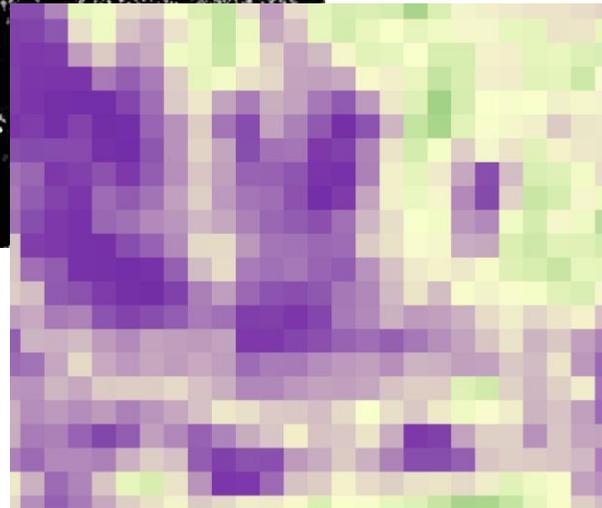
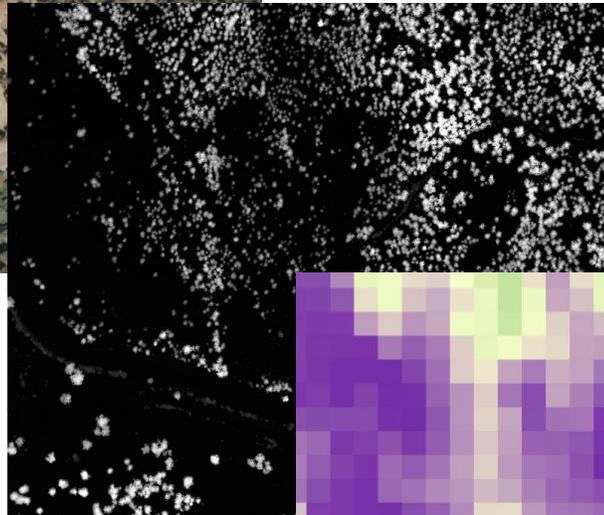
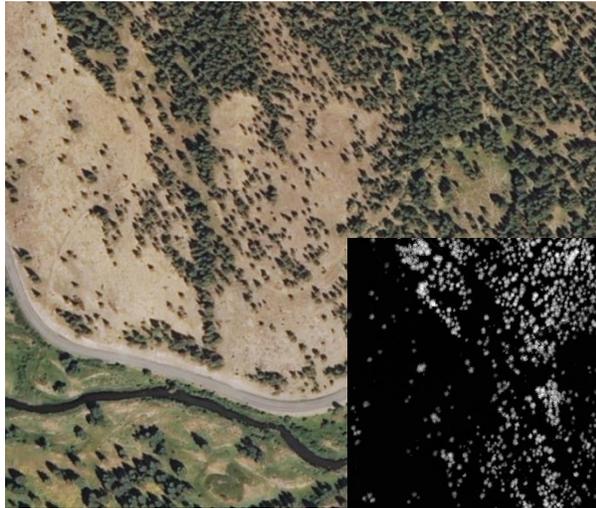




Investigation of LiDAR Data for National Applications



In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotape, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at [How to File a Program Discrimination Complaint](#) and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by: (1) mail: U.S. Department of Agriculture, Office of the Assistant Secretary for Civil Rights, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410; (2) fax: (202) 690-7442; or (3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

Cover image: LiDAR normalized digital surface model (ndsm), NAIP, and tree canopy cover predictions derived using the LiDAR data

Ruefenacht, B.; Heyer, J.P.; Housman, I.W.; Schleeweis, K.; Megown, K.; Bogle, S. 2023. Investigation of LiDAR Data for National Applications. FSIC-GO-10302-RPT1. U.S. Department of Agriculture, Forest Service, Geospatial Office. pp. 25

Executive Summary

The United States Department of Agriculture (USDA), Forest Service (FS), Forest Inventory and Analysis (FIA) is exploring the use of Light Detection and Ranging (LiDAR) data derived geospatial products to enhance landscape assessments. As part of this initiative, the USDA Forest Service Field Services and Innovation Center Geospatial Office (FSIC-GO) investigated pathways for applying recently acquired national LiDAR data to support the tree canopy cover (TCC) national mapping efforts.

Currently, for the Conterminous United States (CONUS), we use response data collected by FIA photo-interpreters using National Agriculture Imagery Program (NAIP) aerial imagery from 2009-2011, along with spectral fitted and topographic predictors to produce CONUS TCC geospatial products through time. For this project, we investigated the potential of using LiDAR data, as response data for creating CONUS geospatial TCC products. Analysis ready LiDAR products were the outcome of a FS research collaboration with the University of Vermont's Spatial Analysis Lab (SAL). They leverage the United States Geological Survey and the National Oceanic and Atmospheric Administration public LiDAR repositories. The analysis-ready LiDAR data, consists of 166,977 chips, each covering approximately 1 km². We used the first percentile returns (Pct1stRtns) to create TCC model training and model validation datasets. We cleaned these data for buildings and water, intersected them with various predictors, and used them to train machine learning models to predict TCC and standard error images. We compared these predicted images to those produced using models trained with FIA photo-interpreted TCC response data. The LiDAR response data models performed best when using seasonal predictors. The quality of the TCC and standard errors for these models outperformed the FIA PI response data models, except for hardwood dominated areas, which were affected by the leaf-off acquisition dates for the LiDAR data.

Authors

Bonnie Ruefenacht (Production Lead) is a contractor employed by RedCastle Resources as a Senior Remote Sensing Specialist in the Field Services and Innovation Center–Geospatial Office (FSIC–GO).

Ian Housman is a contractor employed by RedCastle Resources as a Senior Remote Sensing Specialist in the Field Services and Innovation Center–Geospatial Office (FSIC–GO).

Karen Schleeweis is a Techniques Team Lead for Forest Inventory and Monitoring, Rocky Mountain Research Station

Josh Heyer is a contractor employed by RedCastle Resources as a Geospatial Specialist in the FSIC–GO.

Kevin Megown is a Program Leader for Resource Mapping, Inventory, and Monitoring (RMIM)

Seth Bogle is a contractor employed by RedCastle Resources as a Geospatial Project Manager in the FSIC–GO.

Contents

Executive Summary.....	iii
Authors	iii
Introduction	5
LiDAR Data.....	5
Study Area.....	8
Methods.....	9
Step 1: Reproject.....	10
Step 2: Snap to the NLCD National Grid.....	10
Step 3. Mask Buildings	10
Step 4. Mask Water.....	10
Step 5. Calculate Average TCC Values and Combine with Spectral Data.....	10
Step 6. Create Model Calibration and Validation Datasets.....	11
Step 7. Develop and Apply Random Forest Models	14
Results and Discussion	20
Conclusion.....	24
References.....	24

Introduction

In 2011, the responsibility for generating tree canopy cover (TCC) data for the National Land Cover Database (NLCD) was assigned to the United States (US) Department of Agriculture (USDA), Forest Service (FS) (Homer et al., 2015). Initially, the NLCD TCC datasets were generated every ten years, with releases in 2001 and 2011, followed by five-year intervals in 2016 and 2021 (Homer et al., 2015, 2004; Yang et al., 2018). Due to the escalating demand for more up-to-date information (Yang et al., 2018), we developed approaches to generate NLCD TCC datasets at shorter intervals ensuring timely and accurate information. Through advancements in technology and modeling techniques, we reduced costs, enhanced repeatability, increased rigor, and improved data quality (Housman et al., 2025). However, further improvement is still needed regarding the response data used to build the TCC predictive models.

Currently, for the Conterminous United States (CONUS), we use response data collected by the USDA FS Forest Inventory and Analysis (FIA) – a special collection for the TCC project of photo-interpreted plots using National Agriculture Imagery Program (NAIP) 2009-2011 – along with annual predictors to produce CONUS TCC geospatial products through time. For this project, we investigated the potential of using part of the analysis ready Light Detection and Ranging (LiDAR) data provided through a research collaboration between FIA and the University of Vermont's Spatial Analysis Lab (SAL).

LiDAR Data

The analysis ready LiDAR data include raw LiDAR point clouds at native resolution, normalized point clouds at native resolution, and roughly 100 raster metrics at 10m² resolution. Data in the public National Oceanic and Atmospheric Administration (NOAA) and US Geological Survey (USGS) 3D Elevation Program (3DEP) repositories were clipped to approximately 1 km² over all public plot locations of FIA National Forest Inventory (NFI, not including urban) plots, regardless of forest land status in CONUS and Alaska. The original LiDAR project acquisition was timed +/- 30 months of a FIA plot's measurement year (MEASYEAR) in Western states and +/- 18 months of a plot's MEASYEAR in Eastern states (Figure 1). The resultant data set consisted of 166,977 image chips, collected from 2005 to 2021, along with extensive quality description and quality control metrics. To allow for user specific tolerances and needs a comprehensive set of quality management metrics is part of the analysis ready LiDAR. SAL's automated routine is available at <https://gitlab.uvm.edu/SAL/lidar-dataprep>. The LiDAR chipping and processing criteria for the public LiDAR acquisitions results in the most comprehensive set of analysis ready LiDAR data over FIA NFI plots.

