Can a hierarchical classification of Sentinel-2 data improve the land cover mapping?

EARSeL 2021
European Remote Sensing-New Solutions for Science and Practice
Abstract
Corresponding Author:
adam.wasniewski@igik.edu.pl

Adam Waśniewski[1], Michał Kałuski[1], Agata Hościło[1], Aneta Lewandowska, Milena Chmielewska[1]

[1] Centre of Applied Geomatics, Institute of Geodesy and Cartography, Modzelewskiego 27, 02-679 Warsaw, Poland, adam.wasniewski@igik.edu.pl

Keywords: Sentinel 2, land cover, random forest, hierarchical classification

Abstract

Monitoring of land cover plays an important role in effective environmental management, assessment of natural resources, environmental protection, urban planning and sustainable development. Increasing demand for accurate and repeatable information on land cover and land cover changes causes a rapid development of the advanced, machine learning algorithms dedicated to land cover mapping using satellite images. Free and open access to Sentinel-2 data, characterised with high spatial and temporal resolution, increase the potential to map and to monitor land surface with high accuracy and frequency. Despite a considerable number of published approaches, there is still a problem with clear separation of some land cover classes, for example grasslands, agricultural land and wetlands. The aim of this study is to examine, if a hierarchal classification of Sentinel-2 data can improve the accuracy of land cover mapping and delineation of the complex classes. The study is conducted in the Lodz Province in central Poland. The pixel-based land cover classification is carried out using the machine learning Random Forest (RF) algorithm, based on a time series of Sentinel-2 imagery acquired in the year 2020. The following nine land cover classes are mapped: sealed surfaces, woodland broadleaved, woodland coniferous, shrubland, permanent herbaceous (grassland and pastures), periodically herbaceous (arable land), mosses and wetland, non-vegetated surfaces and water. The land cover classification is conducted following two approaches: 1) all land cover classes are classified together, 2) applying hierarchical approach by separating in the first place the most accurate and stable classes and then classifying the most problematic classes. Hierarchical classification is carried out in several steps, based on results and accuracy for each class in first classification approach. The research also includes the automated selection and verification of the reference, training datasets, which is a very crucial step in the classification process using machine learning classifiers. The national topographic database served as the reference dataset for the classification process. The classifications are performed iteratively to assess the stability of the classification results. The preliminary results showed that by applying the hierarchical approach the classification accuracy can be increased. The overall accuracy of the first classification approach reached above 91%. The confusion matrix revealed that the highest user and producer accuracy was achieved for water and coniferous forest classes, above 94% and 98%, respectively. By contrast, the lowest classification accuracy was obtained for the non-vegetated areas (around 10%), shrublands (above 15%) and mosses and wetlands (above 47%). These classes are often misclassified as periodically herbaceous. The accuracy of grassland and cropland classes reached below 83%. The permanent herbaceous areas were underestimated at the expense of periodically herbaceous. By applying the hierarchical approach, the classification results for the individual land cover classes have improved significantly. For example, classifying the forest into broadleaved and coniferous inside of the forest mask increased the user's and producer's accuracies by 7 percentage points for woodland broadleaved (to 90%) and 1 percentage point for woodland coniferous (to 99%). By using the hierarchical classification, the user's accuracy of sealed surfaces class increased by 13 percentage points from 56% (for the first classification approach) to 69% (for the second classification approach) and the producer's accuracy increased by 4 percentage points, from 87 to 91%, respectively.

These preliminary results confirmed the potential of the hierarchical classification of Sentinel-2 data for improving the accuracy of the land cover mapping. Further development and exploration of this approach is particularly important for accurate delineation of complex classes that are difficult to separate usually the traditional classification method.

The research leading to these results has received funding from the Norway Grants 2014-2021 via the National Center for Research and Development - project InCoNaDa "Enhancing the user uptake of Land Cover / Land Use information derived from the integration of Copernicus services and national databases".

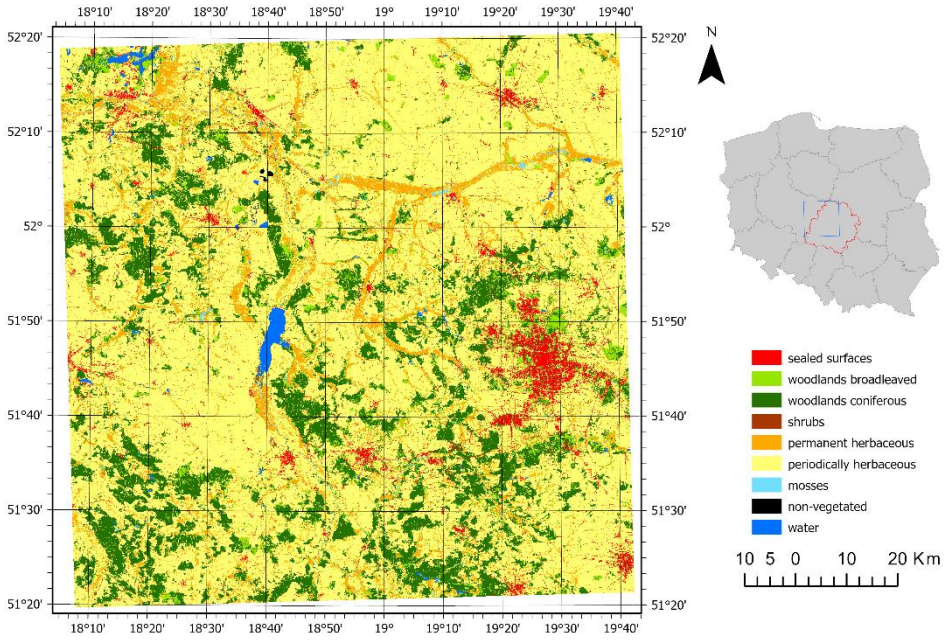

Figure 1. Land cover map as a result of RF classification of all classes together for one Sentinel-2 granule