

EARTH OBSERVATION DATA MINING: A USE CASE FOR FOREST MONITORING

Corneliu Octavian Dumitru¹, Gottfried Schwarz¹, Anna Pulak-Siwiec², Bartosz Kulawik², Jose Lorenzo³,
and Mihai Datcu¹

¹ German Aerospace Center (DLR), Münchener Straße 20, 82234 Weßling, Germany

²SmallGIS, Wadowicka 8a, 30-415, Krakow, Poland

³ATOS SPAIN SA, Calle Albarracín 25, 28037 Madrid, Spain

ABSTRACT

The increased number of free and open satellite images has led to new applications of these data. Among them is the systematic classification of land cover/use types based on patterns of settlements or agriculture recorded by satellite imagers, in particular, the identification and quantification of temporal changes. In this paper, we will present guidelines and practical examples of how to obtain reliable image patch classification results based on data mining techniques for detecting possible changes that can appear within a data set. Here, we will focus on a scenario, namely forest monitoring using Earth observation Synthetic Aperture Radar data acquired by Sentinel-1, and multispectral data acquired by Sentinel-2.

Index Terms—Data mining, Earth observation, forest monitoring, Copernicus data.

1. INTRODUCTION

The Copernicus Access Platform Intermediate Layers Small Scale Demonstrator (CANDELA) [1] is an H2020 research and innovation project, which has among its objectives the development of efficient **data retrieval and mining augmented with machine learning techniques, interoperability** in order to fully benefit from the available assets, the creation of additional value, and subsequently economic growth and development in the European members states.

ECOPOTENTIAL [10] is a large H2020 project that focuses its activities on a targeted set of internationally recognized protected areas, blending Earth observations (EO) from remote sensing and field measurements, data analysis and modelling of current and future ecosystem conditions, and services. The main ECOPOTENTIAL objective is to **improve future ecosystem benefits through Earth observations**.

In this paper, we will concentrate on two forest monitoring scenarios, in which we demonstrate how Earth observation satellite data can be used for forest monitoring using machine learning techniques.

2. DATA SET DESCRIPTION

In the field of remote sensing, there are many satellites for different applications. For the purpose of our two scenarios, we chose various freely available Copernicus Sentinel data sets either from Sentinel-1, a C-band SAR instrument [2], and from Sentinel-2, a 13-band multispectral imager [3]. Besides these Copernicus data, if needed, also third-party very high resolution EO images can be included (e.g., from TerraSAR-X, GeoEye, or Pléiades).

Our first area of interest [4] mainly covers the area affected in 2017 by windstorms near the Bory Tucholskie National Park in Poland (see Figure 1 for an in-situ image from the damaged area).

The park was created in July of 1996. It covers an area of 46.13 km of forests, lakes, meadows and peatlands. The park is located in the northern part of Poland in the heart of the Tuchola Forest, the largest woodland in Poland. In 2010, this park was included in the UNESCO Tuchola Forest Biosphere Reserve [8].



Figure 1: An in-situ image showing forest damages [5].

Our second area of interest is a highly diverse forested mountain domain in Romania that was heavily affected by illegal deforestation. From 2005 to 2009, more than 1,000 hectares around Tarnita Balasaniei were decimated, although the whole area is a fully protected area of the Maramures Natural Park Mountains. Figure 2 shows the deforestation of the forest in this district in 2016, where 220 hectares were cut. As a consequence, a number of natural disasters took place (e.g., landslides and floods), and many houses and agricultural crops were destroyed.

For the investigation of the first area of interest four Sentinel-1 and four Sentinel-2 images were collected from

[3] and [4]. Figure 3 shows the locations of these images, and their common part that was selected for further processing. In the case of the second area, four Sentinel-1 images were collected from [3] and [4]. Figure 4 illustrates their common part selected for further processing.



Figure 2. In-situ data showing the deforestation [9].



Figure 3. Projection of the Sentinel-1 and Sentinel-2 images and the area to be investigated for the first scenario [11].

