Improving Forest Inventory Metrics on Pennsylvania State Game Lands #221

GREEN LEAF CONSULTING SERVICES

Josh Flad, Principal
SAF & ACF Certified Forester, ISA Certified Arborist
[Green Leaf Consulting Services LLC]
Beach Lake, Pennsylvania, USA
570.460.8034

Green Leaf Consulting Services is dedicated to helping landowners achieve their resource management goals, finding innovative solutions and providing guidance for long run sustainability. Josh Flad, SAF CF, ISA CA., founder of GLCS, firmly believes in the "wise use" principles of conservation.

tesera hris BY TESERA

Michael Kieser, Principal
Chief Data Engineer, PEng, BSc. Geomatics Engineering
[Tesera Systems Inc.]
Calgary, Alberta, Canada
403.973.3277

Tesera Systems Inc. is committed to delivering more reliable, scalable and consistent High Resolution Inventory Solutions (HRIS) data to enable sustainable forest management, planning and operations. Michael's passion is turning otherwise disparate geospatial datasets into valuable information for use in advanced analytical processes, like HRIS.

TABLE OF CONTENTS

ABSTRACT

New and innovative forest inventory approaches and solutions continue to evolve. They provide more reliable, scalable and machine learning-based inventories. These new solutions significantly improve the statistical reliability and types of forest inventory attributes. Forest owners and managers are now able to better characterize, plan and manage forest landscapes with greater certainty.

A. INTRODUCTION

In February 2022, Green Leaf Consulting Services (GLCS) retained Tesera Systems Inc. (Tesera) and Forest Metrix to develop a new high resolution forest inventory geopackage for the Dynamic Forest Restoration Block – Pennsylvania State Game Lands No. 221 (SGL No. 221).

The primary objective was to demonstrate how high resolution forest inventory data can enable forestland managers with more comprehensive and reliable information about the current state of the forest. Ultimately, the goal of the new inventory information is to help:

- Design and optimize restorative silvicultural treatments to improve the forest structure and ecological characteristics required by wildlife species.
- Establish greater forest species and structural diversity, and wildlife habitat quality.
- Support healthy populations and distributions of resident and migratory wildlife species.
- Incorporate climate adaptation practices and make tree species less susceptible to current and future climate risks (fire, flood, snow, wind, resistance to frost/hail, insects, disease).
- Improve the forest matrix, so that sport-hunting and outdoor recreation opportunities and experiences are improved and maintained over time.

This article highlights the process and challenges in generating a high resolution forest inventory geopackage for PSGL No. 221.

B. AREA OF INTEREST – PENNSYLVANIA SGL #221

The forest inventory Area Of Interest (AOI) for SGL No. 221 consisted of 4,675 acres, located in Monroe County, north of Mount Pocono, Pennsylvania (see Appendices, Figure 1.).

Pennsylvania's forested landscapes today differ significantly from the structural diversity, functional diversity, forest type and seral stage distributions that characterized the forest landscape in the late 1700's. Societal development, urban land use conversion, timber harvesting and other human activities have contributed to a fragmented forest matrix, and reduced habitat suitability and availability for wildlife species. Deer, bear, turkey, grouse, snowshoe hare, squirrel and rabbits are prevalent game species. Pennsylvania State Game Lands (PSGL) are purchased, managed and maintained primarily for outdoor recreation in the form of sport-hunting. SGL No. 221 is also managed for a range of ecological and timber values.

The current inventory indicated that forest composition within SGL No. 221 is dominated by hardwoods (white oak, northern red oak, red maple, scarlet oak, sweet birch, pignut hickory, white pine, black oak, shagbark hickory, white ash, and sugar maple) and to a lesser extent conifers. The relatively low levels of native conifer species represented in the inventory are believed to be attributable to past harvesting and inventory bias.

Past forest inventory variable radius (VR) ground sampling of SGL No. 221 concentrated largely on upland hardwood acreage, with a focus on implementation of restorative forest stand management prescriptions.

The new forest inventory used VR ground samples that had been recently established on the Buck Hill Falls property, located immediately adjacent to SGL No. 221. GLCS believed these adjacent VR ground plots were reasonably representative of the range of variation and forest structural types that occur on SGL No. 221.

Tesera worked collaboratively with GLCS to ensure that use of ground sample data and remote sensed imagery data would meet quality control and quality assurance specifications. Both parties acknowledged the use of recently established VR ground samples from the adjacent Buck Hill Falls property and existing imagery would introduce some uncertainty regarding the error of estimate.