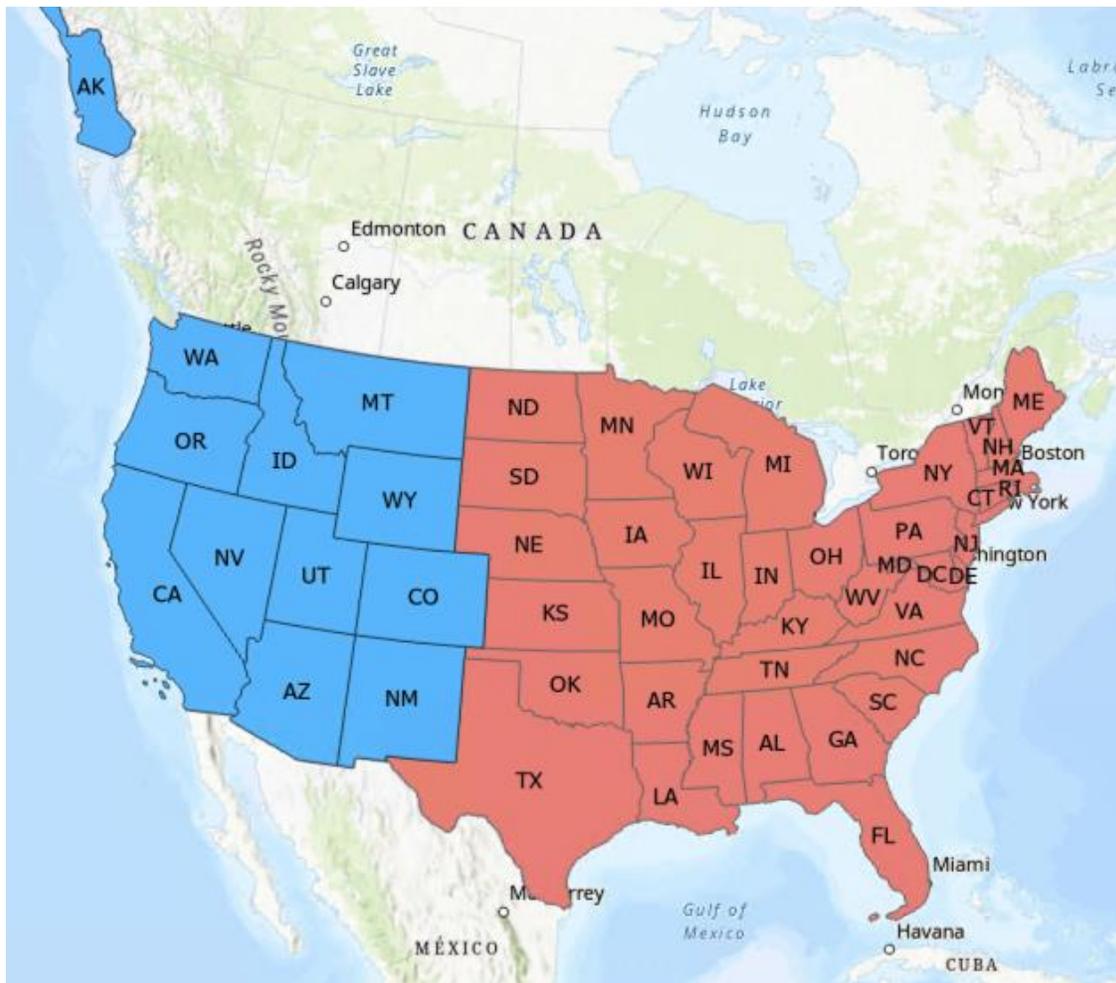


Figure 1.—Western (blue) and Eastern (red) states of the US that define the MEASYEAR for the LiDAR acquisitions.

Given the “inclusive” chipping criteria, there is some amount of re-measured data through time, with the highest being four collects over a FIA latitude and longitude coordinate pair. Over 90 custom and standard lidR package metrics (<https://r-lidar.github.io/lidRbook/>) were generated at 10m spatial resolution. At the time of processing, Fusion software was not available for Linux-based HPC processing (*FUSION/LDV LiDAR Processing and Visualization Software*, 2021). Each image chip was associated with 17 datasets containing LiDAR point clouds, various LiDAR metrics, and quality management information (Rounds et al., 2021).

From these datasets and metrics, we used the percentage of 1st returns above 2 meters (Pct1stRtns) with a spatial resolution of 10 meters for the LiDAR response data for TCC models. This metric captured not only vegetation but also structures, illustrated in Figure 2. We used the LiDAR quality management information to remove incomplete chips (Figure 3), where the LiDAR acquisition didn’t fully overlap the chip centered on the public location of the FIA plot. We discarded 826 incomplete image chips leaving a total of 165,470 “full” image chips.



Figure 2.—Example of the LiDAR dataset (left) metric for 1st returns above 2m in height (Pct1stRtns) identifying structures on the landscape. The image on the right is NAIP 2011 imagery.

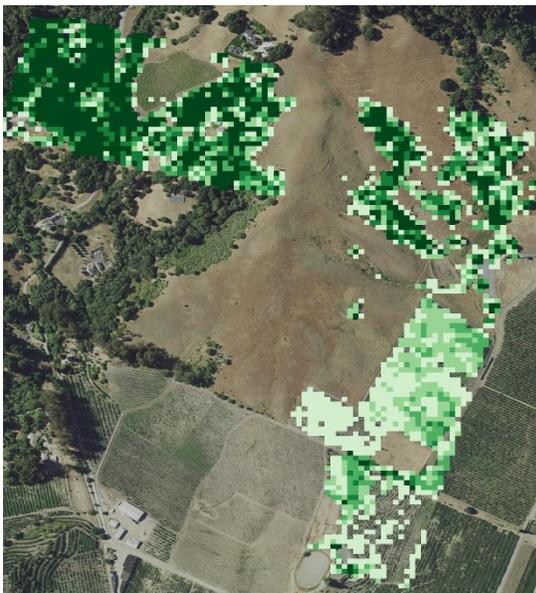


Figure 3.—Example of an analysis ready LiDAR chip not covering the full 1 km² footprint.

LiDAR project data acquisitions can be of long duration; of the analysis ready LiDAR chips used, acquisition periods of 18 to 60 months and spanning various years were recorded. With the data currently available in the original LiDAR project headers, exact acquisition dates for the area chipped is unknown. While an average year of acquisition for each chip is available, the LiDAR data could have been collected during either leaf-on or leaf-off periods. This is problematic for response

data for TCC mapping that relies on growing season image predictors, as leaf-off data can lead to poor modeled relationships, as discussed in later sections. An example of leaf-off LiDAR data is shown in Figure 4. In the area highlighted in red, the LiDAR derived TCC values will be in the 40-80 percent range, while a photo interpreter would likely estimate TCC to be above 90 percent.



Figure 4.—Left: Example of predicted TCC produced from machine learning models using LiDAR response data. Middle: Google Earth image showing the leaf-off conditions (January 2019). Right: 2010 NAIP image showing the area in growing season, which is similar to the imagery a FIA photo-interpreter used during the FIA TCC project data collect. Using the right image, a photo-interpreter would likely assign this area as greater than 90 percent TCC. However, the predicted TCC image on the left shows much lower TCC levels (40 – 80 percent) because of the leaf-off LiDAR data acquisition.

Study Area

We created a tiling system for CONUS, which is used for mapping TCC (Housman et al., 2023). We selected five tiles (7, 14, 19, 24, 38) to map TCC using LiDAR and FIA data combined with different predictor datasets (Figure 5). These tiles were chosen for their collective ecological diversity. Tile 7 presents challenges due to the small trees typical of chaparral and low-density forests. Tile 14 represents the arid Southwest. Tile 19 represents the hardwoods of the southern US. Tile 24 was selected to examine seamline effects between adjacent tiles. Tile 38 was chosen to represent northern hardwood and agricultural areas.

