generated by testing how much prediction error increases when a tested variable is purposely randomized with respect to its actual observations. The idea is that if addition of noise to the data does not increase the prediction error, then the variable must not be very important, but if the error increases, then the information carried by the variable must be important in making accurate predictions. Predictor variable importance scores are computed separately for each forest because each forest is a separate predictor for each response variable. These scores are in the unit of measure of the individual variable and are not comparable between forests. To provide for a reasonable way to rank the overall importance of variables among the forests, they are centered by subtracting their mean and scaled by dividing by the corresponding SD and reported as scaled importance scores (SISs). Boxplots of SISs calculated on the considered predictor variables and sorted on mean SIS decline asymptotically.

In this study, the mean SIS of the most important 8–10 predictors was stable between randomForest runs comprising 500 regression trees. By iteratively running the randomForest selection procedure with varying combinations of the candidate predictors and assessing SISs, we determined that 8–10 variables most effectively balanced parsimony and predictive power. Fewer variables than this tended to yield lower correlation between predicted and observed values, whereas more variables only increased correlations marginally. We manually removed highly redundant variables and reran the procedure until we were satisfied that our choice (and quantity) of predictors was appropriate. We would thus characterize our "guided" approach as a mixture of quantitative and qualitative methods. For consistency, nine predictors were selected for both the LiDAR and Landsat/USGS-DTM models to simplify comparisons between the different models and for the added appeal of symmetry with the nine chosen response variables. When LiDAR data were available, nine LiDAR metrics were selected as predictors from 75 candidate metrics. Across the whole study area, nine predictor variables composed of LandTrendr tasseled cap and USGS-DTM-derived topographic variables were selected from 21 candidate variables (Table 3).

Table 3. Candidate LiDAR metrics and Landsat/USGS-DTM variables considered for imputation modeling.

LiDAR metrics	Landsat and USGS-DTM variables
1. Height and intensity metrics	1. Landsat indices
HMIN, IMIN; height, intensity minimums	Brightness;[a] tasseled cap band 1
HP01; height 1st percentile	Greenness;[a] tasseled cap band 2
HP05; height 5th percentile	Wetness;[a] tasseled cap band 3
HP10; height 10th percentile	2. USGS-DTM variables
HP20; height 20th percentile	DTM;[a] elevation
HP25,[a] IP25; height, intensity 25th percentiles	Slope;[a] slope
HP30; height 30th percentile	Aspect, aspect
HP40; height 40th percentile	Trasp, transformed aspect solar radiation index (Roberts and Cooper 1989)
HMED, IMED; height, intensity medians	PlanCurv;[a] planiform curvature
HP60; height 60th percentile	ProfCurv;[a] profile curvature
HP70; height 70th percentile	Curv; total curvature (PlanCurv + ProfCurv)
HP75, IP75; height, intensity 75th percentiles	SlpCosAsp;[a] slope × cosine(Aspect) (Stage 1976)
HP80; height 80th percentile	SlpSinAsp;[a] slope × sine(Aspect) (Stage 1976)
HP90; height 90th percentile	RelSlpPos;[a] relative slope position
HP95; height 95th percentile	Convolve; DTM convolution derivative
HP99; height 99th percentile	FocalDiv; focal division
HMAX,[a] IMAX; height, intensity maximums	HeatLoad; head load index (McCune and Keon 2002)
HMEAN, IMEAN; height, intensity means	LandBolstad; landform Bolstad (Bolstad and Lillesand 1992)
HSTD, ISTD; height, intensity standard deviations	LandMcNab; landform McNab (McNab 1989)
HVAR, IVAR; height, intensity variances	SurfArea; surface area
HCV, ICV; height, intensity coefficients of variation	SAGAR; surface area to ground area ratio
HSKEW, ISKEW; height, intensity skewnesses	TRI; terrain ruggedness index (Riley et al. 1999)
HKURT, IKURT; height, intensity kurtoses	
HAAD, IAAD; height, intensity average absolute deviations	
HIQ, IIQ; height, intensity interquartile ranges	
HMODE, IMODE; height, intensity modes	
HL1, IL1; height, intensity 1st L-moments (Hosking 1990, Wang 1996)	
HL2, IL2; height, intensity 2nd L-moments (Hosking 1990, Wang 1996)	
HL3, IL3; height, intensity 3rd L-moments (Hosking 1990, Wang 1996)	
HL4, IL4; height, intensity 4th L-moments (Hosking 1990, Wang 1996)	
HLCV, ILCV; height, intensity L-moment coefficients of variation (Hosking 1990, Wang 1996)	
HLSKEW, ILSKEW; height, intensity L-moment skewnesses (Hosking 1990, Wang 1996)	
HLKURT, ILKURT; height, intensity L-moment kurtoses (Hosking 1990, Wang 1996)	
HMADmed; height median of absolute deviations from overall median	
HMADmode; height mode of absolute deviations from overall median	
CRR; canopy relief ratio (Pike and Wilson 1971)	
2. Density metrics	
Pct1stAboveBH; percentage 1st returns above breast height	
Pct1stAboveMean;[a] percentage 1st returns above height mean	
Pct1stAboveMode; percentage 1st returns above height mode	
PctAllAboveBH; percentage all returns above breast height	
PctAllAboveMean; percentage all returns above height mean	
PctAllAboveMode; percentage all returns above height mode	
Stratum0; percentage all returns < 0.15 m	
Stratum1;[a] percentage all returns ≥0.15 and <1.37 m	
Stratum2; Percentage all returns ≥1.37 m and <5 m	
Stratum3;[a] Percentage all returns ≥5 and <10 m	
Stratum4;[a] Percentage all returns ≥10 and <20 m	
Stratum5;[a] Percentage all returns ≥20 and <30 m	
Stratum6;[a] Percentage all returns ≥30 m	
3. Topographic metrics	
TopoElev;[a] elevation	
TopoSlope; slope	
TopoAsp; aspect	
TopoPlan; planiform curvature	
TopoProf; profile curvature	
TopoSolar; solar radiation index	

[a] Selected predictor variables.

Imputation

Neighbor Selection

The GNN, MSN, and RF neighbor selection methods in the R package yaImpute were tested for imputing the nine response variables summarized at the stand level using FVS (Table 1). For brevity, we decided a priori to use only the most accurate of the stand-level GNN, MSN, or RF models to pursue the remaining objectives of this study. Accuracy was assessed using the scaled root mean square difference (RMSD) between imputations and observations (Stage and Crookston 2007, Hudak et al. 2008), computed for a single response variable as follows

$$\text{RMSD} = \left(\sum_{i=1}^{n} (I_i - O_i)^2 / n \right)^{0.5} \quad (1)$$

where I_i is the imputed value of a variable, O_i is the observed value, and n is the number of reference observations. Scaled RMSD is computed by dividing RMSD by the SD of the variable computed