C. HIGH RESOLUTION INVENTORY SOLUTION

A high resolution inventory solution (HRIS) is an area-based inventory that relies on:

- Compiled ground measurement attributes, distributed over precisely defined fixed radius (FR) ground plots that are accurately geolocated.
- Consistently captured, remote sensed, high resolution land base imagery data (e.g. LiDAR and multiband optical imagery) that precisely corresponds geospatially with fixed radius ground plot measurement data.

The move towards higher resolution forest inventory solutions is being driven by:

- Challenges recruiting and retaining skilled labor;
- Improved access to remote sensed imagery;
- Advancements in artificial intelligence (AI) and machine learning (ML) data analytics;
- Climate-related impacts;

- Transparency and accountability ESG reporting and global market forces (forest certification and supply chain requirements).

HRIS models the relationships between measured ground plot unit attributes, as well as characteristics (features) calculated from remote sensed land base data that relate to and are within those FR ground plots. Depending on the input data, the accuracies of the generated models are in the 90-95% confidence level. An outline of the HRIS process is provided in the Appendices, Figure 2.

D. HRIS INVENTORY PROCESS FOR SGL #221

1. **PROJECT AREA AND ATTRIBUTES**: GLCS provided Tesera with a project boundary for SGL No. 221 (4,675 acres). The working coordinate system assigned for the HRIS inventory was UTM83 zone 18 projection (EPSG:26918). A 2km by 2km tiling scheme (six columns by seven rows) was developed over the entire project boundary. The project boundaries were contained within 18 tiles. GLCS provided a list of target attributes (compiled by Forest Metrix) using existing VR ground plot data.

2. **GROUND PLOT DATA:** GLCS also provided VR and FR ground plot sample data:

   - 449 VR ground plots over the Buck Hill Falls adjacent lands, containing ~4,000 treelist records (collected in 2019):
     - Plot centers assumed to be (+/-) 5 meters.
     - The radii for the variable radius plots were calculated by GLCS for each plot. These were based on the maximum distance for the largest tree in each plot. For example, a BAF 10 wedge prism has a 2.75 plot radius factor, therefore a

9 inch tree at a distance of 24.75 foot radius is considered "in". The assumption that the largest tree is also the furthest away (and defines the plot radius) results in some bias in the "per acre" attributes. A partial solution for VR plots is to measure the radius from the plot center to the farthest tree measured in the plot (vs. calculating the radius based on the BAF and largest tree diameter).

  - The radii were applied to each of the plot centers to derive each of the circular polygon plot units.

- 41 FR ground plots (1/10th acre with 37.2 foot radius) over the DFRB SGL 221 AOI, containing 1,018 treelist records (collected in 2022):
  - Plot centers were geolocated within sub-meter accuracy.
  - A 37.2 foot radii was applied to each of the plot centers to derive each of the circular polygon plot units.

The plot treelist data was compiled and aggregated to the plot level by Forest Metrix. The target attributes included: total basal area per acre (BAPA), percent of each species group based on basal area (SPC_PCT), quadratic mean diameter (QMD), mean merchantable height (MHT), trees per acres (TPA), and three types of merchantable volume per acre (MVOLPA).

Tesera determined which species could stand alone, and which should be grouped, as represented in the following seven classes: 1. Red Maple; 2. Black Cherry, Pin Cherry, Sugar Maple, Striped Maple, White Ash, Yellow Birch; 3. Black Oak; 4. Aspen, Black Birch, Black Gum, Black Locust, Basswood, Hophornbeam, Sassafras, Serviceberry, Yellow Poplar; 5. Red Oak, American Chestnut, Hickory, Scarlet Oak, White Oak, Chestnut Oak; 6. Beech; 7. Black Spruce, Cedar, Hemlock, Norway Spruce, Pitch Pine, Red Spruce, White Pine.

3. **LANDBASE DATA:** Existing LiDAR was gathered using the LiDAR Explorer to derive the indices. Publicly available 2019 LiDAR (209 files, covering the project tile extent) was downloaded from the United States Geological Survey. Existing imagery was gathered using the Pennsylvania Imagery Navigator to derive the indices. Publicly available 2019 imagery (12 files, covering the project tile extent) was downloaded from the Pennsylvania Spatial Data Access.

4. **IMAGERY TILING & PRODUCTS:** The acquired imagery was tiled according to the project tiling scheme, and then assessed through a quality control process. A vegetation index is developed for each tile, in order to distinguish stand canopy cover types.

5. **LIDAR TILING & PRODUCTS:** The acquired LiDAR was tiled according to the project tiling scheme, and then assessed through a quality control process. It was then used to develop a Digital Terrain Model (DTM), height normalized LiDAR, and a canopy surface to distinguish stand canopy heights.