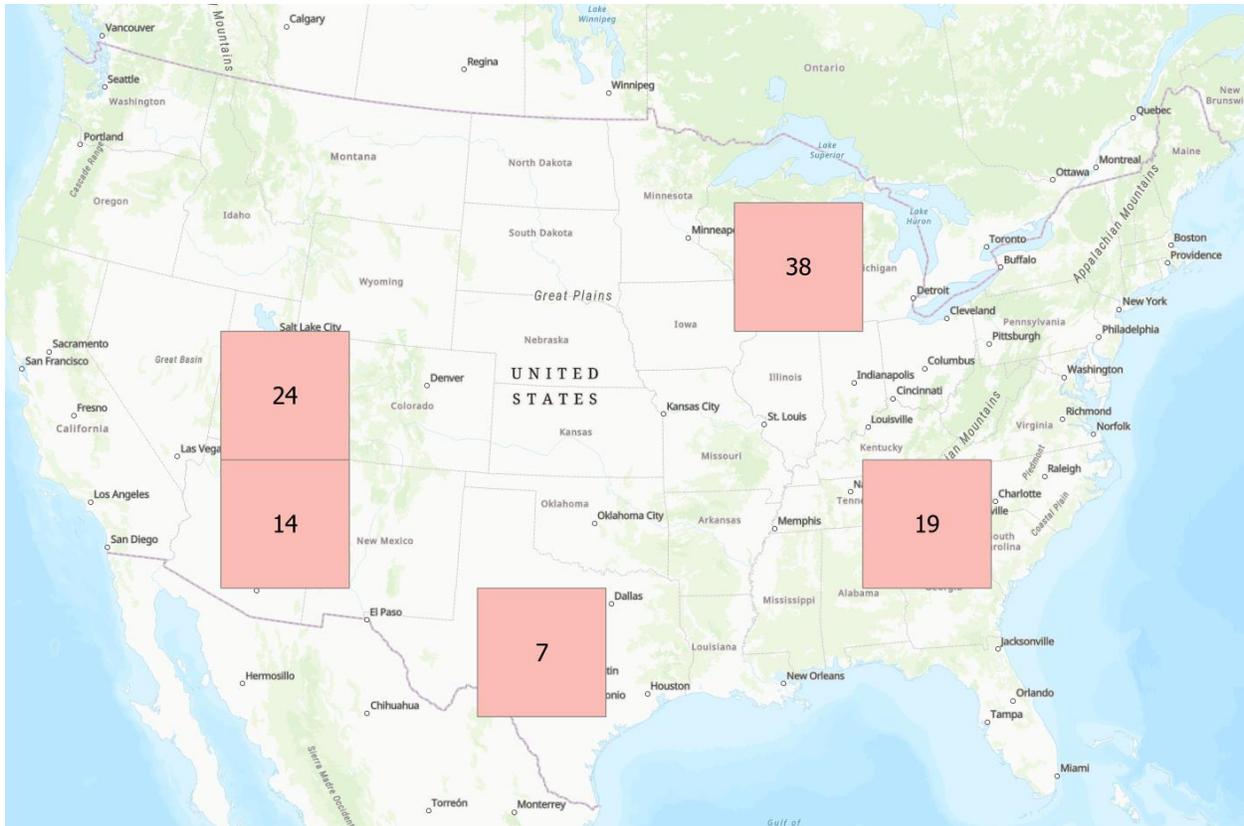


Figure 5.—TCC mapping tiles chosen for this project.

Methods

The following outlines the processing steps we used to create TCC maps from LiDAR data. Further details are provided below. All of the scripts mentioned below are in the FS GitHub:

code.fs.usda.gov/forest-service/TCC_NLCD_USFS/tree/main/lidar_investigation/LiDAR_Natl_Inventory

For each image chip,

1. Reproject to WGS 84 Albers.
2. Snap to the NLCD national grid
3. Mask buildings
4. Mask water
5. Calculate average TCC values and combine with spectral data
6. Create model training and validation datasets
7. Develop and apply random forest models

Step 1: Reproject

The GDAL reproject function was used to reproject the image chips to Albers WGS84, which is the projection for NLCD CONUS products. Since the reproject function assigns background values of zero, binary image masks were created beforehand to differentiate zero background values and zero TCC values. These image masks were also reprojected to Albers WGS 84.

([reproject_image_chips.py](#))

Step 2: Snap to the NLCD National Grid

To align the image chips to the NLCD national grid, the corner coordinates of the original, unreprojected images were transformed to the Albers WGS84 projection. The NLCD grid intersection points for the transformed corner coordinates were calculated. These intersection corner points were used to define the output geotransformation for the reprojected images from step 1. ([reproject_image_chips.py](#))

Step 3. Mask Buildings

Since the LiDAR Pct1stRtns does not differentiate between structures and vegetation, it was essential to find a method and/or dataset to mask structures. We used the step_21_ndsm_building_1m, which is part of the analysis ready LiDAR dataset package, along with the Microsoft Building Footprint dataset (<https://github.com/microsoft/USBuildingFootprints>) to mask buildings. Neither of these datasets perfectly masked buildings but combining the two datasets compensated for most of the dataset's errors of omission.

We downloaded the Microsoft Building Footprint geojsons for all individual states. Using ArcGIS, we converted the geojsons to shapefiles and reprojected them to Albers WGS84. Using Python's rasterio and geopandas modules, we clipped the shapefiles to the image chips and converted them to 1-meter binary rasters that were projected to Albers WGS84 and aligned with the NLCD grid ([2_clip_and_rasterize_mbf.py](#)).

The step_21_ndsm_building_1m dataset was reprojected to Albers WGS84 projection and snapped to the NLCD grid. These data were combined with the Microsoft Building Footprint data created above to create building masks. ([create_lidar_and_mbf_building_masks.py](#))

Step 4. Mask Water

We created a CONUS water binary mask by merging the water and snow/ice categories for NLCD years 2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019, and 2021 (<https://www.mrlc.gov/data>).

These years were chosen because they were all the years available at the time of this project. The water mask was resampled from 30-meters to 10-meters to correspond to the spatial resolution of the LiDAR image chips.

Step 5. Calculate Average TCC Values and Combine with Spectral Data

Using the building masks created in Step 3, the water masks created in Step 4, and the background masks created in Step 1, we examined the eight neighboring pixels around each pixel in the LiDAR image chips for the presence of buildings, water, or background values. If any of these were detected, the pixel was flagged as unavailable for further processing. For the remaining available pixels in the LiDAR Pct1stRtns images, we calculated the average TCC values for 3x3 windows

representing 30-m spatial resolution, from here on referred to simply as LiDAR response data. The corresponding 30-m spatial resolution predictor layers were extracted and this information along with the TCC values was exported to text files.

The predictor layers were the same as those used in the most recent NLCD TCC product (Housman et al., 2023). Two additional types of predictor layers were tested. We added a seasonal component (spring, summer, fall) to the annual LandTrendr fitted composite predictors and we also included the annual LandTrendr fitted composite derivatives dataset.

For FIA PI response data, values were extracted from year 2011 image predictors for training datasets. For LiDAR response data, the averaged year of the LiDAR collect was matched to the image predictors of the same year and compiled into a single training dataset. Temporal duplicate LiDAR observations in the same location were excluded since that would break the statistical assumption of independence.

Step 6. Create Model Calibration and Validation Datasets

The total number of pixel centers available for TCC modeling for each of the study area tiles were as follows: 2,235,753 for tile 7, 2,290,773 for tile 14, 5,787,911 for tile 19, 2,185,103 for tile 24, and 4,358,275 for tile 38 (Figure 6). The next step was to determine the optimal number of response data points to use for TCC modeling. To do this, we started with a random set of 9,489 points per tile. Since the area of each tile is 230,400 km², this corresponds to 1 point per 24 km² or 1 point per 6,000 acres, similar to the “base” intensity of the FIA grid (Bechtold & Patterson, 2005). We multiplied

