

# Masters Program in **Geospatial Technologies**



## ***LULC Map Comparison***

### ***Comparison and validation of land use land cover maps derived from satellite imagery***

Fábio Miguel Domingues da Silva

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

# **LULC Map Comparison**

## **Comparison and validation of land use land cover maps derived from satellite imagery**

Dissertation supervised by:

**Professor Hugo Costa, PhD**

Universidade Nova de Lisboa, Information Management School

Lisboa, Portugal

Co-Supervised by:

**Professor Ana Cristina Costa, PhD**

Universidade Nova de Lisboa, Information Management School

Lisboa, Portugal

**Professor Filiberto Pla Bañón, PhD**

Universitat Jaume I, Institute of New Imaging Technologies

Castellón de la Plana, Spain

February of 2022

## **DECLARATION OF ORIGINALITY**

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or not published) are referenced. This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere.

*Lisbon, February 20th, 2022*

*Fábio Miguel Domingues da Silva*

## ACKNOWLEDGMENTS

First and foremost, I want to show my gratitude to my supervisor Dr. Hugo Costa for being so present and helpful, and for all the guidance that he was able to give me throughout these months. His extensive experience and knowledge in the field of Remote Sensing, as well as his advice were key for the development of this project.

I also want to formally thank Dr. Mário Caetano for guiding me through the definition of this dissertation's topic of research. To my co-supervisors Dr. Ana Cristina Costa and Dr. Filiberto Pla Bañón, I want to show my gratitude for your help, comments, and feedback.

I would like to show my deepest gratitude to my family for supporting my dreams every single day, and guiding me mentally and spiritually through this journey, as well as a very warm thank you to my friends for always being by my side.

Finally, I want to thank all the people that were a part of my journey academically, my colleagues and new friends made, the professors, the welcoming universities, thank you for the memories!

# **LULC Map Comparison**

## **Comparison and validation of land use land cover maps derived from satellite imagery**

### **ABSTRACT**

The technological evolution of remote sensing techniques has allowed for the ever-growing creation of land use land cover maps. Nowadays mapping entities from all around the globe are capable of producing their own products, be it for their region or country, or for a whole continent or even the whole world. However, this raises an issue regarding the comparison of these maps, in one sense direct comparison is difficult due to the lack of harmonization of the data, be it the nomenclature's classification, or technical specifications of the imagery. In another sense, the creation of global cover maps, or continent-wide maps, hinges the ability to accurately classify LULC appropriately due to the complexity of land cover, often leaving specific regions of the map with a less accurate classification, even though the overall one is good. Throughout this study, five maps from five different mapping entities will be compared and evaluated, these maps are COSsim, ELC10, ESA Worldcover, ESRILC and S2GLC. The study area is Continental Portugal, and the main objective is to understand how the international mapping entities' maps compare with the Portuguese map of COSsim, by observing nomenclature differences and accuracy scores. As well as understand what the impact in accuracy is, in European cover or World cover maps, when only analyzing them for the study area of Continental Portugal. The results obtained showed that most international maps proved to have a much smaller accuracy score for Continental Portugal, most of these even having a 20% to 30% drop in their overall accuracy. This research helps understand the necessity for the harmonization of nomenclatures, and at the same time investment necessary for national mapping entities to create their own more accurate maps.

## **KEYWORDS**

Accuracy Assessment

ArcGIS Pro

Continental Portugal

Land Use Land Cover

Nomenclature Harmonization

Remote Sensing

Sentinel

Validation

## ACRONYMS

**LULC** – Land Use Land Cover

**COS** – Carta de Uso e Ocupação do Solo (*Land Use and Land Cover Map*)

**COSSim** – Carta de Ocupação do Solo Simplificada (*Simplified Land Use and Land Cover Map*)

**ELC10** – European Land Cover 10 meters

**S2GLC** – Sentinel-2 Global Land Cover

**ESRILC** – ESRI Land Cover

**CLC** – CORINE Land Cover Map

**ESA** – European Space Agency

**DGT** – Direção Geral do Território

**OA** – Overall Accuracy

**PA** – Producer’s Accuracy

**UA** – User’s Accuracy

## INDEX OF THE TEXT

<b>ACKNOWLEDGMENTS</b> .....	<b>IV</b>
<b>ABSTRACT</b> .....	<b>V</b>
<b>KEYWORDS</b> .....	<b>VI</b>
<b>ACRONYMS</b> .....	<b>VII</b>
<b>INDEX OF THE TEXT</b> .....	<b>VIII</b>
<b>INDEX OF TABLES</b> .....	<b>X</b>
<b>INDEX OF FIGURES</b> .....	<b>XI</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 Theoretical Framework.....	1
1.2 Motivation.....	2
1.3 Research Objectives And Questions .....	3
1.4 Thesis Structure .....	4
<b>2. Background &amp; Related Work</b> .....	<b>5</b>
2.1 Map Comparison.....	5
2.2 Classification Comparison .....	5
2.3 Accuracy Comparison.....	7
<b>3. Materials &amp; Methodology</b> .....	<b>8</b>
3.1 Methodology Approach .....	8
3.2 Data.....	8
3.2.1 Map Selection .....	8
3.2.2 Maps.....	9
3.3 Data Visualization.....	10
3.3.1 Data Adjustments.....	12
3.3.2 Data Inconsistencies.....	13
3.4 Nomenclature Harmonization .....	14
3.4.1 Maps' Nomenclatures .....	14
3.4.2 Unified Nomenclature.....	17
3.4.3 Harmonized Data .....	18



3.4.3.1 Particular Cases .....	27
<b>4. Methods.....</b>	<b>30</b>
4.1 Area Comparison .....	30
4.2 Accuracy Assessment .....	32
4.2.1 Sampling .....	33
4.2.2 Photointerpretation Of Samples .....	35
4.2.3 Confusion Matrix And Results .....	37
4.2.4 Resampling Approach.....	40
4.2.4.1 Confusion Matrix And Results .....	41
4.3 Finding Hot Spots .....	44
<b>5. Discussions And Conclusions .....</b>	<b>50</b>
5.1 Results.....	50
5.2 Conclusions.....	51
5.3 Limitations And Future Work.....	52
<b>BIBIOGRAPHIC REFERENCES .....</b>	<b>54</b>

## INDEX OF TABLES

<b>Table 3.1</b> COSsim' Nomenclature.....	15
<b>Table 3.2</b> ELC10's Nomenclature.....	15
<b>Table 3.3</b> ESRILC's Nomenclature.....	16
<b>Table 3.4</b> Worldcover's Nomenclature.....	16
<b>Table 3.5</b> S2GLC's Nomenclature.....	16
<b>Table 3.6</b> Unified Nomenclature.....	17
<b>Table 3.7</b> Model Matrix.....	19
<b>Table 3.8</b> Agreement matrix matching ELC10 and the unified nomenclature.....	20
<b>Table 3.9</b> Agreement matrix matching S2GLC and the unified nomenclature.....	20
<b>Table 3.10</b> Agreement matrix matching Worldcover and the unified nomenclature.....	21
<b>Table 3.11</b> Agreement matrix matching COSsim and the unified nomenclature.....	21
<b>Table 3.12</b> Agreement matrix matching ESRILC and the unified nomenclature.....	22
<b>Table 3.13</b> ELC10's Harmonized Nomenclature.....	23
<b>Table 3.14</b> S2GLC's Harmonized Nomenclature.....	23
<b>Table 3.15</b> Worldcover's Harmonized Nomenclature.....	24
<b>Table 3.16</b> COSsim's Harmonized Nomenclature.....	24
<b>Table 3.17</b> ESRILC's Harmonized Nomenclature.....	24
<b>Table 4.1</b> Areas (in sq. km) of all maps by class.....	30
<b>Table 4.2</b> Original overall accuracies reported.....	32
<b>Table 4.3</b> COSsim confusion matrix.....	37
<b>Table 4.4</b> ELC10 confusion matrix.....	38
<b>Table 4.5</b> Worldcover confusion matrix.....	38
<b>Table 4.6</b> ESRILC confusion matrix.....	38
<b>Table 4.7</b> S2GLC confusion matrix.....	39
<b>Table 4.8</b> Recalculated accuracies after assessment for independent samples.....	29
<b>Table 4.9</b> COSsim confusion matrix recalculated.....	42
<b>Table 4.10</b> ELC10 confusion matrix recalculated.....	42
<b>Table 4.11</b> Worldcover confusion matrix recalculated.....	43
<b>Table 4.12</b> ESRILC confusion matrix recalculated.....	43
<b>Table 4.13</b> S2GLC confusion matrix recalculated.....	43
<b>Table 4.14</b> Recalculated accuracies after assessment for the resampling approach.....	44

## INDEX OF FIGURES

<b>Figure 3.1</b> COSsim Map.....	11
<b>Figure 3.2</b> ELC10 Map.....	11
<b>Figure 3.3</b> Worldcover Map.....	11
<b>Figure 3.4</b> ESRILC Map.....	11
<b>Figure 3.5</b> S2GLC Map.....	12
<b>Figure 3.6</b> Sample before correction (Dark blue = S2GLC; Light blue = ELC10)....	13
<b>Figure 3.7</b> Sample after correction (Dark blue = S2GLC; Light blue = ELC10).....	13
<b>Figure 3.8</b> Harmonized COSsim map.....	25
<b>Figure 3.9</b> Harmonized ELC10 map.....	25
<b>Figure 3.10</b> Harmonized Worldcover map.....	25
<b>Figure 3.11</b> Harmonized ESRILC map.....	25
<b>Figure 3.12</b> Harmonized S2GLC map.....	25
<b>Figure 3.13</b> Fruit tree areas in COS 2018.....	28
<b>Figure 3.14</b> Olive grove areas in COS 2018.....	29
<b>Figure 4.1</b> COSsim sample points.....	34
<b>Figure 4.2</b> ELC10 sample points.....	34
<b>Figure 4.3</b> Worldcover sample points.....	34
<b>Figure 4.4</b> ESRILC sample points.....	34
<b>Figure 4.5</b> S2GLC sample points.....	35
<b>Figure 4.6</b> Sample points' attribute table.....	35
<b>Figure 4.7</b> Ground truth accuracy assessment process.....	36
<b>Figure 4.8</b> Set of accuracy assessment points for all maps.....	41
<b>Figure 4.9</b> Misclassified points for COSsim accuracy assessment photointerpretation.....	45
<b>Figure 4.10</b> Misclassified points for ELC10 accuracy assessment photointerpretation.....	45
<b>Figure 4.11</b> Misclassified points for Worldcover accuracy assessment photointerpretation.....	45
<b>Figure 4.12</b> Misclassified points for ESRILC accuracy assessment photointerpretation.....	45
<b>Figure 4.13</b> Misclassified points for S2GLC accuracy assessment photointerpretation.....	46
<b>Figure 4.14</b> Hot and cold spots of misclassification of sample points in COSsim.....	47
<b>Figure 4.15</b> Hot and cold spots of misclassification of sample points in ELC10.....	47

**Figure 4.16** Hot and cold spots of misclassification of sample points in  
Worldcover.....48

**Figure 4.17** Hot and cold spots of misclassification of sample points in ESRILC....48

**Figure 4.18** Hot and cold spots of misclassification of sample points in S2GLC.....48

# 1. INTRODUCTION

## 1.1 Theoretical Framework

Throughout recent human history, and due to rapid industrialization of most of our world, mankind has seen an exponential growth in our knowledge about the planet's landmasses, oceans, rivers, and all other geographic features. A major way that industrialization could provide us with this further knowledge was through the invention, creation, and development of Satellites, this technology allows for Humanity to record and observe the Earth's land at almost any given time, it provides high quality imagery for various purposes, being it military purposes, or scientific or civil purposes. Techniques were and are still being developed for enhancing this technology, and it is one that is nowhere near the end of its expansion [1].