6. **SEGMENTATION:** The vegetation index and canopy surface were merged together, as tiled stand delineation rasters. The rasters were segmented into raw **reporting units**, with unique identifiers, that extend beyond the tile boundaries. They were further sub-segmented into hexagons (with unique identifiers) of a similar size to the plot units. The hexagons were then cleaned along tile boundaries, leaving a seamless dataset of tiled **analysis units**. Finally, the plot units and analysis units were merged within each tile,

creating a tiled plot-analysis unit dataset (with unique ids and the parent reporting unit id to which they belong).

7. **VIRTUAL PLOTS:** A **virtual plot** dataset was generated from the plot units and a random sample of 150 analysis units (from each plot-analysis unit tile) for a total of approximately 2,700 virtual plots. These virtual plots were used for land cover modeling and leading species modeling.

8. **FEATURES:** Characteristics (machine learning **features**) were calculated for each plot and analysis unit from the imagery, LiDAR and DTM. There were approximately 2,000 features calculated for every plot and analysis unit. Statistics were calculated across all the feature values for quality control, and to normalize the data.

9. **NORMALIZATION, FEATURE SELECTION & MODELING:** Statistically-based feature selection methods were used to select those features having the strongest relationship with the target attribute. Recursive feature elimination was run over the features (for each target attribute being modeled) to select those features that best explain the target attribute.

   The reference data was split for ML training (70%) and validation (30%). During the ML training process, k-fold cross-validation techniques were used to iteratively adjust the model hyperparameters, with the goal of achieving a model accuracy greater than 95%.

10. **LAND COVER (LC) MODELING:** The virtual plots (see 7. above) with their features were run through a clustering k-means algorithm. The resulting clusters were then converted into land cover labels (e.g. treed, nontreed, water). These labels and features

provide the reference dataset for land cover modeling. The land cover model accuracy was over 95%.

11. **LEADING SPECIES (LSPC) MODELING:** Leading species modeling is a hierarchical child to LC modeling, whereby only virtual plots having a "treed" LC label are analyzed. A cluster analysis (using a k-means algorithm) was run across the virtual-plot feature data, to separate the virtual-plots by canopy leading species. The number of discernable clusters provided the basis for the number of uniquely identified species or species groups. Based on the leading species labels in the plot units, the resulting clusters were then converted into leading species, or leading species group labels, subsequently used for leading species modeling. Feature selection (as detailed above) was run to select those features that best explain the labels. A leading species label model was then developed, based on the leading species labels and selected features in the virtual-plot data. The model accuracy ranged between 80% to 90%.

12. **TARGET ATTRIBUTE MODELING:** Compiled plot unit data and features provided the reference dataset for the target attribute modeling. As target attributes are typically continuous variables (ie. diameters, heights, volumes), feature normalization is critical to ensuring an equal weight to each feature. Feature selection was run to select the features that best explain each target attribute. Feature selection reduced the feature count from ~2,000 down to 250 features. Models were developed for each target attribute based on the plot target attribute data and their selected features. The target attributes were modeled and their associated accuracies are listed below:

- Basal Area Per Acre (BAPA): 93%
- Mean Height (MHT): 75%
- Trees Per Acres (TPA): 84%
- Species Percent (SPC_PCT): 80% to 90%
- Quadratic Mean Diameter (QMD): 90%
- Merch Volume Per Acre (MVOLPA): 87%

13. **PREDICTION:** Following development of the models – land cover, leading species and target attributes were predicted for each analysis unit, using the derived models.

14. **ACCURACY ASSIGNMENT:** The accuracy was derived for every plot unit target attribute. The nearest plot unit (in feature-space and distance) was determined for every analysis unit. The weighted target attribute accuracies were attached to the analysis units.

15. **REPORTING UNIT INVENTORY:** The results of the target attribute predictions and their accuracies (at the analysis unit) were aggregated to the reporting unit, based on their unique identifier. Following aggregation, the predicted unit attribute values were analyzed with the compiled plot data to provide: within range of variation, value data distribution, and general statistical outliers. The inventory metadata and target attribute description fields were added. It was determined that 195 reporting units (2%) contained predicted attribute values that were outliers, and could not be relied upon.

E. PROJECT CHALLENGES

The following project challenges were acknowledged by GLCS and Tesera in producing the HRIS inventory.

- Existing and publicly available data were used to reduce costs and save time.

- Land base LiDAR data was lower density, and higher scan angle, rather than as recommended in the HRIS specifications.
- Land base imagery data was only 3 bands, rather than 4 bands as recommended in the HRIS specifications.
- Acquisition dates of ground plot data and imagery capture did not coincide.
- Use of adjacent VR ground plots and FR ground plots contributed to greater uncertainty.
- A modest budget and project scope needed to be managed accordingly.

