A Methodology to Examine the Strengths and Limitations of using FIA Plot Data and LiDAR-based Structural Metrics for Continuous Mapping in the Sierra Nevada Ecoregion

Noah Krakoff

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2026

Committee:

Van R Kane

Susan Prichard

Hans-Erik Andersen

Maureen C Kennedy

Program Authorized to Offer Degree:

Environmental and Forest Sciences

University of Washington

Abstract

A Methodology to Examine the Strengths and Limitations of using FIA Plot Data and LiDAR-based Structural Metrics for Continuous Mapping in the Sierra Nevada Ecoregion

Noah Krakoff

Chair of the Supervisory Committee:

Van R Kane

Environmental and Forest Sciences

Model-based estimation with airborne LiDAR and the US Forest Service's Forest Inventory and Analysis (FIA) program is a common way to estimate above-ground carbon (AGC), basal area (BA), and wood volume (WV) in forested areas. This study investigates how well the FIA plots capture forested areas within the Sierra Nevada ecoregion and what structures and conditions they fail to capture for effective model-based estimation of AGC, BA, and WV. Random Forest models were used to model these metrics and had a high coefficient of determination (0.79, 0.70, and 0.79 respectively). However, due to strict FIA confidentiality requirements of the true plot locations, the results of this study use the publicly available FIA plot locations, which are guaranteed to be within 1 mile (1.6 km) of the true locations. Results from this study therefore demonstrate the methods, types of results and analysis that could be performed when the true, confidential plot locations are

available. Random Forest models with the public location had relatively low coefficients of determination (0.37, 0.34, and 0.36, respectively). The poor performance our models is likely due to the misalignment between remote sensing and plot locations. Using Meyer and Pebesma’s (2021) concept of area of applicability, the area to which a model can be reliably applied without risking overextrapolation, we determined the effectiveness of the FIA sampling scheme for representing structures across the ecoregion for modeling purposes. Based on an evaluation of area of applicability, FIA plots captured 96% of ecoregion for each forest metric, respectively. We found that we tall forests across a range of canopy covers are not well modeled by the FIA plots. These include very tall, very dense forests with dominant tree height greater than 45m and canopy cover percent more than 50% as well as forests with tall trees greater than 30 m where canopy cover is low. Forests that have recently experienced disturbance are also underrepresented. Model-based estimation of AGC, BA, and WV with FIA plots and LiDAR is effective across much of the ecoregion, but there are some systematic under sampled forest types, structures and conditions where our models may be biased. Overall, this study highlights the importance of quantifying the area in which a sample can be effectively applied for model-based estimation. Without proper quantification of where a map may be biased, researchers, forest managers, and policy makers may unknowingly use maps over areas where they may not be appropriate. Results of this study are preliminary and pending the use of the true FIA coordinates. Regardless, this study provides a valuable demonstration of a workflow to characterize how well a sampling scheme represents an area of interest.

Table of Contents

Table of Figures

Key study limitation

Because of a key limitation, this study should be considered a demonstration of methods to answer our research questions of interests, and all findings should be considered with this in mind. A requirement of using the Forest Inventory and Analysis (FIA) plots is that their true locations are not publicly available and can only be accessed by trained USFS personnel (LaPoint, 2005). Each FIA plot has a public fuzzed location guaranteed to be within one mile (1.6 km) of the true location (LaPoint, 2005). This study was conducted in close collaboration with Dr. Hans-Erik Andersen with the Pacific Northwest Research Station, who we relied on to implement the workflows for the analysis for the true locations, disruptions in the agency prevented any access to any spatial information. Due to the lack of direct access to the true FIA locations, the results of this study use the public fuzzed locations. For more information on how we expect the findings to vary with the true locations, see section 4.6. Our plan is to adapt these methods to the true coordinates when we have access to them.

1 Introduction

Forest ecosystems are highly dynamic over space and time. Long-term monitoring is key for determining the magnitude, direction, and pace of these changes (Lovett *et al.*, 2007). Forest inventories, a key monitoring tool, provide field-based measurements of tree density, diameter and height measurements, from which maps of forest metrics such as above-ground carbon (kg), wood volume (m^3), and basal area (m^2) can be generated. These maps can help managers and policy makers prioritize their efforts to suit a range of applications, including forest restoration strategies, informing sustainable resource management, carbon accounting and monitoring, and wildlife habitat restoration (Zielinski *et al.*, 2006; Lovett *et al.*, 2007; Pan *et al.*, 2011; North, 2012).

With recent widespread structural and compositional changes from drought and wildfires, the Sierra Nevada ecoregion exemplifies the importance of forest inventories for monitoring rapidly changing forest conditions. Fire exclusion throughout much of the 20th century enabled the growth of true fir (ex. White fir [*Abies concolor*]) and other fire-vulnerable tree species and allowed for forest expansion and densification (Collins, Everett and Stephens, 2011a; Lydersen *et al.*, 2013; Van Wagtendonk *et al.*, 2018). Recent droughts have increased vegetation stress and led to widespread tree mortality across the ecoregion with an estimated 48.9% mortality of mature trees on the Eldorado, Stanislaus, Sierra and Sequoia National Forests following drought years in 2014 - 2017 (Fettig *et al.*, 2019). Dense, fire-excluded forests and accumulated canopy and surface fuels have predisposed forested landscapes to large, high severity patches of fire, especially under a more extreme drought and fire environment due to climatic warming (Miller *et al.*, 2009;

Mallek *et al.*, 2013; van Mantgem *et al.*, 2013). Proactive treatments including forest thinning and prescribed burning have been shown to effectively mitigate fire severity and tree mortality associated with recent wildfires and drought (Pollet and Omi, 2002; Fulé *et al.*, 2012; Restaino *et al.*, 2019). Fuel reduction and forest restoration treatments require large investments of time and resources making it difficult to apply them across the entire ecoregion and they are not always appropriate for all forest types (Kolden, 2019; Prichard *et al.*, 2021). With spatially explicit forest inventories, forest managers can map and quantify the health of forests and prioritize treatments accordingly (Krofcheck *et al.*, 2019). Access to high-quality mapping products is therefore fundamental for the long-term adaptive management strategies to maintain and restore resilient forest conditions.

The USFS's Forest Inventory and Analysis (FIA) program offers one of the largest continuously monitored plot networks in the world (Tinkham *et al.*, 2018). A range of forest metrics are measured on these plots at 5-year intervals to allow for consistent inventory and monitoring (*Forest Inventory and Analysis | CAL FIRE*, 2026). Forest inventory estimates that are solely based on FIA field plot data are referred to as design-based. Design-based inference produces unbiased estimators of forest metrics because it assumes population totals are fixed. Uncertainty is a result of only randomness of the sampling design. These estimates have low uncertainty at resolutions where sufficient sample plots exist to estimate averages within each pixel (Lister *et al.*, 2020a). At smaller spatial scales, field plots can be too sparse to effectively support estimation, because the sample size within each pixel is too small to get a reliable estimation of averages.

Advances in remote sensing have allowed for more accurate and spatially explicit forest inventories at smaller spatial scales through model-based estimation, which predicts metrics of interest using field datasets as training data. Model based estimations are unbiased under the assumption that the model is applicable across the entire study area. As airborne LiDAR (light detection and ranging) scanning datasets have become more ubiquitous, model-based estimation of forest structural metrics using LiDAR data have become increasingly common. Specifically, airborne LiDAR scanning (ALS) can capture 3-dimensional (3D) forest structures at a high resolutions up to 30 m across large spatial scales, which forest inventories can leverage for better estimation of forest metrics such as above ground carbon, basal area, and wood volume (Lefsky *et al.*, 1999; Kane *et al.*, 2010; van Leeuwen and Nieuwenhuis, 2010). LiDAR-based forest inventories often follow a two-step inventory framework developed by Næsset, (2002). The first step is creating a model between ALS data and field plot measurements, and the second step is applying that model to the rest of the area of interest. This process enables the creation of high-resolution and spatially explicit maps of forest inventories.

Many forest inventory studies have been conducted using the two-step framework with ALS in the US. FIA plots have been shown to be effective for model development (Sheridan *et al.*, 2015; McRoberts *et al.*, 2018; Lister *et al.*, 2020a; Andersen *et al.*, 2021; Johnson *et al.*, 2022). Many studies have been conducted to evaluate the most accurate modeling approaches. Random Forest, linear regression, and support vector machine have all been found to be effective models, although there is no clear best model in every application (Johnson *et al.*, 2022; Tian *et al.*, 2023; Lamahewage *et al.*, 2025; Ullah *et al.*, 2025).

Auxiliary data, including climatic, topographic and satellite data, have also been found to improve modeling in some studies (Lister *et al.*, 2020a; Ullah *et al.*, 2025).

One major consideration in conducting model-based estimation is the specific sampling scheme of the dataset, including where and how field data are collected. For example, the FIA sample scheme overlays a network of hexagons across the continental U.S. and plot locations are selected within each hexagon at random with plots being measured on a 5-year rolling basis (Bechtold and Patterson, 2005). An optimal sampling scheme is one which adequately represents the range of variability across the study area. A poor sampling scheme, with inadequate capture of the range of variability, will create biased model predictions when the model is forced to extrapolate outside the range of its samples (Hawbaker *et al.*, 2009; Maltamo *et al.*, 2011; Tomppo *et al.*, 2014; Melville, Stone and Turner, 2015; Queinnec *et al.*, 2022; Goodbody *et al.*, 2023). For model-based estimation, estimates would be no longer be unbiased with poor sampling schemes because the assumption that the model is applicable everywhere is no longer ture. For this assumption to hold, sample plots are ideally stratified by forest type, structure, and condition to capture the full range of variability within forests. Some studies have leveraged the high spatial coverage of LiDAR to optimize plot location selection (Hawbaker *et al.*, 2009; Melville, Stone and Turner, 2015; Queinnec *et al.*, 2022). However, because FIA plots are already fixed, there is no way to optimize them for spatial coverage of an area of interest. Although the FIA plot network is extensive, the FIA sampling scheme is not stratified by forest type, structure or condition and thus does not necessarily represent an optimal sampling scheme for our area of interest.

Although the FIA sampling scheme is based on a systematic grid and not stratified by forest type, it is still one of the most comprehensive field-based datasets in the US (Tinkham *et al.*, 2018). The spatial and temporal consistency of the FIA plots mean they are the best choice for large-scale modeling and mapping (Lister *et al.*, 2020a). However, little research has been conducted to understand how well the FIA sampling scheme extrapolates to landscape-scale modeling and mapping. Specifically, no research has been conducted to determine how lack of stratification by forest structure, type and condition affects how well models can be extrapolated across a larger area.

In this study, we adapted a methodology to quantify the areas that the FIA plots fail to capture. Meyer and Pebesma (2021) proposed a method to determine the area in which a model can be applied without extrapolating based on the range of variability in the sample data. They define the area in which a model can be reliably applied as the area of applicability (AOA). Each pixel in a map is assigned a dissimilarity index (DI) based on its difference from the sample data. The AOA is calculated by thresholding this DI. AOA has been used in a variety of ecological models (Khanal et al., 2023) but has only been used once previously for LiDAR-based forest inventory models (Johnson *et al.*, 2022). we are not aware of any other LiDAR based forest inventory studies that assess whether their field plots accurately represent the entire area of interest.

Given the complex variety of structure and composition throughout the Sierra Nevada ecoregion, it is likely that the FIA's sampling design does not capture the full range of structure and composition. With an area of over 50,000 km^2, this ecoregion is large with steep elevational and precipitation gradients and rugged terrain (McNab and Avers, 1994).

The large and varied biogeography of the region creates a variety of climates and biophysical settings, supporting a diversity of forest types and a high complexity of forest structures and species composition (Sleeter, Wilson, and Acevedo, 2012).

In this study, we demonstrate how the AOA can characterize how well the range of forest structures in the FIA network captures the full range of structures present in Sierra Nevada forests. To do so, we created a forest inventory of the Sierra Nevada using the two-step inventory framework based on LiDAR and FIA plot datasets. In the first step, we created a random forest model to predict above-ground carbon (AGC), basal area (BA), and wood volume (WV) using continuous LiDAR data. For the second step, we applied this model to the entire Sierra Nevada ecoregion to create 30-m resolution maps of AGC, BA, and WV as well as maps of the DI and AOA for each metric. we used these maps to determine the strengths and weaknesses of the FIA sampling scheme for modeling my field metrics of interest using LiDAR. This study was guided by three related objectives to: 1) determine the effectiveness of the FIA sampling scheme for representing the structure captured from LiDAR for the entire ecoregion and within common forest types, 2) Investigate the types of structures not well sampled by the FIA data, and 3) determine how well our model capture the structure of recently disturbed areas.

2 Methods

2.1 Study Area

The Sierra Nevada ecoregion encompasses a 53,413 km^2 area of California (Figure 1). The region has a mostly Mediterranean climate with warm, dry summers and cool, moist winters (Van Wagtendonk *et al.*, 2018). Greater precipitation falls in the northern region while the southern region is generally drier (Van Wagtendonk *et al.*, 2018). The region is also characterized by steep elevational gradients, from 150 m in the foothills to 4,421 m at Mt. Whitney. The elevation changes are especially pronounced in the southern region

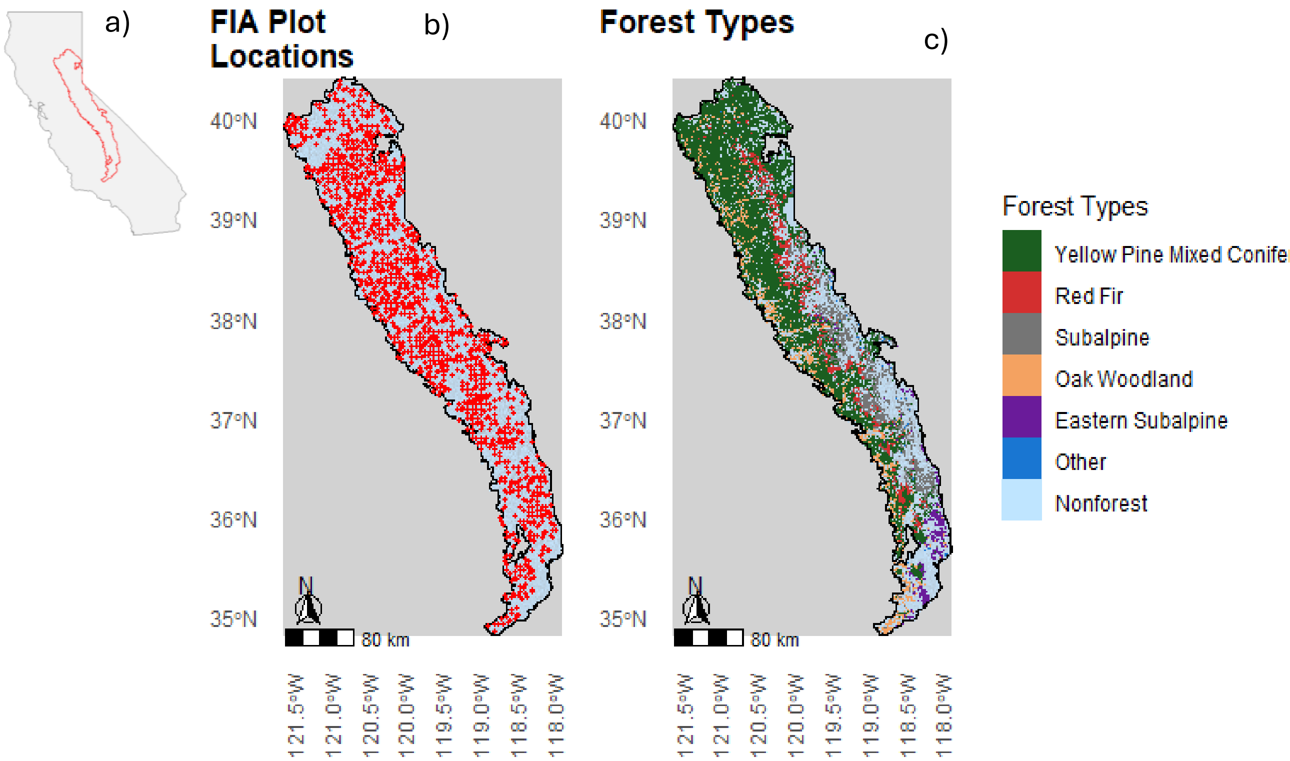


Figure 1: Study Area a) Location of the Sierra Nevada Ecoregion within California, b) Public location of the FIA plots after filtering c) Map of Forest Categories withing the ecoregion

resulting in more precipitation as snow compared to the northern regions (Van Wagtendonk *et al.*, 2018). The Sierra Nevada range creates a rain shadow effect that creates very dry conditions on their east slope. All these factors combine to create a climatically and topographically diverse region.

The climatic and topographic variability are strong drivers of forest composition throughout the region (Van Wagtendonk *et al.*, 2018). The lower western foothills slopes of the Sierra Nevada are primarily comprised of oak woodlands and chaparral shrublands.

Higher in elevation, the lower montane forests are dominated by yellow pine, mixed conifer (YPMC) forests. YPMC forests are primarily composed of ponderosa pine (*Pinus ponderosa*), Jeffrey Pine (*P. jeffreyi*), white fir (*Abies concolor*), incense cedar (*Calocedrus decurrens*) and California black oak (*Quercus kelloggii*) (Van Wagtendonk *et al.*, 2018). In the northern, wetter regions, Douglas-fir (*Pseudotsuga menziessi*) is common within YPMC forests. Above YPMC forests, upper montane forests are mostly composed of white and red fir (*Abies magnifica*). High elevation subalpine forests consist of lodgepole pine (*Pinus contorta*) and several species of white pine including foxtail pine *(P. balfouriana*), limber pine (*P. flexilis*) and whitebark pine (*P. albicaulis*). On the high elevation eastern slopes, the drier climate creates woodlands of pinyon pine (P. *monophylla*) and western juniper (*Juniperus occidentalis*). Across elevational gradients, broadleaf deciduous trees are common in riparian areas and including aspen (*Populus tremuloides*), willow (*Salix* spp.), cottonwood (Populus spp.), and alders (*Alnus* spp.).

For this analysis, I used the Classification and Assessment with Landsat of Visible Ecological Groupings (CalVeg) as a source layer of forest types across the ecoregion (USDA Forest Service, 2019). Using the Society of American Foresters (SAF) forest codes, forest types were grouped into six main categories: woodland, YPMC, red fir, subalpine, eastern subalpine and other (Figure 1c, Eyre, F. H., 1980). These forest categories roughly correspond with elevation and precipitation throughout the region (Appendix 8.3-N).

2.2 LiDAR Data

The USGS 3D Elevation Program (3DEP) program has collected high quality ALS acquisitions over the entire Sierra Nevada ecoregion (*3D Elevation Program | U.S.*

Geological Survey, 2025). Fourteen of these acquisitions were collected over a similar time in 2022 except for Yosemite National Park, where ALS was collected in 2019. All ALS acquisitions are designated as quality level 1 with ≥ 8 pulses per square meter, and a vertical RMSE less than 10 cm. A LiDAR processing tool called Lapis was used to create 30-m grids of LiDAR metrics (Table 1, Kane, 2025). These metrics were selected because they represent three major forest stand attributes important to management of tree height, canopy cover and tree height variation (Lefsky *et al.*, 2005; North, 2012). We selected multiple height metrics because studies have found different heights are better predictors of different forest metrics (White *et al.*, 2013; Queinnec *et al.*, 2022). Metrics were calculated with the assumption that points above 2 m were a part of the forest canopy (Næsset, 2002).

Table 1: Definition of LiDAR-based metrics in this study

LiDAR Metrics	Definition
p95	The 95th percentile height of points >2 m (m).
p50	The 50th percentile height of points >2 m (m).
p25	The 25th percentile height of points >2 m (m).
CV	Coefficient of variation of point height
CC	Canopy cover as measured by Percent of points above 2 m

2.3 Disturbance Data

Maps of past disturbances were generated using the landscape change monitoring system (LCMS) (Sean P. Healey and Zhiqiang Yang, 2023). LCMS provides yearly maps of forest loss across the continental United States at a 30 m resolution. Each forested pixel is classified into 4 groups: fast loss, slow loss, stable, and gain. For this study, we defined disturbance as pixels that were classified as fast loss. One potential limitation of the LCMS dataset is that it focuses on composite images from June 1st to September 1st of each year to determine disturbance, meaning that disturbance events that happen late in the year are often attributed to next year's change. One example that will be relevant in this study is it detects much of the over 100,000 ha of the 2013 Rim Fire in the Sierra Nevada as occurring in 2014.

2.4 FIA data

The Forest Inventory and Analysis (FIA) program surveys a network of field plots across the United States (Bechtold and Patterson, 2005). A hexagonal grid was overlaid across the continental US. Within each hexagon, which has an area of approximately 6,000 acres (2,428 ha), a point was chosen at random to be the center of a field plot. This type of sample is called a systematic unaligned sample. The selection process was not stratified by forest composition, structure, or condition. In the Sierra Nevada ecoregion, plots are measured on a 5-year rolling basis.

FIA plots consist of four subplots with a 24 ft (7.32 m) radius, one in the center and three surrounding the central plot (Appendix 8.1-A). Within each subplot, every tree with a diameter at breast height (DBH) greater than 5 in (12.7 cm) is measured. Each subplot contains a 6.8 ft (2.07 m) radius microplot where saplings (1 in ≤ DBH < 5 in) and seedlings are measured. Exact plot layout is detailed in Appendix 8.1-A. In California, each subplot also has a macroplot with radius 58.9 ft (17.9 m) in which trees greater with DBH greater than 24 in are measured; However, we did not include it the macroplot in this study. We extracted field measurements of above-ground carbon (AGC), basal area (BA), and wood volume (WV) per hectare for each plot as well as the dominant forest category.

To obtain the LiDAR metrics over FIA plots, the point cloud over the subplots of each plot was extracted from the LiDAR data (Appendix 8.1-A). Combining the point clouds over each subplot into one point cloud has been shown to be best for model development (Sheridan *et al.*, 2015). Each point cloud was processed using the R LiDR package to produce LiDAR metrics for each plot (Table 1, Roussel *et al.*, 2026). Notably, because this study uses public FIA locations, the LiDAR metrics produced are not aligned exactly to the FIA measurements.

We removed some FIA plots from the dataset to prevent biased model estimation. Nonforested plots were removed as our primary interest was in forested metrics. We removed any plots where disturbance (as measured with LCMS) occurred between field measurement and LiDAR measurement and plots that had significant noise in their LiDAR metrics. Any plots with <250 points above 2m, where p95 >70m or where sd > 40 m were removed due unstable and unrealistic measurements. After this plot filtering, the final dataset contained 1,260 total plots total for modeling (Figure 1b).