The existence of this ability to record and observe satellite images has opened a door for a type of maps that have proved to be crucial for GIS (Geographic Information Systems) and Remote Sensing, these are land cover land use maps (LULC Maps). Land cover land use maps are a collection of processed satellite images that are analyzed in a way that allows for a division of said image by classes. This division is called "Classification", it is a collection of classes that specify either Earth's, or a certain region's, land use, or land cover, or even both. It can range from classes such as the ones encompassing urban areas, or forests, or even specific types of agriculture and such. These maps have the ability to help people understand the landscape of a certain region, country, or even the whole planet [2]. Moreover, their comparison between a specific set of years, months, decades, allows for a comprehension of Earth's landscape changes throughout the specific time period [3]. Their uses range from helping simply understand the landscape and land uses, to deep diving onto a specialized topic such as monitorization of ecosystems, be it in agriculture, forest conversions, even surface water bodies' changes, and flows, to even visualizing changes in human urban expansion or destruction. These applications of LULC maps mentioned, make them a critical tool in decision making, especially for development and planning [4].

The way to create these maps is through the geo-rectification of satellite images of the study area in question, in order to obtain the base imagery to work on. Afterwards,

either a supervised or unsupervised classification algorithm will be chosen, where the number of classes is chosen. For a supervised classification, the algorithm will compute the map based on training samples, however, the unsupervised classification does not use training samples to compute the map [5]. The training samples allow for an improvement in the supervised map's veracity [7].

## **1.2 Motivation**

In today's world, humanity enjoys a set of various commodities that are associated with the ever-growing globalization of our planet, regarding remote sensing and land use/cover mapping this is no different. Every passing year it is possible to visualize and have access to a set of accurately developed land use/land cover images and maps produced by trustworthy companies and entities, such as Copernicus, ESRI, and others. These maps are often produced for a large area of the Earth, if not all the Earth's surface. However, each country and region, especially in the industrialized world, also develops their own set of maps and images, and for each of these products, different ways of classification are considered.

These different classifications exist due to different regions and areas of the world requiring certain types of classes that are dependent on that area's land cover. For example, if a map is focusing on an area in the Saharan desert, classes regarding specific and varied types of tree areas are not as necessary as they would have been, were it a map focusing on an area in the Indonesian rainforest [6]. This is also valid for the end goal of maps, for example depending on the main goal of the map, classification can be more specific or generalized for certain types of land cover and land uses. If the objective of a certain LULC map is to create a product that allows for agricultural mapping, its classification will focus more on different types of agricultural cover, and not so much on different types of tree cover, or urban cover.

These complex considerations to foresee make it so that mapping entities utilize different ground basis for the creation of their maps, especially in regard to the mapping's classifications [8].

The motivation of this research is drawn from this lack of uniformization of maps and their classifications, and thus a larger difficulty in comparing said maps between themselves. It becomes difficult, and in some cases impossible, to be able to put to

practice similar LULC maps' analysis and frameworks when they are not susceptible of being compared.

### **1.3 Research Objectives and Questions**

For this research, it is important to quantify and qualify the objectives and purposes that need answers. A way to do so is asking research questions that aim to be answered through the research. Due to the theoretical framework and motivation presented in the previous sub chapters, it is now possible to dive into what sort of end goals and objectives the present research has.

This research aims to learn how to compare land use land cover maps of Continental Portugal, measure and compare accuracies, and revise how different mapping entities differ in their maps when only observing the study area of Continental Portugal.

When understanding the issues relying on the comparison of the maps, it is crucial to firstly comprehend how the harmonization of the maps is key for their comparison. An important objective of this research is to utilize classification normalization techniques that make it possible to consider all maps onto the same plain of comparison [9]. However, this cannot take place before the selection of a set of maps that will consist of this work's data. These data consist of LULC maps of either the World, Europe, or Portugal. Another crucial objective is to understand how maps created for large areas perform when looking into only one specific area, this is, how accurate are these World maps when only observing Portugal.

This project focuses on a set of research questions such as:

- Do certain maps, created by International and European agencies lack in accuracy when compared to maps developed by Portuguese national mapping entities?
- How do these maps' accuracies compare between themselves and with actual ground-truth data, when looking only into Continental Portugal?

Moreover, in general, this research aims to aid in the comprehension of the normalization of LULC maps in order to facilitate comparison of such maps and in the end provide more coherent and comparable data. As it is understood that these data can be extremely powerful and helpful for human development and well-being, and it is necessary to have data that can be used interchangeably [10].

## **1.4 Thesis Structure**

The structure of this thesis is divided into five major chapters, being it that firstly, in chapter 1, an introduction to the work will be done with basis not only on the present research, but also a background introduction to the whole topic itself, it also contains an explanation of objectives, aims and research questions. Secondly, chapter 2 contains a literature review that indulges into the research and projects done before on this field, with emphasis on what has been done and what the present project can add that hasn't been explored before, a set of papers will be discussed, integrating their concepts and ideas not only between one another but also with this thesis. Chapter 3 and 4 discuss the actual work, this is the data that exists, how to process and use it to the present work's advantage, analyses to be done and their methods and results. Finally, chapter 5 will focus on conclusions based on the work done in the previous two chapters and it focuses on an in-depth discussion of all the work done and focus on understanding if all the objectives and research questions stated in the beginning were answered with this research, as well as the limitations and problems faced.



## **2. BACKGROUND & RELATED WORK**

The present chapter will indulge in a literature review that will be responsible of shedding light on the discussed topics throughout the analysis, more specifically understanding what work has been done on the topic of land use land cover map comparison, what are the existent comparison techniques discussed in certain papers, how they have been used and how that relates to the present work.

### **2.1 Map Comparison**

Ever since land use land cover maps started to be developed, the need for their comparison started as well. A comparison of two maps is more than observing visual cues that indicate differences. The approach to a comparison firstly depends on the products, or maps, that are being compared, they must represent the same study area. The comparison can be based on a time evolution, this is, a land change comparison, or a land cover comparison, which entails the situation where two or more maps compare the same area in the same time period but are different due to different map classification approaches [11,12]. Throughout this, a set of varied remote sensing comparison approaches were developed throughout many papers.

Often, in remote sensing, the comparison of maps is done based per-pixel and under the measures imposed by the classification agreement [13]. This means that, an approach that can be pursued is by addressing each map's classification, and if this is equal then the comparison can be made directly. If they represent the same, its differences will only differ based mostly on the classification technique, which implicates a difference, or not, in the number of pixels per class. Otherwise, in the most common case, when comparing different maps, it will happen that the classification differs between the two, or more, products. Meaning that there cannot exist a direct comparison based solely on the number of pixels per class, due to these classes representing different things [12].

### **2.2 Classification Comparison**

The case where the maps being compared are classified based on different nomenclatures, makes it that there does not exist a direct comparison, this being that if

for instance, one tries to compare a map that contains classes such as broadleaf tree cover, fruit tree cover, coniferous tree cover, with a map that encompasses all types of trees into one class of forest, issues will arise due to a lack of similarity and harmonization between these classes [14]. Even though, for both products, these classes represent the same general idea, which is in its simplest form just trees, for a pixel-per-pixel comparison, they will not be capable of doing so.

Yang [9] illustrates this issue in the standardization of classification systems and proposes that mapping entities work further in understanding how to “build crosswalks” that offer a solution for the terminological inconsistencies that are seen throughout land use land cover maps [9]. This is still a work in progress and still seen as only a future reality, seen that up to date, there are not any universal nomenclatures that are widely accepted by a significant amount of countries or mapping entities. The reasons as to why this is such a persistent issue rely on the fact that there are still issues in regard to the translation of legends and terms universally, as well as the individualistic approach that often is taken where maps are developed for niche purposes and specific areas, making them noninterchangeable with other maps. Another important factor that significantly troubles the standardization of classification systems is the complexity and variety of Earth’s land cover and its uses. Maps cannot always be susceptible of being compared to every other map developed and should not be always able to do that as well. The heterogeneity of the planet’s cover hinders the creation of a unique classification scheme for large and contrasting areas [15].

Fonte et. al [16], elucidates the procedure of creating a unified nomenclature that allows for the harmonization of classes from different classified maps [16]. This paper takes the classification schemes from CORINE Land Cover (CLC) 2018 and Carta de Uso e Ocupação do Solo (COS) 2018 and works on creating a unified nomenclature that is able to match both map’s specifications, making them comparable. In this case the author presents their set of considered classes that proved to be fitted of comprising the maps’ nomenclatures. The fitting of CLC and COS’s classes was done by matching each class code from these maps onto the class codes of the unified nomenclature.

In Fonte/Fritz et. al [17] the nomenclature harmonization process endured and followed the same methods exposed in Fonte et. al [16] however, in the case of this paper, the comparison of the maps is done between maps of the same entity. This is a time change comparison, i.e., it is a comparison of maps such as Climate Change Initiative Land Cover (CCI LC) between 1992 and 2015, CORINE Land Cover (CLC) between 2006 and 2012, Urban Atlas (UA) between 2006 and 2012 as well, and Carta de Ocupação de Solo (COS) between 2007 and 2010. Nonetheless, the author decided to also harmonize the nomenclature, so that the change results of each map is comparable to all the others. An important detail to note is that most of these maps represent areas larger than their study area of Portugal, meaning that for each map the author needs to assure if certain classes are present, for example permanent icy areas are not present in Portugal and must be then removed from the new harmonized nomenclature all together.

### **2.3 Accuracy Comparison**

A collective concern that impacts all maps regarding their author, mapping entity or classification, is the accuracy that they possess when compared to ground truth data, or in this case, the real world captured through satellite imagery. [18] The accuracy of such maps can be measured through an accuracy assessment analysis, which is done through photointerpretation of samples in a map. This provides the possibility of creating a confusion matrix that contains important values such as the overall accuracy of the map, represented in percentage. [19] The official article of the ELC10 map's publication, describes their methods for obtaining the overall accuracy level of 86% in their LULC map of Europe. [20] However, it also discusses accuracies on a region basis, by collecting samples from a pre-determined area of each climate region of Europe. One of these regions is the Atlantic Mediterranean climate which encompasses the landscape of all Continental Portugal, their independent accuracy assessment for this region showed that Portugal had a much smaller registered overall accuracy than the one for the whole continent, at around 67% [20].

This represents one of the issues the present research will focus on, which are the assessment of the accuracies of these maps for a specific study region, and how accurately this study area represents the map.

### **3. MATERIALS & METHODOLOGY**

#### **3.1 Methodology Approach**

After gathering the required theoretical concepts and ideas that will formulate this thesis and establishing a set of aims and objectives to be answered by this research, a methodology must be defined to take on the work in a practical sense. The methodology for this study is mapped to start in the selection of the products that will be the focus of analysis, this is in the selection of the land use land cover maps that will be compared with each other. Furthermore, these maps will be a target of a set of analysis that will be discussed throughout this chapter.

#### **3.2 Data**

This section of the work reviews the most important practical asset that will entitle this work with its core, the data. In a first approach, the map selection is made taken into consideration a few criteria and the restrictions that came with the available data, secondly, a description is presented, enumerating the maps and their specifications.