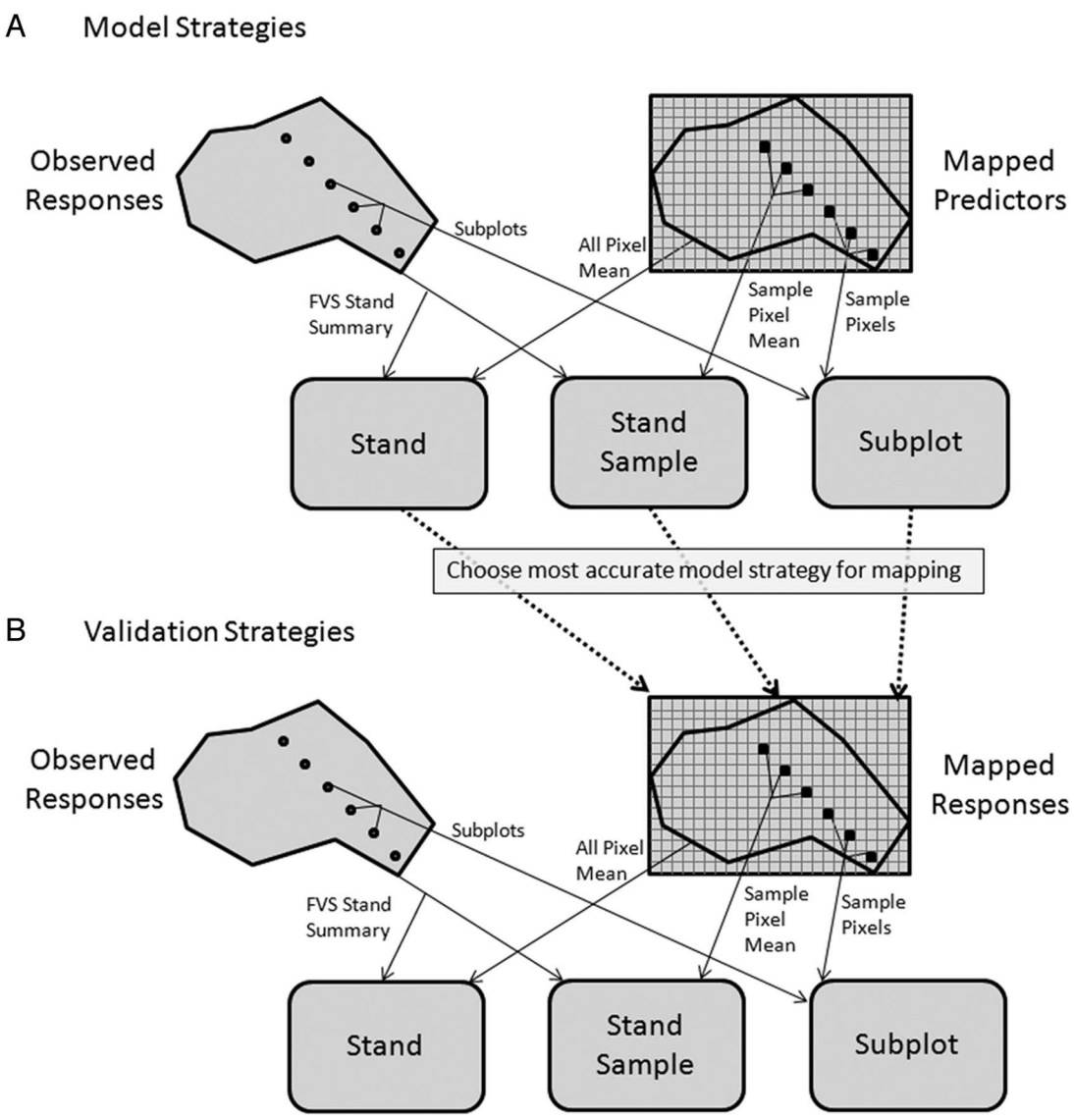

Figure 3. Schematic diagram illustrating the strategies used to impute reference data from remotely sensed data (A) and to validate imputed maps back against the reference data (B).

over the reference observations. Paired *t*-tests were used to assess whether differences in RMSD values calculated and scaled across the response variables were significant between the different stand-level models.

Model Development

Three modeling strategies were tested in this study (Figure 3A). The default strategy was the *stand* model, which was to use FVS to summarize the response variables in each stand, calculate the mean of predictor variable pixels within each stand, and model at the stand level. The *stand sample* model was to use FVS to summarize the response variables in each stand (exactly as with the stand model), calculate the stand-level mean of the predictor variable pixels sampled at estimated subplot locations, and model at the stand level. The *subplot* model was to summarize the response variables at every subplot, extract the predictor variable pixel values at estimated subplot locations, and model at the subplot level, treating the subplots as independent sample units (although the term "subplots" will continue to be used rather than "plots" because stands are, in fact, the sample units). An independent model validation was performed by splitting the full data set in half, fitting the model based on half the data, and then applying the model to the other half of the data.

Map Validation

The best LiDAR and Landsat/USGS-DTM models were used to impute the response variables at 30-m resolution with the yaImpute package. Next, three validation strategies were tested for comparing mapped response variables to reference data, which are very nearly identical to the model strategies just described (Figure 3B). The default strategy was *stand* validation, which was to calculate the mean of all imputed pixels within each stand and compare with the stand-level reference data. The *stand sample* strategy was to calculate the stand-level mean of response variable pixels sampled at the estimated subplot locations and compare these with the stand-level reference data. The *subplot* strategy was to extract imputed response variable pixel values at estimated subplot locations and compare these with the subplot-level reference data. A simple linear regression was fit for each response variable at the stand, stand sample, and subplot levels of aggregation, with observed values modeled as a function of imputed values, using the function *lm* (Chambers 1992)

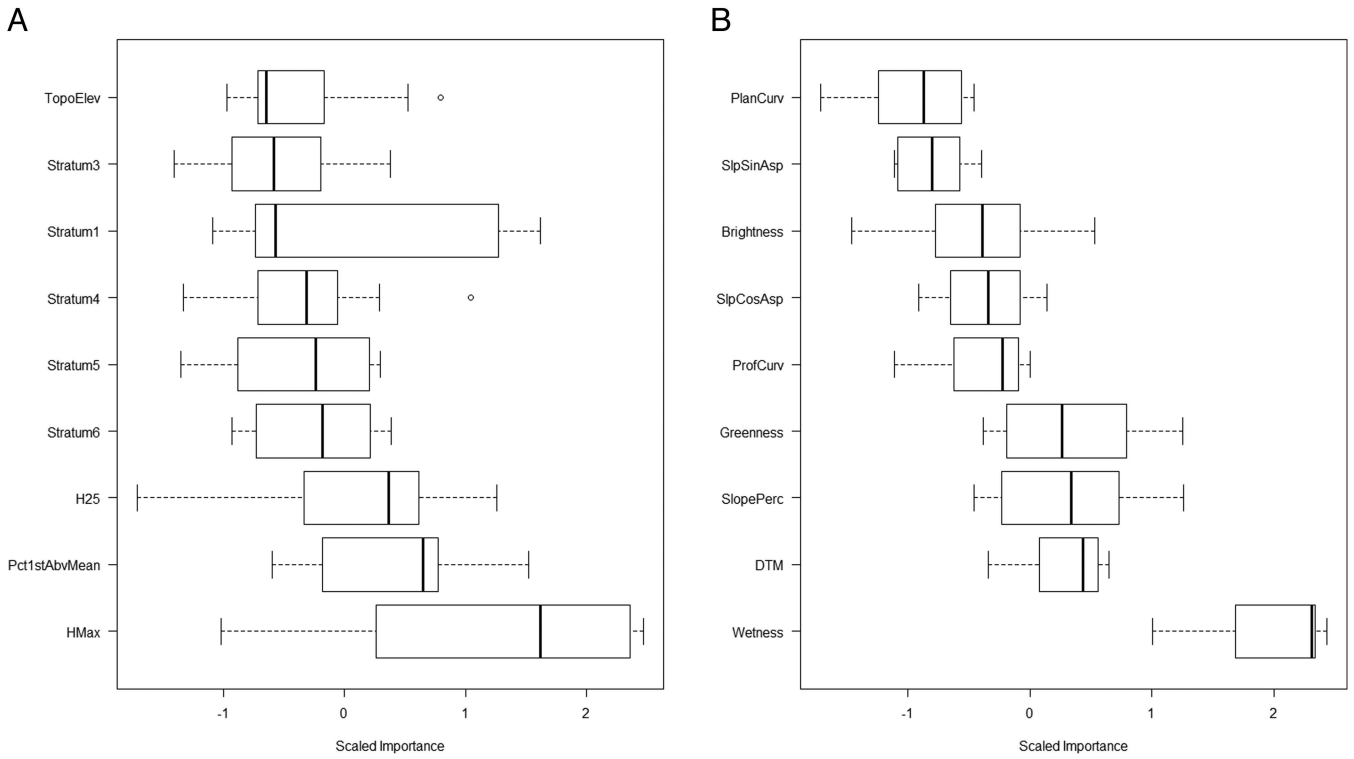

Figure 4. Scaled importance values of nine selected predictor variables (Table 3), derived from either LiDAR (A) or multitemporal Landsat and a USGS DTM (B) using RF imputation. Variables are sorted by mean scaled importance in ascending order from top to bottom.

