

Masters Program in **Geospatial Technologies**



INTRA-ANNUAL LAND COVER MAPPING

AUTOMATIC TRAINING SAMPLE EXTRACTION FROM
OLD MAPS FOR INTRA-ANNUAL LAND COVER MAPPING
AT CENTRAL OF PORTUGAL

William Alexander Martínez Blanco

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*



2019 *Automatic Training Sample extraction From Old Maps for Intra-annual
Land Cover Mapping at Central of Portugal* William Alexander Martinez Blanco

Intra-annual land cover mapping

Automatic Training Sample Extraction From Old Maps for Intra-annual Land Cover Mapping at Central of Portugal.

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

February 24, 2019

Acknowledgment

I want to thank the three supervisors for their contributions which have greatly improved the original version of this thesis. Moreover, I must express my especial gratitude to professor Mario Caetano, for the patient guidance, encouragement and advice he has provided throughout the time frame of this thesis. His office was always open whenever I had a question about my research or writing.

I would also like to thank all the members of staff at DGT and Professors at Nova IMS who helped me in my supervisor's absence. Especially, I would like to thank Daria Lüdtke, Inês Girão and Hugo Costa, and professors Marco Painho and Joel Silva for the suggestions they made in reference to Chapter 4 of this work.

I also leave my most sincere gratitude to the Erasmus Mundus Master program Science in Geospatial technologies, headed by the professors Marco Painho, Christoph Brox, Christian Kray, Joaqun Huerta and Michael Gould. I am very honored for having been one of the recipients of this Erasmus scholarship.

Last but not least, I want to thank all my family who has always supported me unconditionally in every endeavor that I have undertaken in my life. With seven and a half thousand of kilometers away from them, I do not imagine this journey without their calls and constant motivation.

Abstract

Automatic Training Sample extraction From Old Maps for Intra-annual Land Cover Mapping at Central of Portugal.

Making operational efficient the production of Land Use Land cover (LULC) mapping over large areas as the consistency and accuracy keep a high quality is an essential condition for the implementation of applications that require periodic information, such as forest fire propagation, crop monitoring or climate models. The increasing spatial and temporal resolution satellite images, such as those provided by Sentinel 2, open new opportunities for producing accurate datasets that can improve the lack of production of global and regional LULC maps with fine scale and up-to-date information. In this context, while this thesis aimed to make automatic the generation of intra-annual maps implementing a workflow that consists of supervised classification in synergy with automatic extraction of training samples from an old map, it also aimed to use singular and BAP composites. Therefore, after a preliminary selection and preprocessing of the implemented spectral bands in the classification both from single and BAP composites of Sentinel 2 images of 2017, a random selection of training points is extracted from an old reference map; national LULC map of Portugal, COS 2015. We performed a classification scheme using support vector machine (SVM) and Random forest (RF) classifiers with two datasets of six and nine different number of land cover classes. The out-of-date information derived from the old map led us to evaluate the viability of implementing two refining procedures over the data

Abstract

to improve accuracy; one based on margins of NDVI signals and another based on an iterative learning procedure. Since the proposed methodologies did not lead to improving OA on the classification of any of the images of 2017, we questioned for robustness of the classifiers RF and SVM by injecting different levels of noise during the modeling. Finally, the free cloud and phenological maximization of the BAP composites become in a consistent and efficient input for the production of seasonal LULC mapping.

Keywords

Best available pixel

Intra-annual Land Use Land Cover

Support vector machine and random forest

Geographical information systems

Acronyms

DGT - Direcção-Geral de Ordenamento do Território, Portugal

COS Uso e Ocupação do Solo

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

Index Of Figures

3.1	Study area	16
3.2	Available images Sentinel 2, 2017	17
3.3	COS Map 2015, Nomenclature Level 1	18
3.4	Class definition of the two datasets and COS nomenclature 2015 .	20
3.5	Distribution of the number of samples of the external dataset per percentage	21
4.1	General Methodology	24
4.2	Methodology image preprocessing	25
4.3	Methodology composites	26
4.4	Methodology selection points from old map	26
4.5	Methodology extraction features from images	27
4.6	A visualization of the splits of the training dataset during classification	28
4.7	NDVI signals over time for classes of trees	29
4.8	Iterative learning procedure to define informativeness based on boosting and the measurement of information entropy	31
5.1	Spatial distribution of the training data	34
5.2	Baseline July 25 2015, RF classifier	35
5.3	Level of importance variables random forest	36
5.4	Perfomance classifiers RF and SVM, scene-based composites 2017	37
5.5	pixel-based composite Spring	39
5.6	Performance of the classifications, single images vs BPA composites, 9 classes classification	40

Index Of Figures

5.7	Performance of the classifications, single images vs BPA composites, 6 classes classification	40
5.8	Normalized confusion matrix, classification scene-based composite July 29	41
5.9	Seasonal mapping.	43
5.10	NDVI signal for herbaceous classes	44
5.11	Traning data distribution after applying NDVI rules	45
5.12	Cleaning processing NDVI, Performance over scene-based composites 2017	46
5.13	Informativeness score for herbaceous	47
5.14	Example of low informative sample for sealed	47
5.15	Cleaning preprocessing using batch learning, Image July 29 2017 .	48
5.16	Cleaning preprocessing using batch learning per class, Image July 29 2017	50
5.17	Performance of the proposed methodology over a traning data with 30% of noise, scene-based composite July 29	50
5.18	Performance random forest vs SVM using COS data under different levels of noise, scene-based composite July 29	51
5.19	Sensitivy analysis parameters random forest	52
5.20	Sensitivy analysis parameters SVM with RBF kernel	53
1	accumulative values of precipitation per month, 2015. Source data: netCDF files NOAA	63
2	Intersection of the classes of COS map 2015 with the study area, Nomenclature based on recent recommendation from EEA on Copernicus land monitoring services	64
3	Intersection of nine classes of COS map 2015 with the study area, Nomenclature based on recent recommendation from EEA on Copernicus land monitoring services	65

Index Of Tables

5.1 Accuracy assessment 9 classes	42
5.2 Accuracy assessment 6 classes	42

Index Of The Text

Acknowledgment	iii
Abstract	v
Keywords	vii
Acronyms	ix
1 Introduction	1
1.1 Problem statement and Motivation	1
1.2 Objectives	4
1.2.1 Research question	4
1.2.2 Aim	4
1.2.3 Objectives	4
2 Literature Review	7
2.1 Land cover mapping and image classification	7
2.2 Best pixel available	8
2.3 Automatic selection and refinement of training data	9
2.4 Classification algorithms and tuning parameters	10
2.4.1 Support vector machines	11
2.4.2 Random forest	12
2.5 Accuracy assessment in land cover classification	12
3 Data and study area	15
3.1 Study area	15
3.2 Sentinel 2 Imagery	15

Index Of The Text

3.3 COS map	17
3.4 External dataset	19
4 Methodology	23
4.1 Image preprocessing	23
4.2 Seasonal Composites	25
4.3 Automatic training data from a old map	26
4.4 Classification schema	27
4.5 Filtering training data	28
4.5.1 Trimming NDVI signals	29
4.5.2 Filtering training data based on a iterative learning procedure	30
4.6 Evaluation of the robustness of the classifiers	30
5 Results	33
5.1 Automatic training selection from COS map	33
5.2 Classification	34
5.2.1 Baseline	35
5.2.2 Classification using RF and SVM (images 2017)	36
5.3 Single maps vs seasonal maps	37
5.4 Filtering training data	42
5.4.1 Trimming NDVI signals	42
5.4.2 Filtering training data using a iterative learning procedure	46
5.5 Robustness of SVM and RF	50
5.5.1 Effect estimation of the hyper-parameters with Noise in the train data	51
6 Conclusions	55
6.1 Recommendations	56
Bibliographic References	57
Annexes	63

1 Introduction

1.1 Problem statement and Motivation

Characterizing land use land cover (LULC) over large areas is a fundamental task in any environmental, cultural and political study since it provides a baseline for governments to undertake and monitor policies that look for sustainable livelihoods in harmony with the ecosystems. Remote sensing in synergy with image processing makes possible the identification and mapping of the land cover system, and then, the assessment and monitoring of the resources at different temporal and spatial scales ([Rogan and Chen, 2004](#)). After almost four decades of earth observation and development of powerful algorithms in mapping LULC, the research continues adapting new approaches that lead to its update to be operationally efficient and benefit from the massive data available through new technology with open data policy ([Gómez et al., 2016](#)).

The recent operation of the satellites Sentinel 2A and 2B of the European Union's Earth observation program Copernicus can play a crucial role in the new and future generation of LULC maps ([Gómez et al., 2016](#)). With an increase of the revisit frequency and better spatial resolution imagery -as never before- the research can improve the lack of production of global and regional LULC maps with a fine scale and up-to-date information. In this context, the availability of intra-annual maps can be central in a broad spectrum of applications such as forest fire propagation ([Navarro et al., 2017](#)), crop monitoring ([Vuolo et al., 2018](#)), inundation mapping ([Kordelas et al., 2018](#)), and climate change models([Radoux et al., 2014](#)). However, increasing the continuity in time of the characterization of the land cover system over large areas by using this new technology can also

1 Introduction

lead to new operational challenges. Specifically, in making operationally efficient its production while its consistency and accuracy keeps high quality. ([Radoux et al. \(2014\)](#),[Bontemps et al. \(2016\)](#)).

Traditionally, remote sensing image classification is an automatic approach of making LULC mapping where the level of human intervention may vary depending on the classification procedure ([Rogan and Chen \(2004\)](#),[Inglada et al. \(2017\)](#)). In this context, unsupervised procedures (K means, SOM) are attractive for an automatic definition of classes boundaries in comparison with supervised classifiers (RF, SVM, Maximum likelihood) that require the construction of training data ([Mather and Tso, 2016](#)). However, in the context of classification over large areas, high dimensionality and fast reproduction, unsupervised methods can lose effectiveness due to the time-consuming post-processing and complexity in the interpretation of clusters ([Chen and Gong, 2013](#)). Therefore, the better experience with supervised classification over large areas with high dimensionality ([Khatami et al. \(2016\)](#), [Colditz et al. \(2011\)](#)) raise the question if an automatic collection of training labels can overcome its drawback of needing reference data for its implementation.

The availability of previous maps in the study area represents an essential 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 contain no informative labels that can hinder the results in the classification ([Pelletier et al., 2017b](#)). In this context, predefined labels by the old maps can be contributive to the classification of one recent image or not depending on different factors. For example, point random extraction can intersect complexities of a wide diversity of classes that were simplified in one class in the map or intersect outdated labels.

Support Vector Machine (SVM) and Random Forest (RF) represent state of the art algorithms for its application in the production of LULC ([Thanh Nho and Kappas, 2018](#)); important for their ability to handle high dimensionality, being superior to unsupervised approaches and being insensitive to overfitting ([Bishop, 2006](#)). Although these methods are also known for being resistant to anomaly

1.1 Problem statement and Motivation

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 aims to refine the sampling by exploring the viability of implementing two cleaning procedures; one based on margins of NDVI signal and other based on an iterative learning procedure that uses boosting and the measurement of information entropy to control quality data.

While exploring the advantages to work with old maps as reference data for the classification of Sentinel 2 imagery, this thesis also aims to evaluate the consistency and effectiveness of the production of BAP composites as input in the production on intra -annual LULC mapping over large areas. Since the high revisit frequency of Sentinel 2 does not guarantee free-cloud imagery, BAP approach can reproduce imagery without cloud contamination ([White et al., 2014](#)). BAP 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 ([Hermosilla et al., 2015](#)). In this context, based on experiments of [Holben \(1986\)](#) with VHRR time series. That is, making 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.

Therefore, the proposed classification procedures are tested on Sentinel-2 images acquired in 2017. The training data is extracted from COS map of Portugal 2015 ([Caetano et al., 2015](#)) (Uso e Ocupação do Solo), whereas the validation corresponds to two datasets, one out-of-date that corresponds to a fraction of the dataset 2015 and another constructed using satellite image interpretation in 2017.

1.2 Objectives

1.2.1 Research question

This thesis formulated the following research question:

- How useful is the integration of training sampling extracted from old maps for an automatic production of intra-annual land cover mapping?

1.2.2 Aim

To answer the research question we defined that the main objective of this thesis is:

- To evaluate the performance of the integration of training samples extracted from old maps in a supervised classification schema for the production of intra-annual mapping.

1.2.3 Objectives

In order to achieve the formulated general aim, this thesis proposed the following specific objectives:

- To evaluate the performance of RF and SVM in the classification of Sentinel 2 images 2017 using training data extracted from the national LULC map of Portugal COS 2015?
- To evaluate the usability of COS map 2015 in the automation of the classification of Sentinel 2 imagery in 2017.
- To explore the viability of implementing a refining procedure of mislabeled data from old maps based on margins of NDVI signals, and an iterative batch learning procedure that can lead to an improvement of the performance of image classification.
- To evaluate the consistent and efficiency of the BAP composites in the production of intra-annual LULC mapping over large areas?

1.2 Objectives

Finally, this thesis had a focus on processing Sentinel 2 data products in Python and R. Readers interested in the used scripts to achieve the above objectives can visit the following [link in GitHub](#).

2 Literature Review

This Chapter will provide an overview of the literature relevant for the study. The topics covered are land cover mapping and image classification in [Chapter 2.1](#), Best pixel available composites in [Chapter 2.2](#), automatic selection and refinement of training data for supervised classification in [Chapter 2.3](#), and the basis of the implemented classification algorithms in this study in [Chapter 2.4](#).

2.1 Land cover mapping and image classification

The synergy of earth observation and image processing has made possible the monitoring and identification of the land cover system at a global and regional scale ([Rogan and Chen, 2004](#)). Despite the early experience of thematic mapping with Sentinel 2 imagery (data from 2015), it has shown potential in a different number of applications. Specifically, in the production of a new generation of land cover maps that consider fine regional scale and up to date information ([Kordelas et al. \(2018\)](#), [Vuolo et al. \(2018\)](#), [Navarro et al. \(2017\)](#)).

Traditionally, global and regional LULC mapping uses image classification to identify and monitor the land cover system automatically ([Rogan and Chen, 2004](#)). From this angle, the task of image classification can be seen from the perspective of clustering cases by their relative spectral similarity (i.e., unsupervised) or in the localization of cases that match predefined classes that have been characterized spectrally (i.e., supervised). Unsupervised algorithms (SOM, K-means) uses the spectral information derived by the satellite to define clusters. Since the clusters do not necessarily reflect the data classes under analysis, the product pass by post-processing in order to merge the clusters with specific homogeneous

2 Literature Review

variance (Wang and Feng, 2011). Increasing the dimensionality both in features and area of analysis can turn out in a time-consuming and not effective operational task due to the post-processing and the complexity in the interpretation of the clusters (Vesanto et al., 2000). Alternatively, supervised classifiers (SVM, RF, Maximum Likelihood) labels the pixels according to their similarity of samples labeled a priori using reference data (Mather and Tso, 2016). The quality of the reference data is determinant since they must represent the problem extensively; otherwise, the classifier is unable to generalize the problem to unseen pixels (Bousquet et al., 2004). The task of collecting data is usually time-consuming since it is collected manually. Although the accumulation of databases to perform semiautomatic classification, the data that come from the reference map is not error free. This experience can lead to poor results and biased classification (Pal and Mather, 2006).

