

CLASSIFICATION OF SELECTED CORINE CLASSES USING SENTINEL-2 DATA

L. Stara¹, L. Halounova¹

¹ Czech Technical University in Prague, Department of Geomatics, Czech Republic, (lucie.stara, lena.halounova)@fsv.cvut.cz

Commission III, WG III/1

KEY WORDS: CORINE, Sentinel-2, arable land, pastures, natural grassland, land cover, land use, Random Forest

ABSTRACT:

This paper addresses the classification of CORINE classes. Three land use classes (arable land, pastures, and natural grassland) report similar spectral responses which make it challenging to separate. Therefore, we adopted a multitemporal and multispectral approach using Sentinel-2 satellite imagery in combination with the NDVI vegetation index, Haralick's textural measures, and topographic information. The workflow identifies a methodology for combining various groups of input data (optical, NDVI, textural, topographic) and explores the suitable use of the Random Forest classifier for the task. The classification was carried out in three different European locations. The results present a strong separation of arable land (F1 score over 96%) from the other two classes. Pastures and natural grassland classes achieved F1 in the range of 76% to almost 85% in both cases. In conclusion, we discuss the suitability of the CORINE database for such classification problems.

1. INTRODUCTION

The paneuropean land use (LU) and land cover (LC) are captured in Copernicus' CORINE database and it is widely used for classification in Remote Sensing tasks. The database includes five main groups (artificial surfaces, agricultural areas, etc.) which contain 44 classes at the most detailed level. The database covers LU/LC of 39 countries, and every country is responsible for its thematic content from the CORINE nomenclature. The database is created on a national level by visual interpretation.

The classification of these classes becomes challenging towards the detail, particularly for classes of similar LC type (e.g. vegetation). In satellite imagery, classes of similar LC types have spectrally similar characteristics. When it comes to the classification into LU classes, the task becomes more difficult to perform.

Succeeding the Geo-Harmonizer (GH) project (OpenDataScience, n.d.), this study addresses the classification of classes from the CORINE Land Cover database (CLC2018), namely permanently irrigated arable land (212), pastures (231), and natural grassland (321). Arable land and pastures belong to the agricultural areas, while natural grassland is characterized by minimal or no human disturbance. In addition, the pastures are mainly used for grazing or harvesting, whereas natural grassland is located on rough terrain occasionally with rocky areas.

The basis for the discrimination of this phenomenon is the use of multiple inputs. Concerning comparable research, multispectral data revealed features at different wavelengths, whereas multitemporal data captured different LC types at various stages of phenological development (Zhai et al., 2018, Müller et al., 2015, Kyere et al., 2019). Furthermore, in combination with time-series of satellite imagery, the optical data also can be a source for various derivatives that enhance the image, such as vegetation indices or textural measures (McInnes et al., 2015, Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture, 2012).

Additionally, the results depend not only on the input data but also on the classification method. Machine learning methods have come to the front of classification tools in recent years. Among them, the Random Forest (RF) became a popular method. As far as the benefits, RF delivers high accuracy results and handles a large number of features (Breiman, 2001, Gislason et al., 2006). Moreover, it also performs efficiently with multi-source data (e.g., optical, topographic), as demonstrated by the study of combining the MODIS NDVI time-series with textural and topographic data (Melville et al., 2018, Ali et al., 2016).

The presented workflow focuses on the classification of declared LC types using freely available data (Sentinel-2, CORINE), open source software (QGIS, Spyder), and a machine learning approach (Random Forest) according to the GH framework. The objective is to investigate the possibility of mutual separation of the selected CORINE classes. We want to evaluate these classes in different locations. Also, it is desirable to identify the important features and a suitable methodology for the issue presented. The aim is to point out the problem of classifying land cover classes defined by land use and emphasize possible limitations of the reference data.

