Stem Volume Modeling in Eastern Texas Loblolly Pine Forests

L. Malambo[1], S. Popescu[1]

[1]Ecology and Conservation Biology, Texas A&M University, College Station TX 77843, USA
Email: mmoonga@tamu.edu

1. Introduction

Stem volume is a key variable in forest inventory that is useful in the assessment of forest productivity. Accurate estimation of stem volume is therefore critical to support sound economic applications such as timber production (Radtke et al. 2017). Airborne LiDAR data are increasingly being applied for the quantification of forest resource volumes (Wulder et al. 2012). Measures derived from airborne lidar data such as tree heights and crown diameters together with relevant indirectly derived parameters such as diameter at breast height (DBH) are used to estimate stem volumes at tree or stand level based on published allometric equations or using regression analyses (Oono and Tsuyuki 2018). Increasingly, estimates of forest stem volumes are needed over large areas for resource managers to evaluate expected amount of timber from a woodshed for timber marketing and management planning. Airborne lidar, although effective for such a purpose, is usually not available over large areas. With the goal of developing wall to wall stem volume product, this study evaluated regression models relating lidar-based stem volume estimates and multitemporal Landsat 8 image data and ancillary existing vegetation height (EVH) datasets from the LANDFIRE program (Rollins 2009) in Loblolly pine (*Pinus spp.)* forests in eastern Texas. We developed reference stem volume estimates by applying published stem volume allometric equations to lidar derived individual tree measurements, which were then aggregated to 30 m Landsat spatial resolution. XGBoost (Chen et al. 2015) regression models were then set up between reference stem volume, as dependent variable, and Landsat data and ancillary EVH data, as predictors. We tested the performance of the regression models against test data at three Landsat image dates and assessed the benefit of combining multitemporal Landsat data in improving the accuracy of the developed models.

2. Data and Methods

2.1 Study site and data

Our study site (centered on Latitude 30° 27' 14.77" N, Longitude 94° 35' 54.54" W) is in south-eastern Texas covering the area between the Texas-Louisiana border on one side and the Sam Houston National Forest on the other side. Several datasets were collected to support the development and evaluation of models for estimating volume including airborne lidar, Landsat, land cover and disturbance data. Airborne lidar data acquired in 2016 under the 3DEP program (Thatcher, Lukas and Stoker 2020) were obtained from OpenTography.com. These data did not cover the entire study site but provided the needed near-ground truth data for estimating individual tree attributes including tree height and crown width. Landsat 8 surface reflectance data acquired on 03 Jan 2018, 24 May 2017 and 29 September 2017 were obtained from the USGS Earth Explorer website and enabled the development and scaling up of stem volume models to the entire study area. We also obtained 2016 LANDFIRE EVH to provide height information. EVH represents the average height of the dominant vegetation for a 30-m cell and is estimated by combining existing airborne lidar measurements and Landsat data (Rollins 2009). The National Land Cover Dataset (NLCD) was also used to provide species cover data, which enabled development of separate volume models for pines species. Forest cover disturbance data were generated over the study site using the LandTrendr algorithm (Kennedy et al. 2018) as implemented in the Google Earth Engine to facilitate exclusion of changed areas from the analysis.

2.2 Processing airborne lidar data

For adequate processing 100 330 m by 330 m sites in Pine forested areas were randomly selected. Airborne lidar data in each of these sites were processed to remove noise and normalized to aboveground level to enable the estimation of individual tree heights. The aboveground level data were then used for individual tree detection and crown segmentation using automated routines implemented in the lidR

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.
Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.1974

package (Roussel et al. 2020). Local Variable Filtering method (Popescu and Wynne 2004) was applied for individual tree segmentation while a method developed by Silva et al (2016) was used for tree crown segmentation.