2.5 Random Forest Modeling

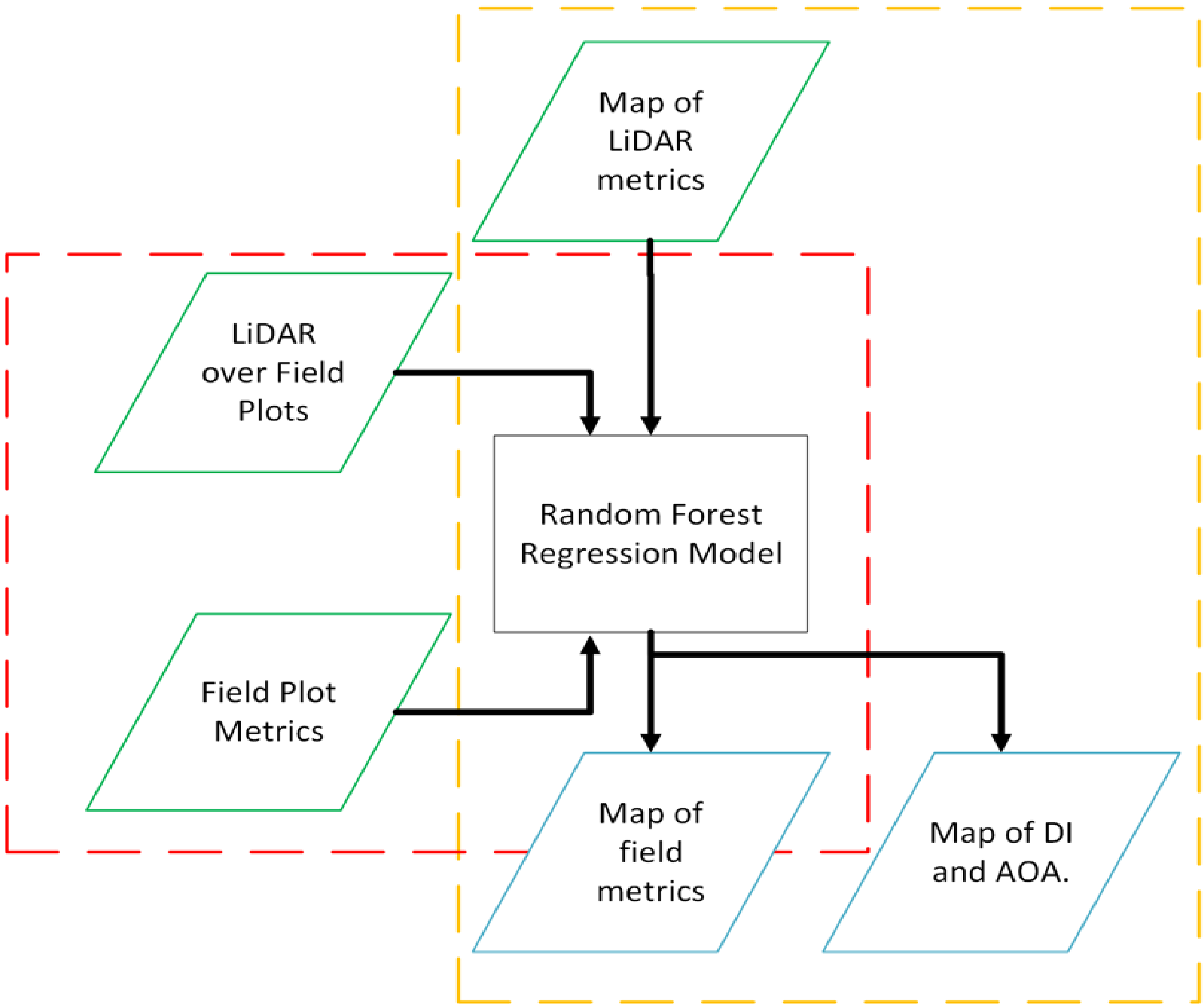


Figure 2: Flowchart of the two-step forest inventory framework. Green boxes are inputs. Blue boxes are outputs. The red dotted box is step 1, the training step. The orange dotted box is step 2, the prediction step.

To create a map of forest metrics across the entire ecoregion, we used a two-step inventory framework (Figure 2). In the first step, we trained a random forest model to predict forest inventory plot metrics based on LiDAR metrics. Random forests is a common machine learning method that has been used effectively in the modeling of forest metrics using ALS data (Cutler *et al.*, 2007; Ullah *et al.*, 2025). Random Forest regression was chosen for this study because it simplified computation (see section 2.6). Before training our model, we reduced our predictor variables to avoid Spearman correlations above 0.6 to allow for interpretation of variable importance plots and to reduce collinearity (Nicodemus *et al.*, 2010; Gregorutti, Michel and Saint-Pierre, 2017). To do this, we removed all predictors with correlations above 0.6 and then added them back, one by one, by selected the variable that had the highest correlation to all other removed variables until no more could be added while keeping correlations below 0.6. To fit our random forest model, we used the R Ranger and caret package with ten-fold cross validation to optimize the number of predictor variables at each split and the minimum number of observations at each terminal node (Meyer *et al.*, 2026; Wright, Wager and Probst, 2026). The second step of the two-step framework is to apply the model across the Sierra Nevada ecoregion using the processed 30-m LiDAR metrics. Any pixels that were not forested as indicated by our CALVEG layer were masked out of the geospatial dataset.

2.6 Computing DI and AOA

To create maps of the dissimilarity index (DI) and area of applicability (AOA), I followed the methodology described in Meyer and Pebesma, (2021), which was developed specifically for random forest regression. We assigned each pixel a DI, calculated by the Euclidean distance to the nearest sample point in statistical space, weighted by variable importance extracted from our random forest models. This value was normalized by dividing by average pairwise sample distance (Equation 1). A DI greater than one means the new pixel is further from its nearest sample point than the average distance of all the sample points. When a pixel's DI was larger than a certain threshold, we classified it as outside the AOA. The threshold was determined by the 10-fold cross-validation step of the random forest model. DIs were calculated for each point outside the fold, and the threshold was set as 1.5 times the interquartile range of this distribution, a standard statistical rule for detecting outliers.

Equation 1: Calculation of Dissimilarity Index

$$\mathrm{X_{j,k}} \coloneqq \text{Normalized ALS metric, k, for sample FIA plot } j$$

$$\mathrm{P_k} \coloneqq \text{Normalized ALS metric, k, for a pixel, P}$$

$$\overline{\mathrm{D}} \coloneqq \text{average pairwise euclidian distance of ALS metrics for sample plots}$$

$$w_k \coloneqq \text{weight of ALS metric, k, as determined by the model}$$

$$\mathrm{Dissimilarity(P)} = \frac{1}{\overline{\mathrm{D}}} \min_{\mathrm{j=1,\ldots,n}} \sqrt{\sum_{\mathrm{k=1}}^{\mathrm{K}} w_k \left(\mathrm{X_{j,k}} - \mathrm{P_k}\right)^2}$$

2.7 Objective 1: Representation by Forest Type

To quantify the effectiveness of the FIA sampling scheme for representing the structure captured from LiDAR for the entire ecoregion and within each forest type, we investigated the size and location of the AOA. First, we calculated the percentage of pixels that fell outside the AOA for each forest metric. We then calculated the percentage of pixels that fell outside the AOA by forest type and forest metric. These data can be compared without a statistical test because they are population-based metrics.

2.8 Objective 2: Underrepresented Structures

To determine the types of structures not well sampled by the FIA dataset, we investigated the percentage of area outside the AOA for the structures found across the Sierra Nevada ecoregion. We focused on two main structures measured from lidar, 95th percentile canopy height (p95) and canopy cover (CC), as they are highly relevant forest management metrics and the most relevant metrics in our model (Hessburg *et al.*, 2020; Moran, Kane and Seielstad, 2020). We then split each metric into bins. The bins for p95 were stratified by 1m increments from 0-1m to 69-70 m. Canopy cover was incremented as integer percentiles (e.g., 0-1% to 99-100%). These bins were large enough to have many sample pixels with each structures from the ecoregion within each bin but small enough for us to investigate nuances in structure. For every combination of bins for p95 and canopy cover, we calculated the percentage of pixels in those bins that are outside the AOA.

2.9 Objective 3: Underrepresented Disturbance

To determine how well the FIA plots capture the structure of recently disturbed areas, I evaluated how disturbance history effects the DI of each pixel. Using LCMS for disturbance detection, I calculated the average DI for all pixels that did not experience disturbance within the years 2012 to 2022. I then classified each pixel by year of most recent disturbance for the sampled decade and calculated the average DI by year. I repeated this process to calculate the percentage of pixels outside the AOA by disturbance year.

3 Results

3.1 Modeling Results

Lidar-based height metrics (p95, p50, and p25) correlation were beyond our 0.6 threshold for inclusion in our model (Figure 3a). All other metrics we considered were not as highly correlated and could be included in our model without risking overfitting (Figure 3a). The variable selection step selected 95th percentile height (p95), canopy cover (CC) and coefficient of variation (CV) for use in our model.

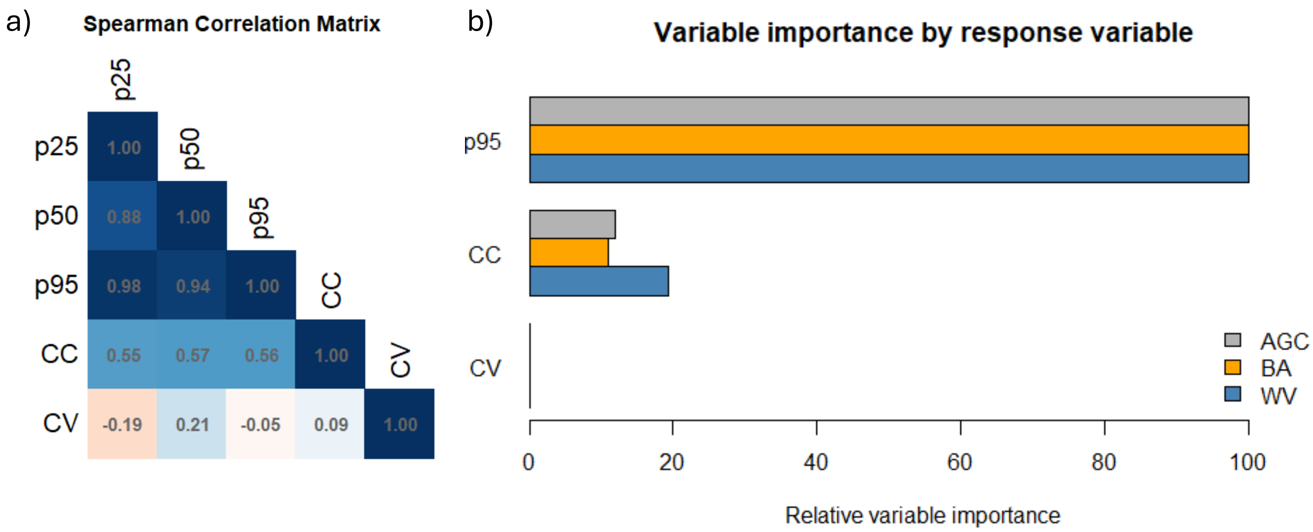


Figure 3: Model Results a) Spearman correlation matrix for the LiDAR metrics across the Sierra Nevada Ecoregion. b) Variable importance plots for Above ground carbon (AGC), basal area (BA) and Wood Volume (WV) as calculated with the public data

Our collaborator was able to assess the random forest regression models for AGC, BA, and WV using the true coordinates which had an R^2 of 0.79, 0.70, and 0.79 respectively. However, given the restrictions to accessing the true FIA plot coordinates, we were unable to complete the analysis of the AOA using the actual plot locations. The following results use the public locations and will serve as a demonstration of methods. Overall, RF models using the public locations had much lower coefficients of determination (R^2 = 0.37, 0.34, and 0.36 respectively). From the variable importance rankings of the random forest models with the public coordinates, 95th percentile height (p95) is most important for all three metrics (Figure 3b). In decreasing variable importance, the model then ranks percent canopy cover (CC), and coefficient of variation (CV) (Figure 3b). CC has some importance but much less compared to p95. CV is not relevant to any metric.

3.2 Objective 1: Representation by Forest Type

Although the FIA plots sample the overall ecoregion, they do not completely represent the entire area. Area of applicability (AOA) is high for all metrics with 96% of the region within the AOA for AGC, BA and WV (Figure 4a). These percentages reflect the DI threshold calculated as 0.064, 0.073 and 0.059 respectively. The areas outside the AOA are concentrated in a small number of forest types. Despite having the most samples, yellow pine mixed conifer (YPMC) forests have the highest percentage of pixels outside the AOA with 5.2%, 5.3%, and 5.5% for AGC, BA and WV respectively (Figure 4b,c). Red fir forests

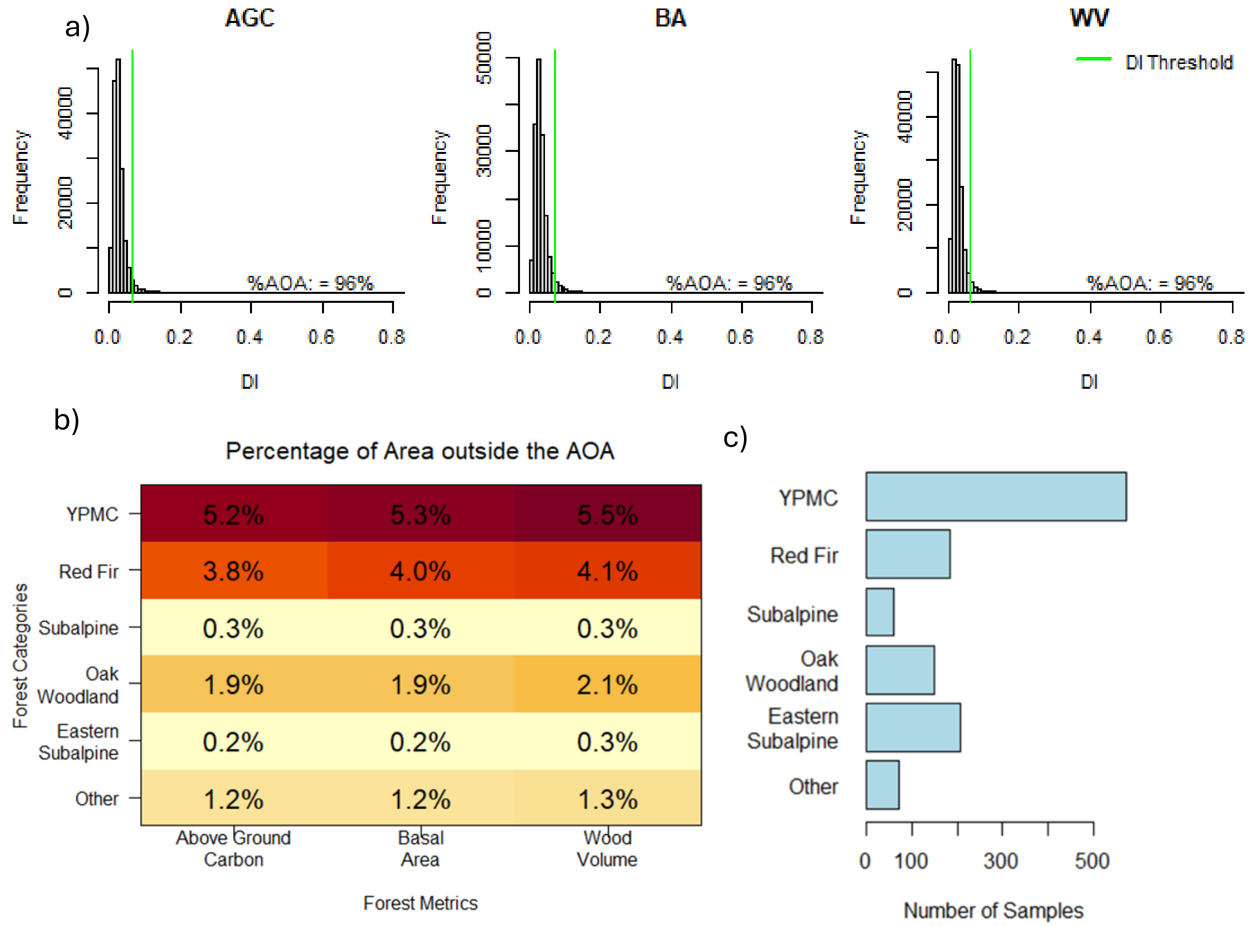


Figure 4: Representation of the Sierra Nevada Ecoregion a) Histogram of Dissimilarity Index (DI) for each forest metric, Aboveground carbon (AGC), Basal Area (BA) and Wood Volume (WV). The green line represents the threshold for being outside the AOA. b) The percentage of pixels outside the AOA by forest type for each forest metric. c) The number of FIA samples we used for modeling for each forest type. Figures are produced from the public FIA locations

have the fewest samples of the forest categories but also have a large percentage of the region outside the AOA with 3.8%, 4.0% and 4.1% (Figure 4b). Oak woodland and the other category have around 2% and 1.2% of the region outside the AOA for each metric (Figure 4b). Subalpine and eastern subalpine have a low (<0.3%) percentage of the region outside the AOA (Figure 4b). Although each forest category has a markedly different number of samples, the number of samples does not appear to affect the AOA percentage (Figure 4c). We found that for forest type the AOA does not appear to differ significantly for AGC, BA, and WV.

3.3 Objective 2: Underrepresented Structures

Pixels outside the AOA are more common with specific values of LiDAR metrics. 95th percentile height (p95) has the largest influence on whether a pixel falls outside the AOA. As p95 increases, the percentage of area outside the AOA increases exponentially for AGC, BA and WV (Figure 5a, Appendix 8.3-A). As a result, when p95 heights are under 30m, less than 1% of pixels are outside the AOA and conversely when p95 heights are greater than 55 m, over 95% of pixels are outside the AOA; However, only about 0.15% of the total pixels have p95 greater than 55m and less than 0.01% have p95 greater than 60m (Figure 5b).

When p95 is between 30 m and 55 m, canopy cover percent (CC) influences whether that pixel falls within the AOA (Figure 5a). Specifically, when p95 is greater than 30 m and less than 40 m, pixels fall outside the AOA when canopy cover percent is less than 25% (Figure 5a). When p95 is between 40m to 45 m, pixels that fall outside the AOA can have CC percent up to 50% (Figure 5a). Between 45 m and 55 m, pixels that fall outside the AOA have CC percent less than 50% or greater than 85%. When p95 is below 30 m or above 55 m, canopy cover has very little effect on the AOA. For all the forest metrics we predicted, the structures that AOA represents are similar (Appendix 8.3-B).

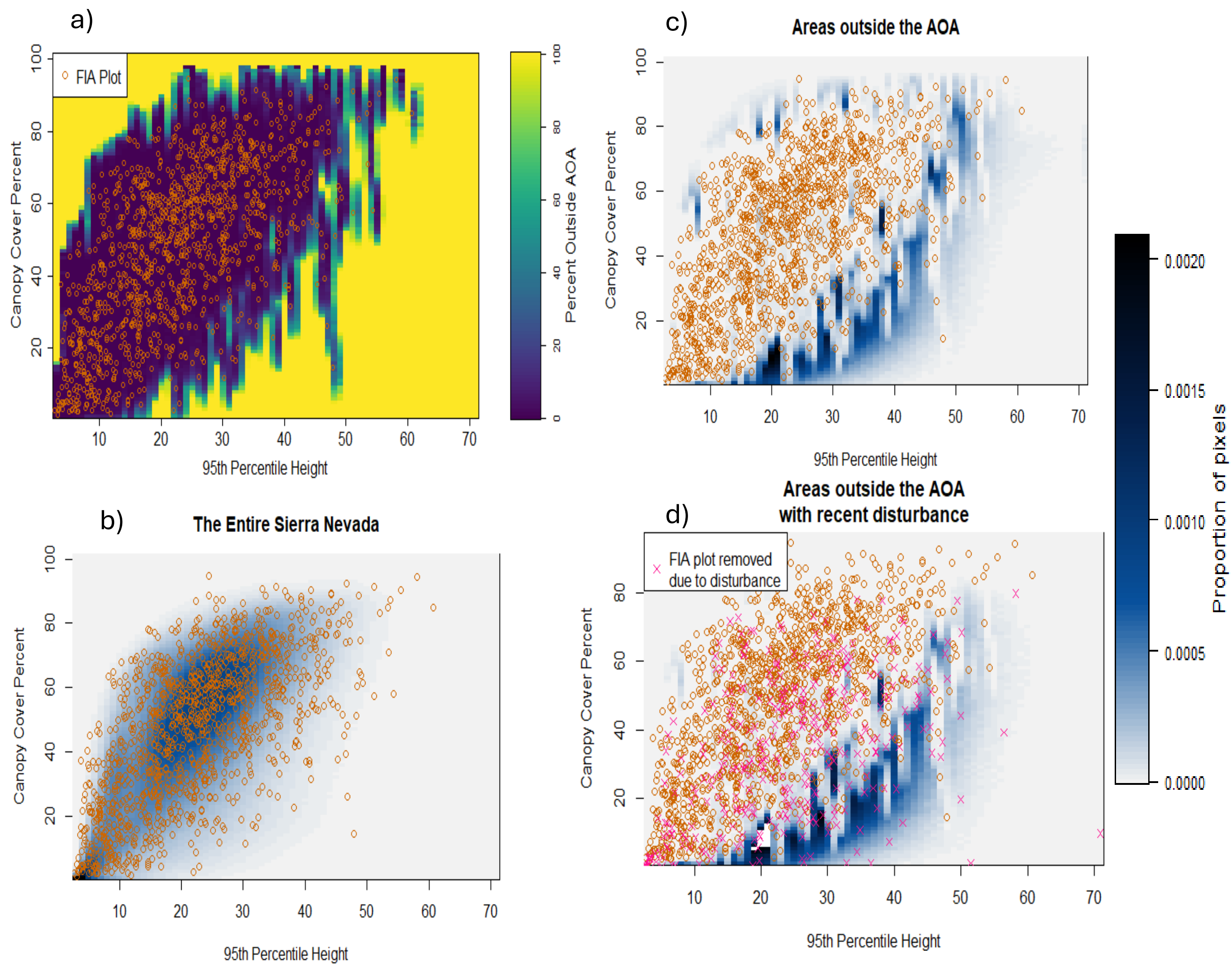


Figure 5: Underrepresented Structures a) A plot of the percent of pixels that falls outside the AOA for different combinations of 95th percentile height and Canopy Cover percent. Orange circles are the 95th percentile height and canopy cover measurements of FIA plots used in our model. b) A 2D density plot of the frequency of pixels with each combination of Canopy Cover percent and 95th percentile height for all pixels across the Sierra Nevada. FIA plots are plotted in orange c) A 2D density plot of the frequency of pixels outside the AOA. FIA plots used for modeling are plotted as orange circles d) A 2D density plot of the frequency of pixels outside the AOA and had recent disturbance. FIA plots used for modeling are plotted as orange circles and FIA plots that had to be removed due to disturbance between field measurement and LiDAR measurement are plotted as pink x's. All plots were created using the above ground carbon's AOA and the public FIA data

The distribution of FIA plots across p95 and CC percent is similar to the distribution of pixels for the entire Sierra Nevada ecoregion (Figure 5b). However, we found two major structures that appear frequently in the ecoregion and do not have a similar FIA plot to represent them and therefore fall consistently outside the AOA (Figure 5c). The first common structure has p95 greater than 30 m and CC less than 50% (Figure 5c). As an example, we found two forest locations with this structure listed in example 2 and 4 of Figure 7. Pixels with this structure account for more than 40% pixels that are outside the AOA. The second structure is characterized by p95 > 45m and CC >50% (Figure 5c). Over 15% of pixels outside the AOA have this structure. Three forests of this structure are listed in examples 1,3 and 5 of Figure 7.