2. METHODOLOGY

2.1 Study area

Three study areas were chosen to compare the results. We observed the distribution of the three selected classes in Europe. The location of all three classes within a single Sentinel-2 tile was important for the selection. In addition, the distribution of the tiles as well as the difference in climate and terrain was crucial for the selection. The combination of the three classes is located rather in the south parts of Europe. Therefore, areas in Spain, North Macedonia, and Turkey were selected (Figure 1).

The study area in Spain is located in the west of the country. The area is devoted mainly to agriculture and the altitudes come up to 1500 m a.s.l. The observed classes are distributed evenly over the area with exception of the arable land, which is located



Figure 1. Study areas.

primarily in the south. The distribution of the classes in the location is displayed in Figure 2.

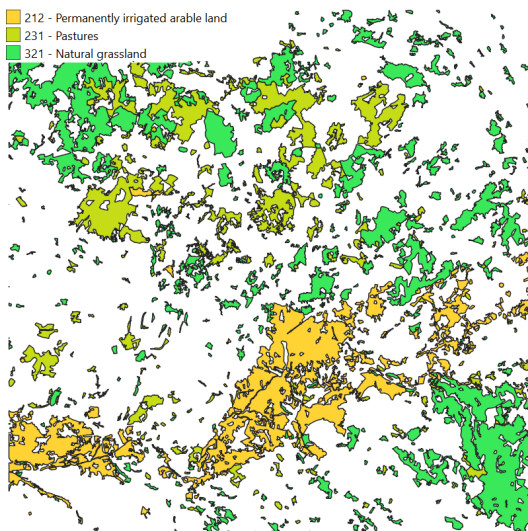


Figure 2. Spain: distribution of LC classes.

In comparison to Spain, the location in North Macedonia is more mountainous with altitudes around 2500 m a.s.l. Again, there is an area dedicated to agriculture (in the west of the location). The rest of the classes is distributed all over the scene as shown in Figure 3.

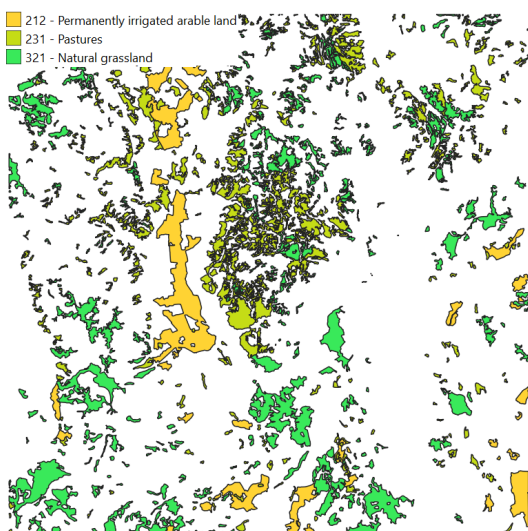


Figure 3. North Macedonia: distribution of LC classes.

In Turkey, a southeast location of the country was chosen for a

study area. The highest altitude is around 2000 m a.s.l. This area is dominated by arable land and natural grasslands. The pastures are mainly located in the north of the location (Figure 4).

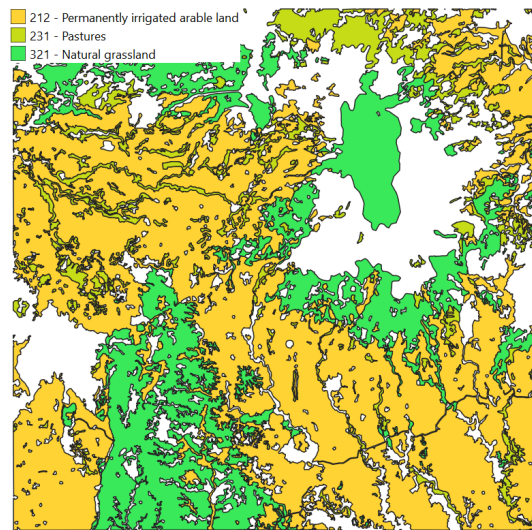


Figure 4. Turkey: distribution of LC classes.

