PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Assessment of the introduction of spatial stratification and manual training in automatic supervised image classification

Moraes, Daniel, Benevides, Pedro, Costa, Hugo, Moreira, Francisco, Caetano, Mário

Daniel Moraes, Pedro Benevides, Hugo Costa, Francisco D. Moreira, Mário Caetano, "Assessment of the introduction of spatial stratification and manual training in automatic supervised image classification," Proc. SPIE 11863, Earth Resources and Environmental Remote Sensing/GIS Applications XII, 1186311 (12 September 2021); doi: 10.1117/12.2599740

SPIE.

Event: SPIE Remote Sensing, 2021, Online Only

Assessment of the introduction of spatial stratification and manual training in automatic supervised image classification

Daniel Moraes*[ab], Pedro Benevides[a], Hugo Costa[ab], Francisco D. Moreira[a], Mário Caetano[ab]
[a]Direção-Geral do Território, Rua Artilharia Um, 107, 1099-052, Lisbon, Portugal; [b]NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312, Lisbon, Portugal

ABSTRACT

The performance of supervised classification depends on the size and quality of the training data. Multiple studies have used reference datasets to extract training data automatically in an efficient way. However, automatic extraction might be inappropriate for some classes. Furthermore, classes can have distinct spectral characteristics across large areas. Thus, dividing the study area into subregions can be beneficial. This study proposes to assess the impact of the introduction of spatial stratification and manually collected training data on classification performance. Two classifications were conducted with the Random Forest classifier and multi-temporal Sentinel-2 data. The classifications' performance was evaluated by accuracy metrics and visual inspection of the maps. The results indicate that introducing spatial stratification and manual training yielded a higher overall accuracy (66.7%) when compared to the accuracy of a benchmark classification (60.2%) conducted without stratification and with training data collected exclusively by automatic methods. Visual inspection of the maps also revealed some advantages of the novel approach, namely constraining some land cover classes to be present only within specific strata, which avoids commission errors of the class to spread freely across the map. Most of the classification improvements were observed in subregions with specific landscapes and spectral patterns, although these strata represent a small fraction of the study area, which might have contributed to the small increase in accuracy.

Keywords: Supervised classification, Random Forest, Spatial Stratification, Sentinel-2

1. INTRODUCTION

Supervised classification has been vastly used to map and monitor Land Cover and Land Use (LCLU), mainly due to its advantages in comparison with unsupervised methods[1]. The performance of supervised methods is heavily dependent on factors such as the size and quality of the training dataset[2]. Therefore, it is essential to develop strategies to acquire sufficient and good quality training samples.

As collecting training data is a costly task in terms of time and resources, multiple studies used existing reference datasets to automatically extract training data[3,4] in a timely and cost-effective manner. Whilst the automatic extraction allows the collection of a large amount of training sampling units, it might be inadequate to capture the spectral particularities of some classes, resulting in reduced classification accuracy. For instance, in Portugal some oaks occur in different spatial patterns (isolated, sparse and dense wood stands) and with diverse vegetation underneath or between trees (crops, pastures and shrubs). Therefore, automatic extraction often fails to capture the oak trees and the diverse spectral pattern brought by the distinct types of vegetation underneath those trees. As an alternative for classes in which automatic extraction does not yield good results, training data collected by experts through visual interpretation of orthophotomaps can be introduced.

Another aspect that can influence the performance of a classification is the spatial extent of the study area. When examining large areas, spectral patterns of a particular species may exhibit significant variations as a result of different environmental conditions, leading to reduced classification accuracy. In order to overcome such problem and increase the accuracy of image classification, a viable approach is to carry out a stratification of the study area into subregions, thus minimizing the heterogeneity of class spectral patterns and reducing intra-class variability. The stratification may take into account aspects such as landscape, soil and vegetation[5]. However, this approach requires adjusting the training

*moraesd90@gmail.com; phone +351 21 381 96 00; https://www.dgterritorio.gov.pt/

Earth Resources and Environmental Remote Sensing/GIS Applications XII,
edited by Karsten Schulz, Proc. of SPIE Vol. 11863, 1186311 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2599740

process to the local conditions, for instance adopting a specific class nomenclature for each subregion (stratum) and ensuring an adequate training sample size.

