GeoTerrace-2021-005

Classification of urban land coverage using satellite data and random forest classifier

***A. Palamar** *(Kryvyi Rih National University)* **M. Malanchuk** *(Lviv Polytechnic National University),* **M. Hanchuk, L. Datsenko** *(Dmytro Motornyi Tavria State Agrotechnological University)*

SUMMARY

Satellite data provides great potential and opportunities for environmental applications. In addition, open-access data is an independent channel for obtaining environmental information. Numerous non-governmental environmental organizations in the world use satellite data for their own activities. In this study, the usefulness of the classification of geoinformation systems of remote sensing for the cartography of urban lands was investigated. The main purpose of the work is to study various vegetation areas in urban settlements and industrial zones, for example, in the Kryvyi Rih city. After completing the production of maps, a visual and field check of the classification accuracy of the buffer zones is performed. Some problems are not excluded due to small smoothing options. Numerous buffer zones between earth classes cause numerous small mosaics on the final map, but these mosaics provide additional information about the differences in the surrounding area.

Keywords: satellite data, GIS, remote sensing, vegetation, mapping

Introduction

Kryvyi Rih is a city in the Dnipropetrovsk region of Ukraine. It is the eighth most popular city in Ukraine (Regional report on the state of the environment in the Dnipropetrovsk region for 2015, 2015). As of 2018, Kryvyi Rih is perhaps the principal city of Eastern Europe, which is a globally important centre of the iron ore mining and metallurgical industry - Kryvbas (Kryvyi Rih Iron Ore Basin). The city stretches from north to south, parallel to the iron ore deposits. The city's environmental safety is a growing problem through abandoned mines and industrial waste pollution. The city is one of the most dangerous places to live due to these problems and according to the Scientific and Hygienic Center of Ukraine (Regional report on the state of the environment in the Dnipropetrovsk region for 2015, 2015).

Thus, Kryvyi Rih is a large industrial centre of Ukraine, which includes enterprises of mining, processing and metallurgical industries. The city has a complex earth cover with many different zones and with a combination of vegetation and hard shell in different proportions. The land cover includes settlements, huge open pits, large industrial zones, post-industrial reclaimed zones and zones with complex soil vegetation (Regional report on the state of the environment in the Dnipropetrovsk region for 2015, 2015).

Field methods for land mapping are very expensive because the area of the city is huge. So, remote sensing techniques are useful and fast. But there are several problems for interpreting the results. The problems are similar to differentiating boulevards (water coastline and flooded areas) from parks and forests.

The differentiation of parks and forests from tree zones in residential areas is also important. For this study, the Environmental Council of Kryvyi Rih used a combination of optical and synthetic radar images for better differentiation of trees and vegetation zones (Wegmann et al., 2016).

Results of investigations

The Environmental Council of Kryvyi Rih used a set of free software projects with open source. Also, data from satellites from open databases of the United States Geological Survey (USGS) and the European Space Agency (ESA) were used. For the separation of interests, the data of the international project Open Street Map (Liaw and Wiener, 2002; Pebesma and Bivand, 2005; Roger et.al, 2013) were used.

We used 10 spectral bands with pixel sizes of 10 and 20 m. Sentinel-2A images, 20 m bands were pulled by 10 m. We also used Landsat-8 thermal infrared bands (10 and 11) and both Sentinel-1A polarization images (VV and VH). Radar and thermal data have also been redistributed to a resolution of up to 10 meters. A Gamma Map filter with a small averaging window (3X3 pixels) was used to protect against excessive anti-aliasing for speckle filtering of radar data. It was evaluated empirically. Larger buffer zones around roofs in single-storey residential areas, buildings and roads provide a wider smoothing radius. But the main purpose of the study was to assess areas with grasses and trees with a lower level of smoothing (Hijmans, 2017).

The training dataset was prepared in several stages. Before studying the polygon that forms the study area, the SAGA GIS Object-Based Image Segmentation algorithm was segmented. The third, fourth, and eighth Sentinel-2A bands and both Sentinel-1A polarizations were used for OBIS.

