# Exploring Spectral Data, Change Detection Information and Trajectories for Land Cover Monitoring over a Fire-Prone Area of Portugal

André Alves[1][a], Daniel Moraes[2,3][b], Bruno Barbosa[1][c], Hugo Costa[2,3][d], Francisco D. Moreira[3][e], Pedro Benevides[3][f], Mário Caetano[2,3][g] and Manuel Campagnolo[1][h]

[1]*Forest Research Centre, Associate Laboratory TERRA, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017 Lisboa, Portugal*

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

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

Keywords: Land Cover Change Classification, Thematic Map, Spectral Composites, NDVI, CCDC, COSc, Earth Observation.

Abstract: Land use/land cover (LULC) change detection and classification in maps based on automated data processing are becoming increasingly sophisticated in Earth Observation (EO). There is a growing number of annual maps available, with diverse but related production structures consisting primarily of classification and post-classification phases, the latter of which deals with inaccuracies of the first. The methodology production of the "*Carta de Ocupação do Solo conjuntural*" (COSc), a thematic land cover map of continental Portugal produced by the Directorate-General for Territory (DGT) mostly based on Sentinel-2 images classification, includes a semi-automatic phase of correction that combines expert knowledge and ancillary data in if-then-else rules validated by photointerpretation. Although this approach reduces misclassifications from an initial Random Forest (RF) prediction map, improving consistency between years and compliance with ecological succession, requires a lot of time-consuming semi-automatic procedures. This work evaluates the relevance of exploring an additional set of variables for automatic classification over disturbance-prone areas. A multitemporal dataset with 124 variables was analysed using data dimensionality reduction techniques, resulting in the identification of 35 major explanatory indicators, which were then used as inputs for RF classification with cross-validation. The estimated importance of the explanatory variables shows that composites of spectral bands, which are already included in the current COSc workflow, in conjunction with the inclusion of additional data namely, historical land cover information and change detection coefficients, from the Continuous Change Detection and Classification (CCDC) algorithm, are relevant for predicting land cover classes after disturbance. Since map updating is a more challenging task for disturbed pixels, we focused our analysis on locations where COSc indicated potential land cover change. Nonetheless, the overall classification accuracy for our experiments was 72.34 % which is similar to the accuracy of COSc for this region of Portugal. The findings suggest new variables that could improve future COSc maps.

## 1 INTRODUCTION

Land use/land cover (LULC) products by remote sensing and satellite image classification are becoming increasingly accurate in land representation. Earth observation (EO) has seen significant practical and theoretical advances as data and machine-learning tools have become more

[a] https://orcid.org/0000-0002-8979-8906
[b] https://orcid.org/0000-0002-4568-8182
[c] https://orcid.org/0000-0001-6298-041X
[d] https://orcid.org/0000-0001-6207-8223
[e] https://orcid.org/0000-0001-7213-5551
[f] https://orcid.org/0000-0001-5858-6815
[g] https://orcid.org/0000-0001-8913-7342
[h] https://orcid.org/0000-0002-9634-3061



Alves, A., Moraes, D., Barbosa, B., Costa, H., Moreira, F., Benevides, P., Caetano, M. and Campagnolo, M.
Exploring Spectral Data, Change Detection Information and Trajectories for Land Cover Monitoring over a Fire-Prone Area of Portugal.
DOI: 10.5220/0011993100003473
In *Proceedings of the 9th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2023)*, pages 87-97
ISBN: 978-989-758-649-1; ISSN: 2184-500X

accessible (Wulder et al., 2018). In the age of big data, these map products are becoming more refined in terms of both thematic and spatial detail, with increased class heterogeneity. Simultaneously, with the growing number of multi-annual maps (Brown et al., 2020; Buchhorn et al., 2020; Hermosilla et al., 2018) and near-real-time products (Brown et al., 2022), the temporal aspect has also received attention. However, measuring vegetation recovery and predicting post-disturbance classes in line with ecological succession principles, as well as other concerns of temporal consistency of land cover change direction, remains a challenge in time series of land cover (Bartels et al., 2016; White et al., 2022; Wulder et al., 2018).

Land succession processes are complex since they can be of high or low magnitude, abrupt or subtle, and completed in months or take several years (Zhu et al., 2022). Also, misclassifications can be numerous because of spectral confusion, land cover heterogeneity within the pixel resolution, different phases of vegetation growth and the broad conceptual definition of classes (e.g., bushland, scrubland, moorland, shrub-steppe, etc., are normally classified in the same class). Several examples of works presenting methodological contributions to post-classification error reduction, improving annual consistency and better-predicting land cover change trajectories, can be found in the literature. Annual filters limiting LULC misclassifications (Franklin et al., 2015), spatial-temporal joint classifications (Cai et al., 2014) and time-series post-classification (Hermosilla et al., 2018), are a few examples that illustrate how diverse the proposed methodologies are. Those examples used several dimensions of data for consistency refinement. Ranging from transition rules based on prior knowledge, spectral and change detection information, historical land cover and class membership data, to contextual information of adjacent pixels, a wide range of potential features can be contemplated to more accurately predict the land cover class in a context of ecological succession.