In this paper we propose to assess the impact of the introduction of stratification of the study area and use of manually collected training sampling units to replace automatically extracted training data in specific land cover classes. Experiments were conducted in a study area in the North of Portugal, using the Random Forest classifier and multi-temporal Sentinel-2 data, aiming to compare a benchmark classification, performed without spatial stratification and with training data collected solely automatically, with a classification carried out with stratification and combination of automatic and manual training data collection. The comparison considered the accuracy metrics and visual inspection of the maps.

2. STUDY AREA AND DATA

The experiments were conducted in the region of Trás-os-Montes, North of Portugal (Figure 1). The region comprises an area of 11,778 km² and is characterized by mountainous land occupied with rocks, forest, bushes and agriculture in the lower lands.

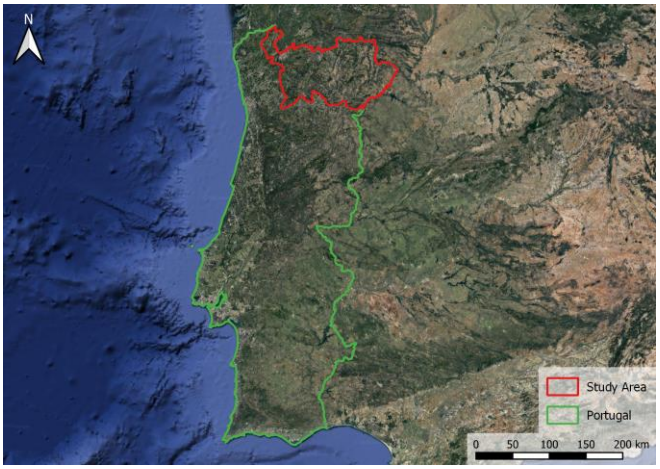

Figure 1. Location of the study area, corresponding to the region of Trás-os-Montes, Portugal.

The datasets used in this study can be grouped into remote sensing data and auxiliary data. The remotely sensed data comprise Sentinel-2 images from the agricultural year of 2018 (October 2017 to September 2018) acquired from the Theia Land Data Centre. In total, 457 images with less than 50% cloud cover were acquired from 6 Sentinel-2 tiles. Pixels with cloud contamination were converted to missing data and monthly composites were produced computing the median value of 10 bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12). Five spectral indices were derived from these bands. Additionally, 7 spectro-temporal metrics (annual mean and quantiles) were calculated for each band and index. Therefore, the final data set used for classification comprises 285 bands.

The auxiliary data consists of multiple datasets used to automatically extract training samples. The national LCLU map for Portugal (COS), the Portuguese Land Parcel Identification System (LPIS) and OpenStreetMap (OSM) road network were used as reference data. Additional datasets, namely the national cartography of burned areas, the Copernicus High Resolution Layers (HRL) products from 2015 and a NDVI-based mask for forest change detection from 2015-2018[6], were also used in order to filter mislabeled pixels, thus refining the quality of the automatically extracted sampling units.

3. METHODS

The methodology consists in comparing two distinct supervised classifications approaches. The first is a benchmark classification, which is conducted for the whole study area without stratification and with the training dataset collected based on a fully automatic process. The second classification is an experimental classification, performed with stratification of the study area and with the training dataset collected through a combination of automatic and manual processes. A different nomenclature is used for the training stage and final map. The training classes are spectral subclasses of the final map nomenclature. Therefore, the final map nomenclature is an aggregation of the training classes

(Table 1) and it is compatible between two classifications. The classifier used in this study is the Random Forest (RF), implemented with the Python Scikit-learn library[7] and parameterized with 500 trees.