of R. The residual standard error (RSE) for these simple linear regressions was computed as follows

$$\mathrm{RSE} = \left(\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / (n-2) \right)^{0.5} \quad (2)$$

where $\hat{y}_i$ is the predicted value, y_i is the observed value, and n is the number of observations − 2 for the 2 *df* lost to estimating the slope b_1 and intercept b_0 parameters of the simple linear model, $\hat{y}_i = b_1 y_i + b_0$.

A problem with calculating the arithmetic mean of imputed QMD pixels is that it can diverge from the quadratic mean in stands where tree stem diameters are highly variable (Curtis and Marshall 2000). Therefore, we calculated QMD as follows

$$\mathrm{QMD} = 100 \left(\left(\sum_{i=1}^{n} \mathrm{BA}_i / n \right) \Big/ \left(\sum_{i=1}^{n} \mathrm{TPH}_i / n \right) \right)^{0.5} \quad (3)$$

where BA (m^2/ha) and TPH (1/ha) are the pixel values, and n is the number of pixels (either all pixels or a sample of pixels corresponding to the number of stand subplots). This is equivalent to the formula used by FVS to calculate stand-level QMD from subplot-level BA and TPH, rather than the arithmetic mean.

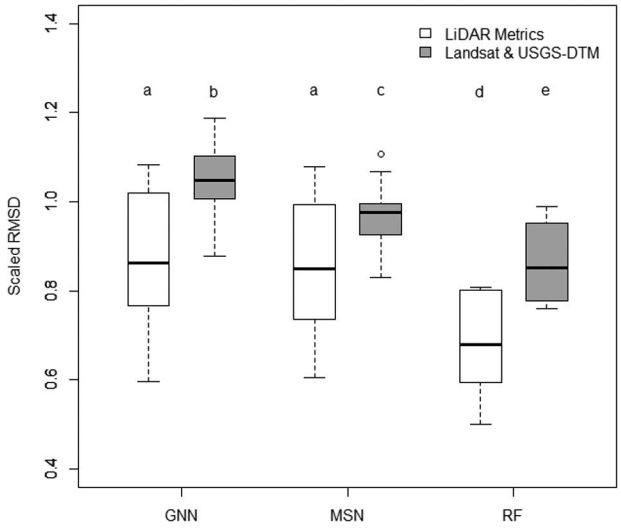

Figure 5. Scaled RMSD values calculated at reference stands from GNN, MSN, or RF imputation models developed to predict nine forest structure attributes at the stand level, using nine predictor variables derived from either LiDAR or multitemporal Landsat and a USGS DTM. Each boxplot displays scaled RMSD values calculated across the nine response variables. Different letters indicate significant differences ($\alpha = 0.05$) as determined from paired *t*-tests.

Bias and Disproportionality

Bias and disproportionality of imputations relative to reference observations were assessed using the equivalence package of R, with observations regressed on imputations in 100 bootstrap iterations of a simple linear model to test null hypotheses of dissimilarity in both the intercept and slope terms (Robinson et al. 2005, Robinson 2010). The equivalence tests of these null hypotheses of dissimilarity are based on bootstrapping to provide more rigorous evidence that the intercept does not significantly differ from 0 and the slope does not significantly differ from 1 and therefore can be considered unbiased and proportional, respectively. The half-length of the region of similarity was set to the default value of 0.25 relative units in the equivalence package, for both the intercept and slope terms, and the

Table 4. Unscaled imputation model RMSD and imputed map RSE statistics for imputing nine response variables using either LiDAR metrics or Landsat/USGS-DTM variables as predictor variables.

	Model RMSD			Map RSE		
Response variable	Stand	Stand sample	Subplot	Stand	Stand sample	Subplot
LiDAR						
n	856	847	14,055	856	847	14,312
TPH (trees/ha)	336.97	354.05	715.84	286.69	302.57	644.18
BA (m^2/ha)	9.91	9.95	23.01	9.17	9.71	20.84
SDI	88.23	87.44	181.92	76.53	79.03	164.53
CCF	56.86	57.82	142.80	51.96	55.10	126.25
Ht (m)	3.68	3.61	7.85	3.21	3.39	7.04
QMD (cm)	7.15	7.94	17.09	5.89	6.12	15.36
TVol (m^3/ha)	132.06	134.49	277.25	117.00	126.82	254.27
MVol (m^3/ha)	115.77	116.78	243.47	101.01	110.78	221.76
TCarb (tonnes/ha)	59.59	61.14	116.78	51.40	54.94	107.21
Landsat/USGS-DTM						
n	1,122	1,122	19,100	1,122	1,122	19,364
TPH (trees/ha)	387.18	447.04	777.13	290.06	316.63	635.90
BA (m^2/ha)	11.74	14.46	23.60	11.28	11.66	21.14
SDI	97.37	113.26	182.20	84.73	88.74	162.96
CCF	65.37	70.60	141.19	57.12	59.08	124.79
Ht (m)	5.39	7.28	9.05	5.07	5.47	8.35
QMD (cm)	8.73	10.04	18.95	7.00	7.55	15.69
TVol (m^3/ha)	160.47	219.78	299.41	161.19	166.60	270.57
MVol (m^3/ha)	144.29	203.13	268.95	145.51	150.28	241.02
TCarb (tonnes/ha)	70.25	88.77	121.56	64.39	66.25	109.43

Both model and map data were summarized at the stand, stand sample, and subplot levels (Figure 3). TPH, trees per ha; BA, basal area; SDI, stand density index; CCF, crown competition factor; Ht, height; QMD, quadratic mean diameter; TVol, total volume; MVol, merchantable volume; TCarb, total carbon.

default significance level of 0.05 was also used. All statistics presented in this article were calculated using R statistical software (R Core Team 2012).

Results

Variable Selection

Nine predictor variables were selected to predict the nine response variables from either LiDAR-based or Landsat/USGS-DTM-based imputation models. Based on the SIS values, HMax was the most important predictor in the LiDAR model, whereas Wetness was the most important predictor in the Landsat/USGS-DTM model (Figure 4). Because highly redundant predictors (i.e., Pearson correlation $r > 0.9$) had been excluded from both models, the highest Pearson correlation between predictors in the case of the LiDAR model was $r = 0.84$ (HMax and Stratum6), whereas in the case of the Landsat/USGS-DTM model it was $r = 0.89$ (Greenness and Brightness).