As for the climate, the locations in Spain and Turkey can be categorized as the Mediterranean, where the summer is hot and winter is mild and wet. On the other hand, the climate in North Macedonia is continental, with mild summer and cold winter.

2.2 Input data

Multiple input features were used to improve the visual interpretation of the classification process. Nine Sentinel-2 bands of 20 m resolution were used (B2, B3, B4, B5, B6, B7, B8a, B11, B12). The NDVI vegetation index was computed based on red and infrared bands (B4, B8a). We also included selected Haralick's texture measures, namely angular second moment (ASM), correlation, homogeneity, and entropy. These were calculated from the B12 Sentinel-2 band. Apart from spectral information, topographic information (DEM, slope) was included using EU-DEM v1.1 product.

The data was collected from five different dates within a growing season of 2018 (Spain, Turkey) and 2019 (North Macedonia) (Figure 1). The year 2018 was selected as it corresponds to the year of CORINE publication. Unfortunately, the scenes of North Macedonia were cloudy in 2018, therefore, we selected 2019 satellite data for this location. Also, as the climate in the locations varies, the growing season is different as well. Five scenes between March and July were selected in Spain and Turkey. In North Macedonia, the scenes capture vegetation from June to October since the growing season is later than in Spain and Turkey. Table 2 displays all input variables.

Spain (2018)	North Macedonia (2019)	Turkey (2018)
28.3.	8.6.	19.3.
17.4.	3.7.	23.4.
17.5.	7.8.	23.5.
16.6.	16.9.	7.6.
16.7.	16.10.	12.7.

Table 1. Multitemporal data - selected scenes.

Category	Features	Total number of features (5 scenes)
Sentinel-2	B2, B3, B4, B5, B6, B7, B8a, B11, B12	45
NDVI	NDVI	5
Textural	ASM, correlation, homogeneity, entropy	20
Topographic	DEM, slope	2

Table 2. Summary of input features.

2.3 Reference data

The CLC reference data was processed in this experiment. The processing included cloud masking and buffering of the training data polygons. As the reference data are generalized, the transition between two adjacent classes may be inaccurate. In order to overcome this issue, the inner buffer of two pixels (40 m) was applied for reference data. This step is visualized in Figure 5. Subsequently, the reference data was transformed into a generalized point layer with 500 m spacing between the points.



(a)



(b)

Figure 5. Editing of CORINE data: a) original polygons b) application of inner buffer.

During the training phase, some removals from reference data were performed. As a consequence of the minimum mapping

unit (MMU) of CORINE¹, particular errors may appear, e.g., vegetation in urban areas. As a solution, pixels were removed based on their NDVI value. Also, a statistical examination was carried out to remove outliers.

In the course of the classification process the number of training points was adjusted as well. At first, the number of points per class was stratified based on the original data. However, the amount of data in every class was imbalanced which can be problematic. Subsequently, the number of points was equalized for every class.

2.4 Classification

Concerning the classification method, the results were improved by identifying the most important features. Not every feature is important and even with a smaller amount of data, the results were able to improve. Also, optimal parameters were obtained by hyperparameter tuning. This process took into consideration especially the number of RF trees in the forest and the maximum depth of the trees.

The reference data were divided randomly into training and testing sets with the 80:20 ratio. To avoid overtraining, the results were validated by the cross-validation method. The results were evaluated using precision and recall metrics. We also included the F1 score which is the harmonic mean of precision and recall. These metrics are more suitable for per class evaluation than overall accuracy. Furthermore, the balance between training and testing sets was observed to prevent the model from overtraining.

