

Inter-annual land cover mapping

**An approach for producing inter-annual maps
by implementing scene based and pixel-based
composites using Sentinel 2 imagery and an
automatic labeling procedure for classification
based on old maps.**

Dissertation supervised by

Dr. Mario Caetano

Professor, Nova Information Management School

University of Nova Lisbon, Portugal

Dr. Edzer Pebesma

Professor, Institute for Geoinformatics

University of Münster, Germany

Dr. Jorge Mateu

Professor, Department of Mathematics

University of Jaume I Castellón, Spain

January 30, 2019

Abstract

Land Use Land Cover (LCLU) is very well established as one of the most efficient approaches for monitoring land cover and its changes (Gómez et al., 2016). Estimating the dynamic of the land cover system under different temporal scales may contribute to improve our understanding of its interaction with different biophysical process according with their temporal variation (Yang et al., 2013). In this thesis, we investigated the viability of producing inter-annual maps by implementing scene-based and pixel-based composites using sentinel 2 imagery, and an automatic labeling procedure for the classification based on old maps. Our results provide initial insights into the benefits and specific issues in this context. We found out that the use of pixel-base composites based on maximization of NDVI can lead to combat cloud contamination and reproduce artificial imagery with good quality for classification. Besides that, the map of reference resulted in an extensive source of training data that also contained an important fraction of noise. To address this issue, we undertook the viability of using two procedures for refining mislabel data, one based on a iterative learning procedure using Random forest and entropies, and another based in the variability of NDVI signals. The proposed methodologies are tested using Sentinel 2 imagery for 2017 and the reference COS map developed by DGT. From the results, we discussed extensively the robustness of classifiers in the presence of different levels od noise data as well as parametrization in an environment of automatic mapping.

Keywords

Composites

Informativeness

Automatic feature selection

Support vector machine and random forest

Geographical information systems

Acronyms

DGT - Direcção-Geral de Ordenamento do Território, Portugal

L1C Level-1C

L2A Level-2A

MMU Minimum Mapping Unit

SVM Support vector machine

RF Random forest

LCLU Land Cover Land Use

BAP Best available pixel

NIR Near Infrared

SWIR Short Wave Infrared

List of Figures

3.1	Study area	12
3.2	Available images Sentinel 2, 2017	14
3.3	Features for the classification process	14
3.4	Sampling dataset COS, 2015	15
3.5	Class definition of the two datasets and COS nomenclature	16
4.1	Methodology image preprocessing	20
4.2	Methodology composites	22
4.3	Train, test and validation split	22
4.4	Cleaning preopocesing forest data	24
4.5	NDVI herbacesu	25
4.6	Traning data distribution after applying NDVI rules	26
4.7	Iterative learning procedure to define informativeness based on boosting and the measurement of information entropy	27
4.8	Informativeness for herbaceous	27
4.9	Sealed example	28
5.1	Perfomance classifiers RF and SVM, scene-based composites 2017	30
5.2	Importance variables	31
5.3	Cleaning processing NDVI, Performance over scene-based compos- ites 2017	33
5.4	Normalized confusion matrix, classification scene-based composite July 29	34
5.5	Cleaning preprocessing using Informativeness, scene-based com- posite July 29	35

List of Figures

5.6	Cleaning preprocessing using Informativeness, scene-based composite December 16	36
5.7	Performance of the proposed methodology over a traning data with 30% of noise, scene-based composite July 29	37
5.8	Sensitivv analysis parameters random forest	38
5.9	Sensitivv analysis parameters SVM with RBF kernel	39
5.10	Performance random forest vs SVM using COS data under different levels of noise, scene-based composite July 29	40
5.11	pixel-based composite Spring	41
5.12	Performance of the classifications, scene-based vs pixel-based composites	42
5.13	Seasonal mapping.	43

List of Tables

5.1 Accuracy assessment 6 classes	44
5.2 Accuracy assessment 6 classes	44

Contents

Abstract	i
Keywords	iii
Acronyms	v
1 Introduction	1
1.1 Problem statement and Motivation	1
1.2 Objectives	3
2 Literature Review	5
2.1 Seasonal land cover mapping	5
2.2 Best pixel available and composites	6
2.3 Automatic selection of training data	6
2.4 Classification algorithms and tuning parameters	7
2.4.1 Support vector machines	8
2.4.2 Random forest	9
2.5 Accuracy assessment in land cover classification	9
3 Data and study area	11
3.1 Study area	11
3.2 Sentinel 2 Imagery	12
3.3 COS data	13
3.4 External dataset	15
4 Methodology	19
4.1 Image preprocessing	19

Contents

4.2 Seasonal Composites	20
4.3 Image classification	21
4.3.1 Distribution of the COS dataset for the modeling	21
4.3.2 Cleaning preprocessing of training data	22
5 Results	29
5.1 Classification performance	29
5.2 Cleaning procedure performance	31
5.2.1 Cleaning preprocessing using NDVI variability	31
5.2.2 Cleaning preprocessing using Informativeness	32
5.3 Parameter optimization	37
5.4 Influence of Noise in the performance of SVM and RF	39
5.5 Single date maps vs seasonal maps	40
5.6 Assessment	42
Bibliography	45
APPENDIX A - Maps composites	51
APPENDIX B - Single Maps	53

1 Introduction

1.1 Problem statement and Motivation

Land Use Land Cover (LCLU) is very well established as one of the most efficient approaches for monitoring land cover and its changes ([Gómez et al., 2016](#)). Estimating the dynamic of the land cover system under different temporal scales may contribute to improve our understanding of its interaction with different biophysical process according with their temporal variation ([Yang et al., 2013](#)). In this context, increasing the observation of LULC by new technology as it is improved the systematic production of LULC maps can lead to address with more consistency the analysis of processes with inter-annual variability, such as fire propagation ([Navarro et al., 2017](#)), crop production ([Vuolo et al., 2018](#)) and climate ([Bontemps et al., 2012](#)).

For the regular characterization in time of the dynamic of the land system by using time series, the increase of dimensionality by the fact of an increase in the number of images to study, the possible differences in accuracy of the products and the clouds persistence in certain periods may determine special constraints in its application ([Lu et al., 2004](#)). With the increase of the revisit of observations of Sentinel 2, the approach of the best pixel available (BPA) offer new opportunities to produce imagery with regular frequency as well as better spectral features than can lead to perform better results in the task of classification ([Gómez et al., 2016](#)). Pixel- based composites have been developed by using different kind of protocols, which mainly depend on the use of NDVI values and distances to the masked clouds to define the BPA ([Hermosilla et al., 2015](#)). In this context, based on experiments of [Holben \(1986\)](#) with VHRR time series. That is, making

1 Introduction

composites by maximizing the NDVI per pixel in an arrange of several scenes in order to capture the state of the vegetation when is more photosynthetically active, we propose to make seasonal composites as a case of study. Besides that, this thesis aims at evaluating the viability of implementing a maximum value composites for classification, by retaining not only maximum NDVI per scene but also the rest of the spectral information associated to the index of the pixel with the highest NDVI.

Besides of giving special interest to the imagery in the dynamic of land cover at inter-annual time frame, the technique of classification may be also fundamental. Support Vector Machine (SVM) and Random Forest (RF) represent state of the art algorithms for its application in the production of LULC ([Thanh Noi and Kappas, 2018](#)); important for their ability to handle high dimensionality, being superior to unsupervised approaches and being insensitive to overfitting ([Bishop, 2006](#)). However, the performance of the supervise algorithms essentially depends on the quality of the labeled data for training ([Tuia et al., 2009](#)). Generally, the strategy for collecting labeled datasets consists of using interpreted data from very high spatial resolution satellite images or aerophotographies. Nonetheless, in most cases this selection turns out an expensive task not operationally efficient in the production of maps under the inter-annual reference ([Inglada et al., 2017](#)).

The availability of previous maps in the study area represent also an important reference ([Colditz et al., 2011](#)), and therefore an effective method for automation of collecting training data. However, even though they can represent a rich source of information, this data may content errors ([Pelletier et al., 2017b](#)). In this context, source of errors can be explained by 1: maps are a generalization of the land cover system, so random samples can fall over complexities of a wide diversity of classes that were simplified in one class in the map, 2: expected changes in the land cover due to the gap between the image acquisition and the date of production of the reference map, and 3: to phenology that can be translated in natural changes of certain communities of vegetation during the year that lead to consider temporarily in the classes. Although support vector machines and random forest are also known for being resistant to anomaly data, a

classifier trained in a set of large amount of wrong labels can turn out in a wrong model ([Pelletier et al., 2017a](#)). Therefore, this thesis aim at refining the sampling by exploring the viability of implementing two cleaning procedures; one based on NDVI signal variability and other based on an iterative learning procedure that uses boosting and the measurement of information entropy.

Therefore, to continue advancing in the state of the art of inter-annual land cover mapping, is therefore important to know how the aforementioned proposals are going to be explored . In this context, this thesis aims at performing classification by using scene-based and pixel-based composites per season of Sentinel 2 images from January 2017 to December 2017 at central of Portugal. Besides that, the implementation of the automatic labeling procedure for supervised classification is based on the reference COS map 2015 developed by DGT.

1.2 Objectives

This thesis aims to research four questions:

- How do the classification approach based on multi-spectral data works over the year 2017 using random forest and support vector machine classifiers?
- Do a refining procedure of mislabeled data can lead to improve the performance of a scene classification.
- How usable is COS dataset in the classification task using multi-temporal and multi-spectral data from sentinel 2 imagery?
- Do a pixel-based image composites analysis achieve better results in classification than a scene based analysis?

2 Literature Review

This thesis aim at making inter-annual maps by using scene-based and pixel-based composites of Sentinel 2 imagery per season and an automatic labeling procedure for supervised classification based on old maps. Therefore, in this chapter we present an overview on the relevant literature related with each proposal. In the Section 2.1, we will discuss about current research in land cover mapping with especial emphasis in the needs for inter-annual mapping. In section 2.2, we will give an overview of what pixel-based composites means. In Section 2.3, we will introduce the theory of automatic selection of training data and the review of two techniques to filter the reference dataset against noise. Finally, the Section 2.4 present the fundamentals of RF and SVM.

2.1 Seasonal land cover mapping

LCLU is well established as environmental variable that can contribute to a better understanding of biophysical problems and therefore to improve the ability to model and predict climate change (Gómez et al., 2016). Increasing the observation of LCLU by new technology as it is improved the systematic production of LCLU maps can lead to address with more consistency the analysis of processes with inter-annual variability, such as fire propagation (Navarro et al., 2017), crop production (Vuolo et al., 2018) and climate (Bontemps et al., 2012). While the production of maps with inter-annual regularity can lead to solve incompatibilities with the temporal stability of certain climate process, it also can benefit the production of maps with better accuracy specifications (Hermosilla et al., 2015). In (Bontemps et al., 2012) is carried out an interesting survey about the needs of

2 Literature Review

users in terms of satellite image products. One of the established requirements is the integration of the dynamic dimension at the inter-annual and seasonal levels. Monitoring land cover in the inter-annual time frame can be efficiently performed by satellite image processing ([Yin et al., 2014](#)). With the opening free access to the sentinel 2 archive, the research benefits with a steady increase of a database that consider high spatial, spectral and temporal resolution, which makes Sentinel 2 a valuable source of data for inter-annual monitoring as never before.

2.2 Best pixel available and composites

The approach of pixel-based composites provides the means to reproduce cloud-free and phenological consistent image composites ([Gómez et al., 2016](#)). Composites have been developed by using different kind of protocols, which mainly depend on the use NDVI values and distances to the masked clouds to define the best pixels available (BPA) ([Hermosilla et al., 2015](#)). The first examples of this approach obeys to applications in the 80s using AVHRR and MODIS images ([Holben, 1986](#)). Essentially, the methods were based on maximizing NDVI or minimizing view angle to select the best observation for a given pixel within a specified compositing period. The limitations derived by the fact of the coarse spatial resolution of sensors that accounted for high temporal resolution. However, with the opening of Landsat and Sentinel 2 archive, the generation of pixel-based composites has become technically feasible. New attempts for the selection of the best pixel aim at implementing diverse rules and including different kind of sensors at the same time. For example, a score protocol proposed in ([White et al., 2014](#)), seek to weight every pixel of the time series according with their proximity to the clouds, date of analysis and type of sensor.

2.3 Automatic selection of training data

The strategy to obtain automatic training data for the process of classification resides in the idea of extracting labels of available governmental databases or

old maps (Inglada et al., 2017). The processing turn out being fast and efficient since leads to the production of the map immediately after the image is released. However, the use of that reference data can be constrained by different reasons as discussed in (Pelletier et al., 2017b). For example, the gap between the production of the map and the image acquisition, or assumptions of spatial continuity in the reference map. The automatic selection relay in the robustness of the classifiers to label under the presence of noise, as for example RF and SVM. However, the systemic presence of erroneous labels can lead to impact negatively the results of the classification (Tolba, 2010). In this context, only few works has addressed the importance of cleaning the training dataset used for classification of time series (Pelletier et al., 2017b). These works are mainly based on iterative learning process that look for scoring the labels according with their informativeness. For example, Büschlenfeld and Ostermann (2012) optimizes the performance of SVM in the classification of 4 classes (Grassland, Forest, Industry, settlement) using Ikonos imagery by using the uncertainty information given by the implementation of SVM. The labels are scored according with probability estimates and proximity of the label to region with the same class (Wu et al., 2004). Therefore, samples with high uncertainty in the predictions are removed to further improve the refinement. Another example is implemented by (Pelletier et al., 2017b) to describe five vegetation classes: straw cereals, maize, sunflower, vineyard and grassland using Sentinel 2 imagery. In this context, it was used the bootstrap mechanism of RF to evaluate the disagreement of different trees in the prediction of certain label. Therefore, a score for each sample is defined based on a similarity measure between instances belonging to the same class.