Neighbor Selection

The best neighbor selection method among the GNN, MSN, and RF techniques tested was RF, which, according to paired *t*-tests, exhibited significantly lower scaled RMSD values for the response variables than either GNN or MSN. The GNN and MSN methods performed with similar accuracy using LiDAR metrics, whereas MSN model RMSD values were significantly lower than GNN model RMSD values using Landsat/USGS-DTM predictors. Moreover, the imputation models based on LiDAR metrics were invariably more accurate than those based on Landsat/USGS-DTM variables (Figure 5). These results supported our first hypothesis that the RF method would be most accurate, so the RF method was chosen to meet the remaining modeling and mapping objectives of this study.

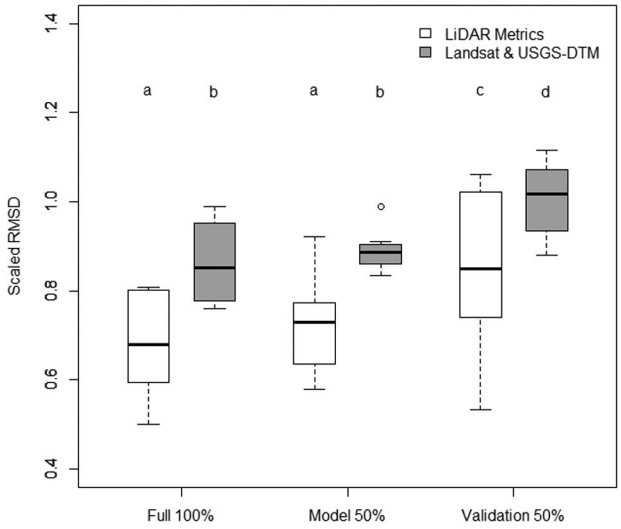

Figure 6. Scaled RMSD values calculated using RF imputation based on all available stands (same RF models as in Figure 5) versus a limited RF model based on 50% of the available stands versus the other 50% of the stands withheld to validate the limited RF model. Each boxplot displays scaled RMSD values calculated across the nine response variables. Different letters indicate significant differences ($\alpha = 0.05$) as determined from paired *t*-tests.

Model Development

Stand-level imputation of the response variables proved much more accurate in terms of the RMSD than subplot-level imputation, using either LiDAR metrics or Landsat/USGS-DTM variables (Table 4). Less markedly, the stand strategy was also more accurate than the stand sample strategy, supporting our second hypothesis. Height predicted from LiDAR was the one exception for which the stand sample strategy was slightly more accurate than the stand strategy (Table 4).

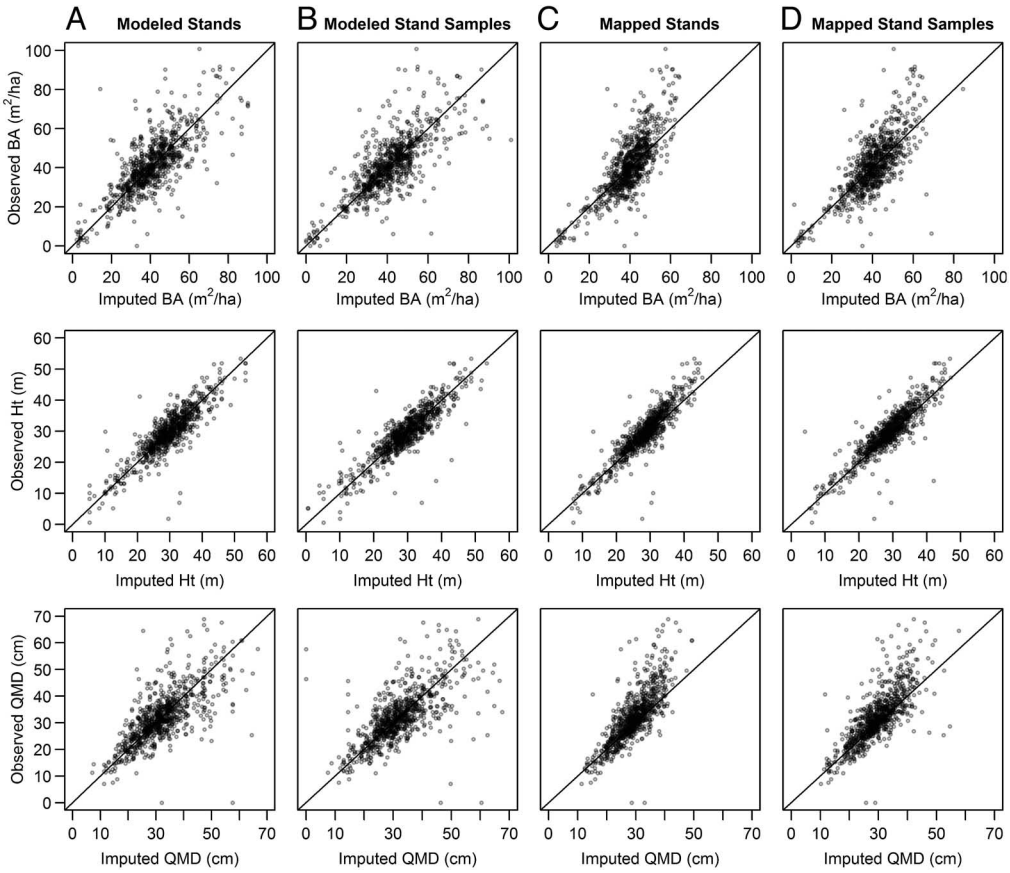

Figure 7. Imputed versus observed scatterplots of three forest structure attributes predicted from nine selected LiDAR metrics using RF to impute stands with all LiDAR metrics pixels (A), stands with a sample of LiDAR metrics pixels (B), all response pixels aggregated to the stand level (C), and a sample of response pixels aggregated to the stand level (D). The 1:1 lines are also plotted.

Model and Map Validation

The stand-level data set was split in half to validate the stand-level model beyond the systematic bootstrapping approach implemented by RF. Comparison of the scaled RMSD values between the full models, based on either all stands with LiDAR data ($n = 856$) or all stands with Landsat/USGS-DTM data ($n = 1{,}122$), versus the same model forms based on half the number of stands, revealed only a slight decrease in accuracy that was not significant, according to paired *t*-tests. However, applying the reduced models to impute values for the withheld validation stands resulted in significant increases in RMSD for the LiDAR model (mean of 15.2%) and the Landsat/USGS model (mean of 14.2%), according to paired *t*-tests (Figure 6).

Having been established as the most accurate, the stand-level models were used to map the response variables at 30-m resolution. These maps were validated at three levels of aggregation (Figure 3B); imputed map accuracies at the three aggregation levels were very consistent with the imputation model accuracies described in the preceding subsection. Stand-level validation was consistently the most accurate in terms of the RSE, supporting our third hypothesis (Table 4).

The stand-level aggregation of pixels reduced the variance of the aggregated map imputations about the mean relative to reference observations, causing the scatter plots to pull away from the 1:1 line compared with the model imputations plotted against the same reference data (Figures 7 and 8). This departure of aggregated imputations from the 1:1 line was most noticeable for the relatively poorly-predicted QMD and least noticeable for the relatively well-predicted Ht, whereas BA exhibited an intermediate level of departure from the 1:1 line.

Bias and Disproportionality

Equivalence tests demonstrated that none of the model imputations were biased for any response variable, under any modeling strategy, based on either LiDAR metrics (Table 5) or Landsat/USGS-DTM predictors (Table 6). Data from the stand and stand sample map validations were also uniformly unbiased with respect to the reference data (Tables 5 and 6).

Compared with tests for estimated bias, the tests for disproportionality produced more mixed results, although the slopes of the relationship between model imputations and observations at the stand level were consistently <1 (below the slope region of similarity). Disproportionality at the stand and stand sample levels was less pronounced if LiDAR metrics were used than if Landsat/USGS-DTM predictors were used and higher for less accurately predicted variables such as TPH, SDI, crown completion factor, and QMD than for more accurately predicted variables such as BA, height, total volume, merchantable volume, and total carbon. At the subplot level, all imputations were disproportionately below the slope region of similarity (Tables 5 and 6).

