Mississippi CRP Forest Remote Sensing with Preliminary Global Ecosystem and Dynamics (GEDI) Mission Derived Data Products

Conservation Biology Institute

Credits

Cite as: Degagne, R., Pizzino, D., Friedrich, H, Gough, M., Joseph, G., Iovanna, R., Smith, C. and Strittholt, J. 2022. Mississippi CRP Forest Remote Sensing with Preliminary Global Ecosystem and Dynamics (GEDI) Mission Derived Data Products. CBI Technical Report 2022-1. 40 pp. (10.6084/m9.figshare.19142147)

For more info please contact: Rebecca.degagne@consbio.org

Table of Contents

Executive Summary

The United States Department of Agriculture's Conservation Reserve Program (CRP) is a federally funded conservation program, whose long-term goal is to re-establish valuable land cover to improve water quality, prevent soil erosion, improve carbon sequestration, and reduce loss of wildlife habitat. This Farm Service Agency-administered program works with farmers and landowners to implement conservation management practices on enrolled lands, with paid contracts ranging from 10 to 15 years in length.

The USDA's CRP has successfully improved the conservation value of private lands; however, the program currently lacks spatially explicit information on land cover and vegetation within CRP-enrolled tracts. Currently, there are over 46,000 CRP contracts in the state of Mississippi alone, making on-the-ground data collection difficult due to the time, resources, and expertise necessary to conduct field vegetation surveys over such extensive holdings. In partnership with the USDA FSA program, the Conservation Biology Institute (CBI) piloted a predictive modeling approach for forested lands in Mississippi participating in the CRP, employing the nuanced relationships between satellite imagery indices, enviro-climatic data, and existing georeferenced vegetation survey data (from USDA's Forest Inventory Assessment) to assess the potential for remote sensing technology to enhance CRP program outcomes.

In this pilot project, CBI initially developed predictive maps of tree height, tree density, biomass, basal area, and forest type using Random Forest machine learning models. Numerous satellite-derived indices from the European Space Agency's (ESA) Sentinel-1 and Sentinel-2 sensors, in addition to soils and topography data, were used as predictor inputs. We then refined these predictive models, focusing primarily on biomass improvements, by implementing new methods for processing Sentinel-1 imagery on the cloud computing platform Google Earth Engine (GEE); significantly updating model code; and incorporating preliminary data products derived from NASA's spaceborne LiDAR mission - the Global Ecosystem Dynamics Investigation (GEDI). We refined the GEDI LiDAR-derived data products and included them in our models, and overall accuracy for the four forest regression models ranged from 57% to 91%. The Biomass model saw the greatest improvement in accuracy with the R2 increasing by 8%, from 49% to 57%. The Basal Area and Tree Height models both had minor 1-2% increases in accuracy, while the Tree Density model had no improvement. The Forest Type classification model had a negligible improvement in overall accuracy, however, the Elm/Ash/Cottonwood class increased in accuracy by ~6%, from 64% to 70%.

The preliminary versions of mapped forest characteristics (without GEDI-enhancements) have been integrated into an easy-to-use online decision support tool, also developed by CBI, that provides USDA's staff and private landowners an opportunity to explore maps and metrics for land enrolled in the CRP. The tool provides access to pertinent spatial information for CRP tracts, as well as the ability to summarize statistics, compare metrics, and download reports for tracts across counties and watersheds.

The new GEDI data products, which are in their early stages of development, combine global GEDI LiDAR measurements with Landsat satellite imagery to provide wall-to-wall estimates of global canopy height. Currently there are large gaps in spatial coverage, but these will get filled as more

GEDI-based imagery is processed and updated. Incorporating these products as they become available into our models will increase accuracy in predicted forest metrics.

There are several pathways for future improvement and refinement of forest modeling techniques. Alternative approaches to machine learning, such as gradient boosting algorithms, may offer increased performance over the currently employed Random Forest method, given the large quantity of FIA training data available and the complex nature of forest remote sensing. Migrating additional workflows to GEE presents an opportunity to overcome current data processing limitations by leveraging distributed, cloud computing power, thus offering potential to improve data resolution and gain more information from input variables. Google Earth Engine also offers a rich, multi-petabyte catalog of satellite imagery, which would readily allow us to test additional value added by different sensors and higher resolution imagery. Lastly, criteria for filtering and selecting FIA plots could be revisited to increase the quantity of training data, which may improve model performance.

Our mapped predictions of forest metrics provide a baseline for characterizing forests within CRP tracts in Mississippi and lay a foundation for quantitatively measuring the success of conservation practices over time. Bringing this and other key data together in the multi-faceted online CRP tool allows relevant information to be analyzed, shared, and downloaded by USDA leadership and CRP managers. Future expansion of analytical and online tool functionality will provide additional information to help guide strategic management actions on existing CRP holdings and to prioritize new enrollment in the CRP. The Conservation Biology Institute's suite of powerful products rolled into an accessible online tool allows CRP to leverage these components to implement cost-effective and scientifically sound decision-making.

1. Introduction

1.1 USDA CRP Program

The United States Department of Agriculture's Conservation Reserve Program (CRP) is a federally funded land conservation program, whose long-term goal is to re-establish valuable land cover to improve water quality, prevent soil erosion, improve carbon sequestration, and reduce loss of wildlife habitat. This Farm Service Agency-administered program works with farmers and landowners to implement conservation management practices on enrolled lands, with paid contracts ranging from 10 to 15 years in length.

The USDA's CRP has successfully improved the conservation value of private lands, however, the program currently lacks spatially explicit information on land cover and vegetation within CRP-enrolled tracts. This lack of data creates challenges for program managers and landowners to characterize landscape conditions and to quantitatively measure the benefits of CRP practices over time. Currently, there are over 46,000 CRP contracts in the state of Mississippi alone, making on-the-ground data collection within CRP plots difficult due to the time, resources, and expertise required to conduct these types of field vegetation surveys. However, recent advances in remote sensing technology offer an alternative means to measuring and monitoring vegetation cover (Xie et al., 2008; Ustin and Middleton, 2021). Predictive modeling approaches are promising means of creating spatially continuous maps of vegetation by employing the nuanced relationships between satellite imagery indices, enviro-climatic data, and existing georeferenced vegetation survey data. CBI's CRP Phase I and Phase II pilot projects use a predictive modeling approach to assess the potential for remote sensing technology to quantify land cover in a cost-efficient manner and to enhance CRP program outcomes.

1.2 CRP Phase I

The purpose of CRP Phase I was to assess the application of the latest remote sensing technology to enhance forest inventory analysis, forest management, and characterize economic value of lands enrolled in the Conservation Reserve Program. For this analysis, CRP lands of interest were required to have participated for at least eight years and to be located in areas with a substantial concentration of acres enrolled under conservation practices devoted to multiple bottomland hardwood and other tree species (i.e., CP03, CP03A, CP11, CP22, CP31, and CP40) (Appendix A1).

We employed indices derived from Sentinel-1 synthetic aperture radar (SAR) data and Sentinel-2 multispectral imagery in Random Forest models to predict key forest metrics. Processing and downloading Sentinel imagery was computationally and time-intensive. The Sentinel-1 and Sentinel-2 data was downloaded through the Copernicus Open Access Data Hub. In total, ~2 terabytes of imagery (3,703 scenes) were downloaded, which took about 39 days to complete. For Sentinel-1, 591 SAR scenes (~560 gigabytes) were downloaded, taking approximately 11 days to complete. To handle the large data volume and requisite pre-processing, CBI built a custom Linux server with 25 GB of RAM, and 6TB of storage. Despite this computing power, all Sentinel imagery was processed at 20 meter resolution, rather than 10 meters, since the higher resolution data would have taken several months of continuous processing on the custom server.

For the Phase I scope of work, we created predictive maps of tree height, tree density, biomass, basal area, and forest type for Mississippi. These spatial datasets lend insight into the potential forest characteristics of lands participating in the CRP Bottomland Hardwoods Initiative and lay a foundation for quantitatively measuring the success of conservation practices over time. The mapped forest characteristics were also integrated into an easy-to-use online decision support tool, developed by the Conservation Biology Institute. This tool provides USDA staff and private landowners an opportunity to explore maps and metrics for land enrolled in the CRP. It provides access to pertinent spatial information for CRP lands, as well as the ability to summarize statistics, compare metrics, and download reports for CRP tracts across counties and watersheds. Our satellite-derived maps of forest characteristics provide a baseline for characterizing vegetation within Mississippi CRP tracts and offer the opportunity to create metrics for quantifying landscape conditions. The CRP tool also allows relevant data to be analyzed, shared, and downloaded through CBI's online mapping platform, Data Basin. Future expansion of online tool functionality will provide additional species and ecological information to help guide strategic management actions on existing CRP holdings and to prioritize lands for new enrollment in the Conservation Reserve Program.

1.3 CRP Phase II Upgrades

In CRP Phase II, we improved methods for processing Sentinel-1 data, significantly refined the model code, and experimentally incorporated GEDI spaceborne LiDAR-derived data products into the existing forest models, with a primary focus on improving biomass predictions. Firstly, the processing for Sentinel-1 SAR data was shifted to Google Earth Engine (GEE). Google Earth Engine is a cloud computing platform that hosts a multi-petabyte catalog of remotely sensed data, provides a suite of data processing functionality to efficiently query and process imagery, and offers a variety of analytical tools for image classification, change detection, time series analysis, and machine learning. In Phase I, computing Sentinel-1 textural indices was extremely computationally and time intensive; however, GEE's highly scalable computations, cloud-based resources, and data catalog facilitated faster processing and allowed for various methodologies to be tested over a short period of time in Phase II. The workflows for Random Forest modeling and input variable zonal statistic calculations were also improved via R scripting (Appendix A2). Subsequently, these updates enabled us to readily add new predictor variables, such as preliminary GEDI-derived global canopy height data, to the models. These workflow upgrades also position us well to incorporate more refined versions of GEDI data products, currently in early stages of development, upon their release over the next eighteen months.

2. Background

2.1 Advances in Mapping Forests with Remote Sensing

Over the past ten years, forest remote sensing has made major strides in the quantity and diversity of freely available, high resolution satellite imagery, as well as huge advances in accessibility of cloud-computing processing platforms. A cultural shift toward data sharing and collaborative work has contributed to increased access to code and resources to implement cutting edge analysis techniques as soon as they become available. In years past, remote sensing efforts were constrained by local computing power, costly software licenses, and expensive satellite imagery. Now, advancements in geospatial data processing facilitated by the advent of distributed cloud computing, coupled with the emergence of multiple, freely available sources of satellite imagery at relevant spatiotemporal resolutions offer new opportunities for earth observation relevant to the USDA's Conservation Reserve Program needs.

Over the past fifty years, NASA's Landsat program has been the longest record of data collection in earth observation. The Landsat mission produced the first imaging satellite to collect earth observation data based upon the spatial, spectral, and temporal characteristics of landscapes (Ustin and Middleton, 2021). The Landsat program, in continuous operation from the 1970s to the present, has created opportunities to track land cover and land use change over a large temporal scale. While Landsat provides data on almost 50 years of forest conversion, these optical, multispectral datasets provide less direct information about the structural metrics of forests, such as the quantity of carbon stored in vegetation. However, Landsat's legacy led to innovations in remote sensing that laid a foundation to execute more advanced earth observing satellite missions, including those to monitor vegetation structure, via deployment of synthetic aperture radar and spaceborne LiDAR, like GEDI.

2.2 Satellite Data Overview

2.2.1 Multispectral

In recent years, increased satellite coverage of near-global remotely sensed data has become available for mapping land cover in areas without existing field survey data, and these data sources have proven crucial for a variety of ecological applications (Ustin and Middleton, 2021). When evaluating remotely sensed data sources for a given application, the spatial resolution, temporal cadence of collection (i.e., return interval), and spectral characteristics of the sensor should all be considered. For optical imagery, NASA satellites offer a generous historical archive - a wide array of spectral characteristics relevant to vegetation mapping at a range of spatial resolutions, from low-resolution (250 m) Moderate Resolution Imaging Spectroradiometer (MODIS) to medium-resolution (30 m) Landsat archive available from 1972-present. The recent addition of the European Space Agency's (ESA) Sentinel-2's higher resolution (10-20 m) multispectral imagery archive is particularly valuable, with various products available from mid-2015 onward (ESA Copernicus, 2020).

2.2.2 Synthetic Aperture Radar (SAR)

In addition to optical imagery, ESA's Sentinel-1 sensor offers synthetic aperture radar data, which is not hindered by cloud cover and can be used to measure land surface texture (ESA Copernicus, 2020). This suite of sensors provides a systematic collection of medium-spatial resolution, remotely sensed data over time, which opens up possibilities for continuous monitoring of vegetation and enables the capture of time-specific snapshots of vegetation type and condition.