2.4 Classification algorithms and tunning parameters

Machine learning algorithms as Random forest and support vector machines are widely used into remote sensing for their special properties of being able to handle high data dimensionality while being insensitive to over-fitting (Mather and Tso,

2 Literature Review

2016). Every classifier has their own type of parameters and versatility to handle different kind of data. Defining which one is better than other may depend on the data and the problem domain (Thanh Noy and Kappas (2018), Vuolo et al. (2018)). Therefore, to achieve an optimal classification we will require eventually an optimization of their parameters. In the following two sections, we will explain which parameters will be optimized and why they matter in the prediction.

2.4.1 Support vector machines

SVM algorithm is a popular supervised classifier that has been widely used in different domains. In the context of remote sensing, it has been extensively used for its good performance (Mountrakis et al., 2011). Particularly, SVM minimize the error of classification by creating an hyperplane among every set of classes, so that it maximizes the distance between the support vectors of every class. The parameter that controls the margin of the hyperplane is called C and generally as higher its value, the better performance for the training data, but with the risk of losing generalization for unseen data. Conversely, a low value of C will neglect possible outliers in the training data, and thus gaining more versatility to over-fitting.

Since the data may be not linearly separable in the original dimension, the separation is done in a higher spectral space controlled by a kernel. RBF kernel is commonly used for its good results (Shi and Yang (2015), Pelletier et al. (2017a)). Therefore, in this thesis will be implemented. The SVM with RBF will require the optimization of a second parameter, γ , that control the shape of the Gaussian kernel function and thus how much jagged or soft the decision boundary will be. The reason by which eventually an optimization of this parameter will matter is because for high estimations of values of γ , we can turn out with a model that works properly for the training dataset, but losing generalization for unseen data.

2.4.2 Random forest

The fundamental idea behind RF is the construction of hundreds of decision trees, considered weak learners, which are then combined to transform them into a strong learner. Its implementation depend on setting up two parameters: number of trees $ntree$ and number of variables randomly sampled as candidates at each split, $mtry$. In several studies, the default parameters of RF, that is 100 trees and $mtry = \sqrt{p}$ (where p is the number of variables), lead in average to the best performance of the classifier.

According with [Breiman \(2002\)](#), the algorithm works as follows: 1) draw $ntree$ bootstrap samples; 2) For each bootstrap sample, grow and un-pruned tree by choosing best split based on a random sample of $mtry$ predictors at each node; 3) Predict new data using majority votes for classification.

2.5 Accuracy assessment in land cover classification

Generally, the judgment of the quality of LCLU maps depends on the evaluation of the derived map against some ground or reference data for validation. In thematic mapping the map quality is function of the degree of correctness of the map that usually is interpreted as accuracy ([Foody, 2002](#)). Accuracy standards can be diverse, from subjective perspectives as the visual appraisal of the final maps to more objective assessments, such as accuracy metrics based on comparisons of the class labels in the thematic map and ground data. Accuracy may be undertaken for different reasons. For example, the general evaluation of the quality of the map or a base for evaluating the performance of different algorithms in the classification ([Congalton and Green, 2008](#)).

According with [Foody \(2002\)](#), the confusion matrix is currently the core of the accuracy assessment in the literature. This matrix consist of a cross-tabulation with the percentages of labels correctly classified. The matrix also provides the means to analysis intraclass confusion, so that it may help studies to pay special attention in the performance of the classification over specific classes. Many metrics of classification accuracy can be derived from a confusion matrix ([Foody,](#)

2 Literature Review

2002). For example, the simple overall accuracy (OA) that determines the total percentage of classes correctly classified; that is, dividing the number of labels correctly classified into the total number of labels. Its simplicity makes it useful for a huge spectrum of applications (Pelletier et al., 2017a). However, in particular for this thesis we highlight its importance for working as base on the comparison of the performance of different algorithms in classification.

In this context, the simplicity of OA may imply its major problem since some users argue that there may be cases where the correct classes were purely classified by chance (Congalton and Green (2008), Pontius (2000)). Therefore, to make balance of the effects of chance agreement, the Cohen's kappa is proposed. Unlike OA, that ranges between 0 and 1, kappa varies between -1 and 1. Even though the kappa coefficient has important features as metric to calculate classification accuracy, its adaptation as standard measurement continue being a debate due to its ability to allow the evaluation of the differences in accuracy is not unique among the accuracy metrics.

3 Data and study area

In this chapter, the study area will be introduced (Section 3.1), as well as the S2 imagery and COS dataset used to conduct this research (Section 3.2). Moreover, in the section 3.3, we will introduce a external dataset developed by DGT, that eventually it will be used to test results.

3.1 Study area

The experiments presented in this thesis are carried out in Lisbon in collaboration with the Direção-Geral de Ordenamento do Território (DGT engl. Directorate-General for the Territorial Development) and direct supervision of Professor Mario Caetano. One S2 tile that overlap the center of Portugal was chosen as study area. According with the figure 3.1 the study area covers an area of 100km by 100km that particularly is intersected by Tagus river in the region of Santarém and Vila Franca de Xira at the head of the long narrow estuary. According with Corine Land Cover 2012, the study area covers a wide variety of land cover types, such as urban fabric, agricultural areas, water-bodies, semi natural areas and forest. The landscape shaped by the river and thus an emerge of fertile lands leads to the neighbor communities to set agricultural practices. Therefore, the wide diverse in the agricultural practices and phenology activity correspond to an indicate scenario where this thesis can be conducted.

3 Data and study area

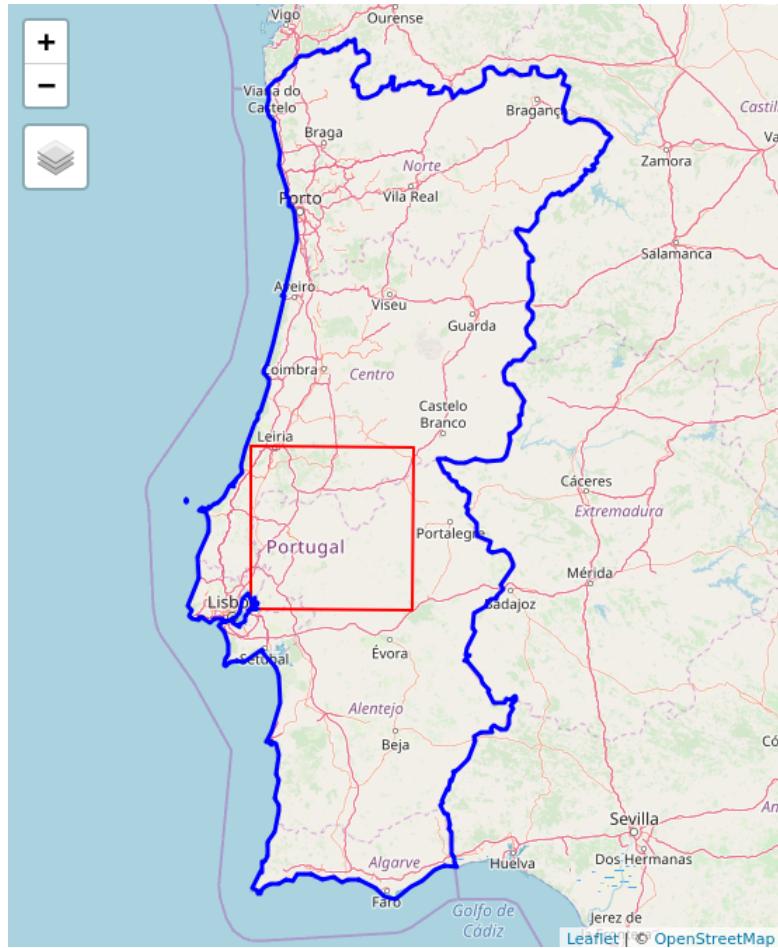


Figure 3.1: Study area

3.2 Sentinel 2 Imagery

This thesis used Sentinel-2 image time series acquired on 2017. The time series, composed of 20 images, is described by the ten Sentinel-2 spectral bands at 10 and 20 meters resolution. The images were downloaded by using the online system Copernicus Open Access Hub developed by ESA. The imagery was completely free. An overview of the images under study, percentage of clouds and level of the available product is shown in the figure 3.2. Data is derived from the S2A and S2B sentinel missions. The lower amount of imagery at the beginning of the year obeys to the release of images from only the first mission; having images every five days was only possible until the second mission S2B was launched on March 2017. Moreover, we distinguish in four colors the imagery used to develop

every seasonal composite and in gray color the excluded images; the exclusion of images obeyed to the high cloud contamination of the image (see percentage of clouds).

The products of Sentinel 2 came in two different levels 1C and 2A. L1C corresponds to a orthorectification of data using as reference the digital elevation model (PlanetDEM 90). Consequently, a preprocessing is carried to offer measurements of reflectance on the top of the atmosphere. Instead, L2A consist of a postprocessing of L1C product to provide reflectance measurements on the bottom of the atmosphere. In this context, products that came in level 1C were subject of preprocessing using the Sen2Cor processor developed by ESA.

Specifically, the bands used for this project were the following: 10m spatial resolution bands B2 (490nm), B3 (560nm), B4 (665nm), and B8-NIR (842nm), and the 20 m spatial resolution bands B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865nm), B11-SWIR (1610nm), and B12-SWIR (2190nm) (ESA, 2017b). The rest of the bands were exclude due to the impact that downscaling may have in the quality of the results. In this context, the spectral information is mainly characterized by channels in the visible/near infrared (VNIR) and short wave infrared spectral range (SWIR).

Finally, the 30 meters resolution of the global DEM developed by NASA was used to create the layer of Slope. Therefore, besides of the spectral data provided by Sentinel 2, we also consider the DEM and Slope after considering resampling to the MMU of analysis. Figure 3.2 shows an overview of the imagery under study for July 29 of 2017.

3.3 COS data

COS dataset corresponds to an stratified random sampling of the COS map with version 2015. COS cartography is composed by polygons, where each polygon represents an homogeneous unity of use and occupation of the soil. According with [Caetano et al. \(2015\)](#), each polygon represents any area of land greater than or equal to the defined minimum mapping unit of 1 ha, with a maximum distance

3 Data and study area

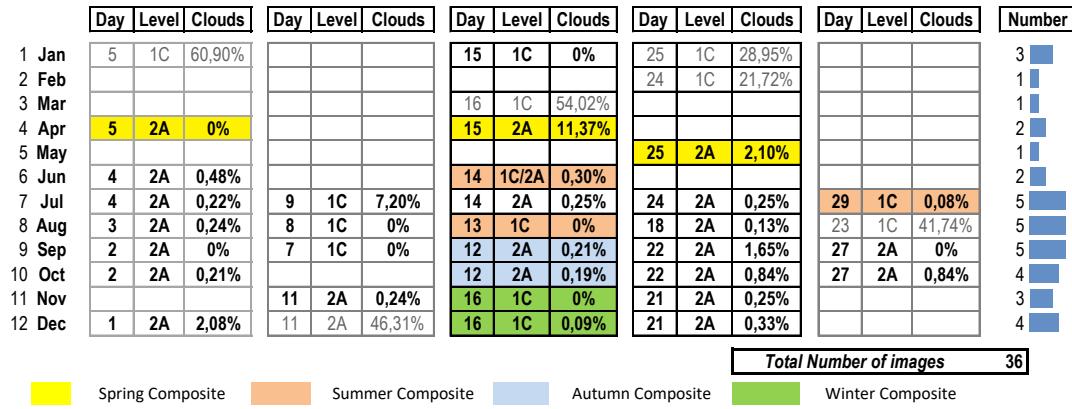


Figure 3.2: Available images Sentinel 2, 2017

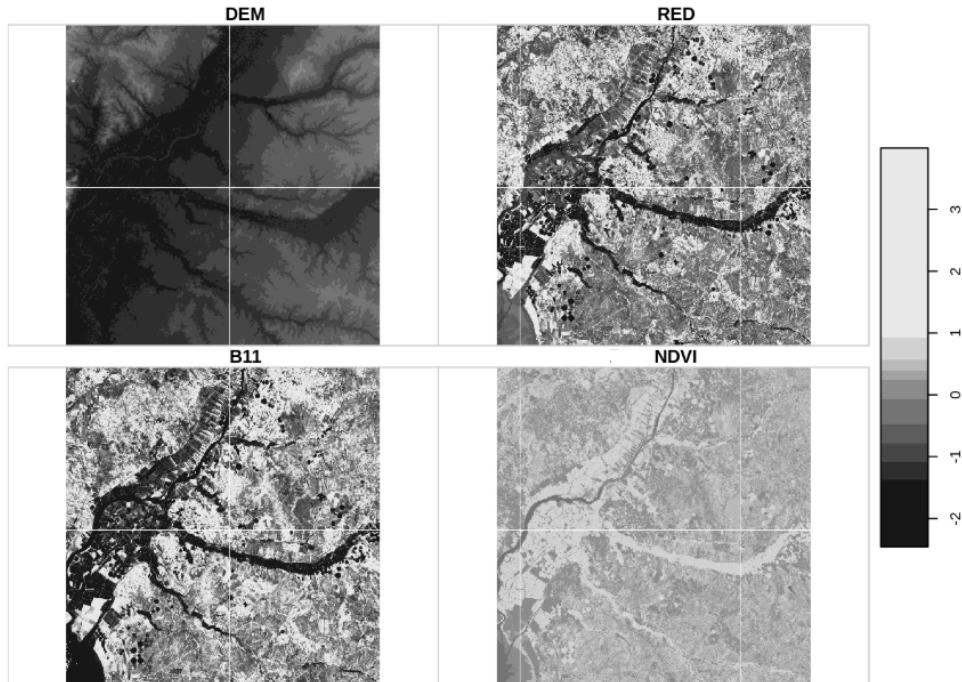


Figure 3.3: Features for the classification process

between lines or equal to 20 m and which percentage of a given class is equal or greater than 75% of the total area. So this product is subject of generalization in order to preserve spatial continuity in the final cartography.