Although modeled imputations were uniformly below the slope region of similarity to varying degrees, the slopes of the relationship between aggregated map pixel imputations and observations were generally >1, with more aggregated imputations within or slightly

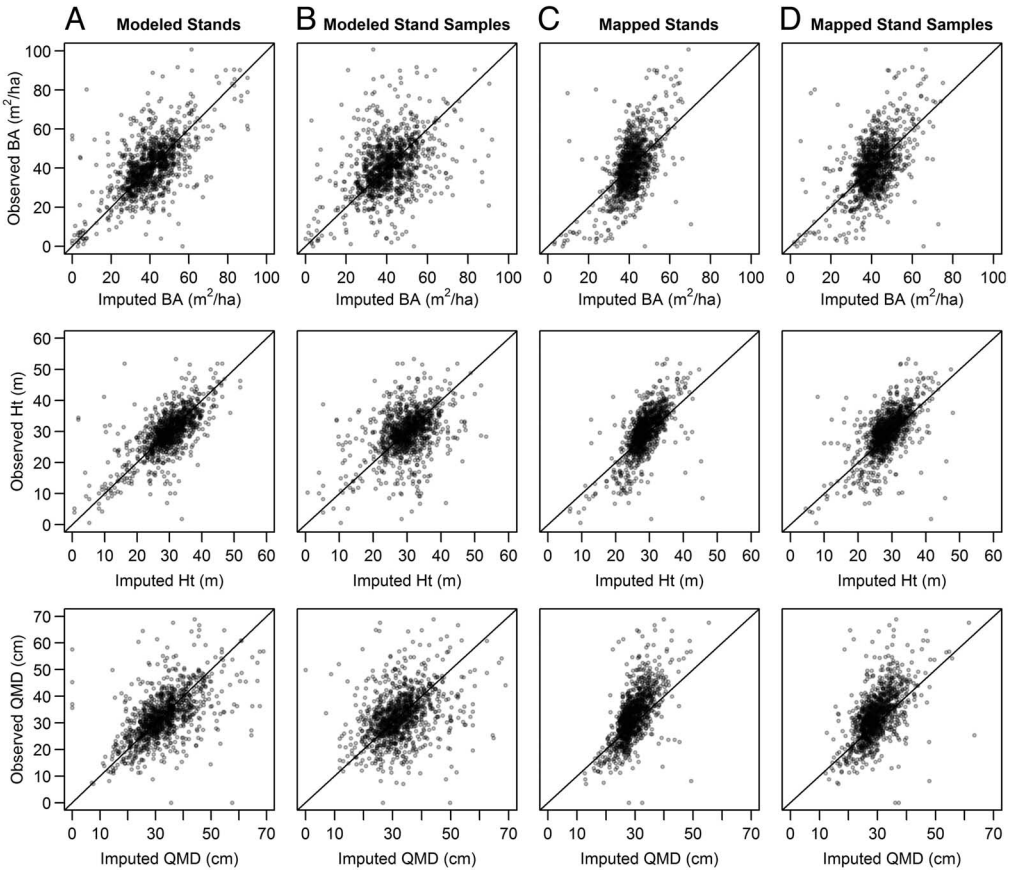

Figure 8. Imputed versus observed scatterplots of three forest structure attributes predicted from nine selected Landsat/USGS-DTM predictors using RF to impute stands with all Landsat/USGS-DTM pixels (A), stands with a sample of Landsat/USGS-DTM pixels (B), all response pixels aggregated to the stand level (C), and a sample of response pixels aggregated to the stand level (D). The 1:1 lines are also plotted.

above the slope region of similarity. This was an unanticipated, beneficial effect of aggregation whether the LiDAR-derived maps or the Landsat/USGS-DTM-derived maps were used. At the subplot level of validation, there was no aggregation of imputed pixels, so the relationships maintained their disproportionality below the slope region of similarity. Predictions below the slope region of similarity equate to underprediction of stands at the high end of the response variable gradient and overprediction at the low end, which is analogous to regression to the mean (Robinson et al. 2005). Only TPH predicted from LiDAR and Landsat/USGS-DTM and aggregated at the stand level exhibited the opposite condition, for which a majority of predictions fall above the slope region of similarity; i.e., overprediction of stands at the high end and underprediction at the low end.

Discussion

The RF method of neighbor selection proved more accurate than the GNN or MSN methods, corroborating the findings of Hudak et al. (2008, 2009). The bootstrapping nature of RF may be the best explanation for its flexibility and utility. The MSN method was developed to impute stand structure attributes (Moeur and Stage 1995), yet RMSD errors of MSN were not significantly lower than those from GNN (Figure 5), which was designed to impute species-level or plant community-type responses along environmental gradients (Ohmann and Gregory 2002, Ohmann et al. 2011). More generally, the accuracy differences between these methods are probably less than the accuracy to be gained by setting $k > 1$ (Muinonen et al. 2001, McRoberts et al. 2002). A consequence of using $k > 1$ nearest neighbors is that it reduces the variance in imputations relative to that in observations (Franco-Lopez et al. 2001), although from a practical standpoint, aggregating imputed pixels to the stand level had a similar smoothing effect (i.e., compare mapped stands to modeled stands) (Figures 7 and 8). Had any estimated biases been significant in this study (Tables 5 and 6), the yaImpute package does offer a bias correction option to select other nearest neighbors that minimize the estimated bias from among k alternative nearest neighbors. A similar strategy might one day be developed to reduce disproportionality. The default settings in the equivalence package served our purposes to test for significant bias and disproportionality, but we advise the user to consider changing these significance thresholds if a stricter or looser interpretation would fit their particular application more practically.

The RMSD and RSE statistics (Table 4) used to assess modeled and mapped imputation accuracies, respectively, are reported in the unscaled units of the response variables. The two statistics are not exactly comparable, because the RSE from a regression model will tend to be less than the RMSD from an imputation model even if based on the same data, because the variance of regression predictions is lower than that of imputations based on a single nearest neighbor (McRoberts et al. 2002, Tuominen et al. 2003, Stage and Crookston 2007, Eskelson et al. 2009). Aggregating the map pixels within stands averaged out much of the variability, thus decreasing

Table 5. Regression-based equivalence tests of estimated bias (intercept) and disproportionality (slope), generated from 100 bootstraps of observations regressed on imputations at the stand, stand sample, and subplot levels, as predicted from LiDAR metrics.