2.2.3 Spaceborne LiDAR

Global Ecosystem Dynamics Investigation (GEDI) is a spaceborne LiDAR instrument that takes high-resolution observations of the earth's surface. Created to take a snapshot of biomass and canopy height of the world's temperate and tropical forests, the GEDI sensor was launched in 2018 from Cape Canaveral, Florida on a planned 24-month mission. The mission will sample ~4% of the earth's surface and, as of September 2021, GEDI has captured ~5 billion of the expected 10 billion cloud-free observations (Duncanson, 2021). Once the mission is complete, the data collected by GEDI will represent the most accurate sample of forest canopy height, canopy structure, and biomass at a global scale. GEDI's measurements of tree height and biomass have been validated by the scientific community using airborne LiDAR scanning observations from around the world.

Methods for estimating biomass have evolved from raw field measurements to local LiDAR, and now, to spaceborne LiDAR like GEDI. Before remote sensing methods were developed to estimate biomass, aboveground biomass estimation was dependent on destructive sampling techniques and allometric equations, or estimation via volumetric methods (Lu, 2006). However, these types of allometric models are known to have high uncertainties due to their reliance on in-situ field measurements, and high model uncertainty depending on the specific allometric model employed (Duncanson et al., 2017; Stovall and Shugart, 2018). In the early 2000s, models using multispectral remotely sensed data, such as Landsat, emerged as a popular method for biomass estimation. However, spectral imagery adds limited value to biomass estimation, since earth observation is frequently hindered by cloud cover and it provides minimal detail about forest structure. In fact, researchers found that textural information in an image was more important than spectral information for predicting biomass in areas with complex vegetation stand structure (Lu, 2006). Due to the limitations associated with obtaining cloud-free spectral imagery in many regions of the world, radar and LiDAR have become the most feasible methods for collecting earth observations, irrespective of weather conditions. LiDAR is an excellent method for characterizing the vertical structures of forests, including canopy height, forest volume, and aboveground biomass (Man et al., 2014). Until recently, LiDAR and radar sampling were expensive and limited to local and regional scales. Fusing local LiDAR with SAR provided a way to upscale biomass estimation over larger spatial scales (Wulder et al., 2012). In the present day, the advent of the GEDI mission ushers in a new era of remotely sensing forest structure, including metrics of canopy height, leaf area index, and biomass, at a global scale (University of Maryland, GLAD, 2021).

GEDI data will serve as a snapshot in time to set a baseline measurement of carbon to inform future carbon losses via deforestation and land conversion. GEDI's precise sampling of forest canopy height, canopy vertical structure, and surface elevation will greatly advance our understanding of carbon and water cycling processes, biodiversity, and habitat structure (Dubayah et al., 2020).

The GEDI mission has been approved for an additional extension and is now expected to continue recording observations through the end of 2022. The GEDI sensor is located on the Japanese Experiment Module-Exposed Facility (JEM-EF) of the International Space Station. Further details on the GEDI mission and sensor are included in Appendix A3.

3. Methods

3.1 Workflow Overview for Modeling

Our overall workflow to derive predictions from each of the forest metric models (Biomass, Basal Area, Tree Height, Tree Density, Forest Type) consisted of processing FIA plot survey training data, processing model input variables (satellite, topography, and soils), extracting model input variable values at FIA plot locations, preparing data for modeling, conducting Random Forest modeling, and lastly, visualizing model predictions as output raster images in the CRP tool (Figure 1).

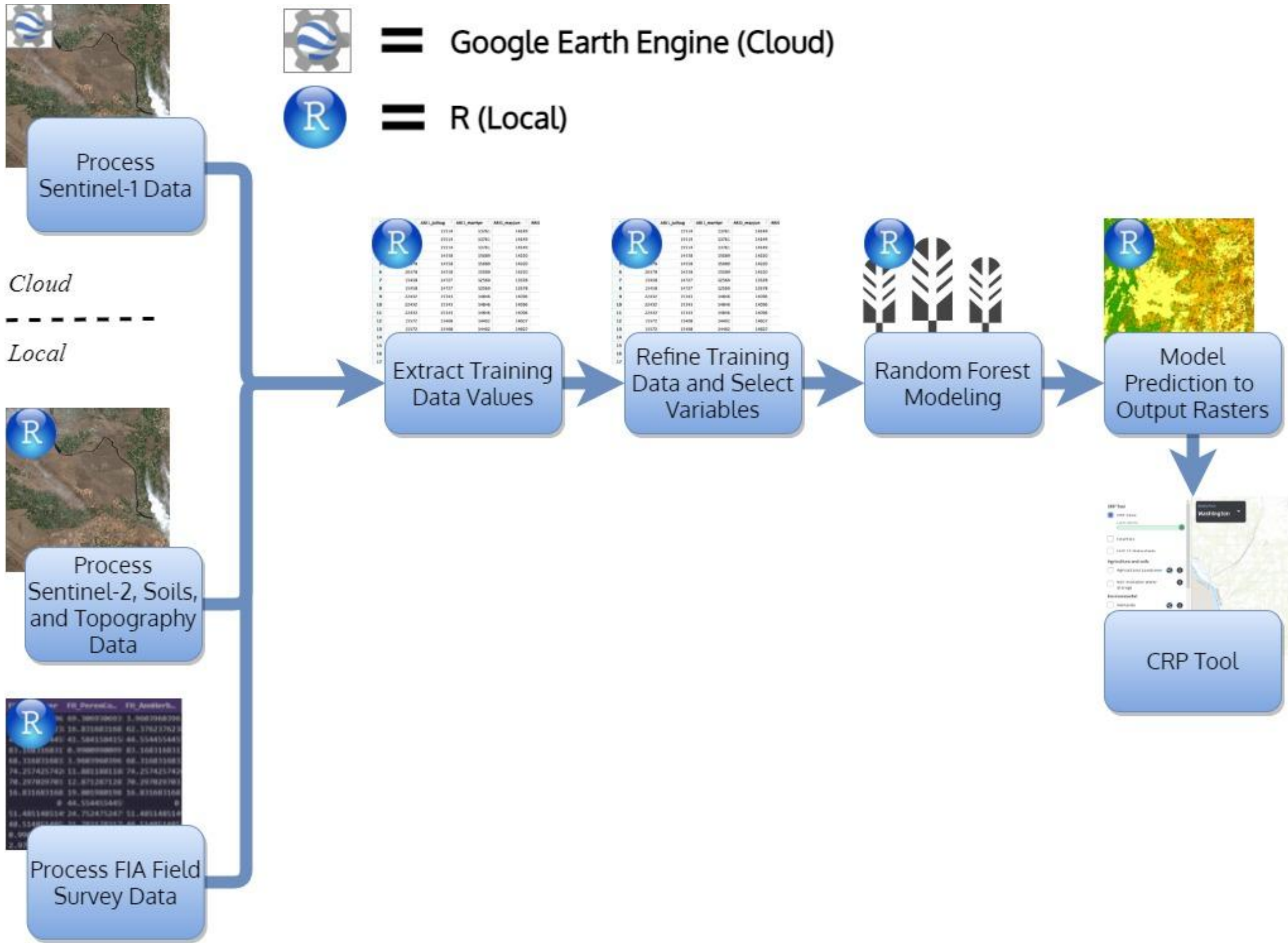

Figure 1. Overview of CRP forest metric modeling workflow to derive predictions for Basal Area, Tree Height, Tree Density, Biomass, and Forest Type.

3.2 FIA Field Survey Training Data

The Forest Inventory and Analysis Program (FIA) is a long-running U.S. Forest Service program that collects, analyzes, and reports information on the status and trends of America's privately owned forests. This information includes data such as the extent and spatial location of forests, the species, size, and health of trees, total tree growth, tree mortality, and removals by harvest ("Forest Inventory and Analysis National Program", 2015). This information can then be used to evaluate wildlife habitat conditions, assess the sustainability of ecosystem management practices, and support planning and decision making activities. As of 1999, the FIA program shifted from a periodic survey to an annual survey, and expanded the scope of their data collection to include an additional suite of attributes on a subsample of their plots such as soils, understory vegetation, tree crown conditions, downed woody material, and invasive species ("Forest Inventory & Analysis: What is Forest Inventory and Analysis?", 2018). FIA plot location and data attributes for privately held lands are highly confidential due to privacy concerns. Further details about FIA sampling design and plot survey design are available in Appendix A4.

Mississippi FIA data with georeferenced survey locations was obtained for the years 2014 - 2017 from the USFS Southern Research Station. Data were filtered so that forested areas had only one condition class across each plot in order to exclude plots that were clear cut or misidentified as forested due to human error or erroneous GPS coordinates. After filtering, a total of 2,090 FIA plots were available for Random Forest model training and validation.

For the regression forest models (Biomass, Tree Height, Tree Density, and Basal Area), training and testing data were pulled from 2017 FIA data, 2019 field data, and digitized plots of cropland, water, and urban development, for a total of 980 observations (Appendix A5). This resulted in 686 training points and 294 testing points when sampled with a 70% training, 30% testing data split.

In comparison, the Forest Type classification model used training and testing data pulled from 2014 - 2017 FIA data, 2019 field data, and digitized areas of cropland, water, and urban development. There were a total of 2,338 data points used in this model, split 70% training, 30% testing.

3.3 Spatial Input Variable Processing

Spatial input variables were obtained from multiple data sources including Sentinel-1, Sentinel-2, soils, topography, and a preliminary GEDI-derived canopy height dataset. Zonal statistics were tabulated in R software for all input variables across each of the 2,090 FIA training plots, and digitized plots of cropland, urban, and water areas.

3.3.1 Sentinel-1 and Sentinel-2

We used Sentinel-1 and Sentinel-2 data (Table 1) available through the European Space Agency (European Commission's Copernicus Earth Observation Program). Due to local computing constraints, all Sentinel imagery was processed at a 20-meter resolution.

Table 1. Spatio-temporal and radiometric resolutions of Sentinel-1 and Sentinel-2.

Sensor	Spatial Resolution	Temporal Resolution	Radiometric Resolution	Collection Start Date	Surface Reflectance Available
Sentinel-1	10 m	6 days	Dual polarization (VV+VH)	April 2014 (Sentinel-1A) April 2016 (Sentinel-1B)	October 2014 onward
Sentinel-2	10 m, 20 m	5 days	Visual bands, NIR, red-edge, SWIR	June 2015 (TOA available)	December 2018 onward

Sentinel-1 imagery was acquired from the GEE data catalog (“Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected”, n.d.) for 2017-2019. The imagery was composited by month for the full time period using four reducers: mean, maximum, minimum, and standard deviation (i.e., for January, the mean value would be an average across all three years of imagery). This produced a total of 48 descriptive input variables. Next, the Sentinel-1 mean monthly composites were used as an input to a gray-level co-occurrence matrix (GLCM) texture analysis within GEE to produce seven indices (Haralick et al., 1973; Appendix A6; Table A6-1). The GLCM analysis used a 1x1 moving window function to aggregate the synthetic aperture radar information, identifying statistical groupings and similarities amongst neighboring pixels.

Sentinel-2 multispectral imagery was downloaded through the Copernicus Open Access Data Hub. Cloud cover and atmospheric effects were removed from the imagery and data were normalized to Bottom-of-Atmosphere reflectance at 20 m resolution (Level 2A). The corrected imagery was then combined into bi-monthly, seasonal (Leaf-off, Greening, Leaf-On, Senescence; Figure 2), and seasonal difference mosaics using the Sen2Mosaic software.