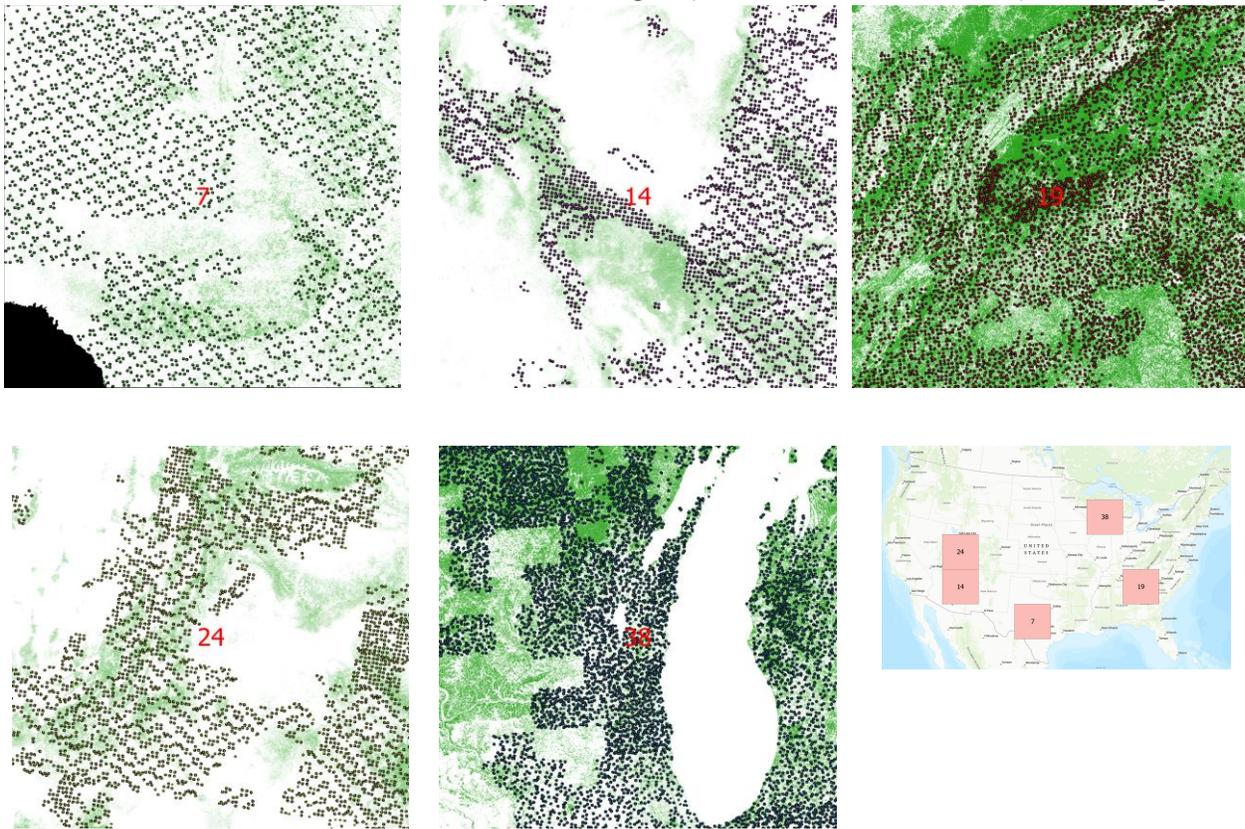


Figure 6.—LiDAR training data for tiles 7,14,19,24,38. NLCD TCC 2021 is the background image.

9,489 by five increasing in increments of five until reaching forty, a process we refer to as point intensification.

We used these random sets of points for training and the remainder for model validation. Using the random forest algorithm in R (Breiman, 2001; R Core Team, 2024), we obtained out-of-bag model performance metrics and calculated root mean square deviation (RMSD) statistics from the model validation datasets. Figures 7 through 9 show the out-of-bag percent variance explained (PVE), the out-of-bag RMSD, and the model validation RMSD. The graphs show there is a decline in the slopes around an intensification of 20, which is 189,781 points ($9,489 \times 20$). This is the number of response data points used for the training datasets. The remaining points were used for model validation.

We developed eight training datasets. Four of the training datasets used a random selection of 189,781 points from the datasets and only differed in the predictor datasets. One dataset used the same predictors as used in the previous NLCD TCC production effort (Housman et al., 2023). Another dataset used seasonal (spring, summer, fall) LandTrendr fitted composites as predictors. The remaining two predictor datasets added annual LandTrendr fitted composite derivatives to the previous two datasets. All datasets included eight topographic variables: slope, elevation, aspect, sine transformation of aspect, cosine transformation of aspect, component of slope in the x direction, component of slope in the y direction, and magnitude of slope.

The other four models used the same predictor datasets as described above but used the FIA photo-interpreted plots collected using NAIP 2009-2011 as the response variable instead of LiDAR.

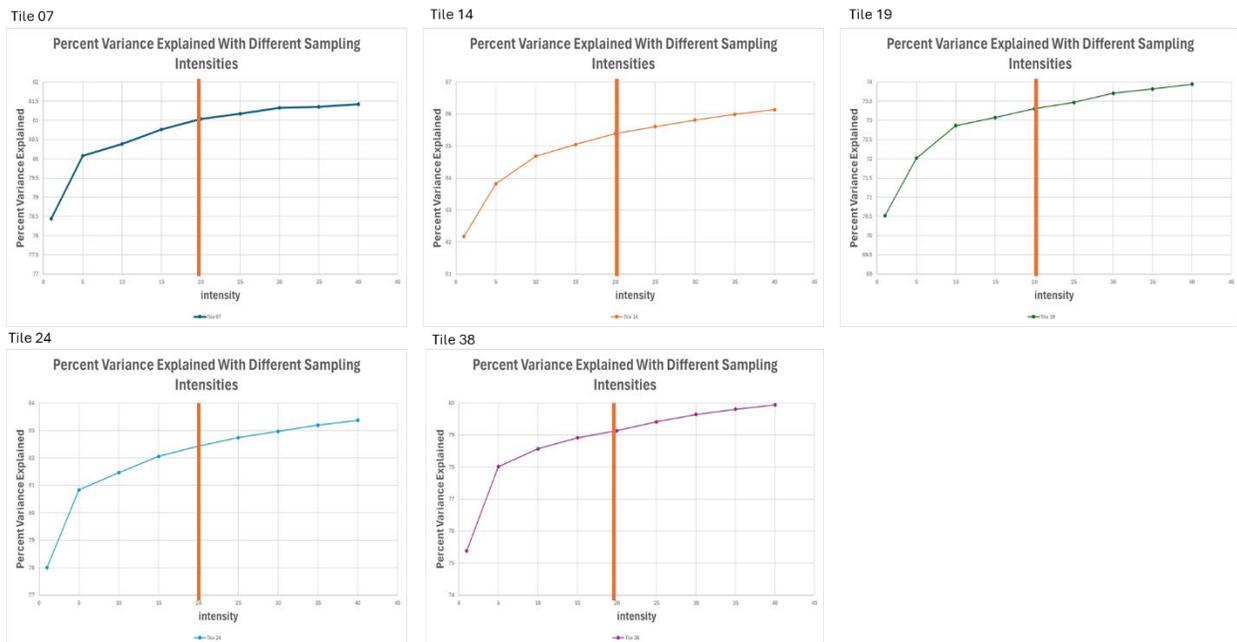
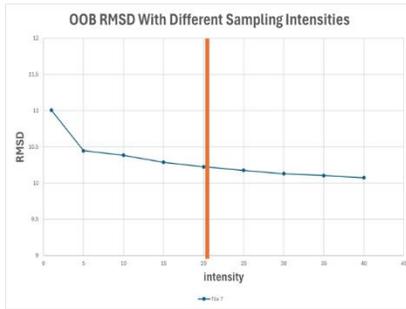
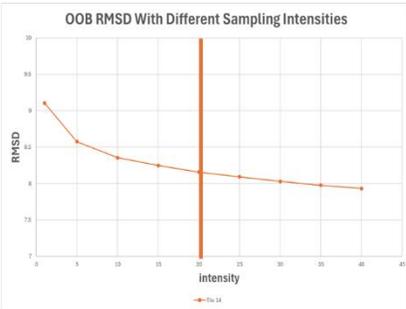


Figure 7.—Random forest out-of-bag percent variance explained (PVE) for models created with different number of response data points. The vertical line depicts the number of response data points ultimately selected for this project.

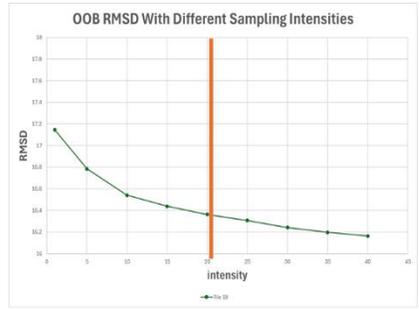
Tile 07



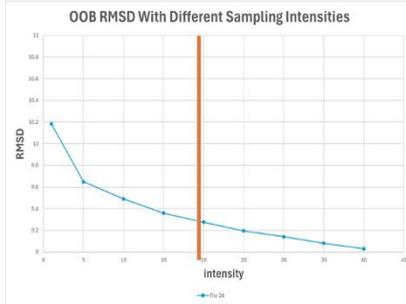
Tile 14



Tile 19



Tile 24



Tile 38

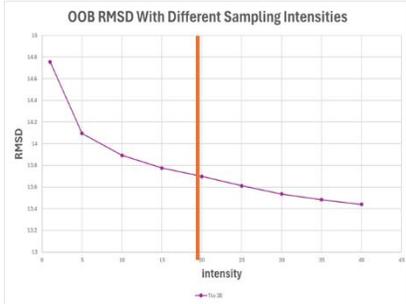
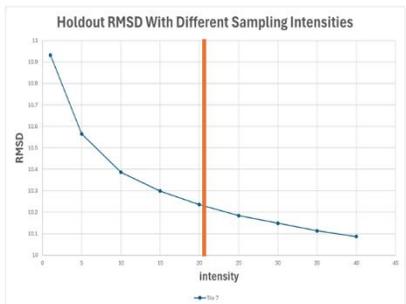
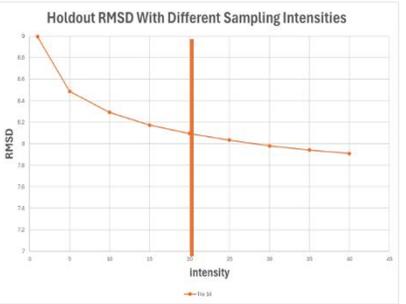


Figure 8.—Random forest out-of-bag RMSD for models created with different number of response data points. The vertical line depicts the number of response data points ultimately selected for this project.