3.1 Benchmark classification and automatic training sample extraction

Here, a single supervised classification is carried out for the whole study area, following a similar methodology applied in a prior study[8]. The training samples are collected based on automatic extraction using the auxiliary datasets. Reference data (COS, LPIS and OSM roads) are used to delineate polygons from which sampling units are extracted. Then, filters based on Copernicus HRL, burned areas and the NDVI-based mask are employed to refine this process, preventing the acquisition of sampling units not related to the intended land cover type. The HRL tree cover density product is used to ensure that training samples of forest classes correspond to actual forested areas and the dominant leaf type product is used to reinforce forest species coherence. Similarly, the burned areas and NDVI-based mask help remove dynamic areas, namely burned vegetation and forest cuts. The resulting polygons are subjected to a spatial constrain, which comprehends a negative 40m buffer with the purpose of eliminating mixed pixels that may exist in the boundaries. Polygons with an area smaller than 1000 m² are removed. Next, training sampling units are extracted from within the resulting polygons and up to 6000 sampling units per class are randomly selected. Finally, the RF classification is performed.

Table 1. Map and training class nomenclature. Training class nomenclature has minor variations from benchmark to experimental classification. Last column identifies the training classes from the experimental classification that adopt manual training sample collection. *Collected within Burned Areas in 2017; **collected within forest cuts 2015-2018

Map Class	**Training Class**		**Manual collection (Experimental only)**
	Benchmark	**Experimental**	
Built up	Built up	Built up	
	Industrial	Industrial	X
	Road network	Road network	
Agriculture	Wheat	Wheat	
	Rye	Rye	
	Oat	Oat	
	Ryegrass	Ryegrass	
	Triticale	Triticale	
	Corn	Corn	
	Sunflower	Sunflower	
	Barley	Barley	
	Managed Grasslands	Managed Grasslands	
Natural Grasslands	Natural Grasslands	Agricultural Natural Grassland	X
		Mountain Natural Grassland	X
		Natural Grasslands Burned Areas 2017*	X
Cork and Holm Oak	Cork Oak	Cork Oak	X
	Holm Oak	Holm Oak	X
Eucalyptus	Eucalyptus	Eucalyptus Young Cuts	X
		Eucalyptus Adult	
		Eucalyptus Burned Areas 2017*	X
		Eucalyptus 1 Year Cuts**	X
Other Broadleaf	Other Broadleaf	Other Broadleaf	
Maritime Pine	Maritime Pine	Maritime Pine	
Stone Pine	Stone Pine	Stone Pine	
Other Coniferous	Other Coniferous	Other Coniferous	
Shrubland	Shrubland	Dense Shrubland	X
		Shrubland Burned Areas 2017*	X
Non-vegetated surfaces	Baresoil	Baresoil	
	Bare Rock	Bare Rock	X
Water	Water	Water	

3.2 Experimental classification

The experimental classification consists in a series of classifications conducted for each subregion of the study area. The stratification process divides the study area into 5 subregions (Figure 2) with distinct land cover characteristics, assisted by the national LCLU and burned areas cartographies: 1) Forest and agro-forestry areas dominated by Cork and Holm Oak, 2) Burned Areas in 2017, 3) Burned Areas in 2016, 4) Forest Cuts from 2015 to 2018 and 5) a Complementary subregion comprising the remaining areas (Table 2). Besides the referred cartographies, the NDVI-based mask was also used in the creation of subregion 4, providing areas where vegetation cuts occurred. In terms of classification, each subregion has a particular set of training classes and is classified independently.