2.3 Generating reference volume data

Published allometric equations were used to estimate tree attributes not directly estimable from airborne lidar and tree-level stem volume. A critical attribute to estimating tree-level stem volume is tree diameter. An allometric equation for Loblolly Pines developed by Popescu (2007) was applied to estimate diameter and breast height (DBH). Given the crown diameter (CD) and tree height both in meters, DBH was calculate per tree according to (1) as:

$$DBH(cm) = \ 0.16 + CD + \ 1.22H \tag{1}$$

Having determined DBH for each tree, stem volume was calculated based on allometric equations in Radke et (2017), which we do not list here due to space limitation. All stem volume estimates at a tree level were then aggregated at the Landsat scale to facilitate retrieval of matching Landsat and EVH data.

2.4 Stem volume modelling using XGBoost

In our preliminary analyses, we evaluated several regression methods approaches including multiple linear regression, machine learning algorithms such as Random Forests, XGBoost and neural networks for predicting stem volume. XGBoost showed better performance and was adopted for this study. XGBoost, for Extreme Gradient Boosting, is an optimized distributed gradient boosting library which provides a parallel tree boosting to solve many data science problems in a fast and accurate way (Chen et al. 2015). Unlike Random Forests which builds independent trees, XGBoost builds trees sequentially, which provides opportunities for accuracy improvement. To facilitate model building, stem volume data from the 100 sites together with corresponding Landsat and EVH data were combined into one dataset. Non-pine and disturbed samples were removed prior to fitting models. To assess the benefit of multitemporal Landsat data, separate regression models were built using reference stem volume, as dependent variable, and each of the three Landsat images and EVH, as independent variables. A fourth model was built that combined all the Landsat 8 data and EVH. For both models, hyper-parameter tuning was carried out using a grid search approach to select optimal values for the learning rate, number of estimates and the maximum depth of the trees. For each of the models, 85% of the data was used for training and 15% for testing the accuracy of the prediction. The performance of the models was evaluated based on coefficients of variation (R^2), mean absolute error (MAE) and mean absolute percent error (MAPE).