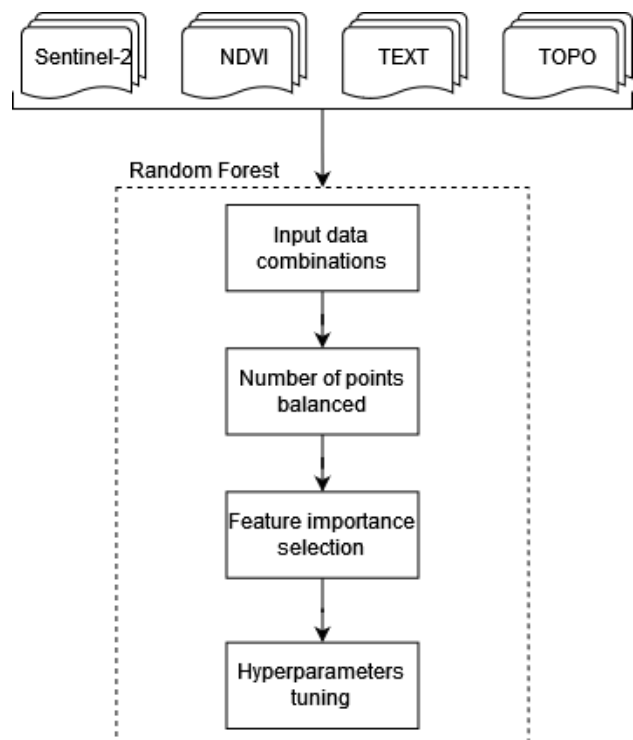


Figure 6. Classification workflow.

The classification was repeated by taking into account various aspects as displayed in Figure 6. In the first stage, various groups of features were taken into account for the classification

¹ 25 ha for areal objects, 100 m for linear body

(optical, NDVI, textural, topographic) and results were evaluated. The best combination was selected based on the mean F1 value. When the best combination of data was selected (e.g., optical+NDVI+topographic, excluding textural), the classification was carried out for a balanced number of points for each class. Subsequently, important features were selected and we performed hyperparameter tuning using these features.

With respect to the Geo-Harmonizer framework, the entire workflow was carried out using freely available data (Sentinel-2, EU-DEM) and open source software and tools (QGIS, Spyder).

3. RESULTS

In this section, the results will be presented by location. Firstly, the most promising combination of features for every location is introduced. Subsequently, we applied the same workflow in every case by balancing the points in every class, selecting the important features, and tuning the optimal parameters of the model.

3.1 Spain

In Spain, the most promising results were obtained for the combination of optical, NDVI, textural, and topographic features. The classification results can be found in Table 3. Permanently irrigated arable land (212) achieved almost 99% F1 score, which makes its classification almost flawless. The F1 values for pastures (231) and natural grassland (321) are lower, however, the values are almost 83% and over 81% respectively. It can be also noted, that the values between precision and recall are approximately balanced which is a sign of a well-trained model.

	recall (%)	precision (%)	F1 (%)
212	98,25	99,24	98,74
231	83,46	82,02	82,73
321	81,20	81,82	81,51

Table 3. Spain: Evaluation metrics (opt+NDVI+text+topo).

Figure 7 displays the 10 most important features in this location. Firstly, the importance of DEM over the rest of the features was significant. Secondly, NDVI channels from different months repeat among the 10 most important features (4-April, 3-March, 5-May). Finally, various Sentinel-2 bands appear, although the bands from July (7) seem to be more important than the bands from other months.

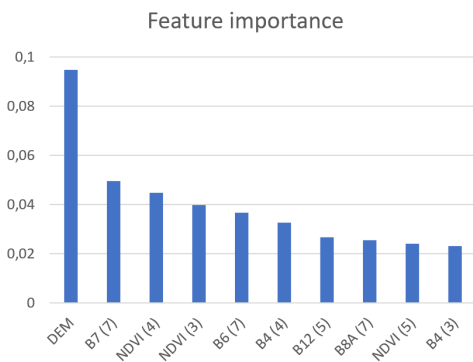


Figure 7. Spain: Feature importance (top 10).

