Advancing Land-Cover Classification in Croatia: Implementation of a Pilot Project for ARKOD+ and CROLIS LU Initial Layer Creation

Matjaž Štanfel[1,], Dragan Vitas[1], Mladen Majcen[2], Dario Perešin[2]*

[1] Sinergise Solutions d.o.o., Cvetkova 29, 1000 Ljubljana, Slovenia, matjaz.stanfel@planet.com, dragan.vitas@planet.com
[2] Agencija za plaćanja u poljoprivredi, ribarstvu i ruralnom razvoju, Ulica grada Vukovara 269d, 10 000 Zagreb, Croatia, mladen.majcen@apprrr.hr, dario.peresin@apprrr.hr
[*] corresponding author

doi: 10.5281/zenodo.11657376

Abstract: The pilot study was conducted in collaboration with the Paying Agency for Agriculture, Fisheries and Rural Development (PAAFRD), with implementation led by Sinergise Solutions d.o.o. and KING ICT d.o.o. as executors. This pilot initiative was integral for fulfilling objectives outlined in the LIFE CROLIS project, with PAAFRD as a key partner. The primary aim of the pilot project was to analyse training data and evaluate methodologies for automating land cover classification within the Republic of Croatia using Sentinel-2 satellite imagery. Additionally, the project aimed to generate an initial vectorized spatial dataset of agricultural land parcels (ARKOD+) by application of a delineation algorithm on aerial imagery. Over a period of 10 months, starting on May 23, 2023, the project consisted of multiple phases organized into work packages, including, among others: analysis of input data and spatial frameworks, developing and applying algorithms for land-cover classification and field delineation, evaluating results, preparing future methodology recommendations for field data collection/validation, and creating a GIS viewer. Additionally, a comparison was conducted between the use of PlanetScope Fusion and Sentinel-2 imagery for calculating markers identifying agricultural use in the Koprivničko-križevačka region. This paper presents the framework, methodologies, and outcomes of this pilot project, shedding light on the feasibility and effectiveness of utilizing satellite data for automated land classification, with implications for land management and policy development.

Keywords: CROLIS; ARKOD+; agriculture; land cover classification; automated delineation.

1 Introduction

The LIFE CROLIS project deals with the development of a harmonised data model for land monitoring in the Republic of Croatia with the aim of developing and establishing the first multi-level and multi-purpose land monitoring system. The establishment of the CROLIS system will enable better climate policy planning (mitigation and adaptation) in various sectors at the level of the Republic of Croatia. Additionally, the application of data from CROLIS will enable more precise reporting and calculation of greenhouse gas emissions and sinks from the LULUCF sector (land use, land-use conversion and forestry) and the CROLIS system will provide a stable basis for planning and implementing climate change mitigation activities through reducing emissions and increasing greenhouse gas sinks in the LULUCF sector.

With new EU legislation, including the adopted Regulation (EU) 2023/839 of the European Parliament and of the Council, which simplified the reporting and compliance rules and set out the targets of the Member States for 2030 on improving monitoring, reporting, and tracking of progress, the establishment of the Land Monitoring System in Croatia becomes an obligation.

One of the most important effects of the establishment of CROLIS is the precise reporting of greenhouse gas sinks in the LULUCF sector. According to the CROLIS project, the entire land area of the Republic of Croatia will be included in a harmonised land monitoring information system and, as a result, new information will be obtained about the connection between Croatian land management practices and emission factors. This will help to introduce land management practices with lower greenhouse gas emissions, and ultimately will help achieve reductions in greenhouse gas emissions.

Since the ARKOD (LPIS) system includes only agricultural areas whose owners submit a request for payment of incentives, for now less than 50% of the total agricultural areas in the Republic of Croatia are registered in ARKOD. After the establishment of CROLIS, all agricultural areas will be determined and officially registered, which will enable better planning of climate change mitigation measures.

2 Materials and methods

2.1 Training data

Training data was collected and analysed for 7 Land Use/Land Cover (LU/LC) classes: forest, water, build-up area, permanent grasslands, arable land, permanent crops, and karst pastures. Validation was based on analysis of NDVI (Normalized Difference Vegetation Index) signals for 2020 and 2022 and manual visual control based on aerial imagery. Data was provided by the State Geodetic Administration,

Croatian State Forest Enterprise, and PAAFRD, and was iteratively selected based on iteration results. The initial dataset was filtered by area and shape with the goal of increasing homogeneity in the LU/LC sample. Objects narrower than 10m were excluded from the sample.