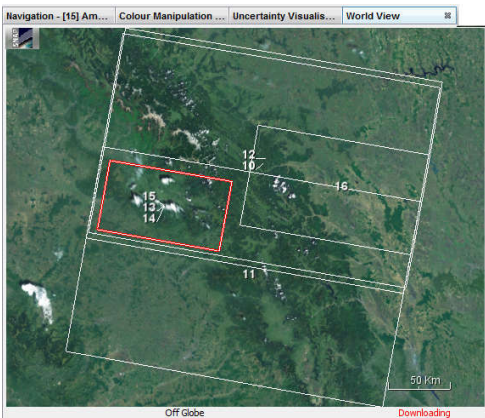


Figure 4. Projection of the Sentinel-1 images and the area to be investigated for the second scenario [11].

3. SCENARIOS

In the case of the first scenario, for the Polish State Forest Holding all updated, latest and relevant information about the forest health conditions is the essential and key information required for forest management. This information has a direct influence on the quality of the

harvested wood, and for the economical income of the forest institution.

Regarding the second scenario, due to massive deforestation in the Maramures region and the Carpathian forests in recent years, a number of floods and landslides occurred and, as a result of these, a large number of houses have been destroyed. In order to manage the existing biomass and to plan future actions, it is necessary to have reliable maps and accurate information. Another goal of image processing is to search for optimized analysis of space-borne images aiming at maximized semantic information content.

Our scenarios are based on determining the areas that have been affected by disasters and how the situations have been managed after their occurrence. This information allows the local authorities to take the necessary measures after the occurrence of an incident. In addition, the data describing the damaged areas (in our case, the Sentinel data) can be enriched with in-situ data or other third-party sensor data. Based on multi-temporal data acquisitions, one can monitor the affected areas by analyzing the changes of semantically annotated corresponding image patches.

Our data mining chain (see Figure 5) comprises six steps and is used for handling and analyzing newly arriving data during routine operations (step 1 to 5), and to quantitatively analyze the data content of images and their changes (step 6) [6], [7]. Note that, during a preparatory phase, experts have to select and download images of the area of interest, and do all the processing in order to generate the reference data that are used later for comparisons during the operational phase. In the case of multi-sensor data (e.g., Sentinel-1 and Sentinel-2) or in the case of partially overlapping data of the area of interest, some preprocessing steps are needed in order to cut out the common area of interest. A fast solution is the use of the SNAP toolbox [11] offered by ESA.

4. EXPERIMENTAL RESULTS

Following the processing chain shown in Figure 5, after the selection of the area of interest for each scenario, the images are tiled into patches of 128×128 pixels (1760 patches/image in the case of the first scenario, and 3354 patches/image in the case of the second scenario). From each patch a feature vector is extracted (we used four multispectral [12] and four SAR images [13]). The next step is the classification of the image patches, by active learning, into semantic categories. Based on the extracted features and the specific patterns of these categories we were able to separate them during classification. The number of categories depends on the sensor and the content of the image.

For the first scenario, Figure 6 illustrates the retrieved categories after the classification projected on each quick-look image (comprising 5128×5723 pixels). In this case, based on the available data, we were able to acquire Sentinel-1 and Sentinel-2 data with close acquisition times.

For the second scenario, Figure 7 (like Figure 6) presents the retrieved categories projected on each quick-look image (with a size of 10053×5623 pixels). For this area, no Sentinel-2 data were available.

Considering the results of the classification for each scenario and sensor separately, we identified a number of patches with different semantic annotations. When comparing the image pairs for each sensor in each scenario, a number of patches with dissimilar semantics were identified [6]. Figures 8 and 9 show the differences between the semantics attributed to each patch of the two images of

the same sensor, taking the first recorded image as reference image.

5. CONCLUSIONS

This type of analysis of the selected data can lead towards important solutions for crisis management in the area of forestry. The outputs of these scenarios can be forest change maps of damaged areas, the percentage of damages/changes within the selected areas, and repetitive maps of forest preservation, etc. Some of these selectable outputs of our processing chain were presented in this paper.

