Tree Height Estimation Using Sentinel-1/2 and LiDAR Data: A Case Study in the Dalseong Wetlands, South Korea

Baek, G.H..[1], Mohamed, S.Y.[2], Jang, H.K.[3] Park, H.S.[4], Choung Y.J. [5], and Jo M.H. *[6]

[1]Researcher, Geo C&I Co., Ltd., South Korea

[2]Researcher, Geo C&I Co., Ltd., South Korea

[3]Researcher, Geo C&I Co., Ltd., South Korea

[4]Researcher, Geo C&I Co., Ltd., South Korea

[5]Researcher, Geo C&I Co., Ltd., South Korea

[6]Director, Geo C&I Co., Ltd., South Korea

*mhjo7312@gmail.com (*Corresponding author's email only)

***Abstract*:** *Accurate estimation of tree height is fundamental to forest resource management, carbon stock assessment, and ecological monitoring. This study presents a remote sensing-based approach to estimate tree height by integrating multi-source satellite data with machine learning, using the Dalseong Wetlands in Daegu, South Korea as a case study area. The region is characterized by heterogeneous forest cover within a protected wetland ecosystem, offering a suitable testbed for evaluating forest height modelling techniques.*

We utilized Sentinel-2-derived NDVI, Sentinel-1 SLC backscatter (VV and VH), and a DSM generated from Sentinel-1 InSAR imagery as predictor variables. The target variable, reference tree height, was extracted from a LiDAR-derived normalized digital surface model (nDSM). A Random Forest regression model was developed using both original and derived features, including NDVI–DSM interaction, DSM squared, logarithmic and square root DSM, square root VH, and the VV/VH backscatter ratio.

After removing invalid or noisy pixels and standardizing the features, the model was trained on 80% of the data and tested on the remaining 20%. The model achieved a root mean square error (RMSE) of 2.0527 meters and a coefficient of determination (R^2) of 0.6161 when compared to LiDAR-based tree heights.

The final tree height predictions were exported as a georeferenced raster map, which may serve as a valuable baseline for long-term monitoring of forest structure, biomass estimation, and habitat mapping in wetland regions. This research demonstrates the effectiveness of combining freely available SAR and optical satellite data with LiDAR and machine learning to estimate forest parameters in ecologically sensitive or data-scarce areas.

Keywords: Dalseong Wetlands, LiDAR, Random forest, Sentinel-1/2, Tree height estimation

Introduction

Forest canopy height is a critical parameter for assessing vegetation structure and managing wildfire risks, particularly in areas surrounding power transmission infrastructure (Belgiu and Drăguţ, 2011). Overhanging branches can serve as ignition points for wildfires and pose threats to the stability of transmission lines. While LiDAR point cloud data provide highly accurate measurements, their high cost and logistical challenges make large-scale deployment difficult (Simard et al., 2011). To overcome these constraints, this study proposes an affordable and scalable approach to canopy height estimation using freely accessible satellite imagery. The methodology integrates Sentinel-2 multispectral data, providing vegetation indices such as NDVI, with Sentinel-1 SAR data, enabling vertical structure characterization through InSAR techniques. A LiDAR-derived normalized Digital Surface Model (nDSM) serves as the reference dataset for model training and validation.

Datasets and study area

The study area encompasses a forested region surrounding the Dalseong Wetlands in Korea, where UAV-based LiDAR point cloud data were acquired in July 2023. Supporting datasets include:

- Sentinel-2 multispectral imagery (May 2023), processed to generate NDVI, providing information on vegetation density and vigor.
- Single Sentinel-1 SLC imagery (July 30, 2023), used to derive calibrated VV and VH backscatter coefficients.
- Paired Sentinel-1 SLC imagery (July 30 and October 10, 2023), utilized for interferometric processing to produce an InSAR-derived DSM.
- LiDAR-based nDSM, obtained by subtracting the DTM from the DSM, serving as the ground truth for canopy height estimation.

All datasets were co-registered and processed to ensure pixel-level spatial alignment for subsequent analysis.