Concerning the specification of the sampling, it consists of an stratified random sampling of 9 classes; each class contains 1000 samples. Samples were also limited to having points that were not closer than 50 meters to each other. Moreover, since this thesis also evaluates the performance of the classification under

two datasets that contains different number of classes, we made a second dataset that merged the *Coniferous*, *Holm and cork trees*, *Eucalyptus* and *Shrubs* in a simple category called *Woody*. It should be noticed, that the merging implied a posterior re-sampling to 1000 samples in order to keep in balance the total data. Besides that, table 3.5 shows per subcategory the number of samples.

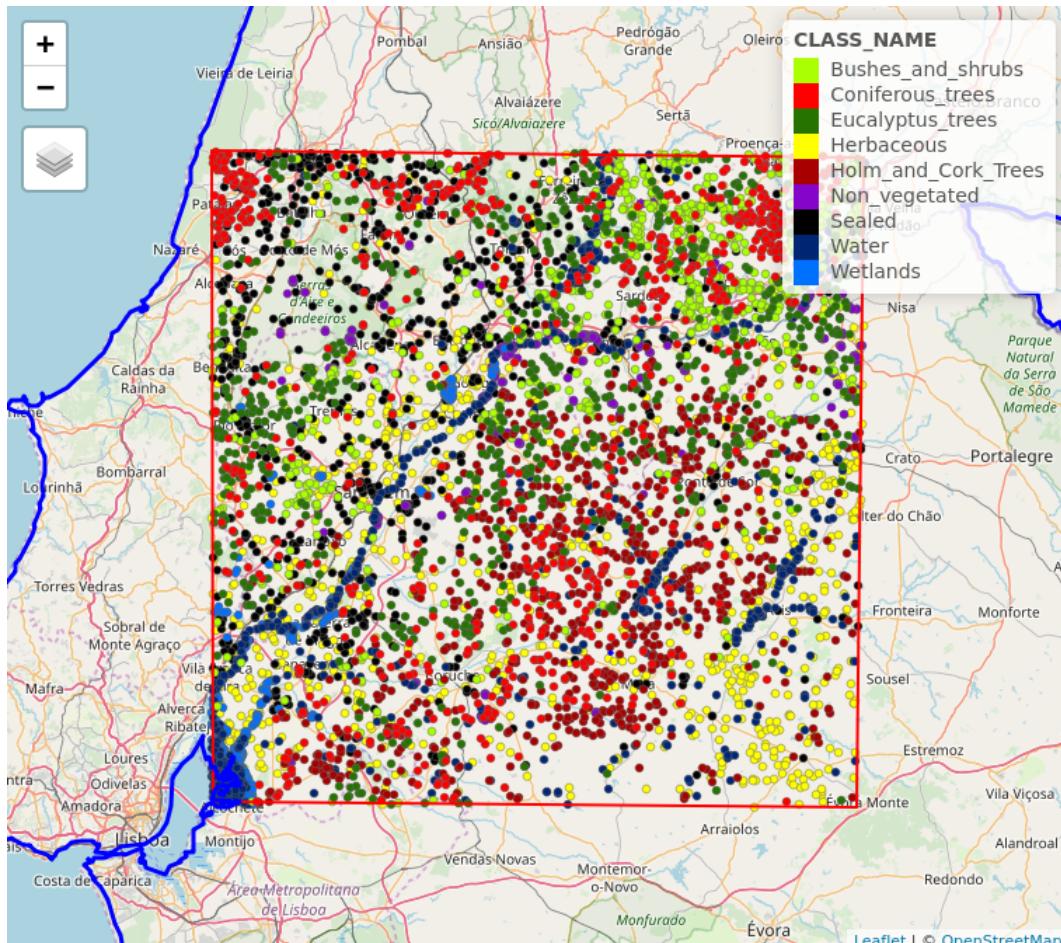


Figure 3.4: Sampling dataset COS, 2015

3.4 External dataset

Since the presence of anomaly data may be general, both in training and testing, and the cleaning processing is only done over the training, the use of a external dataset may be fundamental to compare the predictive power of the trained models against test data that reflect properly the class variation without the ex-

3 Data and study area

Dataset 1		Number	Dataset 2		Number	COS Nomenclature	COS subcategory	Number
Shrubs	1000					Shrubs and herbaceous vegetation association	Matos densos	730
							Matos pouco densos	270
Coniferous trees	1000		Woody	1000		Coniferous trees	Florestas de pinheiro bravo	752
							Florestas de pinheiro manso	248
Eucalyptus trees	1000						Florestas de eucalipto	1000
							Florestas de sobreiro	959
Holm and Cork Trees	1000					Broad leaved forest	Florestas de azinheira	41
Herbaceous	1000		Herbaceous	1000		Temporal crops	Culturas temporárias de sequeiro	205
							Culturas temporárias de regadio	295
						Pastures	Pastagens permanentes	500
Non vegetated	1000		Non vegetated	1000		Open spaces with little or no vegetation	Praias, dunas e areais interiores	383
							Rocha nua	90
							vegetação esparsa	527
							Tecido urbano continuo predominantemente vertical	33
							Tecido urbano continuo predominantemente horizontal	268
							Areas de estacionamentos e logradouros	1
							Tecido urbano descontinuo	335
							Tecido urbano descontinuo esparso	82
							Industria	123
							Comercio	8
							Instalaciones agricolas	42
							Equipamentos publicos e privados	23
							Infraestruturas de produção de energia não renováveis	1
							Infraestruturas de captação, tratamento e abastecimento de aguas para construcciones	1
							Infraestructura para tratamiento de resíduos e efluentes	12
							Rede viária e espaços associados	62
							Rede ferroviária e espaços associados	9
							Cursos de agua naturais	304
							Canais artificiais	5
							Lagos e lagoas interiores artificiais	1
							Lagos e lagoas interiores naturais	3
							Reservatorios de barragens	420
							Reservatorios de represas ou de aludes	18
							Charcas	40
							Desembocaduras fluviais	209
							Pauis	780
							Sapais	115
							Salines	105
								9000
	9000	Total		6000				

Figure 3.5: Class definition of the two datasets and COS nomenclature

position to mislabel data due to phenology. In this case, in collaboration with DGT, we had access to an external dataset that corresponds to a visual interpretation of 557 random samples using the image of 29 of July. The distribution of samples is the following: Bushes and shrubs: 61, coniferous trees:35, Eucalyptus trees: 55, Herbaceous: 217, Holm and cork trees: 38, Non vegetated:69, Sealed:

3.4 External dataset

11, Water:55, Wetlands:16.

4 Methodology

As we have established in the previous chapters, this thesis aimed at producing inter-annual maps by implementing scene-based and pixel-based composites using Sentinel 2 imagery and an automatic labeling procedure for classification based on old maps. In this context, in the section 4.1, we will introduce the methodology for the image processing; this include atmospheric and radiometric correction of the imagery as well as the production of additional features for the classification. In section 4.2, we will explain the methodology for the production of the pixel-based composites based on maximization of NDVI. Finally, in section 4.3, we will discuss about the different strategies for the classification of the images, including the automatic production of training data using old maps and the implementation of two procedures to remove noise in the training data during the modeling.

4.1 Image preprocessing

The preprocessing started with the use of Sen2Cor toolbox allowing the conversion of radiance into reflectance. Specifically, this tool provided a product L2A that considered the atmospheric, terrain and cirrus correction of Top-Of- Atmosphere of Level 1C imagery ([ESA, 2017](#)). Moreover, the process included the correction of also bands of 60 and 20 meters resolution. However, the objective of this thesis was to use only bands of 20 and 10 meters; bands of 20 meters were subject of re-sampling process at 10 meters.

Additionally to the spectral data provided by Sentinel 2, we considered the DEM and the Slope as also features for the modeling. The topography data, with 30 meters resolution, was generated by NASA'S Shuttle Radar Topography

4 Methodology

Mission (SRTM). Consequently, the slope was a byproduct made in QGIS using the DEM as reference. Once both products were obtained, they were subject of re-sampling to 10 meters in order to stack the products with the rest of bands.

Finally, we performed NDVI in order to use it also as additional feature in the task of classification and index for the construction of the seasonal pixel-based composites. Generally, NDVI is calculated from the measured radiance on the near and red infrared part of the electromagnetic spectrum. This index obeys to a process of map algebra using the equation 4.1:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (4.1)$$

Where, ρ_{NIR} correspond to band 8 and ρ_{Red} to band 4 in Sentinel 2 imagery. In the context of working with SVM, all images were subject of normalization, but NDVI. This process was repeated for each image available in 2017.

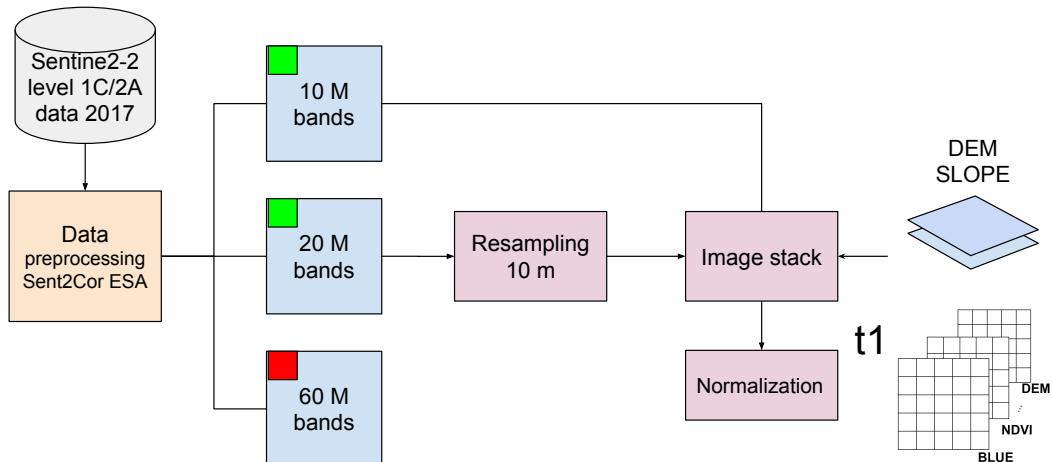


Figure 4.1: Methodology image preprocessing

4.2 Seasonal Composites

The aim at making composites was to reproduce cloud-free and phenological consistent image composites for the seasons of 2017 as a case of study of inter-annual mapping. Essentially, the proposal for the composites was based on ideas of

(Holben, 1986) that consist of retaining the maximum NDVI per scene. However, instead of working with only NDVI, the proposal sought to evaluate the benefits of retrieving the rest of the spectral information associated to the pixel with the highest NDVI in the classification. According with the figure 4.2, we show three series of images highlighting 3 pixels over the time and over different spectral components. The pixel located in the upper left corner associated to the first image contains the highest NDVI over the three images related with that position, so that the composite retrieve the NDVI and the rest of the bands for that time. This methodology was very simple, fast and depended only on the spectral information at level of pixel. The Process was evaluated in 4 suites of images per season. The table 3.3 in the Section 3.3 showed the images used for the composites. It should be noticed, that the imagery was preselected according with the seasonal precipitation conditions in Portugal during the year 2017 (see anexos).

4.3 Image classification

The pixel-based land use classification was carried out in Python to classify the features based on their spectral reflectance and physical properties. LCLU classification involved the following procedures: Distribution of the COS dataset for the modeling, cleaning preprocessing of training data, creating of classification model, accuracy assessment. The last two section are not having discussion here, since in the literature review we highlight the most important aspects about them.

4.3.1 Distribution of the COS dataset for the modeling

Concerning the distribution of the COS dataset for the modeling, this was partitioned into; training, validation and test dataset in the ratio of 55%, 25% and 20% respectively. In the figure 4.3, we highlight in a red box the data used for the implementation of the filter procedure of training data as well as the use of an additional and independent test dataset in green color.