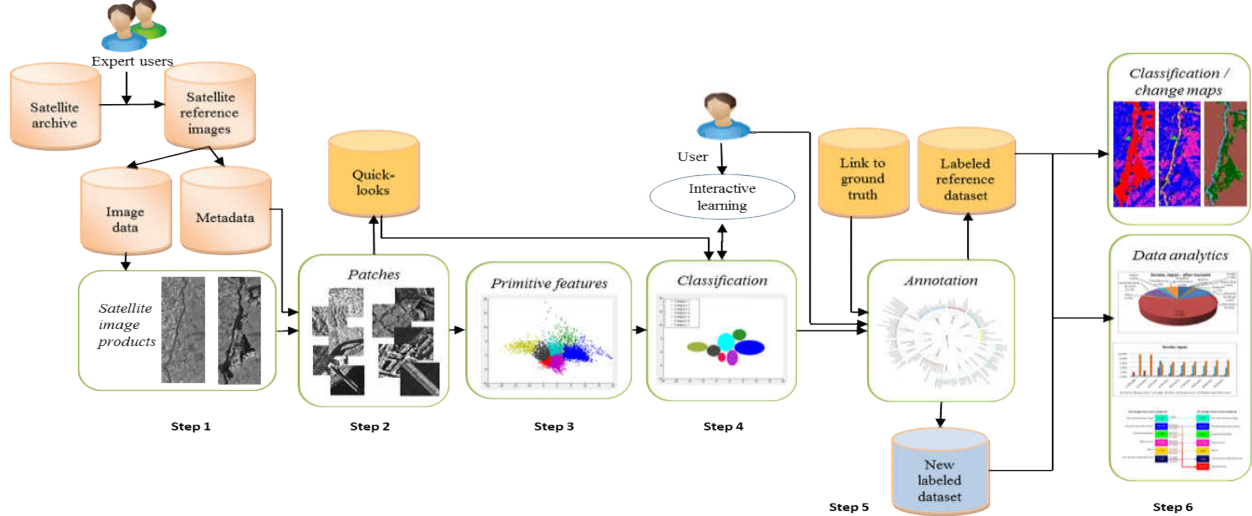


Figure 5: Processing chain used to analyze the content of the data and to generate classification maps, data analytics, and accuracies.

