Stand Density Mapping by Integrating Airborne Laser Scanning Data, Sentinel-1, Sentinel-2 and Topographic Information in Daxinganling Forests

Jinchen W. [1], Xuan M. [2] and Dan Z. [3*]

[1] *Graduate Student, Aerospace Information Research Institute, Chinese Academy of Sciences, China*

[2] *Graduate Student, Aerospace Information Research Institute, Chinese Academy of Sciences, China*

[3] *Associate Researcher, Aerospace Information Research Institute, Chinese Academy of Sciences, China*

*zhaodan@aircas.ac.cn

Introduction

Forests play a crucial role in terrestrial ecosystems, significantly impacting global carbon balance, biodiversity conservation, and climate regulation (Bustamante et al., 2014; Jiang et al., 2021; Keenan et al., 2015). Stand density (SD), the measure of trees per unit area, is key to understanding forest growth, as it influences tree spacing, resource access, and ultimately, ecosystem structure. Accurate SD estimation is essential for predicting forest dynamics and supports sustainable forest management. While traditional SD measurement through field surveys provides precision, it is labor-intensive and impractical for large areas. Unmanned aerial vehicle laser scanning (ULS) and high-resolution remote sensing offer effective alternatives by generating detailed tree distribution data across large regions. For instance, ULS data combined with single-tree segmentation enables fine-grained mapping, though high costs and limited coverage restrict its large-scale use.

Indirect SD measurement methods help address these limitations by leveraging correlations between parameters like vegetation indices and texture characteristics. Open-access data from Sentinel-1 (S1) and Sentinel-2 (S2) satellites, with frequent revisit cycles, provide rich spectral and polarization data beneficial for SD mapping. Research has shown strong correlations between these data sources and metrics like the Normalized Difference Vegetation Index (NDVI) and Fractional Vegetation Cover (FVC) concerning SD. However, these relationships can vary by region, particularly in dense forests with complex canopy structures, which may reduce the precision of indirect methods.

This study integrates field surveys and airborne laser scanning (ALS) to obtain high-accuracy SD data across a target area. Using Sentinel data, supplemented by land cover and topographic information, we apply the random forest (RF) algorithm to estimate SD on a large scale for the Daxinganling forest in Northeast China. This research aims to develop an optimal SD prediction model while examining the effectiveness of variable choices in SD estimation under diverse canopy conditions.

Methodology

Stepwise Multiple Linear Regression (SMLR) and Random Forest (RF) were applied to predict stand density (SD) based on ALS data, Sentinel-1, Sentinel-2, and SRTM topographic data for the Daxinganling forest region. The SMLR model, implemented in R, utilized 60 ALS-based sample plots with extracted height and canopy cover parameters (totaling 48 variables), along with Integration parameters created to bridge height and canopy cover data. The refined SMLR model incorporated 76 variables to explore contributions to SD (Guerra-Hernández et al., 2022).

The study also assessed single-tree partitioning methods, including Canopy Height Model (CHM) and point cloud segmentation for SD estimation. While CHM provides easier implementation, it may smooth canopy shape, limiting understory detection. The point cloud approach, although computationally intense, preserves detail and was selected for use with a range of search grid sizes to enhance segmentation accuracy (Zhang et al., 2022).

For large-scale SD prediction, Sentinel-1 provided backscatter and texture data (e.g., ASM, Contrast, and Entropy), while Sentinel-2 offered multispectral and vegetation indices (e.g., NDVI, EVI, and SAVI), supplemented by SRTM terrain data. Using RF in R, 64 predictor variables were employed to establish an SD regression model. RF's robust handling of outliers and noise enhanced model accuracy, with tuning of Mtry and Ntree parameters to minimize RMSE. Model performance was evaluated using R^2, RMSE (Li et al., 2021, Pötzschner et al., 2022), and relative RMSE. This study demonstrates an optimized approach integrating ALS and satellite data to model SD, contributing insights for forest management and planning across complex canopies.

Findings

The study compared Stepwise Multiple Linear Regression (SMLR) and Single Tree Partitioning (STR) methods for stand density (SD) estimation based on LiDAR parameters. SMLR using ALS data yielded higher accuracy (SMLR: R^2=0.701, SMLR+: R^2=0.750) than STR, which struggled with dense forests (R^2<0.310). SMLR+'s use of advanced canopy cover parameters (AHRCC) improved accuracy slightly, reducing RMSE from 32.126 to 29.511 Trees/pixel and relative RMSE from 25.687% to 23.596%. However, STR segmentation, influenced by search grid sizes (GZ), showed limitations in Sparse and Dense samples due to over- and under-segmentation of trees, respectively.

For broader SD modeling, 4390 ALS-derived samples were used with Sentinel-1 backscatter, Sentinel-2 multispectral indices, and SRTM topography data. Random Forest (RF) was applied, optimizing parameters (Ntree=1000, Mtry=20) to reduce dimensionality to 24 variables, achieving an R^2 of 0.527 and RMSE of 36.564 Trees/pixel. Multispectral data, especially SWIR and red-edge bands, strongly influenced prediction accuracy, contributing 387.19% to %IncMSE. Backscatter data (e.g., VV, VH) and topographical factors like DEM also significantly affected results. Despite good predictions in sparse and moderately dense regions (≤150 Trees/pixel), the RF model underestimated SD in denser areas due to saturation, a limitation of top-down remote sensing in capturing lower canopy layers. These results highlight the potential and challenges of integrating LiDAR and remote sensing data for large-scale SD mapping, particularly in complex forest structures.