##### **3.2.1 Map Selection**

For the selection process it had to be considered that in the previous years there has been a large influx of land cover maps that denotate similar or the same regions on the same technical specifications, more specifically maps deriving from satellite imagery of Sentinel 1 and 2. In light of this offer of maps from the same satellites, a set of maps were chosen on the basis of a few restrictions, such as the pixel size that should see no difference in size, for a more direct approach with no manipulation of the original sources of data. Due to the extensive investment of imagery with a pixel of 10 meters, this was the chosen pixel size. Another restriction is the release date, seen that with the passing of time, maps will see a difference of recorded data from the satellite imagery due to inevitable human expansion and land use changes, the data must all represent equal periods of time, meaning that the years in which these maps were released, and consequently the years in which the satellite imagery were recorded must coincide, with a buffer period of just a couple of years of difference. The study area is Continental Portugal; however, it is important that there are maps coming from not only Portuguese or European mapping agencies, but also from World maps. This

inevitably relates to the proposed aim of understanding whether there is an apparent influence on accuracy between national maps and larger scale maps.

### 3.2.2 Maps

The data obtained is based on various open-source sets of data provided by the entities that have created the studied maps in question. This section enumerates the chosen maps, as well as a background description and their relevant specifications.

- COSsim 2018 (Carta de Ocupação do Solo Simplificada):

This is a national map of Portugal developed by the national mapping agency of Portugal, *DGT (Direção Geral do Território)*. On the contrary of the remaining maps, COSsim represents only Portuguese territory.

The COSsim map has as a year of reference 2018, it was made using Sentinel-2 satellite imagery, with a spatial resolution of 10 meters, and its nomenclature is classified by 13 classes. [21]

- ELC10 2018 (European Land Cover 10 meters):

The European Land Cover map, as the name suggests, is a land use land cover map of the continent of Europe. It produces complete up to date maps for a total of thirty-nine countries, it has as a year of reference 2018, it was made using both Sentinel-1 and Sentinel-2 satellite imagery for a combination of optical and radar imagery which proved to improve the accuracy levels measured, it has a spatial resolution of 10 meters, and its nomenclature is classified by 9 level-1 classes. [20, 22]

- ESA Worldcover 2020 (European Space Agency's Worldcover):

This world map was developed by the European Space Agency, and it is a free open-source map developed under the 5<sup>th</sup> Earth Observation Envelope Programme (EOEP-5). It has as a year of reference 2020, and was released as recently as October 2021, it is based on satellite imagery from both Sentinel-1 and Sentinel2, with a spatial resolution of 10 meters, and its nomenclature is classified by 11 level-1 classes. [23]

- ESRILC 2020 (ESRI Land Cover):

ESRI's land cover map is part of their Living Atlas, it maps the entire globe. This is a map of major global importance due to its goals of making the world's detailed cover available in high resolution, and being open source, as well as able to provide all nations, developed or developing, with tools to fight vital problems in humanity. Released in 2021, it has as a reference year 2020, its imagery is derived from Sentinel-2, and its spatial resolution is at 10 meters. This map's nomenclature is divided by 10 level-1 classes. [24]

- S2GLC 2017 (Sentinel-2 Global Land Cover):

The Sentinel-2 Global Land Cover map encompasses 39 countries in the European continent and was developed with the year 2017 as the reference year. Its imagery is derived from Sentinel-2, supported by Sentinel-1 SAR data. It has a spatial resolution of 10 meters, it contains a total of 13 level-1 classes in its nomenclature. [25]

### **3.3 Data Visualization**

The open-source data for all the land cover land use maps enumerated was downloaded, the visualization of this data was done through using the ArcGIS Pro software. The maps format is a TIFF raster. For each map, the tiles appropriate for my study area had to be selected, to avoid downloading extra amounts of unnecessary data. However, the maps still had to be adjusted to the area of study, which is Continental Portugal, and for this a simple Clip was applied onto a shapefile layer of Continental Portugal from ArcGIS Pro's living atlas library.

Moreover, after being able to visualize the raster files for each map clipped onto my study area, the number of level-1 classes was inserted for each map. The maps' symbology layer was edited based on the classification of each map, crediting the respective number of classes. Visual adjustments were made such as the attribution of colors and labels for each class. Figures 3.1 to 3.5 demonstrate all the maps with the adjustments and styles indicated.

Land Use Land Cover Map - COSSim, 2018

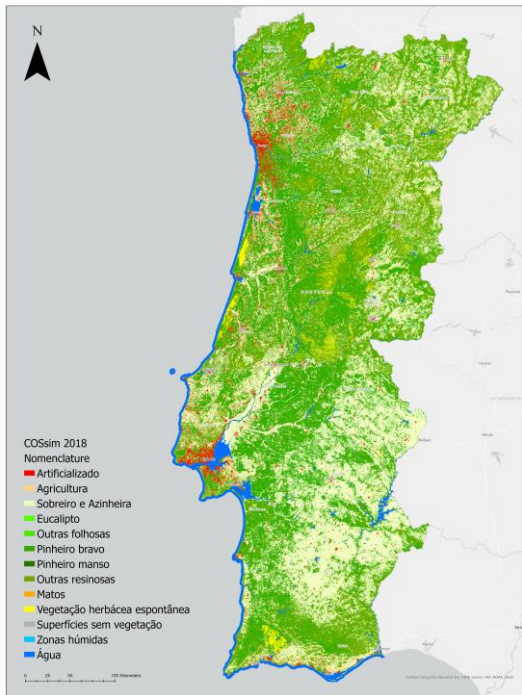


Figure 3.1 – COSSim Map

Land Use Land Cover Map - ELC10, 2018



Figure 3.2 – ELC10 Map

Land Use Land Cover Map - Worldcover, 2020

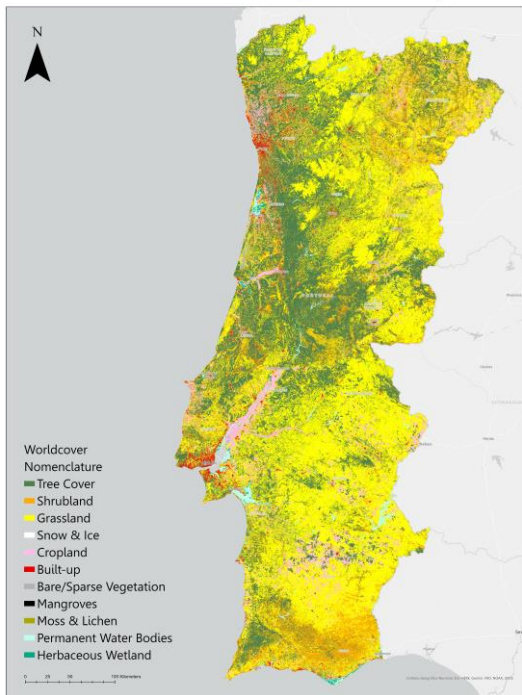


Figure 3.3 – Worldcover Map

Land Use Land Cover Map - ESRILC, 2020

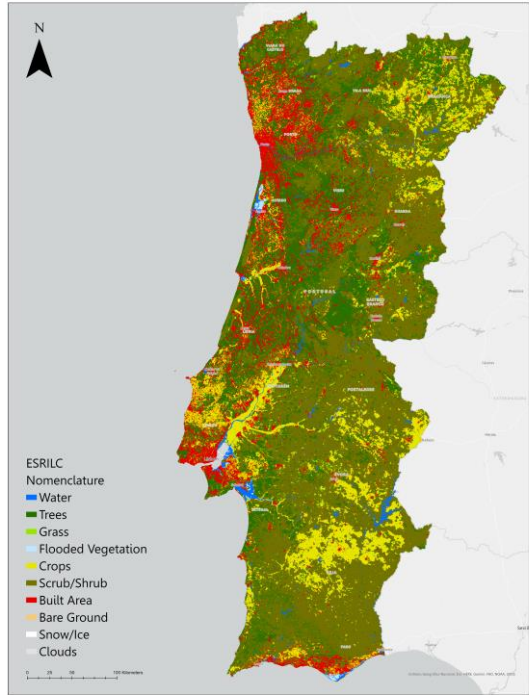


Figure 3.4 – ESRILC Map

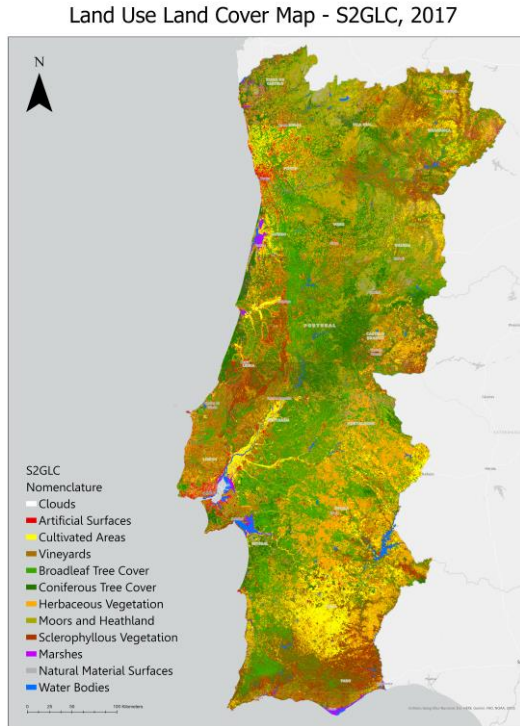


Figure 3.5 – S2GLC Map

### 3.3.1 Data Adjustments

Although now the data represents the study area of Continental Portugal, and has also been clipped to that extent, this does not eliminate the fact that pixel mismatch could and does still occur. To fix this issue it must be brought to attention that the data is not all on the same geographic coordinate system. The maps of Worldcover and ESRILC have the WGS (World Geodetic System) 1984 as their geographic coordinate system, while the maps ELC10 and S2GLC have the ETRS (European Terrestrial Reference System) 1989. The COSSim map data was provided from DGT with the WGS 1984 UTM 29N, coordinate system, based on Portugal’s location. To harmonize the maps and project all of them on the WGS 1984 UTM 29N coordinate system, the Project Raster tool from Data Management Tools on ArcGIS Pro must be used to re-project the two maps onto the desired coordinate system. The WGS 1984 UTM 29 coordinate system was chosen due to it fully encompassing the study area of Continental Portugal. In the following figures (Figure 3.6 and Figure 3.7) it is possible to observe the before and after, respectively, of the pixels mismatch being corrected. These figures represent a very closed up portion of the S2GLC and ESRI maps where the pixels are clearly visible, these two maps before correction were on different



geographic coordinate systems, and this small sample is able to show the general mismatch that was occurring due to that issue.



Figure 3.6 – Sample before correction (Dark blue = S2GLC; Light blue = ELC10)



Figure 3.7 – Sample after correction (Dark blue = S2GLC; Light blue = ELC10)

A set of maps were also using degrees as their main unit of measure and for this reason the cell size, used later on for the area analysis, was not defined in meters. Thus, also using the Project Raster tool from ArcGIS Pro within the same step as the one mentioned above, the cell size (X,Y) was set to the 10 meters by 10 meters measure.

Another detail to include also in the same step within the Project Raster tool, is to consider is that the pixels on the borders of the maps don't fully match, because when clipping the raster's extent to the shapefile of Continental Portugal, different pixel borders can be obtained based on different maps. The COSsim's map extent is used as the pixel extent ground truth and snapped all the remaining maps' rasters to that extent.

### 3.3.2 Data Inconsistencies

Certain inconsistencies were detected, especially regarding the product S2GLC, in which in its original nomenclature contains a class for clouds, however when looking into the study area, it is possible to realize that there are an abnormal amount of perfectly square and rectangular groups of pixels, classified as clouds, almost all in parts of industrialized areas in the Lisbon, Alentejo and Algarve regions, however none of these represent actual clouds, in fact they represent white roofing commonly

found in industrial buildings such as warehouses or department stores, as well as some type of greenhouses. This S2GLC class also considers white foam from waves ashore to be clouds.