3.4 Objective 3: Underrepresented Disturbance

Disturbance history has a significant effect on the AOA. Pixels that have experienced disturbance from the year 2012 to 2022 had an average DI of 0.039 and a percent of the region outside the AOA of 9.2%. Pixels without disturbance had a much lower average DI of 0.026 and lower area outside the AOA at 1.97%. The percentage of pixels outside the AOA from the most recent disturbance years (2021 and 2022) is 10.8% and 12.2%, respectively, which are much higher than most other disturbance years (Figure 6c). These years correlate with a high number of FIA plots removed from the modeling process due to recent disturbance (Figure 6a). For most years this correlation between number of sample plots removed and area outside the AOA is strong (Figure 6a). However, 2014 (mostly representing recently disturbed pixels in the 2013 Rim fire) has the largest percent of pixels outside the AOA with 11.7% which is an exception to this pattern (Figure 6c).

Recently disturbed pixels that are outside the AOA share a common set of forest structural attributes, including a 95th percentile height (p95) greater than 30 m and canopy cover percent less than 50%. These characteristics account for more than 60% of the pixels that both fall outside the AOA and have experienced recent disturbance (Figure 5d). When we removed FIA plots due to disturbance between field measurement and LiDAR measurement, we removed 63 plots that had this common structure (Figure 5d).

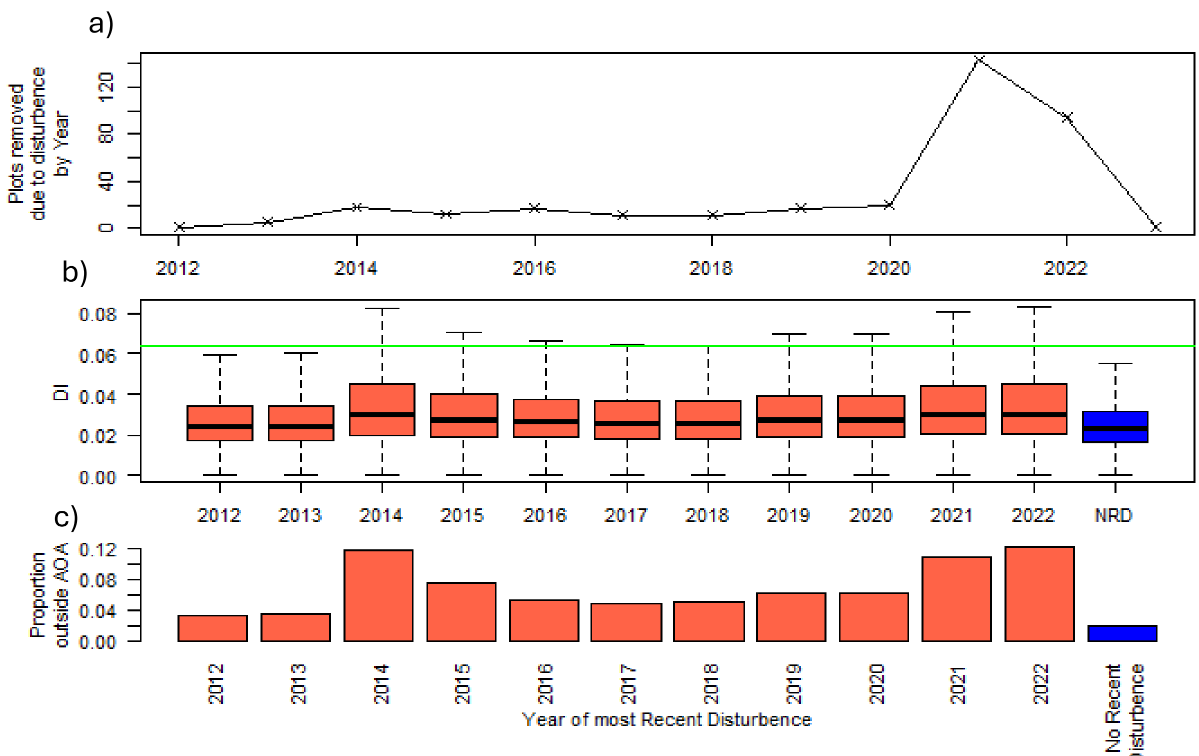


Figure 6: Underrepresented Disturbance a) The number of plots removed due to disturbance occurring at that plot between field measurement and LiDAR measurement. The x-axis is the year that the disturbance occurred at that plot as measured by LCMS. b) Distribution of the Dissimilarity Index (DI) for above ground carbon (AGC) for pixels which experienced recent disturbance. Pixels are grouped by the year they most recently experienced disturbance as measured by LCMS. If there was no disturbance in the listed 10 years that pixel was classified as no recent disturbance. (NRD) c) Proportion of pixels outside the AOA for pixels grouped by the year they most recently experienced disturbance. Figures are produced on the AOA of above ground carbon and from the public FIA locations.

4 Discussion

Given that we used the public fuzzed locations of the FIA plots for our analysis, this study should be considered a demonstration of methods to answer our research questions of interest, and all findings should be considered in this context. We do not know if this AOA analysis would yield similar results using the actual plot locations; However, the AOA method offers a promising approach to analyze sample representation and has not been applied to the FIA plots and LiDAR-based forest inventories to this extent. We demonstrated with this study that the AOA is be a useful methodology, and this discussion is an example of implications that can be drawn with this method.

In this study, we mapped aboveground carbon (AGC), basal area (BA), and wood volume (WV) using LiDAR and FIA plots with the two-step inventory method across the Sierra Nevada ecoregion. Our central goal was to evaluate where our model could be effectively applied without risking extrapolation based on the underlying FIA sample data. We were successfully able to utilize the Area of Applicability (AOA) method described in Meyer and Pebesma (2021) to determine which pixels in our maps could be reliably predicted. We also investigated common structures that were outside the AOA, which include structures not well modeled by the FIA data and LiDAR in which prediction of AGC, WV, and BA may be biased.

Although most of the structure of the Sierra Nevada ecoregion is well represented by the FIA plots for modeling purposes, there are some forest types, structures and conditions that are under sampled. YPMC and red fir forests have the highest proportion of area not well represented of all the forest types. Structurally, models trained with the FIA plots poorly model areas with high 95th percentile height across a range of canopy covers. We also found disturbances, such as fire, insects, and drought, are difficult to model with LiDAR and the FIA data. The more recent the disturbance occurred, the less likely our model can reliably predict forest metrics. The FIA sample is best at modeling montane forests that have low dominant tree height, moderate canopy cover, and are not recently disturbed. In more structurally complex forest stands (like stands with taller trees), we find the FIA plots are not as representative.

4.1 Model Performance

Performance of our random forest regression model on the true FIA plot locations corresponds with other FIA and LiDAR-based modeling studies (Tompalski *et al.*, 2019; Johnson *et al.*, 2022; Queinnec *et al.*, 2022; Adhikari, Montes and Peduzzi, 2023). Unsurprisingly, the performance of our model on public locations is substantially lower than models based on actual plot locations because the exact location the FIA metrics were measured differ from the location where LIDAR metrics were measured. Even so, our models had some predictive power and provided variable importance which aligned with other LiDAR-based forest inventories allowing us to evaluate the effectiveness of our sample.

Using models based on publicly available plot locations, we found that LiDAR-based height metrics were the most important variables for predicting our metrics and that canopy cover, while still important, has less effect than height metrics. Our findings of the importance of height in predicting AGC, BA, and WV are consistent with other studies and in line with what we would expect because larger trees have more disproportionally more AGC, BA, and WV (Lutz *et al.*, 2018a; Mildrexler *et al.*, 2020; Queinnec *et al.*, 2022; Adhikari, Montes and Peduzzi, 2023). Although greater canopy cover is often associated with higher AGC, BA, and WV, this relationship can be overshadowed by tree height (Mildrexler *et al.*, 2020). We also evaluated variability in tree height, represented by the coefficient of variation (CV) and found that it did not contribute to stronger models (Hudak *et al.*, 2020; Queinnec *et al.*, 2022). CV describes relative variability of height, but the specific metrics chosen for our study (AGC, BA, and WV) are more strongly correlated to actual height distributions (Lutz *et al.*, 2012, 2018b; Mildrexler *et al.*, 2020).

4.2 Objective 1: Representation by Forest Type

YPMC and red fir forests exhibiting the largest percent of their area outside the AOA is consistent with structures that fall outside the AOA. Structurally, one of the major characteristics of a forest being outside the AOA is when 95th percentile height is greater than 30 m (Figure 5a). Similarly, areas with recent disturbance are also outside the AOA (Figure 6). YPMC and red fir are both highly productive forest types that are more consistently above 30m. These forest types also have been most impacted by recent disturbances, including fire, insects, and drought-induced tree mortality (Van Wagtendonk *et al.*, 2018; Fricker *et al.*, 2019; Williams *et al.*, 2023). The structures and conditions that are outside the AOA are most found in these forest types, therefore their proportion of area outside the AOA is higher.

It is somewhat surprising that the subalpine, oak woodland, eastern subalpine, and other forest types have such low percentage of area outside the AOA. These forests categories have a much smaller FIA sample size than the YPMC (Figure 4a). It appears that FIA sample size does not necessarily affect the percentage area outside the AOA for each forest category. YPMC and red fir forests have more vertical structural variation than other forest categories, which appears to drive lack of representation more than sample size (Van Wagtendonk *et al.*, 2018; Fricker *et al.*, 2019; Huesca *et al.*, 2019). Conversely, forest categories with comparatively lower horizontal structural variation (e.g., subalpine, woodland, eastern subalpine, and other) have a lower percentage of area outside the AOA (Van Wagtendonk *et al.*, 2018; Huesca *et al.*, 2019).

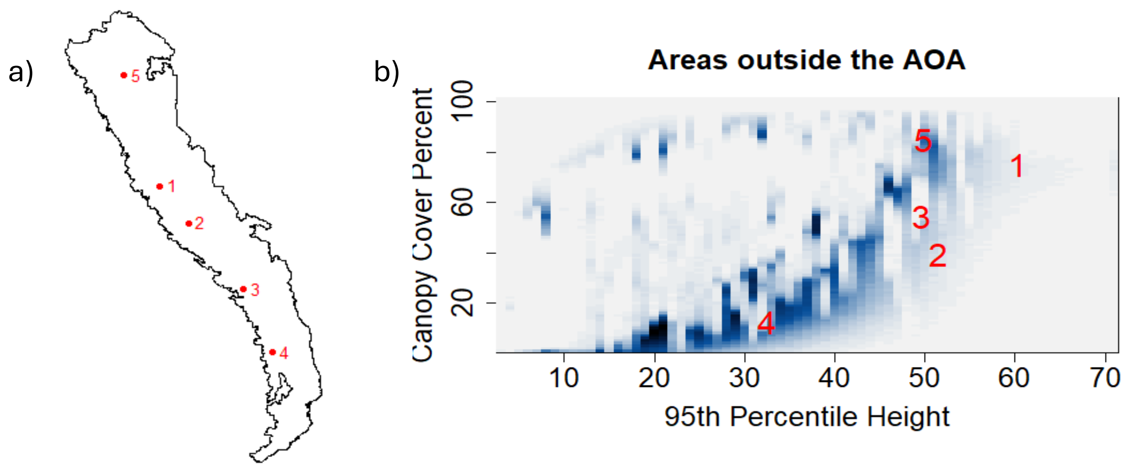
4.3 Objective 2: Underrepresented Structures

It appears that the FIA sampling scheme is highly effective at modeling the most common structures in the ecoregion with high model coefficients of determination (>0.7). In the Sierra Nevada ecoregion, the most common forest stands are characterized by moderately sized trees (trees under 30m) with a moderately dense canopy (35 – 80% canopy cover) (Dolanc *et al.*, 2014; Stephens *et al.*, 2015). Forested stands with these structures fall within the AOA, meaning our models extrapolate predictions less in those forests (Figure 5a). Our findings make sense given that studies have shown that common, simple forest structures have less associated prediction error (Woods, Lim and Treitz, 2008; van Ewijk *et al.*, 2019). Likely, the lower prediction error is a result of the high proportion of these simple forest structures found on the landscape and are sampled m.

These forest structures are common throughout the Sierra Nevada are often associated with industrial forest management and fire suppression (Collins, Everett and Stephens, 2011b; Dolanc *et al.*, 2014; Easterday, McIntyre and Kelly, 2018).

Forests with very tall trees (>45 m) and high canopy cover (>50%) are more likely to fall outside of the AOA (Figure 5a). Tall, dense structure is atypical (~1.5% by area) to the ecoregion but can occur in the northern part of the region where wetter conditions contribute to high forest productivity (Van Wagtendonk *et al.*, 2018). For example, we found that old growth sequoia stands and unburned tall dense YPMC stands have this structure and fall outside the AOA (Figure 7.1 and 7.5). Forests with large trees are highly important to forest managers and policy makers because large trees generally represent old forests with highly valued structure and ecosystem function (Lutz *et al.*, 2012, 2018b; Hessburg *et al.*, 2020). This mapping limitation has consequences for mapping very tall trees and therefore to adaptive forest management strategies. For example, when large trees exist in dense forests, they are more vulnerable to fire and drought driven mortality (Miller *et al.*, 2009; Mallek *et al.*, 2013; Fettig *et al.*, 2019). Proactive management through thinning and prescribed burning are critical to maintaining the health and resilience of these trees (Pollet and Omi, 2002). Thus, providing unbiased forest metric estimates for these stands is of critical importance for management and policy.

Our modeling also does not adequately represent areas with large dominant trees and low canopy cover. This structure type often falls outside the AOA when dominant tree height is > 30m and canopy cover is < 50% (Figure 5c). Forests with this structure can be found in healthy YPMC forests where fire has been reintroduced such as in Teakettle



c)

	1.	2.	3.	4.	5..
Description	An old growth giant sequoia (*Sequoiadendron giganteum*) grove located in Calaveras big trees state park	Forest burned within Yosemite national park by the Rim fire in 2013. Standing Snags are still present.	Located in Teakettle Experimental Forest, this Yellow Pine/Mixed Conifer forest sees extensive management to create resilience and ecosystem functions.	Yellow Pine/ Mixed Conifer Forest burned in the 2020 Castle Fire. Only Snags appear to be standing.	A tall, dense stand of Yellow Pine/ Mixed Conifer forest located in Tahoe National Forest
95th Percentile Height (m)	60.26	51.51	49.60	32.50	49.90
Canopy Cover Percent	75.1%	39.5%	54.5%	12.7%	85.2%
2022 NAIP Imagery					

Figure 7: Example forests that fall outside the AOA. a) The Location within the Sierra Nevada ecoregion where these examples are found. b) Density plot of pixels outside the AOA by 95th percentile Height and Canopy Cover with examples plotted. This is the same density plot found in Figure 5c. c) Table of examples with a description, 95th percentile Height, Canopy Cover Percent, and NAIP Imagery.

Experimental forest (Figure 7.3) or forests with large residual snags from high severity fire (Figure 7.2 and 7.4, Van Wagtendonk *et al.*, 2018). Notably, since 2022 many of these examples have already changed signi ficantly including the Teakettle experimental forest which burned at high severity in 2025. Because these forest structures are important for future resilience and wildlife habitat, they would benefit from unbiased mapping (Lutz *et al.*, 2012).

My evaluation of AOA highlights how model-based estimation with FIA measurements may underrepresent large trees for the mapping and monitoring of AGC, BA, and WV. Policy makers, forest managers, and scientists are interested in assessing the total amount of these forest metrics at landscape scales the size of watersheds, counties, and ecoregions (Gillespie, 1999; Lovett *et al.*, 2007; Pan *et al.*, 2011). One of the major goals of FIA program is to effectively map and monitor these totals (*Forest Inventory and Analysis Strategic Plan | US Forest Service Research and Development*, 2023). However, our model is biased in forest stands with large trees and dense canopies because they are rare forest conditions that can be extremely variable. Despite being rare, these forests contain a disproportionate amount of forest AGC, BA, and WV compared to smaller, less dense forests (Lutz *et al.*, 2012; Stephenson *et al.*, 2014; Mildrexler *et al.*, 2020). To estimate forest metric totals, forests with large trees, and dense canopies need to be predicted accurately to avoid biasing these totals significantly compared to their prevalence on the landscape.

Because the FIA plots use a regular sampling technique, common structures will be sampled more than rare structures. We see this behavior when the distribution of structures sampled by FIA plots matches the distribution of structure seen on the landscape (Figure 5b). For example, the proportion of FIA plots with canopy cover greater than 80% and 95th percentile height greater than 50m matches the proportion of that structure seen on the landscape. As a result, extreme structures are sampled at an appropriate frequency relative to their appearance on the landscape. However, our findings suggest that these extreme structures require more samples than the systematic FIA sampling scheme supports. A sampling scheme stratified by structure would provide more samples at these extreme structures and reduce the dependence of AOA on forest structure.

4.4 Objective 3: Underrepresented Disturbance

When using the FIA samples, the two-step LiDAR based inventory method does not effectively model forest stands that recently experienced disturbance. Because the sampling scheme of the FIA plots measures plots on a 5-year rolling basis, many plots had to be removed from the dataset due to disturbance (*Forest Inventory and Analysis | CAL FIRE*, 2026). Plots that had significant change between when forest metrics were measured to when LiDAR metrics were taken could not be used to train our model as they would add bias due to this temporal mismatch. However, in avoiding this bias, we systematically removed plots with recent disturbance; (Figure 5d). The reduced total number of FIA plots, combined with the complex and unique structure associated with recently disturbed areas, creates a high proportion of pixels outside the AOA (Kane *et al.*, 2019).

The high proportion of pixels outside the AOA are particularly pronounced for more recently disturbed areas disturbed in 2020 and 2021 (Figure 6c). Multiple, uncharacteristically large fires occurred in these years including the 2020 creek fire, 2021 dixie Fire and 2021 caldor fire and since the LiDAR was acquired in 2022, temporal mismatching was very common for plots within these fire footprints (Keeley and Syphard, 2021; Taylor, Harris and Skinner, 2022). As time since disturbance increases, the number of plots requiring removal due to temporal mismatch decreases.

Interestingly, areas disturbed in 2014 appear to be an outlier because they are not well represented by the data despite not having many plots removed that year (Figure 6c). Our hypothesis is that this outlier is due to post-fire management differences for the 2013 Rim Fire which LCMS classifies as occurring in 2014 (see 2.3) ("Rim Fire, California - NASA Science," 2013). Post-fire management was very different for large portions of the Rim Fire due to its occurrence in Yosemite National Park (United States: Agriculture Department: Forest Service *et al.*, 2014). My hypothesis is that this created structure unlike what is seen in the rest of the Sierra Nevada ecoregion which would require more plots to effectively sample (Figure 7 ex 2).

Our model's failure to adequately predict over areas that recently experienced disturbance has major implications for forest mapping and monitoring. Postfire mapping is important to understand how fire changes the landscape at a variety of scales (Lentile *et al.*, 2006; Chu and Guo, 2014; Garcia *et al.*, 2017). Consistent post-fire monitoring has also become extremely important to understanding how different post-fire management strategies effect structure (Robichaud, Beyers and Neary, 2000; Chu and Guo, 2014). With the FIA sampling scheme and our LiDAR based two-step inventory method, we found that areas experiencing recent disturbance have a higher percentage of area outside the AOA; Therefore, our maps may be unreliable at these locations and are not appropriate for post fire mapping and monitoring. Because the Sierra Nevada ecoregion is experiencing increasingly more frequent disturbance at large scales, the area outside the AOA will likely increase over time (van Mantgem *et al.*, 2013; Fettig *et al.*, 2019; Williams *et al.*, 2023).

4.5 Study Implications

The lack of stratification by forest type, structure, or condition by the FIA sampling scheme strongly influences maps produced through model-based estimation in specific forest types, structures and conditions. This is unsurprising as many studies have shown that different sampling methodologies will bias models in different contexts (Ståhl *et al.*, 2024; Kangas, Myllymäki and Packalen, 2025). Many large-scale modeling and mapping studies that use the FIA plots assume they are well representative of their area of study including Sheridan *et al.*, 2015; Huang *et al.*, 2018 and Riley, Grenfell and Finney (2016). In the Sierra Nevada ecoregion, we found that this assumption held for a large majority of the area, but that there were some notable exceptions. We expect that most maps constructed using model-based estimation will have similar areas with biased model performance. Our results compliment studies that show that selecting plots through stratification of structure produce better models (Hawbaker *et al.*, 2009; Melville, Stone and Turner, 2015; Queinnec *et al.*, 2022).

Despite the lack of stratification, FIA plots are still the most comprehensive dataset for large-scale modeling (Tinkham *et al.*, 2018). The FIA sampling scheme has many benefits such as a robust sample size and continues spatial and temporal coverage and its sampling design is likely the most robust for a forest inventory where hundreds of attributes are estimated (Bechtold and Patterson, 2005). FIA plots will remain incredibly valuable as source datasets for large-scale mapping, but it is important to quantify where their models and maps may be misleading because of a lack of data, particularly within landscapes with high structural diversity, increased area burned by high severity fires and high impacted of drought, insects and pathogens.

For managers who want use maps of AGC, WV and BA produced in this study to inform their management practices, they should be aware of the possibility that certain forest types, structures and conditions are not well mapped. If managers are primarily interested in forest metrics for forest stands with large trees across a range of canopy covers or in recently disturbed areas, they should be aware that these maps may not best suit their goals. If managers wish to use maps produced in this study, first they can use the maps of AOA to determine if their study area is well represented. If they find that their study area has much of its pixels outside the AOA, other sources and methods may be better suited to provide their forest metric information.