Besides the experience of classification in LULC mapping according to the basis of the training process, the classifiers can also be categorized on the bases of theoretical models (parametric and not parametric). The primary appeal of using nonparametric models (RF, SVM) in comparison with parametric (ML) is based on its free assumptions of distributions and its high flexibility for adapting to different kind of variables (Mather and Tso, 2016). However, its application depends strictly on the quality and volume of the data (Pal and Mather, 2006).

So far, we have briefly introduced the image classification as a usual method in the production of land cover mapping. Therefore, besides giving particular interest to the classifiers and quality of data, in the next Chapter, we will make a parenthesis in order to talk about the quality of the imagery in the classification.

2.2 Best pixel available

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 several kinds of protocols, which mainly depend on the use of NDVI values and distances to the masked clouds to define

2.3 Automatic selection and refinement of training data

the best available pixel(BAP) ([Hermosilla et al., 2015](#)). The first examples of this approach obey to applications in the 80s using AVHRR and MODIS images ([Holben, 1986](#)). Primarily, 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 BAP composites has become technically feasible. New attempts for the selection of the best pixel aim at implementing several rules respect distance to the clouds and including different kind of sensors at the same time. For example, a scoring protocol proposed in [White et al. \(2014\)](#) that seek to weight every pixel of the time series of images according to their proximity to the clouds, date of analysis and type of sensor.

2.3 Automatic selection and refinement of training data

Since the task of collecting data manually for the process of classification is usually time - consuming, the strategy to obtain it automatically resides in the idea of extracting labels of available databases or old maps ([Inglada et al., 2017](#)). However, the use of that reference data can be constrained by two different reasons as discussed in [Pelletier et al. \(2017b\)](#) and [Radoux et al. \(2014\)](#). On the one hand, the gap between the time production of the map and the date of the image acquisition can turn out in out-up-date labels. On the other hand, effects of spatial continuity in the map can turn out in points retrieving wrong labels in areas of a mixture of classes that were simplified according to the minimum mapping unit of the map. Generally, the presence of this inconsistencies may be not a problem due to the the ability of some classifiers, such as RF and SVM, to deal with anomaly data ([Pelletier et al., 2017a](#)). However, the systemic presence of erroneous labels can lead to impact negatively the results of the classification ([Tolba, 2010](#)).

2 Literature Review

The operational strategies of removing possible outliers in the training data can be seen from the perspective of exploring the spectral dispersion of the data from different dimensions, or from the perspective of iterative learning procedures ([Chandola et al., 2009](#)). On the one hand, distances are crucial in the definition if one point behaves or not like anomaly data. For example, in [Radoux et al. \(2014\)](#) is implemented multiclass border reduction filter (MBRF), where pixels that have the most significant number of neighbors of the same class are inlier for the classification. In other applications such as [Meroni et al. \(2019\)](#),[Xu et al. \(2016\)](#) distances are also crucial; the refining procedures consists of taking out samples that are not within the margins of NDVI signals per class through the year.

On the other hand, the iterative learning procedures consist of integrating to the classification scheme an evaluation of fractions of repetitive misclassified points. In this context, active learning procedures can be roughly divided into online learning and batch learning based techniques ([Hazan et al., 2016](#)). In online learning, a classifier selects and adds iteratively samples that are informative for the definition of cluster boundaries ([Tuia et al. \(2012\)](#),[Tuia et al. \(2009\)](#)). While selecting and updating the necessary train data, it takes out not informative data for the classification. However, in the context of batch learning, the approaches define the contrary. That is, while it removes not informative data, it keeps cluster boundaries ([Büschenfeld and Ostermann \(2012\)](#), [Wu et al. \(2004\)](#),[Pelletier et al. \(2017b\)](#)).

2.4 Classification algorithms and tunning

parameters

Machine learning algorithms as Random forest and support vector machines are widely used into remote sensing for being able to handle high data dimensionality while being insensitive to over-fitting ([Mather and Tso, 2016](#)). In the frame of classification with noise that came from the out-up-date map, RF and SVM can be crucial due to their characteristics of being insensitive to outliers in the

definition of cluster boundaries. Every classifier has its own type of parameters and versatility to handle different types of data. Defining which one is better than others 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 and optimization of their parameters. In the following two chapters, 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 minimizes the error of classification by creating a 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. \(2017a\)](#)). However, it requires 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 due to 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 n_{tree} 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 n_{tree} 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 a 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, such as 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 to [Foody \(2002\)](#), the confusion matrix is currently the core of the accuracy assessment in the literature. This matrix consists 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 to the performance of the classification over specific classes. Many metrics of classification accuracy can be derived from a confusion matrix ([Foody,](#)

2.5 Accuracy assessment in land cover classification

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 vast spectrum of applications (Pelletier et al., 2017a). However, in particular, for this thesis, we highlight its importance for working as a 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 a balance of the effects of chance agreement, Cohen's kappa is proposed. Unlike OA that ranges between 0 and 1, kappa varies between -1 and 1.

3 Data and study area

In this chapter, we introduce the study area and the data to conduct this research. Therefore, In [Chapter 3.1](#), we start with the location of the images to classify and the discussion of the interest of developing the methodology in this region. In [Chapter 3.2](#), we introduce relevant technical specifications of the available images during the period of analysis for the classification. To conduct the automatic selection of training data we introduce in [Chapter 3.3](#) the old map COS 2015. Finally, in [Chapter 3.4](#), we introduce an external and updated dataset developed by DGT in order to test the results of the proposed methodology.

3.1 Study area

We conducted LULC mapping using classification of a series of tiles of Sentinel 2 located at central of Portugal. Figure [3.1](#) shows the location of the study area in red color with a dimension of 100km by 100km. The tiles intersected Tagus river in the region of Santarém and Vila Franca de Xira at the head of the long narrow estuary.

3.2 Sentinel 2 Imagery

This thesis used Sentinel-2 image time series acquired in 2017. The instruments considers 13 spectral channels in the visible/near infrared (VNIR) and short wave infrared spectral range (SWIR); bands came in 10, 20 and 60 meters resolution. The images were downloaded by using the online system Copernicus Open Access Hub developed by ESA. The imagery was utterly free. An overview of the images

3 Data and study area



Figure 3.1: Study area

under study, the percentage of clouds and the level of the available images are shown in 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 the percentage of clouds).

The products of Sentinel 2 came in two different levels 1C and 2A. L1C corresponds to an 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 postprocessing of L1C product to provide reflectance measurements on the bottom of the atmosphere. In this context, products that came in level 1C were

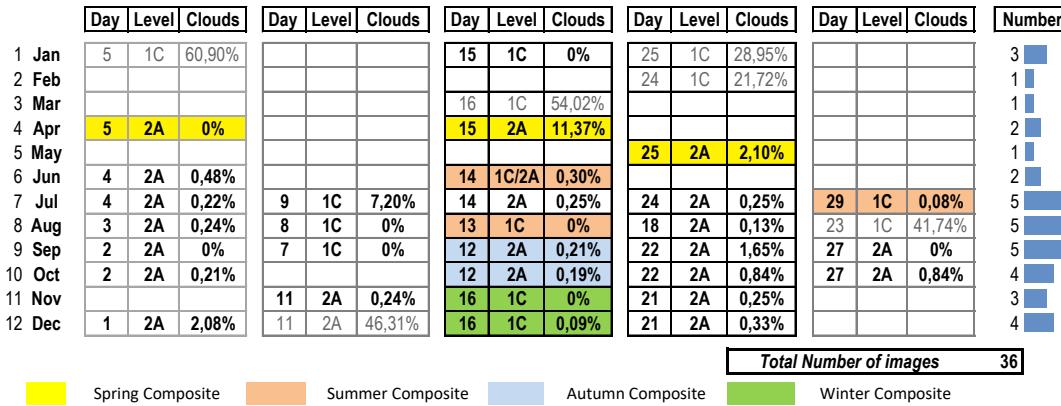


Figure 3.2: Available images Sentinel 2, 2017

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 bands Band 9 Water vapour and Band 10 SWIR Cirrus are not part of this study. 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 digital elevation model (DEM) developed by The National Aeronautics and Space Administration (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 minimum mapping unit (MMU) of analysis.

3.3 COS map

COS map is a national product describing LULC of Portugal with a MMU of 1 Ha and produced using aerial image interpretation. COS map has 5 levels and the level with more detail comprises 48 different classes. DGT is the institute in charge of producing this map and has available four versions (1995, 2007, 2010

3 Data and study area

and 2015). Moreover, COS is a composition of polygons, where each polygon represents a homogeneous unity of use and occupation of the soil. Each polygon represents an area of land greater than or equal to the defined MMU of 1 ha, with a maximum distance between lines or equal to 20 m and which percentage of a given class is equal or greater than 75% of the total area.

According to COS 2015, the study area covered a wide variety of land cover types, such as urban fabric (10.1%), agricultural areas (50.1%), water-bodies (1.4%), wetlands (0.1%), forest and semi-natural areas (38%). 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 diversity in the agricultural practices and phenology activity corresponded to an desirable scenario for conducting this thesis (see figure 3.3).

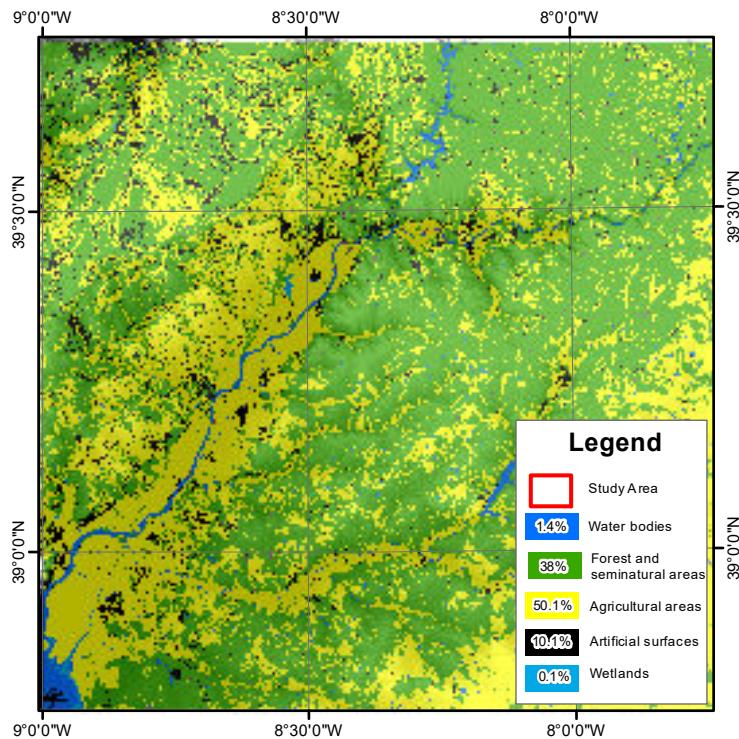


Figure 3.3: COS Map 2015, Nomenclature Level 1

Consequently, the intersection also resulted in 21 different classes with Level 5. This first inspection led to construct table 3.4. According to COS nomenclature (pink color), the categories diversifies depending on their level of description. For

example, the agricultural areas comprise 2 categories for level 2: periodic and permanent crops; in this sense, periodic: diversifies in irrigated and not irrigated crops, and permanent: in rice fields, grasslands, vineyard, and olive trees.

Automation of the classification process requires the production of LULC maps with simple nomenclature that allows both an operational replication and a high quality production. Therefore, this thesis proposed to follow recent recommendation of European Environmental Agency on Copernicus land monitoring in the future generation of Corine land cover maps ([EEA, 2018](#)). In this context, we undertook its relation with COS map nomenclature having as a result the gray section in again table [3.4](#). On the one hand, nomenclature of Dataset 1 matches with the level 1 of nomenclature COS 2015 for the categories of Herbaceous (Agricultural areas), sealed surface (Artificial surfaces), water surfaces (water bodies), non vegetation and wetlands (see Figure [2](#) in Annexes). However, for the specific case of forest and semi-natural areas, dataset 2 diversify the category in 4 subgroups: non-vegetated, shrubs, Coniferous, Eucalyptuses and Holm and Cork trees (see Figure [3](#) in Annexes).

Even though Vinhas and Olivais classes are also part of the intersection of classes in the study area, this thesis do not consider them. Generally, vineyard and olive classes are plantations with fruits. These herbaceous categories have a specific case of sowing in furrows, where the space between them lead to having a mixture of not vegetated and vegetated classes in the pixel. Therefore, since the rest of the categories cover extensively herbaceous class, we did not consider to use them as training data of our models.

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 correctly the class variation without the exposition to mislabel data due to phenology. In this case, in collaboration with

3 Data and study area

Dataset 1	Dataset 2	Nomenclature COS 2015					
		Level 1	Level 2	Level 3	Level 4	Level 5	
Shrubs	Woody	Florestas e meios naturais e semi-naturais	Vegetação arbustiva e herbácea	Matos	Matos	Matos	
Coniferous trees			Florestas	Florestas de resinosas	Florestas de resinosas	Florestas de pinheiro bravo	
Eucalyptus trees				Florestas de folhosas	Florestas de folhosas	Florestas de pinheiro manso	
Holm and Cork Trees			Zonas descobertas e com	Zonas descobertas e com pouca vegetação ou com	Espaços descobertos ou com pouca vegetação	Florestas de eucalipto	
Non vegetated	Non vegetated	Áreas agrícolas e agro-florestais	Culturas temporárias	Culturas temporárias de sequeiro e de	Culturas temporárias de sequeiro e de	Florestas de sobreiro	
Herbaceous	Herbaceous		Culturas permanentes	Arrozais	Arrozais	Florestas de azinheira	
				Pastagens permanentes	Pastagens permanentes	Pastagens permanentes	
				Vinhos	Vinhos	Vinhos	
			Olivais	Olivais	Olivais	Olivais	
Sealed surface	Sealed surface	Territórios artificializados	Tecido urbano	Tecido urbano contínuo	Tecido urbano contínuo	Tecido urbano contínuo	
				Tecido urbano descontínuo	Tecido urbano descontínuo	Tecido urbano descontínuo	
			Indústria, comércio e transportes	Indústria, comércio e equipamentos gerais	Indústria, comércio e equipamentos gerais	Indústria, comércio e equipamentos gerais	
			Redes viárias e ferroviárias e	Redes viárias e ferroviárias e espaços	Redes viárias e ferroviárias e espaços	Redes viárias e ferroviárias e espaços	
Water surfaces	Water surfaces	Corpos de água	Planos de água	Cursos de água	Cursos de água	Cursos de água	
				Planos de água	Planos de água	Planos de água	
			Planos de água	Planos de água	Planos de água	Planos de água	
			Águas marinhas e costeiras	Desembocaduras fluviais	Desembocaduras fluviais	Desembocaduras fluviais	
Wetlands	Wetlands	Zonas húmidas	Zonas húmidas	Zonas húmidas	Zonas húmidas	Zonas húmidas	

Figure 3.4: Class definition of the two datasets and COS nomenclature 2015

DGT, we had access to an external dataset that corresponds to a satellite image interpretation of 557 random samples using the image of 29 of July of 2017. The distribution of samples is the following: Shrubs: 61, Coniferous trees:35, Eucalyptus trees: 55, Herbaceous: 217, Holm and cork trees: 38, Non vegetated:69, Sealed: 11, Water:55, Wetlands:16.

