

EXPLORING THE POTENTIAL OF SENTINEL-2 DATA FOR TREE CROWN MAPPING IN OAK AGRO-FORESTRY SYSTEMS

Hugo Costa^{1,2}, Inês Machado¹, Francisco D. Moreira¹, Pedro Benevides¹,
Daniel Moraes^{1,2}, Mário Caetano^{1,2}

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

²NOVA 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: Costa, H., Machado, I., Moreira, F. D., Benevides, P., Moraes, D., & Caetano, M. (2021). Exploring the Potential of Sentinel-2 Data for Tree Crown Mapping in Oak Agro-Forestry Systems. In *IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium: Proceedings* (pp. 5807-5810). IEEE. <https://doi.org/10.1109/IGARSS47720.2021.9553780>

© 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.

EXPLORING THE POTENTIAL OF SENTINEL-2 DATA FOR TREE CROWN MAPPING IN OAK AGRO-FORESTRY SYSTEMS

Hugo Costa,^{1,2} Inês Machado¹, Francisco D. Moreira¹, Pedro Benevides¹, Daniel Moraes^{1,2}, Mário Caetano^{1,2}

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

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

ABSTRACT

Southern Portugal is characterized by disperse tree cover of Cork and Holm oaks in an agro-forestry system known as *montado*. Mapping these trees has been historically very difficult as they occur in isolation or in groups with different understory vegetation, including grass and shrubland. Automatic classification for binary tree/non-tree map production has been used elsewhere, but with limited success in the context of *montado*. Here, the potential of Sentinel-2 data was explored to map oaks using pure and mixed pixels to train a random forest. The output depicts a gradient of tree cover that can be transformed into a crisp map. The accuracy assessment of the latter shows commission and omission errors of 17% and 18%.

Index Terms— Cork oak, Holm oak, fuzzy, mixed pixels

1. INTRODUCTION

Southern Portugal is largely occupied by agro-forestry systems known as *montado*, which is formed by disperse tree cover of oaks, namely in pure or mixed stands of Cork oak and Holm oak (Fig. 1). The tree density is highly variable in magnitude and spatial patterns, varying between isolated, sparse trees and rather dense wood stands. This agro-forestry system also includes a diversity of low vegetation underneath and between trees, including crops, pastures and occasional patches of shrubs, making this landscape very diverse and heterogeneous.

The value of *montado* is largely acknowledged and appreciated, and hence protected by national and international law. However, the density of trees has declined in the last decades and there are many concerns about its future, especially in the context of climate change.

Mapping trees has been done all over the world [1], including oak stands [2], but typically the spatial resolution of the data used is very high, costly data such as LiDAR is needed, or the quality of the results is limited in the context of *montados* and similar landscapes [3]. The situation is

particularly difficult when coexisting classes are spectrally similar in the optical domain of the wavelength spectrum, such as trees and dense shrubs.

Despite the difficulties, the advent of Sentinel-2 satellites offers an attractive alternative for mapping trees because data of fairly detailed spatial resolution became freely available for almost the entire world. Therefore, Sentinel-2 has been explored to map trees in agro-forestry systems [4]. However, issues remain, and traditional multi-class and hard image classification under-represents oak crowns due to their dispersed distribution. Here, gradient tree mapping [5] based on training classes that express transition from tree to non-tree cover were used to train a Random Forest in a large region of south Portugal. The goal was to produce a map of tree crowns in areas of *montado*.

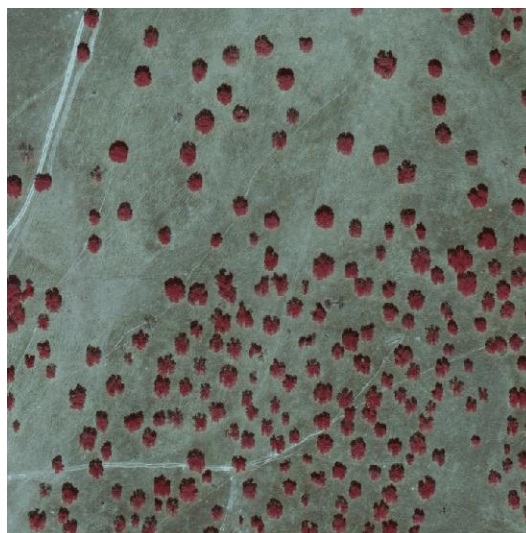


Fig. 1: Example of *montado* in Portugal through an orthophotomapping of 2018 in false colour.