Discussion

In constructing the ALS-based SD model, Stepwise Multiple Linear Regression (SMLR) highlighted the importance of canopy cover (CC), tree height (H), and point cloud density distribution (D) in predicting SD, especially in Daxinganling's forest where limited species diversity affects competition for sunlight. Integration parameters, such as canopy height ratios, were added to capture vertical and horizontal canopy variations, which improved SMLR+ model accuracy. However, predictions varied when SD was low or high, as different competition levels affected tree growth directions, impacting model precision.

For wall-to-wall SD modeling, variable selection involved backscatter, multispectral, vegetation index, and topographical data. SAR backscatter data, sensitive to canopy structure, combined with texture features, enhanced spatial discrimination. Multispectral data, especially SWIR, correlated strongly with SD by distinguishing vegetation from soil moisture. Vegetation indices like NPCI provided insights into chlorophyll content, indicating vegetation density.

Topographic factors such as elevation, slope, and aspect also played roles, influencing temperature, sunlight, and water availability, thus affecting growth rates and vegetation structure on sunny versus shady slopes. Addressing saturation in areas exceeding 150 trees/pixel, 2020’s 1m tree height data from Meta and the World Resources Institute was incorporated. This dataset, combined with the DIONv2 model, improved %IncMSE by 52.16% and increased the model’s R^2 from 0.527 to 0.532, demonstrating the value of high-resolution canopy height data. However, an 8-year lag in data poses limitations, highlighting the need for consistent high-resolution updates to mitigate saturation effects and ensure accurate, large-scale SD predictions.

Conclusion

This study developed an ALS-based stand density (SD) prediction model, applying stepwise multiple linear regression (SMLR) algorithms to assess the effectiveness of multisource remote sensing data in the Daxinganling region. Results highlighted the importance of canopy cover (CC), tree height (H), and point cloud density (D) in SD estimation, especially in areas with low species diversity. The SMLR+ model, enhanced with integrated parameters, improved accuracy, particularly in low and high-density forest samples.

For large-scale SD modeling, combining SAR backscatter, multispectral indices, and topographic factors significantly improved spatial discrimination. To address saturation in high-density areas, 1m resolution tree height data from 2020 was incorporated, enhancing model performance slightly. However, the data's temporal lag limited SD prediction precision, underscoring the need for regular updates to high-resolution tree height data to ensure model stability and reliability.

In conclusion, this study provides an effective approach for large-scale SD estimation in complex forest regions and emphasizes the critical role of integrating high-resolution tree height data to mitigate saturation effects and enhance model performance.

References

Bustamante, M., Robledo-Abad, C., Harper, R., Mbow, C., Ravindranat, N. H., Sperling, F., Smith, P. (2014). Co-benefits, trade-offs, barriers and policies for greenhouse gas mitigation in the agriculture, forestry and other land use (AFOLU) sector. Glob Chang Biol, 20(10), 3270-3290. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24700759. doi:10.1111/gcb.12591

Guerra-Hernández, J., Narine, L. L., Pascual, A., Gonzalez-Ferreiro, E., Botequim, B., Malambo, L., . . . Godinho, S. (2022). Aboveground biomass mapping by integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and topographic information in Mediterranean forests. *GIScience & Remote Sensing, 59*(1), 1509-1533. doi:10.1080/15481603.2022.2115599

Jiang, F., Kutia, M., Ma, K., Chen, S., Long, J., & Sun, H. (2021). Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. Sci Total Environ, 785, 147335. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33933773. doi:10.1016/j.scitotenv.2021.147335

Keenan, R. J., Reams, G. A., Achard, F., de Freitas, J. V., Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management, 352*, 9-20. doi:10.1016/j.foreco.2015.06.014

Li, S., Wang, T., Hou, Z., Gong, Y., Feng, L., & Ge, J. (2021). Harnessing terrestrial laser scanning to predict understory biomass in temperate mixed forests. *Ecological Indicators, 121*. doi:10.1016/j.ecolind.2020.107011

Pötzschner, F., Baumann, M., Gasparri, N. I., Conti, G., Loto, D., Piquer-Rodríguez, M., & Kuemmerle, T. (2022). Ecoregion-wide, multi-sensor biomass mapping highlights a major underestimation of dry forests carbon stocks. *Remote Sensing of Environment, 269*. doi:10.1016/j.rse.2021.112849

Zhang, Z., Dong, X., Tian, J., Tian, Q., Xi, Y., & He, D. (2022). Stand density estimation based on fractional vegetation coverage from Sentinel-2 satellite imagery. *International Journal of Applied Earth Observation and Geoinformation, 108*. doi:10.1016/j.jag.2022.102760