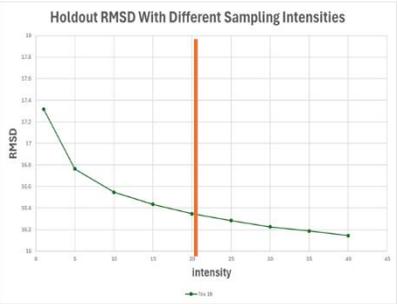
Tile 07



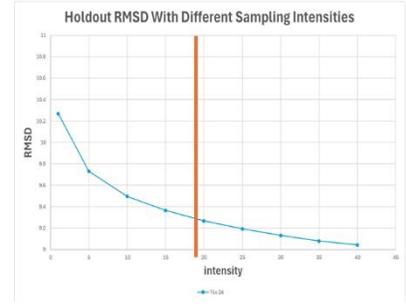
Tile 14



Tile 19



Tile 24



Tile 38

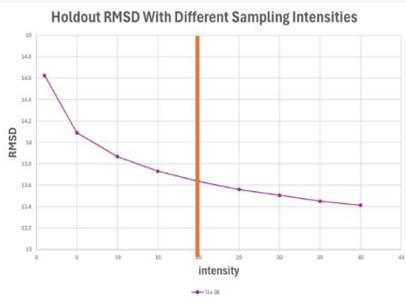


Figure 9.—RMSD calculated using the model validation datasets for models created with different number of response data points. The vertical line depicts the number of response data points ultimately selected for this project.

Step 7. Develop and Apply Random Forest Models

We applied Variable Selection in Random Forest (VSURF) in R to six of the datasets (Genuer et al., 2013; R Core Team, 2024). VSURF was not used for the datasets with annual LandTrendr fitted composites, as they only contained 22 variables. The other datasets had 50 or 106 variables. The variables selected by VSURF for the six modeling datasets and for each tile are shown in Table 1.

We used the random forest model from the CUDA (Compute Unified Device Architecture) machine learning library (cuML) (<https://docs.rapids.ai/api/cuml/stable/>). We developed TCC and standard error images for each of the eight modeling scenarios for each of the five study areas.

Table 1.—VSURF variables chosen for the five tiles and six models. The order of the variables is the VSURF order of importance. A variable with LT in the name is a derivative variable. The names following annual, spring, summer, or fall is a Landsat band name, Landsat derivative name, or TasselCap name. The remaining variables are topographic variables.

Tile 07

LiDAR - Seasonal LandTrendr	LiDAR - Seasonal LandTrendr & Derivatives	LiDAR - Annual LandTrendr & Derivatives	FIA PI - Seasonal LandTrendr	FIA PI - Seasonal LandTrendr & Derivatives	FIA PI - Annual LandTrendr & Derivatives
fall wetness	fall wetness	annual ndvi	fall swir1	fall swir1	annual green
fall swir1	fall swir1	gradient magnitude of slope	fall brightness	fall brightness	annual brightness
summer swir1	fall swir2	slope	fall green	fall green	annual swir1
fall swir2	summer swir1	elevation	fall blue	fall blue	annual blue
summer red	summer red	annual tcanglebg	fall wetness	fall red	annual wetness
spring brightness	fall nbr	red LT diff	elevation	fall wetness	elevation
fall nbr	summer wetness	annual red	fall swir2	elevation	annual swir2
fall ndvi	fall ndvi	red LT mag	fall tcanglebg	fall swir2	annual ndvi
fall ndmi	fall ndmi	swir1 LT mag	summer green	summer green	annual ndmi
summer wetness	fall red	brightness LT mag	summer blue	fall tcanglebg	annual nbr
fall red	spring brightness	swir2 LT mag	spring wetness	summer blue	gradient magnitude of slope
summer green	summer green	ndvi LT diff	spring swir1	summer swir1	annual nir
spring swir1	fall brightness	annual nbr	spring ndsi	spring swir1	x component of slope
spring swir2	summer brightness	ndvi LT mag	summer wetness	spring ndsi	wetness LT dur
fall brightness	spring swir1	ndsi LT mag	slope	slope	
summer ndvi	summer ndvi	nir LT mag	summer ndmi		
elevation	elevation	wetness LT mag			
summer ndmi	slope	blue LT mag			
spring wetness	summer blue	ndsi LT dur			
slope	fall nir	siwr2 LT diff			
spring ndmi	summer ndsi	ndvi LT slope			
aspect - cosine transformed	spring ndsi	wetness LT dur			
	y component of slope	swir1 LT dur			
	swir LT diff	green LT mag			

Tile 14

LiDAR - Seasonal LandTrendr	LiDAR - Seasonal LandTrendr & Derivatives	LiDAR - Annual LandTrendr & Derivatives	FIA PI - Seasonal LandTrendr	FIA PI - Seasonal LandTrendr & Derivatives	FIA PI - Annual LandTrendr & Derivatives
spring red	spring ndvi	annual green	elevation	spring blue	annual blue
spring ndvi	spring red	annual blue	fall ndvi	spring swir2	annual green
fall ndvi	fall ndvi	annual brightness	spring blue	summer green	elevation
spring blue	spring blue	annual red	fall tcanglebg	summer brightness	annual ndvi
summer brightness	summer red	annual ndvi	summer green	fall swir1	annual red
summer red	summer brightness	annual swir1	fall swir1	fall ndvi	annual swir2
spring green	elevation	annual tcanglebg	spring swir2	fall tcanglebg	annual swir1
elevation	spring tcanglebg	elevation		elevation	slope
spring tcanglebg	fall tcanglebg	annual swir2			greenness LT diff
fall tcanglebg	spring brightness	annual wetness			tcAngleBG LT mag
spring brightness	fall swir1	gradient magnitude of slope			ndmi LT mag
fall swir1	summer green	annual nir			ndvi LT mag
summer blue	summer blue	greenness LT diff			ndvi LT diff
summer green	spring nbr	nbr LT diff			
fall swir2	summer ndvi	nbr LT mag			
fall greenness	summer tcanglebg	nir LT mag			
summer swir1	gradient magnitude of slope	red LT mag			
summer ndvi	slope	green LT diff			
spring swir1	summer nbr	ndvi LT mag			
summer swir2	nbr LT diff	ndsi LT mag			
summer tcanglebg		aspect			
summer nbr					
gradient magnitude of slope					
summer ndmi					
slope					
summer ndsi					
fall nir					

Tile 19

LiDAR - Seasonal LandTrendr	LiDAR - Seasonal LandTrendr & Derivatives	LiDAR - Annual LandTrendr & Derivatives	FIA PI - Seasonal LandTrendr	FIA PI - Seasonal LandTrendr & Derivatives	FIA PI - Annual LandTrendr & Derivatives
spring swir1	summer swir2	nbr LT mag	summer green	summer green	annual green
summer swir2	spring swir1	annual swir1	summer red	summer red	annual red
summer swir1	summer swir1	annual swir2	summer swir2	summer swir2	annual swir2
summer wetness	summer wetness	annual wetness	summer nbr	summer nbr	annual tcanglebg
summer green	summer green	greenness LT mag	summer ndvi	summer ndvi	annual wetness
spring brightness	spring brightness	elevation	summer wetness	summer wetness	annual ndvi
fall green	summer red	wetness LT mag	summer swir1	fall green	annual swir1
fall swir1	fall green	swir2 LT mag	fall green	summer tcanglebg	annual ndsi
summer red	fall swir1	nir LT mag	summer tcanglebg	spring brightness	gradient magnitude of slope
fall brightness	fall brightness	gradient magnitude of slope	summer ndmi	summer ndsi	swir2 LT mag
spring wetness	fall ndmi	nbr LT diff	summer greenness	spring red	ndmi LT diff
fall ndmi	spring wetness	ndvi LT slope		spring nir	
summer nbr	fall ndvi	annual green		spring ndvi	
elevation	summer nbr	annual red		gradient magnitude of slope	
fall ndvi	fall red	wetness LT diff		ndmi LT diff yes17	
fall red	elevation	red LT mag			
spring ndsi	summer brightness	siwr2 LT diff			
summer ndsi	slope	annual blue			
gradient magnitude of slope	spring ndsi	swir LT diff			
	summer tcanglebg	ndsi LT mag			
	spring nir	annual greenness			
	wetness LT slope	ndsi LT diff			
	ndmi LT slope	swir1 LT dur			
	spring blue	wetness LT slope			
	ndsi LT diff	green LT diff			
	ndmi LT mag	red LT slope			
	swir2 LT mag	brightness LT diff			
		swir2 LT dur			
		ndmi LT slope			