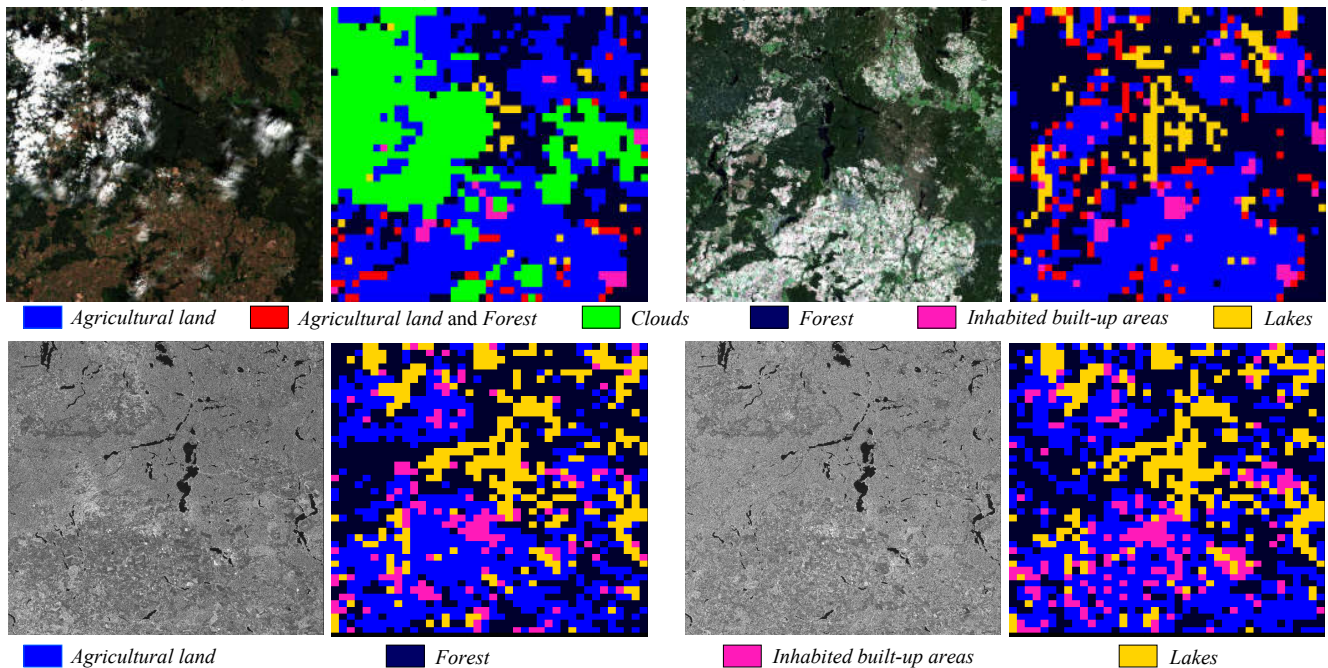


Figure 6: A multi-sensor and a multi-temporal data set for the first scenario: (top - from left to right) a quick-look view of the first Sentinel-2 image from July 30, 2017 and the classification map, and a quick-look view of the second Sentinel-2 image from Sept. 28, 2017 and the classification map; (bottom - from left to right) a quick-look view of the first Sentinel-1 image from Jul. 30, 2017 and the classification map, and a quick-look view of the second Sentinel-1 image from Aug. 29, 2017 and the classification map.

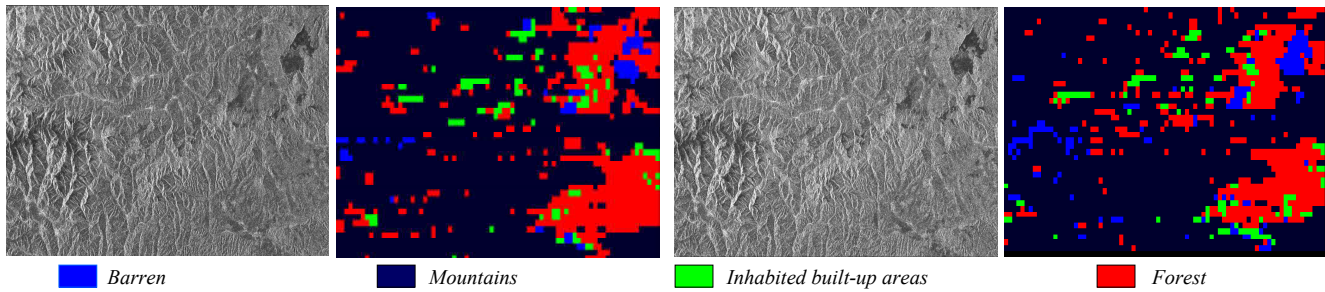


Figure 7: A multi-sensor and a multi-temporal data set for the second scenario: (top - from left to right) a quick-look view of the first Sentinel-1 image from June 27, 2015 and the classification map; (bottom - from left to right) a quick-look view of the second Sentinel-1 image from Sep. 01, 2016 and the classification map.