### 3.4 Nomenclature Harmonization

In this subsection, the nomenclatures of the maps will be discussed. Land use land cover maps are a powerful tool due to their easiness to convey space occupancy and geography with simple and direct nomenclatures. However these are up to human judgment, meaning that most nomenclatures differ from each other. An appropriate practical case of this occurring could be that some might consider water areas and wetland areas similar enough to contain them into the same class, yet others might ponder otherwise and separate them into two different classes. In a case like the present one of comparison of land use land cover maps, this is a dire restraint for a fair comparison. For these exact reasons, a new unified nomenclature is created that can harmonize all others into one “fits all” mold. The necessary considerations and implications of this will be discussed further.

#### 3.4.1 Maps’ Nomenclatures

With the maps presented in section 3.3, it has been possible to observe that they are all classified according to a certain set of nomenclatures. The classes that form the nomenclature of the studied maps are as follows.

<i>COSsim</i>	
<b>Code</b>	<b>Class</b>
100	Artificial Land
200	Agriculture
311	Cork-oak and Evergreen-oak
312	Eucalyptus
321	Maritime pine
322	Stone pine
323	Other resinous
410	Shrubland
420	Spontaneous herbaceous vegetation

500	Non-vegetated surfaces
610	Humid areas
620	Water

Table 3.1. – COSsim’s Nomenclature

<i>ESRILC</i>	
<b>Code</b>	<b>Class</b>
1	Water
2	Trees
3	Grass
4	Flooded vegetation
5	Crops
6	Scrub/shrub
7	Built area
8	Bare ground
9	Snow/ice
10	Clouds

Table 3.2. – ESRILC’s Nomenclature

<i>Worldcover</i>	
<b>Code</b>	<b>Class</b>
10	Tree Cover
20	Shrubland
30	Grassland
40	Cropland
50	Built-up
60	Bare/sparse vegetation
70	Snow and ice
80	Permanent water bodies
90	Herbaceous wetland
100	Mangroves

110	Moss and lichen
-----	-----------------

Table 3.3. – Worldcover’s Nomenclature

<i>ELC10</i>	
<b>Code</b>	<b>Class</b>
0	Not Mapped
1	Artificial Land
2	Cropland
3	Woodland
4	Shrubland
5	Grassland
6	Bare Land
7	Water/permanent snow/ice
8	Wetland

Table 3.4. – ELC10’s Nomenclature

<i>S2GLC</i>	
<b>Code</b>	<b>Class</b>
0	Clouds
62	Artificial surfaces
73	Cultivated areas
75	Vineyards
82	Broadleaf tree
83	Coniferous tree
102	Herbaceous vegetation
103	Moors and Heathland
104	Sclerophyllous vegetation
105	Marshes
106	Peatbogs
121	Natural material surfaces
145	Permanent snow cover
162	Water bodies

Table 3.5. – S2GLC’s Nomenclature

These maps' nomenclatures although varied and unique in each way, all seem to follow the same land features, which is always to divide the Earth's surface by water, natural land of all kinds, natural land controlled by human intervention being it by crops or others, and land which has been erased by human intervention being these cities, industrial areas, and others.

### 3.4.2 Unified Nomenclature

Observing the various nomenclatures that the map products have, it was chosen to harmonize these terms, for easiness of comparison, and through that, the creation of a unified nomenclature that is able to describe and contain all the important land features to be included, while preserving simplicity for easiness and cohesiveness in the comparison work.

For this explicit reason, a general nomenclature, in other words a unified nomenclature has been established. All other nomenclatures will fit into this new mold for creating the possibility of comparison. This nomenclature is as follows.

<i>Unified Nomenclature</i>	
<b>Code</b>	<b>Class</b>
1	Artificial Land
2	Agricultural Land
3	Tree Cover
4	Shrubland & Grassland
5	Bare/Sparsely Vegetated
6	Water & Wetlands
7	Unidentified

Table 3.6. – Unified nomenclature

This nomenclature contains the represented set of 7 classes, for each of them a description is provided. The class of Artificial Land represents built-up non-natural areas, urbanized, paved, industrialized and such. Agricultural Land includes low rise summer and winter crop fields, and year-round crop fields of all types. Tree Cover represents all tree forms, coniferous, broadleaved, and fruit/other agriculturally related trees. To note that these types of trees are considered to be permanent cultivations, in some maps, the so-called permanent cultivations are integrated in their tree cover class, in other this is not the case, and they integrate it within the agricultural class.

Nonetheless, in the case of this study, these types of trees are considered to be a part of the “Tree Cover” class. A specific problematic case is studied in 3.4.3.1. Shrubland & Grassland refers to spontaneous low to medium rise vegetation, shrubs, can be dense or sparse. Bare/Sparsely vegetated areas are the ones that are periodically vegetated with very low density of it, or non-vegetated covered by natural surfaces such as rock, sand, burnt area or artificial bare areas. Water & Wetlands encompasses all types of water surfaces, including rivers, lakes, ocean, either man-made or artificial. Finally, the Unidentified class exists to count and identify error areas, or other non-identifiable, non-mapped areas.

### **3.4.3 Harmonized Data**

In the present section, it must be noted how it is possible to gather a set of diverse nomenclatures and mold them onto a unified one. The methodology of this work goes to the lengths of identifying firstly, from all the nomenclatures, which ones possess classes that are not present in my study area of Continental Portugal, due to the fact that most used data is classified according to a much larger area that contains land characterizations such as permanent snow, mangrove forests or others that are not present in any area of Portugal. In this situation the LULC maps S2GLC, ESA Worldcover and ESRILC all have at least one class that is entirely not present in the study area. The enumeration of these non-present classes, and its specifications is now presented.

- For S2GLC, the classes of Peatbogs and Permanent snow cover were not identified in the raster map.
- For ESA Worldcover, there is no registered presence of the Snow and ice class and the Mangroves class in the study area.
- For ESRILC, the class Snow/ice also does not make itself present for Continental Portugal.

With the realization and understanding of what classes can be directly disposed due to their non-existence, the question of agreement levels arise in the nomenclature harmonization. This is, there is a need to structure and create matrices that match the various maps’ nomenclatures to the unified one. A set of 5 agreement matrices were created, for each map.

Table 3.7 represents the model to be used for the creation of each matrix.

<b>LULC Map Unified Nomenclature</b>	Class 1	Class 2	Class 3	(...)
Artificial Land	X			
Agricultural Land		X		
Tree Cover			X	
Shrubland & Grassland				(...)
Bare/Sparsely Vegetated				
Water & Wetlands				

Table 3.7 – Model matrix

Tables 3.8 to 3.12 represent the matrices with the data of the maps' nomenclatures.

<b>Unified Nomenclature</b>	<b>ELC10</b>	Artificial Land	Cropland	Woodland	Shrubland	Grassland	Bare Land	Water/perm snow/ice	Wetland
Artificial Land		X							
Agricultural Land			X						
Tree Cover				X					
Shrubland & Grassland					X	X			
Bare/Sparsely Vegetated							X		
Water & Wetlands								X	X

Table 3.8 – Agreement matrix matching ELC10 and the unified nomenclature

<b>Unified Nomenclature</b>	<b>S2GLC</b>	Artificial Surfaces	Cultivated Area	Vineyards	Broadleaf tree	Coniferous tree	Herbaceous vegetation	Moors and heathland	Sclerophyllous vegetation	Natural material surfaces	Marshes	Water bodies
Artificial Land		X										
Agricultural Land			X	X								
Tree Cover					X	X						
Shrubland & Grassland							X	X	X			
Bare/Sparsely Vegetated										X		
Water & Wetlands											X	X

Table 3.9 – Agreement matrix matching S2GLC and the unified nomenclature



<b>Worldcover</b> <b>Unified</b> <b>Nomenclature</b>	Built-up	Cropland	Tree cover	Shrubland	Grassland	Bare/Sparse vegetation	Permanent water bodies	Herbaceous wetland
Artificial Land	X							
Agricultural Land		X						
Tree Cover			X					
Shrubland & Grassland				X	X			
Bare/Sparsely Vegetated						X		
Water & Wetlands							X	X

Table 3.10 – Agreement matrix matching Worldcover and the unified nomenclature

<b>COSsim</b> <b>Unified</b> <b>Nomenclature</b>	Artificial Land	Agriculture	Cork-oak and Evergreen-oak	Eucalyptus	Other leafy trees	Maritime pine	Stone pine	Other resinous	Shrubland	Spontaneous herbaceous vegetation	Non-vegetated surfaces	Humid areas	Water
Artificial Land	X												
Agricultural Land		X											
Tree Cover			X	X	X	X	X	X					
Shrubland & Grassland									X	X			
Bare/Sparsely Vegetated											X		
Water & Wetlands												X	X

Table 3.11 – Agreement matrix matching COSsim and the unified nomenclature

<b>ESRILC</b>	Built area	Crops	Trees	Grass	Scrub/ shrub	Bare ground	Flooded vegetation	Water
<b>Unified Nomenclature</b>								
Artificial Land	X							
Agricultural Land		X						
Tree Cover			X					
Shrubland & Grassland				X	X			
Bare/Sparsely Vegetated						X		
Water & Wetlands							X	X

Table 3.12 – Agreement matrix matching ESRILC and the unified nomenclature

The agreement matrices allow for a schematic visualization of how the classes from each map, although varied and different, can be grouped and fitted onto the unified nomenclature, this leaves the nomenclature to be harmonized like such.

<b>Code</b>	<b>Unified Nomenclature</b>	<b>Code</b>	<b>ELC10</b>
1	Artificial Land	1	Artificial Land
2	Agricultural Land	2	Cropland
3	Tree Cover	3	Woodland
4	Shrubland & Grassland	4, 5	Shrubland + Grassland
5	Bare/Sparsely Vegetated	6	Bare Land
6	Water & Wetlands	7, 8	Water/permanent snow/ice + Wetland
7	Unidentified	0	Not mapped

Table 3.13 – ELC10 harmonized nomenclature

<b>Code</b>	<b>Unified Nomenclature</b>	<b>Code</b>	<b>S2GLC</b>
1	Artificial Land	62	Artificial surfaces
2	Agricultural Land	73, 75	Cultivated areas + Vineyards
3	Tree Cover	82, 83	Coniferous tree + Broadleaf tree
4	Shrubland & Grassland	102, 103, 104	Herbaceous vegetation + Moors and heathland + Sclerophyllous vegetation
5	Bare/Sparsely Vegetated	121	Natural material surfaces
6	Water & Wetlands	105, 162	Marshes + Water bodies
7	Unidentified	0	Clouds

Table 3.14 – S2GLC harmonized nomenclature

<b>Code</b>	<b>Unified Nomenclature</b>	<b>Code</b>	<b>Worldcover</b>
1	Artificial Land	50	Built-up
2	Agricultural Land	40	Cropland
3	Tree Cover	10	Tree cover
4	Shrubland & Grassland	20, 30	Shrubland + Grassland
5	Bare/Sparsely	60	Bare/Sparse vegetation

	Vegetated		
6	Water & Wetlands	80, 90	Permanent water bodies + Herbaceous wetland

Table 3.15 – Worldcover harmonized nomenclature

Code	Unified Nomenclature	Code	COSsim
1	Artificial Land	100	Artificial Land
2	Agricultural Land	200	Agriculture
3	Tree Cover	311, 312, 313, 321, 322, 323	Cork-oak + Evergreen-oak + Eucalyptus + Other leafy trees + Maritime pine + Stone pine + Other resinous
4	Shrubland & Grassland	410, 420	Shrubland + Shrubs and herbaceous spontaneous vegetation
5	Bare/Sparsely Vegetated	500	Non-vegetated land
6	Water & Wetlands	610, 620	Humid areas + Water

Table 3.16 – COSsim harmonized nomenclature

Code	Unified Nomenclature	Code	ESRILC
1	Artificial Land	7	Built area
2	Agricultural Land	5	Crops
3	Tree Cover	2	Trees
4	Shrubland & Grassland	3, 6	Grass + Scrub/shrub
5	Bare/Sparsely Vegetated	8	Bare ground
6	Water & Wetlands	1, 4	Water + Flooded vegetation
7	Unidentified	10	Clouds

Table 3.17 – ESRILC harmonized nomenclature

Tables 3.13 to 3.17 show which classes must be integrated, for example, in S2GLC, the classes of Coniferous trees and Broadleaved trees together form the components

of the unified nomenclature class “Tree Cover”. In the cases of ELC10, S2GLC and ESRLC, the classes of “Not mapped” and “Clouds” respectively, were integrated onto the unidentified class due to the fact that, later, it was observed that these account for errors in the data, and these classes have misinterpreted some land use/cover. These phenomena will be further analyzed on sub section 3.4.3.1.