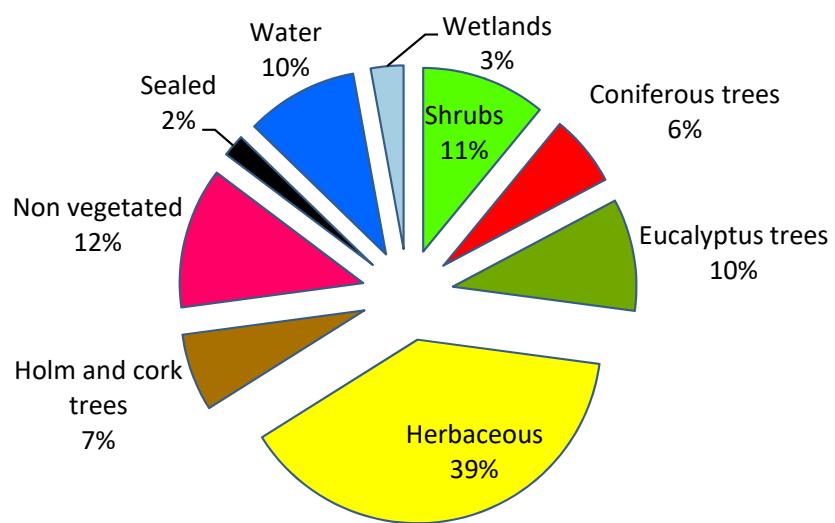


Figure 3.5: Distribution of the number of samples of the external dataset per percentage

4 Methodology

While this thesis aimed to make automatic the production of intra-annual maps implementing a workflow that consisted of supervised classification in synergy with automatic extraction of training samples from an old map, it also aimed to use singular and BAP composites (see Figure 4.1). In this context, after a preliminary selection and preprocessing of the implemented features in the classification ([Chapter 4.1](#)) both from single images and BAP composites ([Chapter 4.2](#)), a random selection of training points is extracted from an old map and intersected with the spectral information of the images ([Chapter 4.3](#)). We performed a classification scheme using SVM and RF classifiers using two datasets with six and nine different number of land cover classes ([Chapter 4.4](#)). The out-of-date information derived from the old map led us to evaluate the viability of implementing two refining procedures over the data in order to improve accuracy ([Chapter 4.5](#)); one based on margins of NDVI signals and another based on an iterative learning procedure. Besides that, the possible inconsistencies of the labeled data led us to evaluate also the robustness of classifiers RF and SVM by injecting different levels of noise ([Chapter 4.6](#)).

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

4 Methodology

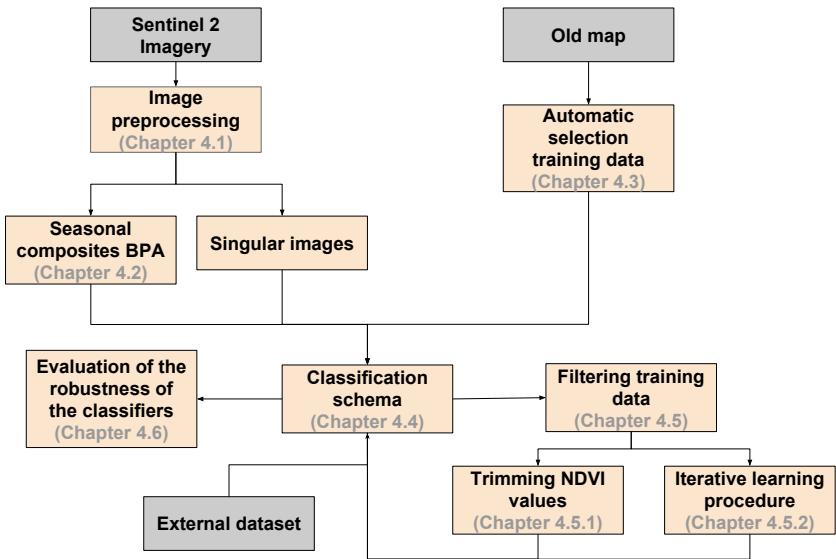


Figure 4.1: General Methodology

was to use only bands of 20 and 10 meters; bands of 20 meters were subject of a re-sampling process at 10 meters.

Besides 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 Mission (SRTM). Consequently, the slope was a byproduct made in QGIS using the DEM as a reference. Once both products were obtained, they were subject of a resampling 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 an additional feature in the task of classification and index for the construction of the seasonal pixel-based composites. Generally, from the measured reflectance on the near and red infrared, we calculated NDVI. 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} corresponded 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 standard normalization, but NDVI. We repeated the same process for each image available in 2017.

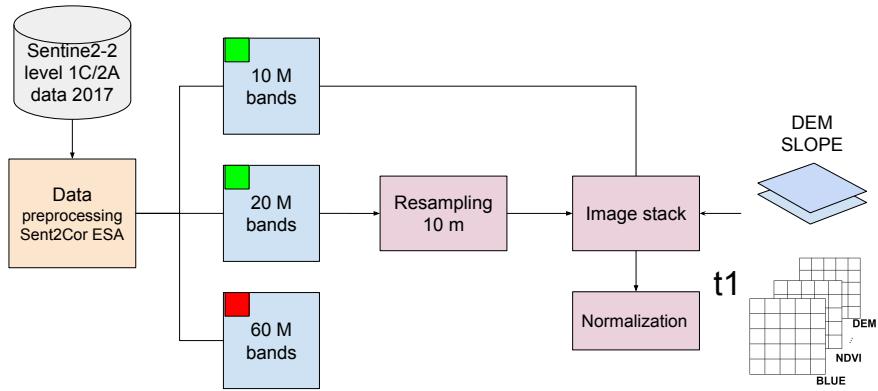


Figure 4.2: Methodology image preprocessing

4.2 Seasonal Composites

The generation of BAP composites was aimed at reproducing cloud-free and phenological consistent image composites for the seasons of 2017 as a case of study of intra-annual mapping. Primarily, 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 with the pixel with the highest NDVI in the classification. According to figure 4.3, we show three series of images highlighting 3 pixels over time and 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 pixels for that time. This methodology was straightforward, fast and depended only on the spectral information at the level of the pixel. The Process was evaluated in 4 suites of images per season (after preprocessing). The table 3.2 in the Section 3.3 showed the images used for the composites. It should be noted, that the imagery was preselected according to the seasonal precipitation conditions in Portugal during the year 2017 (see Annexes 1).

4 Methodology

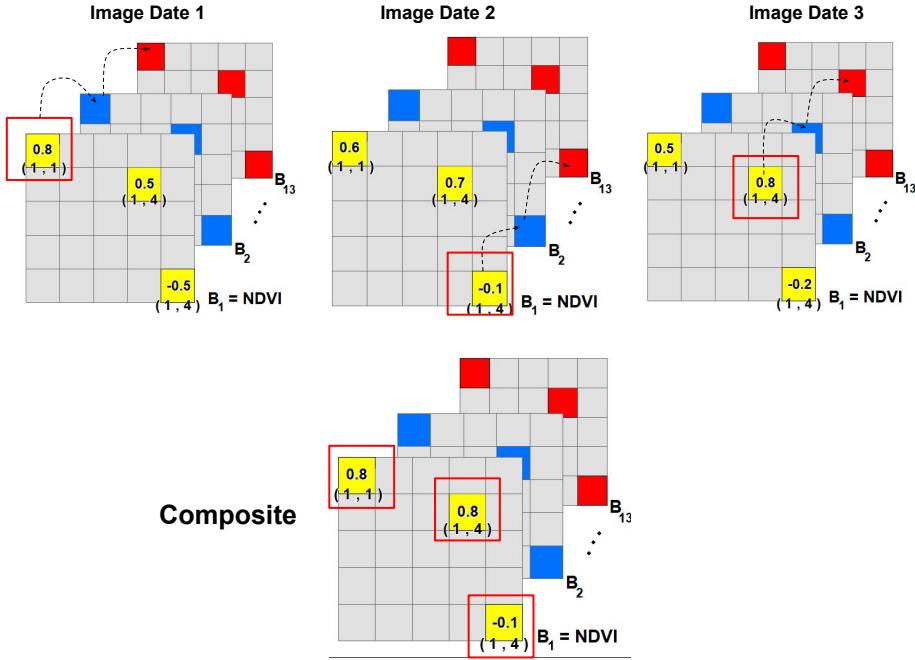


Figure 4.3: Methodology composites

4.3 Automatic training data from a old map

Taking as reference the old map (COS 2015), we randomly sampled the study area. Depending on the number of classes to discriminate, we perform a equalized random selection of 1000 samples per category. Figure 4.4 shows an example of the selection of samples for four type of land cover classes.

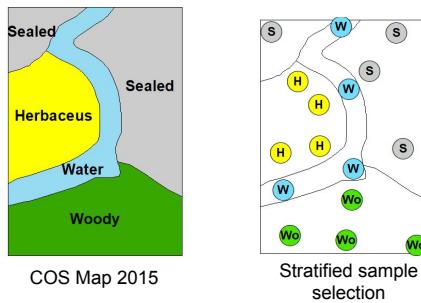


Figure 4.4: Methodology selection points from old map

Moreover, the external test set TB considered the same process. However, instead of having regular number of samples per category, the selection was un-

balance.

Both dataset contained the labels of land cover and coordinates of the points. Therefore, in order to add the spectral and ancillary data information per point over time we performed a spatial and temporal join (see figure 4.5).

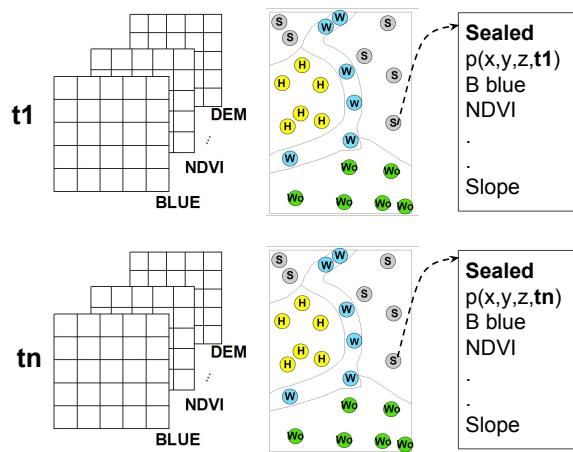


Figure 4.5: Methodology extraction features from images

4.4 Classification schema

In a frame of a supervised classification scheme, traditionally, people split training data during the classification into two sets: Train and Test. Both correspond to a fixed random choose where the model is then iteratively trained using the train set (i.e 70%) and validated using the test set (i.e 30%). Alternatively, there are many ways to split the training dataset, and this arrange is usually called cross validation. Unlike the previous arrange, we can generate multiple splits of the train set, so that we only fixed the initial test set (i.e 30% TA). This fractions parts of the train set correspond to several splits (k-folds) of the train set (i.e 55% train and 25% validation). This allowed us to estimate unbiased parameters and generate stable predictions. In Figure 4.6 we can appreciate how we made this multiple splits during the classification.

We performed a pixel based classification using RF and SV classifiers. The classification models were built from each train set after carefully carrying out a

4 Methodology

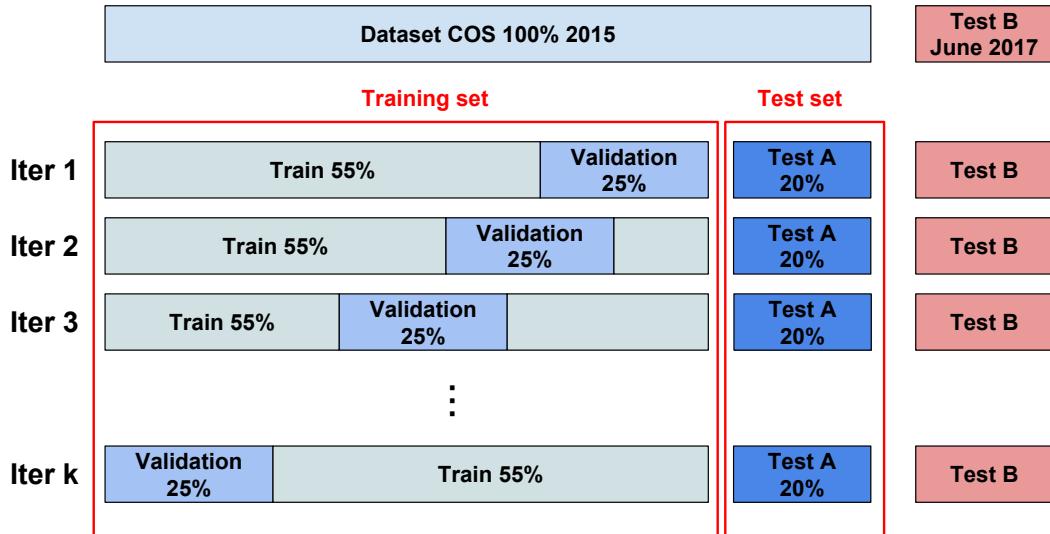


Figure 4.6: A visualization of the splits of the training dataset during classification

sensitive analysis of the parameters. In SVM, two condition were tuned; γ in the insensitive-loss function and the cost of constraints violation C , while the radial kernel function was left default. In RF, the number of variables randomly sampled as candidates at each split ($mtry$) and the number of trees to grow were tuned ($ntree$).

It should be noted, that we only had one independent dataset (TB) for only one image in June 2017. Therefore, the only image with a test using an updated dataset is June 2017, the rest of the images used a fraction part of the sampling of the out-up-date map to test the results.

4.5 Filtering training data

While the training data selection is automatic (see section 3.3), this thesis aimed to refine the training data for each period of analysis in order to evaluate better accuracy in the classification. In this context, we used two different techniques, one based on trimming NDVI values per class depending on how far they are from their normal dispersion and another that consider an iterative learning procedure.

4.5.1 Trimming NDVI signals

We recreated the NDVI values for all points through 2017 year. In this sense, we made box-plots in order to see the dispersion of NDVI values per class and time. Figure 4.7 shows one example of the variability of three woody classes over the year 2017. The box-plots display interquartile ranges (IQR, size of the box), maximum and minimum values (limits of the whiskers) and possible outliers (black points) per scene.