Training polygons for all classes of land cover based on the known territory were selected. Figure 1 shows a fragment of a small part of Kryvyi Rih with segmented polygons of earth cover classes. Nine nominal classes of land cover were selected for mapping. The choice is based mainly on the characteristics of vegetation. Differentiation of hardcover zones occurs outside of these studies. A brief description of the nominal classes is presented in Fig. 2.

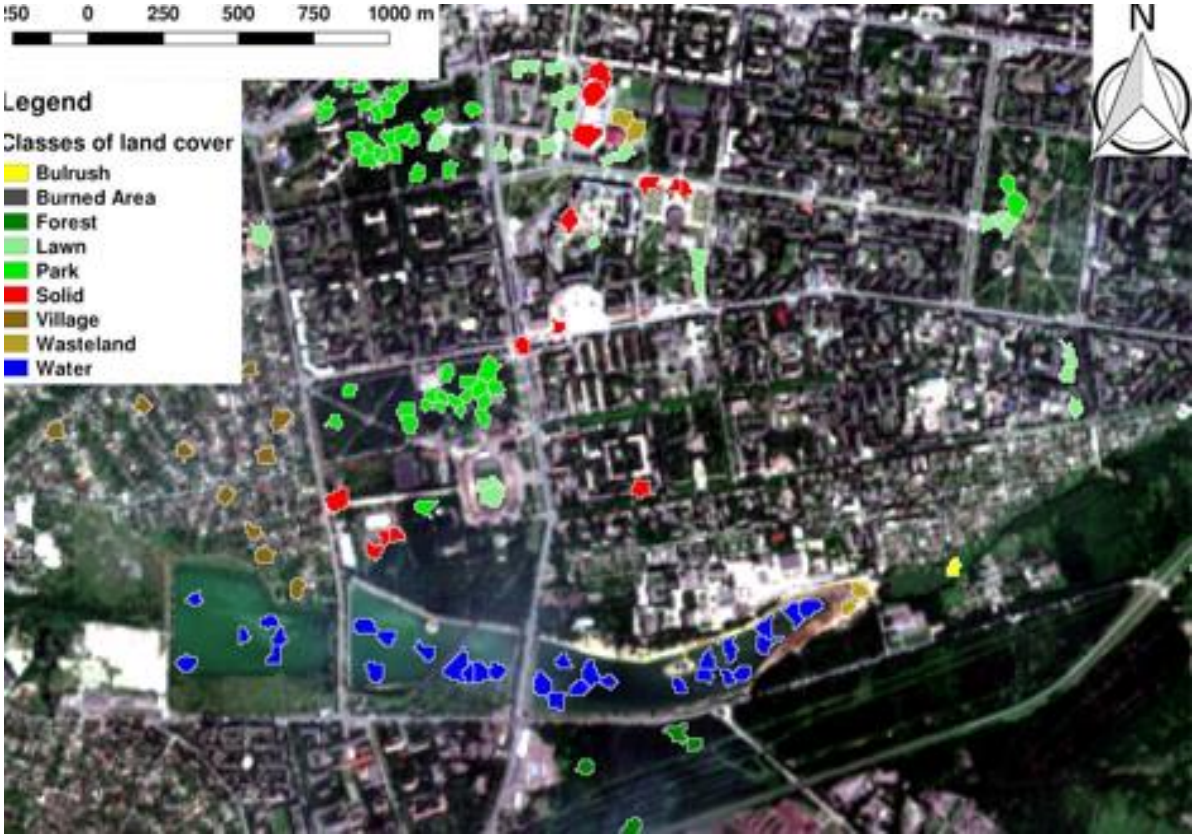

Figure 1 *A fragment of a small part of Kryvyi Rih with segmented polygons of earth cover classes*