3.2 North Macedonia

We selected the combination of optical, NDVI, and topographic features in this location. Table 4 presents the final results. As in the previous location, the arable land was classified most successfully (F1 score 98,51%). The values for pastures and grassland were very similar - almost 78% and 77%, respectively. The unbalanced values between precision and recall can be also a sign of error in reference data.

	recall (%)	precision (%)	F1 (%)
212	99,25	97,78	98,51
231	80,83	74,65	77,62
321	72,93	80,83	76,68

Table 4. North Macedonia: Evaluation metrics (opt+NDVI+topo).

Considering the feature importance (Figure ??), both topographic features were significantly more important than the rest of the features, especially slope. Among other features, the Sentinel-2 band from June (6) appears to be the most important.

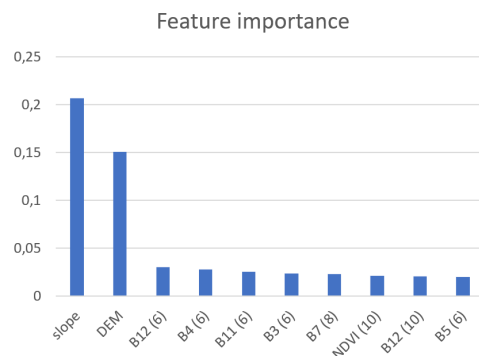


Figure 8. North Macedonia: Feature importance (top 10).

3.3 Turkey

As in North Macedonia, the best results were obtained by combining optical, NDVI, and topographic data. As can be seen in Table 5, the arable land achieved the highest F1 value (96,03%). There was a small confusion with other classes. As for pastures and natural grassland, the F1 scores were almost 85% for both. These classes partly classified into each other.

	recall (%)	precision (%)	F1 (%)
212	96,30	95,77	96,03
231	89,18	81,17	84,99
321	80,00	89,30	84,39

Table 5. Turkey: Evaluation metrics (opt+NDVI+topo).

In contrast to previous locations, the topographic features do not appear among the ten most important ones. According to Figure 5 the April Sentinel-2 bands play the main role as well as NDVI of the same month. Furthermore, visible bands (B2, B4, B3), as well as SWIR (B12), are the most important.

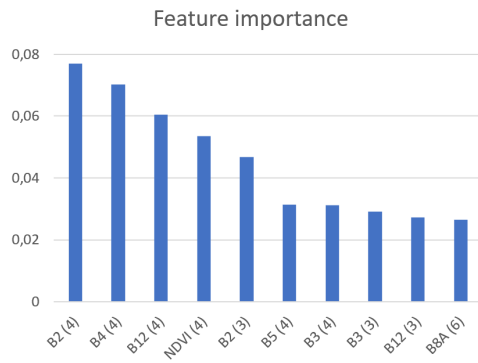


Figure 9. Turkey: Feature importance (top 10).

4. CONCLUSIONS

The task was carried out in three different locations to enable the comparison of the results. Some similarities were observed among all locations despite the differences in climate conditions and terrain.

Firstly, the classification of the arable land resulted in the range of 96% to almost 99% F1 score. The separation of arable land from pastures and natural grassland can be considered successful by this methodology.

The challenge remains in separating pastures and grassland as these classes differ by land use, but their land cover is similar. Concerning pastures, we achieved the F1 score from almost 78% to almost 85%. The natural grassland resulted in almost 77% to over 84%. Also, the values between pastures and grassland in individual locations were similar to each other. This indicates the balanced confusion between these classes. Pastures do not appear more like grassland or vice versa, but the two classes are mutually similar.

Concerning the reference data, the results could be fairly connected to the number of training points. The highest results for pastures and grassland were obtained in Turkey, where the coverage of the observed classes was the largest. While in North Macedonia, where the results were the lowest, the number of points was the smallest as well.