4 Methodology

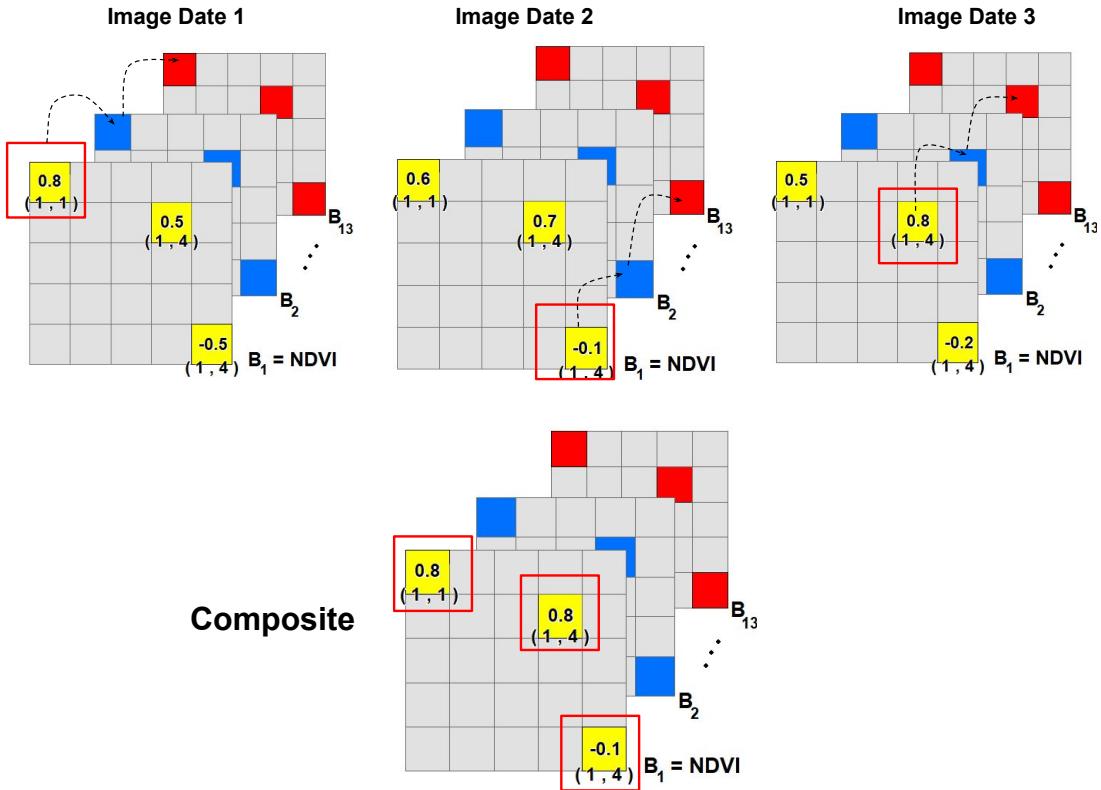


Figure 4.2: Methodology composites

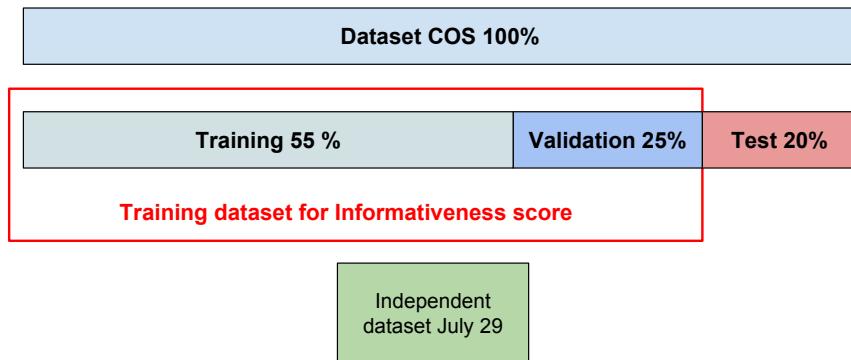


Figure 4.3: Train, test and validation split

4.3.2 Cleaning preprocessing of training data

While the training data selection is automatic (see section 3.3), this thesis aimed at refining the training data for each period of analysis in order to evaluate better accuracy in the classification. Specifically, we used two different techniques, one

based on trends of NDVI signals and another on a iterative learning procedure. On the one hand, we used signals of NDVI to explore normal range of variations per class and time of NDVI values. On the other hand, we used an iterative learning procedure that consisted of scoring the samples according with their level of informativeness for certain period, so that no informative data is separated from the training in the modeling. In the next two sections we will give more details of each approach.

Cleaning preprocessing based in NDVI signals

We recreated the NDVI values for all the imagery available in 2017, so that we used this imagery to retrieve the value of the index per sample of the training data of the classification. Later, we will make emphasis on the distribution of the data used for the modeling, however, it is important the reader understand at this stage that the cleaning procedures were made over a fraction of the data that was used to train the model. In this context, figure 4.4 shows one example of the variability of three woody classes over the year. The boxplot display interquartile ranges (IQR, size of the box), maximum and minimum values (limits of the whiskers) and possible outliers (black points) per scene. As we introduced in the motivation, labels can intersect mixture of classes and, therefore, display different spectral signatures to the usual ones. In the context of trees, depending on the canopy, the distances between trees plantation is diverse. Among the remaining spaces of the trees, other type of vegetation can grow up, such as bushes or grass. Therefore the variation of NDVI may be a result of how saturated of disperse is the biomass for certain pixel. For example, the presence of more outliers at the beginning of the year may obey that in spring most of the vegetation is photosynthetically more active, and therefore, the fractional proportion of the class tree increases in the pixel implying in a separation between what may be truly woody and grass. According with professor Mario Caetano, class trees should display NDVI values larger than 0.3, so that we decided to remove samples with NDVI values lower than this value. Moreover, as it was done in all the rest of the classes under study, the samples with NDVI values beyond 1.5 times the IQR, from the first or third

4 Methodology

quartile were removed from the data.

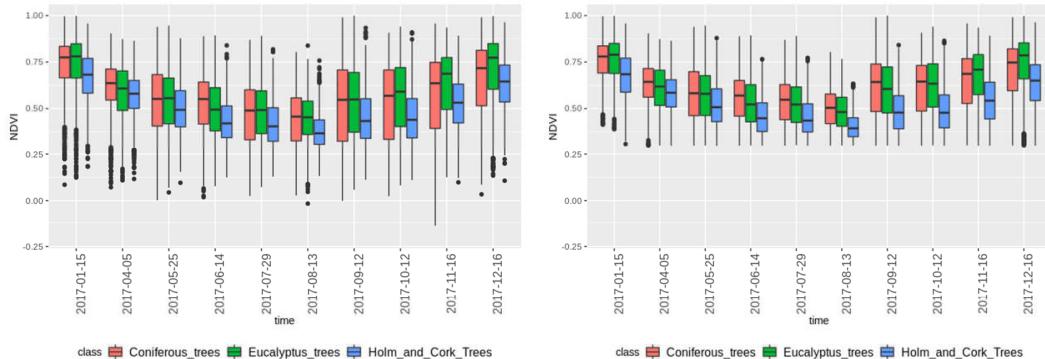


Figure 4.4: Cleaning preopocesing forest data

Since COS map is an annual land cover report, we had a multi-temporal class denominated herbaceous periodic. This class was composed by two signals, agricultural areas with irrigation (Portuguese: Culturas temporarias de regadio) and other without irrigation (Portuguese: Culturas temporarias de sequia). During summer the photosynthetic activity of vegetation is less active, and therefore, the NDVI signal display a decrease in the tendency of the values. In contrary, irrigation controls the production of the crop leading to evergreen samples. Besides herbaceous periodic, we had classes of herbaceous permanent or grass that according with the figure in 4.5 has similar NDVI characteristics in variance to the not irrigated crops.

In this context, since the maps were produced under static modeling, that is a classification model that does not account for spectral difference over the time, the labels of herbaceous samples needed to represent static land covers. Therefore, we set out two steps to use this training data. The first one obeyed to a sub-stratified random sampling of the 3 herbaceous classes (agricultural areas with irrigation, agricultural areas without irrigation and herbaceous permanent) in order to keep in balance their number respect the rest of the classes in the classification. And the second one, a conditional processing to keep only herbaceous with vegetation. Therefore, according with recommendations of professor Mario Caetano, we only keep samples with NDVI values larger than 0.3. That is, values lower than 0.3

corresponds to non-vegetated surfaces while green vegetation canopies correspond to an NDVI values larger than 0.3. The non-vegetation samples product of this experiment were taken out from the training since the original no-vegetation class was enough representative.

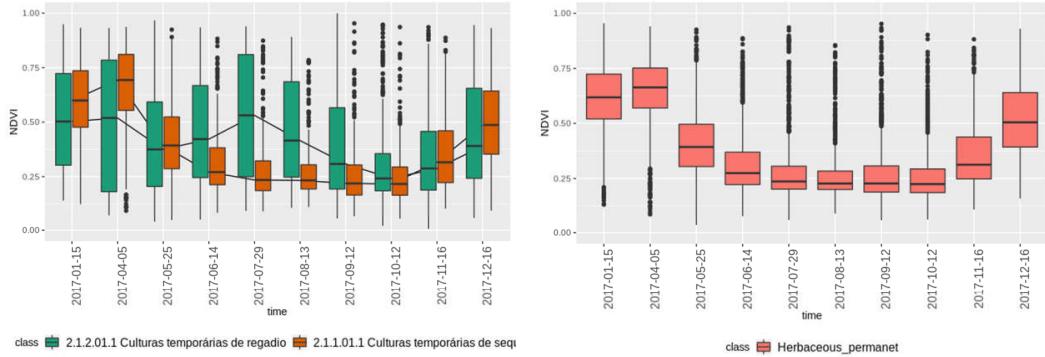


Figure 4.5: NDVI herbacesu

The cleaning preprocessing using NDVI variability led to a new distribution of the training data (see figure 4.6). While the refining of data at the beginning and end of the year represented to leave ranges between 5% and 10% of the data out, in summer we had cases where up to 25% of the information was removed. It should be noticed, that the class with more impact in the number of data per time is herbaceous due to phenology.

Cleaning preprocessing based on informativeness

The calculation of the scores of informativeness is based on an iterative learning process that ensemble the mechanism of boosting and the measurement of information entropy. In this context, according with the figure 4.7, assume that after fitting a model using a fraction of the training data (55% COS dataset), we classify all the samples of the same training data (80% COS dataset). In this first iteration, we took per label the prediction probabilities to calculate the measurement of information entropy (see Equation 4.2).

$$S = \sum_i^n p_i \log(p_i) \quad (4.2)$$

4 Methodology

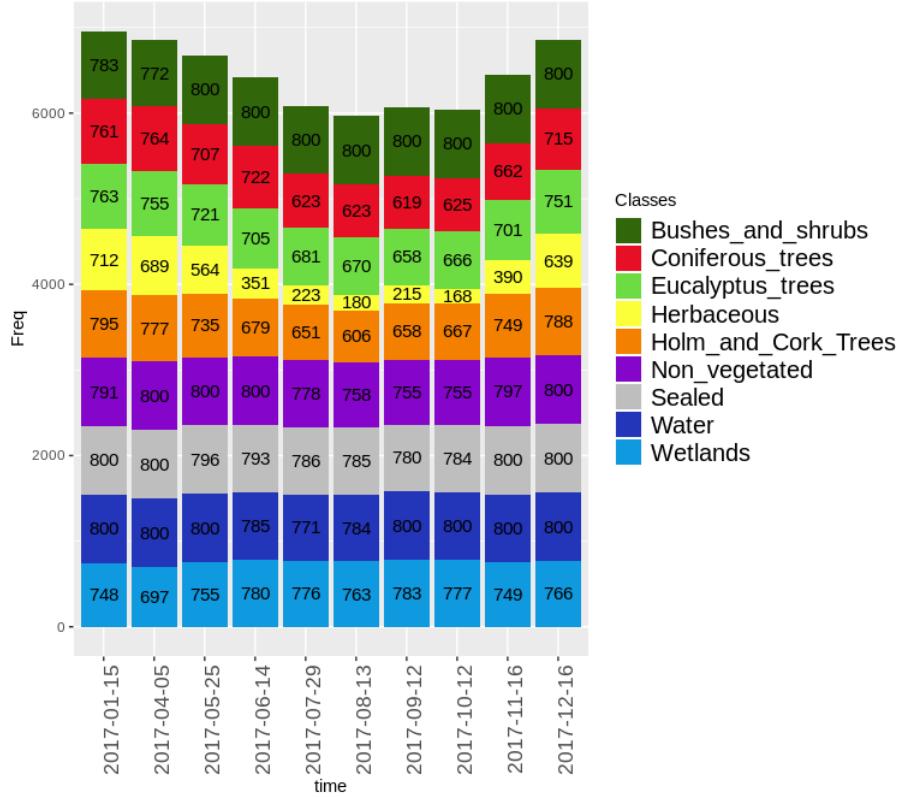


Figure 4.6: Training data distribution after applying NDVI rules

The higher its value, the higher its uncertainty for predicting correctly one of the classes, and therefore low informative. Under this first evidence, we repeated the same process to evaluate if over certain number of simulations all the models agreed in the same average of measurement of information entropy for each label. This process is independent of the image, so it needed to be repeated per scene.