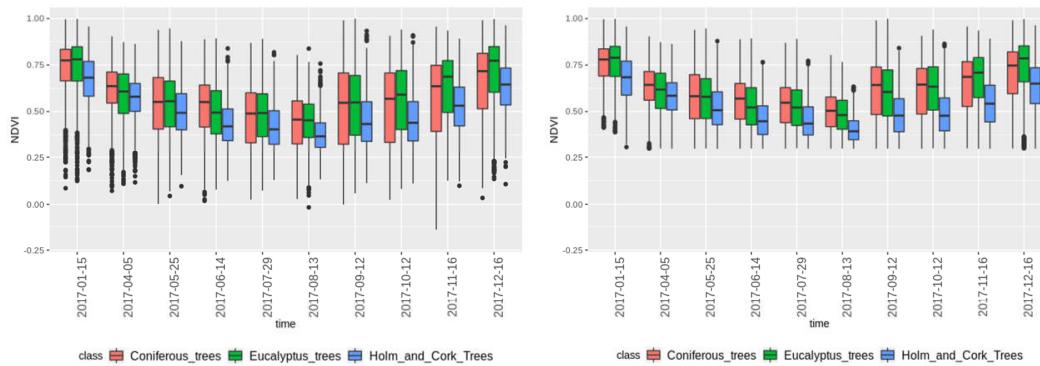


Figure 4.7: NDVI signals over time for classes of trees

Labels can intersect pixels with a 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 types of vegetation can grow up, such as bushes or grass. Therefore the variation of NDVI may be a result of how saturated or dispersed is the biomass for the specific pixel. For example, the slight increase of the median of NDVI values for trees during the beginning and the end of the year correspond to an increase of a photosynthetic activity due to rainy periods. Therefore, the fractional proportion of the class tree in the pixel tends to saturate more the pixel with respect the other fraction that consider other types of vegetation. Generally, 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, samples with NDVI values beyond 1.5 times the IQR, from the first or third quartile were removed from the data. That is, black points located beyond

4 Methodology

the length of the whiskers (see Figure 4.7).

4.5.2 Filtering training data based on a iterative learning procedure

The second strategy for removing possible anomaly data hindering the results of the classification was based on iterative learning procedure. That is, before the learning step, all the train set (80%) is subject to a quality evaluation in order to define the level of contribution of each point in the classification.

Using multiple splits of the training dataset (55% train, 25% validation), we recreated multiple models in order to classify itself. After that, we evaluated the level of uncertainty in every prediction using the measurement of information entropy (see Equation 4.2).

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

If a point was informative for the classification, that is, a point correctly predicted over different splits, then the point depicted low entropies values. After that, we normalized and inverted the entropies values in order to define a score of informativeness per point (see figure 4.8).

As lower the score, the larger the chance of the point to be removed from the training for a specific iteration. Each iteration removed 2.5% of points from the train set, so that, we repeated the process of calculating informativeness and took out points up to the specific definition of points in the train set led to stop a possible increase in the overall accuracy and reduce the predictive power of the classifier.

4.6 Evaluation of the robustness of the classifiers

To asses the classification results, we proposed to add noise to the COS dataset. Independently of the inherent noise of the out-up-date data, we decided to inject in the train set different levels of noise to evaluate their impact in the overall accuracy.

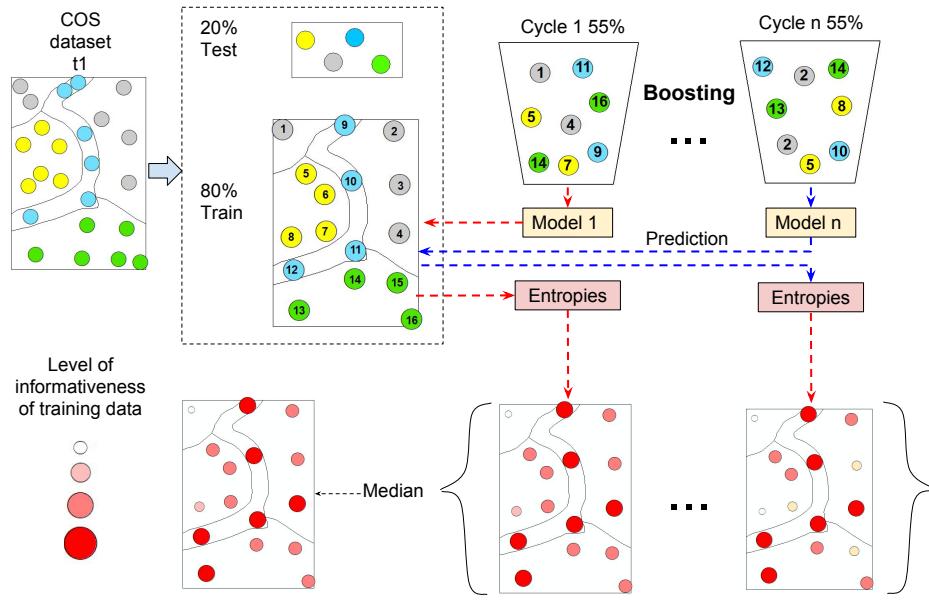


Figure 4.8: Iterative learning procedure to define informativeness based on boosting and the measurement of information entropy

The random and systematic injection of noise consist of selecting a set of points and change their original labels for one of the rest of the classes. Each injection covered 5% of the train. We performed classification using both RF and SVM considering a tune of parameters per injection.

5 Results

While this thesis aimed to make automatic the production of intra-annual maps implementing a workflow that consisted of supervised classification in synergy with automatic extraction of training samples from an old map, it also aimed to use singular and BAP composites. In this context after selecting the training data from COS 2015 automatically ([Chapter 5.1](#)), we defined a baseline to compare the classification results both using imagery that matches the year of production of the map, and imagery for the date of analysis 2017 ([Chapter 5.2](#)). After that, we set out the classification both the seasonal composites and the singular imagery; we performed an accuracy assessment based on the confusion matrix and OA ([Chapter 5.3](#)). The obtained results questioned if quality control over the training data can result in better accuracy ([Chapter 5.4](#)), and how robust are the classifiers in the presence of several levels of noise in the training data during the modeling ([Chapter 5.5](#)).

We implemented the methodology using scripts written in Python and R languages. Readers interested in the application of these procedures can visit the following [link in GitHub](#).

5.1 Automatic training selection from COS map

Figure [5.1](#) shows the spatial distribution of the equalized random sampling done over the study area. For the first dataset composed of nine classes, we selected 1000 samples per class for the categories of Shrubs, Coniferous trees, Eucalyptus trees, Holm and cork trees, Herbaceous, no vegetated areas, sealed surface, water surfaces, and wetlands. Instead, the second dataset composed of six classes, it

5 Results

comprised 1000 samples per class samples for the categories of Woody, Herbaceous, no vegetated areas, sealed surface, water surfaces, and wetlands.

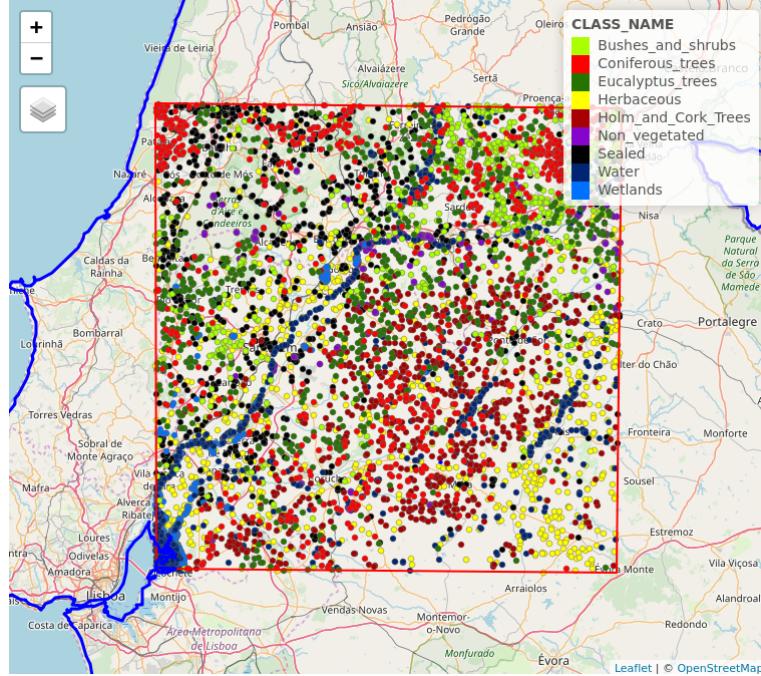


Figure 5.1: Spatial distribution of the training data

In this context, after the atmospheric and radiometric correction of the imagery, we extracted the spectral information per sample using imagery that came from 2015 and 2017.

5.2 Classification

Generally, classification accuracy depends on multiple factors, where the nature of the training samples, the number of features (bands and ancillary data), the number of classes to be discriminated, the spatial resolution of the images and the properties of the classifier are the most important. From this angle, we set out the evaluation of the accuracy of the classification influenced by the previous factors for one image in 2015 in order to use it as a baseline to judge the results over 2017.

5.2.1 Baseline

Figure 5.2 shows a comparison of OA in the classifications of one image on July 25 of 2015 using two different number of classes and two different number of bands(features). On the one hand, from the number of bands, we evaluated the adding value in the OA after including ancillary data in addition to the spectral information that came from Sentinel 2. The three additional features corresponded to NDVI, DEM, and Slope. The addition implied a significant increase in accuracy for the discrimination of classes in both dataset of six and nine classes; 5.0 scores and 3.0 scores respectively. On the other hand, the predictive power of RF increased after merging shrubs and trees variability in only one class called Woody. Therefore, the implementation of updated reference data extracted from COS map in a supervised classification of one image in July 2015 using 13 bands resulted in accuracies of 0.61 and 0.73 for the datasets of six and nine classes respectively.

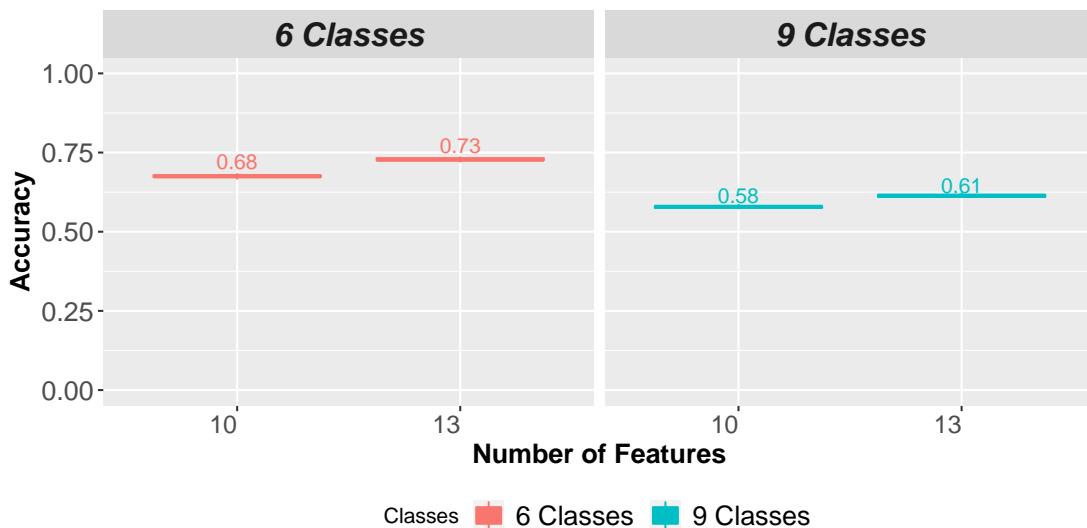


Figure 5.2: Baseline July 25 2015, RF classifier

Even though machine learning algorithms 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 to the documentation of Sklearn (library Python), RF offers a score that measures the

5 Results

importance of the study variables based on the mechanism of out-of-bag-sample error. In figure 5.3, 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 of the variable. Notably, 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 of classes with vegetation. Moreover, the DEM also was a good predictor, determining that the elevation fractionally characterizes the land cover system in the region.

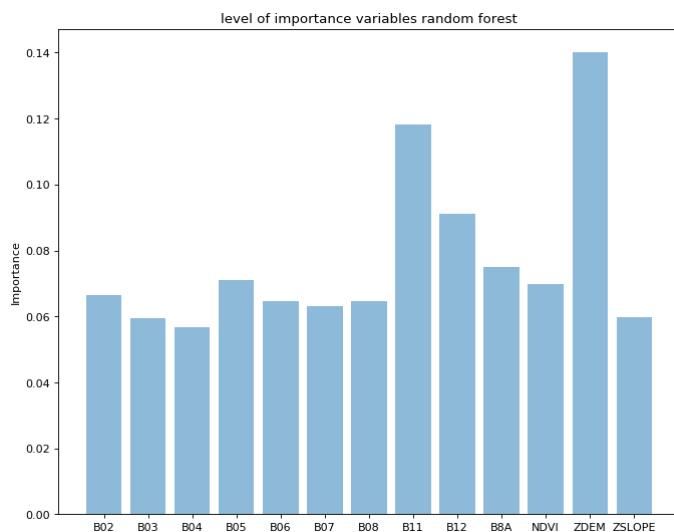


Figure 5.3: Level of importance variables random forest

5.2.2 Classification using RF and SVM (images 2017)

Figure 5.4 shows a general evaluation of the classification of images in 2017 using as reference the training data selected from COS map 2015; classification per image considered a careful sensitive analysis of the parameters of the classifiers. In summary, the implementation of SVM with a RBF kernel outperformed the results obtained with RF over all the imagery of 2017 with average differences of 2 scores.

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 to the technical

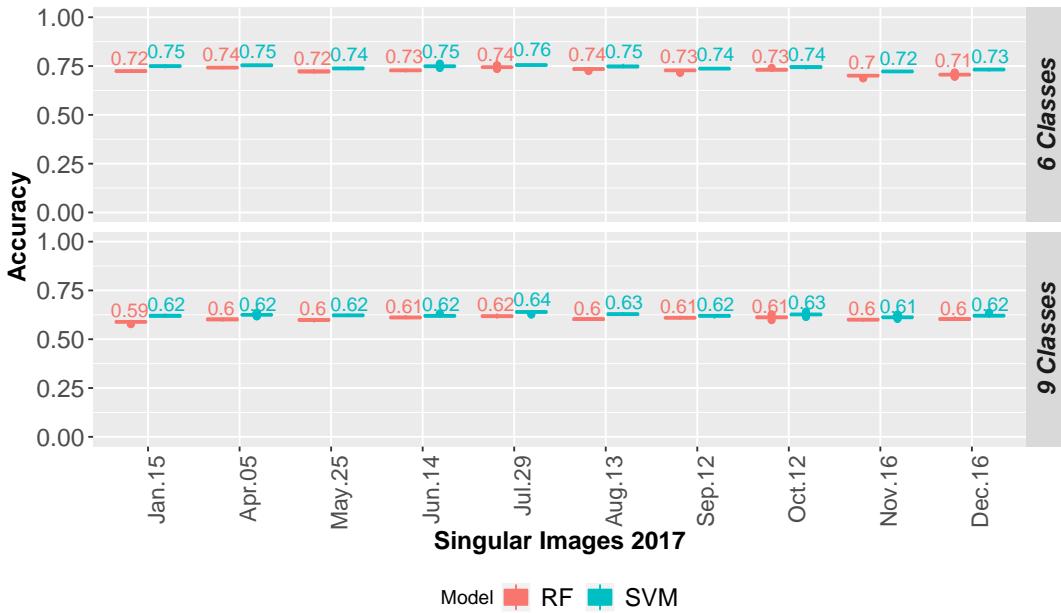


Figure 5.4: Performance classifiers RF and SVM, scene-based composites 2017

report of COS map, the Aereo-photographies were taken during April, May, and June of 2015. The match in season between image acquisition and the date of the images of reference for the production of COS map can be one of the reasons by which around these dates we had slight better results. Generally, the results ranged between the same range of accuracies than for 2015. We highlighted the OA of July 29 where the score was slightly higher respect to the rest of the images: 0.76 and 0.64, dataset with 6 and 9 classes respectively. This scenario showed first insight into the robustness of the classifiers in the presence of possible changes in the reflectance during the year due to phenology.

