Remote sensing and machine learning techniques for above-ground biomass estimation on a regional scale

Abolfazl Abdollahi [a] **and Marta Yebra** [a,b]

[a] *Fenner School of Environment and Society, College of Science, The Australian National University, Canberra, Australia*
[b] *School of Engineering, College of Engineering and Computing Science, The Australian National University, Canberra, Australia*
Email: Abolfazl.Abdollahi@anu.edu.au

Abstract: The estimation of above-ground biomass (AGB) has become a topic of great interest in recent years in light of the increasing concern regarding global warming and the crucial role of forest biomass in carbon sequestration and fuel assessment for forest fires. To better understand the impact of wildfires and environmental deterioration on climate change, an accurate estimate of biomass is necessary. Although biomass generated from measurements of field data is the most precise, it is not a practical method for assessments on a large scale. Collecting ground measurements is a time-consuming and arduous process, making it impractical to obtain the distribution of AGB across regional or global scales solely from ground-based data. Integrating ground data with remotely sensed satellite products with a variety of spectral, temporal, and spatial features is an alternative technique. It allows access to inaccessible locations and can offer data over broad areas at a fraction of the cost associated with extensive sampling. Also, data from remote sensing satellites is available from a number of different platforms at various scales, from local to global.

The current study focuses on the estimation of AGB for the Greater Sydney region, Australia, with 12,368.2 km^2 by combining data from Sentinel-2 and spaceborne Light Detection and Ranging (LiDAR) samples from the Global Ecosystem Dynamics Investigation (GEDI). Sentinel-2 data is a multispectral mission that operates on thirteen different bands and provides information on land surface reflectance from 443-2190 nm, which has acquired popularity in the remote sensing field due to its spatial (10–60m) and temporal (5 days) resolutions. GEDI is a NASA's full-waveform lidar sensor, which its AGB properties have been accessible since April 2019. The availability of GEDI data, which provides lidar-based measurements of forest height and structure has opened up new research avenues for studying AGB at large scales and in remote areas. The high spatial resolution and accuracy of GEDI data make it a valuable tool for monitoring changes in biomass over time, as well as for assessing the impact of natural disturbances such as wildfires, etc. Therefore, we develop the present methodology to generate the AGB map over the study area with the combination of global mapping missions, such as Sentinel-2 and GEDI data. The AGB of GEDI product is used in this study to collect some training samples for the proposed machine learning technique called random forest (RF) and calibrate the Sentinel-2 based AGB model. Machine learning (ML) methods are used in a variety of data processing, and they are promising and current alternatives to analysing remote sensing data. Thus, we use the RF technique as a prediction model to extrapolate the GEDI-based AGB properties over the whole area using multispectral Senitnel-2 data. RF is a popular algorithm in machine learning that works by constructing multiple decision trees and combining their predictions to generate a final output. In addition, our assessment includes an analysis of the impact of spectral and temporal resolutions of multispectral Sentinel-2 satellite data and vegetation indices on AGB prediction. Our results indicate that both multi-temporal Sentinel-2 data and spectral indices play crucial roles, and a model incorporating these factors outperforms a model solely based on the Sentinel-2 spectral bands for AGB prediction. To sum up, by incorporating multi-spectral and temporal Sentinel-2 data trained on GEDI, decision-makers can more effectively monitor changes in AGB over time across regional or global scales. This is particularly important for understanding the dynamics of forest ecosystems and the impacts of disturbances such as wildfires, human activities. etc. Accurate AGB estimates are also essential for carbon accounting and climate change mitigation efforts, as changes in forest biomass play a significant role in the global carbon cycle. Therefore, the use of machine learning and remote sensing data can provide decision-makers with valuable tools to monitor and manage forests and other ecosystems more effectively, ultimately contributing to more sustainable land use practices.

Keywords: *Above-ground biomass, fuel mapping, GEDI, Sentinel-2, remote sensing*