	Proportion of simulation in intercept region of similarity				Proportion of simulation in slope region of similarity			
Response variable	Below	Within	Above	Decision	Below	Within	Above	Decision
LiDAR model imputations								
Stands (n = 856)								
TPH (trees/ha)	0	1	0	Similar	0.54	0.46	0	Weakly dissimilar
BA (m^2/ha)	0	1	0	Similar	0.03	0.97	0	Similar
SDI	0	1	0	Similar	0.81	0.19	0	Weakly dissimilar
CCF	0	1	0	Similar	0.86	0.14	0	Weakly dissimilar
Ht (m)	0	1	0	Similar	0	1	0	Similar
QMD (cm)	0	1	0	Similar	0.66	0.34	0	Weakly dissimilar
TVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
MVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
TCarb (tonnes/ha)	0	1	0	Similar	0.06	0.94	0	Weakly similar
Stand samples (n = 847)								
TPH (trees/ha)	0	1	0	Similar	0.71	0.29	0	Weakly dissimilar
BA (m^2/ha)	0	1	0	Similar	0.02	0.98	0	Similar
SDI	0	1	0	Similar	0.88	0.12	0	Weakly dissimilar
CCF	0	1	0	Similar	0.95	0.05	0	Weakly dissimilar
Ht (m)	0	1	0	Similar	0	1	0	Similar
QMD (cm)	0	1	0	Similar	0.98	0.02	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
MVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
TCarb (tonnes/ha)	0	1	0	Similar	0.09	0.91	0	Weakly similar
Subplots (n = 14,055)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	1	0	0	Dissimilar
QMD (cm)	0	1	0	Similar	1	0	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
MVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
TCarb (tonnes/ha)	0	1	0	Similar	1	0	0	Dissimilar
LiDAR map imputations								
Stands (n = 856)								
TPH (trees/ha)	0	1	0	Similar	0	0.34	0.66	Weakly dissimilar
BA (m^2/ha)	0	1	0	Similar	0	0.97	0.03	Weakly similar
SDI	0	1	0	Similar	0	1	0	Similar
CCF	0	1	0	Similar	0	1	0	Similar
Ht (m)	0	1	0	Similar	0	1	0	Similar
QMD (cm)	0	1	0	Similar	0	0.67	0.33	Weakly similar
TVol (m^3/ha)	0	1	0	Similar	0	0.94	0.06	Weakly similar
MVol (m^3/ha)	0	1	0	Similar	0	0.98	0.02	Similar
TCarb (tonnes/ha)	0	1	0	Similar	0	0.65	0.35	Weakly similar
Stand samples (n = 847)								
TPH (trees/ha)	0	1	0	Similar	0	0.93	0.07	Weakly similar
BA (m^2/ha)	0	1	0	Similar	0	1	0	Similar
SDI	0	1	0	Similar	0	1	0	Similar
CCF	0	1	0	Similar	0	1	0	Similar
Ht (m)	0	1	0	Similar	0	1	0	Similar
QMD (cm)	0	1	0	Similar	0	1	0	Similar
TVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
MVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
TCarb (tonnes/ha)	0	1	0	Similar	0	1	0	Similar
Subplots (n = 14,312)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	0.44	0.56	0	Weakly similar
QMD (cm)	0	1	0	Similar	0.07	0.93	0	Weakly similar
TVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
MVol (m^3/ha)	0	1	0	Similar	0.99	0.01	0	Dissimilar
TCarb (tonnes/ha)	0	1	0	Similar	1	0	0	Dissimilar

TPH, trees per hectare; BA, basal area; SDI, stand density index; CCF, crown competition factor; Ht, height; QMD, quadratic mean diameter; TVol, total volume; MVol, merchantable volume; TCarb, total carbon.

Table 6. Regression-based equivalence tests of estimated bias (intercept) and disproportionality (slope), generated from 100 bootstraps of observations regressed on imputations at the stand, stand sample, and subplot levels, as predicted from Landsat/USGS-DTM variables.

Response variable	Proportion of simulation in intercept region of similarity				Proportion of simulation in slope region of similarity			
	Below	Within	Above	Decision	Below	Within	Above	Decision
Landsat/USGS-DTM model imputations								
Stands (n = 1,122)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	0.86	0.14	0	Weakly dissimilar
QMD (cm)	0	1	0	Similar	1	0	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	0.5	0.5	0	Weakly similar
MVol (m^3/ha)	0	1	0	Similar	0.26	0.74	0	Weakly similar
TCarb (tonnes/ha)	0	1	0	Similar	0.99	0.01	0	Weakly dissimilar
Stand samples (n = 1,122)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	1	0	0	Dissimilar
QMD (cm)	0	1	0	Similar	1	0	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
MVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
TCarb (tonnes/ha)	0	1	0	Similar	1	0	0	Dissimilar
Subplots (n = 19,100)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	1	0	0	Dissimilar
QMD (cm)	0	1	0	Similar	1	0	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
MVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
TCarb (tonnes/ha)	0	1	0	Similar	1	0	0	Dissimilar
Landsat/USGS-DTM map imputations								
Stands (n = 1,122)								
TPH (trees/ha)	0	1	0	Similar	0	0.18	0.82	Weakly dissimilar
BA (m^2/ha)	0	1	0	Similar	0	0.99	0.01	Similar
SDI	0	1	0	Similar	0	0.98	0.02	Similar
CCF	0	1	0	Similar	0	0.89	0.11	Weakly similar
Ht (m)	0	1	0	Similar	0	0.97	0.03	Similar
QMD (cm)	0	1	0	Similar	0	0.49	0.51	Weakly dissimilar
TVol (m^3/ha)	0	1	0	Similar	0	1	0	Similar
MVol (m^3/ha)	0	1	0	Similar	0	0.99	0.01	Similar
TCarb (tonnes/ha)	0	1	0	Similar	0	0.99	0.01	Similar
Stand samples (n = 1,122)								
TPH (trees/ha)	0	1	0	Similar	0	0.99	0.01	Similar
BA (m^2/ha)	0	1	0	Similar	0.05	0.95	0	Weakly similar
SDI	0	1	0	Similar	0.19	0.81	0	Weakly similar
CCF	0	1	0	Similar	0.14	0.86	0	Weakly similar
Ht (m)	0	1	0	Similar	0.01	0.99	0	Similar
QMD (cm)	0	1	0	Similar	0	1	0	Similar
TVol (m^3/ha)	0	1	0	Similar	0.07	0.93	0	Weakly similar
MVol (m^3/ha)	0	1	0	Similar	0.05	0.95	0	Weakly similar
TCarb (tonnes/ha)	0	1	0	Similar	0	1	0	Similar
Subplots (n = 19,364)								
TPH (trees/ha)	0	1	0	Similar	1	0	0	Dissimilar
BA (m^2/ha)	0	1	0	Similar	1	0	0	Dissimilar
SDI	0	1	0	Similar	1	0	0	Dissimilar
CCF	0	1	0	Similar	1	0	0	Dissimilar
Ht (m)	0	1	0	Similar	1	0	0	Dissimilar
QMD (cm)	0	1	0	Similar	1	0	0	Dissimilar
TVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
MVol (m^3/ha)	0	1	0	Similar	1	0	0	Dissimilar
TCarb (tonnes/ha)	0	1	0	Similar	1	0	0	Dissimilar

TPH, trees per hectare; BA, basal area; SDI, stand density index; CCF, crown competition factor; Ht, height; QMD, quadratic mean diameter; TVol, total volume; MVol, merchantable volume; TCarb, total carbon.

the map RSE relative to the model RMSD and making the accuracy of the imputed maps almost always higher than that of the imputed models, the only exceptions being stand-level total volume and merchantable volume predicted from Landsat/USGS variables (Table 4).

The test for dissimilarity in the slope term should not be confused with poor sensitivity of the predictors to the desired response. This behavior is apparent whether one uses LiDAR predictors or Landsat/USGS-DTM predictors (Figures 7 and 8), with the latter predictors exhibiting more scatter in the relationships. Another beneficial effect of aggregating the imputed map pixels was better distribution of the aggregated map imputations within the slope region of similarity, resulting in generally improved proportionality compared with the model imputations (Tables 5 and 6). This improved proportionality is most noticeable at the stand sample level of aggregation, less so at the stand level, and, of course, not at all at the unaggregated subplot level.