We also investigated important features for every classification. The topographic features were the most significant for the classification in two out of three locations (Spain and North Macedonia). This finding indicates that the elevation and slope are essential features in this task. The arable land usually located in flat terrain could be easily distinguished from the classes located in areas with more complex topography. Moreover, the natural grassland is more likely to occur in rough terrain while the pastures have to be more accessible either for agriculture machinery or the cattle.

Overall, the feature selection confirmed that the multitemporal approach was beneficial. On the other hand, the textural features did not manifest their strength except for Spain. Despite that, none of the textural features appeared among the ten most important features. The textural features may give more promising results with finer resolution of satellite imagery.

The CORINE data declares overall accuracy of $\geq 85\%$ which is not very useful for per class evaluation. According to the

CORINE validation report (Moiret, 2021), the accuracy of pastures and natural grassland is between 80% and 85%. However, based on our study, these values can differ in different locations. In conclusion, some classes in the reference data are more reliable than others as well as some locations give better results. Also, as long as the reference data is not 100% accurate, the classification results are limited by its quality. It is the case of CORINE data since the mapping unit is 25 ha where classes of smaller areas are added to their neighboring classes of areas above 25 ha.

As noted earlier, the CORINE database is formed at a national level. Therefore, the classification of pastures and grassland may differ in various climatic conditions as well as different agricultural approaches. For example, the pastures in Turkey are definitely not the same as the pastures in Macedonia. That is the reason why the database is developed on a national level. On the contrary, this can cause problems in paneuropean interpretation of the product. All these limitations have to be taken into account when we work with the CORINE land cover product.

Future research could be focused exclusively on separating pastures and natural grassland in other locations to form more versatile conclusions.

5. ACKNOWLEDGEMENTS

This work was supported by the Institutional grant CTU in Prague, internal nr. SGS22/047/OHK1/1T/11. The research was processed within the Geo-Harmonizer project (OpenDataScience, n.d.).

REFERENCES

- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. *Journal of Plant Ecology*, 9(6), 649-671. 10.1093/jpe/rtw005.
- Breiman, L., 2001. Random Forests. *Mach. Learn.*, 45(1), 5-32. doi.org/10.1023/A:1010933404324.
- Gislason, P. O., Benediktsson, J. A., Sveinsson, J. R., 2006. Random Forests for land cover classification. *Pattern Recognit. Lett.*, 27(4), 294-300. https://doi.org/10.1016/j.patrec.2005.08.011.
- Kyere, I., Astor, T., Graß, R., Wachendorf, M., 2019. Multi-Temporal Agricultural Land-Cover Mapping Using Single-Year and Multi-Year Models Based on Landsat Imagery and IACS Data. *Agronomy*, 9(6). 10.3390/agronomy9060309.
- McInnes, W. S., Smith, B., McDermid, G. J., 2015. Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 8(4), 1395-1403. 10.1109/JSTARS.2015.2416713.
- Melville, B., Lucieer, A., Aryal, J., 2018. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *International Journal of Applied Earth Observation and Geoinformation*, 66, 46-55. https://doi.org/10.1016/j.jag.2017.11.006.

Moiret, A., 2021. Copernicus land monitoring services clc2018 validation report. Technical report, Copernicus.

Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A. J., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sens. Environ.*, 156, 490-499. <https://doi.org/10.1016/j.rse.2014.10.014>.

OpenDataScience, n.d. Geo-harmonizer: Eu-wide automated mapping system for harmonization of open data based on foss4g and machine learning. OpenDataScience <https://opendatascience.eu/geoharmonizer-project/> (14 February 2022).

Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture, 2012. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93-107.

Zhai, Y., Qu, Z., Hao, L., 2018. Land Cover Classification Using Integrated Spectral, Temporal, and Spatial Features Derived from Remotely Sensed Images. *Remote Sens.*, 10(3). <https://doi.org/10.3390/rs10030383>.