4.6 Limitations and Future Work

Our most obvious limitation to our study was the lack of access to true FIA locations, which significantly weakened the predictive power of our random forest regression model.

The major way the AOA will be affected is through changes in how important our model ranks LiDAR metrics, thus changing their weighting for calculating the AOA. We are able to access variable importance plots from the true data because it provides no spatial information, however these cannot be used to calculate the AOA because the dissimilarity index threshold is not tuned to these values. Comparing the variable importance plots, we see true locations have a different variable importance, although these differences are not extreme (Appendix 8.2-B). One major difference between models based on true vs. publicly available locations is that canopy cover percent has a much higher variable importance in true location models. Although we were not allowed to analyze spatial datasets produced by true location models, our hypothesis is that there will still be forest types, structures and conditions that fall outside the AOA because of the FIA plot's lack of stratification and that canopy cover percent will have more of an effect on which stands are within the AOA. We intend to apply this methodology to the true locations and answer the same questions we have done throughout this study.

Another major limitation to our study was focusing on only LiDAR-based structural metrics for our modeling. Many forest inventories utilize climatic and topographic variables to enhance their predictive power and place-based applicability (Lister *et al.*, 2020b; Ullah *et al.*, 2025). We chose to leave these out of our analysis because we were primarily interested in how the FIA sampling scheme represented forest structure. Adding these variables would likely increase the percentage of area outside the AOA. It is extremely unlikely that pixels outside the AOA with our current model would be within the AOA in a model that included more variables because the distance to the nearest sample point can only increase with more variables. However, it is unclear how well the FIA sampling scheme represents climatic and topographic variables.

Finally, the calculation of the DI threshold that the AOA map relies on is another major limitation. We used the calculation specified in Meyer and Pebesma (2021); however, it is unclear how appropriate this threshold is for our application and how much bias a pixel has for when it's outside the AOA. We did a preliminary analysis on the DI threshold by repeatedly withholding FIA plots of a similar structure from our model. We found that FIA plots with similar structure withheld from the model were sometimes still included in that model's AOA indicating that this threshold may be slightly overestimated; however, no conclusive alternative was found. A smaller threshold would increase the percentage of area outside the AOA for all analysis, but trends in how the percentage of area outside the AOA changes in response to different forest types, structures and conditions would remain the same. To conclusively determine a DI threshold, a future analysis needs to be done to determine how predictions at a pixel are affected when that pixel has a specific value of DI. A DI threshold can then be a practical decision based on how much bias those who utilize the maps find acceptable. In doing so, a pixel outside the AOA will have a more concrete interpretation (we will know the exact bias associated with it) than currently described in this study.

5 Conclusion

Forest inventories are a common resource for managers and policy makers to inform their decision making. The FIA dataset is a great resource for training models to map aboveground carbon, basal area, and wood volume using LiDAR-based structure metrics. My study explored the limits of where a model trained using LiDAR and FIA plots can be applied in the Sierra Nevada. Specifically, this preliminary analysis found that YPMC and red fir forests are not as well modeled as other forest categories by the LiDAR and the FIA dataset. Structurally, areas that our model was biased at had large dominant tree heights across a range of canopy covers. We also found that combining the FIA data with LiDAR-based two-step inventory modeling was inappropriate for mapping forests experiencing recent disturbance. It is important for future modelers to be aware that while FIA plots are a valuable resource, there are situations where using them for mapping is inappropriate. Some quantification of the domain that samples are appropriate for should be provided to managers and policy makers so they can best assess when and how to use these maps. When we have better access to the FIA plots' true locations, we will apply this methodology to them to create more precise maps of the aboveground carbon, basal area, and wood volume with their respective area of applicability.

6 Acknowledgments

First, I'd like to express my gratitude to my co-advisors, Van R Kane and Susan Prichard, for their guidance and support for this work. I'd like to thank the rest of my committee: Hans-Erik Andersen for his patience and advice and Maureen Kennedy for her statistical input. This research would not be possible without the funding by The California Department of Forestry and Fire Protection (CAL FIRE). I'd like to thank all faculty, staff and students in the Forest Resiliency Lab, the FLAME lab, and the RSGAL lab for their advice and encouragement. Finally, I'd like to thank my family, friends, and classmates for supporting me through this research.

7 References

3D Elevation Program | U.S. Geological Survey (2025). Available at: https://www.usgs.gov/3d-elevation-program (Accessed: April 27, 2026).

Adhikari, A., Montes, C.R. and Peduzzi, A. (2023) "A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics," *Remote Sensing*, 15(5), p. 1284. Available at: https://doi.org/10.3390/rs15051284.

Andersen, H.-E. *et al.* (2021) "Using advanced airborne remote sensing as a sampling tool to support forest inventory in interior Alaska, USA," pp. 323–326. Available at: https://doi.org/10.34726/wim.2024.

Bechtold, W.A. and Patterson, P.L. (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. 85 p.*, 080. Available at: https://doi.org/10.2737/SRS-GTR-80.

Chu, T. and Guo, X. (2014) "Remote Sensing Techniques in Monitoring Post-Fire Effects and Patterns of Forest Recovery in Boreal Forest Regions: A Review," *Remote Sensing*, 6(1), pp. 470–520. Available at: https://doi.org/10.3390/rs6010470.

Collins, B.M., Everett, R.G. and Stephens, S.L. (2011a) "Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests," *Ecosphere*, 2(4), p. art51. Available at: https://doi.org/10.1890/ES11-00026.1.

Collins, B.M., Everett, R.G. and Stephens, S.L. (2011b) "Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests," *Ecosphere*, 2(4), p. art51. Available at: https://doi.org/10.1890/ES11-00026.1.

Cutler, D.R. *et al.* (2007) "Random Forests for Classification in Ecology," *Ecology*, 88(11), pp. 2783–2792. Available at: https://doi.org/10.1890/07-0539.1.

Dolanc, C.R. *et al.* (2014) "Changing forest structure across the landscape of the Sierra Nevada, CA, USA, since the 1930s," *Ecosphere*, 5(8), p. art101. Available at: https://doi.org/10.1890/ES14-00103.1.

Easterday, K., McIntyre, P. and Kelly, M. (2018) "Land ownership and 20th century changes to forest structure in California," *Forest Ecology and Management*, 422, pp. 137–146. Available at: https://doi.org/10.1016/j.foreco.2018.04.012.

van Ewijk, K. *et al.* (2019) “Forest Site and Type Variability in ALS-Based Forest Resource Inventory Attribute Predictions over Three Ontario Forest Sites,” *Forests*, 10(3), p. 226. Available at: https://doi.org/10.3390/f10030226.

Eyre, F. H. (1980) *Forest Cover Types of the United States and Canada*. Washington, DC.

Fettig, C.J. *et al.* (2019) “Tree mortality following drought in the central and southern Sierra Nevada, California, U.S.,” *Forest Ecology and Management*, 432, pp. 164–178. Available at: https://doi.org/10.1016/j.foreco.2018.09.006.

Forest Inventory and Analysis | CAL FIRE (2026). Available at: https://www.fire.ca.gov/what-we-do/fire-resource-assessment-program/forest-inventory-and-analysis?utm_source=chatgpt.com (Accessed: October 10, 2025).

Forest Inventory and Analysis Strategic Plan | US Forest Service Research and Development (2023). Available at: https://research.fs.usda.gov/understory/forest-inventory-and-analysis-strategic-plan (Accessed: April 27, 2026).

Fricker, G.A. *et al.* (2019) “More than climate? Predictors of tree canopy height vary with scale in complex terrain, Sierra Nevada, CA (USA),” *Forest Ecology and Management. 434: 142-153*, 434, pp. 142–153. Available at: https://doi.org/10.1016/j.foreco.2018.12.006.

Fulé, P.Z. *et al.* (2012) “Do thinning and/or burning treatments in western USA ponderosa or Jeffrey pine-dominated forests help restore natural fire behavior?,” *Forest Ecology and Management*, 269, pp. 68–81. Available at: https://doi.org/10.1016/j.foreco.2011.12.025.

Garcia, M. *et al.* (2017) “Quantifying biomass consumption and carbon release from the California Rim fire by integrating airborne LiDAR and Landsat OLI data,” *Journal of Geophysical Research: Biogeosciences*, 122(2), pp. 340–353. Available at: https://doi.org/10.1002/2015JG003315.

Gillespie, A.J.R. (1999) “Rationale for a National Annual Forest Inventory Program,” *Journal of Forestry*, 97(12), pp. 16–20. Available at: https://doi.org/10.1093/jof/97.12.16.

Goodbody, T.R.H. *et al.* (2023) “sgsR: a structurally guided sampling toolbox for LiDAR-based forest inventories,” *Forestry: An International Journal of Forest Research*, 96(4), pp. 411–424. Available at: https://doi.org/10.1093/forestry/cpac055.

Gregorutti, B., Michel, B. and Saint-Pierre, P. (2017) “Correlation and variable importance in random forests,” *Statistics and Computing*, 27(3), pp. 659–678. Available at: https://doi.org/10.1007/s11222-016-9646-1.

Hawbaker, T. *et al.* (2009) “Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design,” *Journal of Geophysical Research*, 114. Available at: https://doi.org/10.1029/2008JG000870.

Hessburg, P.F. *et al.* (2020) “The 1994 Eastside Screens large-tree harvest limit: review of science relevant to forest planning 25 years later,” *Gen. Tech. Rep. PNW-GTR-990. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 114 p.*, 990. Available at: https://research.fs.usda.gov/treesearch/60635 (Accessed: April 27, 2026).

Huang, S. *et al.* (2018) “F3: Simulating spatiotemporal forest change from field inventory, remote sensing, growth modeling, and management actions,” *Forest Ecology and Management*, 415–416, pp. 26–37. Available at: https://doi.org/10.1016/j.foreco.2018.02.026.

Hudak, A.T. *et al.* (2020) “A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA,” *Environmental Research Letters*, 15(9), p. 095003. Available at: https://doi.org/10.1088/1748-9326/ab93f9.

Huesca, M. *et al.* (2019) “Discrimination of Canopy Structural Types in the Sierra Nevada Mountains in Central California,” *Remote Sensing*, 11(9), p. 1100. Available at: https://doi.org/10.3390/rs11091100.

Johnson, L.K. *et al.* (2022) “Fine-resolution landscape-scale biomass mapping using a spatiotemporal patchwork of LiDAR coverages,” *International Journal of Applied Earth Observation and Geoinformation*, 114, p. 103059. Available at: https://doi.org/10.1016/j.jag.2022.103059.

Kane, J. (2025) “jontkane/Lapis.” Available at: https://github.com/jontkane/Lapis (Accessed: October 13, 2025).

Kane, V. *et al.* (2010) “Comparisons between field- and LiDAR-based measures of stand structural complexity,” *Canadian Journal of Forest Research*, pp. 761–773. Available at: https://doi.org/10.1139/X10-024.

Kane, V.R. *et al.* (2019) “First-entry wildfires can create opening and tree clump patterns characteristic of resilient forests,” *Forest Ecology and Management*, 454, p. 117659. Available at: https://doi.org/10.1016/j.foreco.2019.117659.

Kangas, A., Myllymäki, M. and Packalen, P. (2025) “The effect of sampling design in model-based small area estimation,” *Forestry: An International Journal of Forest Research*, p. cpaf076. Available at: https://doi.org/10.1093/forestry/cpaf076.

Keeley, J.E. and Syphard, A.D. (2021) “Large California wildfires: 2020 fires in historical context,” *Fire Ecology*, 17(1), p. 22. Available at: https://doi.org/10.1186/s42408-021-00110-7.

Khanal, S. *et al.* (2023) “Mapping soil organic carbon stocks in Nepal’s forests,” *Scientific Reports*, 13(1), p. 8090. Available at: https://doi.org/10.1038/s41598-023-34247-z.

Kolden, C.A. (2019) “We’re Not Doing Enough Prescribed Fire in the Western United States to Mitigate Wildfire Risk,” *Fire*, 2(2), p. 30. Available at: https://doi.org/10.3390/fire2020030.

Krofcheck, D. j. *et al.* (2019) “Optimizing Forest Management Stabilizes Carbon Under Projected Climate and Wildfires,” *Journal of Geophysical Research: Biogeosciences*, 124(10), pp. 3075–3087. Available at: https://doi.org/10.1029/2019JG005206.

Lamahewage, S.H.G. *et al.* (2025) “Comparing Machine Learning and Statistical Models for Remote Sensing-Based Forest Aboveground Biomass Estimations,” *Forests*, 16(9), p. 1430. Available at: https://doi.org/10.3390/f16091430.

LaPoint, E. (2005) “Access and Use of FIA Data Through FIA Spatial Data Services,” *In: Proceedings of the fifth annual forest inventory and analysis symposium; 2003 November 18-20; New Orleans, LA. Gen. Tech. Rep. WO-69. Washington, DC: U.S. Department of Agriculture Forest Service. 222p.*, 069. Available at: https://research.fs.usda.gov/treesearch/14295 (Accessed: April 11, 2025).

van Leeuwen, M. and Nieuwenhuis, M. (2010) “Retrieval of forest structural parameters using LiDAR remote sensing,” *European Journal of Forest Research*, 129(4), pp. 749–770. Available at: https://doi.org/10.1007/s10342-010-0381-4.

Lefsky, M.A. *et al.* (1999) “Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests,” *Remote Sensing of Environment*, 70(3), pp. 339–361. Available at: https://doi.org/10.1016/S0034-4257(99)00052-8.

Lefsky, M.A. *et al.* (2005) “Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest,” *Remote Sensing of Environment*, 95(4), pp. 532–548. Available at: https://doi.org/10.1016/j.rse.2005.01.010.

Lentile, L.B. *et al.* (2006) “Remote sensing techniques to assess active fire characteristics and post-fire effects,” *International Journal of Wildland Fire*, 15(3), pp. 319–345. Available at: https://doi.org/10.1071/WF05097.

Lister, A.J. *et al.* (2020a) “Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory,” *Forests*, 11(12), p. 1364. Available at: https://doi.org/10.3390/f11121364.

Lister, A.J. *et al.* (2020b) “Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory,” *Forests*, 11(12), p. 1364. Available at: https://doi.org/10.3390/f11121364.

Lovett, G.M. *et al.* (2007) “Who needs environmental monitoring?,” *Frontiers in Ecology and the Environment*, 5(5), pp. 253–260. Available at: https://doi.org/10.1890/1540-9295(2007)5%5B253:WNEM%5D2.0.CO;2.

Lutz, J.A. *et al.* (2012) “Ecological Importance of Large-Diameter Trees in a Temperate Mixed-Conifer Forest,” *PLOS ONE*, 7(5), p. e36131. Available at: https://doi.org/10.1371/journal.pone.0036131.

Lutz, J.A. *et al.* (2018a) “Global importance of large-diameter trees,” *Global Ecology and Biogeography*, 27(7), pp. 849–864. Available at: https://doi.org/10.1111/geb.12747.

Lutz, J.A. *et al.* (2018b) “Global importance of large-diameter trees,” *Global Ecology and Biogeography*, 27(7), pp. 849–864. Available at: https://doi.org/10.1111/geb.12747.

Lydersen, J.M. *et al.* (2013) “Quantifying spatial patterns of tree groups and gaps in mixed-conifer forests: Reference conditions and long-term changes following fire suppression and logging,” *Forest Ecology and Management*, 304, pp. 370–382. Available at: https://doi.org/10.1016/j.foreco.2013.05.023.

Mallek, C. *et al.* (2013) “Modern departures in fire severity and area vary by forest type, Sierra Nevada and southern Cascades, California, USA,” *Ecosphere*, 4(12), p. art153. Available at: https://doi.org/10.1890/ES13-00217.1.

Maltamo, M. *et al.* (2011) “Different plot selection strategies for field training data in ALS-assisted forest inventory,” *Forestry: An International Journal of Forest Research*, 84(1), pp. 23–31. Available at: https://doi.org/10.1093/forestry/cpq039.

van Mantgem, P.J. *et al.* (2013) “Climatic stress increases forest fire severity across the western United States,” *Ecology Letters*, 16(9), pp. 1151–1156. Available at: https://doi.org/10.1111/ele.12151.

McNab and Avers (1994) *Chapter 33-Ecological Subregions of the United States*, *Ecological Subregions of the United States*. Available at: https://www.fs.usda.gov/land/pubs/ecoregions/ch33.html (Accessed: April 14, 2025).

McRoberts, R.E. *et al.* (2018) “The shelf-life of airborne laser scanning data for enhancing forest inventory inferences,” *Remote Sensing of Environment*, 206, pp. 254–259. Available at: https://doi.org/10.1016/j.rse.2017.12.017.

Melville, G., Stone, C. and Turner, R. (2015) “Application of LiDAR data to maximise the efficiency of inventory plots in softwood plantations,” *New Zealand Journal of Forestry Science*, 45(1), p. 9. Available at: https://doi.org/10.1186/s40490-015-0038-7.

Meyer, H. *et al.* (2026) “CAST: ‘caret’ Applications for Spatial-Temporal Models.” Available at: https://cran.r-project.org/web/packages/CAST/index.html (Accessed: April 27, 2026).

Meyer, H. and Pebesma, E. (2021) “Predicting into unknown space? Estimating the area of applicability of spatial prediction models,” *Methods in Ecology and Evolution*, 12(9), pp. 1620–1633. Available at: https://doi.org/10.1111/2041-210X.13650.

Mildrexler, D.J. *et al.* (2020) “Large Trees Dominate Carbon Storage in Forests East of the Cascade Crest in the United States Pacific Northwest,” *Frontiers in Forests and Global Change*, 3. Available at: https://doi.org/10.3389/ffgc.2020.594274.

Miller, J.D. *et al.* (2009) “Quantitative Evidence for Increasing Forest Fire Severity in the Sierra Nevada and Southern Cascade Mountains, California and Nevada, USA,” *Ecosystems*, 12(1), pp. 16–32. Available at: https://doi.org/10.1007/s10021-008-9201-9.

Moran, C.J., Kane, V.R. and Seielstad, C.A. (2020) “Mapping Forest Canopy Fuels in the Western United States with LiDAR–Landsat Covariance,” *Remote Sensing*, 12(6), p. 1000. Available at: https://doi.org/10.3390/rs12061000.

Næsset, E. (2002) “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data,” *Remote Sensing of Environment*, 80(1), pp. 88–99. Available at: https://doi.org/10.1016/S0034-4257(01)00290-5.

Nicodemus, K.K. *et al.* (2010) “The behaviour of random forest permutation-based variable importance measures under predictor correlation,” *BMC Bioinformatics*, 11(1), p. 110. Available at: https://doi.org/10.1186/1471-2105-11-110.

North, M. (2012) “Managing Sierra Nevada forests,” *Gen. Tech. Rep. PSW-GTR-237. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 184 p*, 237. Available at: https://doi.org/10.2737/PSW-GTR-237.

Pan, Y. *et al.* (2011) “A large and persistent carbon sink in the world’s forests,” *Science. 333: 988-993.*, 333, pp. 988–993. Available at: https://doi.org/10.1126/science.1201609.

Pollet and Omi (2002) “Effect of thinning and prescribed burning on crown fire severity in ponderosa pine forests,” *ResearchGate* [Preprint]. Available at: https://doi.org/10.1071/WF01045.

Prichard, S.J. *et al.* (2021) “Adapting western North American forests to climate change and wildfires: 10 common questions,” *Ecological Applications*, 31(8), p. e02433. Available at: https://doi.org/10.1002/eap.2433.

Queinnec, M. *et al.* (2022) “Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attribute prediction,” *Forestry: An International Journal of Forest Research*, 95(3), pp. 347–362. Available at: https://doi.org/10.1093/forestry/cpab051.

Restaino, C. *et al.* (2019) “Forest structure and climate mediate drought-induced tree mortality in forests of the Sierra Nevada, USA,” *Ecological Applications*, 29(4), p. e01902. Available at: https://doi.org/10.1002/eap.1902.

Riley, K.L., Grenfell, I.C. and Finney, M.A. (2016) “Mapping forest vegetation for the western United States using modified random forests imputation of FIA forest plots,” *Ecosphere*, 7(10), p. e01472. Available at: https://doi.org/10.1002/ecs2.1472.

“Rim Fire, California - NASA Science” (2013), 24 August. Available at: https://science.nasa.gov/earth/earth-observatory/rim-fire-california-81919/ (Accessed: April 27, 2026).

Robichaud, P.R., Beyers, J.L. and Neary, D.G. (2000) “Evaluating the Effectiveness Of Postfire Rehabilitation Treatments,” *General Technical Report RMRS-GTR-63* [Preprint]. Available at: https://research.fs.usda.gov/treesearch/23617 (Accessed: April 27, 2026).

Roussel, J.-R. *et al.* (2026) “lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications.” Available at: https://cran.r-project.org/web/packages/lidR/index.html (Accessed: May 12, 2026).

Sean P. Healey and Zhiqiang Yang (2023) *Landscape Change Monitoring System (LCMS) | US Forest Service Research and Development*. Available at: https://research.fs.usda.gov/rmrs/products/dataandtools/landscape-change-monitoring-system-lcms (Accessed: April 27, 2026).

Sheridan, R.D. *et al.* (2015) “Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest,” *Remote Sensing*, 7(1), pp. 229–255. Available at: https://doi.org/10.3390/rs70100229.

Sleeter, Wilson, and Acevedo (2012) *Status and trends of land change in the Western United States—1973 to 2000*. U.S. Geological Survey. Available at: https://pubs.usgs.gov/pp/1794/a/.

Ståhl, G. *et al.* (2024) “Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time – and how this affects applications,” *Forest Ecosystems*, 11, p. 100164. Available at: https://doi.org/10.1016/j.fecs.2023.100164.

Stephens, S.L. *et al.* (2015) “Historical and current landscape-scale ponderosa pine and mixed conifer forest structure in the Southern Sierra Nevada,” *Ecosphere*, 6(5), p. art79. Available at: https://doi.org/10.1890/ES14-00379.1.

Stephenson, N.L. *et al.* (2014) “Rate of tree carbon accumulation increases continuously with tree size,” *Nature*, 507(7490), pp. 90–93. Available at: https://doi.org/10.1038/nature12914.

Taylor, A.H., Harris, L.B. and Skinner, C.N. (2022) “Severity patterns of the 2021 Dixie Fire exemplify the need to increase low-severity fire treatments in California’s forests,” *Environmental Research Letters*, 17(7), p. 071002. Available at: https://doi.org/10.1088/1748-9326/ac7735.

Tian, L. *et al.* (2023) “Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects,” *Forests*, 14(6), p. 1086. Available at: https://doi.org/10.3390/f14061086.