This study intends to explore an approach to improve the "*Carta de Ocupação do Solo conjuntural*" (COSc). The COSc is a 10-meter raster thematic annual map of land cover for continental Portugal with 15 classes. As the outcome of a supervised classification of satellite images and ancillary data, its production settles on a multi-stage workflow with preliminary maps being produced along the way (Costa et al., 2022b). The first phase is the COScA, a step of automatic data processing that consists of a supervised classification with the Random Forest (RF) algorithm to classify land cover classes based on stratified point samples. Subsequently, a semi-automatic phase takes place, COScR, minimizing misclassifications through expert knowledge implemented by if-then-else rules. It is estimated that COScR map is at least 13 % more accurate than COScA (Costa et al., 2022b). Subsequently, the intra-annual vegetation losses, such as wildfires during summer, are assessed in the COScP phase. Finally, a harmonization process over landscape units concludes the COSc workflow (COScH).

The inputs for the automatic steps of the COSc workflow are partly derived from the time series of Sentinel-2 imagery. COScA uses as inputs monthly composites and COScP relies on detected temporal breaks of a vegetation index to identify potential intra-annual vegetation losses. The COSc map has been made available to users since 2018 and the release of successive yearly products follows a "map updating" strategy, where only pixels with evidence of change are possibly reclassified. More precisely, successive updates focus only on pixels expected to change (e.g., fire scars) or identified as disturbed according to a temporal analysis of the spectral profile and assume that the rest of the map remains unchanged (Costa et al., 2022a). COSc has been updated to 2020, 2021 and the 2022 version is under production.

The current work explores the suitability of a broad range of variables as new input data for the COSc workflow. Towards that end, a dataset was created with a set of variables for land cover classification, which includes spectral data, historical land cover information, Markov chain transition probabilities, and change detection information from the Continuous Change Detection and Classification (CCDC) algorithm, which also models the temporal pattern of the land cover signal between transitions. At the time of writing, none of the existing versions of the COSc map were created using CCDC or past land cover and their transitions, thus their potential for classification is unknown. Since our set of variables is very large and exhibits some high correlations, data was processed to identify intercorrelated groups of variables, reduce the dataset dimensionality, and minimise multicollinearity when assessing for variables' importance.

Our experiments were made using a reference dataset for a forest fire-prone region in the Center of Portugal (tile 29TNE), where high-intensity large fires occurred in 2017. Our reference data included 300 plots, distributed over fire-affected and non-fire-affected areas, and was created through photointerpretation of a temporal series of Sentinel-2 false colour (RGB 843) composites from September 2018 to September 2021 and orthophotos from the summer of 2018 and 2021.

The objectives of the paper are twofold: (i) assess the accuracy of fully automatic land cover classification; (ii) identify the relevance of new candidate variables to be included in the COSc workflow. The remainder of the paper is organized as follows: Section 2 – Data and Methods, Section 3 – Results, Section 4 – Discussion and Section 5 – Conclusions.

# 2 DATA AND METHODS

The methodological approach was supported in Python environment. First, a hierarchical clustering analysis identified groups of highly correlated explanatory variables, that were subsequently generalised by a principal component analysis (PCA) to obtain a parsimonious dataset. The supervised classification was performed using RF. Spectral, thematic, transition probabilities and change detection information were considered as inputs to predict the 2021 class given by photointerpretation. The accuracy of the model was evaluated using traditional metrics and the calculation of Receiver Operating Characteristic (ROC) curves. Feature importance was assessed, and the results were mapped.

## 2.1 Case Study Area

This research was conducted on the Sentinel-2 tile 29TNE, which covers a ~100 km wide swath along the West coast of Central Portugal (Figure 1). This area of continental Portugal landscape is characterized by the coexistence of dense forest, agricultural areas and urban land use bordering wilderness (urban-rural interface areas). According to Alves et al. (2022), there have been significant changes in forest ecosystems in this area over the past few decades. One of the most notable changes has been the decline of maritime pine forests, as the result of the expansion of eucalyptus plantations which now dominate the region (Alves et al., 2022). Furthermore, the study area was affected by large forest fires in 2017 (San-Miguel-Ayanz et al., 2020). As a result, it is a dynamic area with the potential for land change due to vegetation gains.

To run classification tests, the study area was sampled with 300 circular plots with a 200-meter radius. Those plots were randomly located over tile 29TNE, according to a stratification that separated the sample into 150 plots that exhibited a spectral change in the agricultural years from 2018 to 2021 and 150 that did not. The occurrence of change was identified with the CCDC algorithm applied to the full Sentinel-2 Level-2A time series, which initiates in early 2017.

After the 300 plots were chosen as described above, a team of photointerpreters relied on a series of monthly Sentinel-2 temporal composites from September 2018 to September 2021 as well as 25 cm orthophotos from the summer of 2018 and 2021 to segment those plots based on land cover and actual occurrence of change. When a change was identified by the photointerpreter, the thematic class for the polygon pre-change and post-change was registered according to the COSc legend. Therefore, the reference dataset includes the class (minimum mapping unit of 0.5 ha) for polygons that registered changes from September 2018 to September 2021. From these polygons originally 380,951 pixels were derived (10-meter spatial resolution). The official COSc2021 was used to label the polygons identified