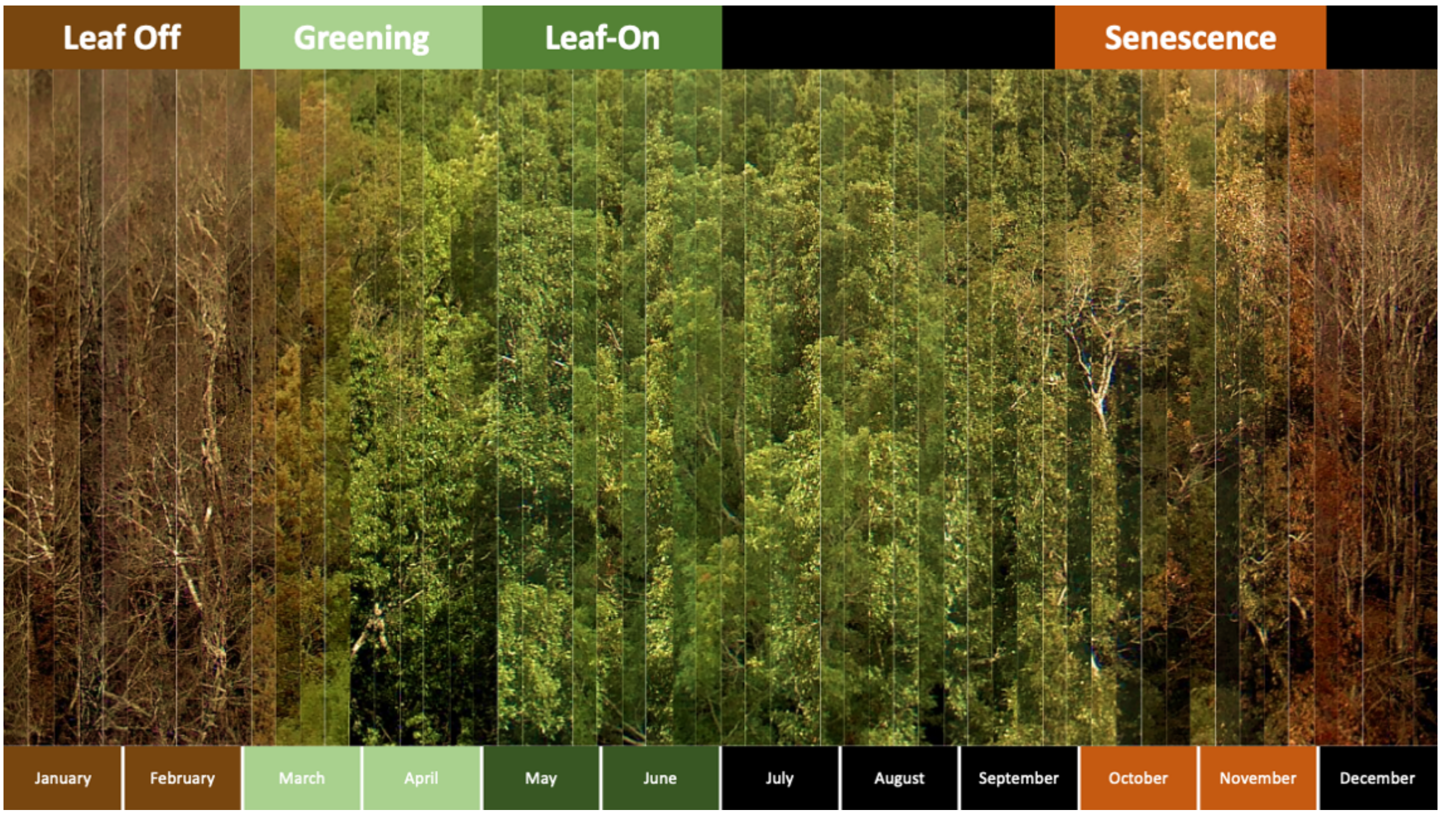

Figure 2. Figure of seasonal breaks for Sentinel-2 imagery and a time slice of deciduous tree phenology across the months of the year. Time slice imagery courtesy of PhenoCam.

From each of these temporal and seasonal mosaics, sixteen vegetation indices were derived using ESRI ArcGIS software and Python scripts (Table A6-2). These indices capture the spectral characteristics of vegetation, helping differentiate forest types and predict vegetation characteristics via Random Forest modeling.

3.3.2 GEDI Derived Global Canopy Height

We used a novel preliminary dataset of global canopy height for 2019 to incorporate 30-meter resolution estimates of tree height into the forest models. The dataset, from Potapov et al., was created using a fusion of GEDI LiDAR observations and Landsat imagery (2021). This fusion data product provides wall-to-wall coverage, whereas GEDI data products have large gaps, spatially; recall that GEDI is a global sample, not a census (Figure 3). Potapov et al. validated the global canopy height map with GEDI data and airborne LiDAR data observations. In this dataset prototype, there are known issues related to GEDI data quality and the availability of Landsat optical time-series data. The raw GEDI data have known errors on steep slopes where forest height is overestimated, and tall buildings can be confounded with tree height in more suburban areas. This data fusion exercise was a proof of concept to create an analysis-ready dataset, and the authors hope to resolve the canopy height saturation problem for trees greater than 30 meters in height in future versions of this data product. However, to temporarily patch the tree and urban building misprediction problem, we masked all tree heights greater than 30 meters before incorporating this data into the forest models. Mississippi does not have many tree species that can grow taller than 100 ft (30.48 m), thus this masking was appropriate as a temporary resolution to the dataset's limitations.

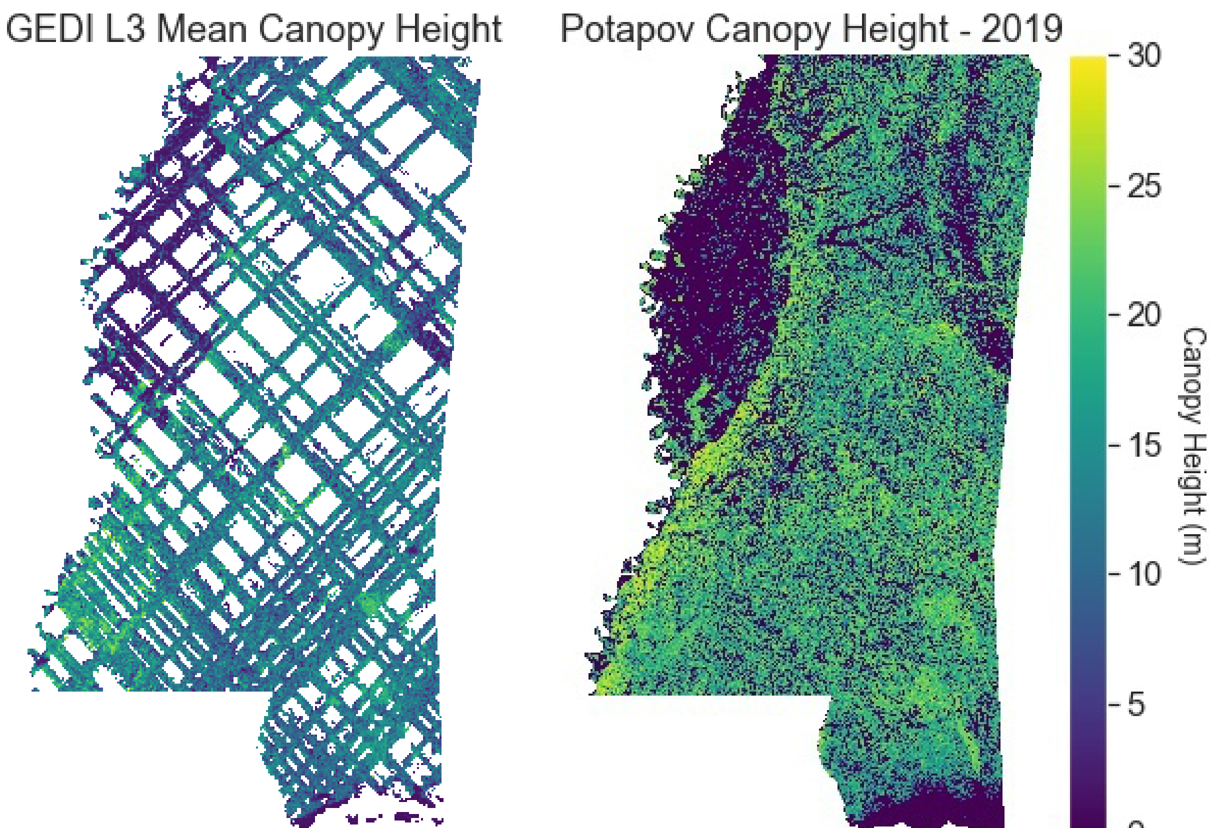

Figure 3. GEDI L3 Gridded Mean Canopy Height (left) versus the Potapov Canopy Height in 2019 (right) coverage in Mississippi. Note that the GEDI L3 data has many large gaps, while the Potapov data has wall-to-wall coverage.

3.3.3 Soils and Topography Data

Soils variables describing soil composition, pH, drainage, and electrical conductivity, among others, were extracted from the NRCS SSURGO soils dataset. Topographic variables including elevation were obtained from the USGS National Elevation Dataset (NED) dataset. Variables such as elevation and soil composition (Table 2) are known to influence the distribution of different tree species and have the potential to add value to the Random Forest models.

Table 2. Soils and Topography data sources and data variables.

Data Source	Variable
SSURGO Soils Data	Soil Percent Clay 0-150cm Soil Percent Sand 0-150cm Soil Percent Silt 0-150cm Soil Organic Matter Soil Dominant Soil Drainage Class Soil Presence of Hydric Components Soil pH (1-to-1 Water) Soil Electrical Conductivity
Topography Data	Elevation (meters) Slope (degrees) Aspect (degrees)

3.4 Random Forest Modeling

The Random Forest modeling algorithm was used to estimate forest type, tree density, basal area, tree height, and aboveground biomass of merchantable live trees across Mississippi forests. This machine learning algorithm is designed to find and model patterns in data. More specifically, it's a supervised ensemble method; an aggregation of individual models trained using known target outputs and whose collective predictive power is greater than any constituent model (Brieman, 2001). Random Forest algorithms have been used successfully with multispectral imagery, synthetic aperture radar, and topographic and environmental variables to model all of the targeted forest metrics. These models work well for applications in remote sensing as they are robust to overfitting, which allow the models to be applied to making predictions outside the training data area. Random Forest models are less sensitive to outliers compared to other methodologies and are nonlinear, allowing them to work with high dimensional data (Belgiu and Drăguţ, 2016). We used the software platform R and the "ranger" Random Forest R package (Wright et al., 2021), which is based on the original "random forest" package from Breiman et al. (2018).

To evaluate the validity of modeled outputs, we performed a cross-validation assessment to document the overall accuracy of each model, using an independent portion of the training data withheld for this purpose. These accuracy metrics were then compared to the original forest modeling results from Phase I (Appendix A7). We assessed model accuracy using R^2, root mean squared error (RMSE), and mean absolute error (MAE) metrics. A model's R^2 provides a "goodness of fit" measure for the predictions as compared to the observations on a scale of 0 to 1. For example, an R^2 value of 0.5 indicates that the model's predictions explain 50% of the variation in observations. In comparison, the RMSE metric is the average deviation of the predictions from the observations. This metric gives weight to large, rare errors in our predictions. RMSE is useful for understanding how well the model is performing, in units of the output variable. The MAE metric is the average of the absolute values of the errors, in units of the output variable. It tells us what size error we can expect from our predictions on average. It is similar to RMSE, but does not weight the errors. Random Forest models have natural variability in accuracy due to the random split of the testing and training datasets. Therefore, averaging the model's R^2, RMSE, and MAE over multiple model iterations informs the model's actual performance.

4. Modeling Results

4.1 Model Accuracy Assessment Comparison

The overall accuracy of the four regression forest models, with the inclusion of the preliminary GEDI fusion data (Potapov Canopy Height), ranged from 57% to 91% (Table 3). The Biomass model saw the greatest improvement in accuracy with the R^2 increasing by 8%, from 49% to 57%. The Basal Area and Tree Height models both had minor 1-2% increases in accuracy, while the Tree Density model had no significant improvements. Furthermore, the root mean squared error (RMSE) and the mean accuracy error (MAE) decreased with the inclusion of the Potapov Canopy Height data in the Basal Area, Tree Height, and Biomass models. The Forest Type classification model had a negligible 0.89% improvement in overall accuracy, however, the Elm/Ash/Cottonwood class increased in accuracy by ~6%, from 64% to 70% (Table A7-1).

Table 3. Forest model improvements with Potapov (2021) Global Canopy Height GEDI fusion data. The Basal Area, Tree Height, and Biomass models gained value from the Potapov data, while the Tree Density model showed no improvement in the R^2.

	Original Models			**Phase II Models**			
Metric	**R^2**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**% Improvement in R^2**
Basal Area (square ft/acre)	0.66	32.53	21.48	0.68	30.58	19.87	1.85
Tree Height (ft)	0.90	9.97	6.61	0.91	9.31	5.84	1.37
Tree Density (trees/acre)	0.67	59.25	37.91	0.65	65.06	41.47	No Improvement
Biomass (Dry Merchantable) (lbs/acre)	0.49	39919.54	25874.74	0.57	35156.67	22666.64	7.45

4.2 Model Variable Importances

The Biomass model was primarily driven by the RENDVI and ARI Senescence seasonal indices, which provide spectral information about leaf pigmentation changes during autumn (Figure 4). The Potapov Global Canopy Height data was found to be the sixth most important input variable on average. Other variables of lesser importance included Greening NDRE and Leaf On ARI multispectral indices, as well as the Sentinel-1 Minimum VV and Minimum VH SAR variables.