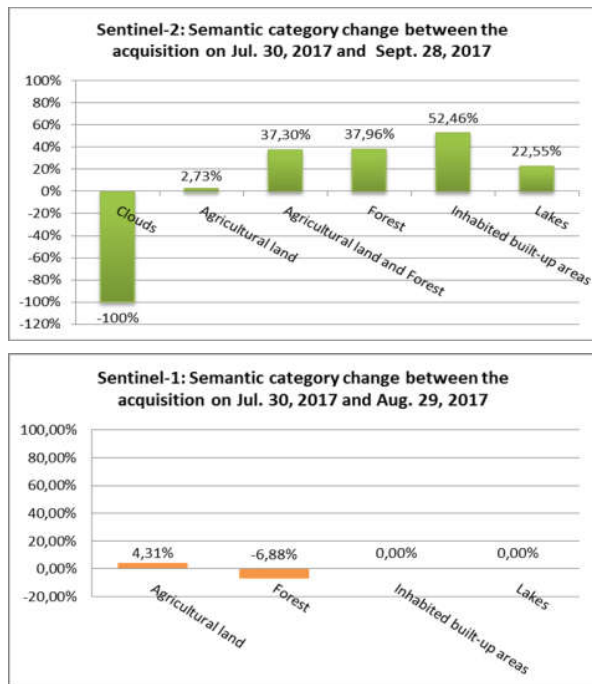


Figure 8: Semantic label change between two Sentinel-2 and two Sentinel-1 images in the case of the first scenario.

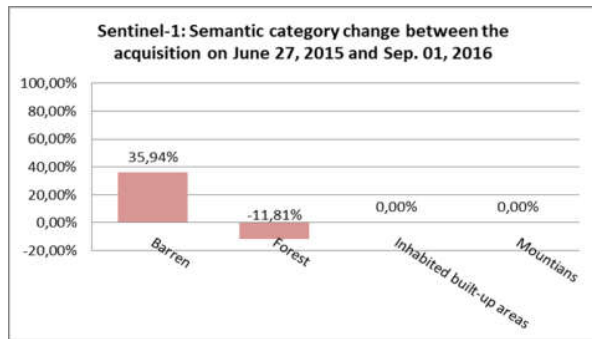


Figure 9: Semantic label change between two Sentinel-1 images in the case of the second scenario.

6. ACKNOWLEDGEMENTS

The work for the first scenario was partially supported by the H2020 CANDELA project under grant agreement No. 776193, while the work for the second scenario was partially supported by the H2020 ECOPOTENTIAL project under grant agreement No. 641762.

7. REFERENCES

- [1] Candela (Copernicus Access Platform Intermediate Layers Small Scale Demonstrator) project. Available online: <http://www.candela-h2020.eu/> (accessed in December 2018).
- [2] Sentinel-1. Available online: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1> (accessed in December 2018).
- [3] Sentinel-2. Available online: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2> accessed in December 2018).
- [4] SmallGIS. CANDELA deliverable D.1.2: Use Case#2 Requirements, 31.10.2018, 19 pages.
- [5] General Directorate of the State Forests. Available: <http://www.lasy.gov.pl/pl/informacje/publikacje/in-english/forests-in-poland/lasy-w-polsce-2017-en.pdf> (accessed in December 2018).
- [6] C. Dumitru, S. Cui, D. Faur, and M. Datcu, "Data Analytics for Rapid Mapping: Case Study of a Flooding Event in Germany and the Tsunami in Japan Using Very High Resolution SAR Images", IEEE JSTARS, 8(1), pp. 114-129, 2015.
- [7] C. Dumitru, G. Schwarz, and M. Datcu, "SAR Image Land Cover Datasets for Classification Benchmarking of Temporal Changes", IEEE JSTARS, 11(5), pp. 1571-1592, 2018.
- [8] Bory Tucholskie National Park. Available online: https://en.wikipedia.org/wiki/Bory_Tucholskie_National_Park (accessed in May 2019).
- [9] (Romanian) Ministry of the Environment; © 2019. Available online: <http://www.mmediu.ro> (accessed in February 2017).
- [10] ECOPOTENTIAL project. Available online: <http://www.ecopotential-project.eu/> (accessed in May 2019).
- [11] Sentinel Application Platform (SNAP). Available online: <https://step.esa.int/main/toolboxes/snap/> (accessed in May 2019).
- [12] F.A. Georgescu, D. Raducanu, and M. Datcu, "New MPEG-7 Scalable Color Descriptor Based on Polar Coordinates for Multispectral Earth Observation Image Analysis", IEEE Geoscience and Remote Sensing Letters, 14 (7), pp. 987-991, 2017.
- [13] B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", IEEE PAMI, 18(8), pp. 837-842, 1996.