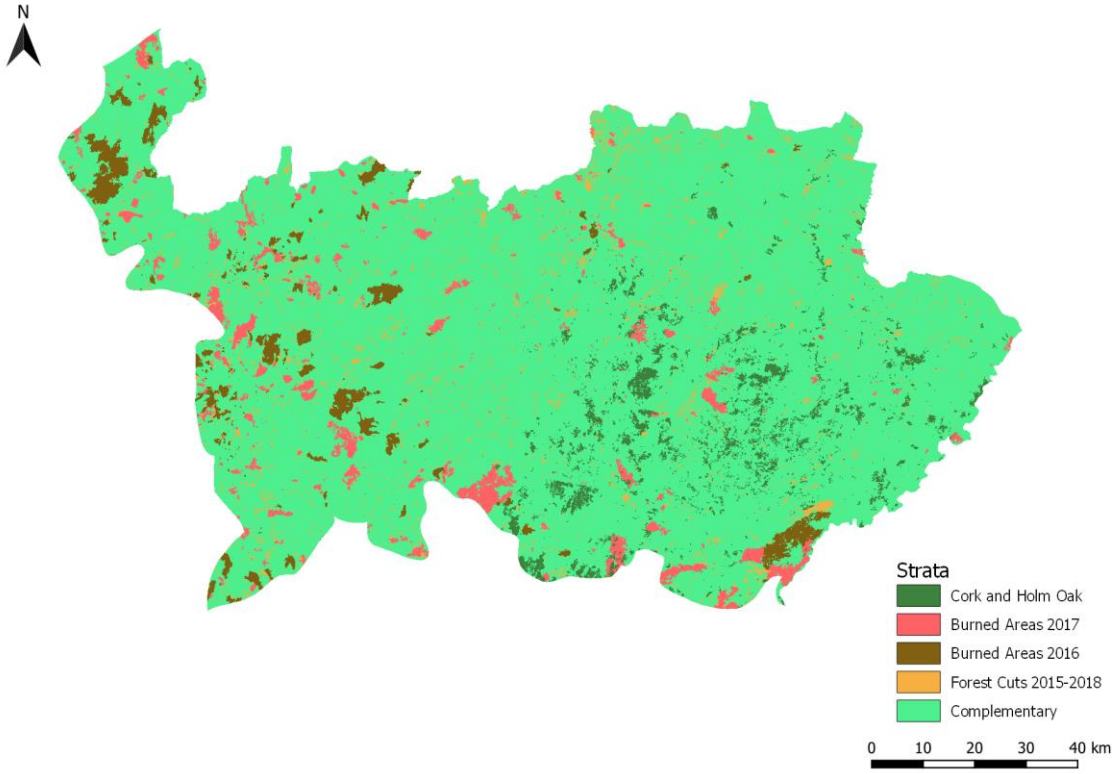

Figure 2. Stratification of the study area into 5 subregions or strata.

Another aspect of the experimental classification is modifying the training sample collection method of specific classes from automatic to manual collection. This approach aims to improve the quality of the training dataset since previous experiments indicated that the automatic collection using reference datasets is insufficient to capture specific spectral characteristics, such as burn scars in the vegetation, or include the spectral diversity of considerably heterogeneous classes, which can increase classification errors. The training classes in which the manual method is employed can be seen in Table 1. The manual collection process consists in the acquisition of training data through digitization of polygons representing the target land cover type, through photointerpretation with a 25 cm orthophotomap. Similarly to the benchmark procedure, the experimental training dataset is composed of up to 6000 sampling units per class.

Table 2. List of subregions and respective extension.

Number	Subregion	Area (ha)	%
1	Cork and Holm Oak	39113	3.3
2	Burned Areas 2017	42981	3.6
3	Burned Areas 2016	35418	3.0
4	Forest Cuts 2015-2018	45513	3.9
5	Complementary	1014777	86.2
	Total	1177802	100.0

Regarding the classification, 5 RF models are trained to classify the respective subregion using the Sentinel-2 composites. The classes used in the model training vary according to the subregion (Table 3). Despite the stratification, training samples are collected from the whole study region regardless of the strata, except for the subregion-specific classes (e.g. Natural Grasslands Burned Areas 2017).

Table 3. Training classes adopted in the classification of each subregion.

Map Class	Training Class	Subregion				
		1	2	3	4	5
Built up	Built up	X	X	X	X	X
	Industrial	X	X	X	X	X
	Road Network	X	X	X	X	X
Agriculture	Oat	X	X	X	X	X
	Wheat	X	X	X	X	X
	Barley	X	X	X	X	X
	Ryegrass	X	X	X	X	X
	Triticale	X	X	X	X	X
	Rye	X	X	X	X	X
	Corn	X	X	X	X	X
	Sunflower	X	X	X	X	X
	Managed Grasslands	X	X	X	X	X
Natural Grasslands	Agricultural Natural Grassland	X	X	X	X	X
	Mountain Natural Grassland	X	X	X	X	X
	Natural Grasslands Burned Areas 2017		X			
Cork and Holm Oak	Cork Oak				X	
	Holm Oak				X	
Eucalyptus	Eucalyptus Adult	X	X	X	X	X
	Eucalyptus Burned Areas 2017		X			
	Eucalyptus 1 year cuts			X		
	Eucalyptus Young Cuts			X		
Other Broadleaf	Other Broadleaf	X	X	X	X	X
Maritime Pine	Maritime Pine	X	X	X	X	X
Stone Pine	Stone Pine	X	X	X	X	X
Other Coniferous	Other Coniferous	X	X	X	X	X
Shrubland	Dense Shrubland	X	X	X	X	X
	Shrubland Burned Areas 2017		X			
Non-vegetated surfaces	Baresoil	X	X	X	X	X
	Bare Rock	X	X	X	X	X
Water	Water	X	X	X	X	X