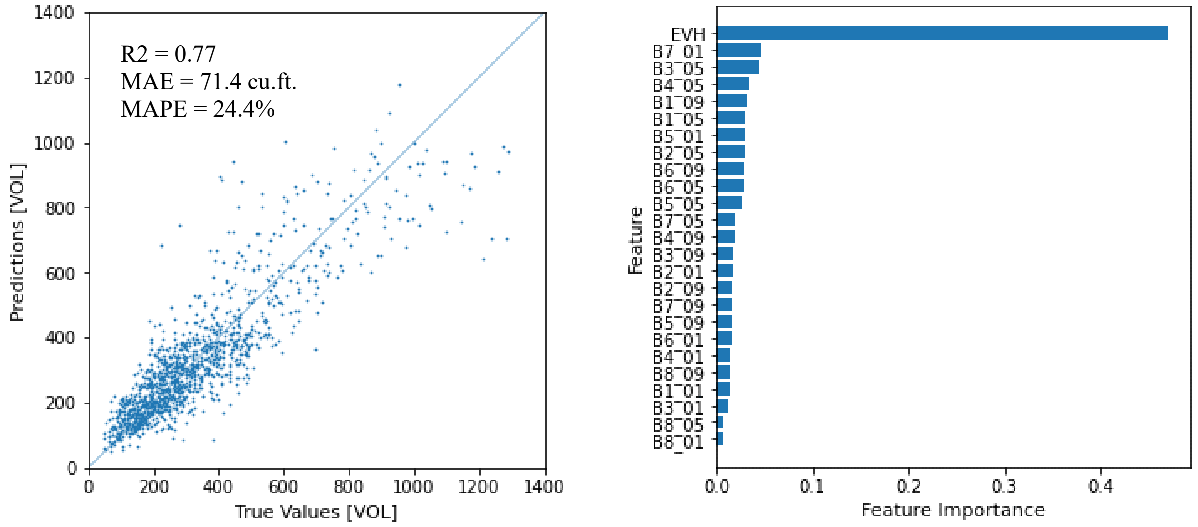
3. Results and Discussion

The total number of samples collected from the 100 sites was 8454. Of this, 7186 were used for training the models and 1268 for testing. Table 1 summarizes the performance of the four regression models trained for predicting stem volume. Model performance varied by Landsat 8 date with the model II trained with data acquired on 05/24/2017 showing the best performance among separate models in terms of R^2 and MAE values. R^2 values ranged from 0.71 to 0.77 and MAE values ranged from 71.4 cubic feet (cu.ft.) to 81.7 cu.ft. estimates. Model performance improved when combined Landsat data were used. All model predictions were within 24 -29% of corresponding reference stem volume.

Table 1: Summary of model performance

Model	Landsat 8 data	R^2	MAE (cu.ft.)	MAPE (%)
1	3-Jan-18	0.71	81.7	29.79
11	24-May-17	0.74	74.5	26.36
111	29-Sep-17	0.72	79.3	28.19
1V	Combined data	0.77	71.4	24.44

In terms of variable importance, the EVH variable was overwhelming significant in all models. However, the variable importance for individual Landsat bands fluctuated with time which is indicative of impact of seasonal changes on forest structure.



a) Predicted vs. Reference stem volume

b) Variable importance for model IV

Figure 1: Combined Regression Model Performance. a) Scatter plot of predicted vs reference stem volume values, b) Variable importance for combined model. B indicates Landsat bands, suffixes _01, _05 and _09 indicate the respect image dates (Jan, May and September)

4. Conclusion

Results from this study show that there is a high potential for developing wall to wall product by leveraging available airborne lidar and multitemporal image data. The improved performance of the developed stem volume models indicates that there is a benefit in applying multitemporal image data, though the gain was not that large in our case. While promising results were obtained in this study, it is expected that even better performance could be achieved by extraction of more features from the EVH and Landsat data such as spectral indices, principal components and other transformations.

References

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang & H. Cho (2015) Xgboost: extreme gradient boosting. R package version 0.4-2, 1.

Kennedy, R. E., Z. Yang, N. Gorelick, J. Braaten, L. Cavalcante, W. B. Cohen & S. Healey (2018) Implementation of the LandTrendr algorithm on google earth engine. Remote Sensing, 10, 691.

Oono, K. & S. Tsuyuki (2018) Estimating individual tree diameter and stem volume using airborne LiDAR in Saga Prefecture, Japan. Open Journal of Forestry, 8, 205.

Popescu, S. C. & R. H. Wynne (2004) Seeing the trees in the forest. Photogrammetric Engineering & Remote Sensing, 70, 589-604.

Radtke, P., D. Walker, J. Frank, A. Weiskittel, C. DeYoung, D. MacFarlane, G. Domke, C. Woodall, J. Coulston & J. Westfall (2017) Improved accuracy of aboveground biomass and carbon estimates for live trees in forests of the eastern United States. Forestry: An International Journal of Forest Research, 90, 32-46.

Rollins, M. G. (2009) LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. International Journal of Wildland Fire, 18, 235-249.

Roussel, J.-R., D. Auty, N. C. Coops, P. Tompalski, T. R. Goodbody, A. S. Meador, J.-F. Bourdon, F. de Boissieu & A. Achim (2020) lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. Remote Sensing of Environment, 251, 112061.

Silva, C. A., A. T. Hudak, L. A. Vierling, E. L. Loudermilk, J. J. O'Brien, J. K. Hiers, S. B. Jack, C. Gonzalez-Benecke, H. Lee & M. J. Falkowski (2016) Imputation of individual longleaf pine (Pinus palustris Mill.) tree attributes from field and LiDAR data. Canadian journal of remote sensing, 42, 554-573.

Thatcher, C. A., V. Lukas & J. M. Stoker. 2020. The 3D Elevation Program and energy for the Nation. US Geological Survey.

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