Special attention was given to the training data for agricultural parcels, as it was crucial not only for CROLIS, but also for training the model for ARKOD+, where borders of each agricultural parcel had to follow visible parcel borders on aerial imagery. For this case, PAAFRD controls

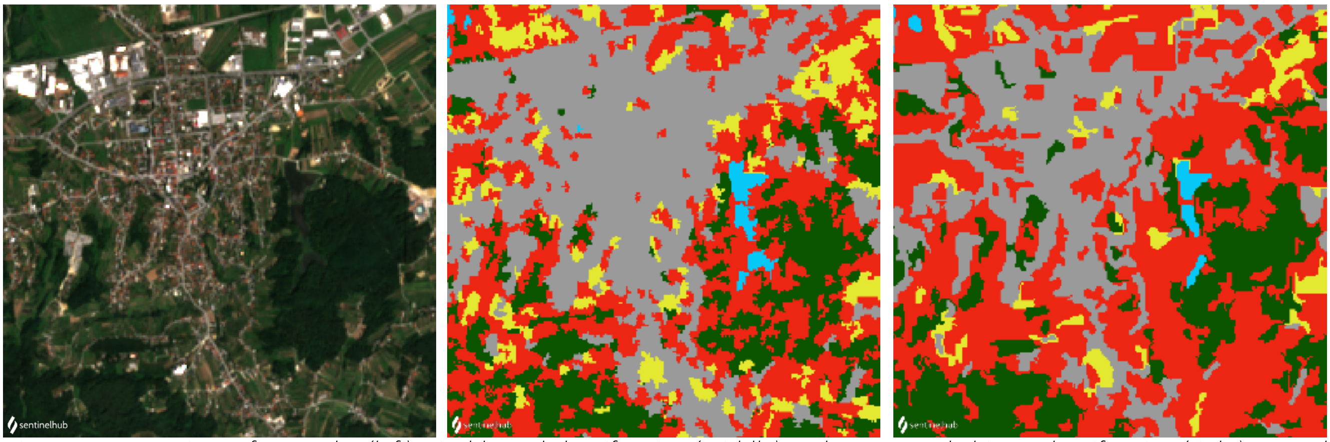

Figure 1. Comparison of true color (left), pixel-based classification (middle) and segmented objects classification (right) around city of Ivanec for 2022.

data was used. Out of 22,410 agricultural parcels, 5,375 (24%) were marked as acceptable. Additionally, 587 parcels were added after editing of quality assurance control made by the executor making a total of 5,962 samples distributed over Croatia. Non-agricultural samples were used as “negative” training data.

For validation purposes, collaboration with the Croatian State Forest Enterprise was established, and with the usage of drones, aerial imagery was created and used to validate land cover changes on a small sample.

2.2 Processing

2.2.1 Land-use / land-cover classification

We compared the machine-learning (pixel-based LightGBM) approach, as showcased in blog-post series (Lubej 2018) with the object-based model.

We first tested LU/LC prediction per pixel, which resulted in substantial salt-and-pepper noise. Sieving was applied to smooth the prediction, but this led to artefacts, primarily seen on linear features. This issue was partly addressed by applying coregistration to the temporal stack of images, but the results still contained patches of obvious misclassifications. As an alternative, an object-based approach was attempted. We applied Felzenszwalb Segmentation (Felzenszwalb and Huttenlocher 2004) to the maxNDVI image for the whole year (separately for 2020 and 2022), segmenting areas into objects. NDVI signals of pixels within objects were averaged, and prediction was run for each object. The results proved more suitable for comparison

between years, and therefore this method was selected for the pilot (Figure 1).

In the next iteration, segmentation into objects was done by considering quarterly values of NDVI, Normalized Difference Water Index and Normalized Difference Built-up Index to reduce influence of changes within objects during the year.

Classification for the whole country was made in 2 iterations. In the first iteration, it was identified that the model could benefit from additional samples from agricultural areas, which were then added in the second iteration. Although we saw significant improvements in the second iteration when comparing confusion matrices, interpretation should be done with caution, as we observed that the performance on the validation set may not fully reflect the predictive capabilities across the entire dataset. This discrepancy is likely attributable to the high quality ("clean") nature of the reference data, which comprises samples exclusively representing a single land cover type.

It is noteworthy that our analysis is conducted by using Sentinel-2 imagery, where significant signal mixing can occur. Notably, a single 10m x 10m pixel often captures data from multiple land cover types simultaneously, influenced by numerous factors that determine the predominant signal within that pixel.