5.3 Single maps vs seasonal maps

We proposed two examples of intra-annual mapping, one using single images and other using seasonal composites. Regarding the construction of the composites, the computational performance of a methodology based on maximum NDVI for their construction was quite fast. Depending on the number of spectral bands and images per date to consider, the scripts developed on Python could create

5 Results

composites with Sentinel 2 imagery in terms of few minutes. Figure 5.5 shows one example of how the composites combated the cloud contamination of some images during the production of one composite in Spring. 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 allows to retrieve the best pixels associated to high values of NDVI, and therefore, the construction of artificial images free of clouds.

The production of BAP composites raised the curiosity for comparison in the classification performance between single images and composites. In the case of classification of a dataset with six categories, Figure 5.6 shows the classification evaluation using OA for the best singular images per season vs. the BAP composites. The results showed equivalent OA for all the images based in the same range of variation of the OA after cross-validation.

Moreover, with the same above specifications we compared the accuracies of the dataset with nine classes (see Figure 5.7). In summary, composites seem to give the same valuable information than pure single images for the discrimination of land cover classes. Unlike, single images, BAP composites per season are cloud-free and phenomenologically more consistent for the production of seasonal LULC mapping.

To describe the performance of the classification models per classes we created the normalized confusion matrix (see Figure 5.8). Particularly, this matrix represented the predictive power of the classifier SVM with RBF kernel in the discrimination of nine classes using an image of July 29 of 2017. From this angle, the values of the diagonal elements represented the degree of correctly predicted classes. The confusion is expressed by the false classified off-diagonal elements, since they may be mistakenly confused with the rest of classes. We warn the readers for possible underestimation on accuracy since the validation of these results are based on a test set that also contains samples from an old map. Besides that, we calculated the kappa index and accuracies for omission and commission (see table 5.1). The accuracies in the diagonal of the confusion matrix corresponded

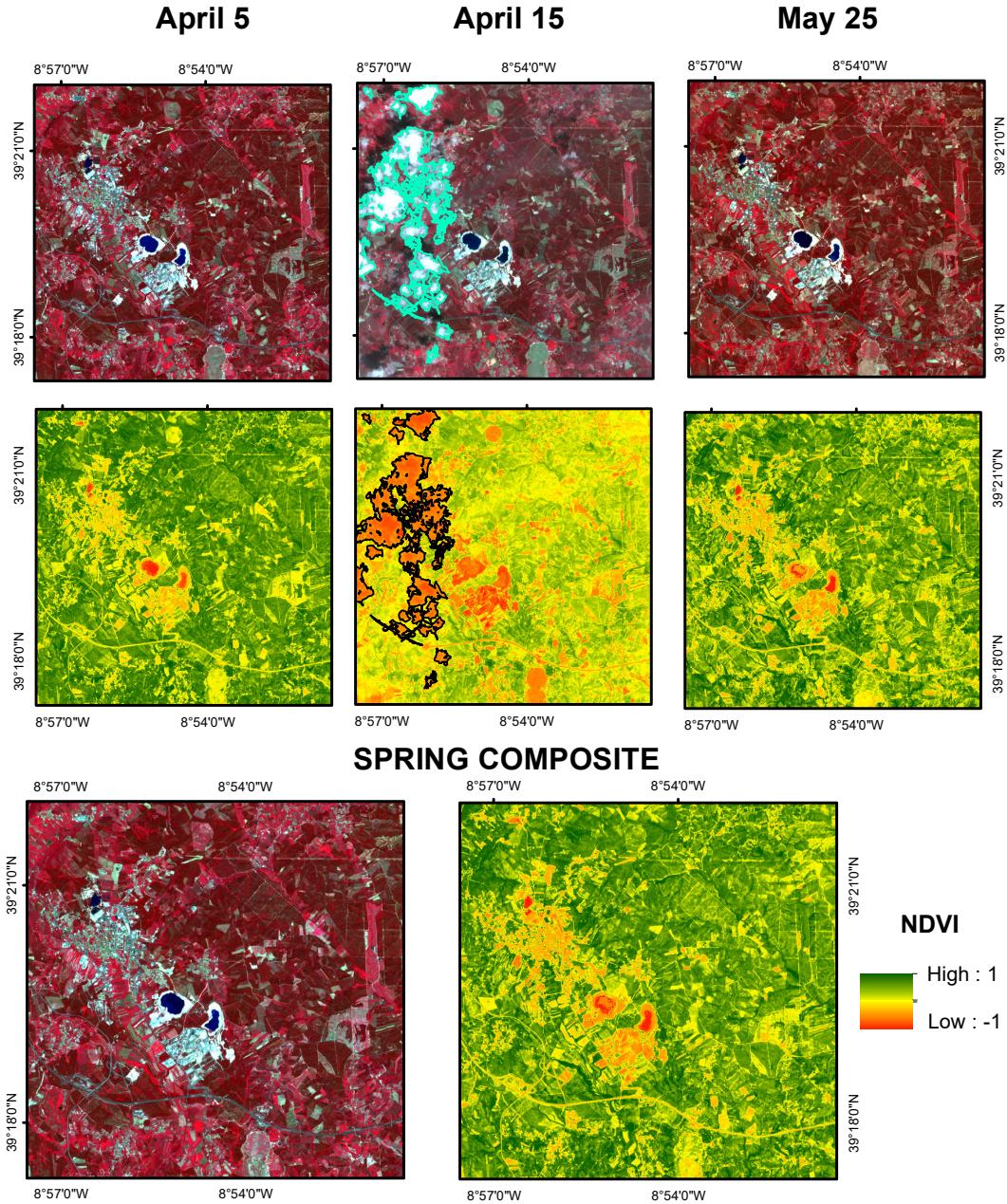


Figure 5.5: pixel-based composite Spring

to the producer accuracy. The proximity of Cohen's kappa and OA statistics explained a high agreement between the user and producer accuracy. That is, classification accuracies both for omission and commission have similar results.

Regarding the classification of the dataset with nine classes, We could divide the predictive power of the model in three categories. Firstly, classes that contained woody such as Shrubs and Trees are showing a low predictive power

5 Results

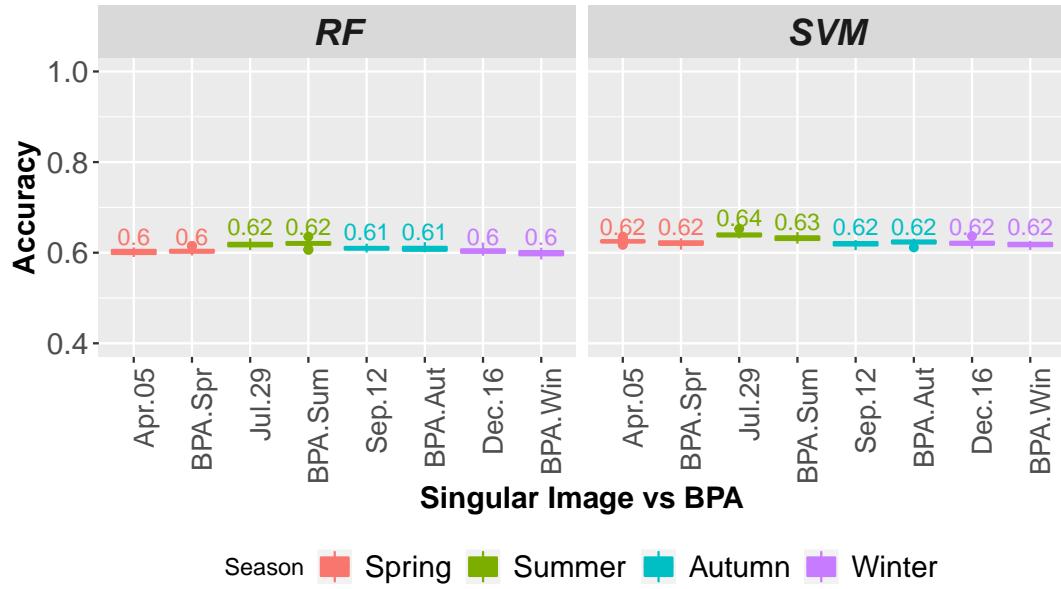


Figure 5.6: Performance of the classifications, single images vs BPA composites, 9 classes classification

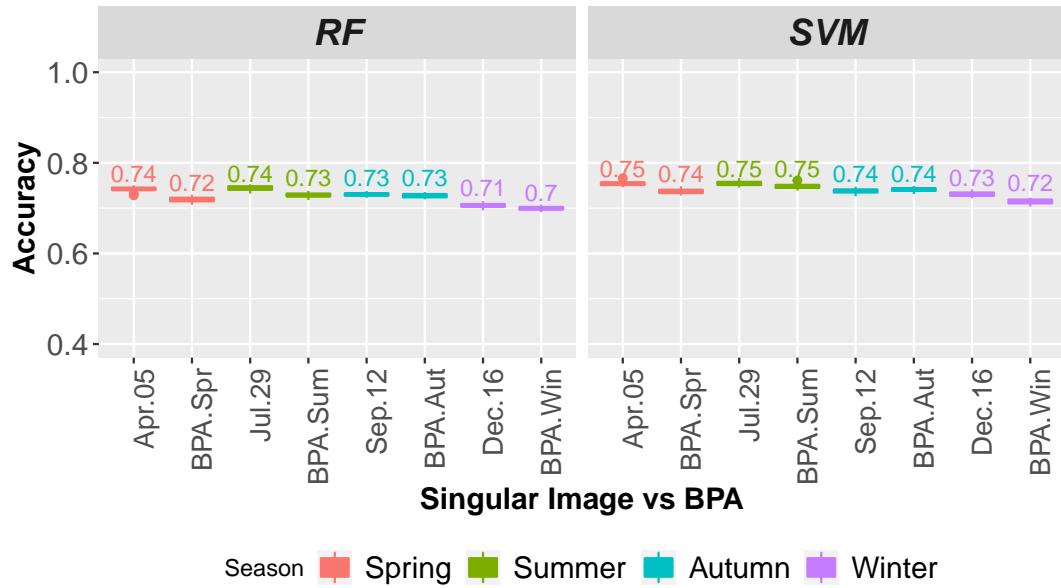


Figure 5.7: Performance of the classifications, single images vs BPA composites, 6 classes classification

of the model to discriminate and identify them; with ranges between 0.48 and 0.58. Specifically, the confusion of these classes was concentrated among them; that is why a merge of these classes in Woody for the second dataset ended up in

5.3 Single maps vs seasonal maps

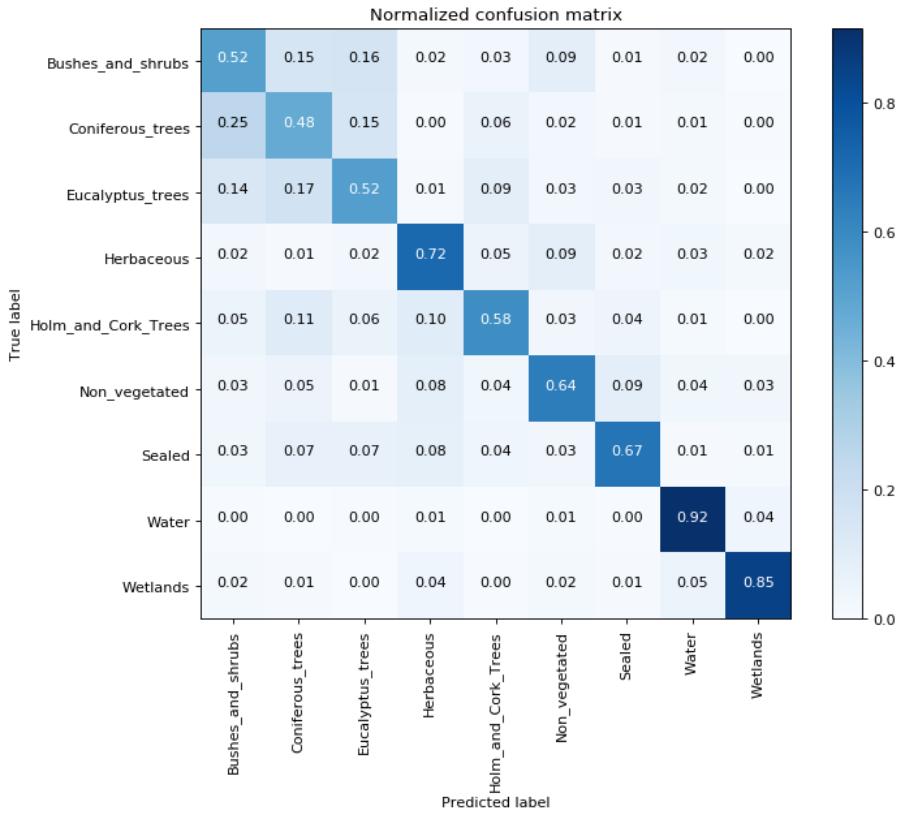


Figure 5.8: Normalized confusion matrix, classification scene-based composite

July 29

an increase of the accuracies for all the classes (see table 5.2). Secondly, classes such as wetlands and water depicted high predictive power of the model to predict them in comparison with the above classes; 0.85 and 0.92 respectively. The good perform for wetlands, and water surfaces could be a consequence of the proportion of the sampling respect to the size of the land cover classes. For example, Water and wetlands sum up an area that covered only 1.5% of the study area in comparison to 38% of classes with woody. Therefore, even though the classification was in balance keeping the same number of samples per class, the representativeness of the sampling in proportion is entirely different. Finally, the predictive power of the model to classify herbaceous, sealed and no vegetated corresponded to level medium. The model still weakly predicted these classes probably due to the presence of mixture classes in the pixels. Therefore, instead of pixel-based classification, future attempts can approach sub-pixel models.

5 Results

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 5.1: Accuracy assessment 9 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

Table 5.2: Accuracy assessment 6 classes

Figure 5.9 shows an example of the classification over a fraction part of the study area. To compare the results visually, we used NDVI and false color composition (near-8, red-4, green-3 in Sentinel 2) and the NDVI. The final seasonal maps showed that there was an apparent overestimation of herbaceous vegetation.

5.4 Filtering training data

In the next two chapters we present the results of the two cleaning procedures over the training data.