In the figure 4.8, we show a normalization of the level of informativeness for the training data associated to herbaceous under different seasons. The change in the level of informativeness of the samples depend on how representative they are from one scene to another. For example, assume a label of herbaceous is representing a rice field. Depending on its state of production, the crop can be flooded during the Spring, vegetated during the Summer and Autumn, or not vegetated during the Winter; when there is no production (see rice production in Portugal [USDA \(2019\)](#)). The dynamic of this kind of crops lead to define different states of the label, and therefore raise the question of whether removing

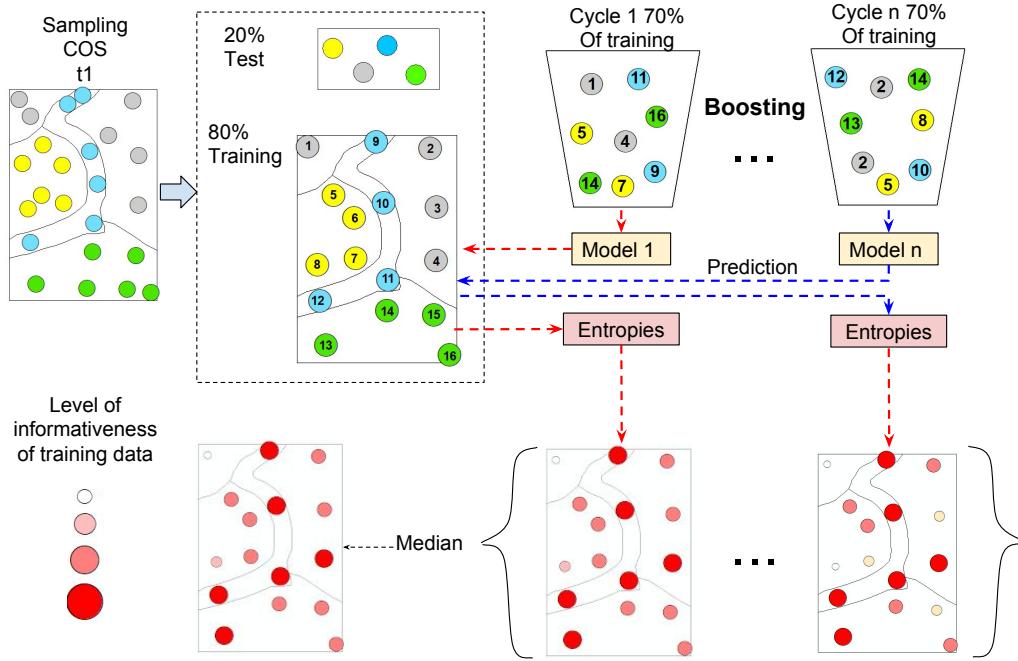


Figure 4.7: Iterative learning procedure to define informativeness based on boosting and the measurement of information entropy

mislabeled data by this effect can imply an improvement in accuracy classification.

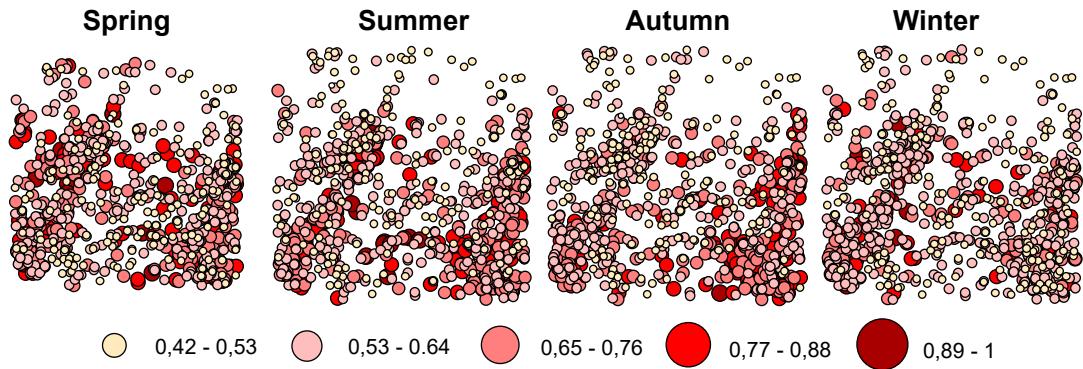


Figure 4.8: Informativeness for herbaceous

Another example derives from generalization of the land cover system in the production of the reference data, so random samples can fall over complexities of a wide diversity of classes that were simplified in one class in the map. According with this example, we can have two scenarios, one where the sample falls over a pixel that spectrally differs completely of the label, and another one where the pixel correspond to a mixture of different classes 4.9. Concerning the first

4 Methodology

scenario, we expect those labels contain the lowest values of informativeness. However, concerning the mixture of classes in one pixel, this can eventually turn out in one limitation of the proposed methodology since it does not consider fractional proportions of classes per pixel.

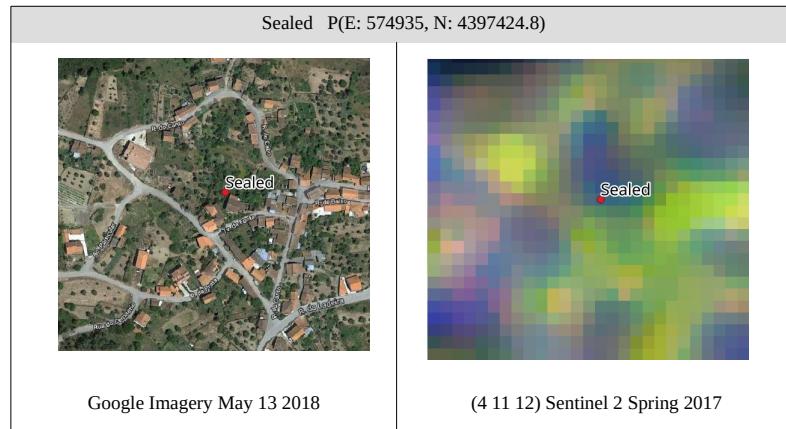


Figure 4.9: Sealed example

5 Results

This chapter presents the final results of the proposed methodology. In Section 5.1, we discuss about the results of the classification for all the images in 2017 using RF and SVM. This results represented the baseline to implement the two operational strategies of cleaning training data in section 5.2. In Section 5.3, we discuss about the parametrization of RF and SVM and its importance in the automatic classification. In section 5.4, we discuss about the usability of COS data using two versions of the dataset with different number of classes to predict as well as the influence of different levels of noise that eventually can hinder the results when COS data is used. At this stage, we show results based on the scene-based composite with the best scores of OA in the classification, so that in section 5.5 we make a comparison between classification using seasonal composites and single date images. Besides that, we show the benefits and limitations of implementing maximum NDVI for the composition of images. Finally, in 5.6, we interpret the accuracies per classes and give some recommendations for future work.

5.1 Classification performance

In consideration to the first objective, the figure ?? show OA of the classifications over the best scene-based composites per month by using two classification algorithms. The classification models were built from dataset 1 (9 classes) after careful carrying out sensitivity analysis to ensure optimal parameter were obtained per image (In section 3 we will discuss about the parametrization). In this context, the implementation of SVM with a RBF kernel outperformed the results obtained with RF over all the imagery of 2017 with average differences of 3 scores.

5 Results

In practice, the selection of the images are conditioned for the level of cloud contamination, and therefore best images are selected from the time of the year less affected by cloud cover like those from Summer. According with the technical report of COS map, the aereo-photographies were taken during the months of April, May and June of 2015. This imply the best prediction may obey to images around these dates, however, except by July 29 with a score of 0.64, in general all the imagery depicted similar results. This scenario show how robust the classifiers can be against possible changes in the reflectance captured by the label during the year due to phenology.

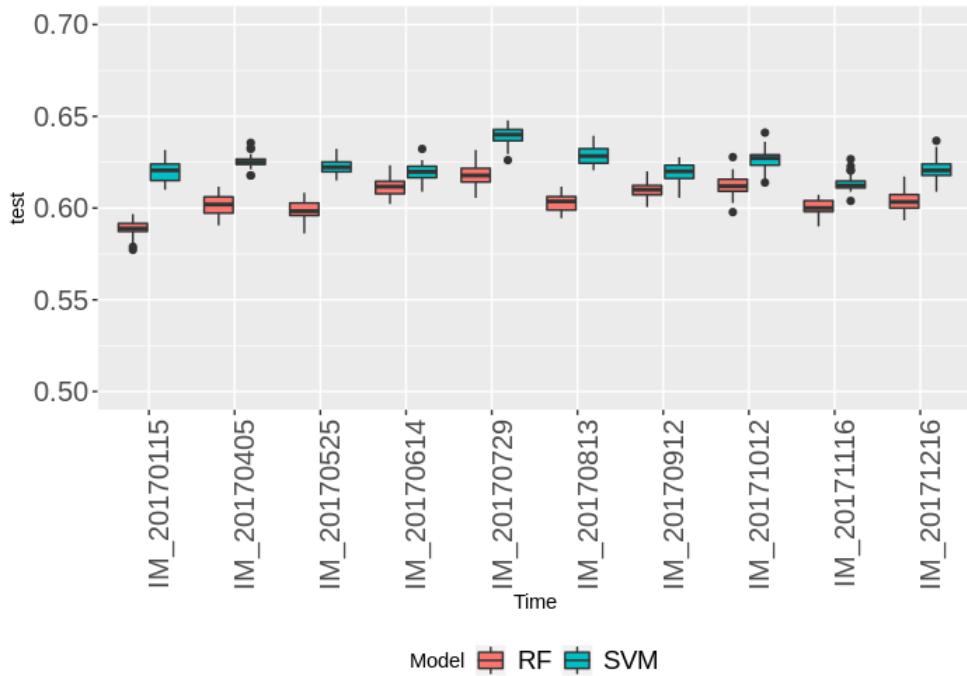


Figure 5.1: Performomance classifiers RF and SVM, scene-based composites 2017

Even though RF and SVM can be considered methods based on a black-box supervised learning, they can also give an intuitive explanation about the predictive power of the features in the classification. According with the documentation of Sklearn (library Python), RF offers an score that measure the importances of the study variables based on the mechanism of out-of-bag-sample error (cite). In the figure ??, we show scores of the variables after implementing RF (results were the same for all the imagery). As higher the score, higher the predictive power

5.2 Cleaning procedure performance

of the variable. Particularly, in terms of spectral information, the bands NDVI, B11-SWIR and B12-SWIR, showed the highest scores. This pattern may obey to the fact that the dataset was mainly composed by classes with vegetation. Moreover, the DEM also was a good predictor, determining that the land cover system in the region is fractionally characterized by the elevation.

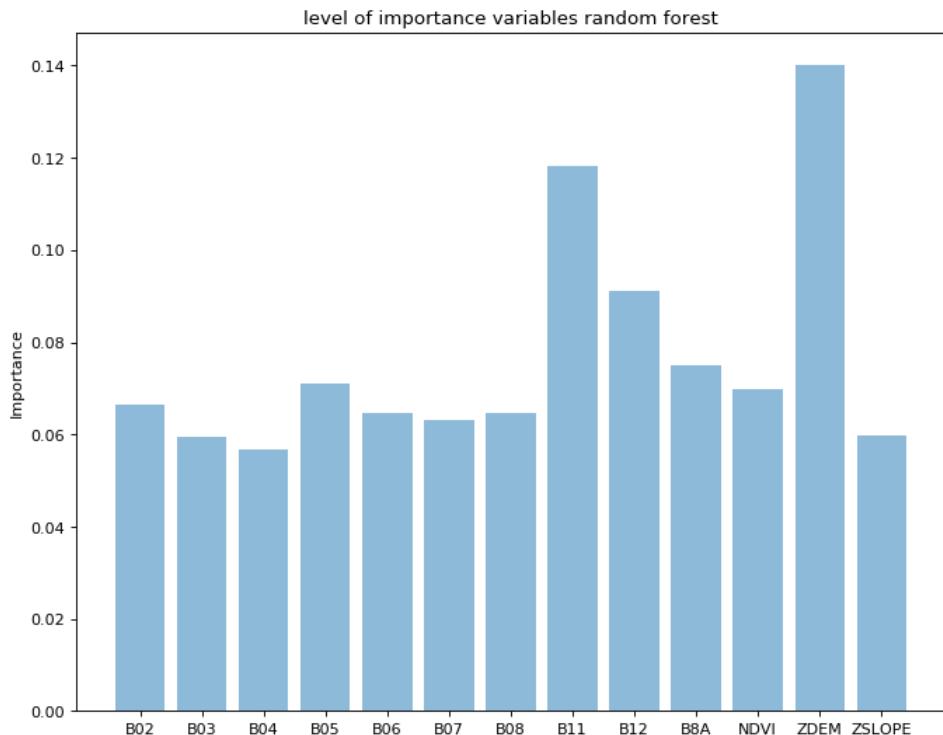


Figure 5.2: Importance variables

5.2 Cleaning procedure performance

In the next two sections we will show the results of the two proposed methodologies for refining data.

5.2.1 Cleaning preprocessing using NDVI variability

The figure 5.3 shows the comparison of the a classification with and without implementing the first strategy of filtering mislabel data, that is, analysis of NDVI signal variability. The classification was performed using SVM. Based on com-

5 Results

parisons of OA, the lower performance in the classification over the year let to establish that the first strategy of a cleaning preprocessing was not viable. The reason for this scenario may obey to two reasons. The first one may correspond to issues in classification due to unbalanced data. As we saw in methodology, the cleaning preprocessing implied a reduction in the number of samples for vegetation, especially during summer, and therefore possible not enough representation of them. Although the impact of unbalanced data is high in Summer, in Spring the reduction of erroneous labels was very low. However, regardless whether the impact was low in Spring, a classification without filtering continued being better. That result allowed us to give an additional interpretation of the low performance of the strategy, that is, loss of generalization. SVM and RF are considered non-parametric learning algorithms. That is, mechanisms that are able to adapt any functional form from the trained data and without making any assumption of its size or distribution. This flexibility may turn out in a limitation in the face of strong changes in the amount of training data that eventually can lead to define new functionals forms in the prediction, and therefore, to lose ability to generalize unseen data.

So far, we have introduced the results of the first sampling refining method. These results were fundamental to undertake more sophisticated strategies during the development of this thesis and to evaluate the ability of the predictors under different assumptions of the data. For example, a variation in the number of classes to predict or the level of noise the models can resist.