F. USE OF VARIABLE AND FIXED RADIUS GROUND PLOTS

As mentioned above, a noteworthy project challenge was how to use both variable radius (VR) and fixed radius (FR) ground plots. Much has been written about the relative differences between VR and FR ground plot sampling approaches. Historically, VR ground plots have been the predominant approach used across the USA to sample timber-related values (tree species, basal area, height and quality metrics) and to predict volume and market value. However, forest inventory sampling is now shifting towards the use of FR ground plots to predict, report and monitor a much wider range of forest attributes, beyond timber values. For example, many foresters are seeking better ways to support the design and implementation of restorative silvicultural treatments, habitat management, biodiversity, predict carbon sequestration, water management and fire management – in addition to timber values.

Tesera's HRIS standard for ground sampling is based on a 10 meter FR ground plot, in which all trees are sampled. The number of plots required to properly sample a project area is proportional to the range of variation in the forest including the number of species, the range in diameters, and the range in heights to model/predict. Generally, the number of FR ground plots required for an

HRIS inventory is less than the number of variable radius plots that are typically sampled for forest inventories. HRIS inventory FR ground plots measure all trees within a fixed area and locate (with sub-meter accuracy) the plot center. The accurate geolocation is used to align and fuse remotely sensed data (LiDAR and multispectral imagery) with the corresponding surface geolocation of the FR ground plot data. The compiled ground measurements are then matched to the land base data captured over the same area. Forty one **(41) FR ground plots were established.**

The compiled attributes for the original VR plots could not be attached to a precise area, causing significant challenges using remotely sensed data in the modeling. Ideally, compiled ground plot attributes and remote sensed data should coincide exactly with sub-meter plot centers and associated FR plot areas. The project team is unaware of a reliable method for accurately relating a VR plot into a defined fixed area.

G. KEY FINDINGS AND RESULTS

Initially, the project team used only existing VR plots from an adjacent property. Subsequently, FR area plots were established on the AOI to refine the modeling and analysis. The team compared results using VR plots versus the use of FR plots. The findings indicated that an HRIS inventory using FR plots provides superior inventory results (e.g., statistically reliable attributes, value and utility in forest management decision making), when compared to VR plots. This is a key finding for forest resource managers when considering the design of future forest inventories.

Examples of attributes with average accuracies, for an **Overall Accuracy of: 95%**

- Diameter: 96%
- Height: 94%
- Basal Area Per Acre: 97%
- Trees Per Acre: 95%
- Merchantable Volume Per Acre: 95%
- Above Ground Biomass Per Acre: 94%

The accuracy for the reporting units were derived as follows: R-squared (R2) values were calculated for every plot attribute. Utilizing a nearest neighbor algorithm, every reporting unit was compared to every plot unit. The R2 values (0-1) for the nearest plot were applied to each reporting unit. Then, the Pearson Correlation Coefficient (PCC) was calculated (0-1) for each reporting unit relative to its nearest plot. The R2 values were multiplied by the PCC to derive the accuracy for each attribute. The overall accuracy for each reporting unit is the arithmetic average of each of the individual attribute accuracies. Further analysis determined that 195 reporting units contained predicted attribute values that were outliers and could not be relied upon, and were flagged as such.

This project demonstrates how high resolution forest inventory data can enable forestland managers with more comprehensive and reliable information about the current state of the forest to support the design of restorative forest stand management prescriptions.

A key finding was that a fewer number of well distributed FR ground plots, in combination with HRIS and standard field reconnaissance and mapping, can deliver a high quality and reliable forest inventory dataset.

Additional results and information on the SGL No. 221 inventory is provided in the Appendices.

APPENDICES

Figure 1. Pennsylvania State Game Lands No. 221 (Area of Interest)

