ANNUAL CROP CLASSIFICATION EXPERIMENTS IN PORTUGAL USING SENTINEL-2

Pedro Benevides[1], Hugo Costa[1,2], Francisco D. Moreira[1], Daniel Moraes[1,2], Mário Caetano[1,2]

[1]Direção-Geral do Território, Rua da Artilharia Um, 107, 1099-052 Lisboa, Portugal

[2]NOVA Information Management School (NOVA IMS), Universidade Nova Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

This is the accepted version of the conference paper published by IEEE at *IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium: Proceedings:*

How to cite: Benevides, P., Costa, H., Moreira, F. D., Moraes, D., & Caetano, M. (2021). Annual Crop Classification Experiments in Portugal Using Sentinel-2. In IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium: Proceedings (pp. 5838-5841). IEEE. https://doi.org/10.1109/IGARSS47720.2021.9555009

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

ANNUAL CROP CLASSIFICATION EXPERIMENTS IN PORTUGAL USING SENTINEL-2

Pedro Benevides[1], Hugo Costa[1,2], Francisco D. Moreira[1], Daniel Moraes[1,2], Mário Caetano[1,2]

[1]Direção-Geral do Território, Rua da Artilharia Um, 107, 1099-052 Lisboa, Portugal
[2]NOVA Information Management School (NOVA IMS), Universidade Nova Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

ABSTRACT

This paper presents an experimental crop classification of the 10 most abundant annual crop types in Portugal, using a study area located in Alentejo region. This region has great diversity of land uses as well as multiple crop types. Sentinel-2 2018 intra-annual time-series imagery is considered in the experiment. The Portuguese Land Parcel Identification System (LPIS) is used to extract automatic training samples. LPIS information is automatically processed with the help of auxiliary datasets to filter out crop areas more likely to have been mislabeled. Classification is obtained using random forest. Validation is performed using an independent dataset also based on LPIS. A global accuracy of 76% is obtained. The novelty of the methodology here presented shows that LPIS can be used together with auxiliary data for crop type mapping, helping to characterize the agriculture land diversity in Portugal.

Index Terms— Sentinel-2, crop mapping, random forest, time-series, Portugal

1. INTRODUCTION

The availability of free and more frequent satellite data brought by the Sentinel-2 program, combined with the utilization of supervised classification techniques, promotes new approaches for land cover classification. In particular, the monitoring of phenological changes that commonly characterize annual crops became feasible. Presently, there are some studies that rely on automatic land cover classification based on satellite data at an operational and national level [1-4]. Studies using crop classification can also be found either combined with more common classes [5] or performing crop mapping alone [6,7]. Yet, gathering training data still remains a challenge and its quality may have a significant effect on the land cover map accuracy [1,2]. Training data collection is usually very demanding on resources, requiring a team of specialized technicians to manually collect land cover information, being also a time-consuming process. Automatic training extraction can overcome this problem, allowing the selection of a large number of samples in a much shorter period.

Land Parcel Identification System (LPIS) datasets have been used to execute this automatic training extraction in several European countries. Most of these studies consider small test areas or a reduced number of crop types for classification. However, even using a quality reference database, problems may arise since a large number of undesired samples may be selected due to generalization in reference data, i.e. points located in bare soil or natural grassland inside orchards, roads, water reservoirs or trees inside agricultural crops. To minimize the samples mislabelling, a filtering process can be applied to the reference datasets with the help of auxiliary datasets, such as burnt areas, clear cuts, imperviousness, vegetation density information, among others.

This paper presents an experimental method to produce a crop type map at regional level using LPIS as a reference training dataset. The main novelty is the application of a pre-process filtering using a set of auxiliary datasets, prior to the automatic sample extraction from LPIS. Additionally, manual training extraction for some classes (e.g. natural herbaceous vegetation, bare soil, shrubland) is used to reduce the misclassification of crop types. Multi-temporal Sentinel-2 imagery dataset is preprocessed to be inputted into a random forest supervised classification processor. The study area is located in the south of Portugal. The selection of the study area took into consideration its relatively large dimension at a regional scale (1.7 million ha) and also the existence of large and diverse agricultural areas, including multiple crop types. This is the first experiment to test the feasibility to produce an annual crop map at national level using an automatic methodology. Therefore, the 10 most abundant annual crops types at a country level are considered. The proposed methodology serves as a test to evaluate if the use of LPIS dataset combined with an automatic training extraction and data filtering process is reliable for the production of a crop type map in Portugal.