Tile 24

LiDAR - Seasonal LandTrendr	LiDAR - Seasonal LandTrendr & Derivatives	LiDAR - Annual LandTrendr & Derivatives	FIA PI - Seasonal LandTrendr	FIA PI - Seasonal LandTrendr & Derivatives	FIA PI - Annual LandTrendr & Derivatives
summer green	summer green	annual green	fall ndvi	fall ndvi	annual ndvi
summer swir1	summer swir1	annual blue	summer tcanglebg	summer tcanglebg	annual tcanglebg
fall tcanglebg	fall ndvi	annual brightness	summer ndvi	summer ndvi	slope
fall ndvi	fall tcanglebg	slope	elevation	elevation	annual green
summer red	summer red	annual ndvi	summer green	summer green	annual brightness
summer brightness	summer brightness	elevation	slope	fall tcanglebg	elevation
fall greenness	fall greenness	annual tcanglebg	fall tcanglebg	slope	annual blue
summer swir2	summer swir2	annual swir1	fall greenness	fall greenness	annual swir1
summer wetness	slope	annual swir2	gradient magnitude of slope	gradient magnitude of slope	annual ndmi
slope	summer wetness	annual wetness	summer swir1	summer swir1	swir LT diff
summer ndvi	summer blue	annual nbr	fall brightness	fall brightness	
elevation	summer ndvi	red LT diff	summer ndmi	summer ndmi	
fall swir2	elevation	red LT mag	fall green	fall green	
fall brightness	fall swir2	ndsi LT mag	fall red	fall nir	
summer tcanglebg	summer tcanglebg	swir2 LT mag	fall swir2	fall swir2	
spring swir1	summer ndmi	annual ndsi	spring greenness		
spring tcanglebg	fall brightness	ndsi LT diff	fall wetness		
fall swir1	fall swir1	nir LT mag			
summer ndmi	spring tcanglebg	wetness LT dur			
spring greenness	spring swir1	wetness LT mag			
spring brightness	spring greenness	ndmi LT mag			
spring nir	spring brightness	ndmi LT diff			
fall ndsi	spring nir				
x component of slope	fall ndsi greenness LT diff				
	nir LT mag x component of slope aspect - sine tranformed				

Tile 38

LiDAR - Seasonal LandTrendr	LiDAR - Seasonal LandTrendr & Derivatives	LiDAR - Annual LandTrendr & Derivatives	FIA PI - Seasonal LandTrendr	FIA PI - Seasonal LandTrendr & Derivatives	FIA PI - Annual LandTrendr & Derivatives
summer red	summer red	nbr LT mag	summer red	summer red	annual red
summer green	summer green	annual red	summer green	summer green	annual green
summer swir2	summer swir2	annual green	summer swir2	summer swir2	annual ndvi
fall swir1	fall swir1	swir2 LT mag	summer ndvi	summer ndvi	annual nbr
summer swir1	summer wetness	annual blue	summer nbr	summer nbr	annual swir2
summer wetness	summer swir1	greenness LT mag	spring ndvi	spring ndvi	annual tcanglebg
spring red	spring red	siwr2 LT diff	fall green	summer tcanglebg	annual brightness
spring ndvi	spring ndvi	nir LT mag	spring brightness	fall green	annual swir1
spring swir2	spring swir2	nbr LT diff	summer wetness	summer wetness	annual nir
fall wetness	fall wetness	annual brightness	summer blue	summer blue	wetness LT mag
spring swir1	spring swir1	wetness LT mag	spring greenness	spring brightness	greenness LT mag
spring brightness	spring green	wetness LT diff	fall swir1	fall swir1	tcAngleBG LT mag
spring green	spring brightness	ndmi LT mag		fall swir2	slope
fall ndsi	fall green	nbr LT slope		spring green	
fall brightness	summer ndmi	ndvi LT mag		summer swir1	
summer ndmi	fall brightness	elevation		wetness LT mag	
summer nbr	summer nbr	swir2 LT dur		fall greenness	
spring nbr	fall ndmi	red LT mag		ndmi LT mag	
fall greenness	spring greenness	annual swir1		swir1 LT mag	
spring ndsi	nir LT mag	ndvi LT slope			
slope	fall greenness	swir2 LT slope			
gradient magnitude of slope	ndmi LT mag	gradient magnitude of slope			
elevation	nbr LT mag	red LT slope			
	gradient magnitude of slope	ndsi LT diff			
	nbr LT diff	red LT dur			
		nbr LT dur			
		wetness LT dur			

Results and Discussion

Table 2 shows the model performance metrics (PVE and RMSD) and RMSD from the independent model validations. The models using LiDAR response data and annual Landtrendr composites and derivatives generally performed better than the same models but without the derivatives. For example, for tile 19, the OOB PVE, OOB RMSD, and holdout RMSD for the LiDAR – Annual Landtrendr & Derivates model was 52.2, 21.9, and 20.2, respectively, and 29.5, 26.6, and 22.9,

Table 2.—Out-of-bag percent variance explained (OOB PVE), out-of-bag RMSD (OOB RMSD), and the RMSD from the model validation (Holdout RMSD) for the different models and study area tiles.

Tile		OOB PVE	OOB RMSD	Holdout RMSD
Tile 07	LiDAR - Seasonal LandTrendr	80.6	10.3	9.4
	LiDAR - Seasonal LandTrendr & Derivatives	80.5	10.4	9.5
	LiDAR - Annual LandTrendr	32.3	19.3	14.3
	LiDAR - Annual LandTrendr & Derivatives	53.6	16	12.8
	FIA PI - Seasonal LandTrendr	63	16.9	15.3
	FIA PI - Seasonal LandTrendr & Derivatives	NA	NA	NA
	FIA PI - Annual LandTrendr	58.4	17.9	15.3
	FIA PI - Annual LandTrendr & Derivatives	60	18	15.4
Tile 14	LiDAR - Seasonal LandTrendr	84.7	8.4	7.7
	LiDAR - Seasonal LandTrendr & Derivatives	84.7	8.4	7.6
	LiDAR - Annual LandTrendr	76.2	10.4	8.8
	LiDAR - Annual LandTrendr & Derivatives	78.6	9.9	8.4
	FIA PI - Seasonal LandTrendr	66.8	8.1	6.7
	FIA PI - Seasonal LandTrendr & Derivatives	NA	NA	NA
	FIA PI - Annual LandTrendr	60.7	8.8	7.6
	FIA PI - Annual LandTrendr & Derivatives	62.2	8.7	7.4
Tile 19	LiDAR - Seasonal LandTrendr	72.1	16.7	16.8
	LiDAR - Seasonal LandTrendr & Derivatives	73	16.5	16.6
	LiDAR - Annual LandTrendr	29.5	26.6	22.9
	LiDAR - Annual LandTrendr & Derivatives	52.2	21.9	20.2
	FIA PI - Seasonal LandTrendr	90.3	12.1	12.0
	FIA PI - Seasonal LandTrendr & Derivatives	90.5	12	11.7
	FIA PI - Annual LandTrendr	90.1	12.2	11.6
	FIA PI - Annual LandTrendr & Derivatives	90.1	12.1	11.8
Tile 24	LiDAR - Seasonal LandTrendr	81.7	9.5	9
	LiDAR - Seasonal LandTrendr & Derivatives	81.8	9.4	9
	LiDAR - Annual LandTrendr	58.9	14.2	10.9
	LiDAR - Annual LandTrendr & Derivatives	62.5	13.5	11.1
	FIA PI - Seasonal LandTrendr	69.2	11.1	9.9
	FIA PI - Seasonal LandTrendr & Derivatives	69.3	11.1	9.9
	FIA PI - Annual LandTrendr	63.6	12.1	10.6
	FIA PI - Annual LandTrendr & Derivatives	63.7	12.1	10.6
Tile 38	LiDAR - Seasonal LandTrendr	78.1	14	15.6
	LiDAR - Seasonal LandTrendr & Derivatives	77.7	14.2	15.7
	LiDAR - Annual LandTrendr	36	24	20.4
	LiDAR - Annual LandTrendr & Derivatives	56.5	19.8	18.5
	FIA PI - Seasonal LandTrendr	85.3	14.2	14.2
	FIA PI - Seasonal LandTrendr & Derivatives	86	13.9	13.9
	FIA PI - Annual LandTrendr	83.4	15.1	14.9
	FIA PI - Annual LandTrendr & Derivatives	83.8	14.9	14.4

respectively, for the same model but without the derivatives (LiDAR – Annual LandTrendr). For all the other tiles and models, there were minimal difference between models with and without derivatives for both the LiDAR and FIA PI trained models. VSURF did not select any derivative variables for tile 7 and tile 14 for the FIA PI models using seasonal and derivative predictors. The models using annual LandTrendr fitted composites with LiDAR response variables performed worse compared to those using seasonal LandTrendr fitted composites. The OOB PVE for the models using annual Landtrendr fitted composites was < 77 and, except for tile 14, had considerably lower OOB PVE as compared to the other models. For tiles 19 and 38, the FIA PI version with seasonal LandTrendr fitted composites outperformed the LiDAR version. However, for the other tiles, the LiDAR version with seasonal LandTrendr fitted composites performed better than the FIA PI version. The models using FIA PI as response variables and Seasonal Landtrendr for predictors outperformed the models using Annual Landtrendr.

Figure 10 shows the predicted TCC versus the model validation observed LiDAR TCC for the eight model versions for the five tiles. As noted in Table 2, the results from the models using derivative predictors are very similar to those without. For tiles 19 and 38, the predicted TCC values above 60 percent from models using FIA PI response data were higher than those from the models using LiDAR response data. This is caused by the presence of leaf-off TCC in the LiDAR data.