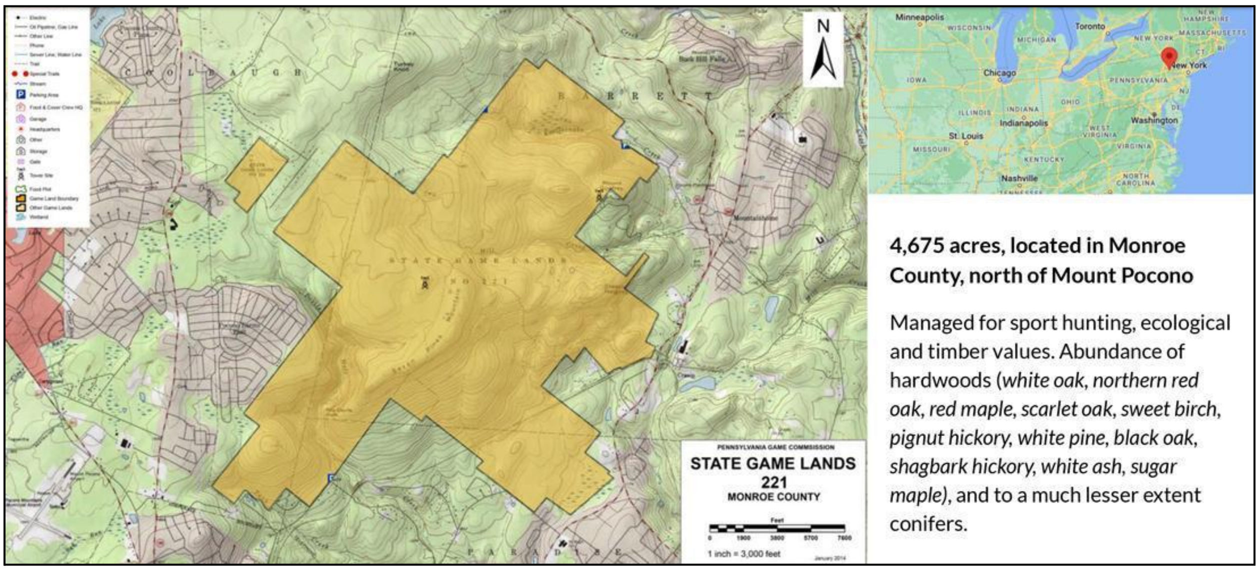


Figure 2. Outline of HRIS Process

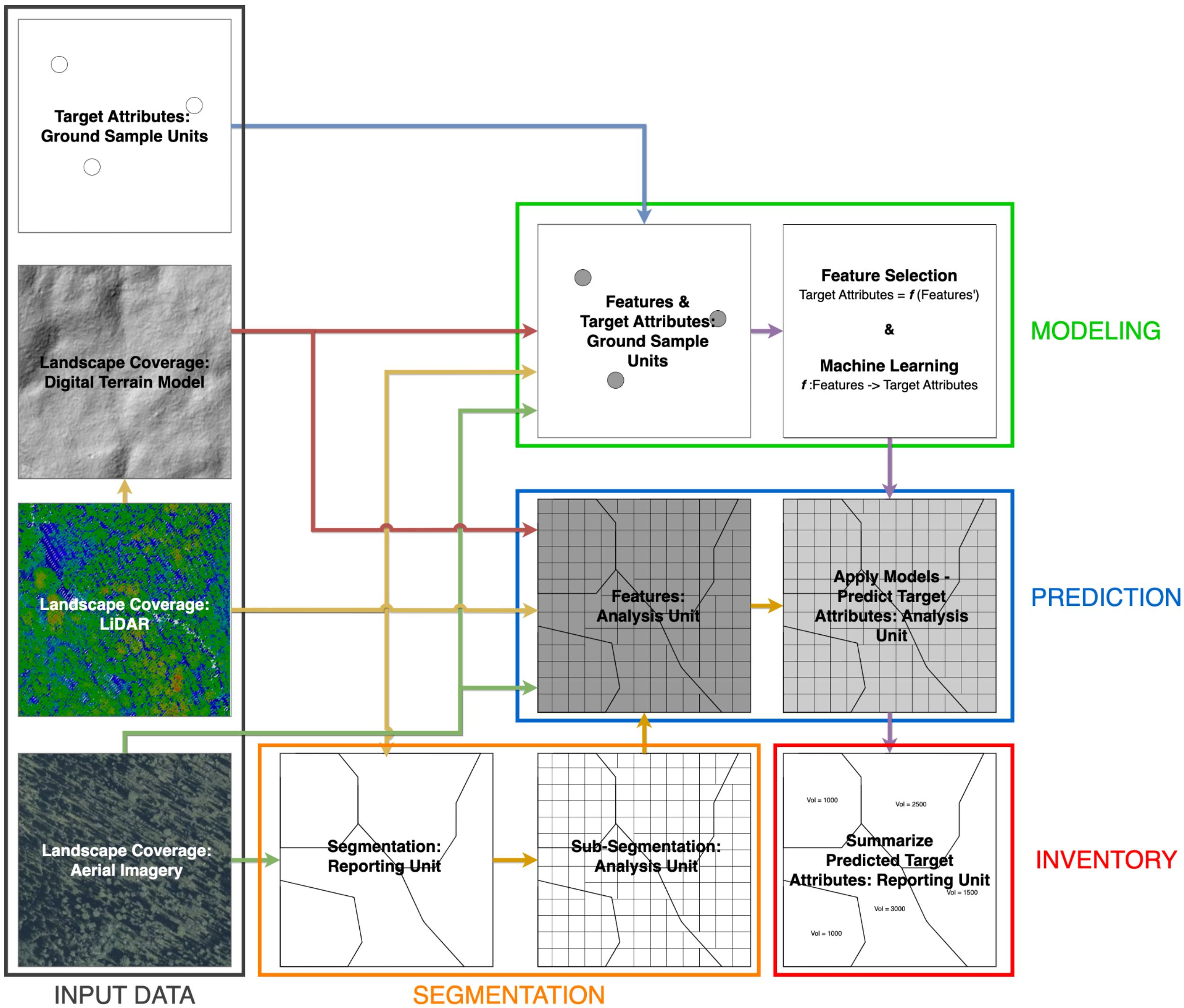


Figure 3. SGL #221: Reporting and Analysis Units

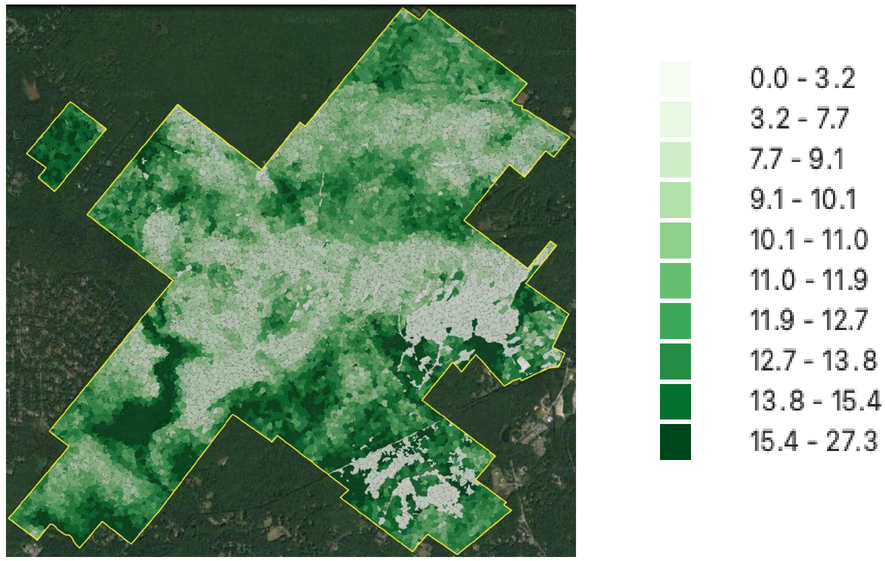
Number of Reporting Units = 9,781

Number of Analysis Units = 62,000

Figure 4. Inventory Diameter: Comparing FR and VR Plots