2. STUDY AREA

South Portugal has been divided in landscape units for land cover classification from remotely sensed data [6], and one

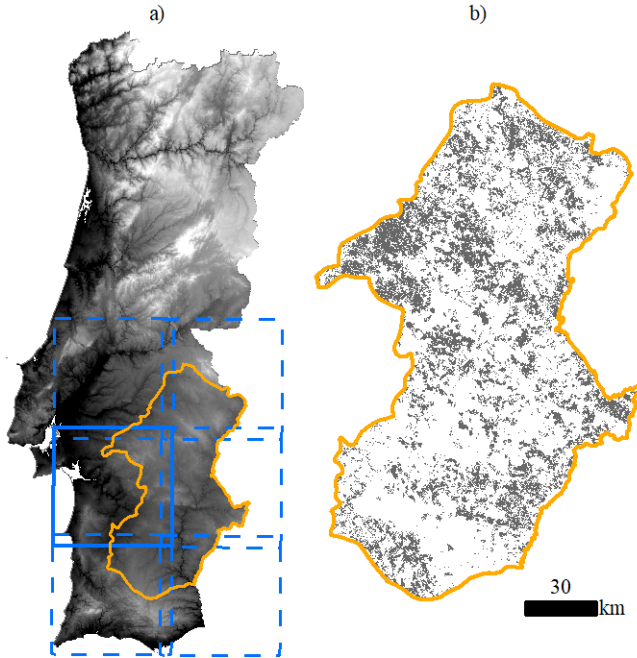


Fig. 2: a) Landscape unit (yellow outline) in Portugal (DEM showed as background) and overlapping Sentinel-2 granules (blue outlines); b) Study area (in grey) inside the landscape unit.

landscape unit in the countryside of region *Alentejo* was selected for this study (Fig. 2a). Inside the unit the analysis focused on areas mapped as *montado* in the national land cover/land use map of Portugal for 2018 (hereafter referred as COS), which is a vector map with a minimum mapping unit of 1 ha. In COS, *montado* has a minimum of 10% tree cover. A total of 4 526 km² was included in the analysis (Fig. 2b).

In addition, the area of 100 x 100 km that corresponds to granule T29SNC of Sentinel-2 data (blue solid line in Fig. 2a) was used for evaluating the results. This was done as this granule covers two landscape units of interest in preliminary research, although only one landscape unit is presented.

3. DATA AND METHODS

Sentinel-2A (S2A) and Sentinel-2B (S2B) data acquired in summer were used to take advantage of the spectral contrast between the vegetation that dries during summer and evergreen vegetation. The latter is the case of oaks while low vegetation such as grass and crops get dry and harvested. Therefore, S2A and S2B level 2A data were downloaded from the Theia data centre, namely S2A images of the 8th and 18th of August 2018, and S2B images of the 3th, 13th, and 23th of the same month. These images were used to produce a single image composite representative of the complete month. This was done to deal with missing data caused by clouds and cloud shadows and other noise. The

images were combined band by band with the median, except bands 1, 9 and 10 which were not used. The monthly composite also allowed the production of a seamless image made by the six Sentinel granules that cover the landscape unit (Fig. 2a).

Training points were collected manually across the study area and guided by orthophotomaps of 2018 with 25 cm spatial resolution. Each training point was collected to represent a pixel of 10 m and associated with one of the training classes defined. These classes express gradual levels of abundance of four elements: oaks, shrubs, grass and (bare) soil. The training protocol, however, assumed for each pixel gradient levels between oaks and only one of the remaining three elements. Therefore, each of the training pixels includes only a maximum of two elements. Moreover, the abundance of each element in the pixels was defined in four categories according to the order of presence of oaks: 100%, 80%, 20% and 0% of cover. As a result, a total of 12 classes were defined (Table 1).

Table 1. Training classes and sample size.

Class id	Class code	% oak	Sample size
1	Oak100_shrub	100	51
2	Oak080_shrub	80	53
3	Oak100_soil	100	54
4	Oak080_soil	80	59
5	Oak100_grass	100	60
6	Oak080_grass	80	63
7	Shrub100_oak	0	53
8	Shrub080_oak	20	62
9	Soil100_oak	0	54
10	Soil080_oak	20	55
11	Grass100_oak	0	53
12	Grass080_oak	20	58

The training protocol admitted classes apparently equivalent, specifically classes 1, 3 and 5 which represent pixels with a pure cover of oaks. However, different elements occur in the understory or vicinity (shrub, soil and grass, respectively), and hence the spectral content of the pixels might vary as a function of the spectral contamination from the understory or neighbouring pixels.