The agreement matrices allowed for a harmonization of all nomenclatures, at least theoretically, now this leaves the question of, how can it possible to gather this information and put it to the test in a practical manner? While interpreting the visualization of the data, the maps itself, there seems to exist a lack of visual cohesiveness due to their different classifications, however, adapting each nomenclature and harmonizing it onto the unified one, by grouping the values of each class using the symbology section of ArcGIS Pro, births a new set of maps that contain the same classifications of land use. These maps will be presented in figures 3.8 to 3.12.

Land Use Land Cover Map - Harmonized COSSim

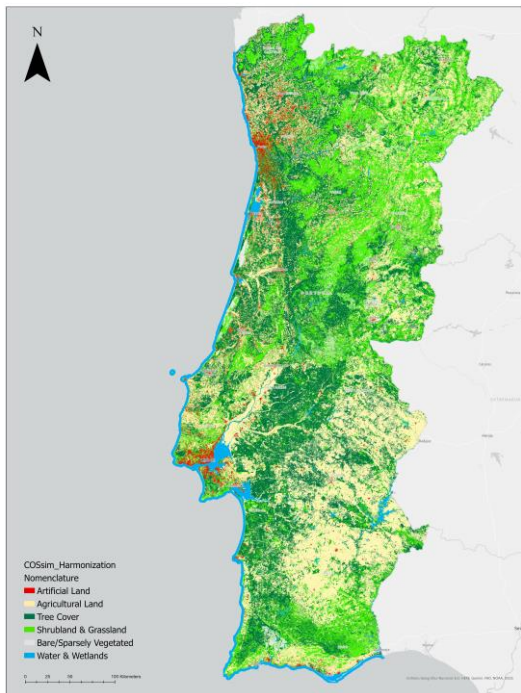


Figure 3.8 – Harmonized COSSim map

Land Use Land Cover Map - Harmonized ELC10

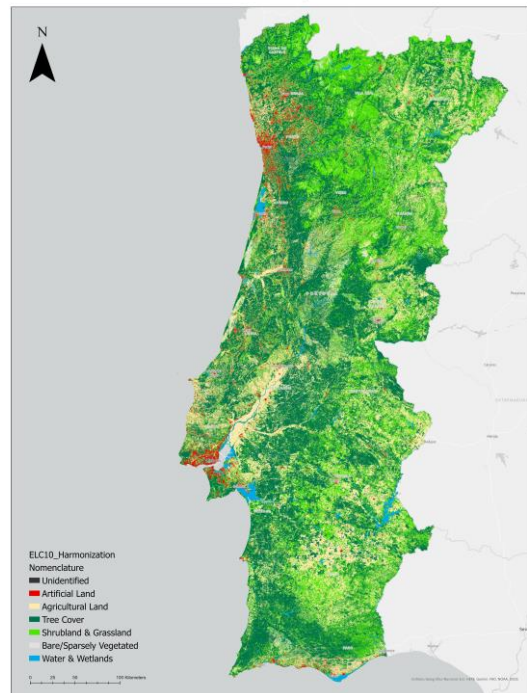


Figure 3.9 – Harmonized ELC10 map

Land Use Land Cover Map - Harmonized Worldcover

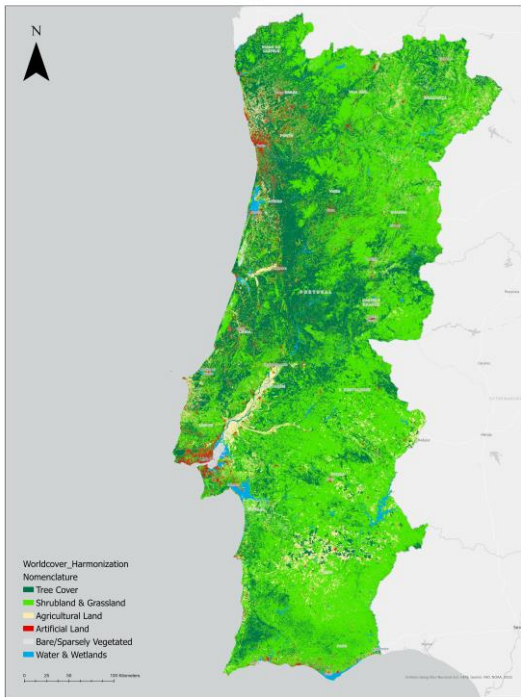


Figure 3.10 – Harmonized Worldcover map

Land Use Land Cover Map - Harmonized ESRI LC

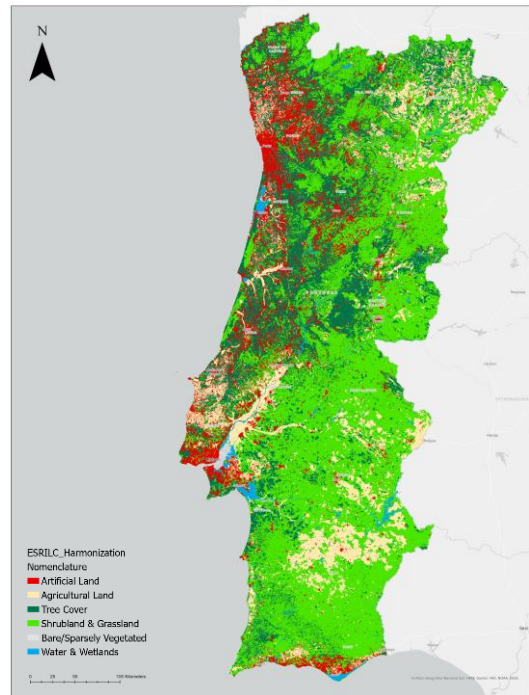


Figure 3.11 – Harmonized ESRI LC map

Land Use Land Cover Map - Harmonized S2GLC

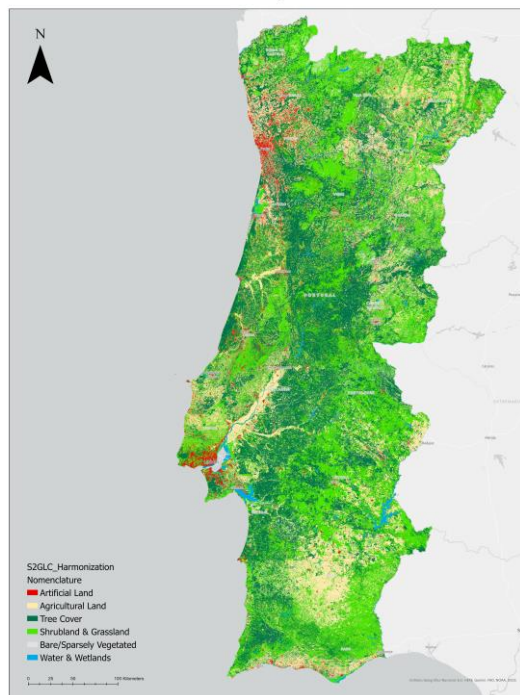


Figure 3.12 – Harmonized S2GLC map

Observing the maps with the harmonized nomenclatures it is possible to note that certain patterns are always present, like the major metropolitan areas of Lisbon or Porto dotting themselves and the surrounding areas with bright reds from artificial land use, or the central green areas of the Central Mountain Range (*Cordilheira Central*). But some areas are also strikingly different between the maps. This is due to a large pool of reasons, may it be due to the different classification algorithms used, average annual date of the processed satellite images, or different types and number of samples, and many others.

### **3.4.3.1 Particular Cases**

Although it is possible to directly connect most classes onto certain types of land use, there are always certain exceptions and differences in what some might consider part of “Tree cover” for example. Proof to this lies directly in the COSsim map, here it is crucial to note that in COSsim’s nomenclature, all fruit trees and trees used for agricultural purposes Orchards (*Pomares*) are included in their agricultural class. On the contrary in the other nomenclatures as well as the unified one, all types of trees are included into the classes of tree cover. This creates an accuracy issue in where the extent of COSsim’s agricultural class cannot be considered to be 100% contained within the unified nomenclature’s agricultural land class. Due to the fact that other maps consider fruit trees to be in some way a sub class of a Forest/Tree cover class, the unified nomenclature opted to include these types of trees into its Tree cover class as well. This causes a certain inconsistency in the re-classification of the COSsim map for the areas where fruit trees are most located. To observe which areas are most affected by this, Figure 3.13 dots all areas where the “*Pomares*” or fruit trees sub-class exists in the COS 2018 map.

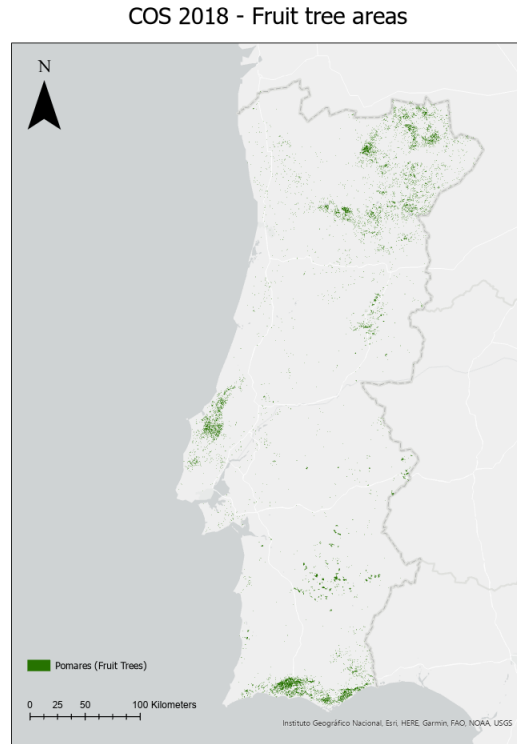


Figure 3.13 – Fruit tree areas in COS 2018

Figure 3.13 dots the fruit tree areas which are a target of misclassification in the re-classification work, and as mentioned, the data comes from COS 2018 (Land use land cover map of Portugal), and not the COSsim 2018 (Simplified Land use land cover map of Portugal) that is being used for the present work. The COS 2018 map is created through photointerpretation of orthophoto maps and auxiliary information with national cartographic bases. This map is being used for this due to COSsim not having a sub class that separates Agricultural land onto its own sub classes such as fruit trees and others, but despite that its formula classifies fruit trees to be a part of agricultural land. On the other hand, COS does have this separation, and has a sub-class from the main Agricultural Land class that considers fruit trees, this is the “*Pomares*” class of level-4. For these exact reasons, the fruit tree areas were taken from COS and not COSsim.

The regions most affected by this misclassification issue are mostly the southern Algarve, the Oeste regions, and the Douro valleys and Trás-dos-Montes region in northeastern Portugal. Some dots also surge throughout Alentejo and the Centro region.



In Figure 3.14, the same issue occurs, in which olive groves are classified as agricultural land, however, in the unified nomenclature these are encompassed within the “Tree Cover” class.