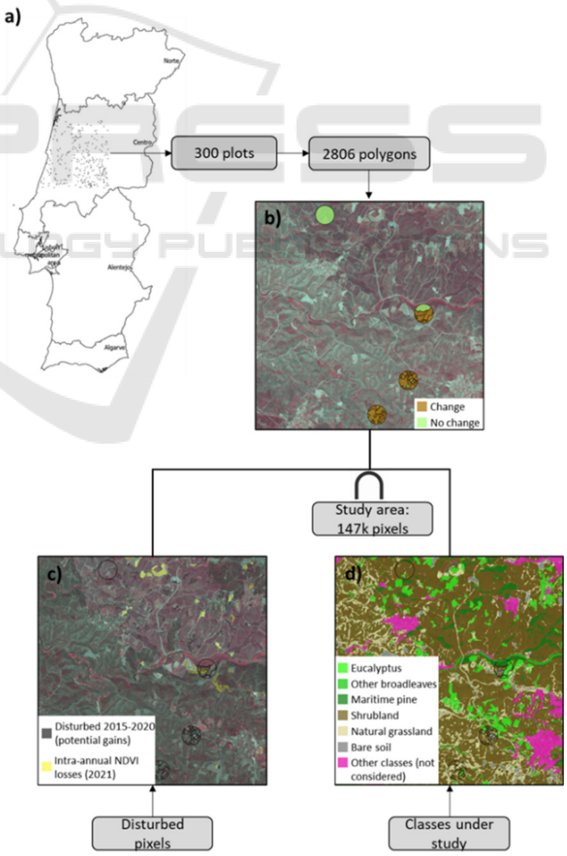


Figure 1: Definition of the study area: a) Plots in tile 29TNE; b) Change/No-change polygons; c) Disturbed pixels under study; d) COSc map land cover classes considered.

as "no change" when the dominant class remained equal to 2020 and covered more than 80 % of the polygon area. "No change" polygons not meeting that threshold were not included in our study area. In addition, only pixels with potential vegetation gains were kept (COScA i.e., areas where a disturbance occurred in the last years (2015-2020) and COScP i.e., losses (clear-cuts and fires) in the 2021 agricultural year). From the 15 classes of the COSc legend, only 6 classes were relevant over the study area: bare soil, natural grassland, shrubland, eucalyptus, other broadleaves and maritime pine. As a result, the analysis focused on those 6 land cover classes (see Figure 1).

After these conditions were imposed, 147,060 pixels remained for the classification experiments which covered about 14,7 km$^2$. From 2020 to 2021, according to the photointerpreters, 73945 pixels changed land cover class, of which 20688 registered a loss of vegetation while 53257 had some type of growth.

## 2.2 Data

The multidimensional dataset (Table 1) was based on a literature review of studies that address multitemporal map production, annual consistency refinement and ecological succession (Abercrombie & Friedl, 2016; Franklin et al., 2015; Hermosilla et al., 2018; Liu et al., 2021; Xian et al., 2022; Xie et al., 2022). The 124 selected variables arose from a compromise between information that is not considered for the current COSc classification and that could be easily generated without disrupting the existing workflow. Furthermore, the new variables were supposed to inform about the ecological succession trajectory and the type of land cover change. All information was resampled to the 10-m spatial resolution of Sentinel-2 visible and near-infrared (NIR) bands.

Table 1: Per-pixel raw data.

| Dimension | Variable |
|---|---|
| Historical land cover | COSc2018 and 2020 land cover classes |
| Class transition likelihood | Markovian Conditional Probabilities 2018-2020 |
| Spectral images | B2, B3, B4, B8, B11, B12 and Normalized Difference Vegetation Index (NDVI) monthly composites |
| Change detection information (CCDC) | Annual Cosine term of B2, B3, B4, B8, B11, B12 and NDVI |
| | Annual Sine term of mentioned bands and NDVI |
| | Magnitude change (mentioned bands and NDVI) |
| | Time trend slope (mentioned bands and NDVI) |
| | Time trend intercept (mentioned bands and NDVI) |
| | Time duration of the last segment |
| | Number of detected breaks |
| | Change in the year 2021 |

### 2.2.1 Historical Land Cover Information

The incorporation of prior land cover classes constitutes information relevant for predicting the next class and its inclusion in classification and post-classification processes is not atypical (Cai et al., 2014; Reis et al., 2020). The inclusion of previous classes can be informative for two main reasons. On the one hand because if a specific tree species class has already occurred in a determined location, it is more likely to occur in the future than other tree species. On the other hand, the model can learn what were the most common transitions to occur in the past, giving it greater confidence in predicting the future class. In this sense, the COSc2018 and 2020 classes were used as inputs for classification.

### 2.2.2 Class Transition Likelihood

In research aimed at enhancing the consistency of annual maps, transition probabilities are typically used as input variables in post-classification schemes (Gong et al., 2017). In our approach transition probabilities were derived using Markov Chains based on the Markovian transition estimator in IDRISI Selva. The reference images used were the 2018 and 2020 COSc maps of the study area, but since our objective was to predict the land cover class in 2021 it was defined that the period to project forward from the second image was only one year. The probability calculation assumed that the existing thematic error in COSc did not propagate over time.

### 2.2.3 Spectral Images

The spectral variables used correspond to the Sentinel-2 bands 2 (blue), 3 (green), 4 (red), 8 (near-infrared), 11 and 12 (short-wave infrared). In addition, the Normalized Difference Vegetation Index (NDVI), with a known strong capacity for discriminating phenological profiles of different land covers (Balata et al., 2022; García et al., 2019), was calculated. These spectral data consisted of monthly composites computed using the median value of observations covering the 2021 agricultural year

(October 2020 until September 2021), with the Sentinel-2 Surface Reflectance image collection from Earth Engine Datasets. The s2cloudless was used as the cloud filter and gaps due to missing data were filled based on interpolation using a harmonic model.