Diameter: Fixed Radius Plots (Inches; >80% accuracy)

(using FR plots from Pennsylvania State Game Lands No. 221)



Diameter: Variable Radius Plots (Inches; ~60% accuracy)

(using VR plots from adjacent Buck Hill Falls property)

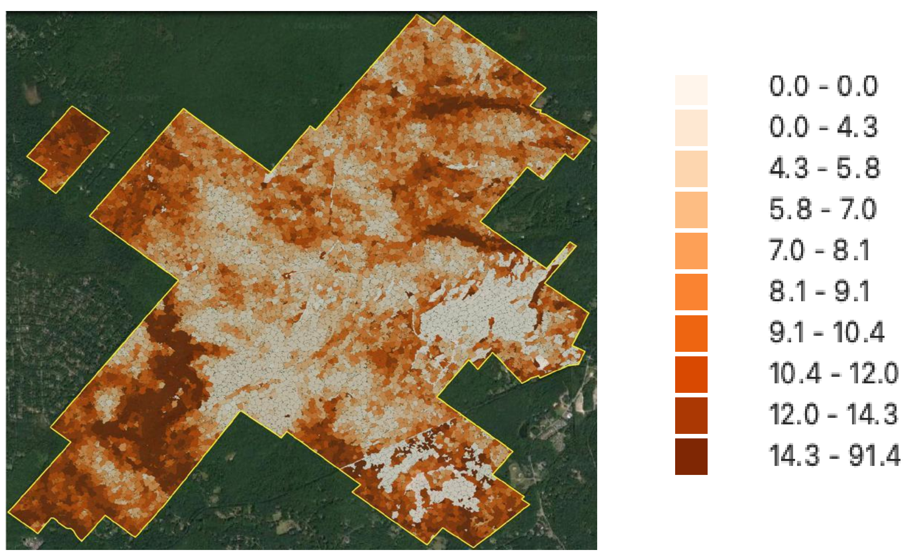


Figure 5. Inventory Height (Feet; >80% accuracy)