Tinkham, W.T. *et al.* (2018) “Applications of the United States Forest Inventory and Analysis dataset: A review and future directions,” *Canadian Journal of Forest Research. 48: 1251-1268.*, 48, pp. 1251–1268. Available at: https://doi.org/10.1139/cjfr-2018-0196.

Tompalski, P. *et al.* (2019) “Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data,” *Remote Sensing of Environment*, 227, pp. 110–124. Available at: https://doi.org/10.1016/j.rse.2019.04.006.

Tomppo, E. *et al.* (2014) “A sampling design for a large area forest inventory: case Tanzania,” *Canadian Journal of Forest Research*, 44(8), pp. 931–948. Available at: https://doi.org/10.1139/cjfr-2013-0490.

Ullah, S. *et al.* (2025) “Remote sensing for aboveground biomass monitoring in terrestrial ecosystems: A systematic review,” *Remote Sensing Applications: Society and Environment*, 39, p. 101635. Available at: https://doi.org/10.1016/j.rsase.2025.101635.

United States: Agriculture Department: Forest Service *et al.* (2014) *Forest Vegetation Report: Rim Fire Recovery Project*. Agriculture Department. Available at: https://purl.fdlp.gov/GPO/gpo53918 (Accessed: April 27, 2026).

USDA Forest Service (2019) *CALVEG: Classification and Assessment with Landsat of Visible Ecological Groupings.*, *USDA Forest Service, Enterprise Data Warehouse*. Available at: https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=calveg (Accessed: October 10, 2025).

Van Wagtendonk, J.W. *et al.* (2018) “Chapter Fourteen. Sierra Nevada Bioregion,” in J.W. van Wagtendonk (ed.) *Fire in California’s Ecosystems*. University of California Press, pp. 249–278. Available at: https://www.degruyterbrill.com/document/doi/10.1525/9780520961913-017/html?srsltid=AfmBOooEoWzt2HE_EbJ9u6p9Ut3UPTXUsNp81Y8VxnFFUdZcfOXGrjpq (Accessed: October 10, 2025).

White, J.C. *et al.* (2013) “A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach,” *The Forestry Chronicle*, 89(06), pp. 722–723. Available at: https://doi.org/10.5558/tfc2013-132.

Williams, J.N. *et al.* (2023) “High-severity burned area and proportion exceed historic conditions in Sierra Nevada, California, and adjacent ranges,” *Ecosphere. 14(1): e4397.*, 14, p. e4397. Available at: https://doi.org/10.1002/ecs2.4397.

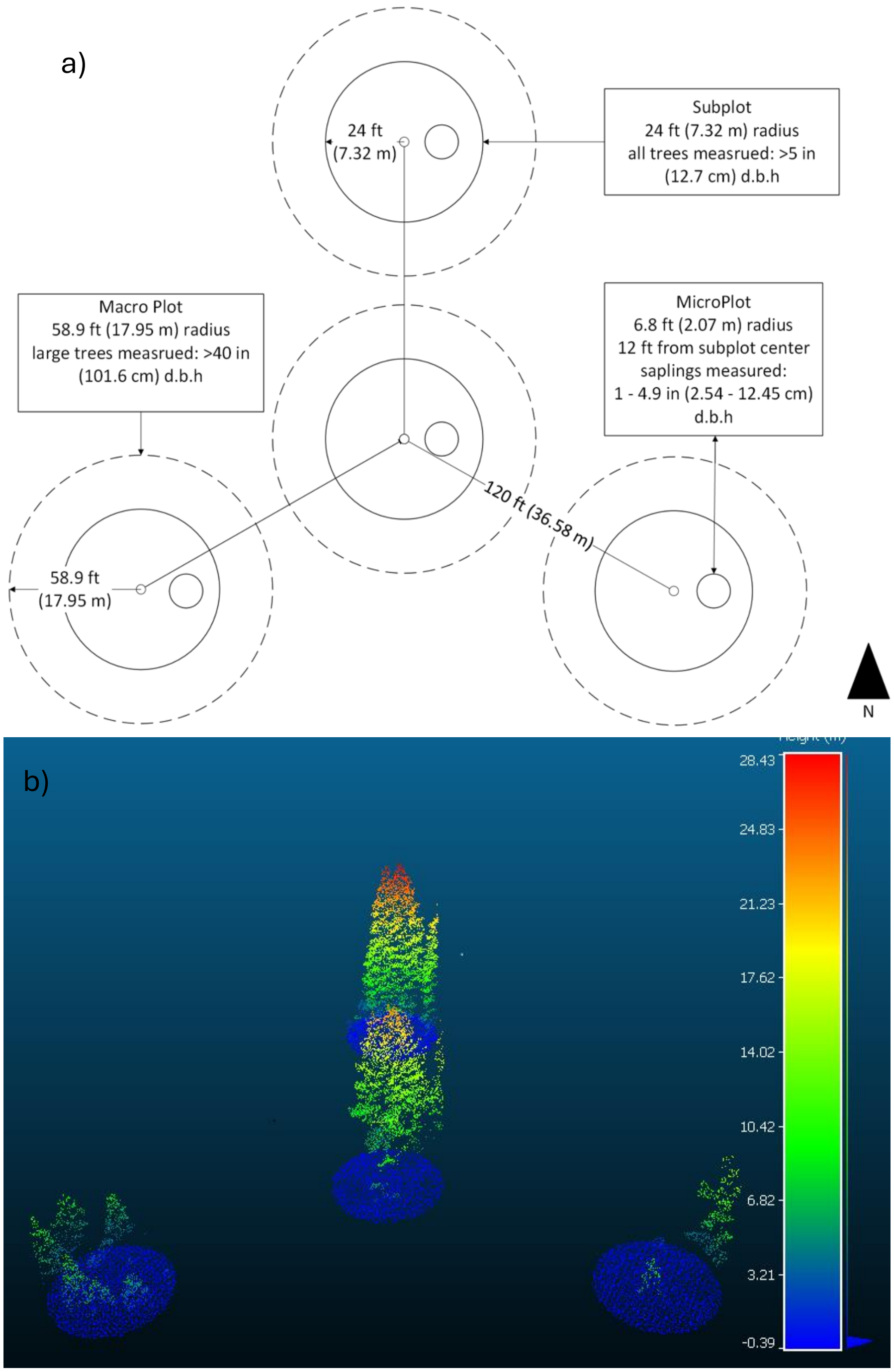
Woods, M., Lim, K. and Treitz, P. (2008) “Predicting forest stand variables from LiDAR data in the Great Lakes – St. Lawrence forest of Ontario,” *The Forestry Chronicle*, 84(6), pp. 827–839. Available at: https://doi.org/10.5558/tfc84827-6.

Wright, M.N., Wager, S. and Probst, P. (2026) “ranger: A Fast Implementation of Random Forests.” Available at: https://cran.r-project.org/web/packages/ranger/index.html (Accessed: April 27, 2026).

Zielinski, W.J. *et al.* (2006) “Using forest inventory data to assess fisher resting habitat suitability in California.,” *Ecological Applications, 16(3): 1010-1025*, 16, pp. 1010–1025.

8 Appendix

8.1 FIA Information



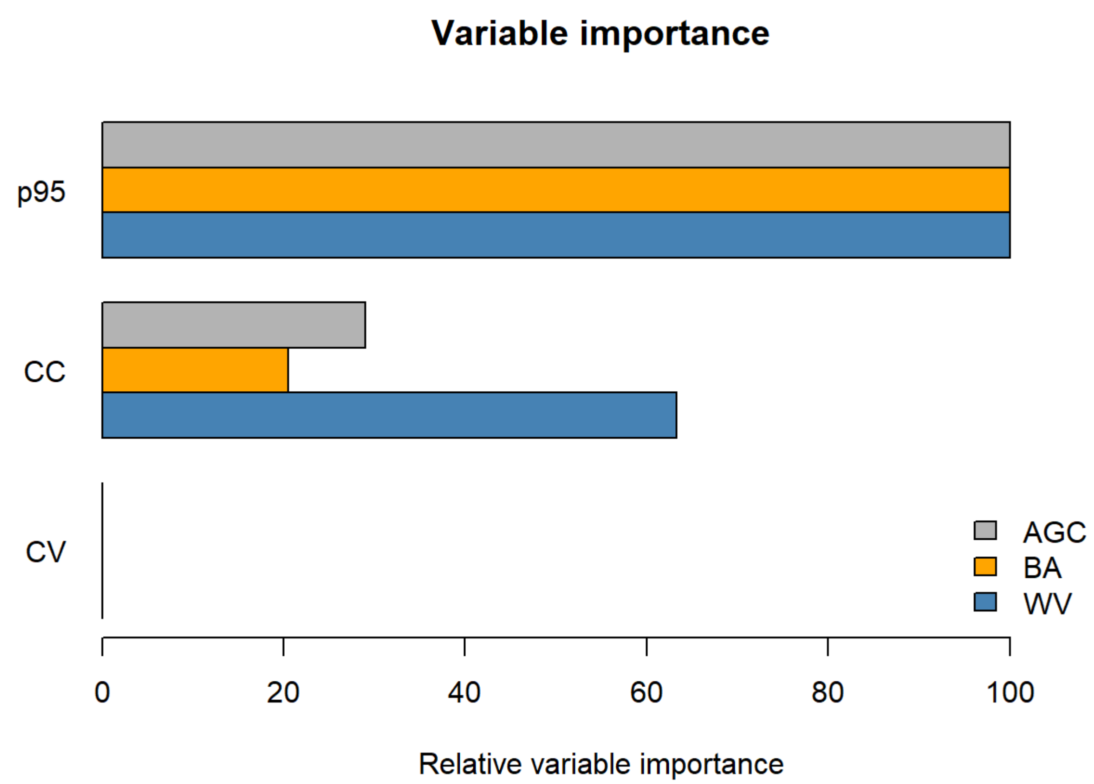
Appendix 8.1-A: a) The FIA plot footprint. The plot consists of 4 subplots each subplot contains a microplot b) An example LiDAR point cloud from over a plot footprint. LiDAR metrics were calculated over these plot footprints. This plot is based on the public locations.

8.2 Modeling Results

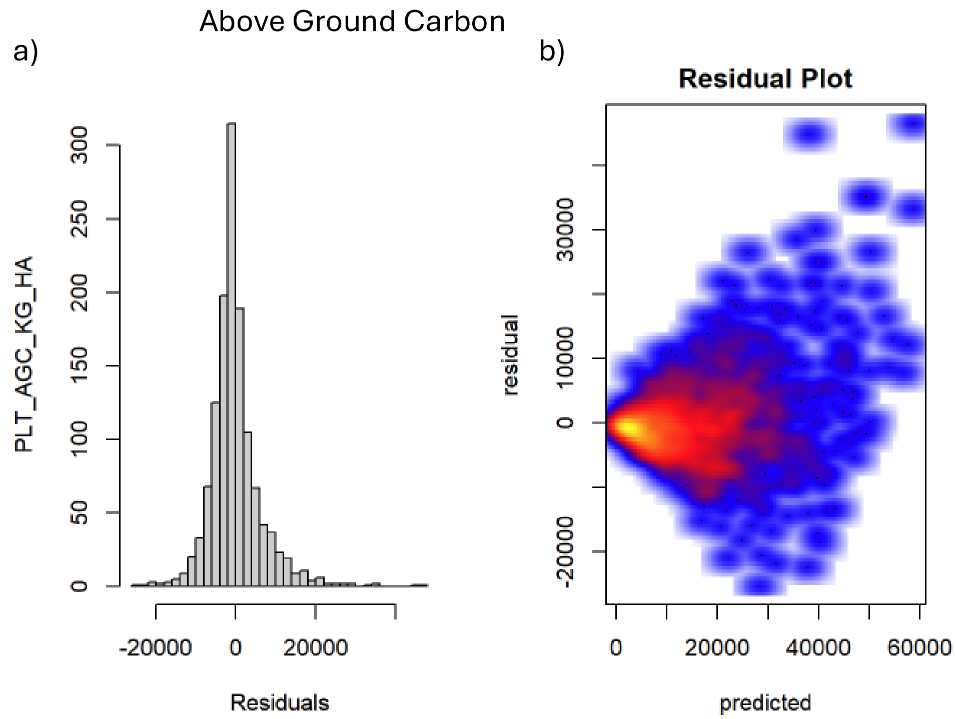
8.2.1 Modeling Results: True Locations

Metric	R^2	RMSE	Mtry	Nodesize	DI threshold
AGC	0.79	6542	1	10	0.084
BA	0.70	3.42	1	10	0.073
WV	0.79	36.1	1	10	0.097

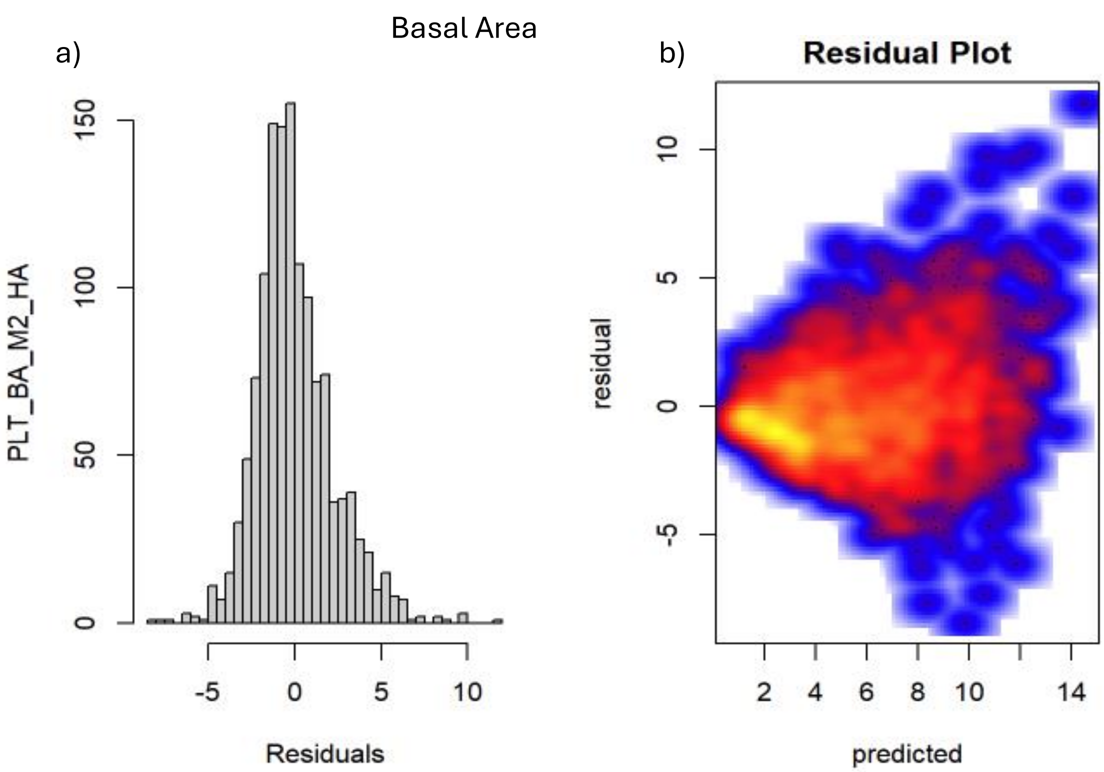
Appendix 8.2-A: Table of hyperparameters and outputs found to be best our random forest mode with the True coordinates.



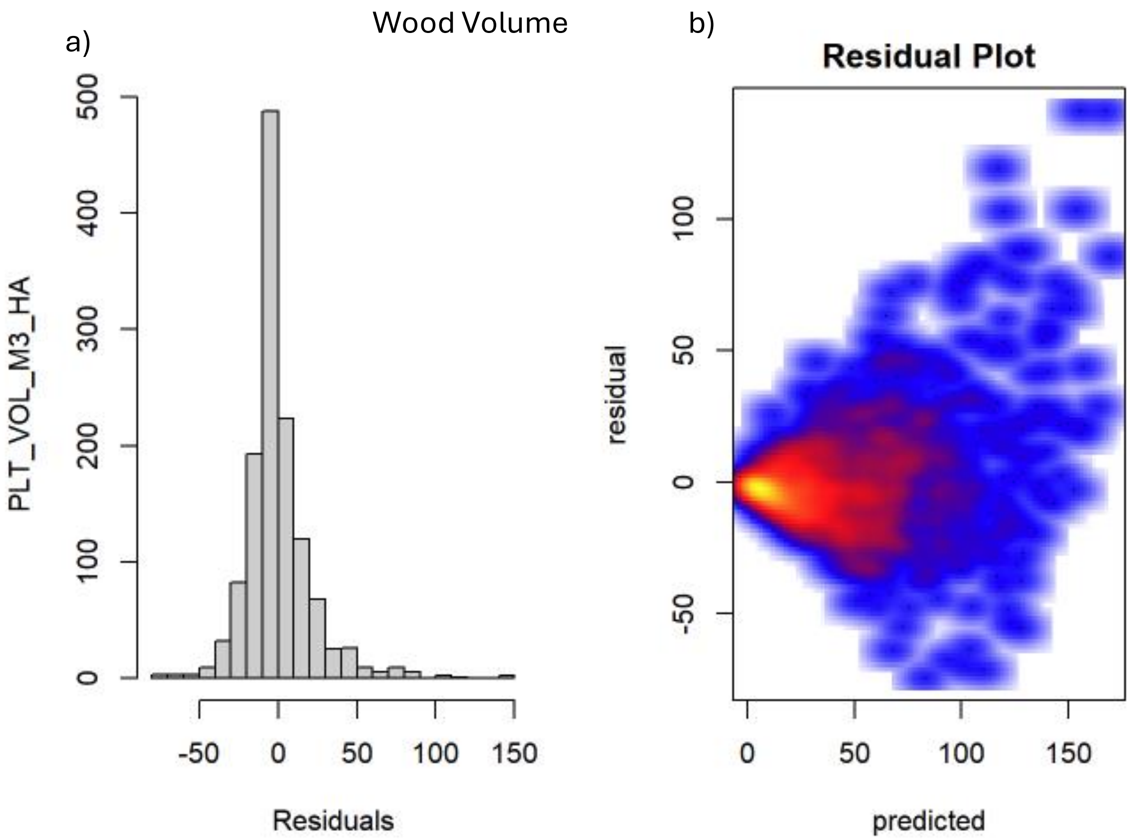
Appendix 8.2-B: The Variable Importance plot for our random forest model with the true coordinates



Appendix 8.2-C: a) Histogram of Residuals from our model for above ground carbon (AGC) b) A residual plot for our random forest model of AGC. These figures use models trained on the true locations.



Appendix 8.2-D: a) A histogram of Residuals for Basal Area produced by our model. b) a residual plot produced for our BA model. These figures use models trained on the true locations.

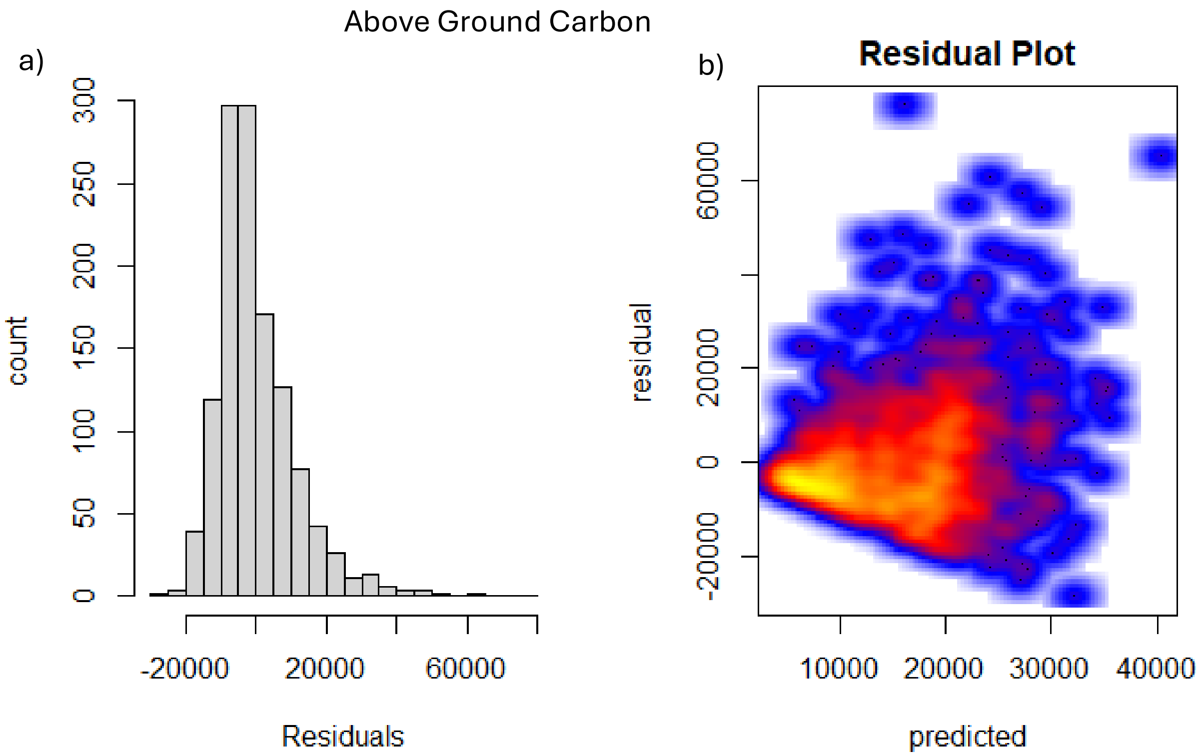


Appendix 8.2-E: a) Histogram of Residuals from our model for wood volume (WV) b) A residual plot for our random forest model of WV. These figures use models trained on the true locations.