As expected, the stand-level models that made full use of the available remotely sensed data proved to be most accurate. We ascribe the very poor performance of the subplot-level models to the poor geolocation accuracy of the estimated subplot locations. However, the stand sample models produced imputation results nearly as accurate as those for the stand models (Table 4). That the stand sample models used only a small sample of pixels in proportion to the number of subplots, rather than all pixels within stand polygons, provides a strong argument for satellite-borne LiDAR systems that sample rather than survey the landscape (Hudak et al. 2002, Lefsky et al. 2011). LiDAR samples of canopy height systematically arranged across a landscape would be much less expensive to process than a comparable number and distribution of field plots; this would reduce reliance on relatively expensive ground data collection.

We found that subplot positions within stands may not need to be accurately geolocated to generate reasonable relationships to remotely sensed data, if aggregated at the stand level. If the subplots together well represent the structural variability within the stand, as intended in a stand inventory, then it follows that an equal number of sample pixels should also well represent the conditions throughout the stand. A random selection of pixels with no regard to subplot locations should result in poorer relationships at the stand level. Conversely, accurate colocation of these stand subplots and sample pixels should serve to strengthen the relationship and, if sufficiently accurate, enable useful subplot-level models, as we were unable to develop with this data set.

Our results demonstrate that stand-level data can be used to impute 30-m-resolution maps across spatial extents of high relevance to forest managers. Withholding 50% of the reference stands (Figure 6) for independent validation provided a more realistic estimate of the reduced accuracy (−15%) ODF managers can expect imputing to target stands, which at Tillamook District represents the majority of stands ($n = 4{,}750$), as in most managed forests. One could argue for using Forest Inventory and Analysis plot data collected on a systematic national grid (Riemann et al. 2010). However, problems with using Forest Inventory and Analysis data include the plot spatial sampling frequency being too sparse to provide adequate reference data for a typical project area extent and the need for accurate plot locations.

The Landsat/USGS-DTM-based models and maps were generally inferior to those based on LiDAR, corroborating previous studies that compared Landsat to LiDAR for predicting forest structure attributes (e.g., Lefsky et al. 2001, Hudak et al. 2006). The reason is that LiDAR is less sensitive to structure attributes such as stem diameter (QMD) and density (TPH) that vary in the horizontal domain than to more height-driven structure attributes (height, BA, total volume, and merchantable volume) that vary in the vertical domain. It is worth mentioning that we resampled the 30-m LandTrendr tasseled cap bands to 10 m and calculated 10-m metrics from the LiDAR and the 10-m USGS-DTM, which allowed us to generate response variable maps at 10-m resolution and validate these using the same three aggregation strategies (Figure 3). Although we omitted these results for brevity, we can say that the relationships to the reference data showed the same trends as aggregating the 30-m maps, but errors were slightly higher (increase in RMSD of 1–2% per response variable), which we attribute to the higher initial variability in the 10-m maps than in the 30-m maps, before aggregation. Our finding that traditional stand inventory data may be imputed from LiDAR or other remotely sensed data at higher resolution than the reference data still holds, but a conservative approach argues for mapping at the lowest resolution of the predictor variables, which in this case were the 30-m LandTrendr tasseled cap bands. Furthermore, given the geolocation uncertainties of the estimated subplot locations in this analysis, relating 10-m pixel data to the subplot-level attributes following our stand sample strategy could be more problematic.

Conclusion

As hypothesized, we found RF to be a more accurate neighbor selection method than either GNN or MSN. We also supported our hypotheses that stand-level model and map validation strategies that made use of all available remotely sensed data would be most accurate. A subplot-level strategy suffered from subplot geolocation uncertainties and did not benefit from aggregation to average out variability within stands. On the other hand, a stand sample strategy of using only a sample of pixels from estimated subplot locations within stands produced accuracies comparable to those with the stand strategy. Moreover, the stand sample strategy of data aggregation may improve the proportionality of predictions relative to observations, by basing each aggregation on similar sampling intensities. We conclude that a sample of pixels with size similar to that of inventory stand subplots may be sufficient to characterize structure variability within a stand remotely, just as subplots are considered sufficient to characterize structure variability on the ground. We recommend that forest researchers and managers not forego using traditional stand inventory data to model and map structural attributes of interest from LiDAR, Landsat, and probably other remotely sensed data.

Literature Cited

BOLSTAD, P.V., AND T.M. LILLESAND. 1992. Improved classification of forest vegetation in northern Wisconsin through a rule-based combination of soils, terrain, and Landsat TM data. *For. Sci.* 38(1):5–20.

BREIMAN, L. 2001. Random forests. *Machine Learning* 45:5–32.

CANTY, M.J., A.A. NIELSEN, AND M. SCHMIDT. 2004. Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sens. Environ.* 91:441–451.

CHAMBERS, J.M. 1992. Linear models. Chapter 4 of *Statistical models in S*, Chambers, J.M., and T.J. Hastie (eds.). Wadsworth & Brooks/Cole, Pacific Grove, CA. 608 p.

CHAVEZ, P.S. JR. 1996. Image-based atmospheric corrections—Revisited and improved. *Photogramm. Eng. Remote Sens.* 62:1025–1036.

CRIST, E.P. 1985. A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* 17:301–306.

CROOKSTON, N.L., AND G.E. DIXON. 2005. The Forest Vegetation Simulator: A review of its structure, content, and applications. *Comput. Electron. Agr.* 49:60–80.

CROOKSTON, N.L., AND A. FINLEY. 2008. yaImpute: An R package for k-NN imputation. *J. Stat. Softw.* 23:1–16.

CURTIS, R.O., AND D.D. MARSHALL. 2000. Why quadratic mean diameter? *West. J. Appl. For.* 15:137–139.

DIXON, G.E. (COMP.). 2002 (revised Aug. 10, 2011). *Essential FVS: A user's guide to the Forest Vegetation Simulator.* USDA For. Serv., Internal Rep., Forest Management Service Center, Fort Collins, CO. 204 p.

ESKELSON, B.N.I., H. TEMESGEN, V. LEMAY, T.M. BARRETT, N.L. CROOKSTON, AND A.T. HUDAK. 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. *Scand. J. For. Res.* 24: 235–246.

FRANKLIN, J.F., AND C.T. DYRNESS. 1973. *Natural vegetation of Oregon and Washington*. Oregon State University Press, Corvallis, OR. 468 p.

FRANCO-LOPEZ, H., A.R. EK, AND M.E. BAUER. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* 77:251–274.

HOSKING, J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. B.* 52(1):105–124.

HUDAK, A.T., N.L. CROOKSTON, J.S. EVANS, M.J. FALKOWSKI, A.M.S. SMITH, P.E. GESSLER, AND P. MORGAN. 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral data. *Can. J. Remote Sens.* 32:126–138.

HUDAK, A.T., N.L. CROOKSTON, J.S. EVANS, D.E. HALL, AND M.J. FALKOWSKI. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens. Environ.* 112(5):2232–2245.

HUDAK, A.T., N.L. CROOKSTON, J.S. EVANS, D.E. HALL, AND M.J. FALKOWSKI. 2009. Corrigendum to "Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data" *Remote Sens. Environ.* 113(1):289–290.

HUDAK, A.T., M.A. LEFSKY, W.B. COHEN, AND M. BERTERRETCHE. 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sens. Environ.* 82:397–416.

HUDAK, A.T., E.K. STRAND, L.A. VIERLING, J.C. BYRNE, J. EITEL, S. MARTINUZZI, AND M.J. FALKOWSKI. 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sens. Environ.* 123:25–40.

HUMMEL, S., A.T. HUDAK, E.H. UEBLER, M.J. FALKOWSKI, AND K.A. MEGOWN. 2011. A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. *J. For.* 109:267–273.

ISENBURG, M. 2012. *LAStools—Efficient tools for LiDAR processing.* Available online at lastools.org; last accessed Dec. 13, 2011.