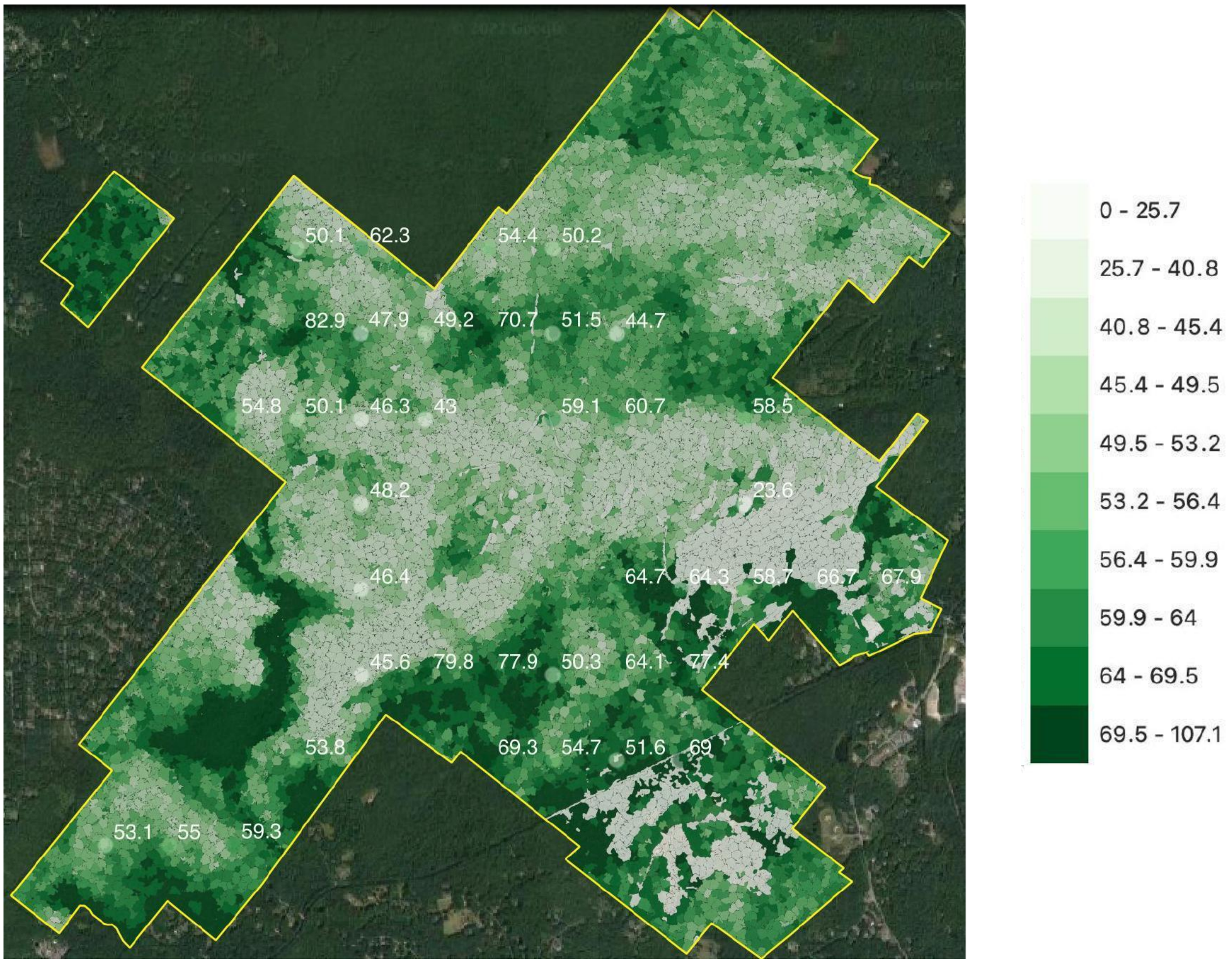
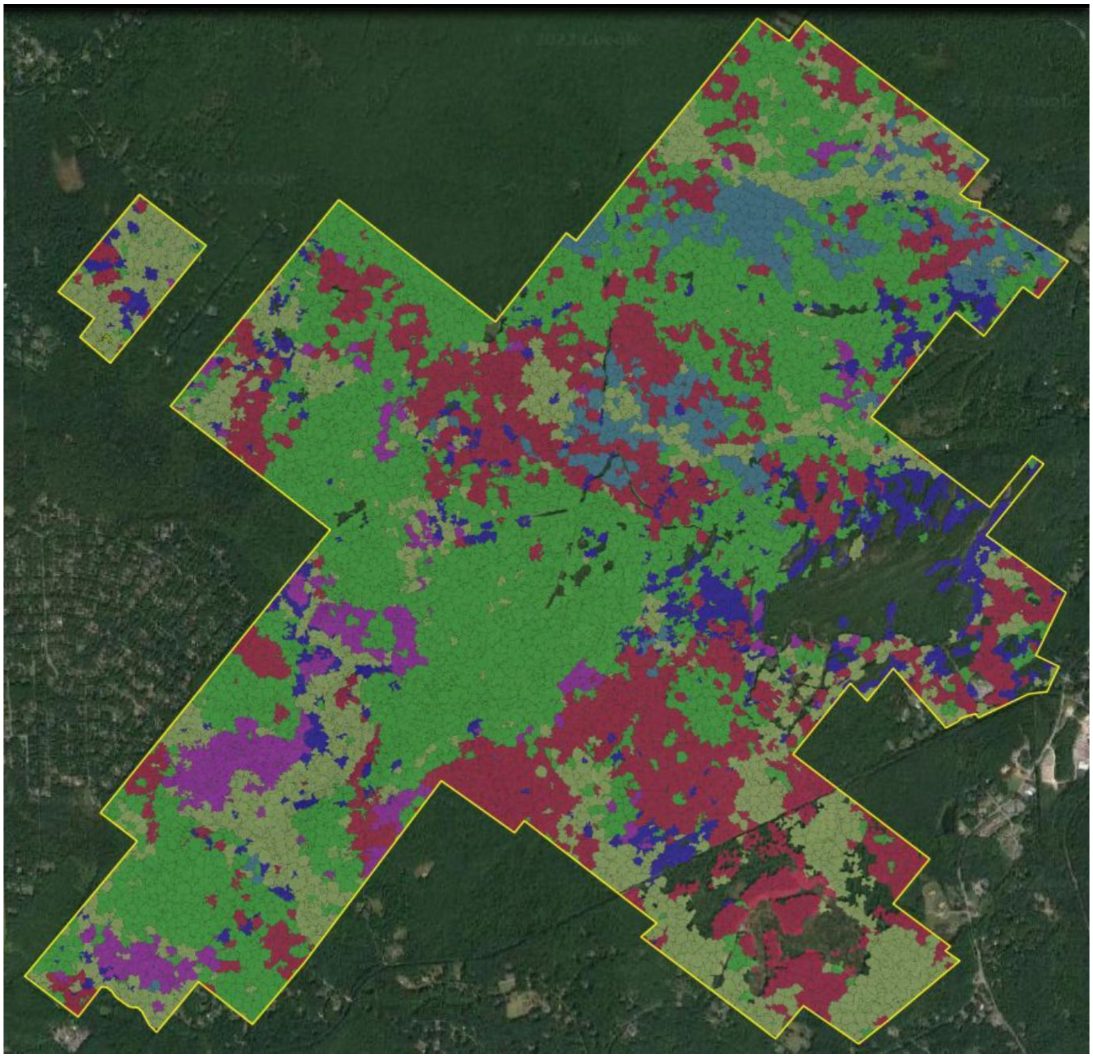


Figure 6. Inventory Leading Species Groups (>80% accuracy)



Red Maple

Black–Pin Cherry, Sugar Maple, Striped Maple, White Ash, Yellow Birch

Black Oak

Aspen, Black Birch, Black Gum, Black Locust, Basswood, Hophornbeam, Sassafras, Serviceberry, Yellow Poplar

Red Oak, American Chestnut, Hickory, Scarlet–White–Chestnut Oak

Beech

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the collaborative efforts and support of:

Advisor:

Dr. Jeff Larkin, Professor – Indiana University of Pennsylvania

Sources of Remotely Sensed Data:

Imagery – Pennsylvania Spatial Data Access

LiDAR – Pennsylvania Spatial Data Access

LiDAR – United States Geological Survey

Ground Plot Measurements:

Josh Flad, Senior Consultant – Green Leaf Consulting Services

Cruise Plot Compilation:

Donn Downey, Forestry Lead – Forest Metrix

High Resolution Inventory Data Analytics:

Michael Kieser, Chief Data Engineer – Tesera Systems

Vipul Malhotra, Data Scientist – Tesera Systems

Tonja Dwyer, Senior Geospatial Developer – Tesera Systems