Land cover changes were detected by intersecting LU/LC classes from 2020 and 2022 and retaining results where predicted classes changed. Changes were retained only if pseudo-probability for the predicted group in each year

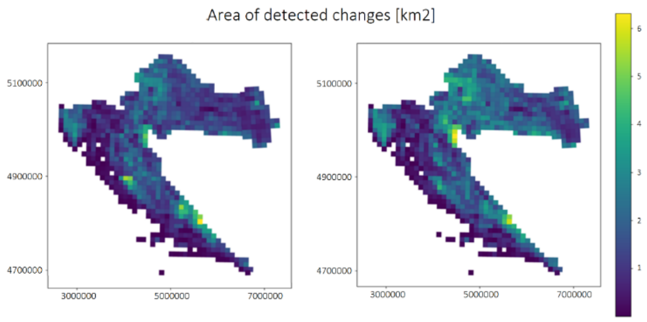

Figure 2. LU/LC changes between 2020 and 2022. First iteration left, second iteration on the right.

was significant (above 0.98), and change was larger than 4 Sentinel-2 pixels (Figure 2).

2.2.2 Field delineation of agricultural parcels

The eo-grow framework (Batič et al. 2023) was used to tackle the issues of processing scalability, by enabling coordination of clusters to run the EO workflows over large areas.

Delineation of agricultural fields was done on normalized aerial imagery with 0.5m resolution made in 2022 and 2021, which included spectral bands for red (R), green (G), blue (B), and near-infrared (NIR). The area of the country was divided into 230,000 tiles with dimensions 1100x1100 pixels and an overlap of 100 pixels.

We rasterized the positive polygons (training data for agricultural areas) to create extent and boundary masks. As there was more negative reference data (in terms of total area), to avoid overrepresentation of the negative reference pixels, we computed the number of "negative" and "positive" pixels in each of the tiles and then down-sampled the negative pixels in a way that the ratio between the two is roughly equalized.

Different experiments were done: training models with different datasets (changing ratio of negative vs positive samples), changing learning rate (from 0.001 to 0.0001), precision (from 16 to 32), and epochs (from 75 to 150).

The evaluation of the models was done on 1200 locations distributed across the country. The best model achieved IoU 0.82 and accuracy 0.93 on the evaluation dataset (Figure 3).

The model was run on the entirety of Croatia, and results were simplified in postprocessing, where holes and parcels with an area of less than 100m2 removed.

2.2.3 Trend of agricultural activities on delineated parcels

On an area of 24x24km within Koprivničko-križevačka region, Sentinel-2 and PlanetScope Fusion NDVI signals were extracted for time range 2018–2023 for each parcel. Bare-soil and mowing markers were then calculated to monitor agricultural trends. The marker records an observation on the signal, and contains both the nature of the observation and the time it was manifested on the parcel (Devos et al. 2021). Out of 24,145 parcels, 27% (6,491) were monitored only with PlanetScope Fusion due to Sentinel's resolution limitation. 97 parcels were too small to contain even a single PlanetScope Fusion pixel (3x3m), and therefore were not analysed. 17,557 (73%) parcels had signals from both sources. A commercial solution, provided by executor, Sinergise, was used for marker calculation (Sinergise 2024).

Bare-soil marker did not show significant trend changes with around 8,000–11,000 parcels having at least 1 bare-soil observation through the year.

Differences in marker results were noticed between PlanetScope Fusion and Sentinel-2 on parcels with heterogenous crops. The PlanetScope Fusion bare-soil marker was more sensitive to partial changes within parcels.

3 Results

Field delineation resulted in 2,848,790 agricultural parcels with a total area of 1,822,550 ha.

It was noticeable that higher accuracy results were achieved in continental parts of Croatia. In coastal areas and islands, initial results would require manual editing (as expected) to precisely follow borders of agricultural parcels.

Karst pasture borders represent a challenge, as this kind of land use often does not have borders that are clearly visible from aerial imagery. Karst pastures are often declared over forested or bare-soil areas with sparse vegetation, with no clear contrast between two neighbouring fields - making it hard to determine field boundaries from aerial imagery alone.

Preliminary LU/LC classification results proved that cautiously-selected training data can significantly improve results, and with an established framework and methodology, iterations could be calculated, often providing a fast feedback loop important for improvement of ground truth.

Lastly, the monitoring of land-use changes and maintenance of the ARKOD+ layer (delineated parcels) is possible with markers which could be used to prove agricultural use of land, or as a potential change-detector (homogeneity, forest probability etc.). However, the 10m resolution of Sentinel-2 creates an issue for small or narrow parcels which don't contain at least 1 whole Sentinel-2 pixel.