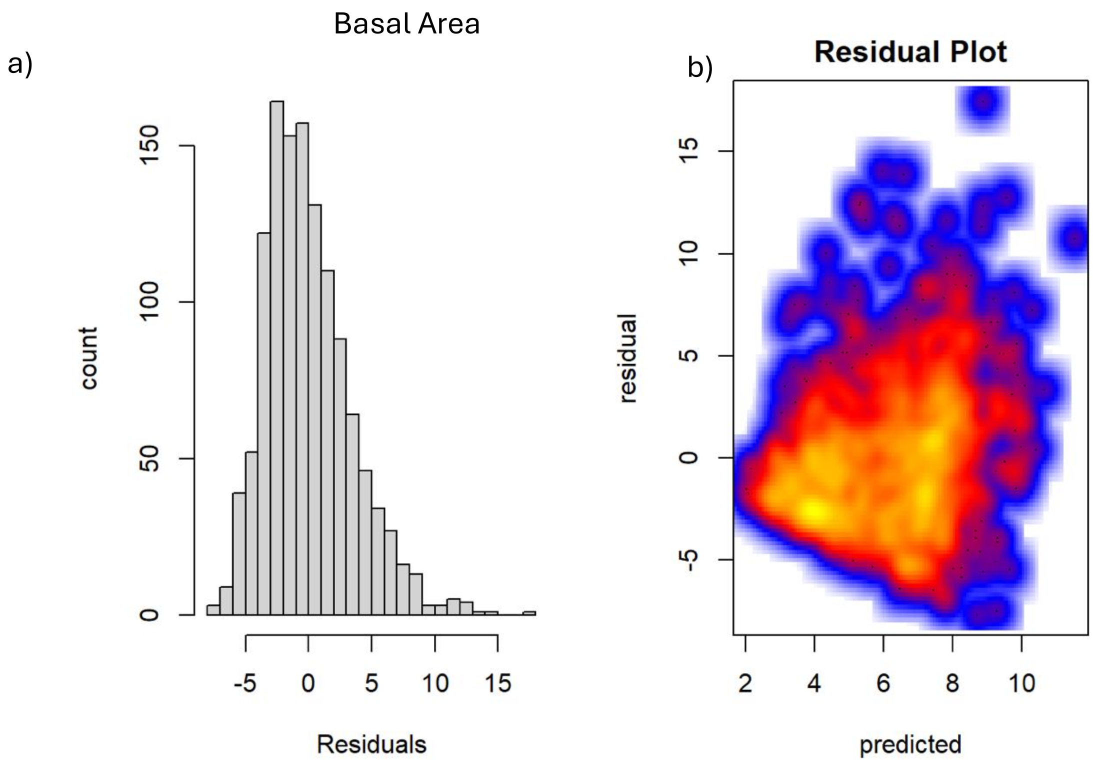
8.2.2 Modeling Results: Public Locations

Metric	R^2	RMSE	Mtry	Nodesize	DI threshold
AGC	0.36	11571	1	20	0.063
BA	0.33	3.42	1	20	0.072
WV	0.37	36.1	1	20	0.058

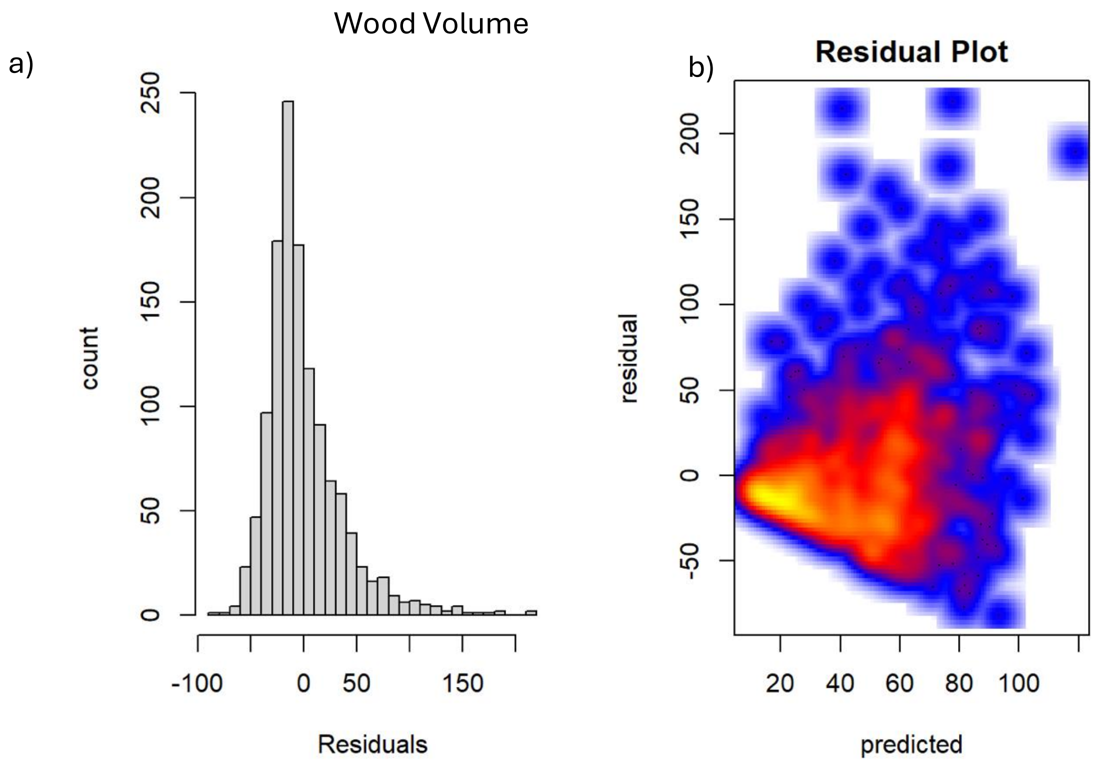
Appendix 8.2-F: Table of hyperparameters and outputs found to be best our random forest mode with the public coordinates.



Appendix 8.2-G: a) Histogram of Residuals from our model for above ground carbon (AGC) b) A residual plot for our random forest model of AGC. These figures use models trained on the public locations.

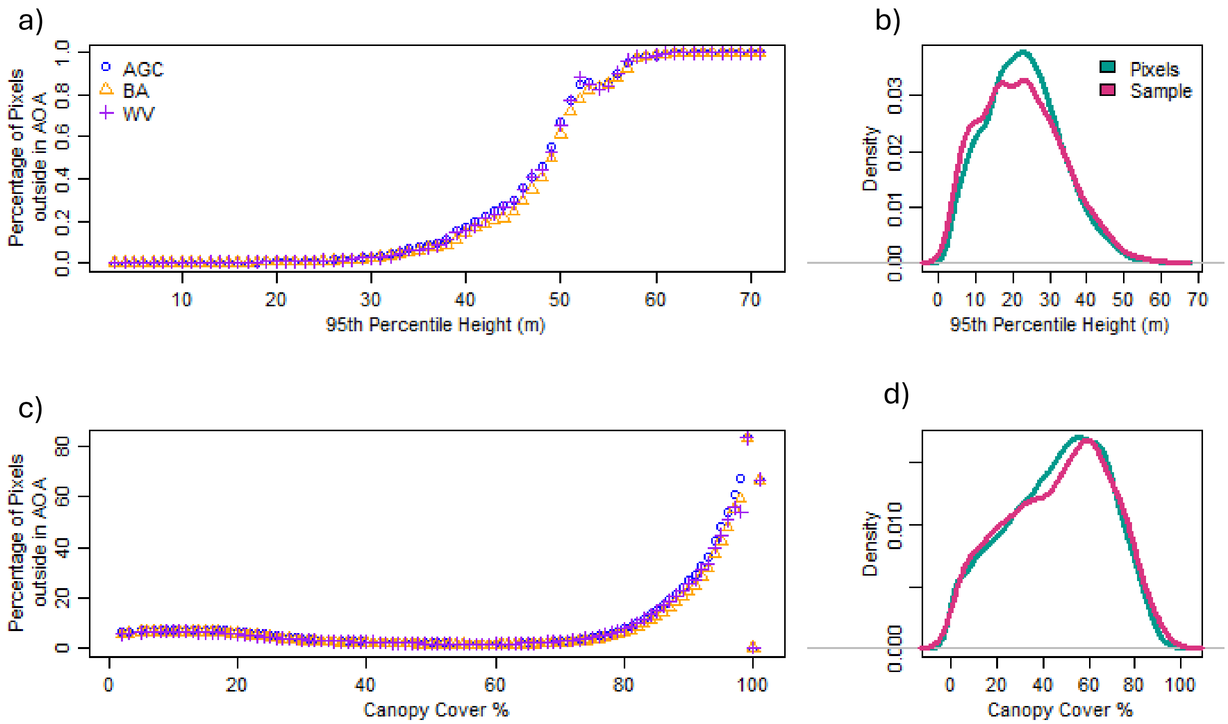


Appendix 8.2-H: A histogram of Residuals for Basal Area produced by our model. b) a residual plot produced for our BA model. These figures use models trained on the public locations.

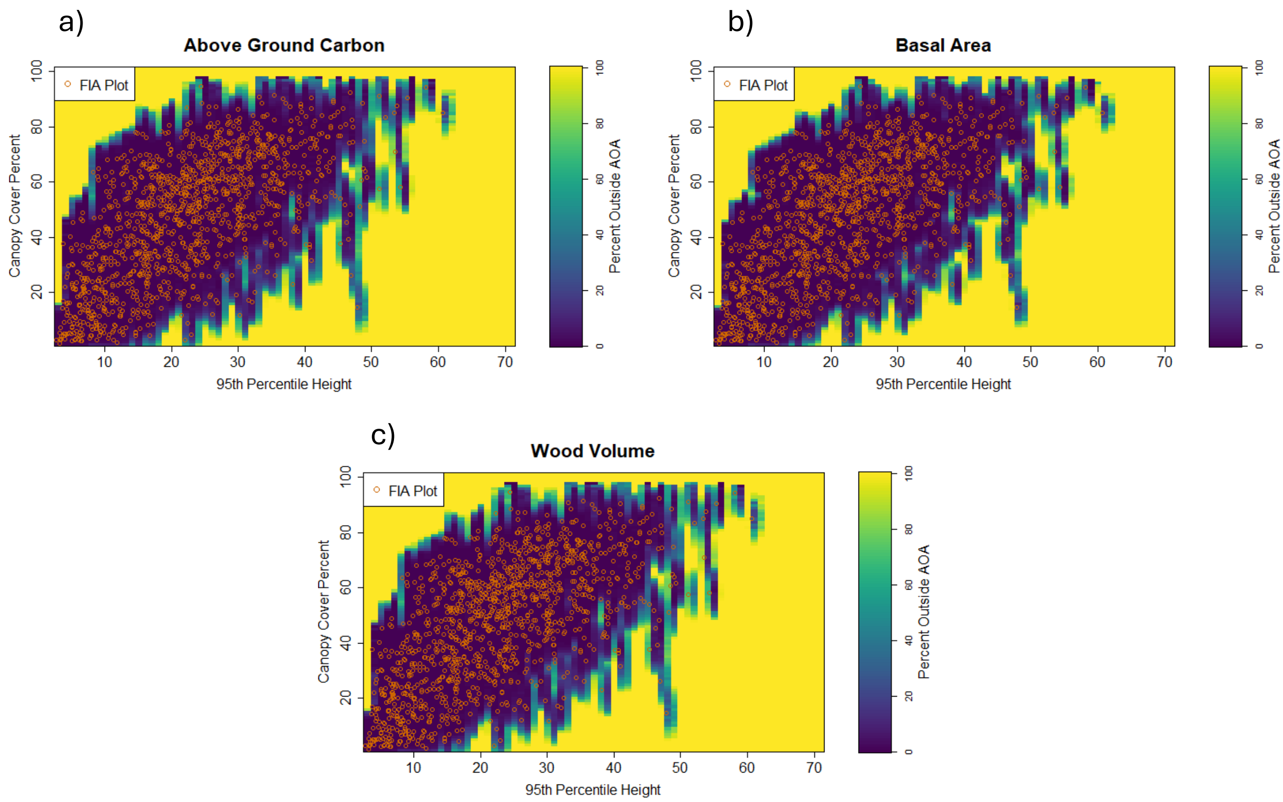


Appendix 8.2-I: a) Histogram of Residuals from our model for wood volume (WV) b) A residual plot for our random forest model of WV. These figures use models trained on the public locations.

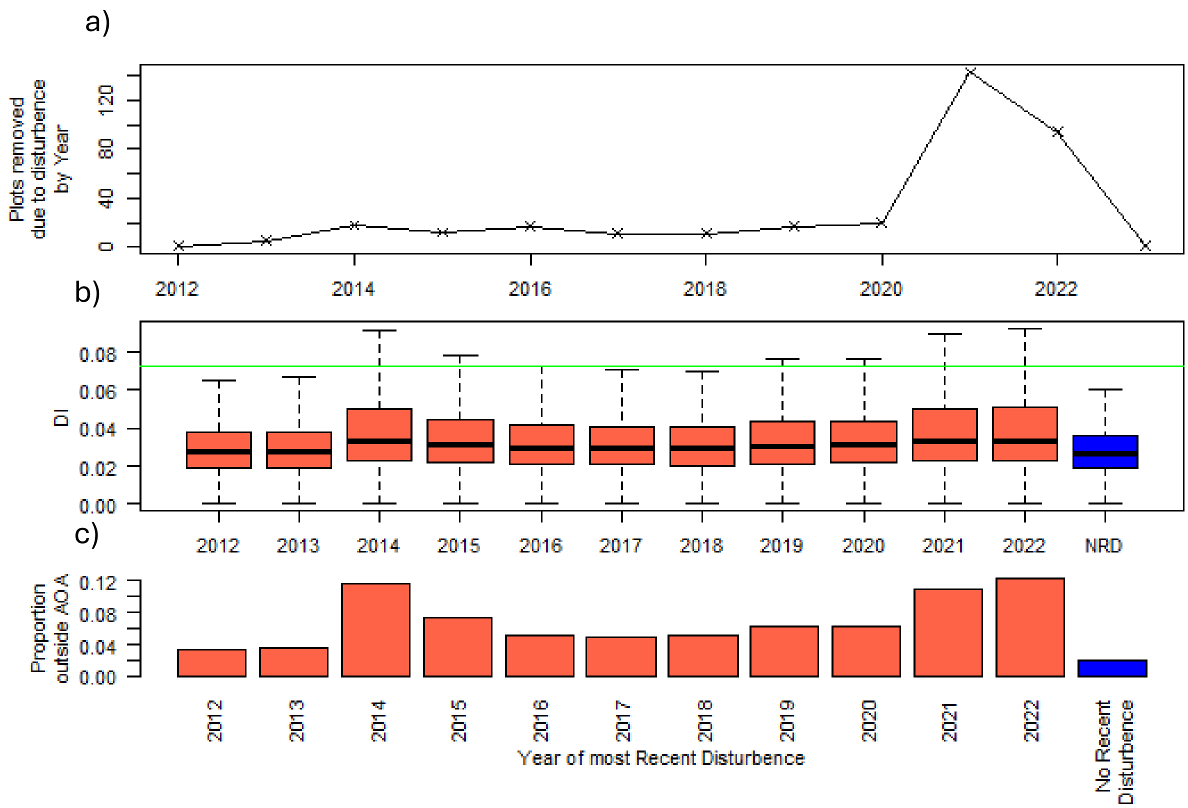
8.3 AOA Results



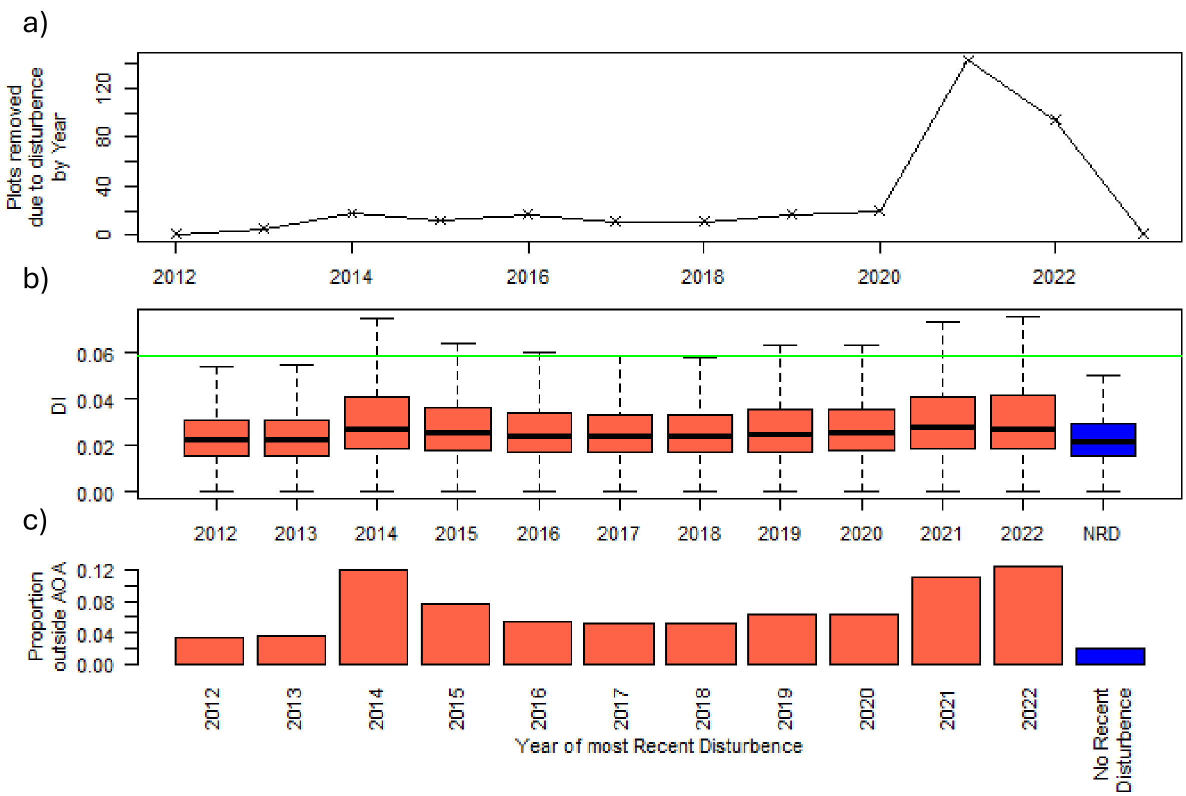
Appendix 8.2-A: a) The percentage of Pixels outside the AOA for above ground carbon (AGC), basal area (BA) and wood volume (WV) versus 95th percentile height. b) The distribution of 95th percentile heights across both the pixels and the samples. c) The percentage of Pixels outside the AOA for AGC, BA, and WV versus canopy cover percent. The distribution of canopy cover percent across both the pixels and the samples.



Appendix 8.3-B: Distribution of AOA versus canopy cover and 95th percentile height for above ground carbon. b) Distribution of the AOA for basal area c) Distribution of the AOA for wood volume.



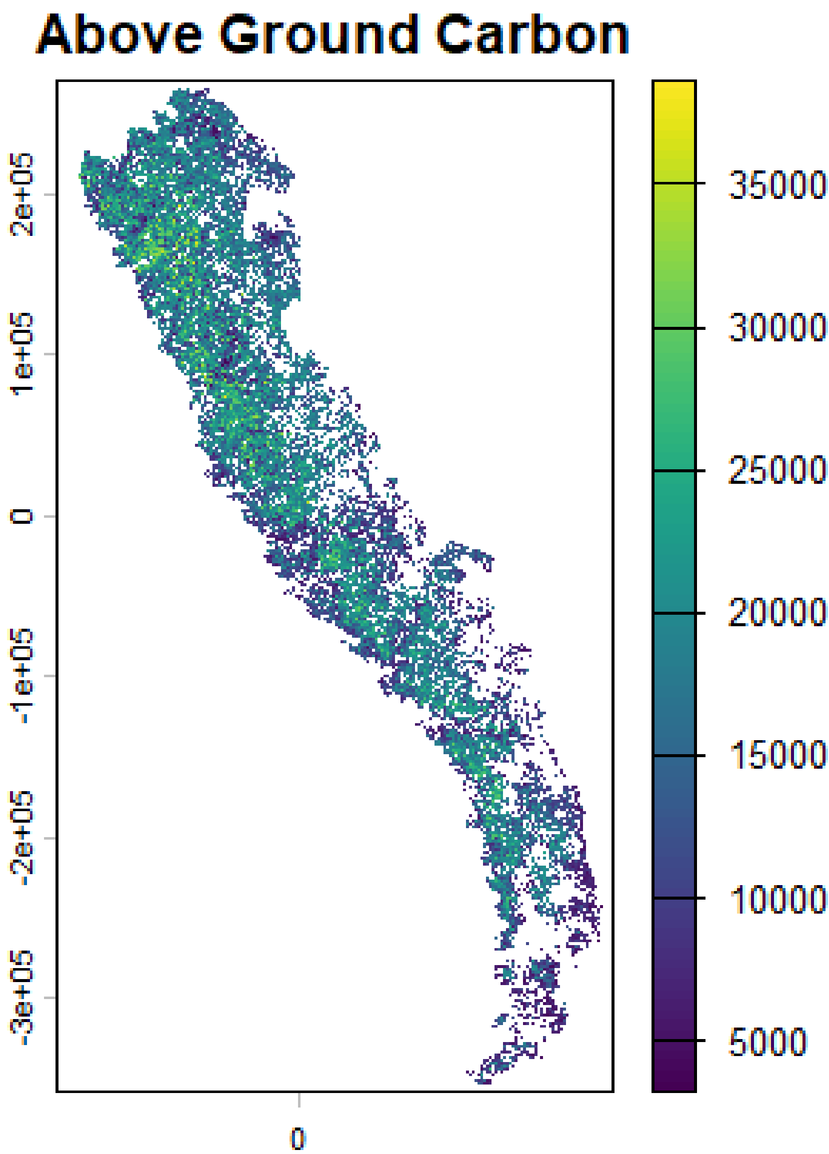
Appendix 8.3-C: A recreation of figure 6 but with the AOA for basal area. a) The number of plots removed due to disturbance occurring at that plot between field measurement and LiDAR measurement. The x-axis is the year that the disturbance occurred at that plot as measured by LCMS. b) Distribution of the Dissimilarity Index (DI) for above ground carbon (AGC) for pixels which experienced recent disturbance. Pixels are grouped by the year they most recently experienced disturbance as measured by LCMS. If there was no disturbance in the listed 10 years that pixel was classified as no recent disturbance. (NRD) c) Proportion of pixels outside the AOA for pixels grouped by the year they most recently experienced disturbance. Figures are produced on the AOA of above ground carbon and from the public FIA locations.