3.3 Accuracy assessment and comparison

An independent validation dataset composed of 600 sampling units acquired by stratified random sampling is used to compute the classification accuracy of the benchmark and experimental classification maps. Labeling of the validation dataset was done by visual interpretation of a 25 cm orthophotomap. Validation is conducted considering the map classes rather than the training classes. The accuracy assessment is conducted using metrics such as the overall accuracy, precision, recall and f1-score. These metrics are then used to compare both classifications and evaluate whether the introduction of stratification and manual training is beneficial.

4. RESULTS AND DISCUSSION

The benchmark classification exhibited an overall accuracy of 60.2%, whilst the experimental classification scored 66.7%, meaning a fair increase in accuracy. The analysis of the accuracy metrics per class (Table 4) reveals that, in terms of f1-score, the inclusion of stratification and manual training benefited all classes, except for built up, non-vegetated surfaces and natural grasslands, even though the last two were expected to benefit from the manual training which covers their spectral subclasses. On the other hand, cork and holm oak, eucalyptus and other broadleaf were the classes that experienced the most benefits. A substantial reduction in the commission error, as seen in the precision, was observed in

the cork and holm oak subregion. This can be explained by the spatial constraint imposed by the stratification, which means that this class can only occur in the area within subregion 1.

Table 4. Comparison of benchmark and experimental classification accuracy metrics per class.

Class	Precision (%)		Recall (%)		F1-score (%)	
	Benchmark	Exp.	Benchmark	Exp.	Benchmark	Exp.
Built up	47.06	43.24	100.00	94.12	64.00	59.26
Agriculture	57.45	50.91	67.50	91.80	62.07	65.50
Natural Grasslands	78.22	82.50	65.83	36.67	71.49	50.77
Cork and Holm Oak	7.14	83.33	85.71	83.33	13.19	83.33
Eucalyptus	66.67	42.86	7.14	25.00	12.90	31.58
Other Broadleaf	100.00	97.44	29.58	55.88	45.65	71.03
Maritime Pine	74.76	77.06	59.69	67.74	66.38	72.10
Other Coniferous	25.00	28.95	33.33	33.33	28.57	30.99
Shrubland	65.42	68.09	72.16	82.76	68.63	74.71
Non-vegetated Surfaces	57.14	41.18	44.44	63.64	50.00	50.00
Water	100.00	100.00	98.00	98.00	98.99	98.99