### 2.2.4 Change Detection Information

The inclusion of change information in land cover classification can enable a result with a lower degree of misclassifications when considering areas experiencing changes since it is informative about the trajectory of each pixel. Change information was calculated using the CCDC algorithm (Zhu & Woodcock, 2014) in Google Earth Engine. CCDC uses linear harmonic models to detect breaks in time series based on EO data. In particular, the harmonic terms are modelled by periodic functions of sine and cosine for varying periods, although we only use the annual terms (see Table 1). The algorithm was executed on the Sentinel-2 Surface Reflectance image collection and its parameterization is exhibited in Table 2.

The CCDC algorithm was not used to detect new areas with disturbance, since the study area mask was already defined based on the current COSc framework, but for deriving relevant information for classification. Although CCDC is designed to detect more than one change per pixel, and model all temporal segments between detected breaks in the time series, we restricted our analysis just to the most recent segment.

Table 2: Parameters used to run the CCDC algorithm.

| Parameter | Value |
|---|---|
| Lambda | 50 |
| Chi-square | 0.995 |
| Minimum number of years factor | 1 |
| Minimum observations | 6 |
| Maximum iterations | 25000 |
| Breakpoint bands | B3, B12, NDVI |
| TMask bands | B3, B12 |

The CCDC estimated coefficients (intercept, trend, and annual periodicity fitted with sine and cosine terms) model the pattern of the time series after the most recent disturbance and are suitable to land cover classification processes (Xian et al., 2022). The coefficients represent a temporal segment between two breaks. If no disturbance is detected by the algorithm since the beginning of the time series (which is early 2017 for Sentinel-2 surface reflectance data) the full series corresponds to the most recent segment. However, we used the total number of detected breaks for the whole time series as a proxy for the overall frequency of disturbance at the pixel location.

## 2.3 Data Dimensionality Reduction

To obtain a parsimonious model, avoiding data redundancy and multicollinearity when measuring the feature importance, a double-step dimensionality reduction approach was used. First, a hierarchical cluster analysis was performed using all variables' correlation as the distance matrix, applying Ward's aggregation rule. This allowed us to identify groups of highly intercorrelated variables. The variable that represents each group was defined as its principal component according to a PCA. Due to their scales, the prior land cover class (2 variables), Markovian probabilities (6 variables, 1 for each class), and the binary variable of change in 2021 were set aside and added later to the representative variables determined as described above. This approach resulted in a total of 26 new variables (derived from clustering and PCA) plus the 9 ones that had been set aside (see Table 1).

## 2.4 Random Forest Classification and Accuracy Assessment

We used the sklearn library in Python (Pedregosa et al., 2011) to run RF classifications with stratified 10-fold cross-validation. Data were partitioned into training and testing polygons stratified by the 2021 class, ensuring that polygons used for training would not be used for testing, and vice versa. Parametrization ensured a maximum number of 300 trees and $\sqrt{n}$ as the number of features available at each split.

The measurement of variable importance was obtained by the feature permutation algorithm, determining the mean decrease in accuracy for each variable. The permutation method to estimate variable importance takes an explanatory variable $x$ and randomly shuffles its values in the dataset before re-fitting the model. Then, it measures the increase in prediction error regarding the model fitted to the original dataset. Repeating this procedure for each variable $x$ at a time estimates its importance and compares it to the remaining variables.

# 3 RESULTS

Results highlight that some variables, due to their significance, have the potential for inclusion in future COSc map production, specifically, by aiding in the integration of land cover trajectory-related parameters into the classification.

## 3.1 Accuracy

Table 3 shows the cross-validation accuracy assessment with the RF classifier. The parsimonious model achieved an accuracy of 72.34 % (± 1.86 % at the 95 % confidence level). Pixels that were identified by the photointerpreters as “change” exhibited a slightly lower accuracy, particularly in the case of vegetation gains.

Table 3: Model accuracy.

| Type of pixels | Correctly classified pixels (%) |
|---|---|
| Global | 72.34 |
| No change in 2018-2021 | 77.30 |
| Change in 2018-2021 | 67.61 |
| Vegetation gains 2020-2021 | 65.31 |
| Vegetation losses 2020-2021 | 73.52 |

Land cover changes leading to loss of vegetation (e.g., eucalyptus to bare soil; shrubland to natural grassland, etc.) had higher accuracy than vegetation gains. However, the model’s capacity to correctly classify vegetation gains seems to have been influenced by the type of occurrence that caused the disturbance. According to Table 4, a greater percentage of incorrectly classified pixels was related to wildfires (60 %), whereas polygons with clear-cuts had lower misclassification occurrences. Even though these proportions apply to all pixels and not only those with vegetation gains, it appears that the classification outcome was influenced by the fact that burned scars recover more slowly than areas that had tree cutting.

Table 4: Error distribution by disturbance processes.

| Type of disturbance | Incorrectly classified pixels (%) |
|---|---|
| No disturbance detected in 2018-2021 | 2.16 |
| Fire in 2018-2021 | 59.89 |
| Clear-cut in 2018-2021 | 37.96 |

The land cover classes with the greatest accuracy were shrubland and eucalyptus (Figure 2). The classification of maritime pine and other broadleaves, which are not abundant in the study area, were more prone to errors.