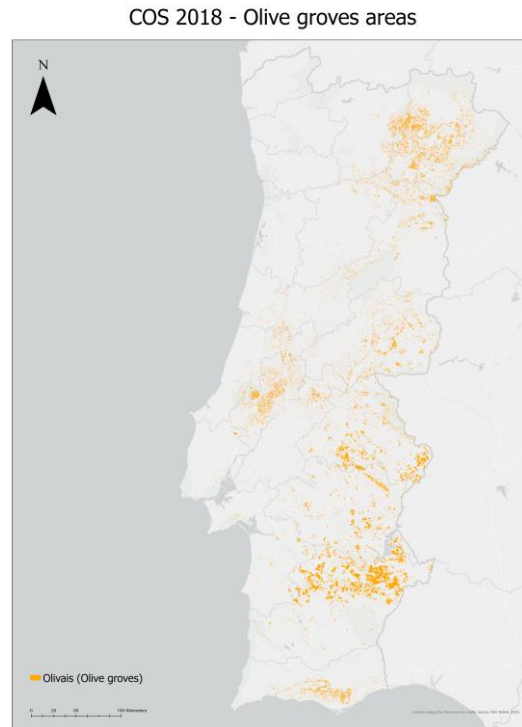


Figure 3.14 – Olive groves areas in COS 2018

As observed, these areas most affected by this misclassification are located in the interior of the country, most noticeably in Central and North Alentejo, in the Ribatejo region, as well as the northwestern Trás-dos-Montes region.

Vineyards are also seen as another issue, due to the lack of harmonization in its classification, for example, in the ESA Worldcover map, this class is seen as part of the Shrubland category, in the case of the S2GLC this is seen as agricultural land. For the purpose of this study vineyards were seen as agricultural land, however, inconsistencies in this classification may arise due to the lack of harmonization.

## 4. METHODS

In this chapter, results deriving from the various analyses done are to be presented, the analyses mentioned are outlined in sections 3.3 and 3.4, from the previous chapter, chapter 3. Results are crucial to the understanding and comparison of the LULC maps, the following sections contains different type of results, regarding areas, errors, accuracies, and validations.

### 4.1 Area Comparison

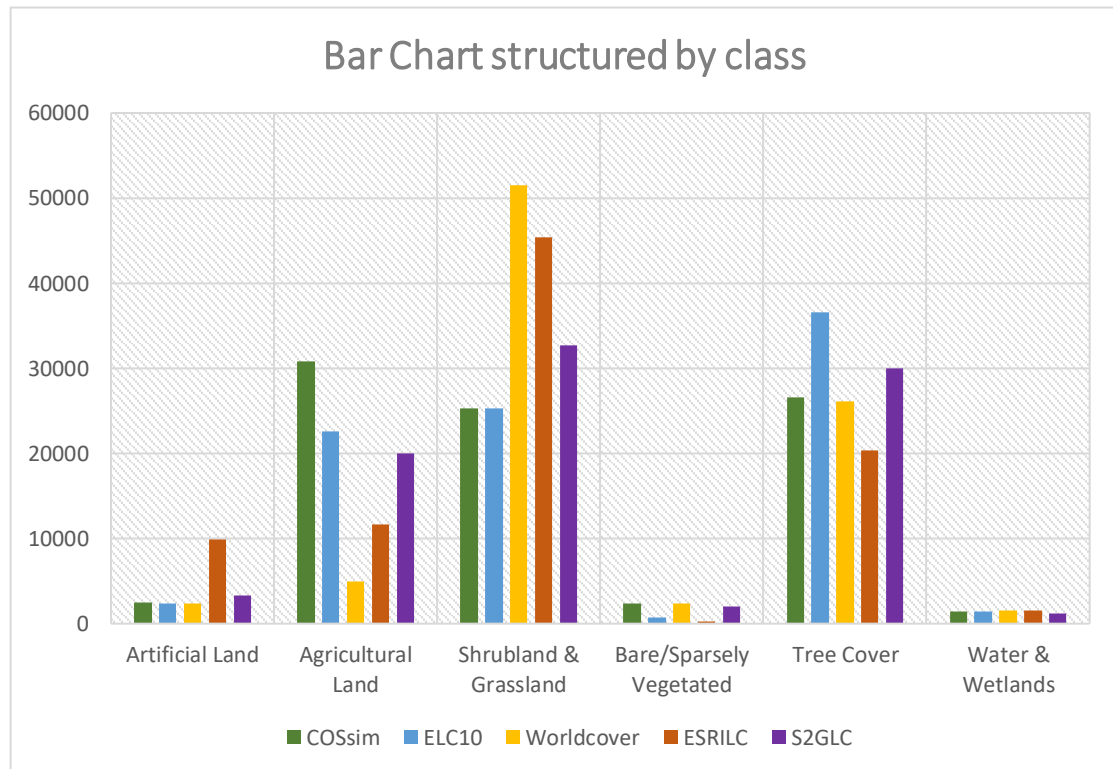
With the harmonization of the data accomplished, now it becomes important to indulge into techniques of comparison. One way of executing this is looking into the areas of each map per class, this is, now that all maps are classifying land uses regarding the same criteria, it is possible to observe the area that each class occupies and its weight in the overall map.

For this, in ArcGIS Pro, after the corrections of all the data onto the same 10x10 meters cell size, the count of pixels from each map was multiplied by the cell size values and then divided by 1000000 to obtain the number of square kilometers for each class (Table 4.1). Another step important for validation of these maps is assuring the total value of square kilometers per map matches the actual true size of Continental Portugal, which lies around 89 thousand square kilometers. After verification all maps do lie around this number as well, the only exception being to the COSsim map, and this is justified by the fact that this product considers all ocean areas within 2 kilometers of the coast, however by clipping the map onto the Portuguese border shapefile this is corrected.

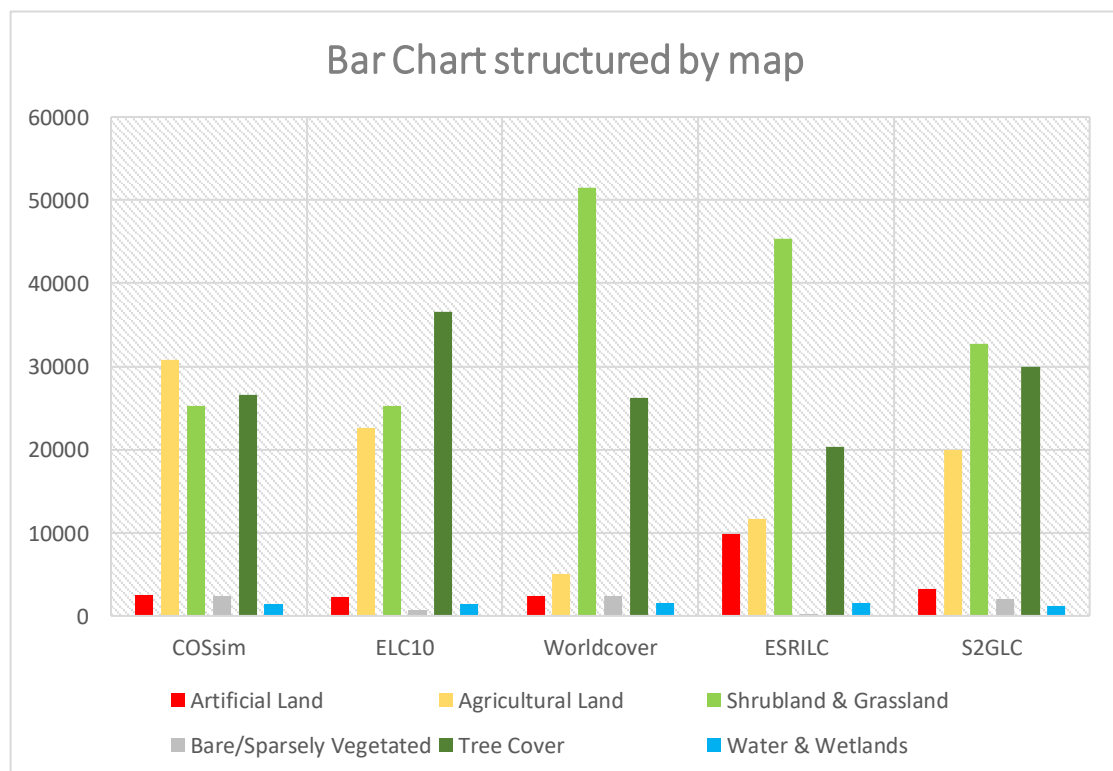
<b>Classes</b>	<b>COSsim</b>	<b>ELC10</b>	<b>Worldcover</b>	<b>ESRILC</b>	<b>S2GLC</b>
Artificial Land	2 510	2 338	2 425	9 902	3 271
Agricultural Land	30 786	22 616	5 008	11 705	19 954
Shrubland & Grassland	25 289	25 300	51 498	45 363	32 698
Bare/Sparsely Vegetated	2 419	744	2 387	199	2 019
Tree Cover	26 594	36 607	26 165	20 325	29 969
Water & Wetlands	1 386	1 407	1 529	1 519	1 136

Table 4.1 – Areas (in sq. km) of all maps by class

The values presented in Table 4.1, all in square kilometers, demonstrate that most maps follow similar values for their areas in certain classes.



Bar Chart 4.1 – Area (in sq. km) by class



Bar Chart 4.2 – Area (in sq. km) by map

The classes Bare/Sparsely Vegetated and Water & Wetlands always have the smallest areas, and Shrubland & Grassland has mostly the largest area in Portugal. Granted that most values follow the same patterns, however there are a set of exceptions that do not follow the general tendency. For instance, the fact that the ESRILC map has an extremely larger area for Artificial Land comparing to all other maps. When observing the map, it is very immediate and easy to realize that the highlighted artificial areas are much larger than all other maps, this will be further discussed in section 4.2 when looking into ESRILC’s accuracy analysis. Another particular case where a value does not follow any identical pattern to others is in the case of the Worldcover map’s class of Agricultural Land, since it identifies this area to be much smaller than in any other map.

## 4.2 Accuracy Assessment

A major question remains throughout the concretization of this project, this question has been highlighted multiple times in this research and aims to be answered with the appropriate results and conclusions. The question itself is regarding the accuracies of the maps studied. The maps’ original overall accuracy reported by the producing entities for the total study areas of values are presented in Table 4.2.

<b>Overall Accuracy</b>	<b>COSSim</b>	<b>ELC10</b>	<b>Worldcover</b>	<b>ESRILC</b>	<b>S2GLC</b>
Reported OA	81%	86%	75%	86%	86%

Table 4.2 – Original overall accuracies reported

Seen that these maps are all developed for an area much larger than Continental Portugal, with the exception of COSsim, their registered accuracies, mainly their overall accuracy is as well in regard to that large study area. This poses the question of, are these overall accuracy values the same when looking only into this project’s study area of Continental Portugal?

This will be answered by using probabilistic sampling for validation and computing the confusion matrix with the accuracy values.

### **4.2.1 Sampling**

For the verification and validation of the maps, an analysis must be performed where each LULC Map will be given 500 random points using a Stratified Random sampling strategy in the “Create Accuracy Assessment Points” tool in ArcGIS Pro. The Stratified Random sampling strategy distributes the amount of sample points proportionally to the size of each class, this means that out of each 500 points created, the largest number of them will fall onto the largest class, and inversely, the smallest class will contain the smallest number of sampling points.

Another sampling strategy would be fully random points on the image or Equalized Stratified Random which would give each class of the map the same number of points.