5.2.2 Cleaning preprocessing using Informativeness

In the previous section, we used a fractional part of the COS dataset to test the results. However, the presence of anomaly data may be general, both in training and testing. Since the cleaning processing is done over the training, the use of an external dataset let us to see the results over data that suppose to be representative for specific date. Therefore, in collaboration with DGT, we had access to an external dataset (see Section data). This dataset corresponds to a visual interpretation of 557 samples for the image of 29 of July. In this sense, the

5.2 Cleaning procedure performance

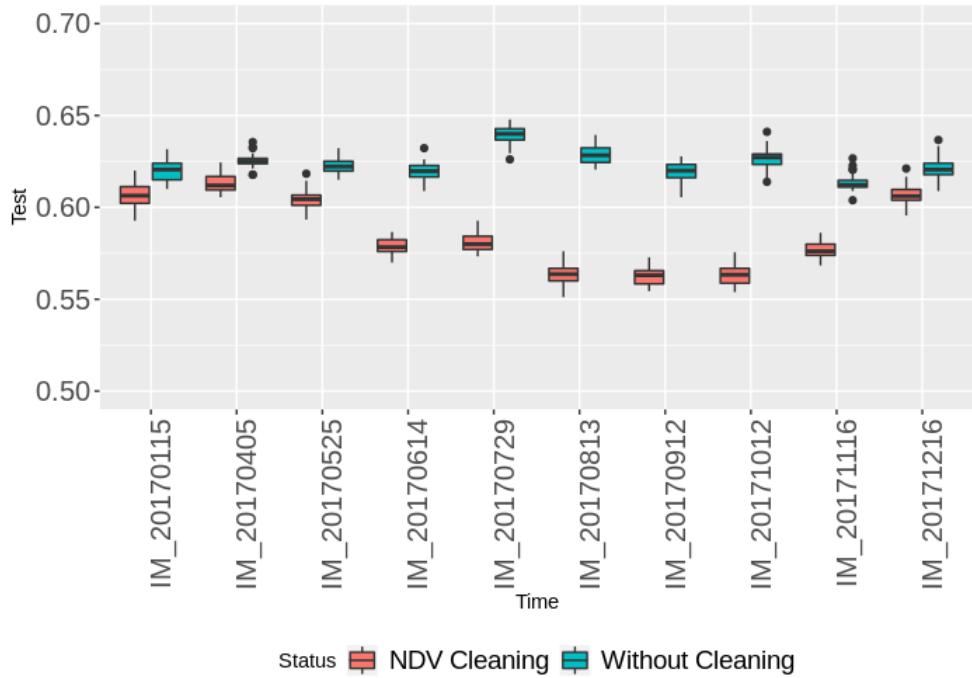


Figure 5.3: Cleaning processing NDVI, Performance over scene-based composites 2017

second strategy was evaluated making especial emphasis in this date.

Figure 5.5.a shows the performance of SVM classifier by evaluating COS test (test 1 red line) and the external data test (test 2 green line) under different number of observations in the modelling of training data for July 29. The graphic shows in x-axis from right to left the percentage of data removed according with their level of informativeness and in the y-axis their correspond OA for test 1, test2 and the validation data (blue line). The box-plot are a demonstration of the stability of the results of classifications under different number of observations of the training data. The lower performance on the external data set may be due to the use of unbalanced data, the external data set was mainly characterized by samples of herbaceous and woody classes. That is, classes that the model had low predictive power.

As we saw in the methodology, this strategy allowed to score the training data according with its contribution in the classification of a particular image. While the results of the cross validation increased after every iteration since the

5 Results

model was validated using less noisy data, the OA of both test datasets keep constant over certain range of reduction of samples. After a reduction of 20% of samples, we could appreciate that the classifier start losing ability to generalize and predict properly the the test datasets. Even though the filter did not lead to a better performance of the classification for this date, the graphic allowed us to appreciate how the methodology successfully allow to identify which samples are not contributing to a better performance in the classification either because they may be redundant or they do not continue representing the class on the ground (This was also verified by visual inspection).

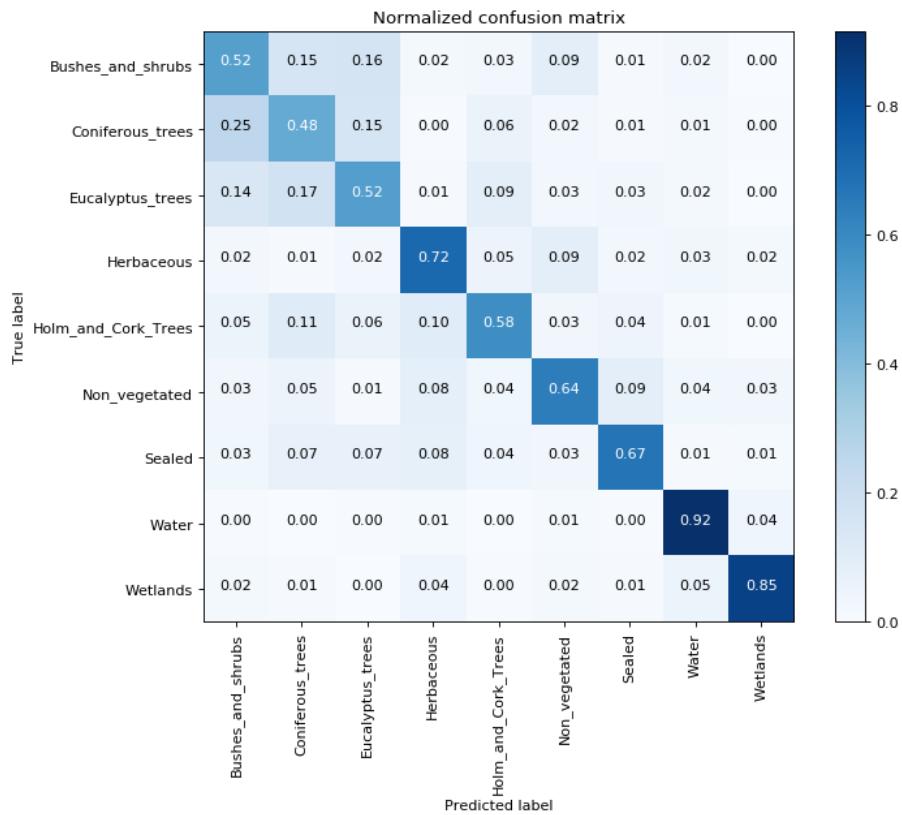


Figure 5.4: Normalized confusion matrix, classification scene-based composite

July 29

The similar spectral variation of classes such as, coniferous, Eucalyptus, Holm and cork and shrubs let the classification model to present high confusion in the discrimination of these classes (see normalized confusion matrix in figure 5.4). This scenario raised the question if the results could maintain the same or improve

5.2 Cleaning procedure performance

after merging land cover types with similar spectral characteristics. Therefore, we undertook the creation of a second dataset that merged the aforementioned classes in a new class called woody. Generally, the merge let to provide better results in terms of OA in almost 10 scores (see figure 5.5.b). As it was done with 9 classes, the informativeness analysis had the same considerations, except by the fact that dataset 2 offered better performance by narrowing the problem to only 6 classes.

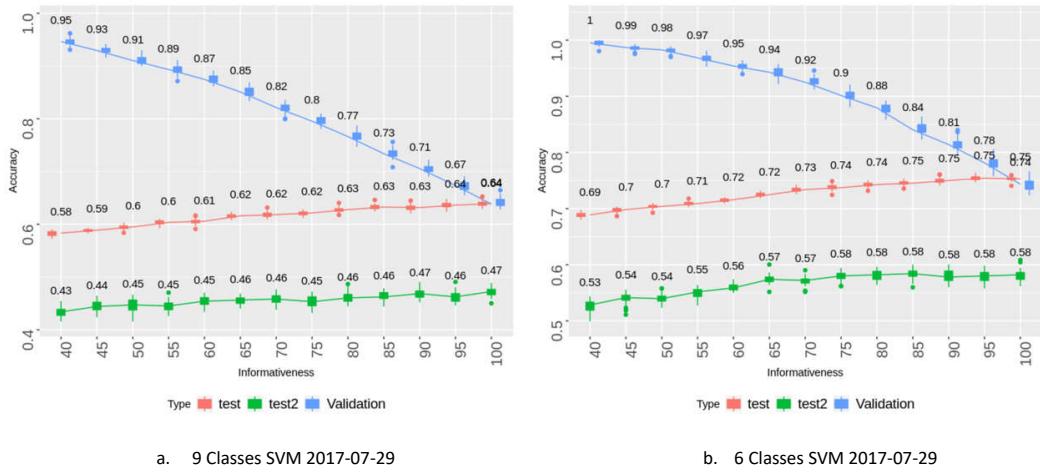


Figure 5.5: Cleaning preprocessing using Informativeness, scene-based composite
July 29

It should be notice, that so far we narrowed the viability of using the second strategy based on the classification results of a scene-based composite of July 29. This scenario raised the question of whether the implementation of the criterion of informativeness over a different image, for example in Winter, where the OA is relatively lower than in summer, could lead to increase accuracy.

Therefore, we undertook the implementation of the methodology over one image in Winter, specifically December 16. The limitation here was the lack of a test dataset for this date as it was implemented for July 29. We performed the same methodology including this information with the warning that this dataset should be used for experiments in summer (see figure 5.6). Unlike this limitation we performed the strategy of scoring data according with its informativeness. This turned out in similar results, except than in this case the filter allowed to

5 Results

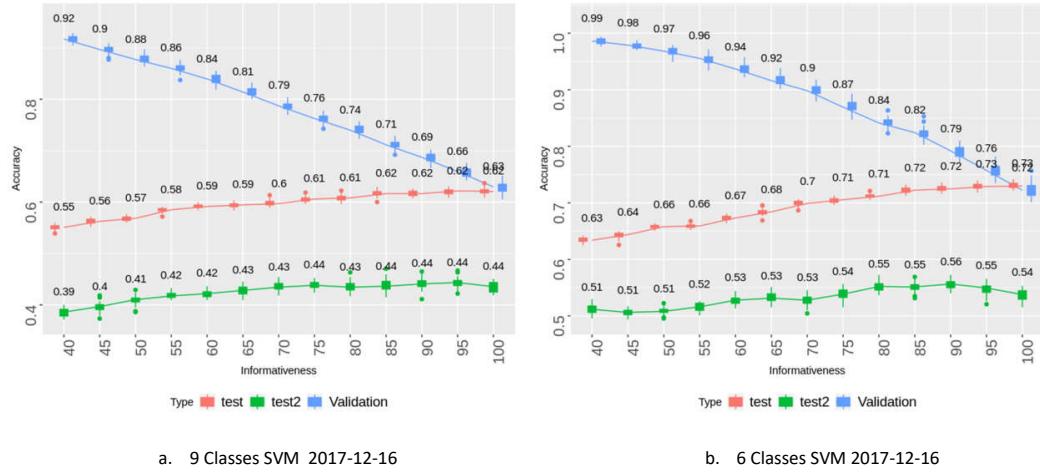


Figure 5.6: Cleaning preprocessing using Informativeness, scene-based composite December 16

increase slightly the OA over the external dataset.

In this context, since the results of the previous experiment were not completely reliable due to the assumption about using the external dataset in an experiment of winter, we undertook another strategy. This new idea consisted of adding noise to the training data used for the classification of July 29, so that it corresponded to a parallel experiment in winter where training data may be more impacted by mislabeled data, but with a real external test data.

After calculating again the score of informativeness over a new training data with an additional random noise of 30% to the inherent error associated from the automatic selection of labels using the reference COS map, the strategy for refining data started benefiting the classification. According with the graphic 5.7, coincidentally, the OA started increasing gradually as we approached to the 30% of data removed. Which means that the score of informativeness was working properly to remove data from the training set. After this threshold the model started losing the ability to generalize, and therefore produce results with lower accuracies. Simultaneously, the external test dataset also benefited with the reduction of noise at some point that it started giving the same results than in the graphic 5.5.

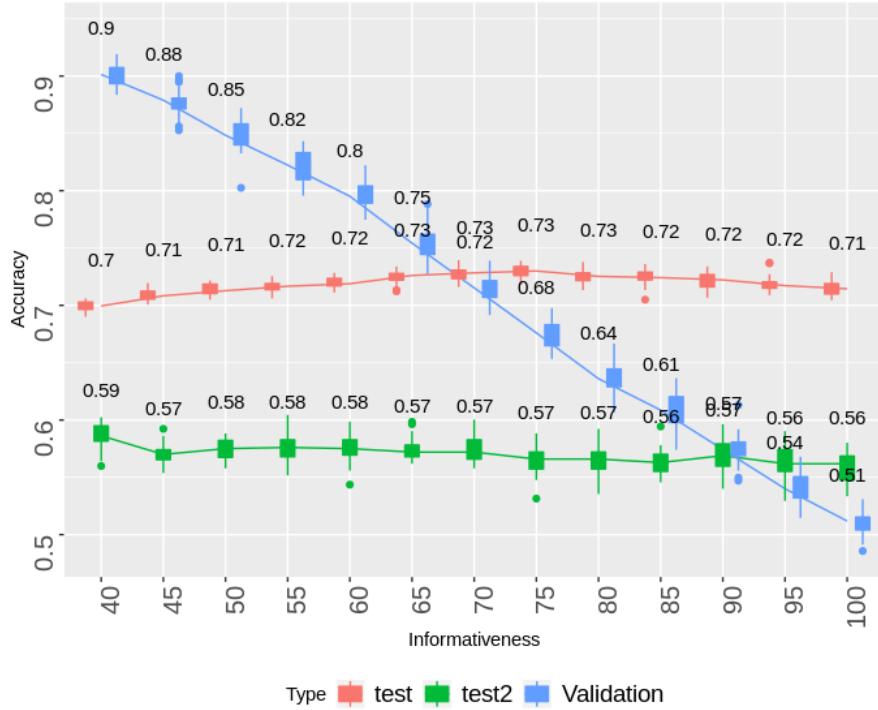


Figure 5.7: Performance of the proposed methodology over a training data with 30% of noise, scene-based composite July 29

5.3 Parameter optimization

While a sensitivity analysis of the parameters lead essentially to optimize the performance of an algorithm, their estimations may vary from one date to another. To evaluate how stable or similar is the parametrization in that scenario, we recreated a sensitivity analysis over different levels of noise and different number of classes for each classifier in the classification of the image of July 29.