The Potapov Canopy Height data was found to be in the top six most important variables for predicting biomass, basal area, tree height, and tree density, but was not found in the top fifty most important variables for the Forest Type model. Variables composited by seasonality (Senescence, Greening, Leaf-on, and Leaf-off) were the top three most important for multiple forest models (Appendix A8). We found Senescence to be the most important seasonal variable for the Tree Height, Tree Density, Biomass, and Basal Area models, while Greening was the most important for the Forest Type model.

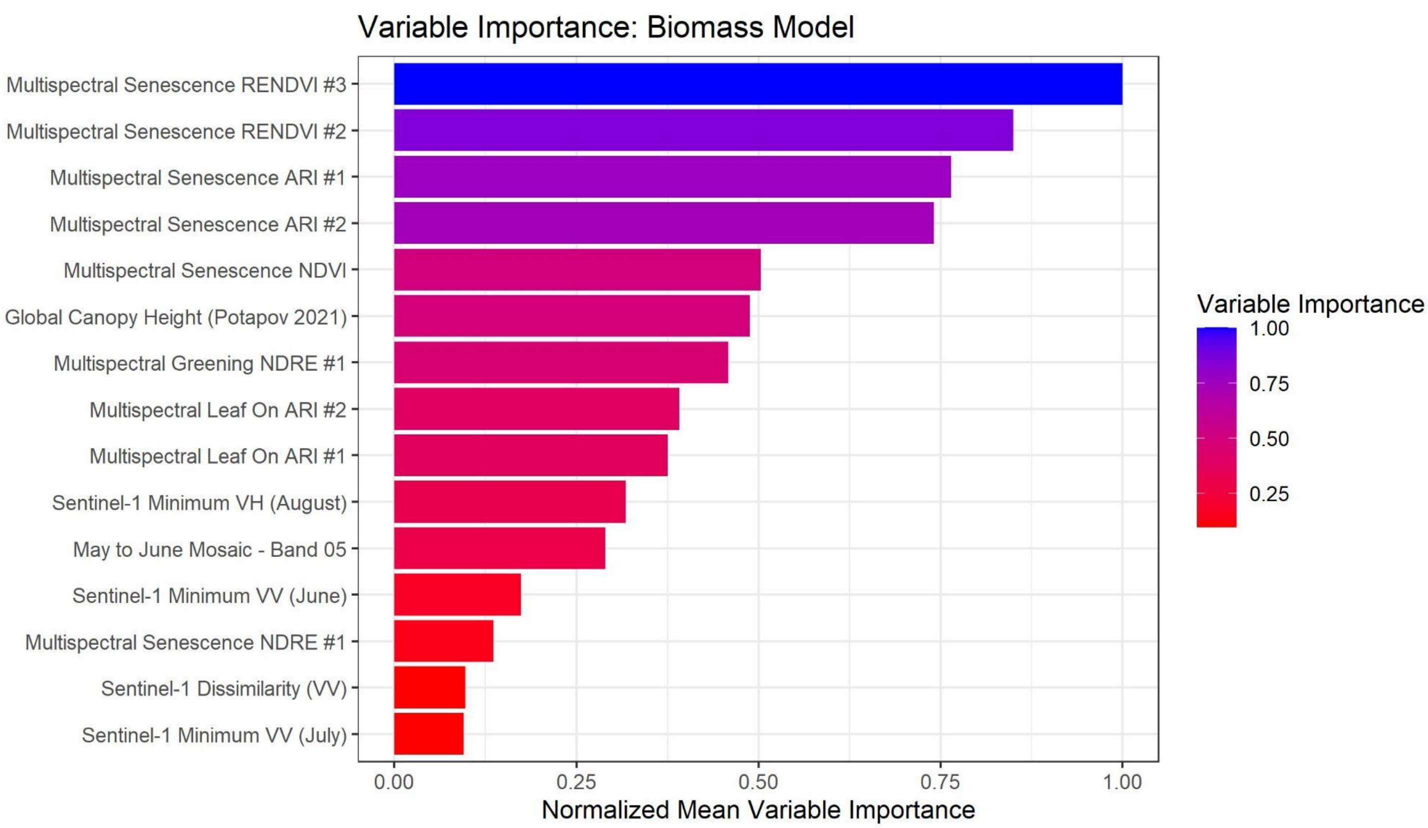

Figure 4. Chart of normalized mean variable importance for the fifteen most important variables, averaged across five iterations of the Biomass model.

5. Discussion & Conclusion

5.1 Outcomes - Increased Model Performance

Phase II modeling saw model performance significantly improve in terms of processing efficacy, runtime, and accuracy, and also saw the inclusion of a preliminary, novel GEDI-derived global canopy height dataset. The Biomass model likely gained the most value from this dataset because forest biomass is greatly influenced by tree height (Proulx, 2021). The Basal Area and Tree Density models could likely gain more value from other GEDI-derived datasets that describe additional metrics of vertical forest structure.

Our findings of increased model accuracy after inclusion of GEDI-derived data are in line with promising results from other early studies that used GEDI to quantify forest structure and disturbance (Spracklen and Spracklen, 2021; Guerra-Hernández and Pascual, 2021). Since only a small proportion of GEDI data are currently available from the total 10 billion cloud-free observations that the mission plans to collect, the complete GEDI LiDAR dataset will likely add even more value to modeling CRP forest characteristics.

5.2 GEDI Limitations

While GEDI shows promise for global tree height and biomass estimation, NASA's direct GEDI data products have limitations due to the nature of the sensor and data acquisition methods. Because the sensor is renting space on the International Space Station, the pattern and coverage of data collection across the surface of the earth is entirely random, resulting in large gaps in coverage. These gaps can be observed as no-data areas in both the footprint level and gridded products (Figure 3). As of now, NASA has no plans to create a GEDI product with wall-to-wall coverage (i.e., coverage without gaps), as GEDI was intended to be a survey, not a census, of the world's biomass (Healey et al., 2020). However, independent researchers are tackling this limitation by combining GEDI with data from other satellite sensors to create exciting data fusion products (Duncanson, 2021).

5.3 GEDI Fusion Potential

We anticipate new GEDI data fusion products will add subsequent value to the forest models that are heavily influenced by forest structure or forest height. Given the 2-7% improvement in our R^2 metrics with inclusion of the GEDI-derived Global 2019 Canopy Height data from Potapov et al., refined GEDI fusion datasets may add even more value (Appendix A9). While GEDI fusion data using optical imagery like Landsat has been the most forthcoming, GEDI X SAR fusion datasets (e.g. TanDEMX and GEDI fusion) will likely have greater potential because of the added structural information that SAR provides. We also expect high-resolution GEDI fusion products, such as GEDI X Sentinel-2 maps of canopy height and biomass, to be released in the next few years (Lang et al., 2021). Looking toward the future, these GEDI fusion datasets might also act as a source for validating tree height and biomass data to supplement field measurements from annual FIA surveys.

6. Future Work

6.1 Model Refinements

In addition to inclusion of forthcoming GEDI-derived data fusion products, there are several areas identified for future improvement and refinement of forest modeling techniques. Alternative approaches to machine learning, such as gradient boosting algorithms, may offer increased performance over the currently employed Random Forest method, given the large quantity of FIA training data available and the complex nature of forest remote sensing.

Migrating additional workflows to GEE and utilizing its Python API presents an opportunity to overcome current data processing limitations by leveraging distributed, cloud computing power, thus offering potential to improve data resolution and gain more information from input variables. In particular, computation of SAR textural indices could be improved, since source code for computing more sophisticated SAR indices are now open source.

Google Earth Engine also offers a rich, multi-petabyte catalog of satellite imagery, which would readily allow us to test additional value added by different sensors and higher resolution imagery, which would allow us to test the potential additional value added by 10-meter resolution Sentinel imagery. Lastly, criteria for filtering and selecting FIA plots could be revisited to increase the quantity of training data, which may improve model performance.

6.2 Integration of Forest Mapping into the Online CRP Tool

Our mapped predictions of forest metrics provide a baseline for characterizing forests within CRP tracts in Mississippi and lay a foundation for quantitatively measuring the success of conservation practices over time. Bringing this and other key data together in the multi-faceted online CRP tool allows relevant information to be analyzed, shared, and downloaded by USDA leadership and CRP managers. Future expansion of analytical and online tool functionality will provide additional information to help guide strategic management actions on existing CRP holdings and to prioritize new enrollment in the CRP. The Conservation Biology Institute's suite of powerful products rolled into an accessible online tool allows CRP to leverage these components to implement cost-effective and scientifically sound decision-making.

7. References

Baston, D. 2021. Package ‘exactextractr.’

Belgiu, M., and L. Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114:24–31.

Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2018. Package ‘randomForest.’

Burkman, B. 2005, February 3. Forest inventory and analysis sampling and plot design.

Dubayah, R., J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock, P. Jantz, S. Marselis, P. L. Patterson, W. Qi, and C. Silva. 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography. Science of Remote Sensing 1:100002.

Dubayah, R. O. 2017. Global Ecosystem Dynamics Investigation mission status.

Dubayah, R. O. 2020. GEDI L1B geolocated waveforms product (GEDI01_B).

Duncanson, L. 2021, September 23. The Global Ecosystems Dynamics Investigation (GEDI) mission: Concepts, methods and applications. Remote.

Duncanson, L., W. Huang, K. Johnson, A. Swatantran, R. E. McRoberts, and R. Dubayah. 2017. Implications of allometric model selection for county-level biomass mapping. Carbon Balance and Management 12:18.

Forest Inventory & Analysis: What is Forest Inventory and Analysis? 2018, November 2. USDA Forest Service.

Forest Inventory and Analysis National Program. 2015, November 9. https://www.fia.fs.fed.us/.

Guerra-Hernández, J., and A. Pascual. 2021. Using GEDI lidar data and airborne laser scanning to assess height growth dynamics in fast-growing species: a showcase in Spain. Forest Ecosystems 8:14.

Haralick, R., K. Shanmugam, and I. Dinstein. (1973). Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, 3(6):610-621.

Healey, S. P., Z. Yang, N. Gorelick, and S. Ilyushchenko. 2020. Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation. Remote Sensing 12:2840.

Herbert, J. K. (n.d.). Tandem-L Interferometric Radar Mission. ESA Earth Observation Portal.

Lang, N., N. Kalischek, J. Armston, K. Schindler, R. Dubayah, and J. D. Wegner. 2021. Global canopy height estimation with GEDI LIDAR waveforms and Bayesian deep learning. arXiv:2103.03975 [physics].

Lu, D. 2006. The potential and challenge of remote sensing-based biomass estimation. International

Journal of Remote Sensing 27:1297–1328.

Man, Q., P. Dong, H. Guo, G. Liu, and R. Shi. 2014. Light detection and ranging and hyperspectral data for estimation of forest biomass: a review. Journal of Applied Remote Sensing 8:081598.

Potapov, P., X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C. E. Silva, J. Armston, R. Dubayah, J. B. Blair, and M. Hofton. 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment 253:112165.

Proulx, R. 2021. On the general relationship between plant height and aboveground biomass of vegetation stands in contrasted ecosystems. PLOS ONE 16:e0252080.

Qi, W., S. Saarela, J. Armston, G. Ståhl, and R. Dubayah. 2019. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. Remote Sensing of Environment 232:111283.

Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected. (n.d.). https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD.

Spracklen, B., and D. V. Spracklen. 2021. Determination of Structural Characteristics of Old-Growth Forest in Ukraine Using Spaceborne LiDAR. Remote Sensing 13:1233.

Stovall, A. E. L., and H. H. Shugart. 2018. Improved Biomass Calibration and Validation With Terrestrial LiDAR: Implications for Future LiDAR and SAR Missions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11:3527–3537.

United States Department of Agriculture. 2018. Agricultural Resource Conservation Program 2-CRP (Revision 5). Retrieved from https://www.fsa.usda.gov/Internet/FSA_File/2-crp_r05_a35.pdf

University of Maryland, GLAD. 2021, September 29. GEDI Data Products. https://gedi.umd.edu/data/products/.

Ustin, S. L., and E. M. Middleton. 2021. Current and near-term advances in Earth observation for ecological applications. Ecological Processes 10:1.

Wright, M. N., S. Wager, and P. Probst. 2021. Package ‘ranger.’

Wulder, M. A., J. C. White, R. F. Nelson, E. Næsset, H. O. Ørka, N. C. Coops, T. Hilker, C. W. Bater, and T. Gobakken. 2012. Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment 121:196–209.

Xie, Y., Z. Sha, and M. Yu. 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology 1:9–23.

8 Appendices

A1. CRP Forest Practices

Table A1-1. CRP forest practices and brief descriptions (USDA, 2018).