5.4.1 Trimming NDVI signals

During the automatic selection of training data for herbaceous, we considered an equalized random sampling of all categories of level 5 of COS map. That is, agricultural areas with irrigation and not irrigation, rice fields and herbaceous permanent. However, the presence of herbaceous periodic questioned if the dynamic for phenology could impact notably the number of right samples representing herbaceous over the year. Therefore, to perform classification using representative labels of herbaceous for every specific time of analysis, we set out a cleaning procedure implementing thresholds over the dispersion of NDVI values. Figure 5.10 shows an example of signals of NDVI over time for classes of permanent,

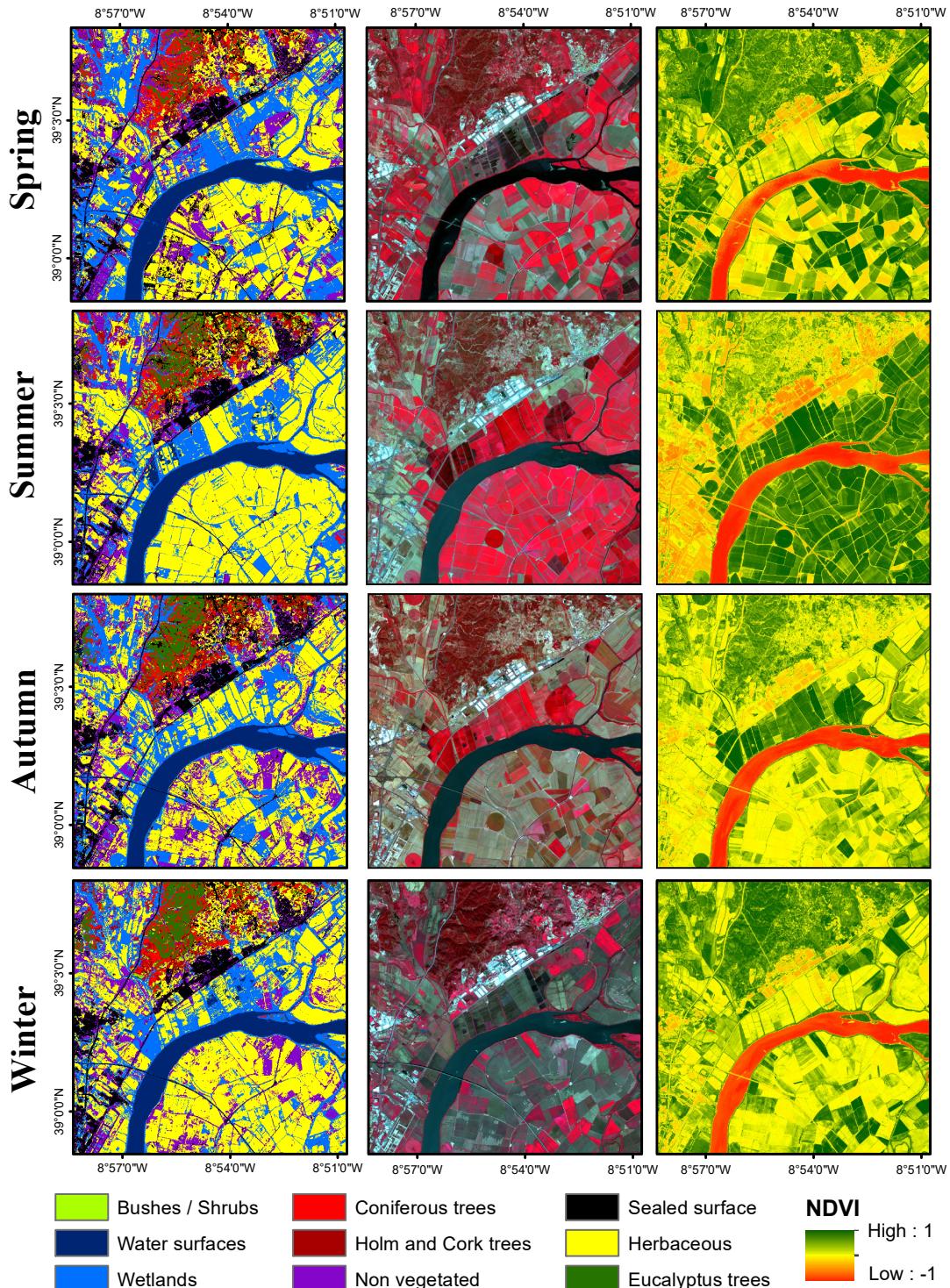


Figure 5.9: Seasonal mapping.

irrigated and not irrigated herbaceous. Particularly, these signals allow exploring how the photosynthetic activity of Herbaceous permanent and not irrigated

5 Results

crops tend to decrease during the summer. While values lower than 0.3 tend corresponded to non-vegetated surfaces, samples with NDVI values larger than 0.3 corresponded to herbaceous. Therefore, we set out the removal of samples with NDVI values lower than 0.3.

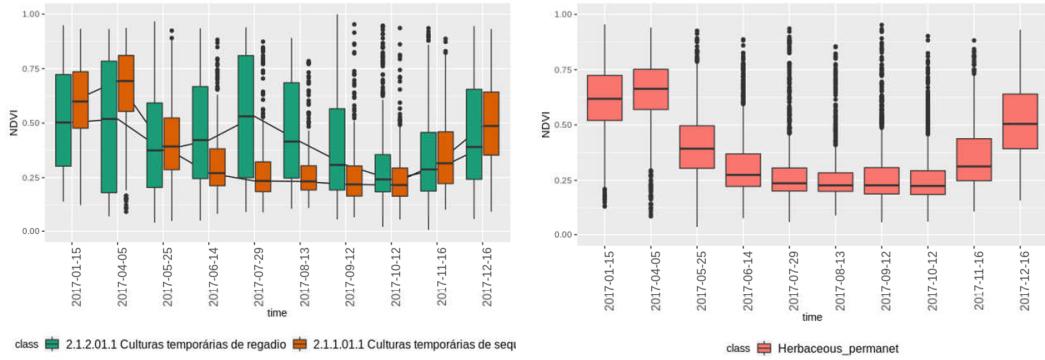


Figure 5.10: NDVI signal for herbaceous classes

The above procedure was particularly for herbaceous class. However, the rest of the classes were subject of a refinement procedure using IQR procedure. The cleaning preprocessing using NDVI variability led to a new distribution of the training data (see figure 5.11). 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 noted, that the class with more impact on the number of data per time was herbaceous due to phenology.

The figure 5.12 shows the comparison of the classifications with and without implementing the first strategy of filtering mislabel data. The classification was performed using SVM and using the two versions of test sets; TA that corresponds to a fractional part of the traning data COS 2015, and TB that correspond to an updated test for an image in July 29 of 2017. Based on comparisons of OA, the low OA in the classification over the year led to establishing 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

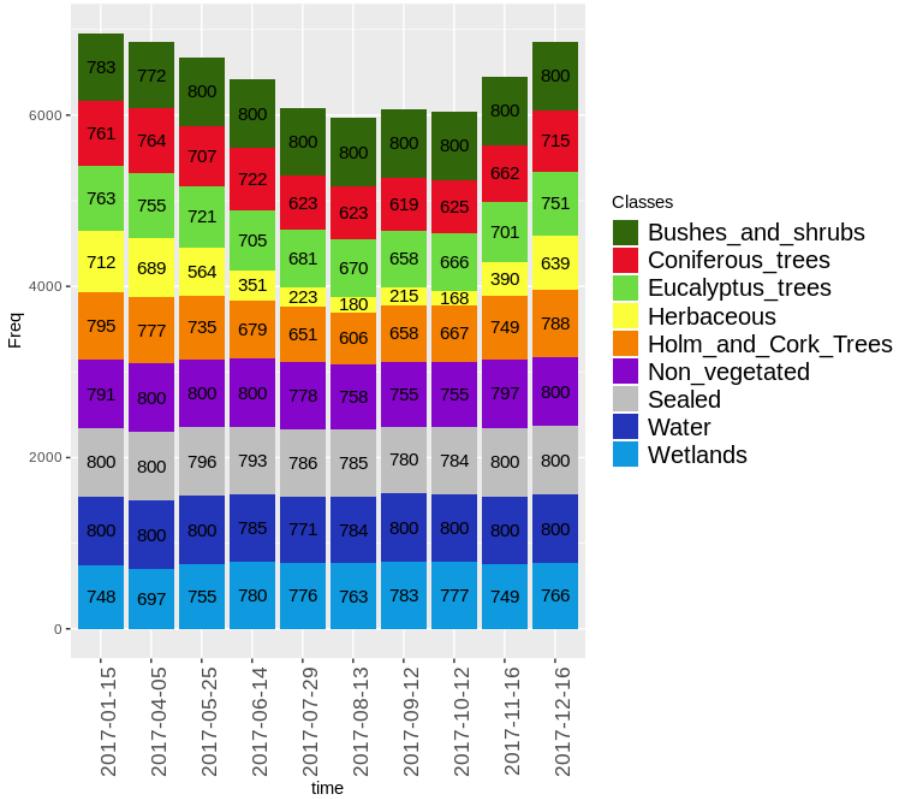


Figure 5.11: Training data distribution after applying NDVI rules

a reduction in the number of samples for vegetation, especially during summer, and therefore possibly not enough representation of them. Although the impact of unbalanced data is high in Summer, in Spring the reduction of erroneous labels was meager. However, regardless of 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 can 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 substantial changes in the amount of training data that eventually can lead to defining new functionals forms in the prediction, and therefore, to lose the ability to generalize unseen data.

So far, we have introduced the results of the first sampling refining method.

5 Results

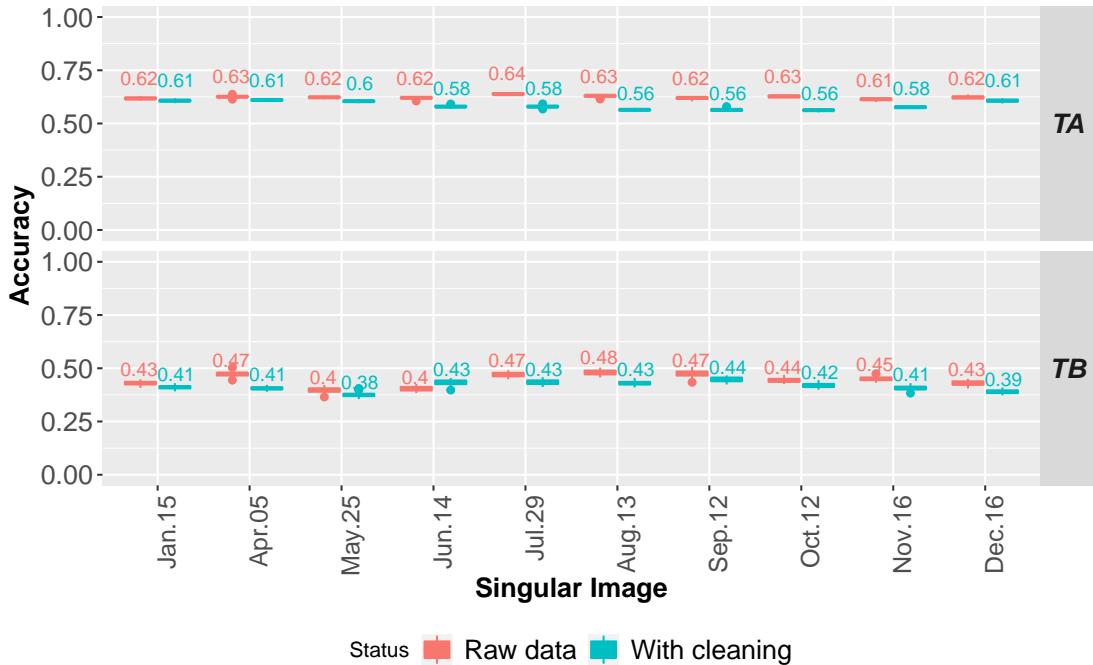


Figure 5.12: Cleaning processing NDVI, Performance over scene-based composites 2017

These results were fundamental to undertake a more sophisticated strategy during the development of this thesis.

5.4.2 Filtering training data using a iterative learning procedure

In figure 5.13, we show a normalization of the level of informativeness for the training data associated with herbaceous under different seasons. The change in the level of informativeness of the samples depends 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 led to define different states of the label, and therefore raise the question of whether removing mislabeled data by this effect could have implied an improvement in accuracy

classification on an image of a specific time.

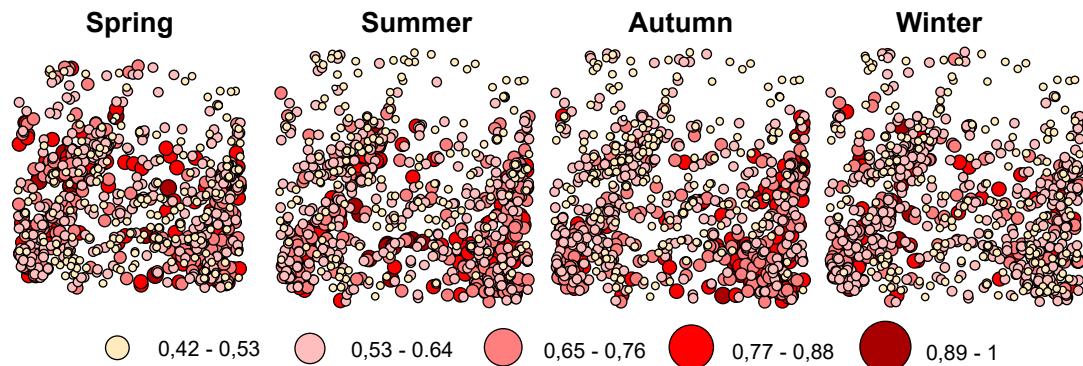


Figure 5.13: Informativeness score for herbaceous

Another example derived from the spatial continuity 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 to 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 overlap a mixture of different classes 5.14. We expected that labels misrepresenting the pixel produced the lowest values of informativeness. However, concerning the mixture of classes in one pixel, this turned out in one constraint of the proposed methodology since it did not consider fractional proportions of classes per pixel.

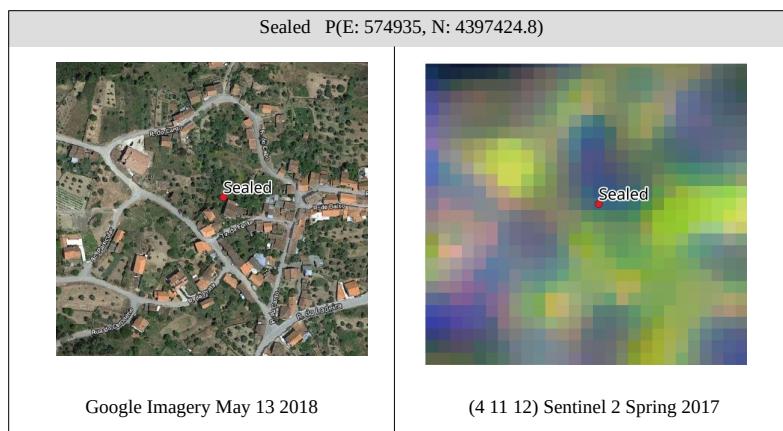


Figure 5.14: Example of low informative sample for sealed

5 Results

In the previous chapter, 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 led us to see the results over data that suppose to be representative for a specific date. Therefore, in collaboration with DGT, we had access to an external dataset (TB). This dataset corresponds to a visual interpretation of 557 samples for the image of 29 of July. In this sense, the second strategy was evaluated making particular emphasis on this date.

