Using machine learning and GEOBIA for the automation of land cover classification with a time series of Landsat and sentinel 2 data

Masroor Hussain¹, * Wolfgang Lueck¹

¹. PCI Geomatics, 490 Boulevard Saint-Joseph, Gatineau, Québec, Canada, QC J8Y 3Y7 hussain@pcigeomatics.com, luck@pcigeomatics.com

Land cover classifications mainly using single date satellite imagery have been dependent on manual, expensive, and inconsistent qualitative classification techniques. The past decade has seen substantial advancements in: Geographic Object Based Image Analysis (GEOBIA), quantitative image analysis based on high resolution optical satellite imagery, machine learning image classification techniques, and time series analysis of satellite imagery. The rapid availability of frequent observations from sensors and satellite missions - such as Sentinel 2 and Landsat 8, designed for the monitoring of natural resources and land cover, has driven the operationalization of remote sensing techniques.

The authors hereby present a highly automated workflow based methodology, implemented in the PCI Geomatica 2017 software package, to achieve the generation of a highly accurate land cover classification, using imagery acquired over the eastern part of South Africa. A time series of Sentinel 2 and Landsat 8 imagery acquired over a period of one year is used to demonstrate the workflow. All the images were automatically preprocessed to provide radiometrically normalized and quality flagged data that was then reliably used in a data stack for the time series analysis. This preprocessing entails the atmospheric correction of imagery using radiative transfer modelling, spectral pre-classification for the masking of cloud, cloud shadow, water and surface classes with distinct BRDF characteristics. This classification layer was then used for a topographic normalization and BRDF correction over an area with high topographic relief. A set of indices and spectral features were calculated for each dataset and used in a time series analysis. The data was subsequently segmented on the Sentinel 2, 10m GSD bands for this purpose using Geomatica’s region-growing segmentation algorithm. Various objects features were calculated from the entire time series data-stack consisting of both Landsat 8 OLI and Sentinel 2 MSI sensor for pure pixels only, disregarding sensor measurements along the boundary of segments. Thematic features were calculated from temporal matrices that were then used in a supervised Support Vector Machine (SVM) classification.

Classification accuracies obtained are around 90% mean accuracy, and Kappa values larger than 0.8.