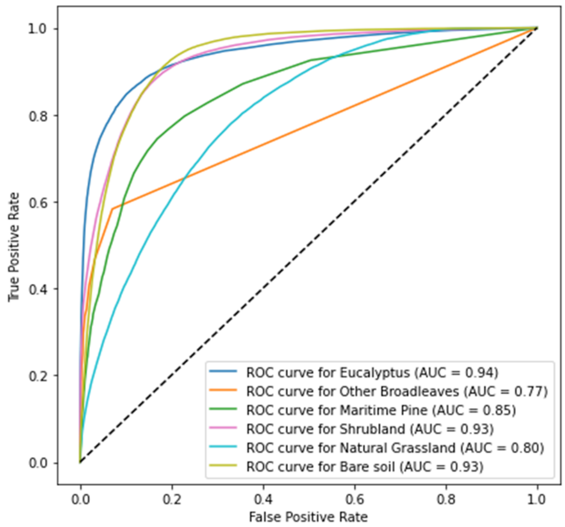


Figure 2: Classification ROC curves.

## 3.2 Dimensionality Reduction and Feature Importance

The data dimensionality reduction approach led to the reduction of the original dataset of 124 variables to meaningful groups of just 35 variables, which are listed in Figure 3. For instance, the first group includes monthly vegetation indices (NDVI) for all spring and summer months, while the "Visible summer” group contains all Sentinel-2 visible bands for the summer months. Considering the variable selection approach, our results show that NDVI during spring and summer months, prior land cover classes (particularly COSc2018), the length of the CCDC's last segment and the number of valid breaks had the larger importance for classification (Figure 3). Indicators from the CCDC algorithm and spectral information stand out as those with the greatest contribution to land cover monitoring.

Generally, the variables corresponding to winter and fall months were of less importance. Transition probabilities based on Markov Chains appeared with low importance and those that seem to have been the most informative for the model correspond to class change to other broadleaves and eucalyptus, two classes with spectral similarity issues.

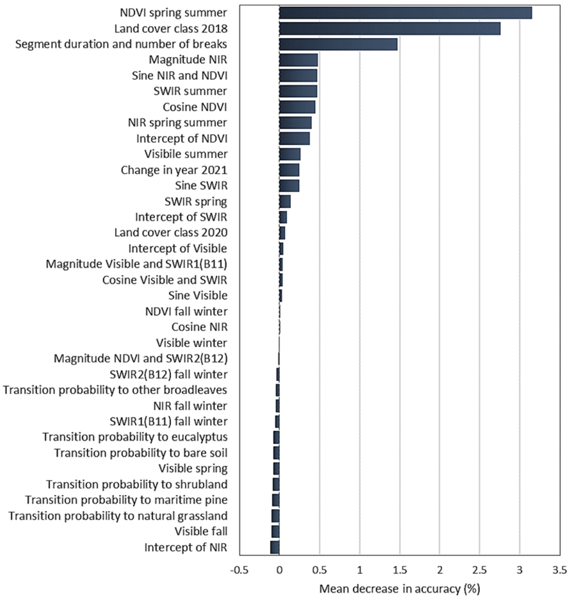


Figure 3: Permutation importance given by the mean decrease in cross-validation estimated accuracy.

## 3.3 Spatial Error Distribution

The spatial distribution of classification errors highlights the predominance of misclassifications related to smaller polygons (Figure 4) and pixels near polygon boundaries (Figure 5). In other words, the model performed better on larger polygons and in pixels distant from borders (i.e., less heterogeneous). Despite the error rates being lower in non-forest classes (Figure 2), their higher prevalence in the study area caused the most misclassifications. More specifically, the absolute number of misclassifications occurred predominantly in the classes of natural grassland (23.69 %), shrubland (34.90 %), and bare soil (29.78 %).

Polygons with high land cover heterogeneity and diverse types of forest species at distinct stages of evolution appear to have penalized the model since several plots with error patches have been observed as in Figure 6. This figure includes several polygons that were misclassified as shrubland, suggesting that our model tended to select that class over other classes with a similar spectral signal. It also illustrates an error type that may arise from the fact that photointerpretation sampling had a minimum mapping unit, which resulted in the inclusion of multiple land covers inside the same polygon. In the example in Figure 6, a eucalyptus stand is dominant in a polygon but it is mixed with shrubland leading to incorrect classification. The same explanation is valid for other classes that exhibit a large enough degree of heterogeneity.

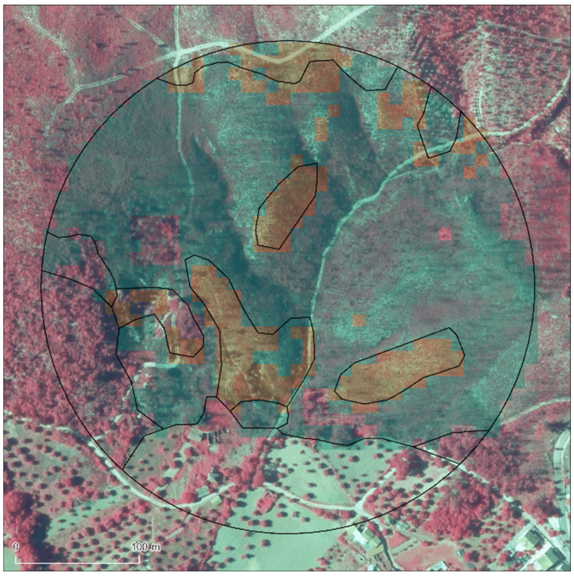

Figure 4: Small polygons error pattern (green – correctly classified; brown – classification error). The base map is a DGT Orthophoto for 2021.