Figure 5.15 shows the implementation of the iterative learning procedure for the refinement of training data using the datasets with 6 and 9 classes in one image of July 29 of 2017. It shows in x-axis from left to right the percentage of data removed according with their level of informativeness and in the y-axis their correspond OA for validation (blue), TA(red) and TB(green). The box-plot are a demonstration of the stability of the results of classifications under different number of observations of the training data.

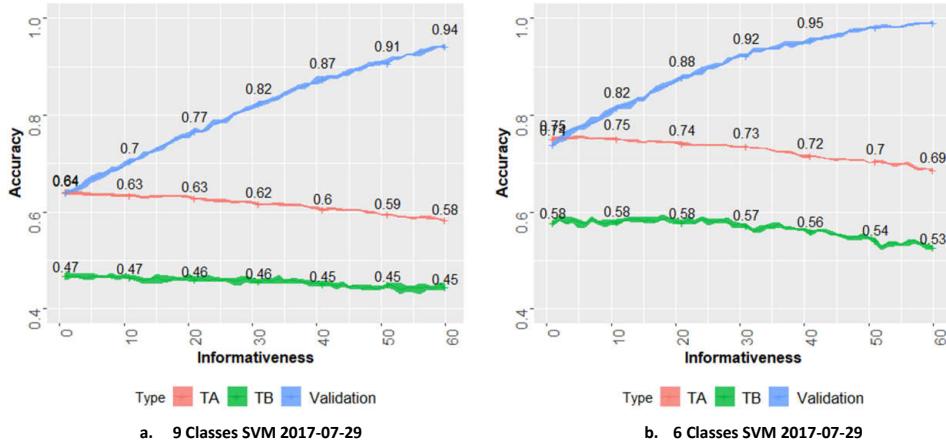


Figure 5.15: Cleaning preprocessing using batch learning, Image July 29 2017

While the results of the cross-validation (blue margin) increased after every iteration since the model was validated using less noisy data, the OA of both test datasets kept constant over a specific range of reduction of samples. After a reduction of 20% of samples, we could appreciate that the classifier started losing the ability to generalize and predict the test datasets accurately. Even though

the filter did not lead to a better performance of the classification for this date, the graphics show how the methodology can define a fraction of samples as not informative for the classification either because they may be redundant or they did not continue representing the class on the ground (This was also verified by visual inspection).

Further inspection of the methodology performance was done over the accuracies per class. Figure 5.16) shows the producer accuracy for the nine classes of dataset one using as reference the image of July 29 of 2017 and TB. The number over each line correspond to the accuracy per class at a specific percentage of data removed. For example, the class water that on average had a producer accuracy of 0.87 benefited in 3 scores with a reduction of 15% of samples. However, this did not happen with the class sealed, since a reduction of 8% of the lowest informative samples for this class ended up decreasing the predictive power of the model to classify it. In general, classes with vegetation are highly impacted in number for the refinement procedure; it should be noted, the predictive power for classes with vegetation is affected negatively by the methodology after removing 20% of the total information.

These results questioned if the implementation of the iterative learning procedure was not viable for any data, or the results corresponded to a particular scenario where the methodology was not necessary. To conduct this question, we decided to inject a specific level of noise in addition to the inherent error associated with the automatic selection of labels using an old map. With 30% of noise over each class, we performed the same above process. After new iterations, the strategy for refining data started benefiting the classification. According to Figure 5.17, the OA for the margins of TA and TB started increasing gradually as we approached to the 30% of data removed. The overall accuracy for TA increased on average from 0.7 to 0.73. Similarly, for TB the OA increased from 0.55 to 0.58. After the 30% of removed samples, our model started losing the ability to generalize, and therefore produce results with lower accuracies.

5 Results

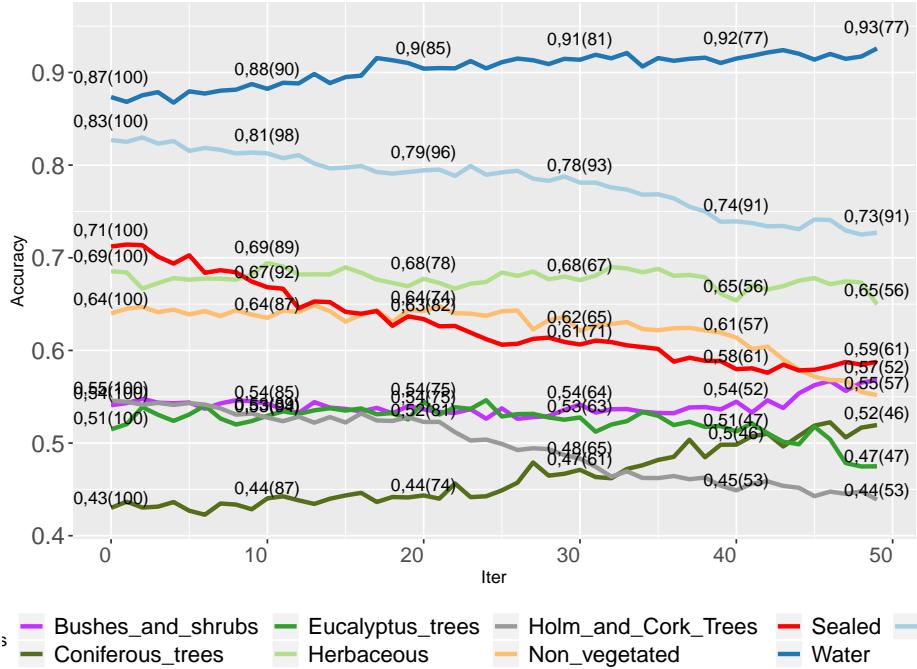


Figure 5.16: Cleaning preprocessing using batch learning per class, Image July 29
2017

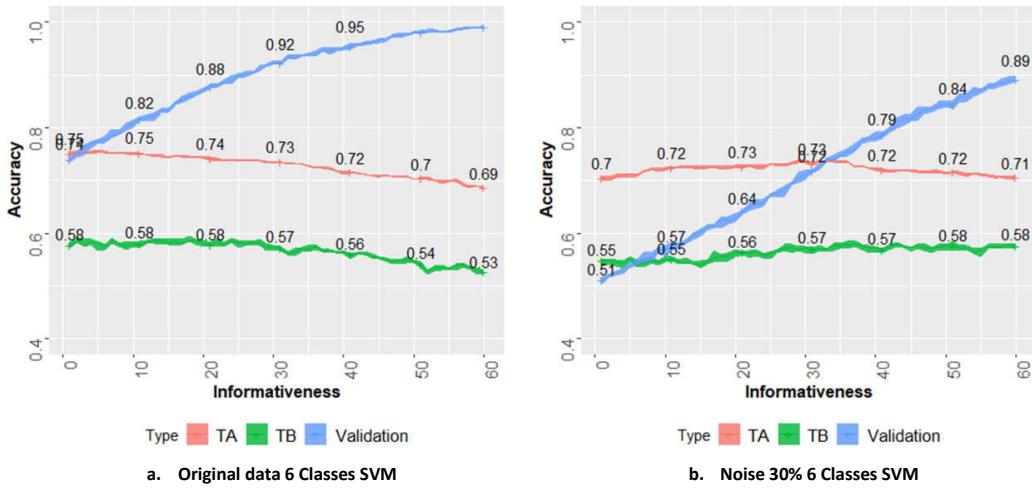


Figure 5.17: Performance of the proposed methodology over a training data with 30% of noise, scene-based composite July 29

5.5 Robustness of SVM and RF

The previous example was indicative of raising the question of how robust the classifier could be in the presence of different levels of noise in the training data.

Therefore, random level class noise influence is evaluated using the dataset with six and nine classes, and test TA. Figure 5.18 shows the average OA over several levels of noise. OA start being highly impacted after the level of noise overcome the 20% of the total data. After that threshold, the models start losing the ability to generalize and therefore predict correctly.

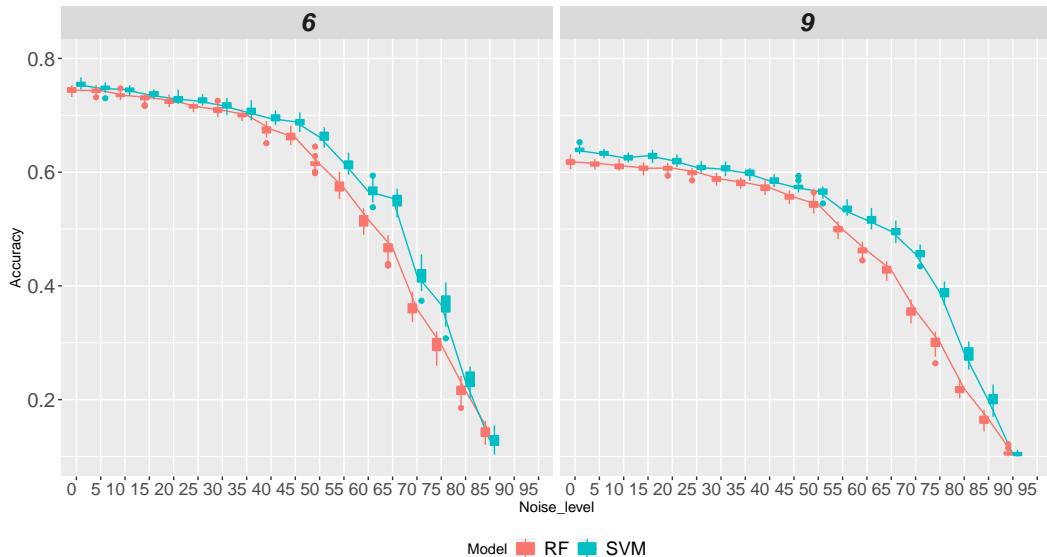


Figure 5.18: Performance random forest vs SVM using COS data under different levels of noise, scene-based composite July 29

5.5.1 Effect estimation of the hyper-parameters with Noise in the train data

While a sensitivity analysis of the parameters leads essentially to optimize the performance of an algorithm, their estimations may vary from one level of noise to another. To evaluate how stable or similar is the parametrization in the scenario of noise data, we recreated a sensitivity analysis over different levels of noise and two 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 5.19 shows that the optimal values found by two-folds cross-validation kept constant over 0%, 30% and 50% percentage of noise. Besides that, we implemented the same study over

5 Results

the dataset with six classes. The result is that the number of trees parameter equal or larger than 100 is a good estimation, no matter how noisy the data can be as a product of the natural changes of land cover over the year or types of misregistrations. However, regarding *mtry*, there was not a definite value to pick up. The default parameter of *mtry* equal to the square of the number of variables, in this case, *mtry* between 3 and 4, does not lead to better or worst results.

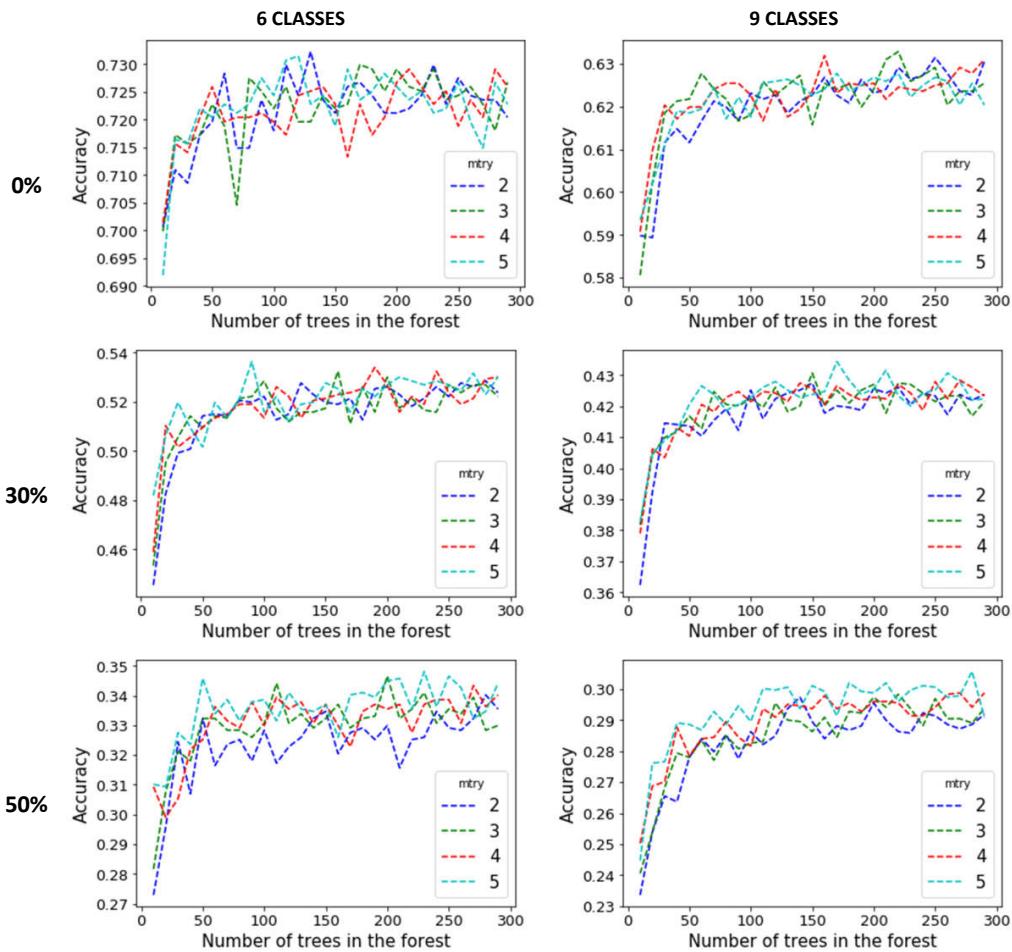


Figure 5.19: Sensitivity analysis parameters random forest

On the other hand, in the same study frame of random forest, we made a sensitive analysis of the parameters for SVM with a kernel RBF. Figure 5.20 shows that high values of **C** tested similar best overall accuracies under several levels of noise. A large value for **C** implied a lower chance for misclassifying the training

data; however, this could have resulted in bias in the modeling, and therefore possible the wrong prediction of the unseen test data. Therefore, we made sure of selecting a joint estimation of low values of 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.