A total of 675 training points were collected, with a fairly balanced distribution across the classes (Table 1). The randomForest package [7] of R programming language [8] was used to predict the gradual class of oak cover in the entire study area. The spectral bands of the monthly composite were used as classification variables in addition to five spectral indices: NDVI, NDBI, NDWI, NBR, and NBR2 [9]. A national digital terrain model (DTM) was also used.

The map produced was analyzed to understand which classes should be considered as corresponding mostly to oak crowns. All but classes 7, 9 and 11 were considered as oak and grouped together to do a binary map of tree crowns.

The accuracy of the results was assessed for Cork and Holm oaks independently. This was done to understand if the results could vary as a function of the oak species even if they are extremely similar and not expected to be distinguishable from each other with remote sensing data. Therefore, two validation samples were drawn from areas mapped as either Cork oak or Holm oak in COS. In each case, 100 points were randomly allocated inside the polygons of COS of the corresponding species and intersecting Sentinel-2 granule T29SNC (Fig. 2a). Sampling was random stratified by the map classes, using 25 points distributed by four strata: oaks, shrubs, grass and soil. The ground-truth class was determined through visual interpretation of the orthophotomaps of 2018. Some few points were found in land cover types not relevant for the study, such as narrow rivers and roads crossing the landscape. These random points were ignored during validation.

4. RESULTS

The classification of the study area resulted in a map of 12 classes later merged in tree and non-tree classes. Fig. 3 shows the binary map, but class “trees” is presented with three levels of oak cover. The spatial patterns that immediately arise are the heterogeneous distribution of pixels classified as including oak trees with variable levels of tree cover. This is a characteristic typical of *montado*.

The two validation samples were used to assess the classification accuracy (considering the 12 classes of Table 1), and 41% and 65% of the points were correctly classified, respectively for Cork and Holm oaks. However, substantial misclassification is observed between classes that represent presence or absence of oaks, which is innocuous for a binary tree/non-tree map. The accuracy of the binary map is larger, and 78% and 85% of accuracy is estimated from the two samples (Cork and Holm oaks, respectively). The two samples combined show an accuracy of 81%.

The error matrix of the two samples combined for the binary classification is shown in Table 2, which reveals equivalent levels of commission and omission error (around 20%) for both classes (presence and absence of oaks). For the specific class of presence of oak trees, Recall and Precision are 83% and 82%, resulting in a F1 score of 83%.

Table 2. Error matrix.

	Trees	Non-trees	Commission error
Trees	86	17	17%
Non-trees	19	68	22%
Omission error	18%	20%	

5. DISCUSSION AND CONCLUSIONS

Mapping trees is a challenging task that often requires complex analyses and data. However, promising results were obtained only with optical data of 10 and 20 m spatial resolution collected during a short period of summer 2018. This was possible because mixed pixels were used at the training stage contributing with information on spectral patterns related to the presence of trees. Using mixed pixels and sub-pixel information has been proved valuable for image classification when the pixel size is larger than the object of interest [10].

Omission and commission errors of around 20% were obtained. One important source of error is the scattered spatial distribution of oaks. Where trees occur in isolation, the detection rate of trees decreases considerably, and it is not rare to find punctuated trees unnoticed in the map (Fig. 3). These may be caused by the small size of the tree crowns relative to the size of the pixel, especially if trees are young or unhealthy, making the spectral information barely different from neighbouring pixels with no trees. When trees stand together, their presence becomes more evident (Fig. 3).

Another source of error is the spectral similarity between some of the classes addressed, such as oaks and dense shrubs. In this case, there is no contrast between the tree crowns and the understory, and the subtle spectral differences between the two types of vegetation are often insufficient for a successful classification. Confusion between trees and shrubs and other classes is a long-standing classification problem, namely in Mediterranean landscapes such as south Portugal [11].

Although the good results achieved with the proposed methodology there is room for further improvement. One front to explore is image texture. The spatial arrangement of isolated and sparse trees is a characteristic of *montado* and similar landscapes [3] and can be important for image classification [4]. Usage of data other than Sentinel-2 may also be an important asset, although with the potential caveat of increasing too much the complexity of the analysis, production time or costs involved. An effective but yet relatively simple approach for mapping trees is relevant for operational programmes that typically map large areas with moderate to high spatial resolution of low or no cost.