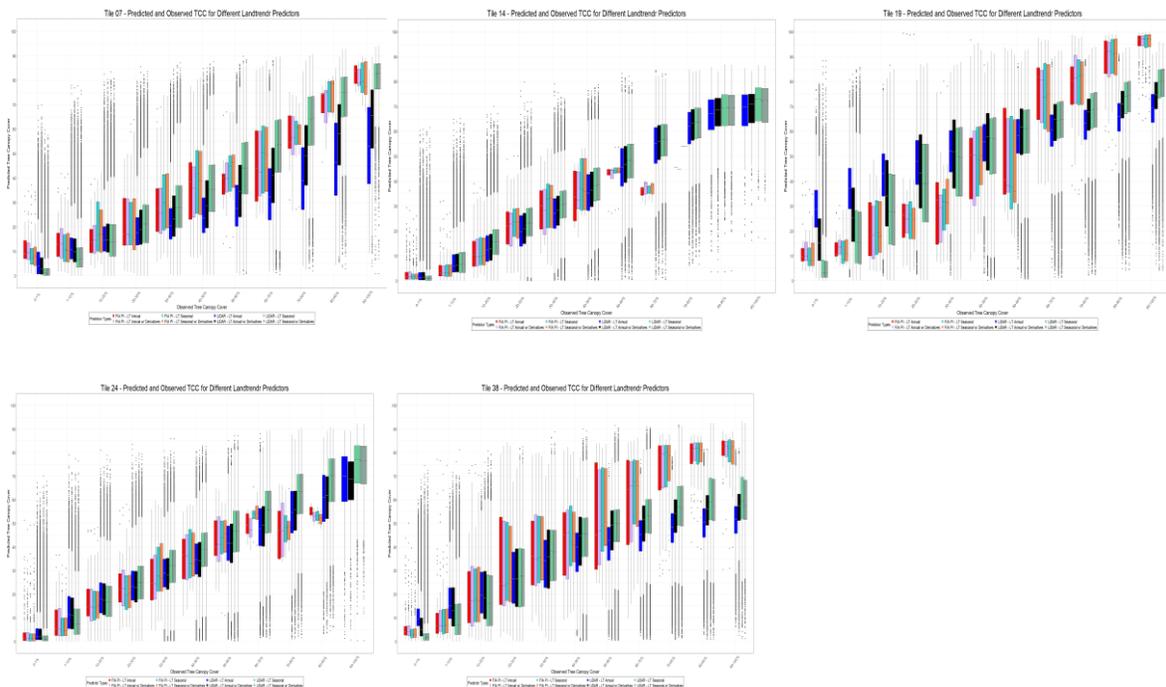


Figure 10.— Predicted vs Observed TCC for the eight models for the five study area tiles.

Figure 11 shows predicted TCC against the standard errors which were calculated from the 500 random forest decision trees for each prediction. As has been observed, the difference between predictions and standard errors for models with and without LandTrendr derivatives is minimal. Predictions developed from models using LiDAR response data and annual predictors had high standard errors as compared to predictions from all the other models. For tiles 7, 14, 24, and 38 predictions developed using LiDAR response data with seasonal predictors tended to have lower

standard errors compared to predictions from the other models. For tile 19, the predictions developed from the model using LiDAR response data and seasonal predictors had higher standard errors for predictions in the 10 to 50 percent range as compared to predictions developed from models using FIA PI response data. For the other ranges, the standard errors were more similar for the FIA PI and LiDAR versions, though the predictions developed using LiDAR response data generally had slightly lower standard errors than predictions developed using FIA PI response data. For all tiles, predictions developed from models using FIA PI response data and seasonal Landtrendr predictors had lower standard errors than those developed with annual LandTrendr predictors.

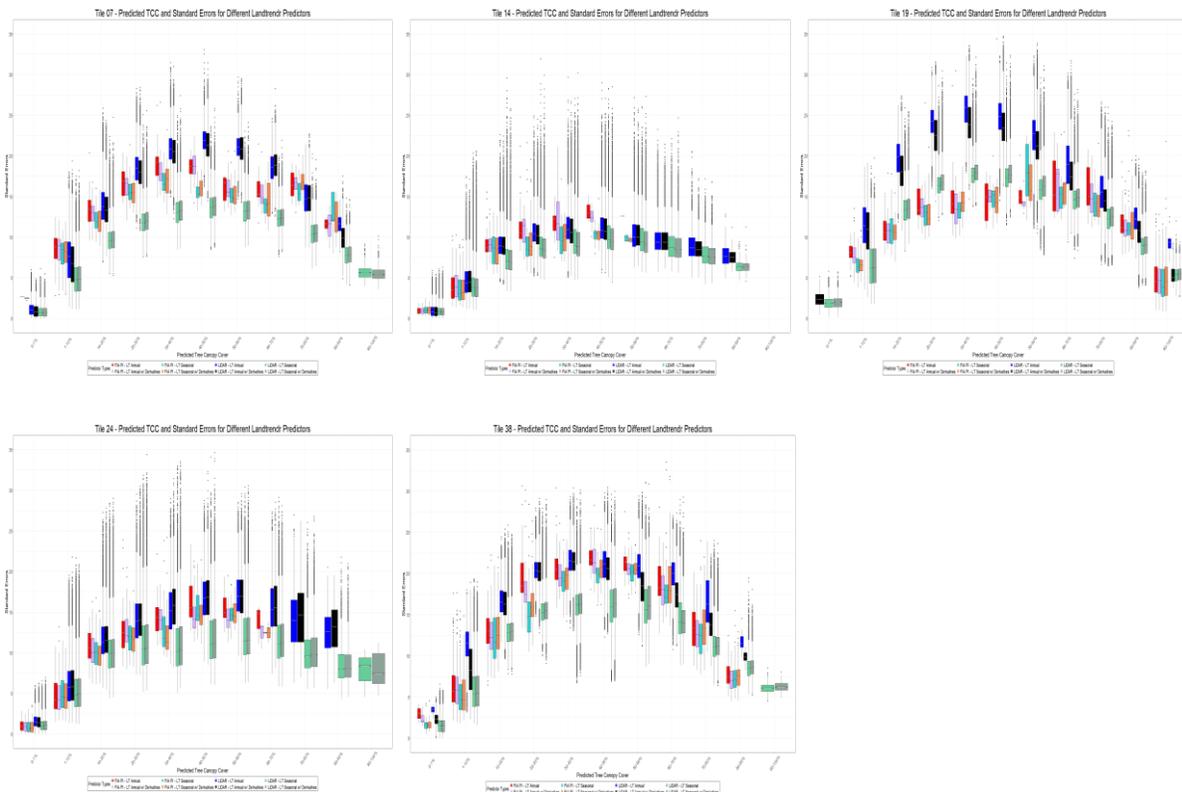


Figure 11.—Predicted TCC vs standard errors for the eight models for the five study area tiles.

Figures 12 and 13 are plot densities of image values for TCC and standard errors, respectively. These figures focus on the models using LiDAR and FIA PI response data and with seasonal predictors and annual predictors. For comparison purposes, the TCC and standard error values from the v2021.4 dataset are displayed. The datasets for the LandTrendr derivatives were excluded from these figures since the previous analysis demonstrated that LandTrendr derivatives did not affect model outputs. The TCC distributions for tiles 7, 14, and 24 are similar in form. The TCC distributions for tiles 19 and 38 are different because of the leaf-off TCC values in the LiDAR response data. For tile 19 and 38, the TCC distribution of predictions from the models using the

LiDAR response data peak in the 50 – 70 percent TCC range, while those from the models using the FIA PI response data peak in the 80 – 100 percent TCC range.