For each map, an independent sample was created, thus, each contains an attribute table with a nomenclature code for each 500 points. These points are randomly placed around the map according to the “Stratified Random” sampling strategy mentioned before, and each point will fall onto a certain pixel, each pixel containing a nomenclature code. The proposed task is to observe each point’s classification code and note in a new table column next to it if it matches with ground truth data from satellite imagery or not. For example, if a certain pixel is characterized with code number 1 meaning it is an artificial surface, and later in the verification, it is noted that that certain pixel actually belongs to a water surface, then in the new column it should be inserted the code for the water class which would be 6. On the contrary, if it is noted that its correct and it is actually an artificial surface then it should be inserted in the new column the same code number.

There are a total of 25 hundred points in a total of 5 maps (Figures 4.1 to 4.5).

Accuracy Assessment Points for COSsim

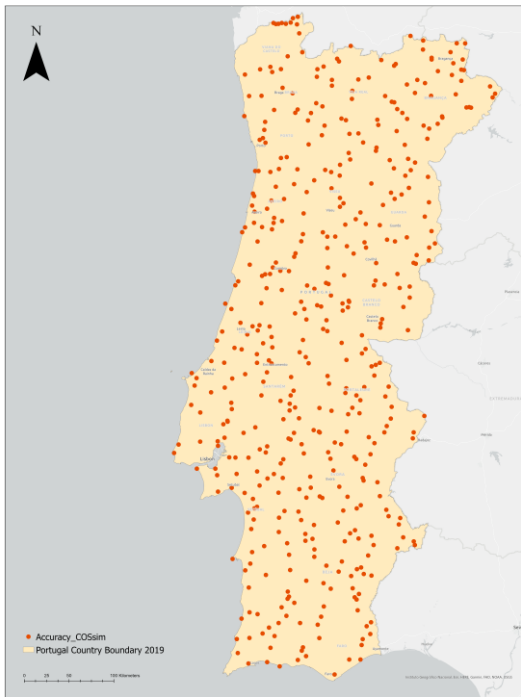


Figure 4.1 – COSsim sample points

Accuracy Assessment Points for ELC10

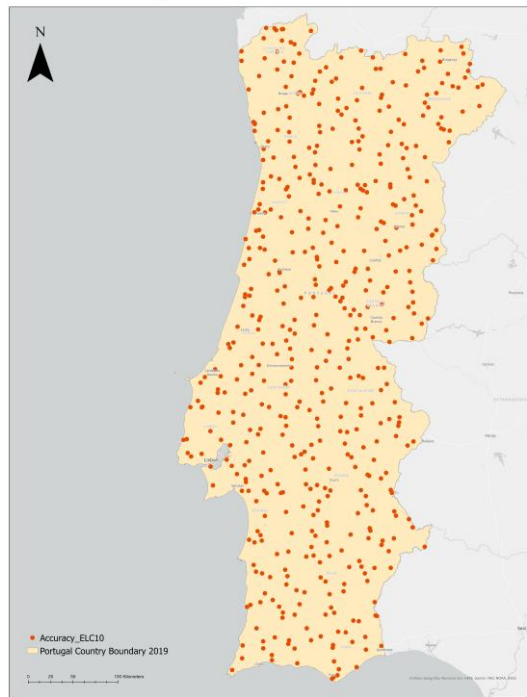


Figure 4.2 – ELC10 sample points

Accuracy Assessment Points for Worldcover

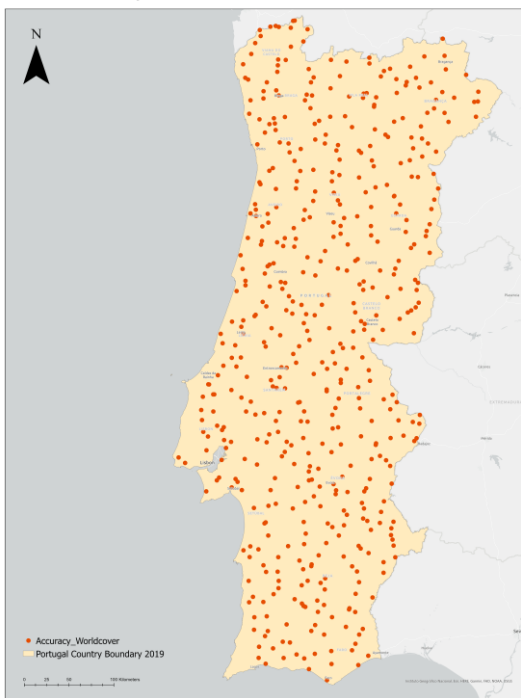


Figure 4.3 – Worldcover sample points

Accuracy Assessment Points for ESRILC

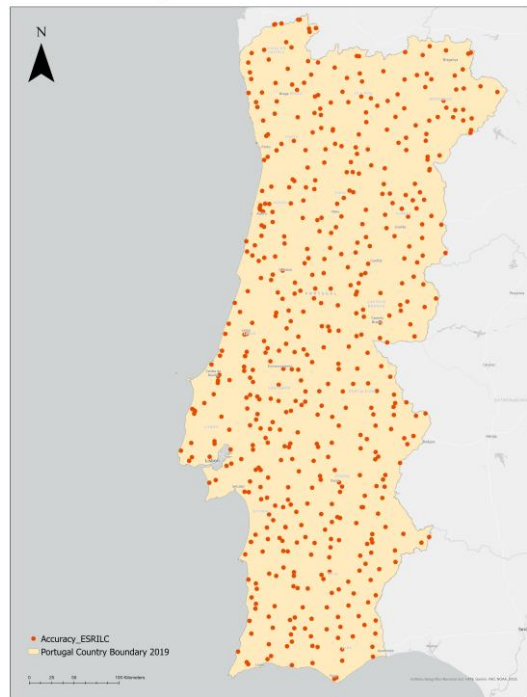


Figure 4.4 – ESRILC sample points

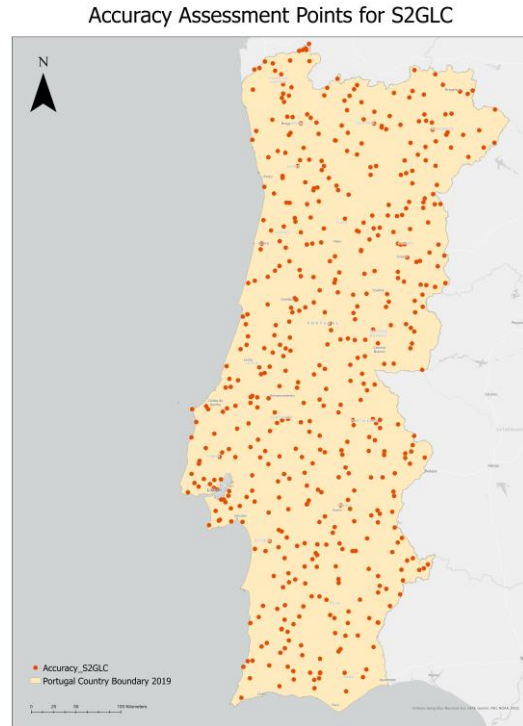


Figure 4.5 – S2GLC sample points

## 4.2.2 Photointerpretation of Samples

The process of photointerpretation of the samples, as mentioned in 4.2.1, starts when a new map layer is created containing the sampling points. Firstly, it is crucial to open the attribute table of each map, and understand which columns are identifying which class each point represents (Figure 4.6).

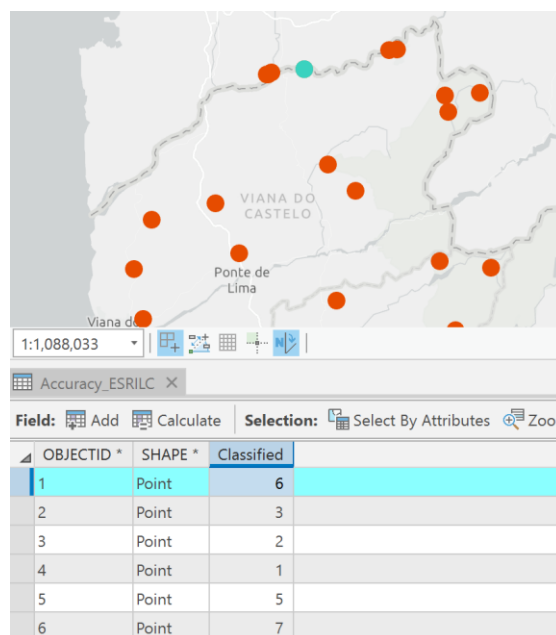


Figure 4.6 – Sample points' attribute table

Figure 4.6 is what is seen when opening the attribute table of, in this case, the ESRILC map, for each of all the 500 points, numbered in “OBJECTID” 1 through 500, each has a corresponding field named “Classified” in which a number is given, depending on the classification. In figure 4.6, point 1 is classified as class 6, which represents Water & Wetlands.

The next step is to create a new column named “GrndTruth” where the accurate classification will be inserted comparing to the ground truth, this is, for each point it is observed, comparing with satellite imagery, if the map’s classification is accurate to reality (Figure 4.7).

It is crucial to mention that, although for this case the “Classified” column is visible, it is just for a matter of visual presentation. In the actual assessment of these samples, the “Classified” column is hidden to not influence the person assessing the ground truth data.

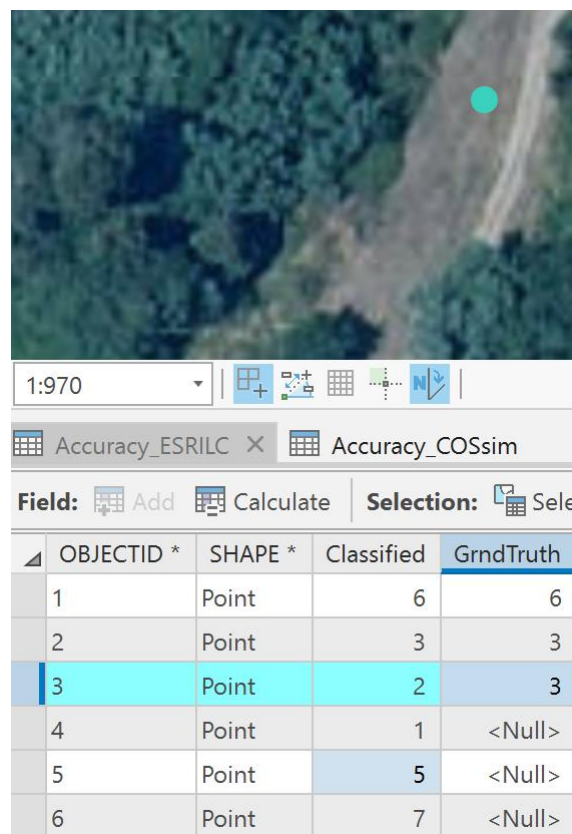


Figure 4.7 – Ground truth accuracy assessment process

In Figure 4.7, it is possible to observe that it is necessary to go through each sampling point and compare its classification with the actual ground truth satellite imagery data, in this case, point number 3 is located in a pixel that is classified as agricultural land,



when in reality is belongs to an area of shrubland & grassland. This process is then repeated by double clicking each point, looking at the area it is located and its classification, and fill the “GrndTruth” column with the correct class code, be it the same as the classified one or different.

### 4.2.3 Confusion Matrix and Results

With the samples being examined one by one based on their veracity, the attribute table for each layer will then contain the fully filled column of the ground truth data. Now in this stage of the analysis it is possible to compute a confusion matrix for each set of sample points, it is done by using the “Compute Confusion Matrix” tool from ArcGIS Pro, and it provides a new table containing errors of omission and commission, producer’s accuracy, and user’s accuracy as well as the most important index, the overall accuracy. These matrices are represented in Tables 4.3 to 4.7.