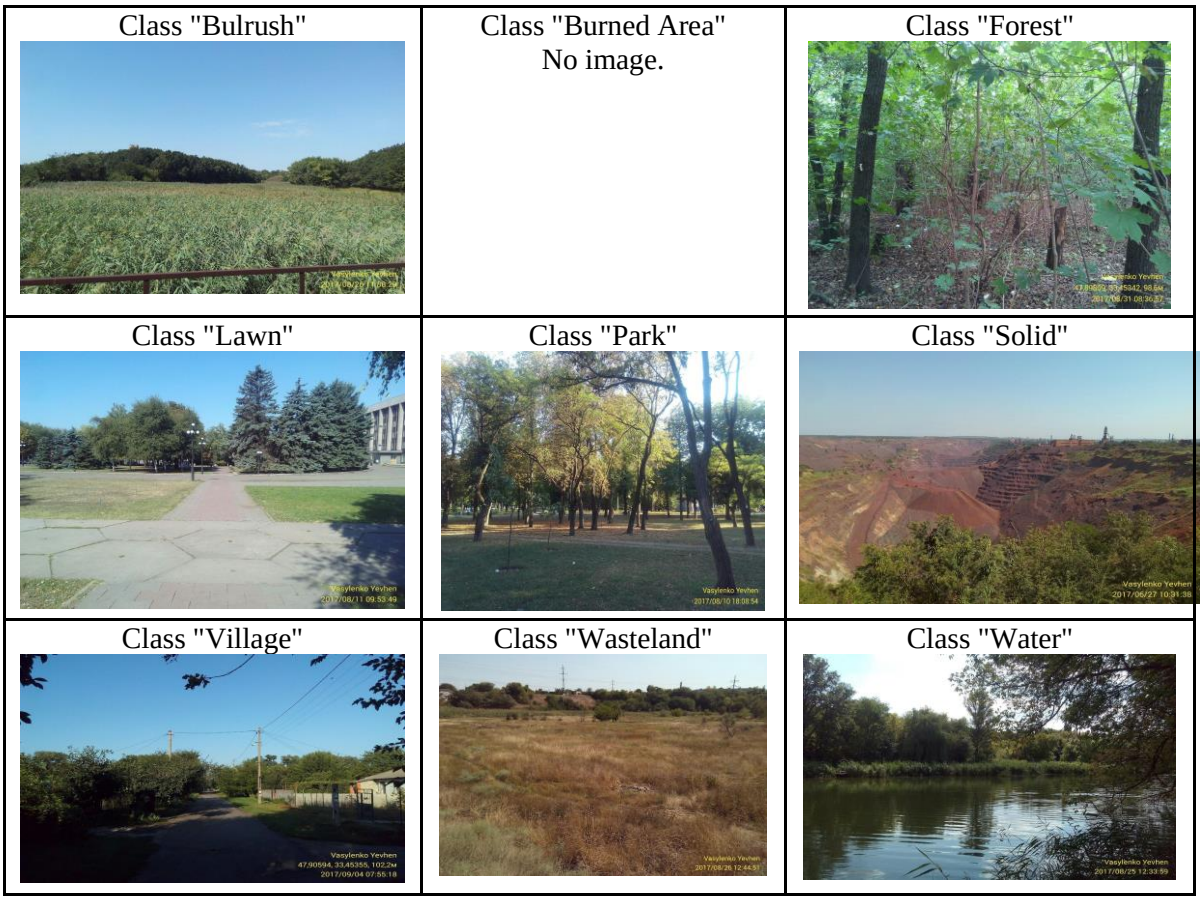

Figure 2 *Nominal classes of land cover*

All types of land have blurred boundaries with a smooth transition from one to another. They will be fixed in numerous prediction errors for buffer zones after classification. A simple diagram of statistical blocks shows that all nominal classes have blurred boundaries with interclass buffer zones, as shown in Figure 3.