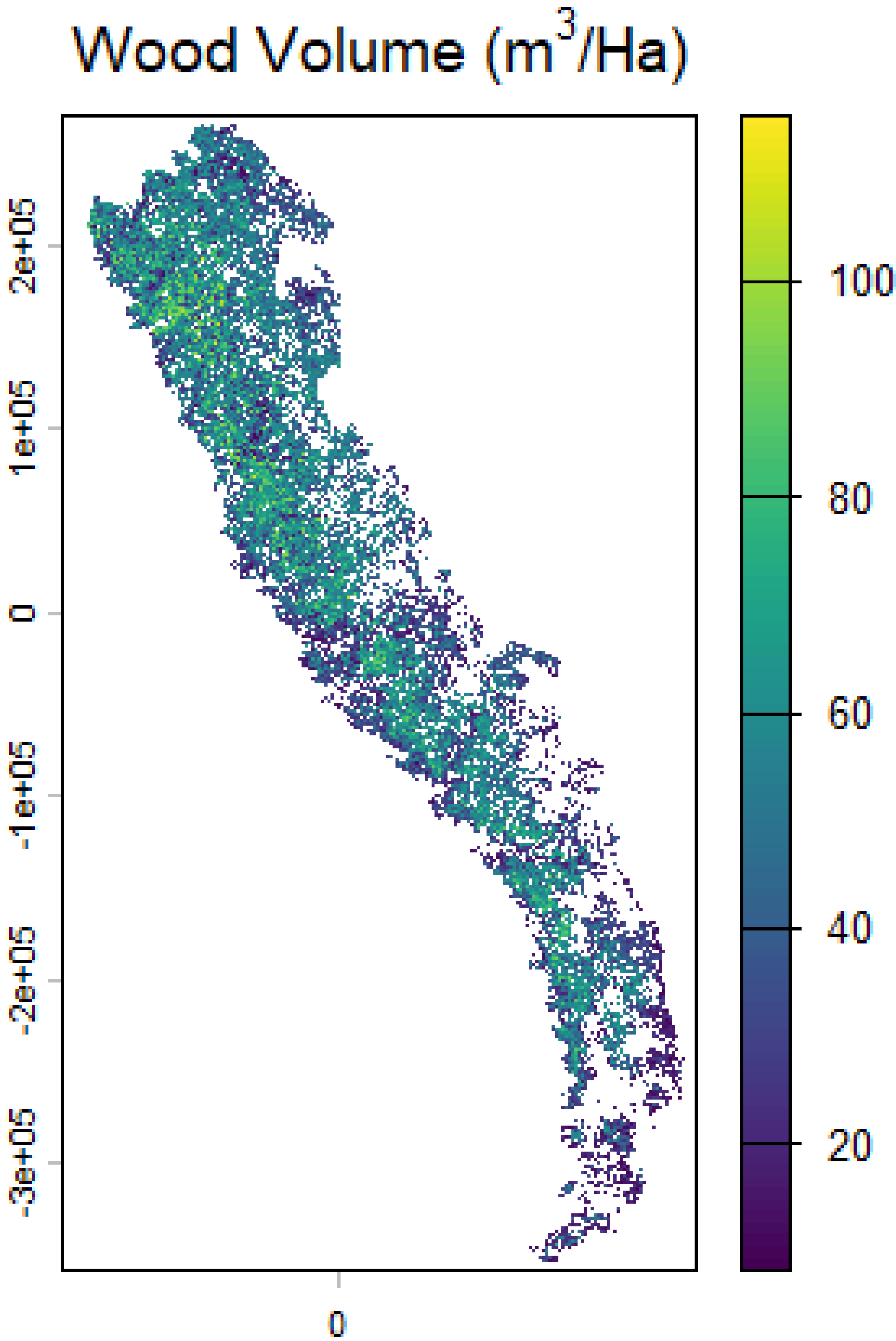
Appendix 8.3-D: A recreation of figure 6 but with Basal Area. a) The number of plots removed due to disturbance occurring at that plot between field measurement and LiDAR measurement. The x-axis is the year that the disturbance occurred at that plot as measured by LCMS. b) Distribution of the Dissimilarity Index (DI) for above ground carbon (AGC) for pixels which experienced recent disturbance. Pixels are grouped by the year they most recently experienced disturbance as measured by LCMS. If there was no disturbance in the listed 10 years that pixel was classified as no recent disturbance. (NRD) c) Proportion of pixels outside the AOA for pixels grouped by the year they most recently experienced disturbance. Figures are produced on the AOA of above ground carbon and from the public FIA locations.

8.4 Study Area Maps

8.4.1 Maps of Forest Metrics

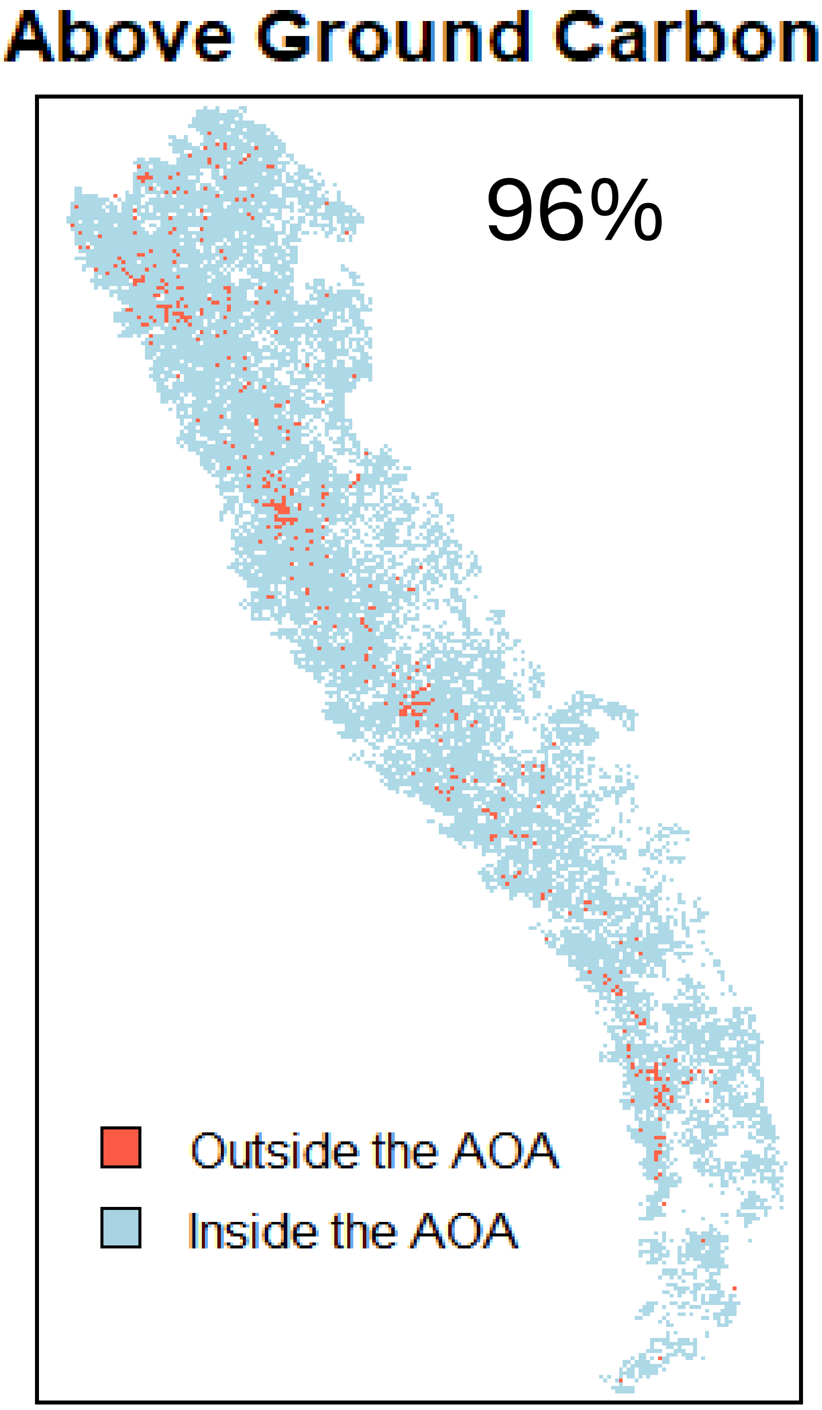


Appendix 8.4-A: Map of Above Ground Carbon (kg/Ha) produced by our models.

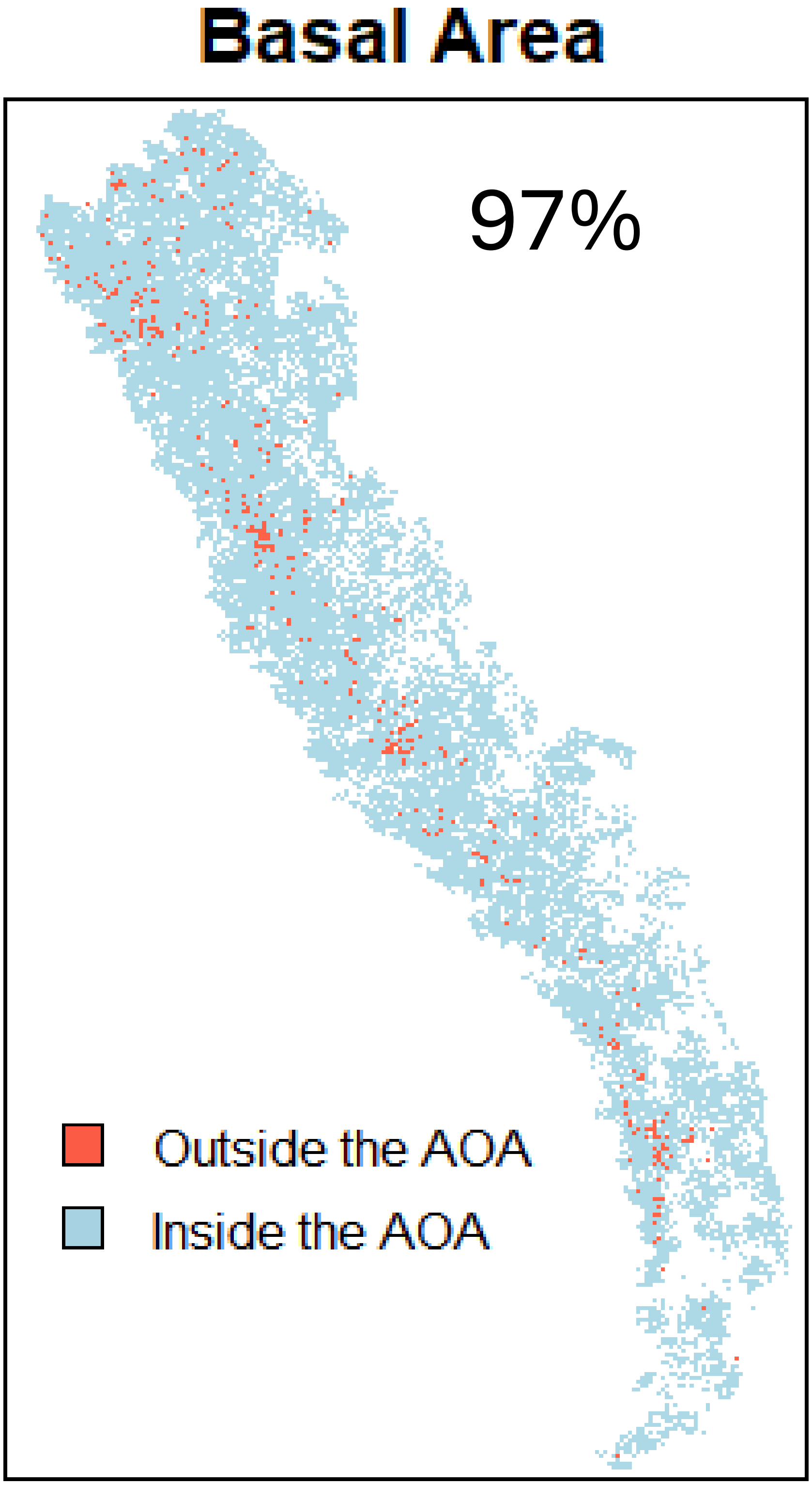


Appendix 8.4-C: Map of Wood Volume (m^3 / Ha) produced by our models

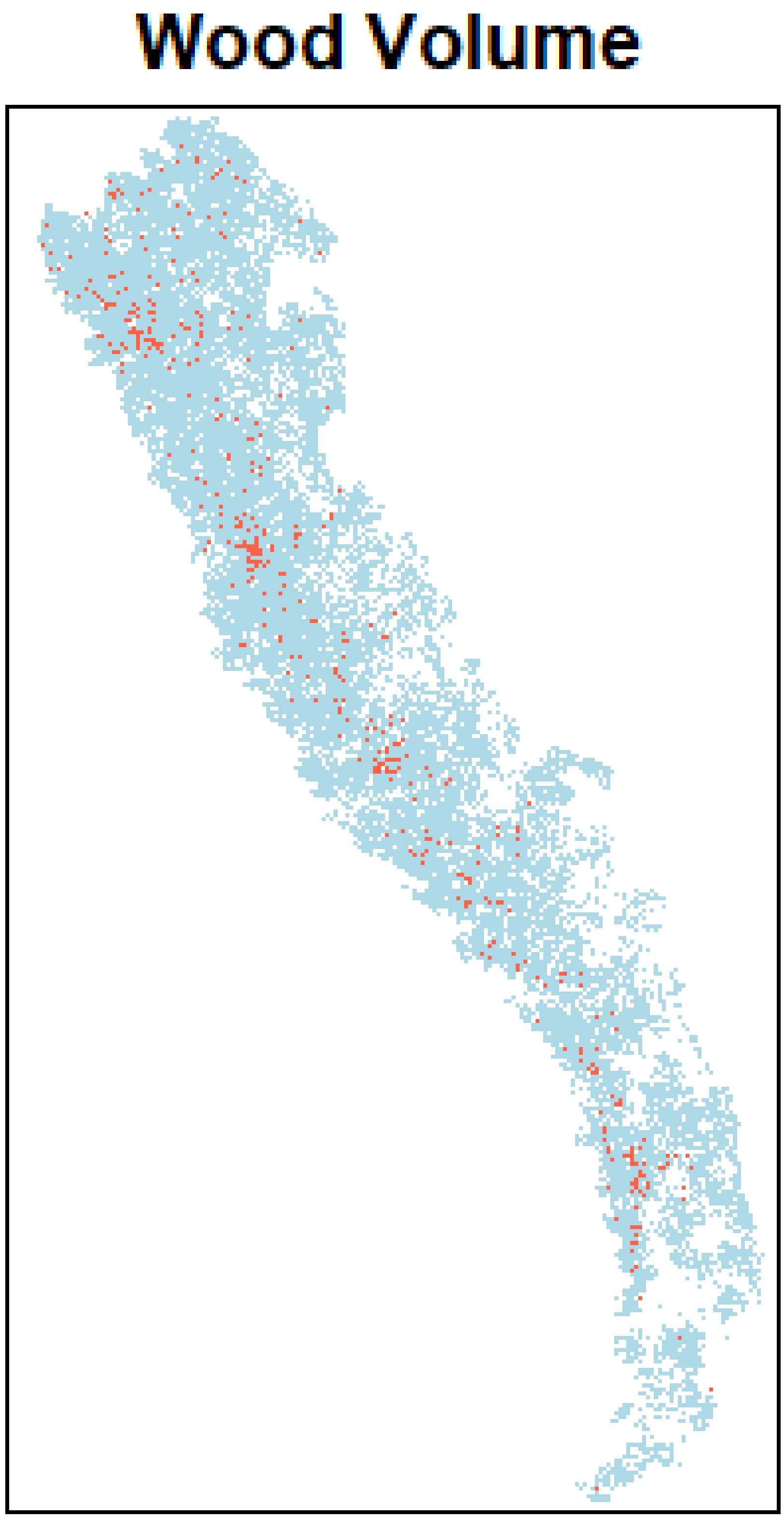
8.4.2 Maps of the AOA



Appendix 8.4-D: Map of the AOA for above ground carbon. 96% of the region is within the AOA

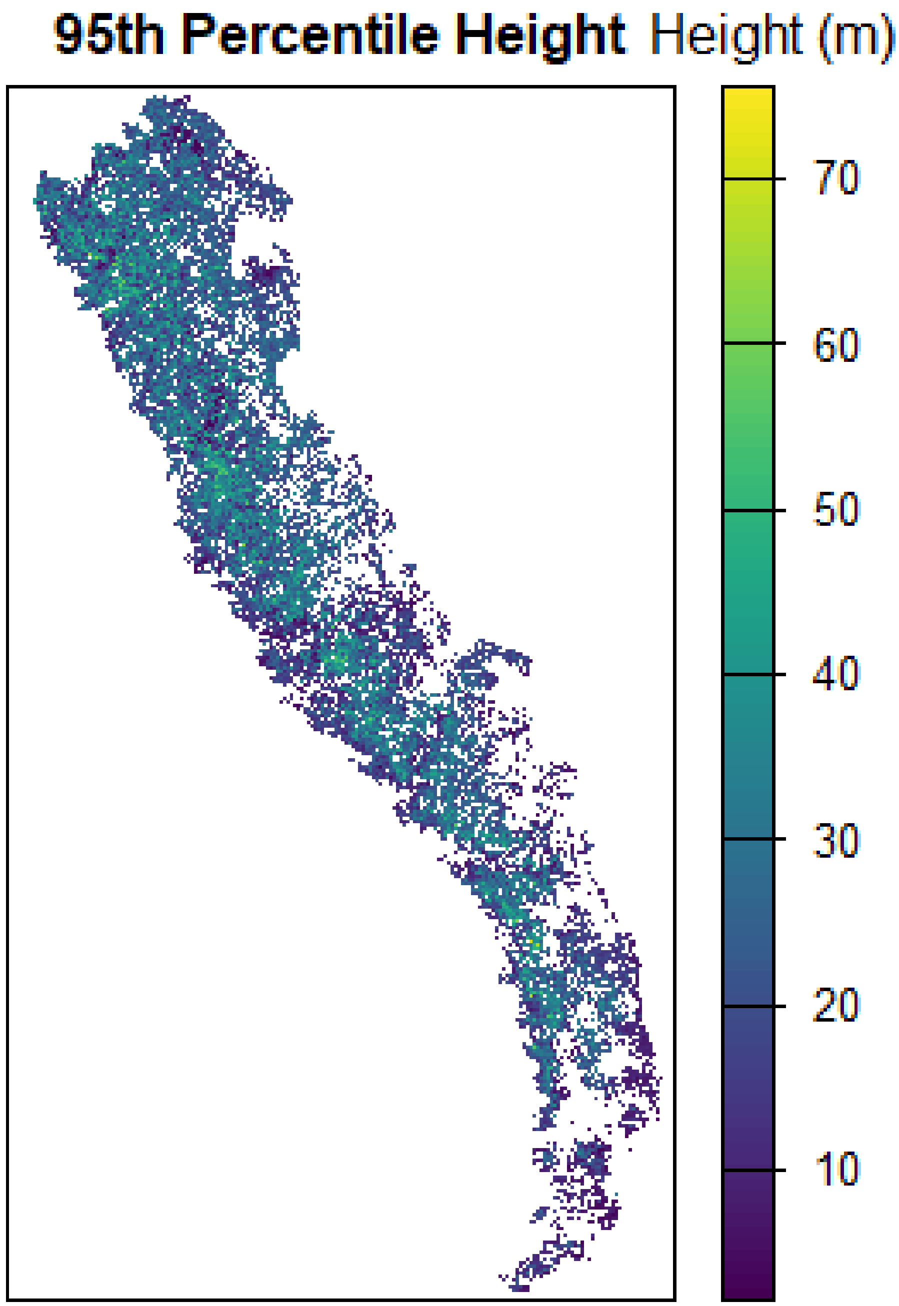


Appendix 8.4-E: Map of the AOA for Basal Area. 96% of the region is within the AOA

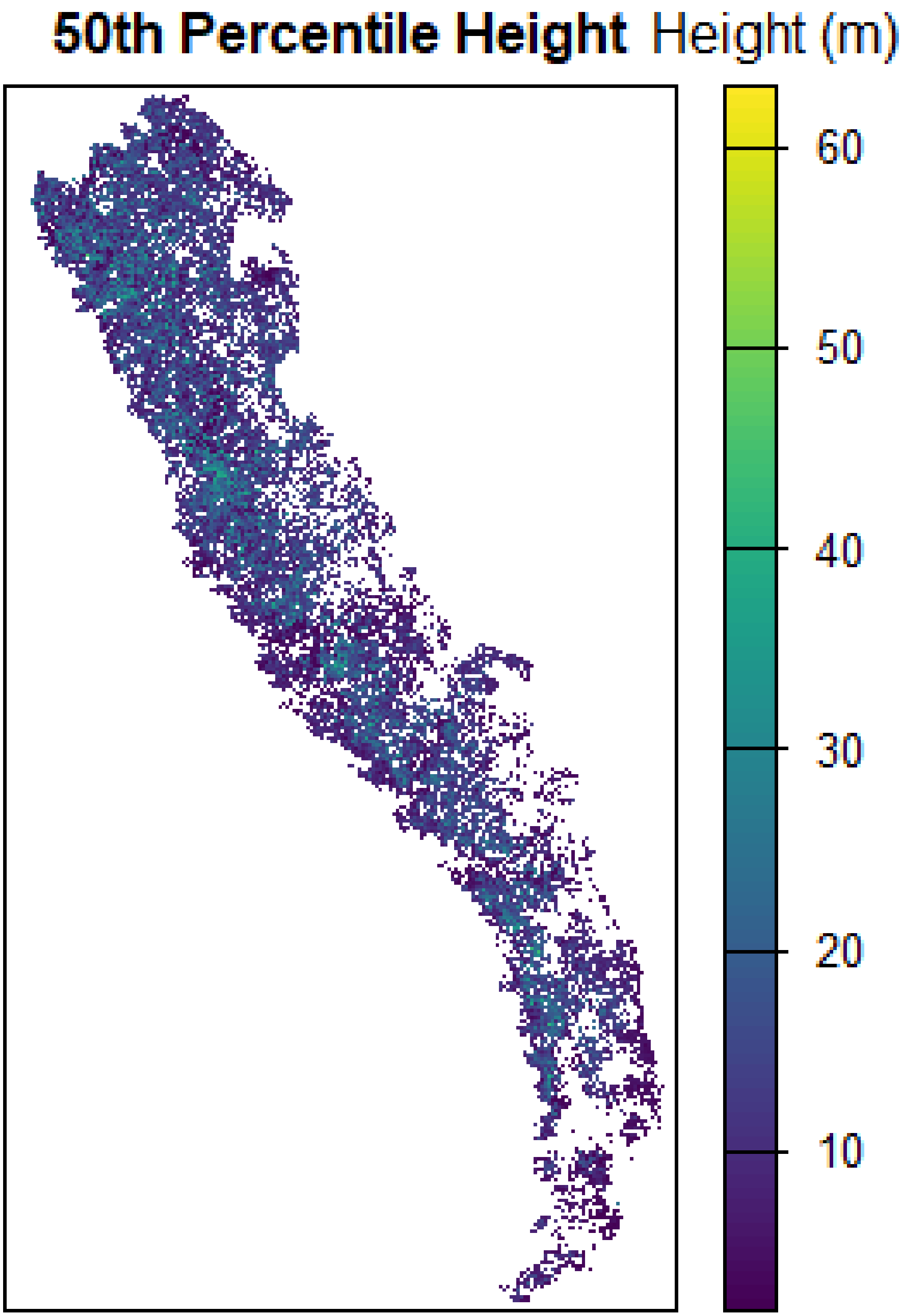


Appendix 8.4-F: Map of the AOA for wood volume. 96% of the ecoregion is within the AOA