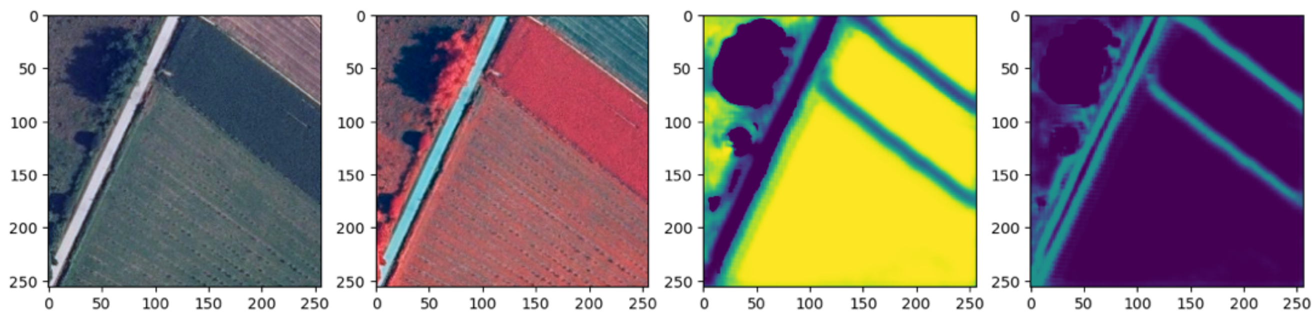

Figure 3. Training the delineation model. From left to right; aerial imagery true color, aerial imagery NIR, predicted parcel extent, predicted parcel boundary.

Marker analysis revealed a negative trend in the number of mowed parcels. On both Sentinel-2 and PlanetScope Fusion, the number of mowed parcels decreased. On average, approximately 8,000 of the 17,557 parcels were mowed in 2018, decreasing to approximately 5,174–7,863 mowed parcels in 2023, depending on the source.

4 Discussion

The area of delineated parcels is smaller than the PAAFRD estimate of 2,695,037 ha (Narodne novine 2013), likely due to results for karst pastures, for which visible boundaries are hard or even impossible to precisely detect with the field delineation model. Results seem to be mainly suitable for arable land, with room for improvement for permanent crops and grasslands. A larger and more diverse sample of training data is needed to improve the quality, especially over coastal areas. The need for larger and more diverse sample data is also true for LU/LC classification. Although initial results look promising, it seems there is still room for improvement in segregation between arable land and permanent crops in coastal areas where we did not have significant distribution of training data.

As for agricultural monitoring purposes, PlanetScope Fusion showcased as advantageous compared to Sentinel-2 on small and narrow parcels. On larger parcels, lower revisit time did not significantly improve results. As changes are expected to occur on delineated parcels over the years, it seems sensible to monitor parcels homogeneity as an indicator of potential parcel border change.

5 Conclusions

The pilot study resulted in important guidelines for the collection and preparation of training data, which proved highly significant. Preparation of the training data in advance will accelerate the process of carrying out necessary activities to reach the objectives of the CROLIS project. This acceleration may ultimately support the establishment of an information system to realize these goals.

Also, the results that were obtained through the delivery of the pilot project determined the guidelines surrounding the methodology that will be applied during the establishment of the future information system. The initial results of ARKOD + parcels through the automatic classification method confirm significant progress in the creation of LU agricultural land, and for the purposes of achieving the goals derived from the CROLIS project.

6 References

Batič, M., Lukšič, Ž., Milčinski, G., 2023. eo-grow-Earth Observation framework for scaled-up processing in Python. In EGU General Assembly Conference Abstracts (pp. EGU-16802).

Devos, W., Sima, A. and Milenov, P., 2021. Conceptual basis of checks by monitoring. JRC127678

Felzenszwalb, P. F., Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. International journal of computer vision 59, 167-181.

Lubej, M., 2018. Land Cover Classification with eo-learn: Part 1, https://medium .com/sentinel-hub/land-cover-classification-with-eo-learn-part-1-2471e8098195. (Accessed 30 April, 2024)

Narodne novine, 2013. Strategija upravljanja i raspolaganja imovinom u vlasništvu Republike Hrvatske za razdoblje od 2013. do 2017. Godine, https://narodne-novine.nn.hr/clanci/sluzbeni/2013_06_76_1532.html. (Accessed 30 April, 2024)

Sinergise, 2024. Area monitoring - markers, https://area-monitoring.sinergise.com /docs/markers. (Accessed 30 April, 2024)