2. STUDY AREA AND DATA

The study area is located in the south of Portugal and it includes a large part of the Alentejo region. It was obtained from a landscape unit subdivision of the Portuguese Continental area that groups regions for their land characteristics [8]. This area is mostly characterized by a flat topography, Mediterranean climate and diverse land cover occupations, including several types of agricultural areas such as pastures, *montado* (an agro-forestry system), natural grassland and several crop types. The Alentejo region also includes one of the largest artificial lakes of Europe, arising with the construction of the Alqueva dam in 2002. Since then, the landscape has been significantly affected causing land cover and land use transformation. In particular, the construction of irrigated systems supplied by Alqueva has increased the spring/summer crop production. The area has a total of 1 680 967 ha, 50% of which is related to agriculture according to National Land Cover Land Use Map (COS) of 2018.

Multi-temporal Sentinel-2 imagery is used in this work. The chosen period includes images acquired between October 2017 and September 2018, corresponding to the agricultural year of 2018 in Portugal. A total of 6 Sentinel-2 granules located over the study area are used. Level-2 products are downloaded from Theia Land Data Centre. Each image is composed by 10 bands, being the ones with larger pixel size disaggregated to 10 m size. To avoid missing data resulting from the atmospheric corrections, images are preprocessed using the pixel median during a month, resulting in 12 monthly composite images. Additionally, 5 spectral indices are processed for each composite following a previous study [9]. A set of spectral-temporal metrics are also computed, using all the bands from the monthly composites and the 5 spectral indices, resulting in 7 quantile metrics. A digital terrain model is also used. To summarize, a total of 286 input variables are gathered for the supervised classification.

3. METHODS

Even though we are interested in an agricultural map considering only crop types from LPIS, non-agriculture land cover needs also to be considered. This is a fundamental step, because other land cover classes can be frequently observed inside agriculture parcels (artificial structures, water bodies, tree species, shrubland or herbaceous vegetation). Therefore, training information regarding non-agriculture classes was extracted from the reference dataset COS. COS is the Portuguese reference map for land use land cover, being available since 1995 and having a minimum mapping unit of 1 ha. Crop type information regarding agriculture is retrieved from the Portuguese official LPIS, called Sistema de Informação Parcelar (SIP). SIP has higher spatial detail, including parcels smaller than 1000 m^2. Nevertheless, the training information is sampled to produce a map based on Sentinel-2 spatial resolution (100 m^2). Therefore, the filtering process for automatic training extraction becomes a crucial step to minimize the selection of mislabelled pixels. The use of auxiliary data is also considered for the filtering process. In particular, auxiliary datasets that characterize the vegetation type (leaf type, density or its absence) and soil proprieties (imperviousness) are used providing training data with the spectral information suitable to distinguish between crops and other land cover classes. Auxiliary datasets related to burnt areas and clear-cuts are also used to detect abrupt changes in the terrain [8,9]. Figure 1 presents a diagram of methodology used in this work.

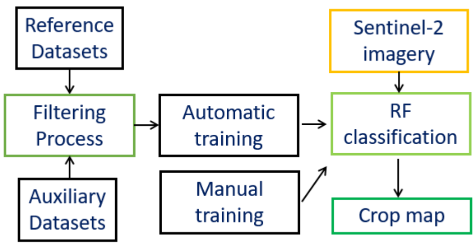

Fig. 1: Diagram of the methodology.

After the filtering process, training sample units are obtained from automatic random extraction of these areas, stratified by the land cover and crop type classes. The 10 most abundant crop types in Portugal are listed in Table 1. The collected samples are used to retrieve the spectral information from the Sentinel-2 intra-annual dataset. A large number of training samples is retrieved, up to 6000 per class. Initial classification tests revealed that some land cover classes with automatic training extraction did not produce the expected results, as they were being confused with some crop types. Consequently, for some classes (i.e. natural grassland, shrubland and bare rock) manual samples were collected by photo-interpretation of 2018 orthophotos and monthly Sentinel-2 images.

A land cover crop map is produced for the study area, using random forest (RF) algorithm for supervised classification [10]. Since SIP data does not provide information for all the agriculture parcels (only those submitted for financial support to the European Union), the classes in COS that identify annual crops are used to define the map limits. An additional step was performed removing isolated contiguous areas smaller than 5 ha. Validation is performed using an independent dataset extracted from SIP.

Table 1. Characterization of map crop type classes.

Crop class	Growing season
Oat	autumn/winter
Wheat	autumn/winter
Barley	autumn/winter
Ryegrass	autumn/winter
Triticale	autumn/winter
Rye	autumn/winter
Maize	spring/summer
Rice	spring/summer
Tomato	spring/summer
Sunflower	spring/summer