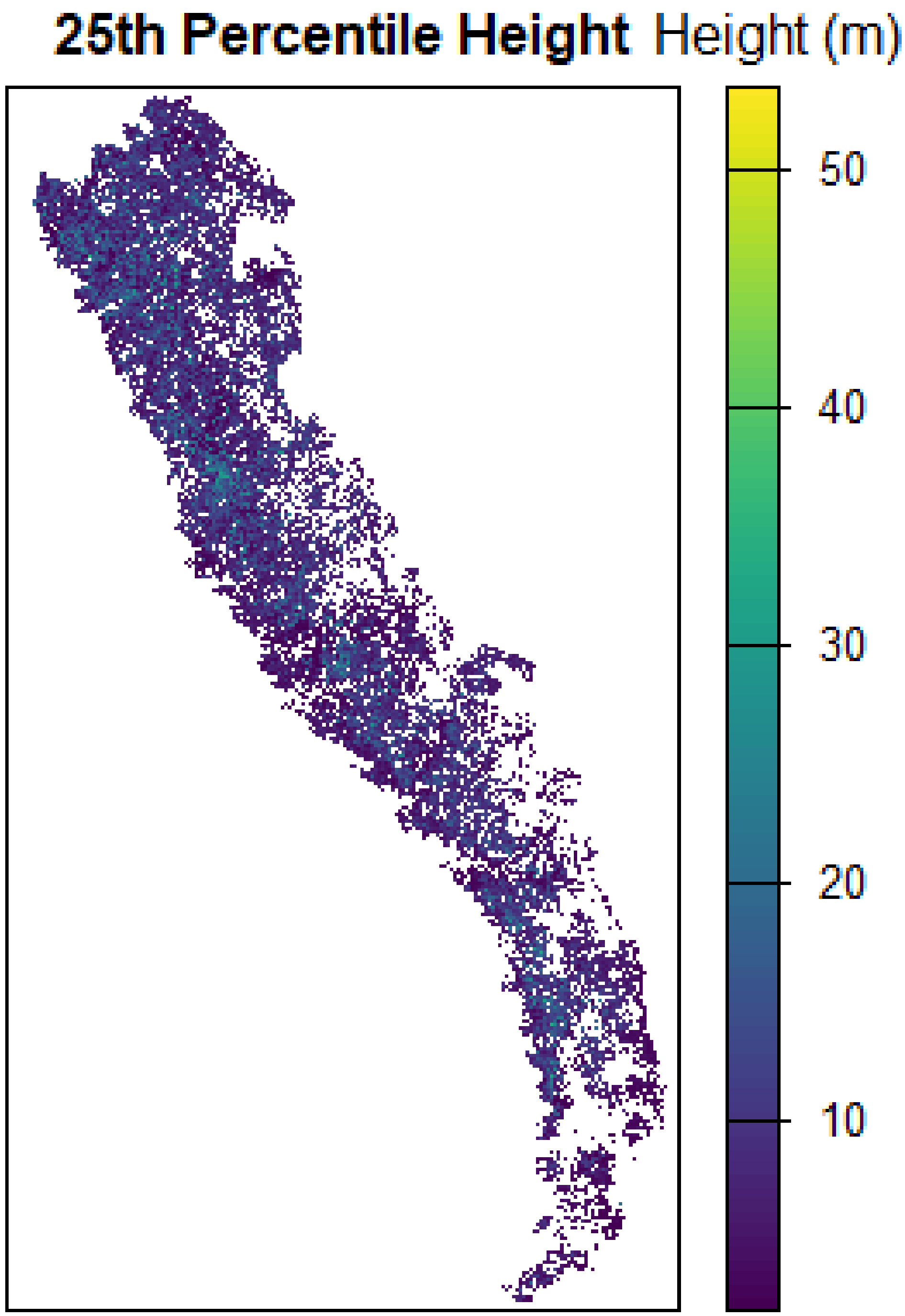
8.4.3 Maps of LiDAR Metrics



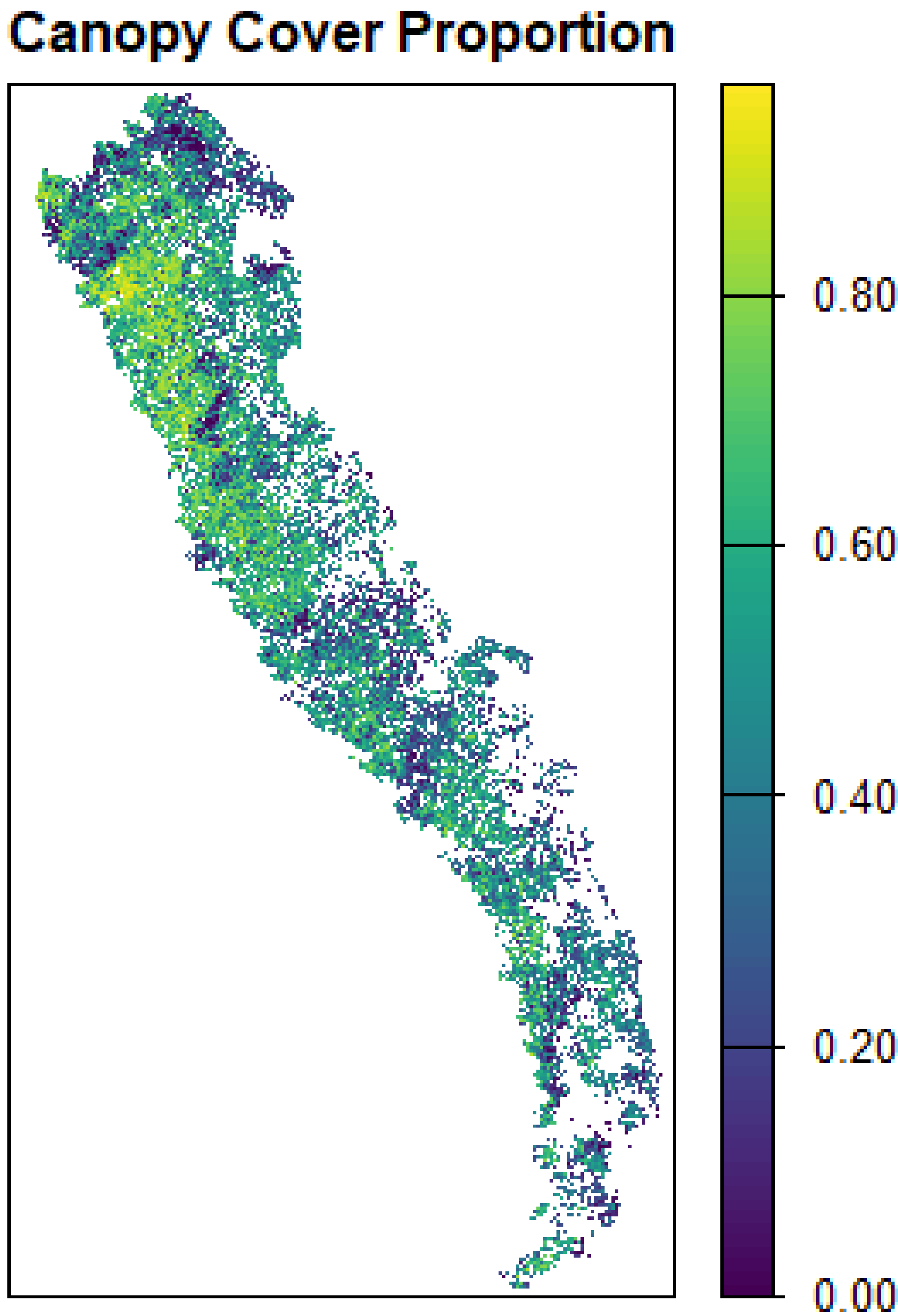
Appendix 8.4-G: Map of 95th percentile Height produced by our LiDAR processing



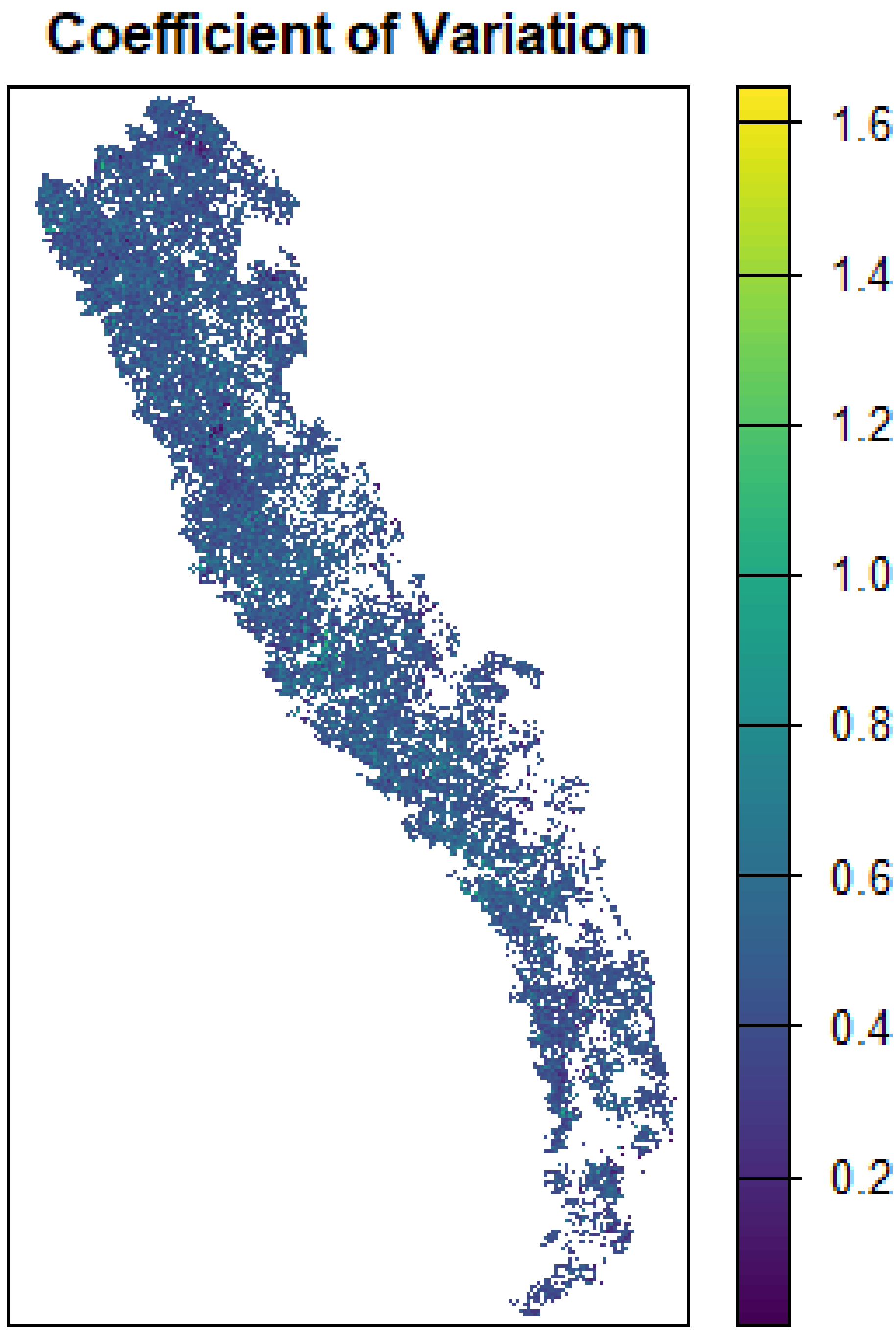
Appendix 8.4-H: Map of 50th percentile height produced by our LiDAR processing



Appendix 8.4-I: Map of 25th Percentile Height from our LIDAR processing

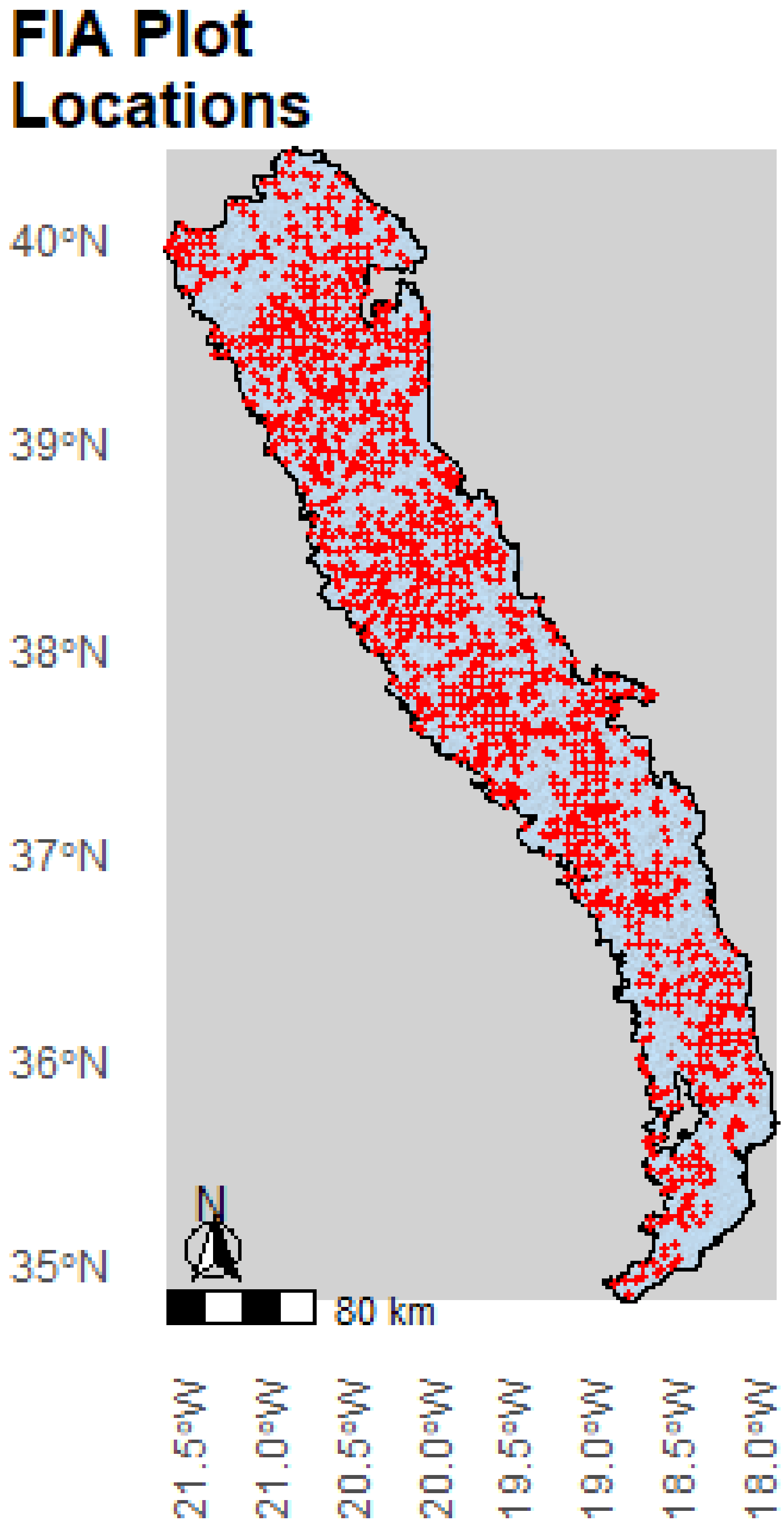


Appendix 8.4-J: Map of Canopy Cover Proportion from our LiDAR procesing

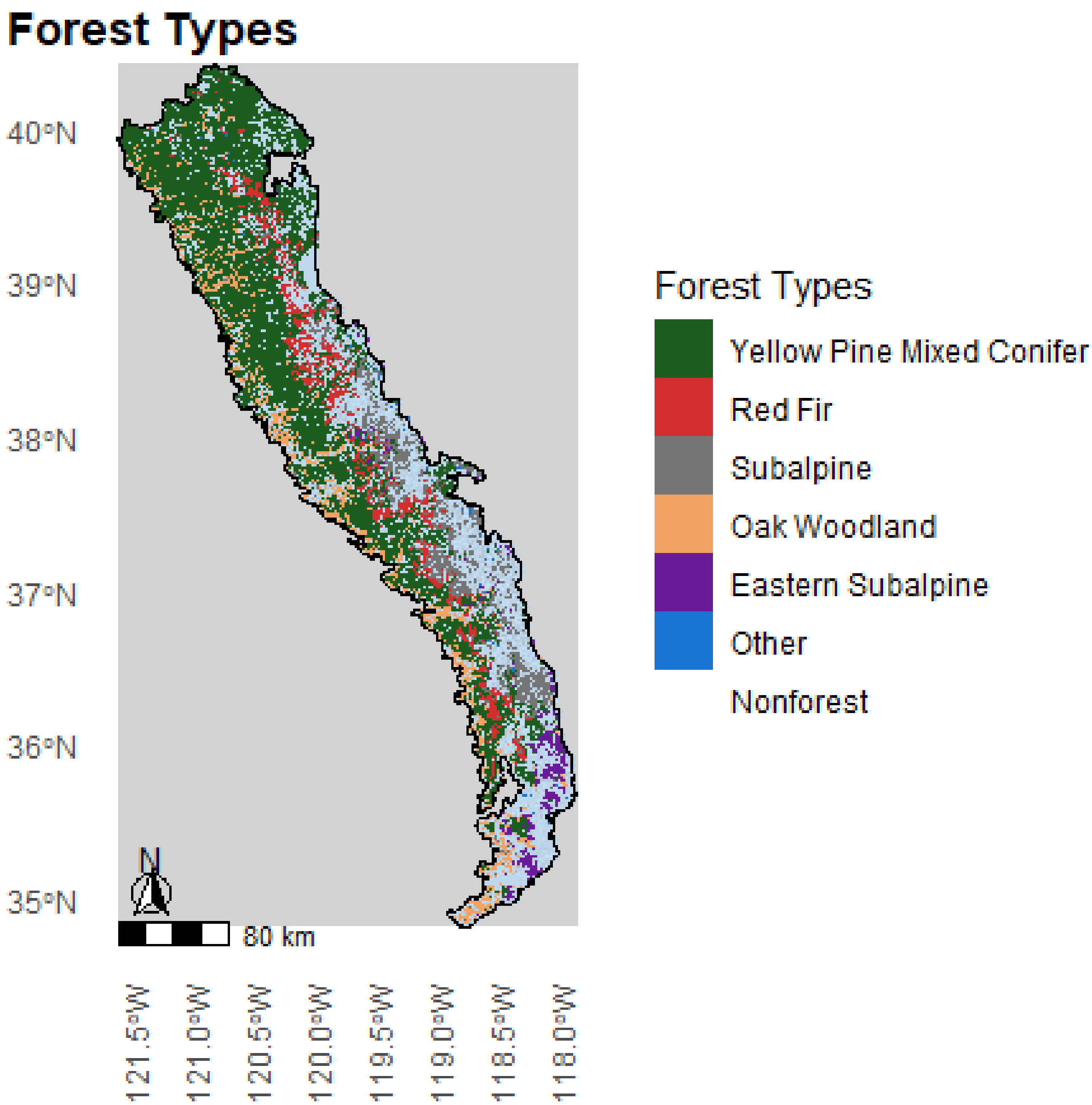


Appendix 8.4-K: Map of Coefficient of Variation from our LiDAR Processing

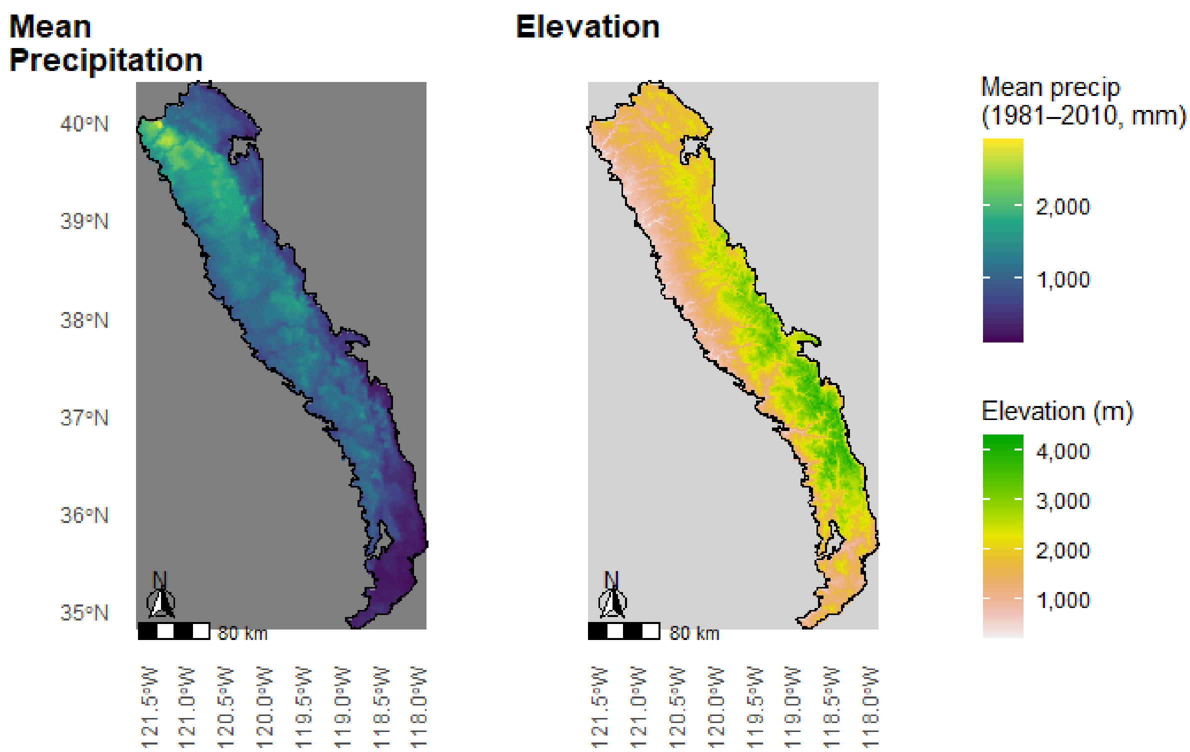
8.4.4 Other study Area Maps



Appendix 8.4-L: Full page map of FIA plots. This figure is a larger version of figure 1b. These points are the public locations



Appendix 8.4-M: Map of the Forest Types across the ecoregion. This is a larger version figure 1c



Appendix 8.4-N: Mean Precipitation and Elevation Gradient across the ecoregion. These roughly align with forest types seen in figure 1c