We stress that the validation partitioning approach that was followed to obtain the results described above ensured that pixels in the same polygon could not be used for both training and testing, mitigating spatial autocorrelation and potential overestimation of classification accuracy.

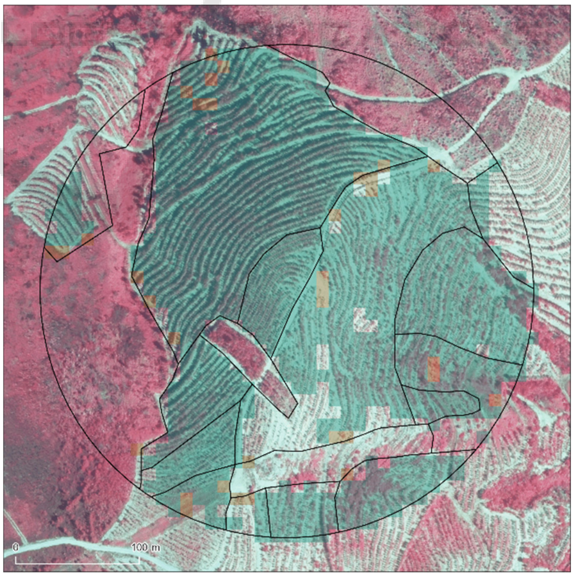

Figure 5: Boundary error pattern (green – correctly classified; brown – classification error). The base map is a DGT Orthophoto for 2021.

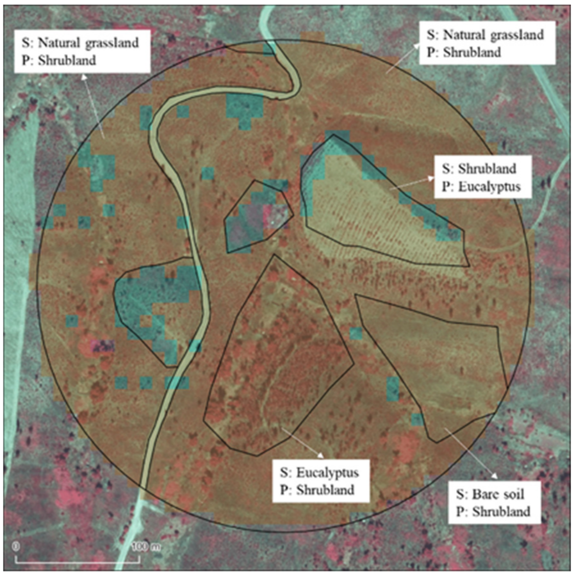


Figure 6: Mixed polygons error pattern (green – correctly classified; brown – classification error; S – sample; P - prediction). The base map is a DGT Orthophoto for 2021.

# 4 DISCUSSION

In this research, we have explored a set of variables that can potentially improve the COSc annual map classification and also reduce the manual work in the post-classification stage. A multidimensional dataset including time-series spectral and change detection information, historical land cover, and transition probabilities was created for this purpose. We demonstrated how this methodological approach yielded insights about information useful for a better land cover change prediction.

## 4.1 Major Findings

Costa et al. (2022b) concluded that the accuracy of the semi-automatic COSc thematic map for 2018 for our study area (tile 29TNE) is close to 75 %. Even if the results are not fully comparable (Costa et al., 2022a), our findings suggest that it is possible to obtain a similar accuracy using an automatic approach by adding new variables to the COSc workflow.

We showed (see Table 3) that for pixels where no recent changes had been identified the accuracy is highest (77 %) even if we restricted our analysis to areas where evidence of change exists for the period 2015-2021 according to the COSc workflow. This points out that automatic classification tends to have higher accuracy when no recent change occurs, which is expectable for change detection methods like CCDC that perform better with a sufficiently long series of observations to model the land cover signal after a change. Since classification tends to be more accurate for stable land covers, the obtained accuracy results are likely to underestimate overall accuracy since our reference dataset has a proportion of changed land cover (50.3 %, according to the photointerpreters) much higher than the average proportion of changed pixels for Portugal over 2018-2020 which is 5.4 % according to Costa et al. (2022a).

Change information, which Wulder et al. (2018) stated as an essential component of modern land cover monitoring, was especially pertinent to discriminate the classes under study. CCDC variables' high importance confirms the potential of this change detection technique to generate pertinent data for land cover classification. The relevance of CCDC outputs for classification has already been demonstrated in LCMAP production (Xian et al., 2022). However, the major novelty of our work in this domain is the use of a short time series of Sentinel-2 images, while most work implementing CCDC relied on long series of Landsat data (Franklin et al., 2015; Xian et al., 2022; Xie et al., 2022). The importance of the temporal component in the classification process, considered by Gómez et al. (2016) as indispensable in the current state of EO, was confirmed by the CCDC because the coefficients used were of the last segment, i.e., information after the last disturbance. Other variables informed about trajectory parameters and ecological succession. The historical information revealed itself to be especially useful for the model to learn the class that follows, both in terms of vegetation losses and gains. In terms of spectral information, the months that mark the biggest difference in vegetation greenness of the land cover classes (spring and summer) were the most important. This outcome is not difficult to comprehend for the Portuguese Mediterranean climate since the interannual greenness variability of some classes is more pronounced at the end of summer, when the maximum dryness is reached, and at the beginning of fall, when the growth in greenness resumes. Transition probabilities were the least important data dimension. Its limited contribution may be attributed to its exclusive consideration of the study area's overall transition probabilities, thereby failing to distinguish contextual characteristics of pixel sets.