Conservation Practice	Description
CP03	Tree Planting
CP03A	Hardwood Tree Planting
CP11	Vegetative Cover, Trees Already Established
CP31	Bottomland Timber Establishment on Wetlands
CP40	FWP Aquaculture Wetland Restoration

A2. Random Forest Model & Code Info

To improve data extraction processing times, all variables from Sentinel 1 and 2, and other raster datasets were tabulated for each FIA training plot using the ExactExtractR R package (Baston 2021). The ExactExtractR package summarizes the raster values within each polygon. Unlike traditional methods for calculating zonal statistics, the exact_extract function can even handle fractions of raster cells that are covered by a polygon. This new package allowed us to retire the “trim mean” function which previously excluded the top and bottom 20% of raster values from the summary mean calculation. In addition, this R package is built upon a C++ framework which greatly improves the processing time to extract data.

Second, to improve the clarity and readability of the existing model code, each model was modularized to a separate script. The code for each model was reviewed and edited to become more concise, which in turn improved the processing times for all models. These changes provide additional benefits for future iterations of these models by improving the ability to incorporate new data variables.

A3. GEDI Technical Details

GEDI has three lasers that take 8 beam transects that are spaced 600m apart, sampling a 25m footprint every 60 meters along each transect (Figure A3-1). Four of the beams are full power lasers that take the most accurate measurements, while the last four lasers are less powerful but provide more coverage.

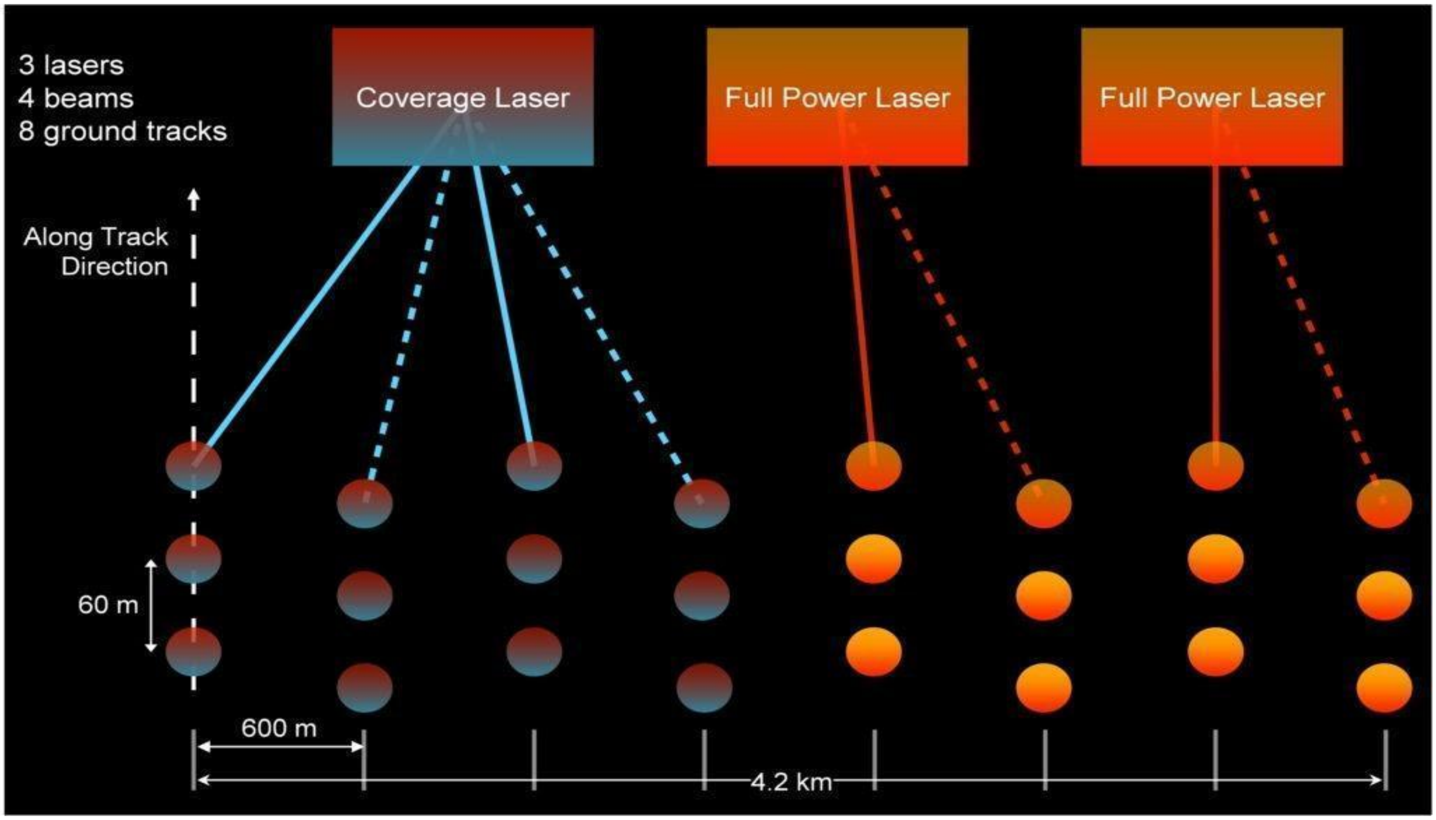

Figure A3-1. The GEDI beam pattern consists of three lasers (Dubayah 2020).

GEDI's lasers fire pulses of near-infrared energy at the surface of the earth, where each pulse is reflected by leaves, branches, and the ground within each 25 m laser footprint (Figure A3-2). The returned waveform can be thought of as a histogram of reflecting surface heights that represent the different layers of the forest canopy and the ground. The GEDI waveforms are processed to derive metrics of canopy height, relative height, and ground elevation. Additional processing is applied to extract metrics like leaf density, canopy cover, and aboveground biomass (Dubayah 2020). GEDI's lasers are not able to see through clouds, which limits the ability to take observations in cloudy regions of the world.

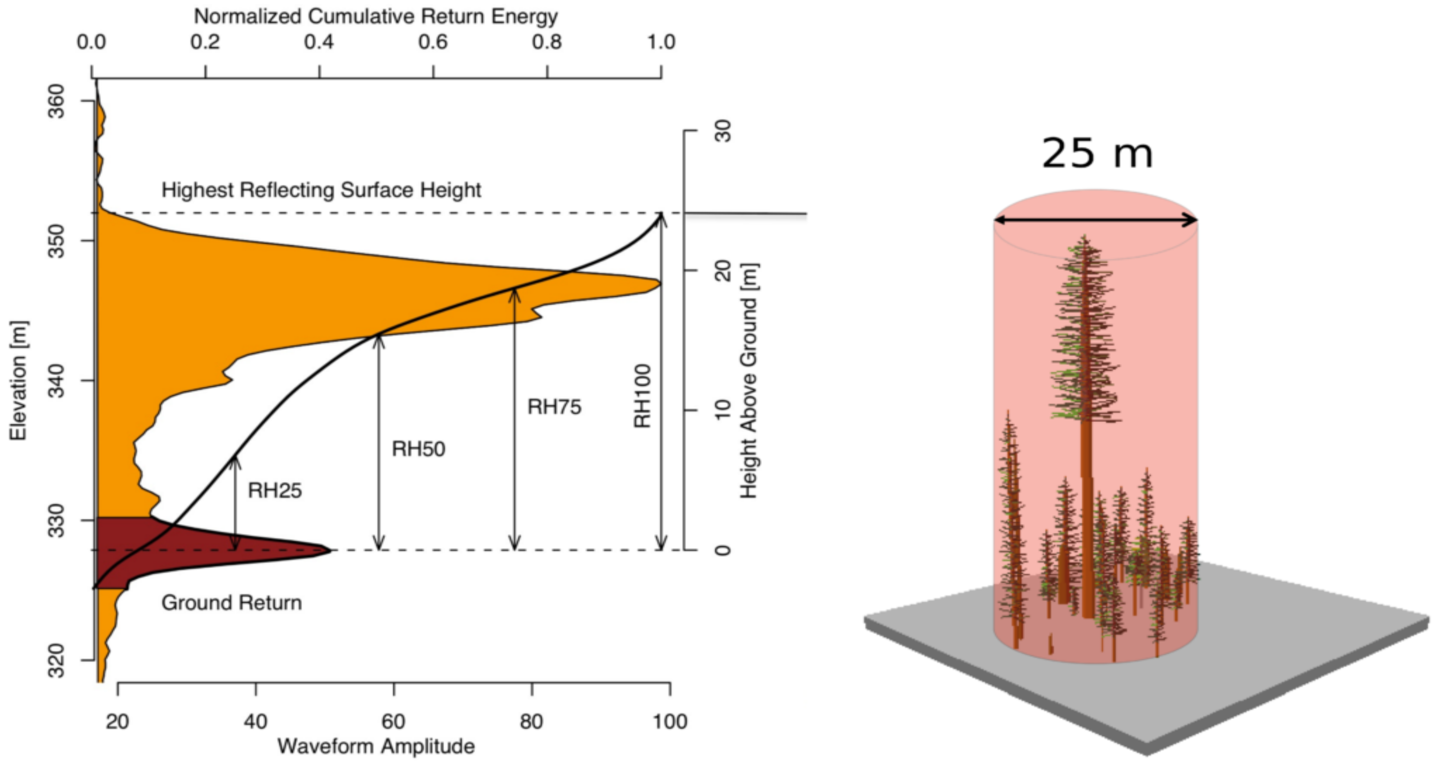

Figure A3-2. The GEDI return waveform and pulse footprint (Dubayah et al. 2020).

A4. FIA Plot Surveying

The Forest Inventory Analysis (FIA) Program collects, analyzes, and reports information on the status, trends, and condition of America's forests ("Forest Inventory and Analysis National Program", 2015). This data includes the quantity and spatial location of forests, and metrics of vegetation cover change, tree harvest, and regrowth. FIA plots are sampled following a national sampling design that includes all forested land in all 50 States, plus all US territories. The FIA program covers all public and private forest lands such as wilderness, National Parks, and National Forests.

A4.1 Survey Design

The FIA program has three main phases that make up the survey design:

Phase 1: Remote Sensing, is the aspect of data collection related to remotely sensed data in the form of aerial photographs, digital orthoquads, and satellite imagery. A Phase 1 "photo point" is characterized as forest or non-forest. A subset of the photo points are selected for field data collection in Phase 2.

Phase 2: Forest Inventory, consists of one field sample site for every 6,000 acres. Field crews collect data on forest type, site attributes, tree species, tree size, and overall tree condition on accessible forest land.

Phase 3: Forest Health, measures for a broader suite of forest health attributes within a subset of Phase 2 sample plots. There is approximately one Phase 3 plot for every sixteen Phase 2 plots. The forest health attributes that are measured include tree crown conditions, lichen community composition, understory vegetation, down woody debris, and soil attributes. Soil samples are sent to a laboratory for chemical analysis.

A4.2 Plot Design

An FIA plot consists of a cluster of four circular subplots spaced out in a fixed pattern within a sampling radius. The plot provides a sampling frame for all Phase 2 and Phase 3 measurements. Most tree measurements are taken within the subplots while seedlings, saplings, and other vegetation are measured within the microplots.

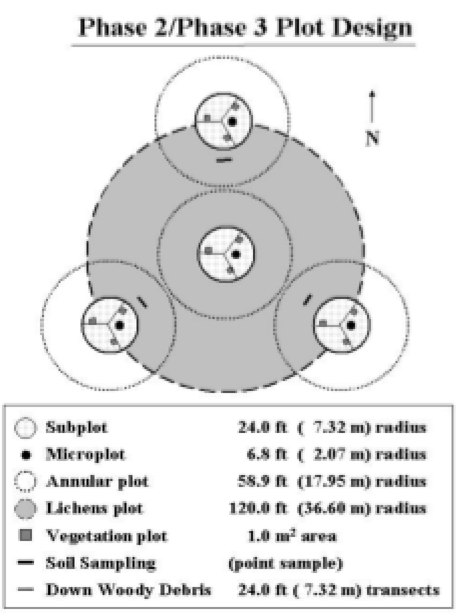

Figure A4.2-1. The FIA plot design layout.

Subplots are never reconfigured or moved after siting. Plots may straddle more than one ‘condition class’ such as two different forest types, or a forest and a meadow. The forest service defines a ‘condition class’ as a specific combination of attributes such as land use, forest type, stand age, and other attributes which collectively describe a homogeneous area. Every plot exists in at least one condition class, and may include more than one. In the case that multiple condition classes are present on a plot, each condition class is described separately. Forested condition classes are further classified by the following characteristics: reserved status, owner group, forest type, stand size class, regeneration status, and tree density.

A4.3 Sampling Intensity

The FIA’s sampling intensity is 20% of the plots in each state, each year, achieved through a federal-state partnership. The FIA program has a national set of core measurements, including some forest health variables collected on a subset of the plots (Burkman, 2005).