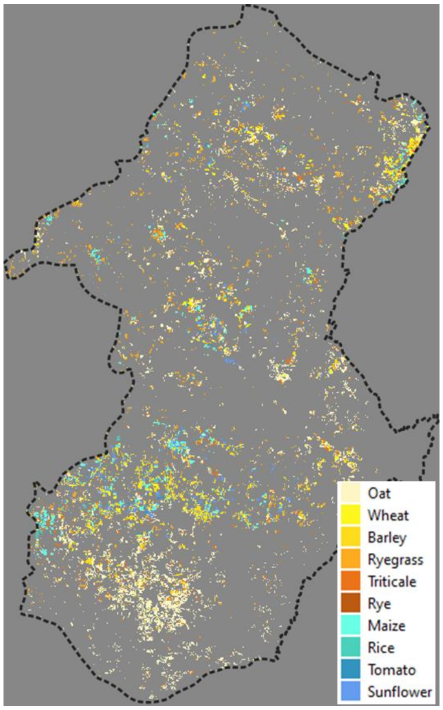

Fig. 2: Annual crop map overview**.**

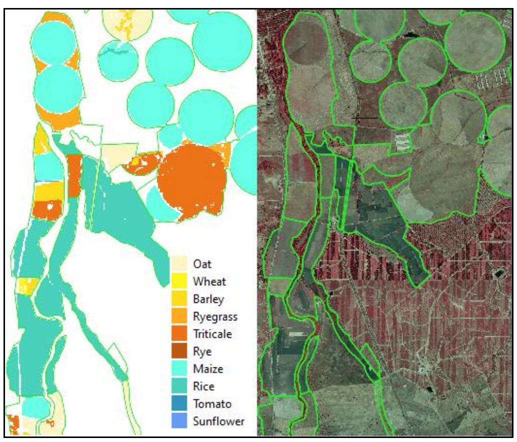

Fig. 3: Crop map close-up and orthophoto map of 2018 in false color. Green line shows COS annual crop classes.

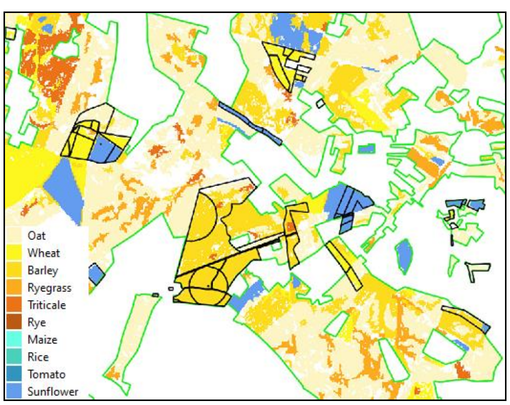

Fig. 4: Crop map close-up with COS agriculture classes.

4. RESULTS AND DISCUSSION

The annual crop map for the study area is presented in Figure 2. A close-up example of the annual crop map is shown in Figure 3. The autumn/winter crops types are colored in yellow to orange tones while spring/summer crops are colored in blue tones (Table 1). Oat is the most abundant crop type and Rye the least abundant, being both autumn/winter crops. The autumn/winter crop types are dominant. Sunflower is the spring/summer annual crop with the highest area, while Rice is the least abundant spring/summer crop type.

The crop map validation was performed using a subset of the SIP parcels, defined as SIP validation, which is obtained by means of photo-interpretation and fieldwork. This subset corresponds to about 5% of the total SIP reference dataset, and it was removed from the automatic training extraction to keep the SIP validation dataset independent. The close-up example in Figure 4 shows the SIP reference dataset parcels along the crop map. It is noticeable that most of these parcels contain only one crop type, in contrast to the COS agriculture parcels, which contain mostly a mix of crop types. An intersection between the crop map and the SIP validation dataset was performed to evaluate the classification accuracy. Only parcels corresponding to the crop types (Table 1) are selected for this intersection. A confusion matrix table comparing the crop map classification with the SIP validation areas was therefore calculated along with the producer's accuracy (PA) and user's accuracy (UA) percentages (Figure 5).

	Oat	**Wheat**	**Barley**	**Ryegrass**	**Triticale**	**Rye**	**Maize**	**Rice**	**Tomato**	**Sunflower**	Total	UA (%)
Oat	842	116	133	28	213		0			2	1334	**63**
Wheat	55	538	26	2	76		0			1	698	**77**
Barley	37	104	687	0	9		2		0	3	842	**82**
Ryegrass	91	11	3	307	7		3			0	422	**73**
Triticale	81	51	41	6	178		0			0	358	**50**
Rye											0	-
Maize		1	7		3		525			0	536	**98**
Rice		0					1			0	1	**0**
Tomato			0						16	8	25	**67**
Sunflower	4	5	5	0	0		17			483	515	**94**
Total	1109	826	902	343	486	0	548	0	16	497		
PA(%)	**76**	**65**	**76**	**89**	**37**	-	**96**	-	**99**	**97**		

Fig.5: Confusion matrix table for the intersected area (ha) between crop map (lines) and SIP validation (columns).