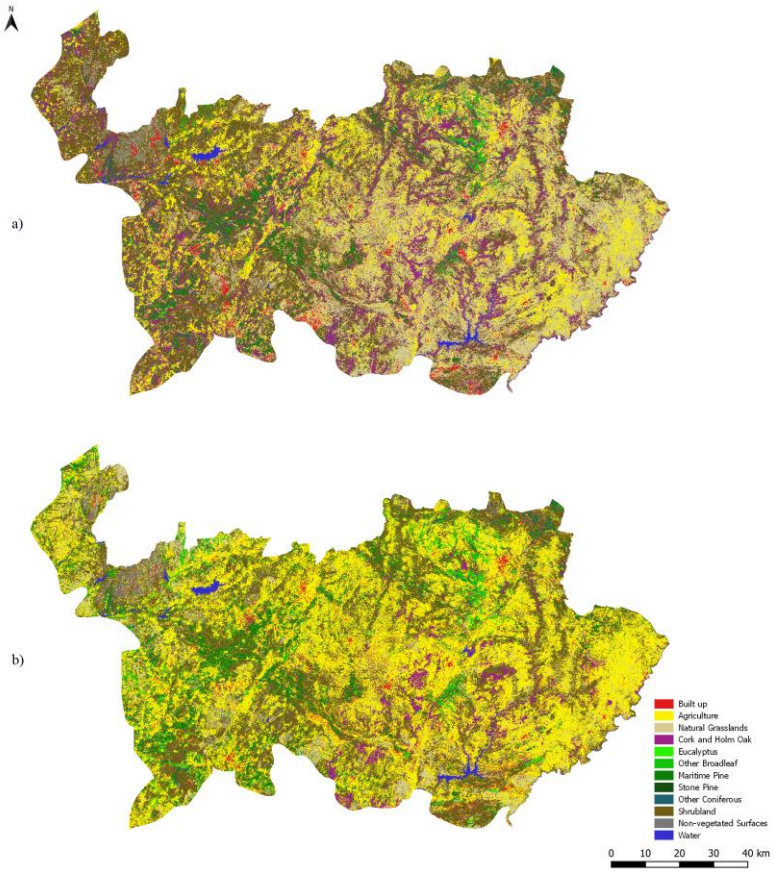

Figure 3. Land cover maps of the classifications – a) benchmark map; b) experimental map.

Besides the accuracy assessment, a visual inspection of the maps was conducted. Figure 4 depicts the classifications of an area affected by forest fires in 2017 (subregion 2), revealing that the stratification and introduction of manual training for burned natural grasslands and eucalyptus (Figure 4c) reduced the misclassification of built up within burned areas.

Another example of how stratification and manual training might have enhanced the classification is presented in Figure 5. Within subregion 4 (forest cuts 2015-2018), an area identified as eucalyptus forest by the 2018 national LCLU

cartography was mapped mostly as shrubland and cork and holm oak in the benchmark map (Figure 5b). On the other hand, the experimental map classified correctly most of the eucalyptus (Figure 5c), which can be explained by the introduction of the manual training class eucalyptus young cuts and eucalyptus 1 year cuts together with the strategy of mapping forest cuts in a separate subregion.

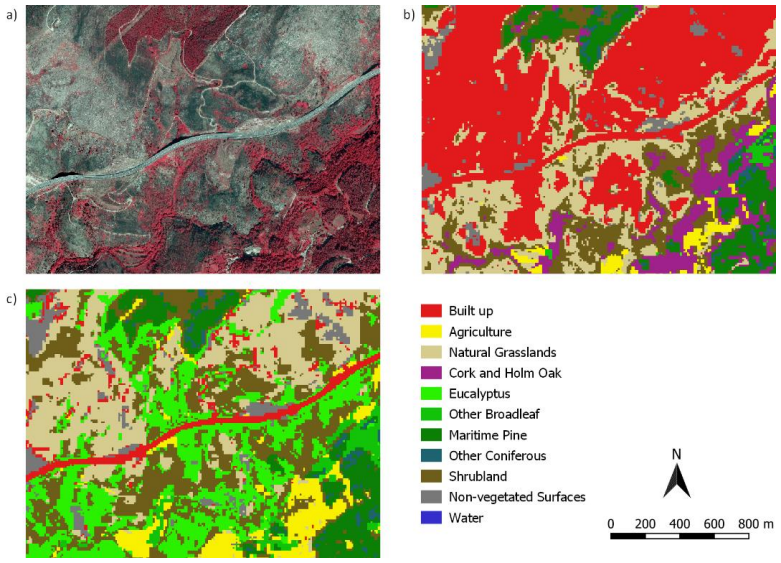

Figure 4. Benefits of stratification and manual training – a) orthophoto of an area affected by fires in 2017 (subregion 2); b) benchmark classification; c) experimental classification.