A5. FIA Sample Data Summary Statistics

Table A5-1. Summary statistics for the 636 FIA plots (2017 and 2019) used for training and testing the Random Forest models.

Variable	Mean	SD	Min	Max
Biomass (Dry Merchantable lbs/acre)	80310.576	49343.643	138.359	302004.198
Basal Area (square ft/acre)	88.493	41.691	0.821	234.548
Tree Density (trees/acre)	164.463	92.535	6.018	559.678
Tree Height (ft)	60.877	13.349	15	104.267
Potapov (2021) Global Canopy Height (m)	18.57	5.393	0	28.619

Table A5-2. Count of FIA plot surveys from 2017 and 2019, versus digitized cropland, urban, and water plots used in the Biomass, Tree Height, Tree Density, and Basal Area regression models.

Data Source	N
FIA Plots	636
Digitized Cropland	200
Digitized Developed	100
Digitized Water	44
Total	980

Table A5-3. Summary statistics of FIA forest metric variables, and the Potapov (2021) Canopy Height variable within FIA plots used for training and testing in the classification Random Forest model.

Forest Type	Variable	Mean	SD	Min	Max
Longleaf/Slash Pine Group (140)	Potapov Global Canopy Height	19.193	3.679	0	24.642
	Basal Area (square ft/acre)	80.596	36.275	1.813	157.546
	Tree Density (trees/acre)	151.958	91.432	12.036	502.635
	Tree Height (ft)	62.892	12.98	24.286	86.75
	Biomass (Dry Merchantable lbs/acre)	79404.631	42433.256	326.599	188968.436

Loblolly/Shortleaf Pine Group (160)	Potapov Global Canopy Height	18.103	4.506	0	28.619
	Basal Area (square ft/acre)	92.618	43.739	0.821	254.062
	Tree Density (trees/acre)	194.21	110.042	6.018	613.841
	Tree Height (ft)	59.342	14.613	13	120.333
	Biomass (Dry Merchantable lbs/acre)	76832.036	47865.701	38.787	274006.206
Oak/Pine Group (400)	Potapov Global Canopy Height	19.347	4.941	0	27
	Basal Area (square ft/acre)	81.404	39.686	1.702	199.861
	Tree Density (trees/acre)	140.308	60.296	6.018	324.974
	Tree Height (ft)	59.56	12.051	24	96.308
	Biomass (Dry Merchantable lbs/acre)	76596.455	52575.258	410.738	296254.248
Oak/Hickory Group (500)	Potapov Global Canopy Height	20.482	5.326	0	29.304
	Basal Area (square ft/acre)	79.729	40.481	1.344	219.788
	Tree Density (trees/acre)	132.868	63.257	6.018	421.263
	Tree Height (ft)	61.009	11.769	15	94
	Biomass (Dry Merchantable lbs/acre)	78594.98	51496.137	720.916	267274.395
Oak/Gum/Cypress Group (600)	Potapov Global Canopy Height	19.719	5.143	0	27.072
	Basal Area (square ft/acre)	86.697	45.402	0.821	235.534
	Tree Density (trees/acre)	135.667	69.707	6.018	385.155
	Tree Height (ft)	59.642	13.811	15	104.267
	Biomass (Dry Merchantable lbs/acre)	81688.048	56359.21	138.359	302004.198
Elm/Ash/Cottonwood Group (700)	Potapov Global Canopy Height	18.283	6.413	0	26.336
	Basal Area (square ft/acre)	70.169	40.303	1.88	205.116
	Tree Density (trees/acre)	118.006	65.245	6.018	312.938
	Tree Height (ft)	57.356	10.843	21	87.35
	Biomass (Dry Merchantable lbs/acre)	58373.997	40678.232	161.241	199300.457

Table A5-4. Count of FIA plots versus digitized cropland, urban, and water plots used in the Forest Type classification model. FIA survey data came from the years 2014 - 2017.

Data Source	N
FIA Plots	1994
Digitized Cropland	200
Digitized Developed	100
Digitized Water	40
Total	2338

A6. Spectral and Textural Indices

Table A6-1. GLCM textural indices computed for Sentinel-1 mosaics.

Indices	Description
Angular Second Moment (ASM)	ASM is known as uniformity or energy. It measures the uniformity of an image. When pixels are very similar, the ASM value will be large.
Contrast	Contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbor
Dissimilarity	Dissimilarity is a measure of distance between pairs of objects (pixels) in the region of interest.
Entropy	Amount of irremediable chaos or disorder
Correlation	Correlation shows the linear dependency of gray level values in the co-occurrence matrix. It presents how a reference pixel is related to its neighbor, 0 is uncorrelated, 1 is perfectly correlated.
Variance	Variance is a measure of the dispersion of the values around the mean of combinations of reference and neighbor pixels.
Difference	Measures the dispersion (with regard to the mean) of the gray level difference distribution of the image.

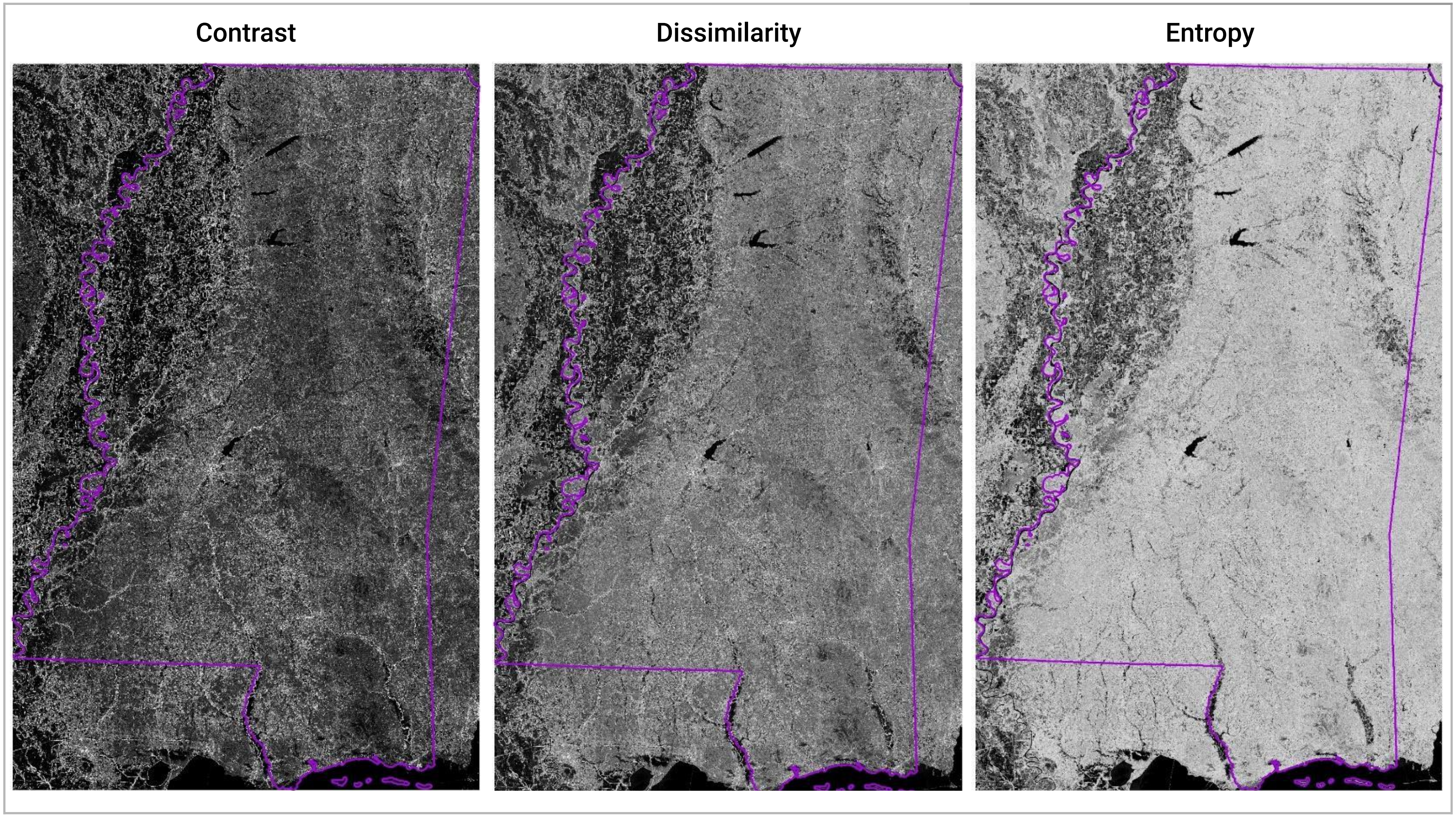

Figure A6-1. Images show the Contrast, Dissimilarity, Energy, and Entropy VH indices for the Leaf Off time period over the state of Mississippi

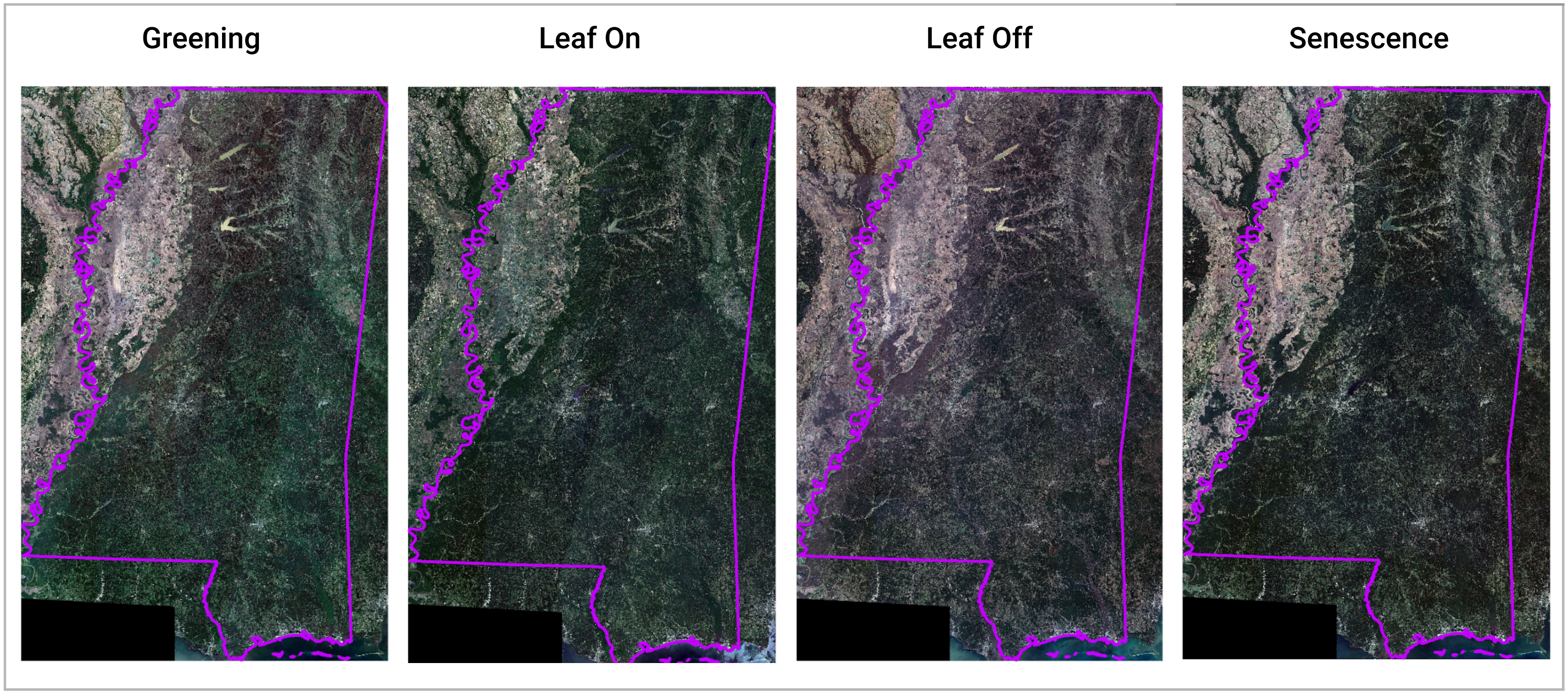

Figure A6-2. RGB images show phenological changes over Mississippi, from left to right: Greening, Leaf On, Leaf Off, and Senescence.

Table A6-2. Multispectral indices were computed for each Sentinel-2 mosaic.