The spatial distribution of misclassifications reflects essentially three situations: boundary pattern, small polygon, and heterogeneous polygon. Boundaries between different geographic features can cause diverse spatial patterns in the occurrence of errors. Since the study samples comprised polygons with different shapes and sizes, the characteristics of these polygons (such as their size, shape, and spatial arrangement) influence the error patterns (Corcoran et al., 2015; Moraes et al., 2021). In addition, due to the minimum mapping unit of the polygons, a certain level of spectral diversity was inevitable, which made it more difficult to classify polygons with more intricate land cover patterns.

Regarding the classification errors by land cover types, the phenological profiles of natural grassland and shrublands have common behaviours and spectral similarities. The model overestimated the number of shrubland pixels however, this class had high accuracy. Because we are dealing with a dynamic environment, the transition from herbaceous to shrubland can be gradual and undetectable. As a result, the model underperformed in classifying natural grassland, with more than 40 % of pixels misclassified as shrubland or bare soil. Since abrupt changes (losses) were better classified than gains (slower transitions), there is still a need to investigate a wider collection of metrics that can assist in monitoring more subtle changes and longer succession processes.

## 4.2 Limitations and Future Development

The ultimate goal of this research is to improve the COSc workflow for the whole Portuguese territory. Our study area was limited to a disturbance-prone region previously identified by the COSc processing steps. In future work, we will explore the suitability of the new candidate variables not only for classification but also for identifying pixels for map updating.

Also, our approach was limited in describing long-term ecological succession. The Markovian transition probabilities, which we had anticipated to be highly informative due to their importance in previously cited studies, were found to have relatively low levels of significance. This is partly because the Markov Chain processes are memoryless, meaning that the land cover transition probabilities to 2021 were independent of the previous classes in instances before the period 2018-2020. Although with residual importance, Figure 3 shows that the two most important probabilities from the six variables were transitions to other broadleaves and eucalyptus. These two classes are often difficult to distinguish in the COSc mapping due to their spectral similarities and the presence of mixed forest patches in the study area, which suggests that this type of stochastic transition element could be explored to discriminate forest species. If these probabilities are complemented by additional information, such as per-pixel vegetation growth suitability, their relevance may increase for dealing with the unique context of some pixel sets.

Future work should focus on extending these experiments to a larger area and with a higher number of land cover classes. Additional spectral bands relevant to land cover monitoring, such as the red-edge, should be explored. It may also be crucial to assess potential accuracy gains from the use of a classifier with spatial dependency or the incorporation of variables with contextual information from neighbouring pixels.

# 5 CONCLUSIONS

The main goal of this work was to identify new variables to improve the automatic steps of the COSc workflow. This was accomplished by testing a set of variables across multiple dimensions, including previous land cover data, transition probabilities, spectral and change detection information. The analysis applied to a fire-prone area showed that some variables not yet included in the COSc workflow, specifically land cover classes in previous years and change detection information produced by the CCDC, have a high potential for improving the classification. It was discovered that, when combined with Sentinel-2 temporal composites, CCDC coefficients have high importance, particularly the duration of the last segment post-disturbance and the number of breaks in the study period. Variable selection combining hierarchical clustering and PCA effectively resulted in a more parsimonious model without compromising classification performance.

# ACKNOWLEDGEMENTS

We thank the Directorate General for Territory's young research fellows for their effort in drawing and identifying the different types of land cover in polygons.

# FUNDING

This research was conducted under the collaboration contract DGT-ISA 261/2021 with funding from Compete2020 (POCI-05-5762-FSE-000368), supported by the European Social Fund, and Centro de Investigação em Gestão de Informação (MagIC), Project UIDB/00239/2020 (Forest Research Centre), both supported by the Portuguese Foundation for Science and Technology (FCT).

## REFERENCES

Abercrombie, S. P., & Friedl, M. A. (2016). Improving the Consistency of Multitemporal Land Cover Maps Using a Hidden Markov Model. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(2), 703–713. https://doi.org/10.1109/TGRS.2015.2463689

Alves, A., Marcelino, F., Gomes, E., Rocha, J., & Caetano, M. (2022). Spatiotemporal Land-Use Dynamics in Continental Portugal 1995–2018. *Sustainability, 14*(23), 30. https://doi.org/10.3390/su142315540

Balata, D., Gama, I., Domingos, T., & Proença, V. (2022). Using Satellite NDVI Time-Series to Monitor Grazing Effects on Vegetation Productivity and Phenology in Heterogeneous Mediterranean Forests. *Remote Sensing*, *14*(10), 2322. https://doi.org/10.3390/rs14102322

Bartels, S. F., Chen, H. Y. H., Wulder, M. A., & White, J. C. (2016). Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. *Forest Ecology and Management*, *361*, 194–207. https://doi.org/10.1016/j.foreco.2015.11.015

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1), 251. https://doi.org/10.1038/s41597-022-01307-4

Brown, J. F., Tollerud, H. J., Barber, C. P., Zhou, Q., Dwyer, J. L., Vogelmann, J. E., Loveland, T. R., Woodcock, C. E., Stehman, S. V., Zhu, Z., Pengra, B. W., Smith, K., Horton, J. A., Xian, G., Auch, R. F., Sohl, T. L., Sayler, K. L., Gallant, A. L., Zelenak, D., … Rover, J. (2020). Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach. *Remote Sensing of Environment*, *238*, 111356. https://doi.org/10.1016/j.rse.2019.111356