Methodology

1) NDVI Generation:

NDVI was calculated from Sentinel-2 multispectral imagery using the standard formula (1). This vegetation index provides spatial information on canopy density and overall plant health, supporting the assessment of horizontal vegetation distribution.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

2) Generation of nDSM using LiDAR Point Cloud Data:

A Digital Surface Model (nDSM) for the forest area of the Dalseong Wetland was produced using our company's drone LiDAR point cloud data (July 2023) and derived from DSM-

DTM, serving as reference data.

LiDAR-derived nDSM is calculated as:

$$nDSM = DSM - DTM$$

- DSM: Digital Surface Model from first return

- DTM: Digital Terrain Model from ground points (PMF algorithm used)

- Provides precise reference canopy heights for supervised learning

3) Single Sentinel-1 SLC Image Generation

a. Sentinel-1 SAR Backscatter Processing:

Single-look complex (SLC) Sentinel-1 imagery acquired on July 30, 2023, was processed to extract calibrated VH and VV backscatter coefficients. The preprocessing workflow included:

- Orbit file application
- Data splitting and debursting
- Radiometric calibration
- Multilooking and terrain flattening
- Geometric terrain correction

The resulting calibrated VV and VH backscatter bands were used as input features for model development.

c. InSAR-Derived DSM Generation:

Vertical structural information was obtained through interferometric processing of paired Sentinel-1 SLC images collected 72 days apart (July 30 and October 10, 2023). The InSAR-based DSM was produced through:

- Co-registration of SLC image pairs
- Interferogram formation
- Phase unwrapping
- Conversion of phase values to elevation
- Geometric terrain correction

This DSM captures three-dimensional surface height variations associated with forest canopy structure.

4) Data Integration and Feature Preparation:

Independent variables included:

- NDVI from Sentinel-2 imagery
- VV and VH backscatter from single-date Sentinel-1
- DSM from InSAR processing

The dependent variable was the LiDAR-derived normalized DSM (nDSM = DSM – DTM, Figure 1). After removing null values and applying standardization, a total of 587-pixel samples were prepared for model training and evaluation.

5) Model Training with Random Forest:

A Random Forest regressor was employed to estimate tree height, chosen for its robustness in handling diverse feature types and nonlinear relationships. Model performance was assessed using root mean square error (RMSE) and the coefficient of determination (R^2).

Figure 1. LiDAR-derived nDSM

Results and Discussion

The Random Forest model achieved a root mean square error (RMSE) of 1.82 m and a coefficient of determination (R^2) of 0.63, reflecting a moderately strong predictive capability. Model accuracy is expected to improve with the integration of additional topographic variables and multi-temporal vegetation indices.

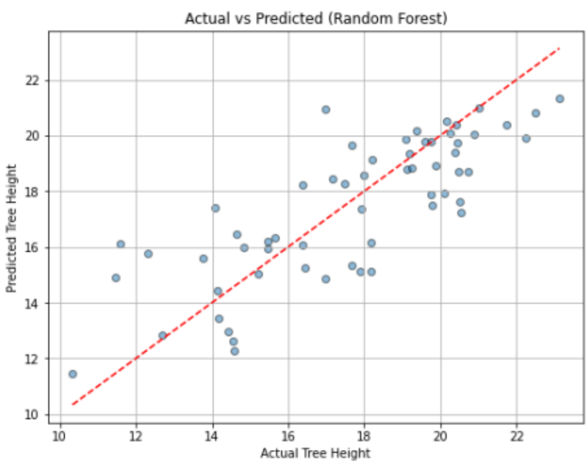

Figure 2. Scatterplot of predicted vs. LiDAR nDSM.

Figure 2 presents a scatterplot comparing predicted canopy heights with LiDAR-derived nDSM values. The distribution of points demonstrates a clear positive relationship, indicating that the model captures key height patterns with reasonable accuracy.

Conclusion

This study demonstrates that satellite imagery from Sentinel-1 and Sentinel-2 can effectively be used to estimate forest canopy height, offering a low-cost alternative to LiDAR in areas where traditional surveys are difficult or expensive. The method holds promise for regional-scale forest monitoring and wildfire risk assessment.

References

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

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research: Biogeosciences, 116, pp. G04021.