Indices	Description
Normalized Difference Vegetation Index (NDVI)	Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs).
Red Edge Normalized Difference Vegetation Index (RENDVI) Variations #1, #2, #3	This index is a modification of the traditional broadband NDVI. Applications include precision agriculture, forest monitoring, and vegetation stress detection. RENDVI capitalizes on the sensitivity of the vegetation red edge to small changes in canopy foliage content, gap fraction, and senescence.
Normalized Difference Red Edge (NDRE) Index Variations #1, #2, #3	NDRE is a spectral index that is built as a blend of several bands: Nea-InfraRed (NIR) spectrum and a band that uses a narrow spectral range between visible Red and NIR. NDRE is more sensitive than NDVI for a certain period of plant maturation.
Enhanced Vegetation Index (EVI)	EVI can be used to quantify vegetation greenness. EVI corrects for some atmospheric conditions and canopy background noise and is sensitive in areas with dense vegetation.
Anthocyanin Reflectance Index (ARI)	Increases in ARI indicate canopy changes like new foliage growth or folidage death. Anthocyanins are water-soluble pigments abundant in newly forming leaves and those undergoing senescence.
Structure Insensitive Pigment Index (SIPI)	Increases in SIPI are thought to indicate increased canopy stress. Applications of this index include vegetation health monitoring, plant physiological stress detection, and crop production.
Modified Chlorophyll Absorption in Reflectance Index (MCARI)	MCARI gives a measure of the depth of chlorophyll absorption and is very sensitive to variations in chlorophyll concentrations as well as variations in Leaf Area Index (LAI). MCARI values are not affected by illumination conditions, the background reflectance from soil, and other non-photosynthetic materials observed.
Tasseled Cap Wetness (TCWET)	TCWET is interpreted as an index of “wetness” (e.g., soil or surface moisture) or “yellowness” (e.g., amount of dead/dried vegetation).
Tasseled Cap Vegetation (TCV)	TCV corresponds to “greenness” and is typically used as an index of photosynthetically-active vegetation.
Tasseled Cap MSS Green Vegetation (TCMSSV)	TCMSSV describes variations in the rigor of green vegetation.
Tasseled Cap Brightness (TCB)	TCB corresponds to the overall brightness of the image.
Tasseled Cap MSS Soil Brightness (TCMSSBRI)	TCMSSBRI describes variations in soil background reflectance.

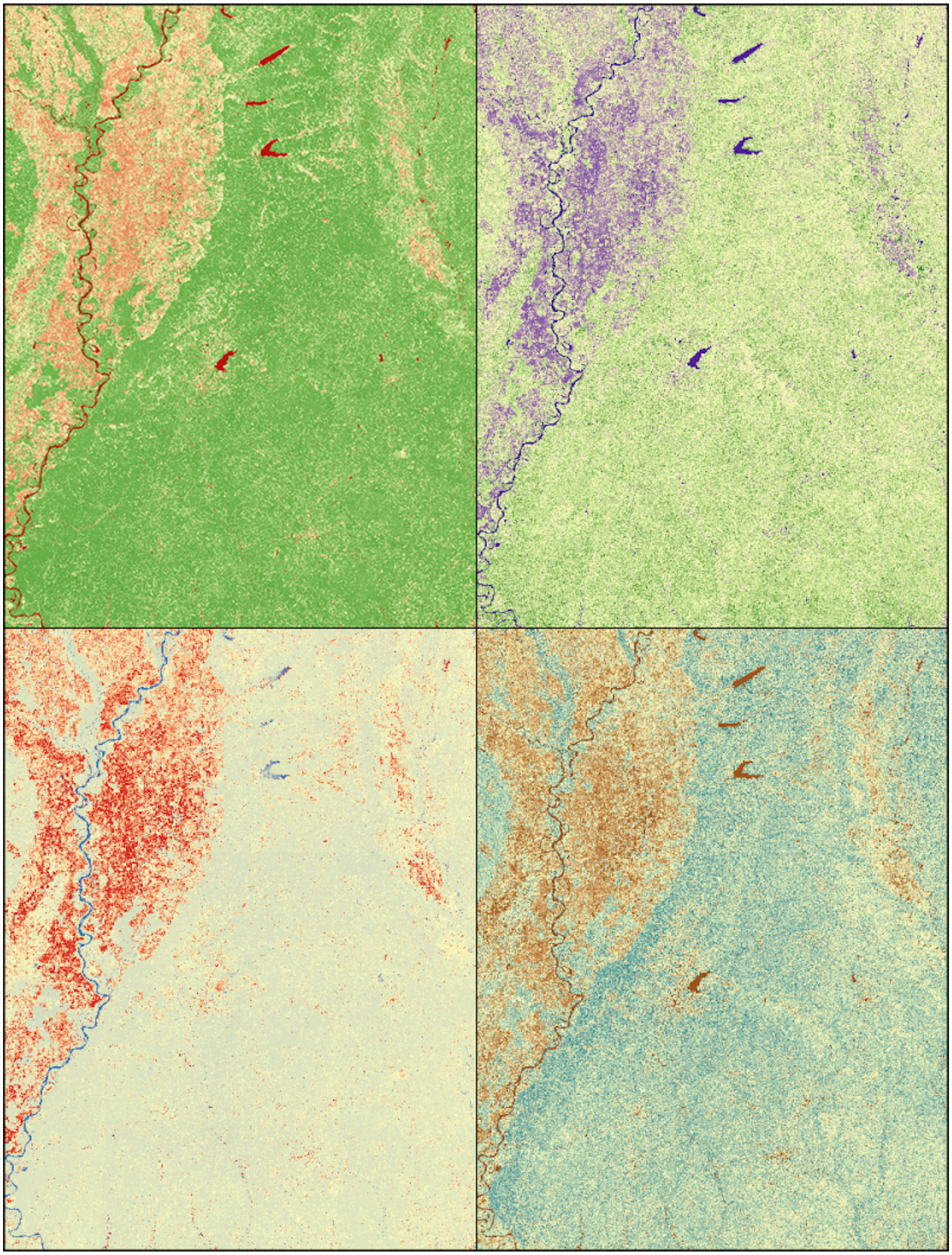

Figure A6-3. Examples of Sentinel-2 multispectral vegetation indices (Senescence, listed clockwise from top left: Normalized Difference Vegetation Index, Normalized Difference Red Edge Index Variation #2, Structure Insensitive Pigment Index, Tasseled Cap Vegetation)

A7. Model Accuracy Results

Table A7-1. Comparison between the original forest type model results and the new forest type model results with the Potapov (2021) Global Canopy height data.

Metric	Original Kappa Value	Original Accuracy Rate	Phase II Kappa Value	Phase II Accuracy Rate	Accuracy Improvement
Forest Type Group - Overall	0.68	0.74	0.68	0.75	0.86
Class Metrics	**Original Detection Prevalence (Area Coverage)**	**Original Accuracy Rate**	**Detection Prevalence (Area Coverage)**	**Phase II Accuracy Rate**	**Accuracy Improvement**
Developed (10)	0.04	0.98	0.04	0.99	0.15
Water (20)	0.02	1.00	0.02	1.00	0.00
Cropland/Rangeland (30)	0.10	0.98	0.08	0.99	1.59
Longleaf/Slash Pine Group (140)	0.05	0.79	0.05	0.82	2.25
Loblolly/Shortleaf Pine Group (160)	0.37	0.90	0.37	0.89	-0.96
Oak/Pine Group (400)	0.06	0.60	0.09	0.67	7.29
Oak/Hickory Group (500)	0.22	0.84	0.20	0.82	-2.09
Oak/Gum/Cypress Group (600)	0.12	0.81	0.12	0.77	-3.73
Elm/Ash/Cottonwood Group (700)	0.02	0.64	0.03	0.70	6.64

A8. Importance of Model Variables

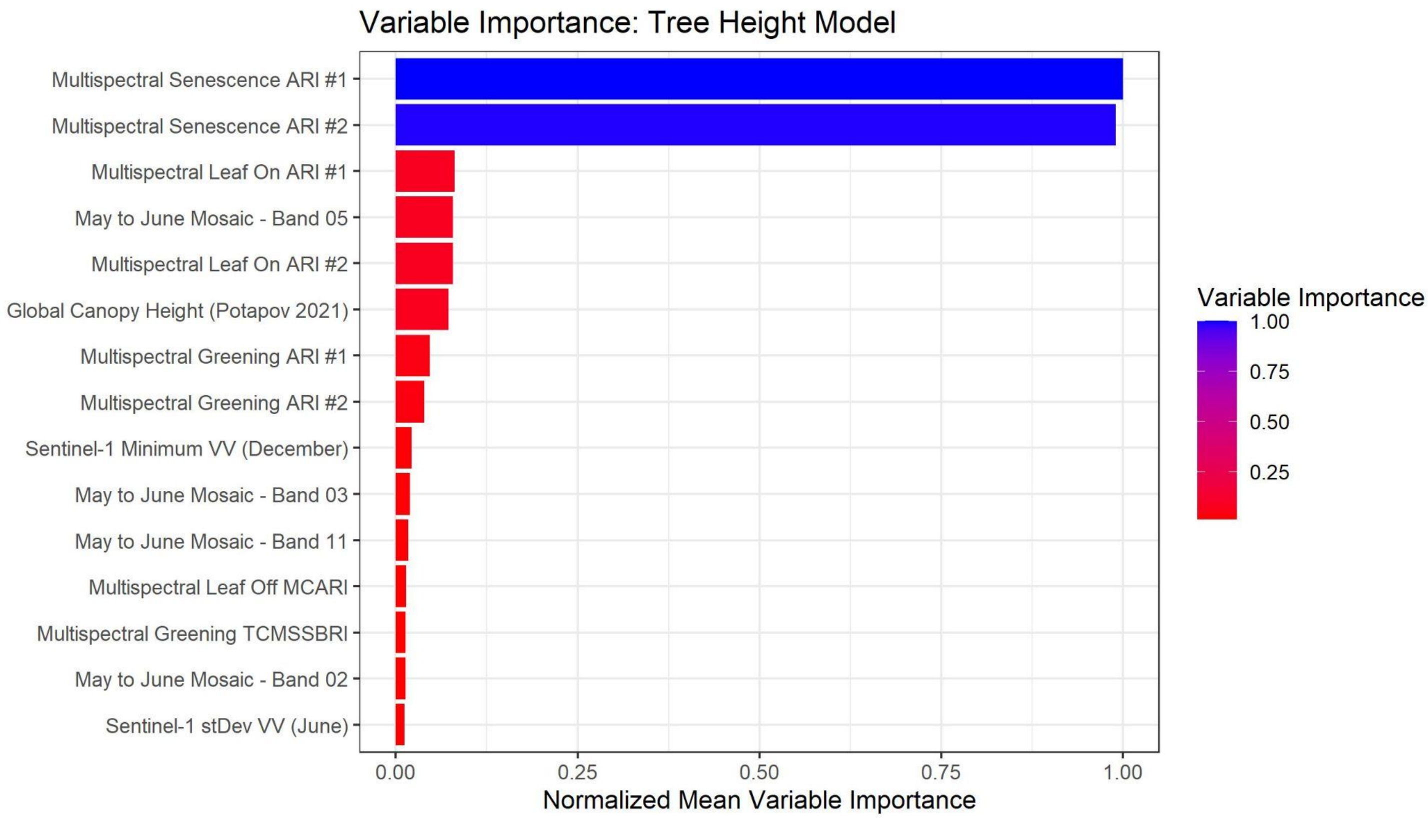

Figure A8-1. Chart of normalized mean variable importance for the fifteen most important variables from the five iterations of the Tree Height model.