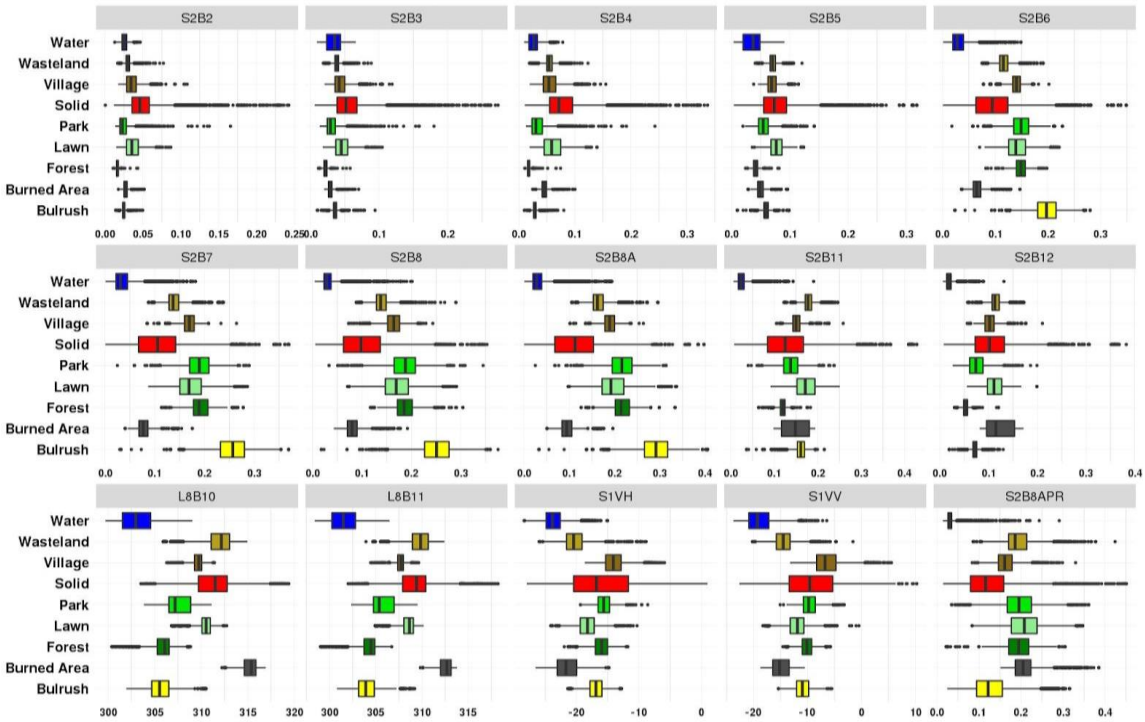

Figure 3 *Optical and radar characteristics of nominal classes of urban land cover in Kryvyi Rih*

The Random Forest classifier from the RandomForest library of the R programming language was used as an algorithm for information training by the Environmental Council of Kryvyi Rih (Wickham, 2009). Pre-set random values "12210" were used to reproduce the studies. The following parameters were used for growing trees: TRUE replacement, node size 2, number of variables in each tree 3, the total number of trees 500. The errors Out of Bag for calculating steps are shown in Fig. 4.

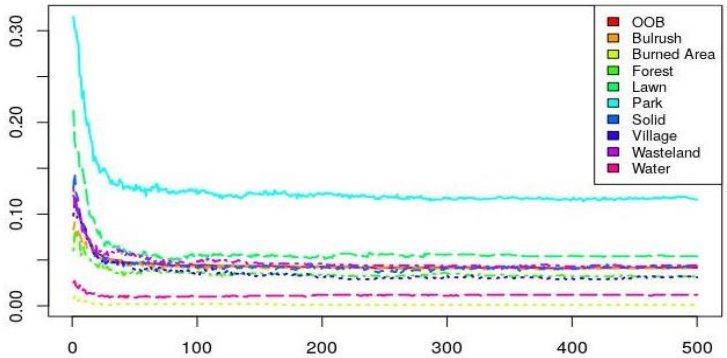

Figure 4 *Overall estimate of the error rate*

The overall estimate of the Out of Bag error rate is 4.17%. As shown above, after 70-100 trees had grown, the OOB error was still low. But the project was provided as an educational project in GIS, remote sensing, so a standard batch of parameters "500 trees" was chosen for the Random Forest classifier. At the end of the classification, the information of the significance of each factor is checked using standard GIS functions.