On the one hand, in the context of random forest, the figure 5.8 shows that the optimal values found by two-folds cross validation keep constant over 0%, 30% and 50% percentage of noise. Besides that, we implemented the same study over the dataset with 6 classes. The result is that the number of trees parameter equal or larger than 100 is a good estimation, not matter how noise the data can be as product of the natural changes of land cover over the year or types of mis registrations. However, regarding *mtry*, there was not a clear value to pick up. The default parameter of *mtry* equal to the square of the number of variables,

5 Results

in this case $mtry$ between 3 and 4, does not lead to better or worst results.

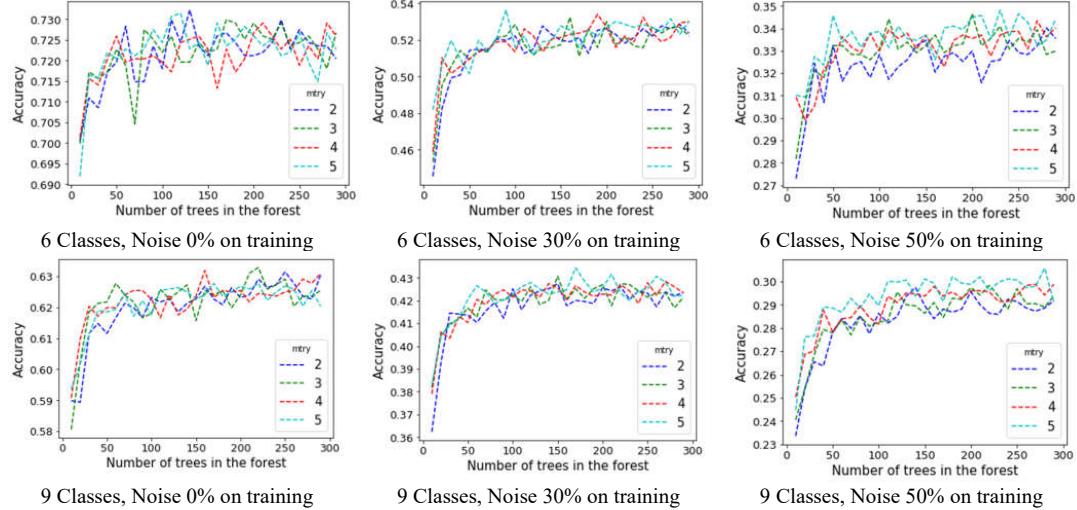


Figure 5.8: Sensitivity analysis parameters random forest

On the other hand, in the same frame of study of random forest, we made a sensitive analysis of the parameters for SVM with a kernel RBF. The results in figure 5.9 showed that high values of \mathbf{C} tested similar best overall accuracies under different level of noise. A large value for \mathbf{C} implied a lower chance for misclassifying the training data, however this could have resulted in bias in the modeling, and therefore possible wrong prediction of the unseen test data. Therefore, we made sure of selecting a common estimation of low values of \mathbf{C} parameter under different levels of noise that at the same time guarantee good performance. Moreover, regarding γ , its estimates seem to be also limited for the three levels of noise and also for the number of classes in each prediction.

After carefully carrying on a sensitivity analysis to ensure optimal parameters in both classifiers, we found out that we can keep similar parameters for different levels of noise of the data and different number of classes; at least for the scenarios of 6 and 9 classes.

5.4 Influence of Noise in the performance of SVM and RF

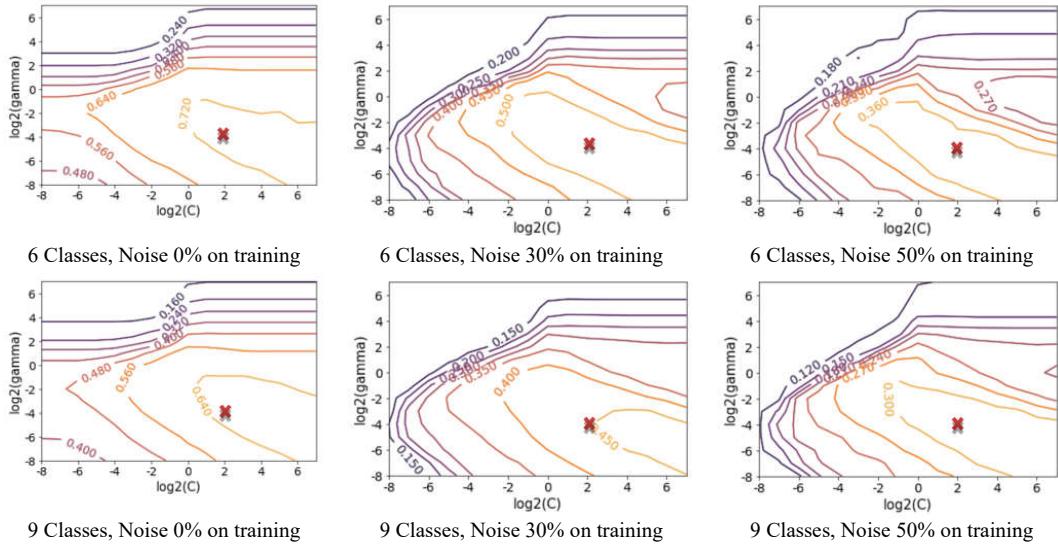


Figure 5.9: Sensitivity analysis parameters SVM with RBF kernel

5.4 Influence of Noise in the performance of SVM and RF

So far, we evaluated all our results using SVM. However, our final attempt aim at showing the predictive power of the classifiers under different levels of noise. Figure 5.10 shows in red and blue line the medians of producer accuracies respect the level of noise introduced in the COS data using support vector machine with a kernel RBF and random forest classifier respectively. The results show how the performance of both classifiers decreased as the level of noise increased. Comparing both classifiers, SVM looked less sensitive to the presence of noise data than RF. Both learning graphics have in common a inflection point at the level of 50% of noise in the data. This implies that both classifiers have good ability for generalization under scenarios of high content of noise data.

So far, we have shown the predictive power of the classifiers under study, the viability of working with COS data as reference data in the classification and the viability of using a refining procedure for mislabeled data. In this context, in the next section, we will discuss about the classification performance using pixel-based composites and how their results differ from scene-based composite

5 Results

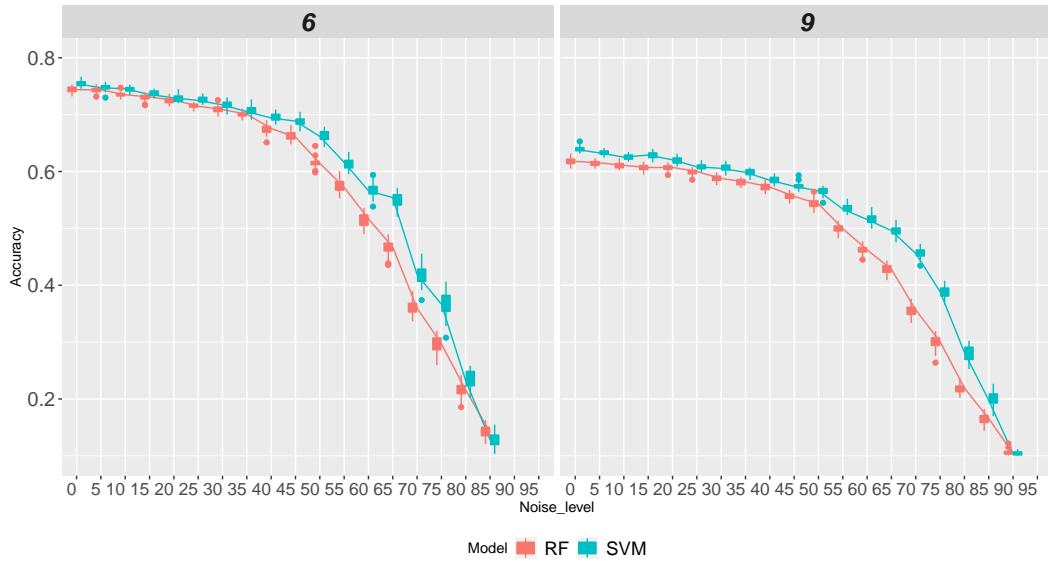


Figure 5.10: Performance random forest vs SVM using COS data under different levels of noise, scene-based composite July 29

classification.

5.5 Single date maps vs seasonal maps

The computational performance of pixel-based composites based on maximum NDVI for their construction were computationally fast. Depending on the number of spectral bands and images per date to consider in the time series, the scripts developed on python could create composites with Sentinel 2 imagery in terms of few minutes.

In the figure 5.11, we show one example of how the composites could combat the cloud contamination. Clouds tend to depict red reflectance somehow larger than the near infrared. This slight difference turns out in low positive or negative values of NDVI. When the clouds passed over land covers such as vegetation, the maximization of NDVI was a simple mechanism that allow to retrieve true pixels associated to high values of NDVI, and therefore, the construction of artificial images free of clouds.

Figure 5.12 shows the performance of the modeling by evaluating overall accuracies over two classifiers, two versions of the dataset with different number

5.5 Single date maps vs seasonal maps

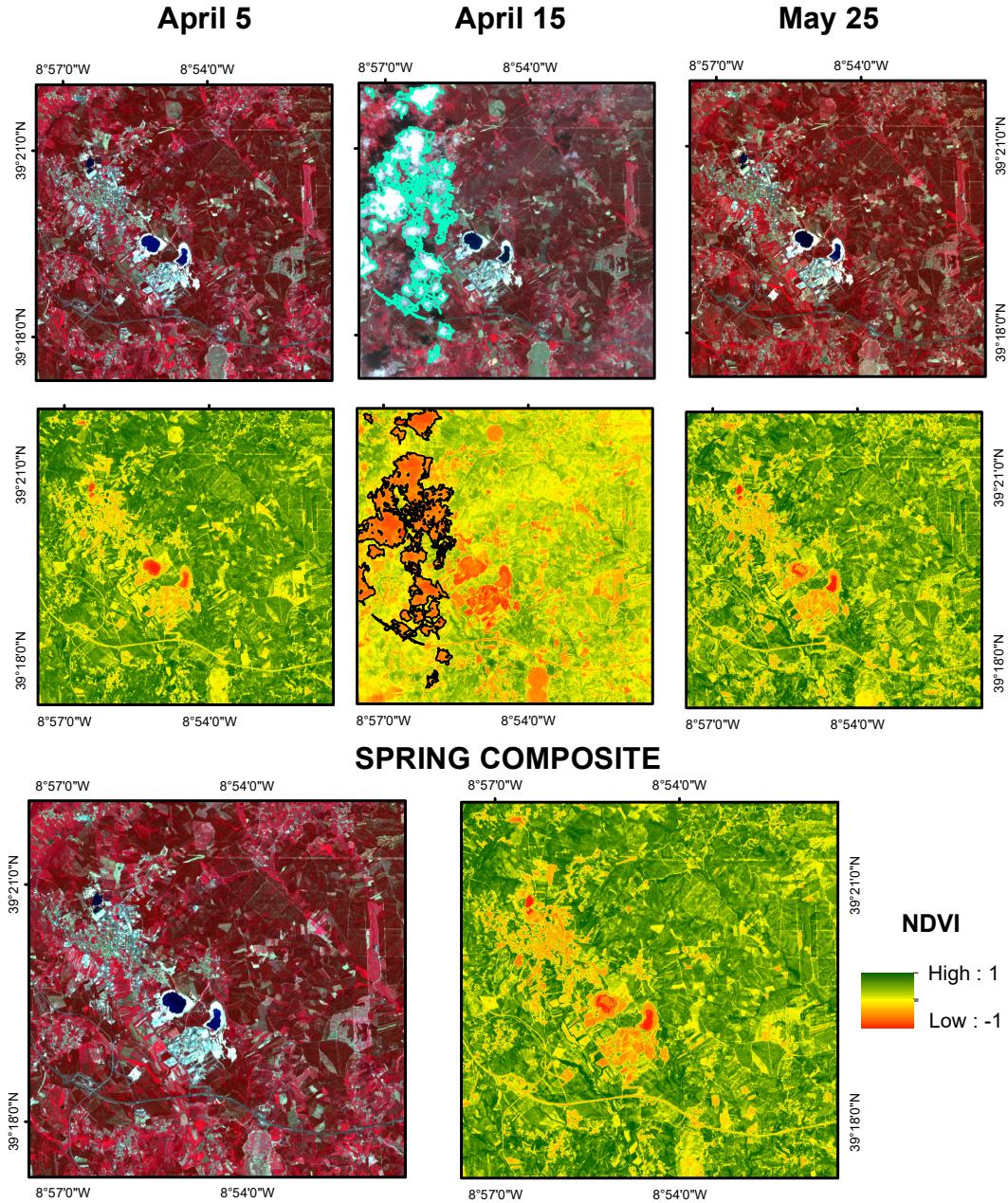


Figure 5.11: pixel-based composite Spring

of classes and over different images. Regarding the images, we highlighted the best performance of the scene based composites per season, so that we had a reference to compare the results against the pixel based composites. We did not implement any refining procedure over the data due to previous considerations.

Firstly, in the context of the performance by classifiers using composites, SVM with a RBF kernel always over-performed the results in relation to random forest

5 Results

after carefully optimizing their parameters. Secondly, as we explained before, now also in the composites, a merging of the woody classes implied to improve the ability of the classifier to predict correctly all the classes, but evidently narrowing the problem to only 6 classes.