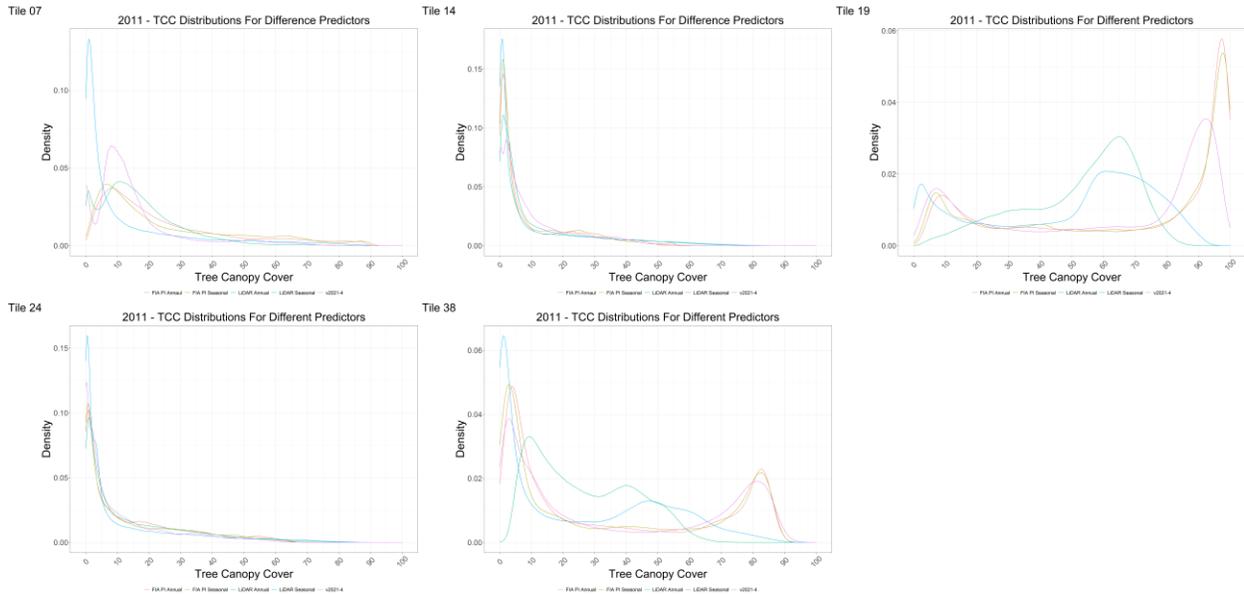


Figure 12.—Density plots for TCC predictions. Response data used were either LiDAR or FIA PI. The predictors were either annual or seasonal spectral imagery along with topographic datasets.

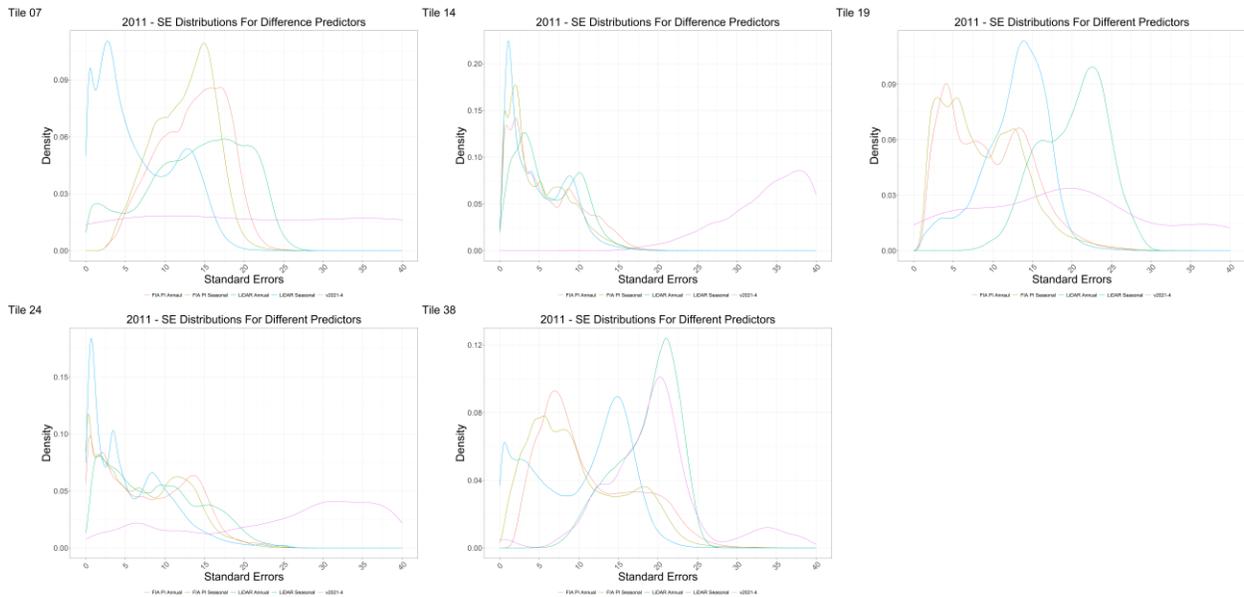


Figure 13.—Density plots for standard errors of the TCC predictions. Response data used were either LiDAR or FIA PI. The predictors were either annual or seasonal spectral imagery along with topographic datasets.

The standard error distributions are different for every tile (Figure 13). For tile 7, the standard error distribution for the LiDAR response data and 11 seasonal LandTrendr predictors had the most values

in the lower ranges while the other distributions had the most values in the middle ranges. The standard error distribution for the v2021.4 product for tile 7 was consistent across the range. For tiles 14 and 24, the v2021.4 product had more standard errors in the high range than all the other standard error distributions. For tiles 19 and 38, the standard error distributions for the LiDAR response data had higher standard errors than for the FIA PI response data. For tile 38, the standard error distribution for the v2021.4 product resembles the standard error distribution for the LiDAR response data and annual LandTrendr predictors.

These figures illustrate the impact of using LiDAR leaf-off data acquisitions on mapping TCC. TCC predictions for tiles 19 and 38, dominated by hardwood forests, exceeded 80 percent for both FIA PI response data models but peaked in the 60 to 70 percent range for LiDAR response models, with very few values above 90 percent. Tiles 7, 14, and 24, where coniferous tree species comprise the majority of forests, did not exhibit the issues shown in tiles 19 and 38. The LiDAR response model for tiles 7, 14, and 24 showed more TCC predictions below two percent, suggesting that leaf-off LiDAR may be the cause for more low values as compared to the FIA PI response data models.

Conclusion

Historically, the FS TCC project has relied on FIA photo-interpreted plot data collected for the project circa 2011. The TCC workflow was overhauled for the 2021.4 release to allow for time invariant modeling, where a model built with one year of response and predictor data is applied to multiple years of annual predictor data (Housman et al., 2023). The rebuilt workflow and additional LiDAR testing also allow for time agnostic modeling (response and predictors from the same year, are mixed across years, in one model). These TCC workflow modifications plus the addition of a large analysis-ready LiDAR dataset, demonstrate that a viable alternative exists to the FIA TCC project FIA PI collect, and/or the time invariant modeling used in the last TCC production cycle (v2021.4). However, challenges remain. There currently is no method to derive pixel-wise dates from LiDAR acquisitions in the public USGS and NOAA repositories, making it difficult to differentiate between leaf-on and leaf-off points. As seen in tiles 19 and 38, the absence of filtering for leaf-off conditions impacted TCC quality in areas with more hardwoods, which are more sensitive to leaf-off data. Multiple TCC science projects are investigating possible solutions suitable for applications at the national project level. Here we provide conclusive evidence that incorporating seasonal predictors with the LiDAR response data and FIA PI response data achieves higher-quality results compared to single growing season only image predictors, especially where leaf-off LiDAR is confounding models.

References

Bechtold, W. A., & Patterson, P. L. (Eds.). (2005). *The enhanced forest inventory and analysis program—National sampling design and estimation procedures. Gen. Tech. Rep. SRS-80*. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. Retrieved from <https://doi.org/10.2737/SRS-GTR-80>

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi: 10.1023/A:1010933404324

FUSION/LDV LiDAR processing and visualization software. (2021). United States Department of Agriculture, Forest Service. Retrieved from <https://research.fs.usda.gov/pnw/products/dataandtools/fusion/ldv-lidar-processing-and-visualization-software-version-440>

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2013). *VSURF: Variable Selection Using Random Forests* (p. 1.2.0) [Data set]. doi: 10.32614/CRAN.package.VSURF

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering & Remote Sensing*, 81(5), 345–354. doi: 10.1016/S0099-1112(15)30100-2

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 National Land Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70(7), 829–840. doi: 10.14358/PERS.70.7.829

Housman, I., Schleeweis, K., Heyer, J., Ruefenacht, B., Bender, S., Megown, K., Goetz, W., & Bogle, S. (2023). *National Land Cover Database Tree Canopy Cover Methods v2021.4. GTAC-10268-RPT1* (No. GTAC-10268-RPT1; p. 29). Salt Lake City, UT: Field Services and Innovation Center- Geospatial Office. Retrieved from Field Services and Innovation Center- Geospatial Office website: https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/docs/TCC_v2021-4_Methods.pdf

Housman, I. W., Heyer, J. P., Ruefenacht, B., Schleeweis, K., Megown, K., Bogle, S., Reischmann, J., & Ryerson, D. (2025). *National Land Cover Database Tree Canopy Cover Methods v2023.5. FSIC-GO-10268-RPT2* (pp. 1–29). Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Field Services and Innovation Center- Geospatial Office. Retrieved from U.S. Department of Agriculture, Forest Service, Field Services and Innovation Center- Geospatial Office website: <https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/index.php>

R Core Team. (2024). *R: A language and environment for statistical computing* [R]. Vienna, Austria: R Foundation for Statistical Computing,. Retrieved from <https://www.R-project.org/>

Rounds, E., Cate, N., McCallum, K., Beaty, M., & Mitchell, B. (2021). *First Order LiDAR Metrics: A Supporting Document for LiDAR Deliverables* (pp. 1–15). Salt Lake City, UT: U.S. Department of Agriculture, Geospatial Applications and Technology Center.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., & Xian, G. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 108–123. doi: 10.1016/j.isprsjprs.2018.09.006