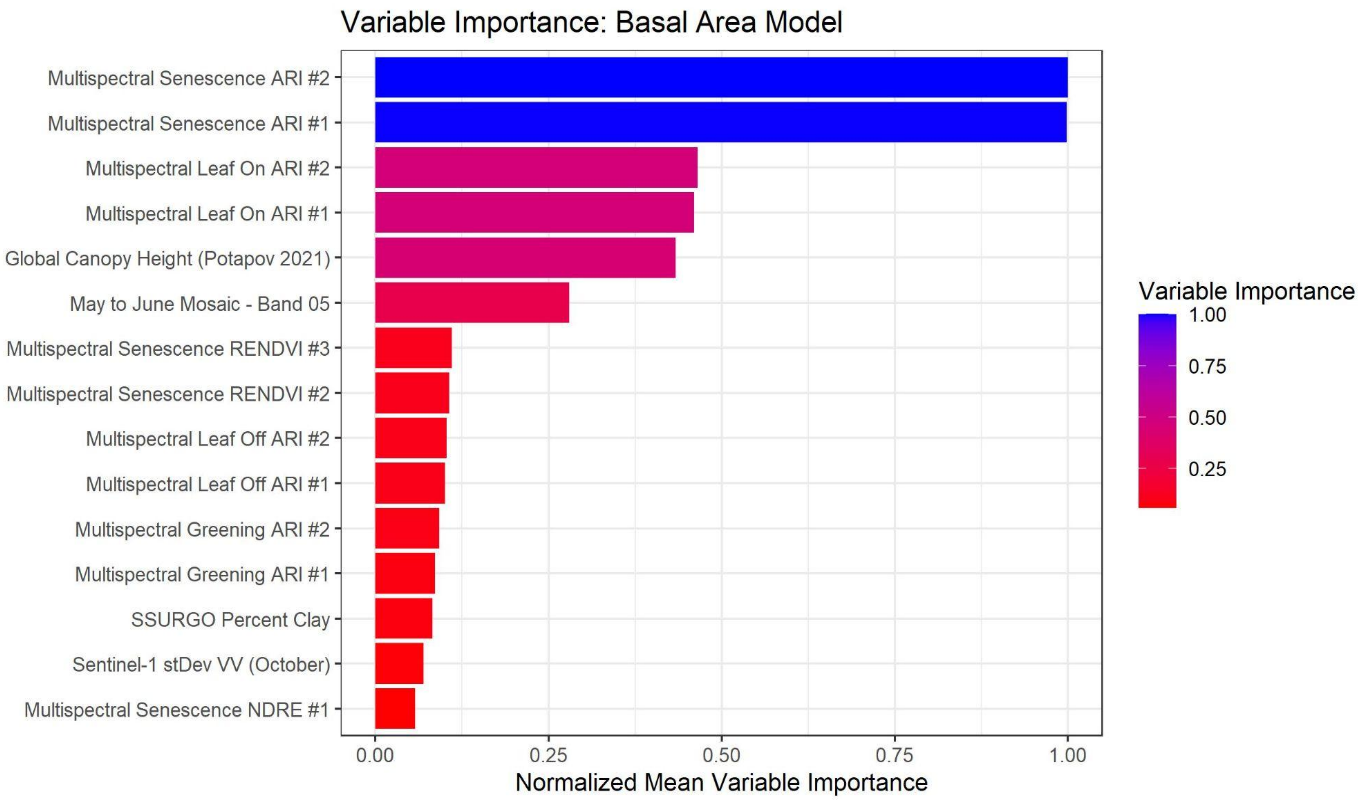

Figure A8-2. Chart of normalized mean variable importance for the fifteen most important variables from the five iterations of the Basal Area model.

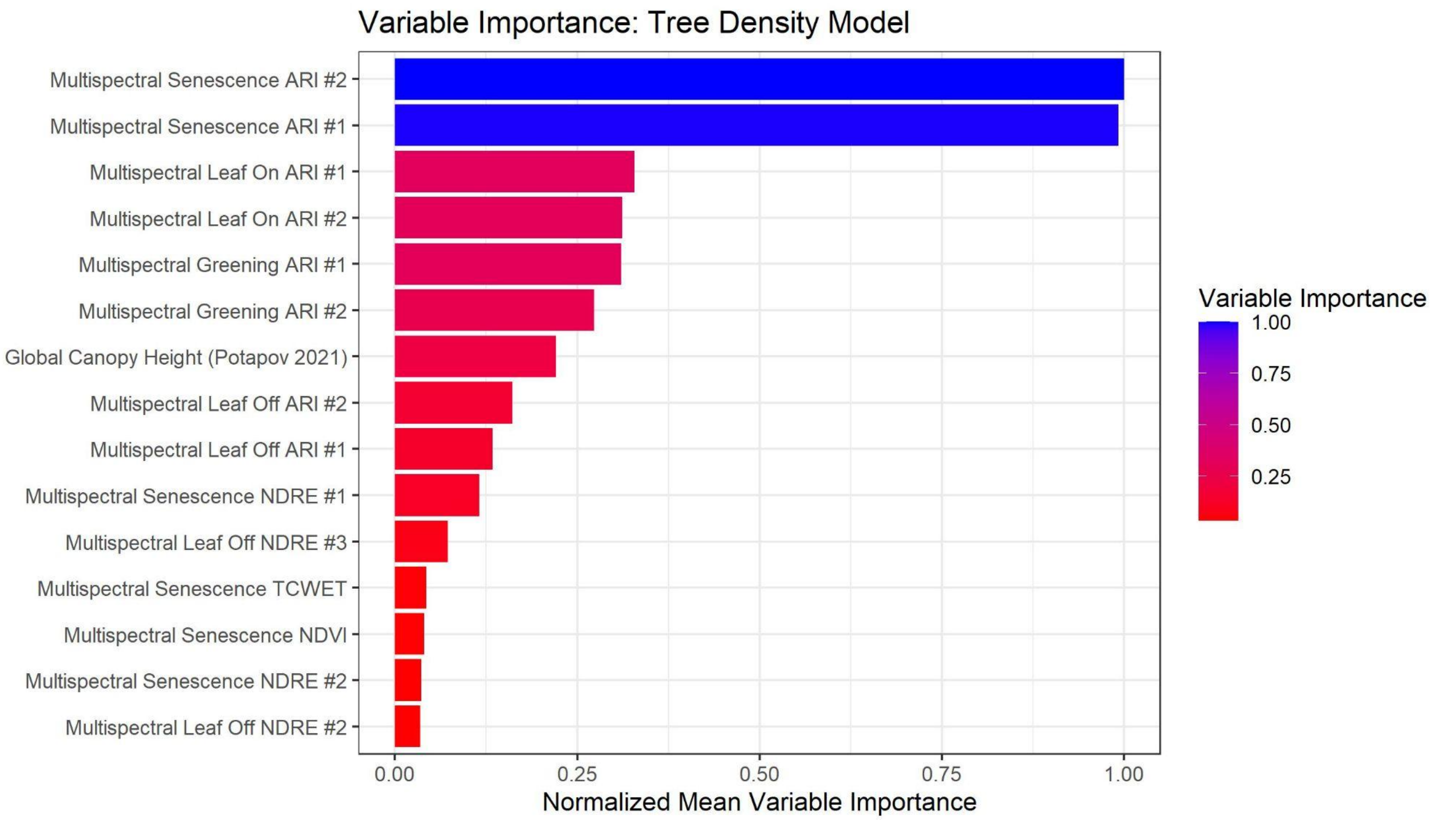

Figure A8-3. Chart of normalized mean variable importance for the fifteen most important variables from the five iterations of the Tree Density model.

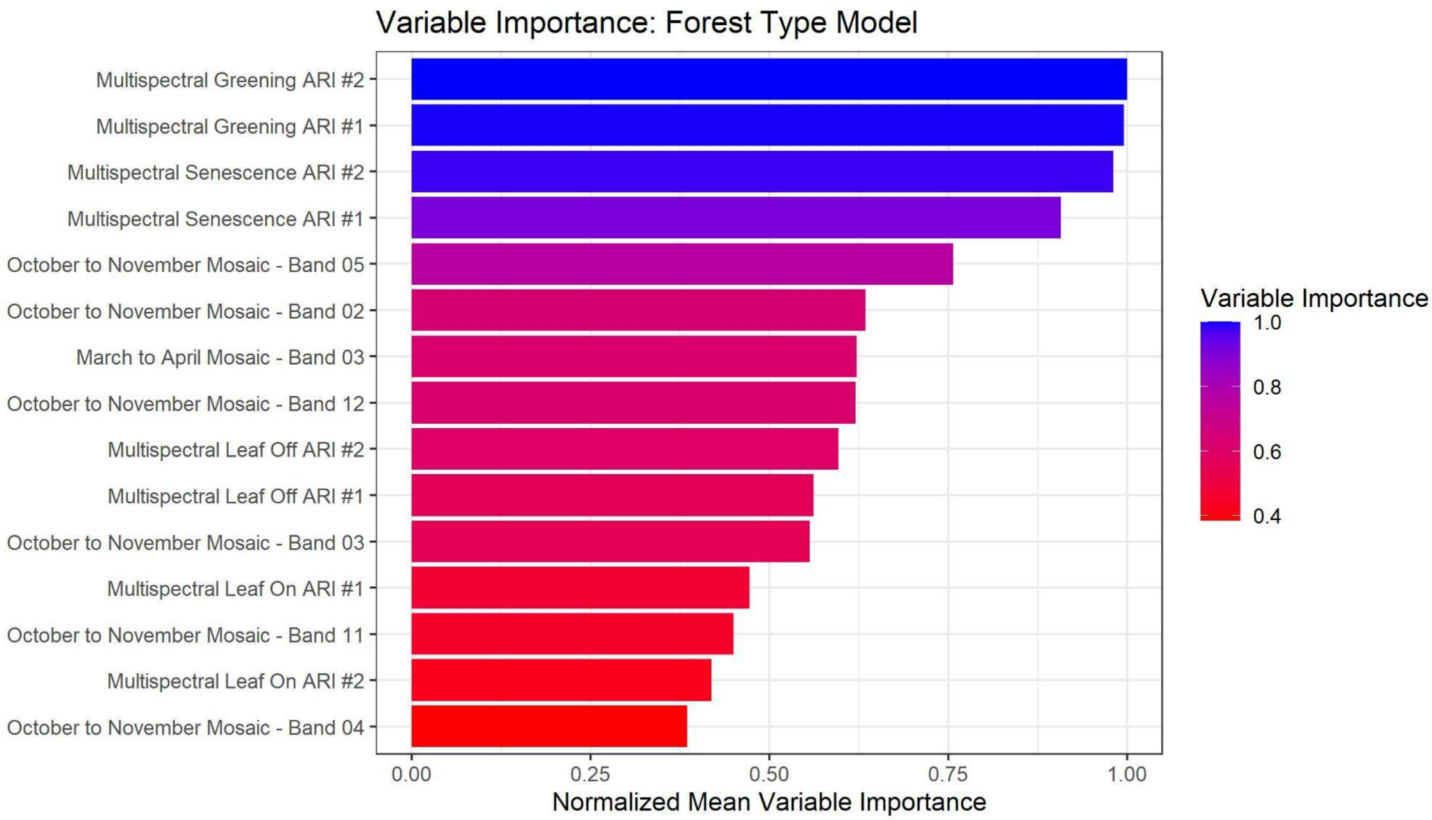

Figure A8-4. Chart of normalized mean variable importance for the fifteen most important variables from the five iterations of the Forest Type model.

A9. GEDI Fusion Data Product Details

While the immediate GEDI data products will not provide wall-to-wall data coverage for canopy height or biomass, independent research groups are actively working on GEDI fusion products using Landsat and Sentinel. These products will use robust modeling approaches to estimate the gaps in data coverage, and even have the possibility to apply the same models in reverse, backward in time, to estimate biomass in years past. Currently, research teams are working to create demonstrative data products that utilize GEDI-derived metrics for domain-specific applications.

A9.1 GEDI X TanDEM Fusion

New GEDI fusion products that incorporate data from Synthetic Aperture Radar are expected to be released in the next few years. The TanDEM-X and TanDEM-L SAR missions fused with GEDI have the potential for high-resolution wall-to-wall maps of biomass and canopy height. A formal collaboration between GEDI and the German Space Agency (DLR) is creating advanced algorithms to combine the wall-to-wall mapping capability of TanDEM-X (12 meter resolution) with the spatial sampling of GEDI LiDAR. This fusion is expected to improve height and biomass estimates at a finer resolution and accuracy than either mission can achieve alone (University of Maryland, GLAD, 2021). The TanDEM-X mission has two spacecraft that image the earth simultaneously, with an interferometric synthetic aperture radar (InSAR) sensor (Qi et al., 2019). The TanDEM-L Inferomatic Radar mission, planned to launch in 2024, will provide estimates of canopy height and biomass at a 12 meter resolution. The Synthetic Aperture Radar techniques used by TanDEM-L are not constrained by weather and daylight, and the L band's wavelength (23.6 cm) is optimal for imaging the 3-D structure of vegetation (Herbert J., n.d.). Unlike TanDEM-X, the L-band of TanDEM-L will be able to penetrate to the ground even in extremely dense forest environments (Dubayah, 2017). Similar to a GEDI X TanDEM-X fusion, a GEDI, and TanDEM-L fusion product will improve the accuracy of biomass and canopy height estimation even in dense forests.

A9.2 The Ecosystem Demography Model

The Ecosystem Demography model will provide fine resolution estimates of carbon stocks and fluxes over expansive areas. The Ecosystem Demography model will incorporate GEDI vegetation structure data to produce high spatial resolution estimates of the carbon sequestration potential of tropical and temperate forests under multiple Intergovernmental Panel on Climate Change (IPCC) climate and land-use-change (University of Maryland, GLAD, 2021). Through these simulations, researchers can assess the impacts of policies on CO2 emissions and land use (Dubayah et al., 2020).