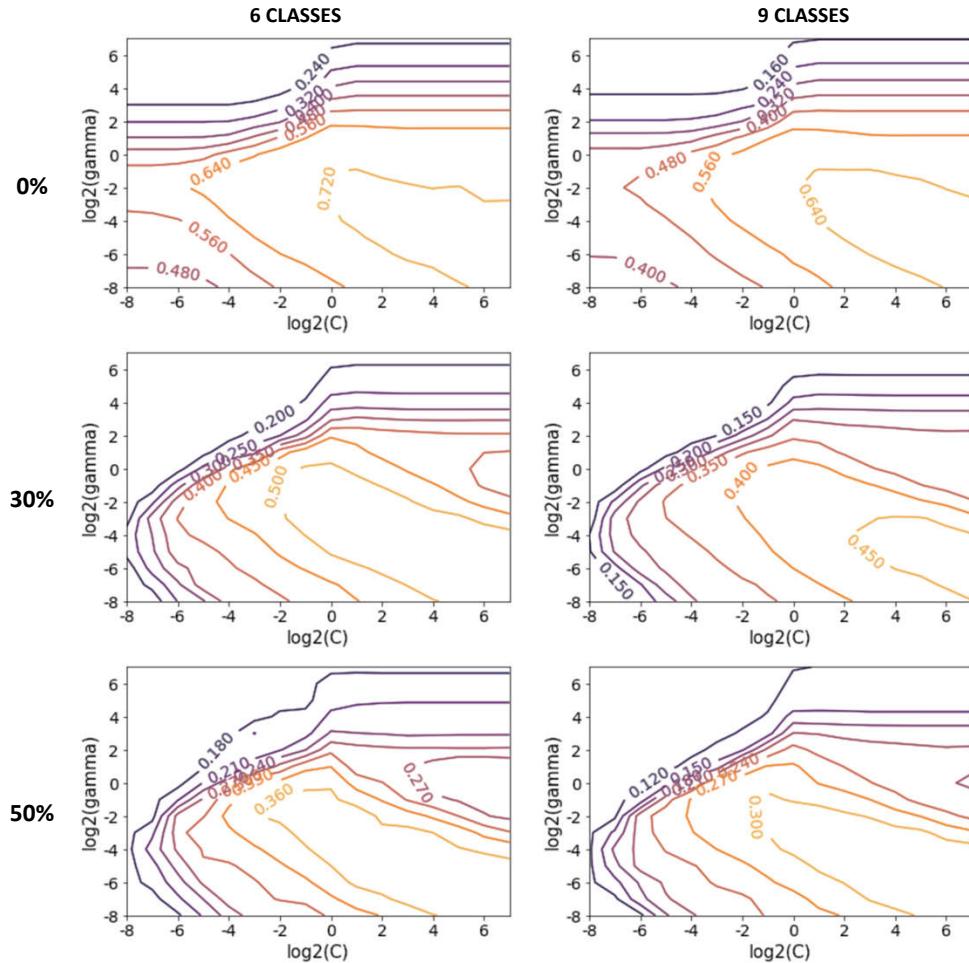


Figure 5.20: Sensitivity analysis parameters SVM with RBF kernel

6 Conclusions

This thesis aimed to answer one research question about the benefit of the integration of training samples extracted from old maps for automatic production of intra-annual land cover mapping. To do so, we proposed to explore the usability of training samples obtained from the LULC national map of Portugal (COS) to automatize the construction of training data, that usually is collected manually, in the supervised classification of singular and BAP composites of Sentinel 2. In this context, this thesis implemented different strategies to evaluate the use of the training data with version 2015 in the classification of images of 2017. Generally, the performance of image classification depended on multiple factors, where the nature of the training samples, the classification performance, the properties of the classifiers and quality of the composites were conducted in this thesis.

Firstly, after attempts for a refinement of the training data without positive impact in the increase of the OA of the classification of singular and seasonal composites of 2017, a simulation of the classification performance under different levels of noise in Cos map allowed us to infer that Cos map is a good source where to extract training data for classification. The still poor accuracy in the results and overestimation of herbaceous in the LULC maps may be related to other factors such as representativeness of the sampling per class, the number of classes to discriminate, the pixel-based approach and the spectral features that we used for the classification of images where the woody class was dominant.

Secondly, the application of RF and SVM was crucial in the image classification using training data extracted from old maps due to their desirable properties of robustness in the presence of mislabeled data. Moreover, the stability in the estimation of the parameters of both classifiers after different levels of noise also

6 Conclusions

led us to infer the balance of the parametrization for replication of intra-annual maps.

Finally, the free-cloud and phenological maximization of BAP composites become in a consistent and efficient input for the production of seasonal LULC mapping with 10 meters of spatial resolution. Its construction is computationally fast, simple and can be extended to different intra-annual frequencies besides the seasonal one, depending on the time frame of analysis.

6.1 Recommendations

We proposed the viability of implementing a batch learning procedure in the refinement of training data extracted from old maps for the classification of Sentinel 2 imagery. That is a classification schema that initially accounts with all samples at once and iteratively reduce the training set while keeping cluster boundaries. Future attempts can research the viability of implementing active learning procedures based on the contrary. That is online learning procedures. The possible benefit of this methodology can be related to a different interpretation of the measurement of information entropy. Since as more uncertain the classification of the pixel, more informative is the pixel to be selected by an oracle for later being corrected and updated. Therefore, its application can extend beyond a refinement of the training dataset; instead, it can work as a mechanism to detect areas where COS map needs to be updated.

It should be noted that the use of COS dataset is conditioned to the year 2017. Any other future use of the data for the classification of recent images will require a similar validation as the one conducted in this thesis.

Bibliographic References

- Bishop, C. M., 2006. Pattern recognition and machine learning (information science and statistics) springer-verlag new york. Inc. Secaucus, NJ, USA.
- Bontemps, S., Arino, O., Bicheron, P., Carsten, C., Leroy, M., Vancutsem, C., Defourny, P., et al., 2016. Operational service demonstration for global land-cover mapping: the globcover and globcorine experiences for 2005 and 2009. In: Remote sensing of land use and land cover. CRC press, pp. 262–283.
- Bousquet, O., Boucheron, S., Lugosi, G., 2004. Introduction to statistical learning theory. In: Advanced lectures on machine learning. Springer, pp. 169–207.
- 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.
- Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM computing surveys (CSUR) 41 (3), 15.
- Chen, Y., Gong, P., 2013. Clustering based on eigenspace transformation–cbest

Bibliographic References

- for efficient classification. ISPRS journal of photogrammetry and remote sensing 83, 64–80.
- 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.
- EEA, 2018. Technical specifications for implementation of a new land-monitoring concept based on eagle. <https://land.copernicus.eu/user-corner/technical-library/clcplus-draft-technical-specifications-v4>.
- 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.
- Hazan, E., et al., 2016. Introduction to online convex optimization. Foundations and Trends® in Optimization 2 (3-4), 157–325.
- 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.
- Khatami, R., Mountrakis, G., Stehman, S. V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment* 177, 89–100.
- Kordelas, G., Manakos, I., Aragonés, D., Díaz-Delgado, R., Bustamante, J., 2018. Fast and automatic data-driven thresholding for inundation mapping with sentinel-2 data. *Remote Sensing* 10 (6), 910.
- Mather, P., Tso, B., 2016. Classification methods for remotely sensed data. CRC press.
- Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., Klisch, A., 2019. Near real-time vegetation anomaly detection with modis ndvi: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sensing of Environment* 221, 508–521.
- 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.
- Pal, M., Mather, P., 2006. Some issues in the classification of dais hyperspectral data. *International Journal of Remote Sensing* 27 (14), 2895–2916.
- 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.

Bibliographic References

- 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.
- Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated training sample extraction for global land cover mapping. *Remote Sensing* 6 (5), 3965–3987.
- 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.
- 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., Muñoz-Marí, J., Camps-Valls, G., 2012. Remote sensing image segmentation by active queries. *Pattern Recognition* 45 (6), 2180–2192.
- 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.

Bibliographic References

- Vesanto, J., Alhoniemi, E., et al., 2000. Clustering of the self-organizing map. *IEEE Transactions on neural networks* 11 (3), 586–600.
- 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.
- Wang, Y.-C., Feng, C.-C., 2011. Patterns and trends in land-use land-cover change research explored using self-organizing map. *International Journal of Remote Sensing* 32 (13), 3765–3790.
- 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.
- Xu, L., Li, B., Yuan, Y., Gao, X., Zhang, T., Sun, Q., 2016. Detecting different types of directional land cover changes using modis ndvi time series dataset. *Remote Sensing* 8 (6), 495.

Annexes

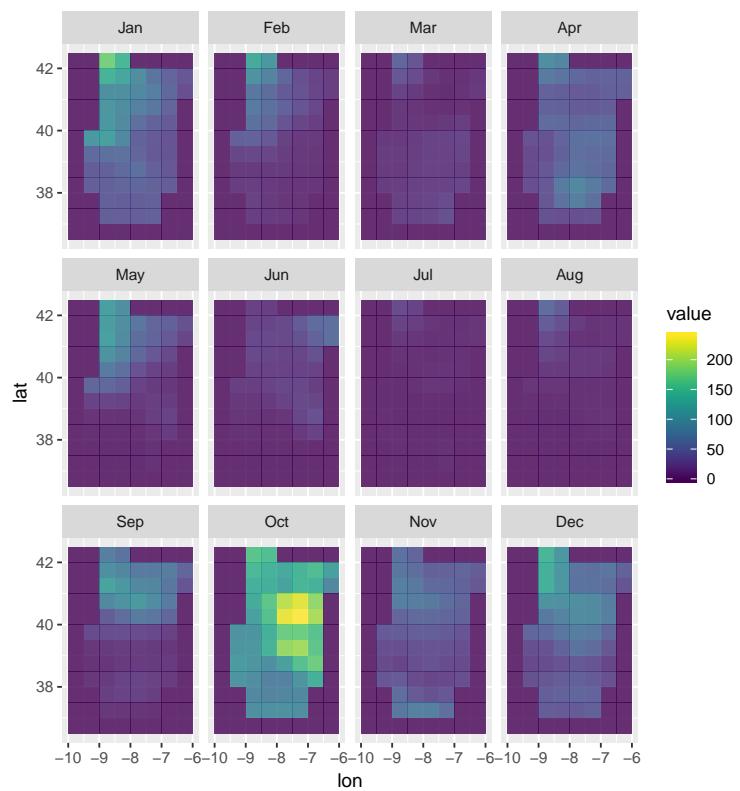


Figure 1: accumulative values of precipitation per month, 2015. Source data:
netCDF files NOAA

Annexes

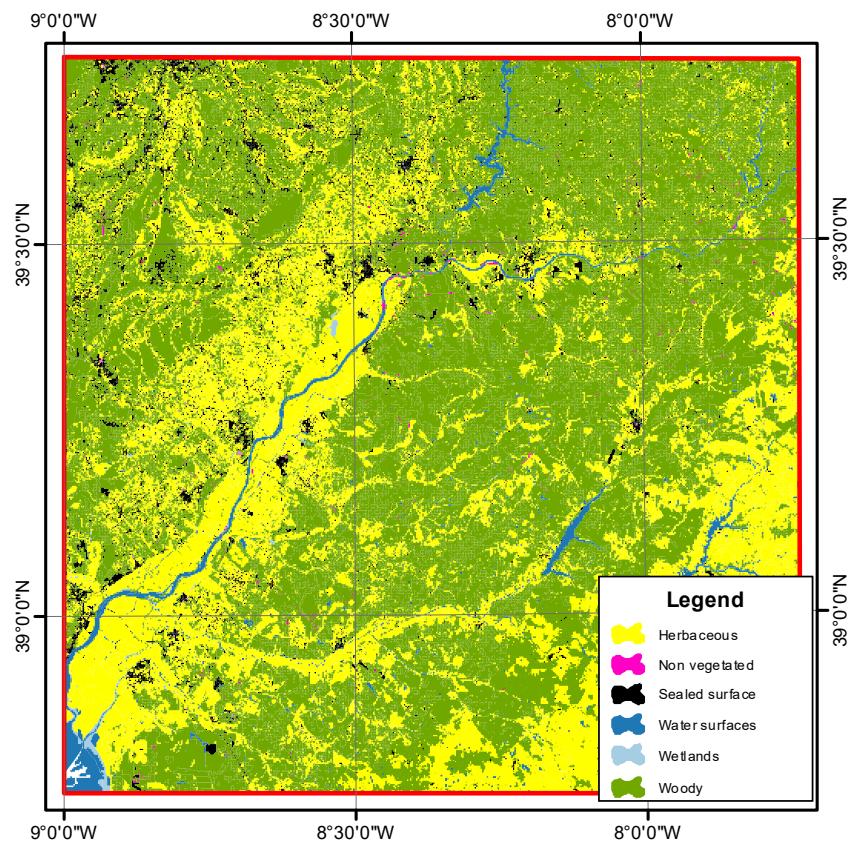


Figure 2: Intersection of the classes of COS map 2015 with the study area,
Nomenclature based on recent recommendation from EEA on Coperni-
cus land monitoring services

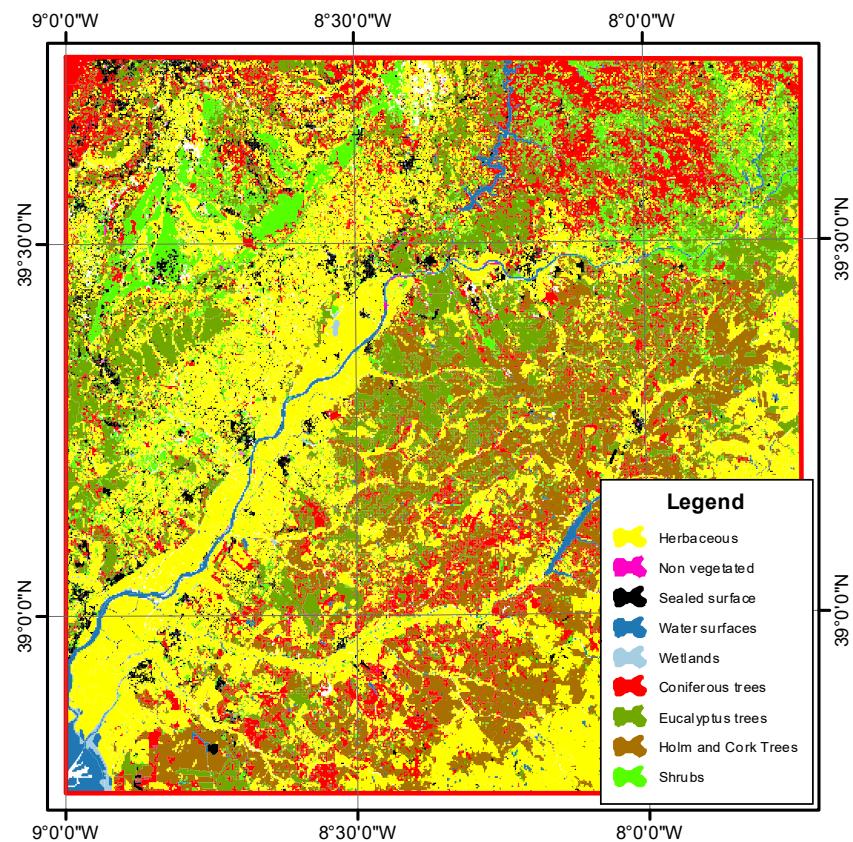


Figure 3: Intersection of nine classes of COS map 2015 with the study area,
Nomenclature based on recent recommendation from EEA on Copernicus land monitoring services

Masters Program in **Geospatial Technologies**