JOHNSON, M.D. 1998. *Field procedures for the current vegetation system*, version 2.02b. USDA For. Serv., Region 6 Inventory and Monitoring System, Pacific Northwest Region, Portland, OR. 152 p.

KENNEDY, R.E., Z. YANG, AND W.B. COHEN. 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sens. Environ.* 114:2897–2910.

KEYSER, C.E. 2008 (revised May 9, 2012). *Pacific Northwest Coast (PN) variant overview—Forest Vegetation Simulator.* USDA For. Serv., Internal Rep., Forest Management Service Center, Fort Collins, CO. 49 p.

LEFSKY, M.A., W.B. COHEN, AND T.A. SPIES. 2001. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Can. J. For. Res.* 31:78–87.

LEFSKY, M.A., T. RAMOND, AND C.S. WEIMER. 2011. Alternate spatial sampling approaches for ecosystem structure inventory using spaceborne lidar. *Remote Sens. Environ.* 115:1361–1368.

LIAW, A., AND M. WIENER. 2002. Classification and regression by randomForest. *R News* 2:18–22.

MCGAUGHEY, R.J. 2012. *FUSION/LDV: Software for LIDAR data analysis and visualization*, version 3.10. USDA For. Serv., Pacific Northwest Research Station, Portland, OR. 170 p.

MCNAB, H.W. 1989. Terrain shape index: Quantifying effect of minor landforms on tree height. *For. Sci.* 35(1):91–104.

MCROBERTS, R.E. 2008. Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. *Remote Sens. Environ.* 112:2212–2221.

MCROBERTS, R.E., M.D. NELSON, AND D.G. WENDT. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. *Remote Sens. Environ.* 82:457–468.

MILES, P.D., AND W.B. SMITH. 2009. *Specific gravity and other properties of wood and bark for 156 tree species found in North America*. USDA For. Serv., Res. Note NRS-38, Newtown Square, PA. 35 p.

MOEUR, M., AND A.R. STAGE. 1995. Most similar neighbor: An improved sampling inference procedure for natural resource planning. *For. Sci.* 41:337–359.

MUINONEN, E., M. MALTAMO, H. HYPPANEN, AND V. VAINIKAINEN. 2001. Forest stand characteristics estimation using a most similar neighbor approach and image spatial structure information. *Remote Sens. Environ.* 78:223–228.

OHMANN, J.L., AND M.J. GREGORY. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA. *Can. J. For. Res.* 32:725–741.

OHMANN, J.L., M.J. GREGORY, E.B. HENDERSON, AND H.M. ROBERTS. 2011. Mapping gradients of community composition with nearest-neighbour imputation: Extending plot data for landscape analysis. *J. Veg. Sci.* 22:660–676.

OREGON DEPARTMENT OF FORESTRY. 2004. *Stand level inventory field guide*. Oregon Department of Forestry, Salem, OR. 180 p.

OREGON DEPARTMENT OF FORESTRY. 2006. *Harvest and habitat model final report.* Oregon Department of Forestry, Salem, OR. 145 p.

OREGON DEPARTMENT OF FORESTRY. 2009. *Tillamook district implementation plan*. Oregon Department of Forestry, Salem, OR. 86 p.

OREGON DEPARTMENT OF FORESTRY. 2010. *Northwest Oregon state forests management plan.* Oregon Department of Forestry, Salem, OR. 581 p.

OREGON STATE SERVICE CENTER FOR GIS. 1997. *7.5 minute digital elevation data.* Oregon State Service Center, Salem, OR. 14 p.

PIKE, R.J., AND S.E. WILSON. 1971. Elevation relief ratio, hypsometric integral, and geomorphic area altitude analysis. *Geol. Soc. Am. Bull.* 82:1079–1084.

R CORE TEAM. 2012. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available online at www.Rproject.org; last accessed Mar. 4, 2013.

REBAIN, S.A. (COMP.). 2010 (revised Mar. 20, 2012). *The fire and fuels extension to the Forest Vegetation Simulator: Updated model documentation.* USDA For. Serv., Internal Rep., Forest Management Service Center, Fort Collins, CO. 397 p.

RIEMANN, R., B.T. WILSON, A. LISTER, AND S. PARKS. 2010. An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. *Remote Sens. Environ.* 114:2337–2352.

RILEY, S.J., S.D. DEGLORIA, AND R. ELLIOT. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermount. J. Sci.* 5(1–4): 1999.

ROBERTS, D.W., AND S.V COOPER. 1989. Concepts and techniques of vegetation mapping. P. 90–96 in *Land classifications based on vegetation: Applications for resource management*. USDA For. Serv., GTRINT-257, Intermountain Research Station, Ogden, UT.

ROBINSON, A.P. 2010. *Equivalence: Provides tests and graphics for assessing tests of equivalence.* R Foundation for Statistical Computing, R package version 0.5.6, Vienna, Austria. Available online at CRAN.R-project.org/package=equivalence; last accessed Mar. 16, 2013.

ROBINSON, A.P., R.A. DUURSMA, AND J.D. MARSHALL. 2005. A regression-based equivalence test for model validation: Shifting the burden of proof. *Tree Phys.* 25:903–913.

RUEFENACHT, B. 2014. *Digital elevation model derivatives*. USDA For. Serv., Remote Sensing Applications Center, Salt Lake City, UT. Available online at www.fs.fed.us/eng/rsac/; last accessed June 7, 2012.

SCHROEDER, T.A., W.B. COHEN, C. SONG, M.J. CANTY, AND Z. YANG. 2006. Radiometric calibration of Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sens. Environ.* 103:16–26.

SOIL SURVEY STAFF. 2013. *Web oil survey.* US Department of Agriculture Natural Resources Conservation Service. Available online at websoilsurvey.nrcs.usda.gov/; last accessed Jan. 16, 2013.

STAGE, A.R. 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. *For. Sci.* 22(3):457–460.

STAGE, A.R., AND N.L. CROOKSTON. 2007. Partitioning error components for accuracy assessment of near-neighbor methods of imputation. *For. Sci.* 53:62–72.

TEMESGEN, H., V.M. LEMAY, K.L. FROESE, AND P.L. MARSHALL. 2003. Imputing tree-lists from aerial attributes for complex stands of southeastern British Columbia. *For. Ecol. Manage.* 177:277–285.

THORSON, T.D., S.A. BRYCE, D.A. LAMMERS, A.J. WOODS, J.M. OMERNIK, J. KAGAN, D.E. PATER, AND J.A. COMSTOCK. 2003. *Ecoregions of Oregon* [color poster with map, descriptive text, summary tables, and photographs; map scale 1:1,500,000]. US Geological Survey, Reston, VA.

TOMPPO, E., H. OLSSON, G. STÅHL, M. NILSSON, O. HAGNER, AND M. KATILA. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* 112: 1982–1999.

TUOMINEN, S., S. FISH, AND S. POSO. 2003. Combining remote sensing, data from earlier inventories, and geostatistical interpolation in multisource forest inventory. *Can. J. For. Res.* 33:624–634.

WANG, O.J. 1996. Direct sample estimators of L moments. *Water Res.* 32(12):3617–3619.

WANG, T., A. HAMANN, D.L. SPITTLEHOUSE, AND T.Q. MURDOCK. 2012. ClimateWNA—High-resolution spatial climate data for Western North America. *J. Appl. Meteorol. Climatol.* 51:16–29.

WATERSHED SCIENCES. 2009. *LiDAR remote sensing data collection reports*. Watershed Sciences, Portland, OR. 16 p.

WILSON, B.T., A.J. LISTER, AND R.I. RIEMANN. 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. *For. Ecol. Manage.* 271:182–198.