6. ACKNOWLEDGEMENTS

The work has been supported by projects IPSTERS (DSAIPA/AI/0100/2018), foRESTER (PCIF/SSI/0102/2017), and SCAPE FIRE (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

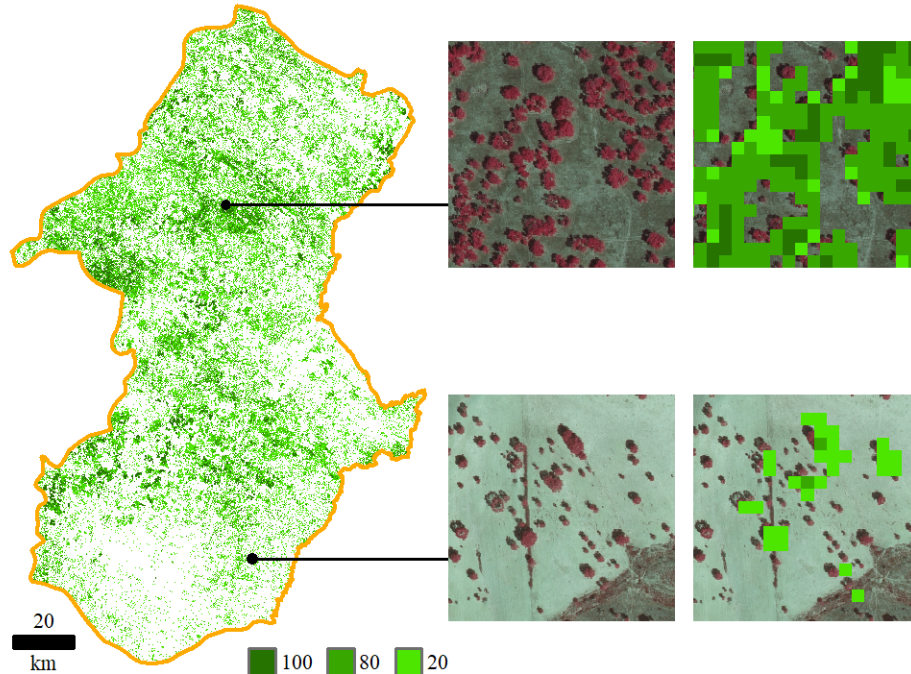


Fig. 3: Tree crown map of Cork and Holm oaks with three levels of crown cover. Levels 100, 80 and 20 correspond to the classes of the same oak crown cover (e.g. level 100 are classes 1, 3 and 5 together in Table 1). Insets show contrasting examples of classification success with orthophotomaps of 2018 as background.

processing uses algorithms developed by Theia's Scientific Expertise Centres.

7. REFERENCES

- [1] M. C. Hansen *et al.*, "High-Resolution Global Maps of 21st-Century Forest Cover Change," *Science (80-.)*, vol. 342, no. 6160, pp. 850–853, Nov. 2013.
- [2] P. Surovy, N. Almeida Ribeiro, and D. Panagiotidis, "Estimation of positions and heights from UAV-sensed imagery in tree plantations in agrosilvopastoral systems," *Int. J. Remote Sens.*, vol. 39, no. 14, pp. 4786–4800, Aug. 2018.
- [3] U. Gessner, M. Machwitz, C. Conrad, and S. Dech, "Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles," *Remote Sens. Environ.*, vol. 129, pp. 90–102, Feb. 2013.
- [4] S. Godinho, N. Guiomar, and A. Gil, "Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm," *Int. J. Remote Sens.*, vol. 39, no. 14, pp. 4640–4662, Aug. 2018.
- [5] H. Feilhauer *et al.*, "Let your maps be fuzzy!—Class probabilities and floristic gradients as alternatives to crisp mapping for remote sensing of vegetation," *Remote Sens. Ecol. Conserv.*, pp. 1–14, 2020.
- [6] H. Costa, "Estrategia multi-temporal para producao automatica de cartografia de ocupacao do solo com imagens AWiFS," Instituto Superior de Estatistica e Gestao da informacao, 2009.
- [7] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R news*, vol. 2, no. December, pp. 18–22, 2002.
- [8] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2019.
- [9] I. Hernandez, P. Benevides, H. Costa, and M. Caetano, "Exploring Sentinel-2 for land cover and crop mapping in Portugal," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLIII-B3-2, pp. 83–89, Aug. 2020.
- [10] G. M. Foody, "Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network," *Neural Comput. Appl.*, vol. 5, no. 4, pp. 238–247, 1997.
- [11] H. Carrao, P. Goncalves, and M. Caetano, "Contribution of multispectral and multitemporal information from MODIS images to land cover classification," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 986–997, Mar. 2008.