Enzo Zerega

Assessing edge pixel classification and growing stock volume estimation in forest stands using a machine learning algorithm and Sentinel-2 data

Sustainable forest management requires accurate and up-to-date baseline data regarding forest structural parameters and the definition of forest stand units. A crucial component in stand characterization is the identification of tree species and to correctly define the stand boundaries. The most common procedure used for this is land cover classification using remote sensing data. However, when classifying forests, misclassification of edge areas can be substantial, yet classification studies often exclude these areas from the accuracy assessment. With the launch of the Sentinel-2 (S2) mission, which provides medium-high spatial resolution satellite imagery, global coverage and high revisit times, new methodologies for defining and estimating stand parameters are being developed. In combination with machine learning algorithms, the use of S2 data to predict forest variables has demonstrated to generate highly accurate results. Random Forest (RF) is one of these algorithms, and has become increasingly popular in environmental studies during the last decade. The combination of the multi-temporal higher resolution S2 images with the ability of the RF algorithm to detect outliers may contribute to improve the classification accuracy of forest edge areas.

This study presents a methodology based on a combination of field data and S2 multi-temporal imagery that were analyzed using the RF algorithm in southern Sweden. The aim was to perform a land cover classification to identify forest patches of three tree species (Scots pine, Norway spruce and birch), and to test the inclusion of edge pixels as training data to improve the accuracy at edge areas. For this, a segmented accuracy assessment was proposed, where the accuracy was assessed for interior, intermediate and edge areas, as well as for entire forest patches. The RF algorithm was also used to estimate growing stock volume (GSV), another important forest stand parameter. The results indicate that higher accuracies at edge areas can be obtained when edge pixels are included in the training set. Moreover, the findings describe how different allocation schemes of the training and validation data affect the results. Bands covering the Red Edge, SWIR and a narrow segment of the NIR proved to be beneficial, together with the use of multi-temporal scenes. The GSV estimation yielded inferior results but was able to distinguish the S2 bands most correlated with tree volume. The present study contributes to a better characterization of forest stands and, consequently, to facilitate the generation of forest data required by environmental scientists and the forestry sector.

Keywords: Physical Geography, Ecosystem Analysis, Random Forest, remote sensing, accuracy assessment, land cover classification, training stage, multi-temporal, edge areas

Supervisor: **Lars Eklundh**
Co-supervisor: **Abdulhakim M. Abdi**
Master degree project 30 credits in *Geomatics*, 2018
Department of Physical Geography and Ecosystem Science, Lund University. Student thesis series INES nr 458