After the production of the maps, a visual and field check of the classification accuracy was performed, which is shown in Fig. 5. Some problems are not excluded due to small smoothing

parameters. Numerous buffer zones between classes of earth cover cause numerous small mosaics on the final map. But these mosaics provide additional information about the differences around the territory.

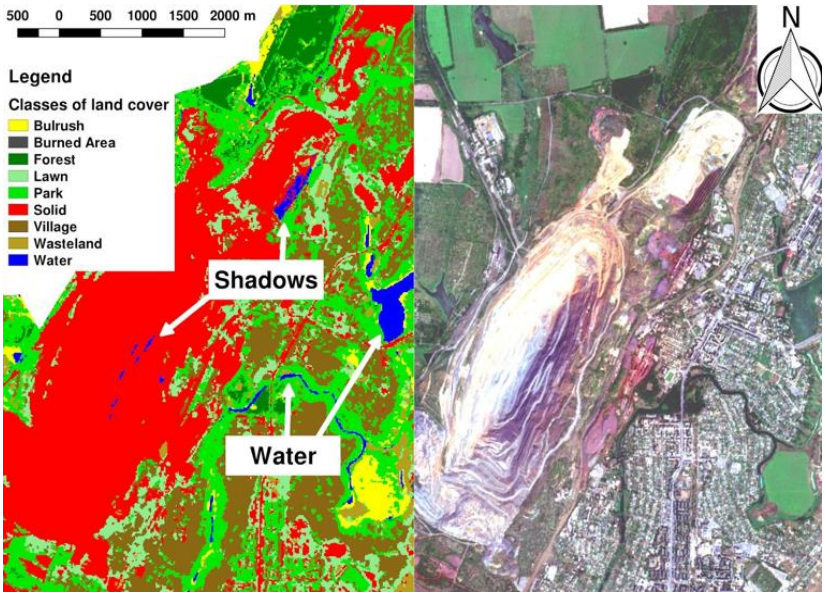

Figure 5 *Comparison of the photo with parts of the map*

After analyzing the maps, the problem was that the study was not focused on identifying agricultural products. In turn, on the border of Kryvyi Rih, there are several fields with sunflowers and other crops. Sunflower fields were classified as a park, as a lawn and as a village. Thus, the classification model in these agricultural areas is not consistent with reality. To avoid this error, more time intervals should be used to retrieve the data. Moreover, the inclusion of satellite images from several periods will provide high accuracy for the classification of both types of landscape: urban and agricultural.

Conclusions

This study is based on open-source remote sensing data and software. This work was supported by the non-governmental organization for environmental protection Environmental Council of Kryvyi Rih.

References

Hijmans R. J. (2017) Raster: Geographic Data Analysis and Modeling. R package version 2.6–7.

Liaw A., Wiener M. (2002). Classification and Regression by Random Forest. *R News* 2(3), 18-22

Pebesma E.J., Bivand R.S. (2005). Classes and methods for spatial data in R. *R News* 5 (2)

Regional report on the state of the environment in the Dnipropetrovsk region for 2015. (2015). (In Ukrainian) https://adm.dp.gov.ua/storage/app/uploads/public/605/06f/469/60506f4698e21842633226.pdf

Roger S.B., Pebesma E.J., Gomez-Rubio V. (2013) Applied spatial data analysis with R, the Second edition. Springer, New York.

Wegmann M., Leuther B., Dech S. (2016) Remote Sensing and GIS for Ecologists Using Open Software. Series - Data in the Wild. *Pelagic Publishing*.

Wickham H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.