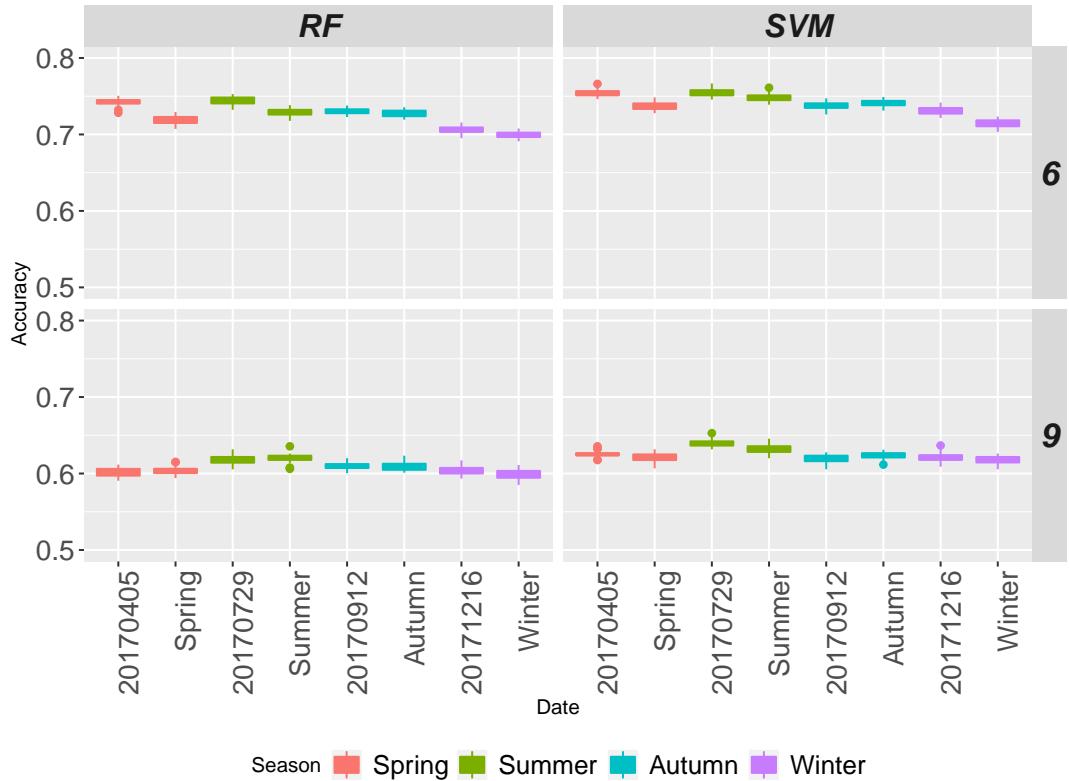


Figure 5.12: Performance of the classifications, scene-based vs pixel-based composites

5.6 Assessment

It should be noticed, that the appreciation of a similar performance of the single date imagery respect of composites was based on the metric OA. As we introduced in the methodology, there are a diverse ways to assess the results besides OA. Therefore, we started giving a visual appraisal of the results in the figure 5.13. This graphic depict the classification of 9 classes during the 4 seasons using the pixel-based composites; next to each prediction we find the false color composition (near-8, red-4, green-3 in Sentinel 2) and the NDVI. The results were

quite satisfactory after visual inspection.

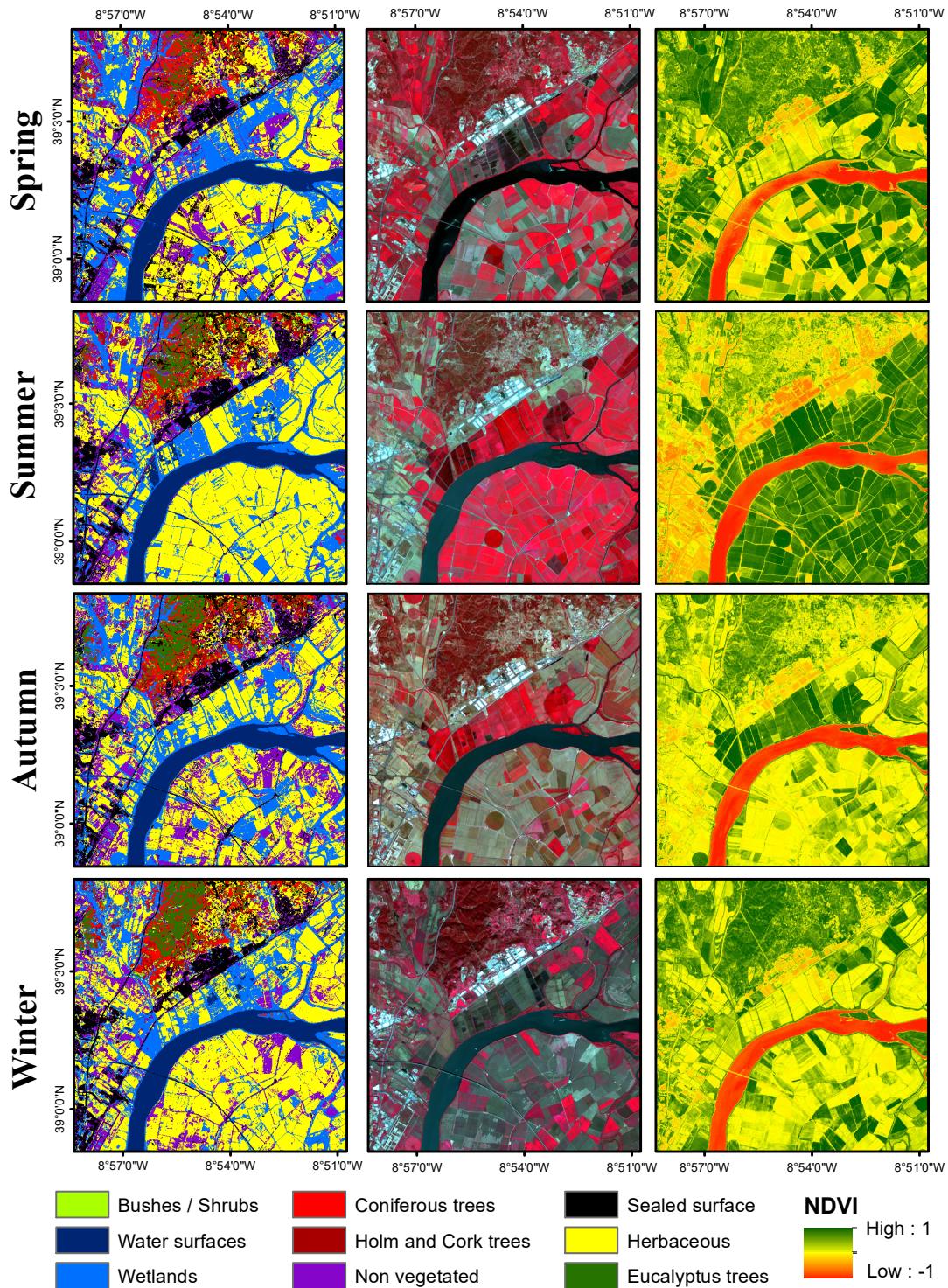


Figure 5.13: Seasonal mapping.

In consideration to the fourth objective, working for obtaining images free

5 Results

of clouds ended up the only reason why we can highlight the use of composites in this thesis. Even though the literature emphasize in the additional improvements that the methodology can bring in the discrimination of vegetation classes, the results per class according with 5.1 are very similar respect single imagery. In general, for the classification of 9 classes, the model showed a high ability to discriminate classes of water, wetland, sealed and not vegetated. As we explained before, the low performance and high confusion among Woody classes let to consider a classification with a lower number of classes (see table 5.2). In this context, no only the class of woody reached better accuracies, specifically herbaceous benefited with the new structure.

Table 5.1: Accuracy assessment 6 classes

Date	Overall Accuracy	Kappa	Accuracy	Bushes and shrubs	Coniferous trees	Eucalyptus trees	Herbaceous	Holm and Cork Trees	Non vegetated	Sealed	Water	Wetlands
2017 July 29	0.64	0.59	User	0.55	0.43	0.53	0.67	0.54	0.64	0.71	0.87	0.83
			Producer	0.55	0.41	0.48	0.64	0.69	0.67	0.7	0.72	0.89
Summer	0.64	0.59	User	0.53	0.42	0.54	0.68	0.54	0.66	0.67	0.88	0.84
			Producer	0.52	0.38	0.51	0.61	0.71	0.64	0.74	0.7	0.89

Table 5.2: Accuracy assessment 6 classes

Date	Overall Accuracy	Kappa	Accuracy	Herbaceous	Non vegetated	Sealed	Water	Wetland	Woody
2017 July 29	0.75	0.69	User	0.7	0.69	0.74	0.9	0.83	0.69
			Producer	0.71	0.71	0.77	0.77	0.88	0.7
Summer	0.74	0.68	User	0.69	0.69	0.72	0.89	0.83	0.68
			Producer	0.68	0.69	0.76	0.76	0.87	0.71

Bibliography

Bishop, C. M., 2006. Pattern recognition and machine learning (information science and statistics) springer-verlag new york. Inc. Secaucus, NJ, USA.

Bontemps, S., Herold, M., Kooistra, L., Van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., Defourny, P., 2012. Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences* 9 (6).

Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3. 1. Statistics Department University of California Berkeley, CA, USA 1.

Büschenfeld, T., Ostermann, J., 2012. Automatic refinement of training data for classification of satellite imagery. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 1–7.

Caetano, M., Nunes, A., Dinis, J., Pereira, M., Marrecas, P., 2015. Carta de uso e ocupação do solo de portugal continental para 2007-cos2007. do Território, Direcção-Geral.

Colditz, R., Schmidt, M., Conrad, C., Hansen, M., Dech, S., 2011. Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions. *Remote Sensing of Environment* 115 (12), 3264–3275.

Congalton, R. G., Green, K., 2008. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

Bibliography

- Deng, J., Wang, K., Deng, Y., Qi, G., 2008. Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing* 29 (16), 4823–4838.
- ESA, 2017. Sen2cor. <http://step.esa.int/main/third-party-plugins-2/sen2cor/>.
- Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment* 80 (1), 185–201.
- 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.
- Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., 2015. An integrated landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sensing of Environment* 158, 220–234.
- Holben, B. N., 1986. Characteristics of maximum-value composite images from temporal avhrr data. *International journal of remote sensing* 7 (11), 1417–1434.
- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing* 9 (1), 95.
- Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. *International journal of remote sensing* 25 (12), 2365–2401.
- Mather, P., Tso, B., 2016. Classification methods for remotely sensed data. CRC press.
- Meier, J., Zabel, F., Mauser, W., 2018. A global approach to estimate irrigated areas—a comparison between different data and statistics. *Hydrology and Earth System Sciences* 22 (2), 1119–1133.

Bibliography

- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3), 247–259.
- Navarro, G., Caballero, I., Silva, G., Parra, P.-C., Vázquez, Á., Caldeira, R., 2017. Evaluation of forest fire on madeira island using sentinel-2a msi imagery. *International Journal of Applied Earth Observation and Geoinformation* 58, 97–106.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C., Dedieu, G., 2017a. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing* 9 (2), 173.
- Pelletier, C., Valero, S., Inglada, J., Dedieu, G., Champion, N., 2017b. Filtering mislabeled data for improving time series classification. In: 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, pp. 1–4.
- Pontius, R. G., 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric engineering and remote sensing* 66 (8), 1011–1016.
- Rogan, J., Chen, D., 2004. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in planning* 61 (4), 301–325.
- Shi, D., Yang, X., 2015. Support vector machines for land cover mapping from remote sensor imagery. In: *Monitoring and Modeling of Global Changes: A Geomatics Perspective*. Springer, pp. 265–279.
- Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing* 10 (6), 989–1003.
- Sinha, P., Kumar, L., Reid, N., 2012. Seasonal variation in land-cover classification accuracy in a diverse region. *Photogrammetric Engineering & Remote Sensing* 78 (3), 271–280.

Bibliography

- Thanh Noi, P., Kappas, M., 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors* 18 (1), 18.
- Tolba, A., 2010. Manifolds for training set selection through outlier detection. In: *Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on*. IEEE, pp. 467–472.
- Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., Emery, W. J., 2009. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 47 (7), 2218.
- USDA, 2019. Crop calendar for europe. https://ipad.fas.usda.gov/rssiws/al/crop_calendar/europe.aspx.
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.-T., 2018. How much does multi-temporal sentinel-2 data improve crop type classification? *International journal of applied earth observation and geoinformation* 72, 122–130.
- White, J., Wulder, M., Hobart, G., Luther, J., Hermosilla, T., Griffiths, P., Coops, N., Hall, R., Hostert, P., Dyk, A., et al., 2014. Pixel-based image compositing for large-area dense time series applications and science. *Canadian Journal of Remote Sensing* 40 (3), 192–212.
- Wu, T.-F., Lin, C.-J., Weng, R. C., 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research* 5 (Aug), 975–1005.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. *Nature climate change* 3 (10), 875.
- Yin, H., Pflugmacher, D., Kennedy, R. E., Sulla-Menashe, D., Hostert, P., 2014. Mapping annual land use and land cover changes using modis time series. *IEEE*

Bibliography

Journal of selected topics in applied earth observations and remote sensing
7 (8), 3421–3427.

APPENDIX A - Maps composites

Here some maps

APPENDIX B - Single Maps