Cai, S., Liu, D., Sulla-Menashe, D., & Friedl, M. A. (2014). Enhancing MODIS land cover product with a spatial–temporal modeling algorithm. *Remote Sensing of Environment*, *147*, 243–255. https://doi.org/10.1016/j.rse.2014.03.012

Corcoran, J., Knight, J., Pelletier, K., Rampi, L., & Wang, Y. (2015). The Effects of Point or Polygon Based Training Data on RandomForest Classification Accuracy of Wetlands. *Remote Sensing*, *7*(4), 4002–4025. https://doi.org/10.3390/rs70404002

Costa, H., Benevides, P. J., Moreira, F. D., & Caetano, M. R. (2022a). Detection and classification of changes in agriculture, forest, and shrublands for land cover map updating in Portugal. Em C. M. Neale & A. Maltese (Eds.), *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXIV* (p. 19). SPIE. https://doi.org/10.1117/12.2636127

Costa, H., Benevides, P., Moreira, F. D., Moraes, D., & Caetano, M. (2022b). Spatially Stratified and Multi-Stage Approach for National Land Cover Mapping Based on Sentinel-2 Data and Expert Knowledge. *Remote Sensing*, *14*(8), 1865. https://doi.org/10.3390/rs14081865

Franklin, S. E., Ahmed, O. S., Wulder, M. A., White, J. C., Hermosilla, T., & Coops, N. C. (2015). Large Area Mapping of Annual Land Cover Dynamics Using Multitemporal Change Detection and Classification of Landsat Time Series Data. *Canadian Journal of Remote Sensing*, *41*(4), 293–314. https://doi.org/10.1080/07038992.2015.1089401

García, M., Moutahir, H., Casady, G., Bautista, S., & Rodríguez, F. (2019). Using Hidden Markov Models for Land Surface Phenology: An Evaluation Across a Range of Land Cover Types in Southeast Spain. *Remote Sensing*, *11*(5), 507. https://doi.org/10.3390/rs11050507

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Gong, W., Fang, S., Yang, G., & Ge, M. (2017). Using a Hidden Markov Model for Improving the Spatial-Temporal Consistency of Time Series Land Cover Classification. *ISPRS International Journal of Geo-Information*, *6*(10), 292. https://doi.org/10.3390/ijgi6100292

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2018). 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. https://doi.org/10.1080/07038992.2018.1437719

Liu, C., Song, W., Lu, C., & Xia, J. (2021). Spatial-Temporal Hidden Markov Model for Land Cover Classification Using Multitemporal Satellite Images. *IEEE Access*, *9*, 76493–76502. https://doi.org/10.1109/ACCESS.2021.3080926

Marcel Buchhorn, Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing*, *12*(6), 1044. https://doi.org/10.3390/rs12061044

Moraes, D., Benevides, P., Costa, H., Moreira, F. D., & Caetano, M. (2021). Influence of Sample Size in Land Cover Classification Accuracy Using Random Forest and Sentinel-2 Data in Portugal. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4232–4235. https://doi.org/10.1109/IGARSS47720.2021.9553924

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., & Duchesnay, É. (2011). Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.*, *12*, 2825–2830.

Reis, M. S., Dutra, L. V., Escada, M. I. S., & Sant'anna, S. J. S. (2020). Avoiding Invalid Transitions in Land Cover Trajectory Classification With a Compound Maximum a Posteriori Approach. *IEEE Access*, *8*, 98787–98799. https://doi.org/10.1109/ACCESS.2020.2997019

San-Miguel-Ayanz, J., Oom, D., Artes, T., Viegas, D., Fernandes, P., Faivre, N., Freire, S., Moore, P., Rego, F., & Castellnou, M. (2020). Forest fires in Portugal in 2017. In *Science for Disaster Risk Management 2020: Acting today, protecting tomorrow*. Publications Office of the European Union. https://doi.org/10.2760/571085

White, J. C., Hermosilla, T., Wulder, M. A., & Coops, N. C. (2022). Mapping, validating, and interpreting spatio-temporal trends in post-disturbance forest recovery. *Remote Sensing of Environment*, *271*, 112904. https://doi.org/10.1016/j.rse.2022.112904

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284. https://doi.org/10.1080/01431161.2018.1452075

Xian, G. Z., Smith, K., Wellington, D., Horton, J., Zhou, Q., Li, C., Auch, R., Brown, J. F., Zhu, Z., & Reker, R. R. (2022). Implementation of the CCDC algorithm to produce the LCMAP Collection 1.0 annual land surface change product. *Earth System Science Data*, *14*(1), 143–162. https://doi.org/10.5194/essd-14-143-2022

Xie, S., Liu, L., Zhang, X., & Yang, J. (2022). Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack. *ISPRS Journal of Photogrammetry and Remote Sensing*, *185*, 201–218. https://doi.org/10.1016/j.isprsjprs.2022.01.014

Zhu, Z., Qiu, S., & Ye, S. (2022). Remote sensing of land change: A multifaceted perspective. *Remote Sensing of Environment*, *282*, 113266. https://doi.org/10.1016/j.rse.2022.113266

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152–171. https://doi.org/10.1016 /j.rse.2014.01.011