The PA highest accuracies are obtained for spring/summer crops, with values close to 100%, being Tomato the highest. The UA are also high for the spring/summer crops, with the exception of Tomato with 67% and Rice that has all the classified parcels outside the correspondent parcels in SIP. Indeed, SIP validation dataset did not include any Rice and also Rye parcels for this study area. As for the autumn/winter crops, the best results are obtained for Ryegrass (PA) and Barley (UA) and the worst result for Triticale (PA and UA). The crop map global accuracy is found to be 76%. If one considers only the autumn/winter crops and the spring/summer crops as the main classes, the UA and PA are even higher (98% or higher).

Figure 6 presents all the SIP validation parcels (lines) that are intersected with the crop map (Crop map agreement with SIP) and the remaining SIP parcels (not intersected by the crop map). It is also shown the percentage of the SIP parcels intersected/covered by the crop map. It was found that most of the SIP validation parcels are observed inside the same crop map classification, varying from a minimum of 70% (Oat) and a maximum of 96% (Sunflower). This reinforces that the classification is able to identify quite well the crop type and also to reproduce accurately its spatial distribution. It also shows that a significant percentage of the crop map is located inside SIP validation parcels of other crop types not considered in training: 54% for other autumn/winter crops and 65% for other spring/summer crops. This can indicate that when considering the 10 most abundant crops at a country level, other crops types are omitted. Moreover, some COS agricultural areas are absent of crop type classification (Figures 3 and 4). These can correspond to other land cover types frequently observed in crop areas, and to crops wrongly identified as annual crops in this dataset (grassland is often confused with autumn/spring crops). All these constraints have to be taken in consideration when defining an operational workflow for the crop map production at national level.

	Oat	Wheat	Barley	Ryegrass	Triticale	Rye	Maize	Rice	Tomato	Sunflower	Other autumn/winter	Other spring/summer
Crop map agreement w/ SIP	1109	826	902	343	486		548		16	497	332	535
Remaining SIP (no map)	472	153	144	145	189		22		1	18	283	292
% SIP in Cropmap	70	84	86	70	72		96		92	96	54	65

Fig.6: Crop map area table agreement (ha) with the whole SIP validation parcels by class.

5. CONCLUSIONS

The innovative methodology presented in this paper shows that a crop type abundance map can be generated using a supervised classification in South Portugal. The preprocessing of LPIS dataset with auxiliary information for automatic training extraction combined with manual training for specific classes are crucial steps for achieving an accurate crop map. Nonetheless, the methodology has room for improvement. Specific crop type classes can be used from SIP dataset, instead of using the 10 most abundant crops at country level. Mapping areas outside COS annual crops is also important to identify crops in smaller areas (<1 ha). A more focused approach on other agricultural types, such as permanent or greenhouse crops can give a wider perspective of the local agricultural activity. Nevertheless, the experiment here presented shows that this methodology can be useful for the production of a national annual crop map at an operational level.

6. ACKNOWLEDGEMENTS

This work has been supported by projects IPSTERS (DSAIPA/AI/0100/2018), foRESTER (PCIF/SSI/0102/2017), and SCAPEFIRE (PCIF/MOS/0046/2017), and by Centro de Investigação em Gestão de Informação (MagIC), all funded by the Portuguese Foundation for Science and Technology (FCT). Value-added data processed by CNES for the Theia data centre www.theia-land.fr using Copernicus products. The satellite image pre-processing uses algorithms developed by Theia's Scientific Expertise Centers. SIP validation data was kindly provided by Instituto de Financiamento da Agricultura e Pescas.

7. REFERENCES

[1] J. Inglada et al., “Operational high resolution land cover map production at the country scale using satellite image time series”, *Remote Sensing*, 9(1), 95, 2017.
[2] T. Hermosilla et al., “Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year landsat time series”, *Canadian Journal of Remote Sensing*, *44*(1), 67-87, 2018.
[3] D. Pflugmacher et al., “Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey”, *Remote sensing of environment*, 221, 583-595, 2019.
[4] R. Malinowski, et al., “Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery”, *Remote Sensing*, 12(21), 3523, 2020.
[5] P. Griffiths et al., “Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping”, *Remote sensing of environment*, 220, 135-151, 2019.
[6] F. Vuolo et al., “How much does multi-temporal Sentinel-2 data improve crop type classification?”, *International journal of applied earth observation and geoinformation*, *72*, 122-130, 2018.
[7] P. Defourny et al., “Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world”, *Remote sensing of environment*, 221, 551-568, 2019.
[8] I. Hernandez, P. Benevides, H. Costa and M. Caetano, “Exploring SENTINEL-2 for Land Cover and Crop Mapping in Portugal”, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 83-89, 2020.
[9] H. Costa, P. Benevides, F. Marcelino and M. Caetano, “Introducing automatic satellite image processing into land cover mapping by photo-interpretation of airborne data”, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., 42, 29–34, 2020.
[10] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2019.