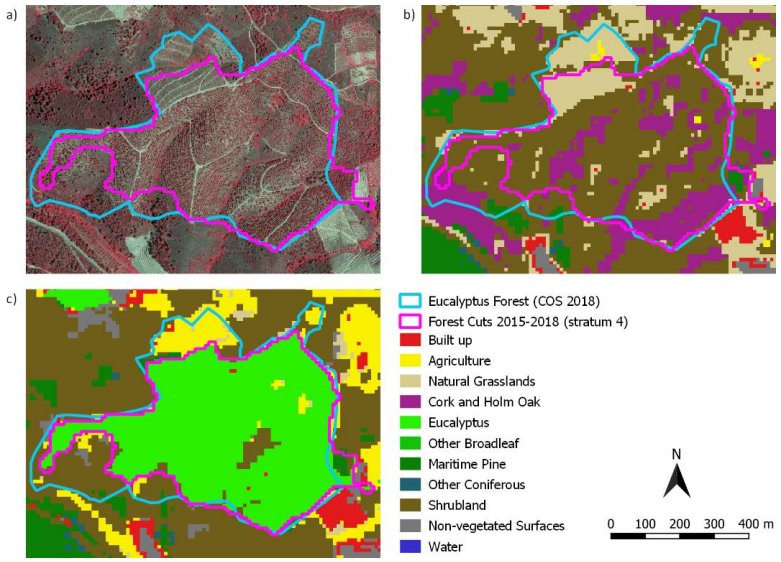

Figure 5. Benefits of stratification and manual training – a) orthophoto of an area where forest cuts occurred (subregion 4); b) benchmark classification; c) experimental classification.

Visual inspection of the maps provided evidence of improvements in the classification when introducing stratification and manual training. The results of the quantitative analysis might have been influenced by the predominance of the complementary subregion in the study area, as it represents 86.2% of the area. Furthermore, 89.17% of the validation sampling units belong to this subregion. Since the complementary subregion corresponds to generic spectral characteristics rather than distinguishing particular landscapes or dynamics (e.g. burned areas), its classification process is similar to the benchmark approach. In fact, most examples of classification improvement observed by visual inspection of the maps were noticed in areas outside the complementary subregion.

5. CONCLUSION

This study proposed to assess the impact of stratification and introduction of manually collected training samples on automatic supervised classification using the Random Forest classifier and multi-temporal Sentinel-2 imagery. A benchmark classification, conducted without stratification and with training data collected by a fully automatic process, was compared to an experimental classification, carried out with stratification and combination of automatic and manual training data. The accuracy assessment of both classifications, conducted with an independent validation dataset, revealed that the experimental classification increased the overall accuracy. Visual inspection of the maps also presented evidence that the experimental product improved the classification in comparison with the benchmark map, especially in areas within the limits of certain subregions. A larger increase in classification accuracy might have been prevented by the predominance of the complementary subregion in the study region, which accounts for over 86% of the area and represents generic spectral patterns instead of the specific patterns present in the other subregions. However, spatial stratification and manual training appear to be largely beneficial to locations associated to challenging dynamics such as wildfires.

ACKNOWLEDGEMENTS

This work has been supported by project foRESTER (PCIF/SSI/0102/2017), 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 processing uses algorithms developed by Theia's Scientific Expertise Centres.

REFERENCES

[1] Maxwell, A. E., Warner, T. A. and Fang, F., “Implementation of machine-learning classification in remote sensing: an applied review”, International Journal of Remote Sensing 39(9), 2784- 2817 (2018)

[2] Belgiu, M. and Drăguţ, L., “Random forest in remote sensing: A review of applications and future directions”, ISPRS Journal of Photogrammetry and Remote Sensing 114, 24-31 (2016).

[3] Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I., “Operational high resolution land cover map production at the country scale using satellite image time series”, Remote Sensing 9(1), 95 (2017).

[4] Griffiths, P., Nendel, C., and Hostert, P., “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).

[5] Cano, E., Denux, J., Bisquert, M., HubertMoy, L. and Chéret, V., “Improved forest-cover mapping based on MODIS time series and landscape stratification”, International Journal of Remote Sensing 38(7), 1865-1888 (2017)

[6] Costa, H., Benevides, P., Marcelino, F. and Caetano, M., “Introducing automatic satellite image processing into land cover mapping by photo-interpretation of airborne data”, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-3/W11, 29–34 (2020).

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., “Scikit-learn: machine learning in Python”, Journal of Machine Learning Research 12, 2825-2830 (2011).

[8] Hernandez, I., Benevides, P., Costa, H. and Caetano, M., “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).