COSsim								
Class Value	C_1	C_2	C_3	C_4	C_5	C_6	Total	U_Accuracy
C_1	22	0	0	0	0	0	22	1
C_2	0	133	10	15	0	0	158	0.875
C_3	0	7	140	7	0	0	154	0.909
C_4	8	4	5	115	6	0	138	0.858
C_5	0	0	0	1	11	0	12	0.916
C_6	0	1	0	0	0	25	26	0.961
<b>Total</b>	30	145	155	138	17	25	500	
<b>P_Accuracy</b>	0.733	0.917	0.903	0.83	0.647	1		<b>0.838</b>

Table 4.3 – COSsim confusion matrix

ELC10									
Class Value	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Total	U_Accuracy
C_1	12	0	0	1	0	0	0	13	0.923
C_2	8	71	7	36	5	0	0	127	0.559
C_3	1	3	178	21	1	1	0	205	0.868
C_4	5	10	11	111	4	1	0	142	0.781
C_5	0	1	0	0	9	0	0	10	0.9
C_6	0	0	2	1	0	17	0	20	0.85

<b>C_7</b>	10	0	0	0	0	0	0	10	0
<b>Total</b>	36	85	198	170	19	19	0	500	
<b>P_Accuracy</b>	0.33	0.835	0.899	0.653	0.474	0.894	0		<b>0.583</b>

Table 4.4 – ELC10 confusion matrix

<b>Worldcover</b>								
<b>Class Value</b>	<b>C_1</b>	<b>C_2</b>	<b>C_3</b>	<b>C_4</b>	<b>C_5</b>	<b>C_6</b>	<b>Total</b>	<b>U_Accuracy</b>
<b>C_1</b>	12	0	0	0	0	0	12	1
<b>C_2</b>	0	20	0	4	0	0	24	0.833
<b>C_3</b>	0	0	113	27	0	0	140	0.807
<b>C_4</b>	2	7	18	263	5	8	303	0.882
<b>C_5</b>	0	1	0	1	11	0	13	0.846
<b>C_6</b>	0	0	0	2	0	18	20	0.9
<b>Total</b>	14	28	131	297	16	27	500	
<b>P_Accuracy</b>	0.857	0.714	0.863	0.886	0.688	0.667		<b>0.779</b>

Table 4.5 – Worldcover confusion matrix

<b>ESRILC</b>									
<b>Class Value</b>	<b>C_1</b>	<b>C_2</b>	<b>C_3</b>	<b>C_4</b>	<b>C_5</b>	<b>C_6</b>	<b>C_7</b>	<b>Total</b>	<b>U_Accuracy</b>
<b>C_1</b>	26	9	8	12	1	0	0	56	0.464
<b>C_2</b>	1	51	5	8	1	0	0	66	0.772
<b>C_3</b>	1	2	92	19	0	0	0	114	0.807
<b>C_4</b>	1	13	40	193	8	2	0	257	0.750
<b>C_5</b>	1	0	1	1	7	0	0	10	0.7
<b>C_6</b>	0	0	1	1	0	18	0	20	0.9
<b>C_7</b>	7	1	0	1	1	0	0	10	0
<b>Total</b>	37	76	147	235	18	20	0	500	
<b>P_Accuracy</b>	0.702	0.671	0.626	0.821	0.389	0.9	0		<b>0.587</b>

Table 4.6 – ESRILC confusion matrix

<b>S2GLC</b>									
<b>Class Value</b>	<b>C_1</b>	<b>C_2</b>	<b>C_3</b>	<b>C_4</b>	<b>C_5</b>	<b>C_6</b>	<b>C_7</b>	<b>Total</b>	<b>U_Accuracy</b>
<b>C_1</b>	13	4	0	0	1	0	0	18	0.722
<b>C_2</b>	7	58	17	26	4	0	0	112	0.517
<b>C_3</b>	2	4	146	14	1	2	0	169	0.863

<b>C_4</b>	4	16	12	139	7	1	0	179	0.776
<b>C_5</b>	1	1	1	1	7	0	0	11	0.636
<b>C_6</b>	1	0	3	0	1	15	0	20	0.75
<b>C_7</b>	7	0	1	0	1	1	0	10	0
<b>Total</b>	35	83	180	180	22	19	0	500	
<b>P_Accuracy</b>	0.371	0.698	0.811	0.772	0.318	0.79	0		<b>0.537</b>

The overall accuracies' values obtained through this analysis show that, when observing only Portugal, most maps provide lower accuracies than in a larger study area. This is due to a set of reasons, one reason that often affects the accuracies registered in European areas like Portugal or other southern countries is their very specific and complex landscapes [20]. European maps are made for very large areas and the landscape existent in Portugal (and the south) is often times not uniform. Some maps like ELC10, ESRILC and S2GLC had a significant number of points representing classes that were, when comparing to ground truth, not present in Portugal. Furthermore, in S2GLC, a significant portion of artificial surfaces with white roofing fell under the class of clouds when in reality that's not true, so when assessing its veracity, it was noted that it had the wrong class, reducing accuracy levels.

<b>Overall Accuracy</b>	<b>COSsim</b>	<b>ELC10</b>	<b>Worldcover</b>	<b>ESRILC</b>	<b>S2GLC</b>
Recalculated OA	84%	58%	78%	59%	54%

Table 4.8 – Recalculated accuracies after assessment for independent samples

The recalculated accuracies show that COSsim obtained the highest accuracy score 84% for the 6 classes of the unified nomenclature, 3 percentual points higher than the original. This means that there is credit to give due to its high levels of veracity in the mapping. It proved to be the map with the highest accuracy.

On the other hand, ELC10, ESRILC and S2GLC register a much smaller accuracy level for Continental Portugal, but ESA Worldcover map registered a higher score by 3 percentual points, its original accuracy level is 75%. This represents a very important result, and for this reason, the Worldcover map deserves extra attention. Despite being a world map, it showed to be capable to accurately classify Continental

Portugal in a way that no other international map on this list did. Its accuracy was measured to be almost as close as COSsim's one, which is a national map. This goes to show that ESA's Worldcover is, for these reasons, the best international map studied in this project.

When observing the accuracies per class, more specifically the producer's accuracy, it is notable that the Bare/Sparsely Vegetated class had the lower levels of accuracy at an average of just 50% between all maps. However, there is a need to be careful with these values, due to the use of a Stratified Random sampling, the smaller classes have less points, which leads to bigger confidence intervals, which can deviate a lot the value of their accuracy.

#### **4.2.4 Resampling Approach**

The sampling approach taken so far, has only considered independent samples, this is, for each map evaluated, a different set of accuracy assessment points were created. This can cause some statistical uncertainty, due to each map being attributed a new accuracy value that is dependent on a set of points that is different for other maps. For example, for a certain map, that set of 500 points created could have fallen onto certain pixels located in geographical areas that are more prone for error, such as areas that see changes throughout the year, be it rotative agricultural fields or others. For another map, the set of 500 pixels could have fallen onto much more stable geographical features that do not see quick changes, such as grasslands, roads, lakes, and others.

Due to this exact reason, it was found necessary to assess another approach of sampling that could add to the one already done, and that could help assure less inconsistencies in the accuracy results. For this, a fixed set of 500 points were created using once again the "Create Accuracy Assessment Points" tool in ArcGIS Pro. Figure 4.8 represents the accuracy assessment points.

Accuracy Assessment Points - Approach for all maps

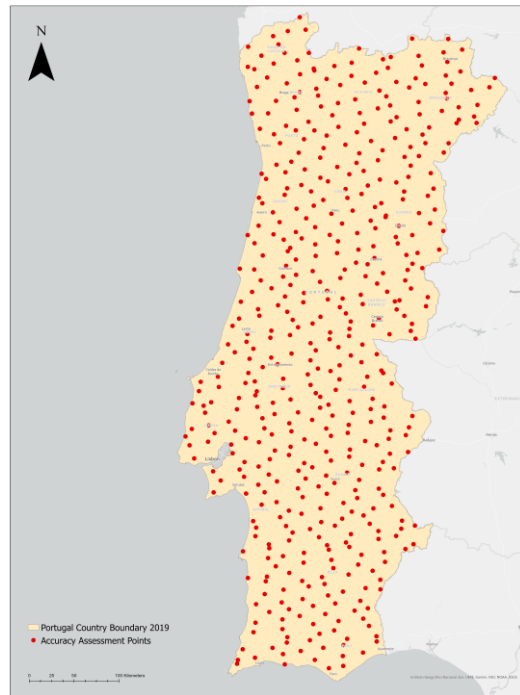


Figure 4.8 – Set of accuracy assessment points for all maps

The sampling strategy chosen for the distribution in this case has to be fully Random, because as discussed in 4.2.1, both Stratified Random and Equalized Stratified Random sampling approaches depend on the size or amount of classes. In this case, seen that the desire is to have the same 500 points for all maps, the distribution of the points cannot depend on the classes of any map. The Random sampling approach ensures that the distribution is fully random on the image, the image in this case being the boundaries of the study area of Continental Portugal.

The procedure of recalculating the accuracy levels is the same as done previously for the independent samples, it will be done by photointerpretation of the samples. This will entail the creation of 5 new layers, one for each map, all of them with the same set of points presented in Figure 4.8. The photointerpretation process is the same as described in 4.2.2.

#### **4.2.4.1 Confusion Matrix and Results**

This section will, once again, state all the confusion matrices that result from the accuracy assessment photointerpretation. The “Compute Confusion Matrix” tool from ArcGIS Pro is used to calculate these matrices that will be able to provide crucial data

and values such as, errors of omission and commission, producer's accuracy, and user's accuracy as well as the most important index, the overall accuracy (Tables 4.9 to 4.13).

COSsim								
Class Value	C_1	C_2	C_3	C_4	C_5	C_6	Total	U_Accuracy
C_1	30	0	0	0	0	0	30	1
C_2	0	133	4	10	0	0	147	0.904
C_3	2	7	140	3	0	0	152	0.921
C_4	5	3	5	96	6	5	120	0.8
C_5	1	0	0	1	11	0	13	0.846
C_6	0	0	0	0	0	25	25	1
<b>Total</b>	38	143	160	110	19	30	500	
<b>P_Accuracy</b>	0.789	0.930	0.875	0.872	0.578	0.833		<b>0.813</b>

Table 4.9 – COSsim confusion matrix recalculated

ELC10									
Class Value	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Total	U_Accuracy
C_1	22	0	3	2	0	0	0	27	0.814
C_2	5	81	0	12	16	0	0	114	0.710
C_3	0	0	88	8	0	0	0	96	0.916
C_4	0	0	18	108	13	10	0	149	0.724
C_5	1	17	0	5	36	0	0	59	0.610
C_6	0	2	0	7	0	14	0	23	0.608
C_7	5	0	21	0	0	5	0	31	0
<b>Total</b>	33	101	130	142	65	29	0	500	
<b>P_Accuracy</b>	0.667	0.801	0.676	0.760	0.553	0.482	0		<b>0.563</b>

Table 4.10 – ELC10 confusion matrix recalculated

Worldcover								
Class Value	C_1	C_2	C_3	C_4	C_5	C_6	Total	U_Accuracy
C_1	6	0	0	0	3	0	9	0.667
C_2	0	74	3	0	0	0	77	0.961
C_3	0	0	145	27	0	2	174	0.833
C_4	4	15	13	120	0	8	160	0.75
C_5	0	8	4	1	29	0	42	0.690

<b>C_6</b>	0	0	0	2	6	30	38	0.789
<b>Total</b>	10	97	165	150	38	40	500	
<b>P_Accuracy</b>	0.6	0.762	0.878	0.8	0.763	0.75		<b>0.759</b>

Table 4.11 – Worldcover confusion matrix recalculated

<b>ESRILC</b>									
<b>Class Value</b>	<b>C_1</b>	<b>C_2</b>	<b>C_3</b>	<b>C_4</b>	<b>C_5</b>	<b>C_6</b>	<b>C_7</b>	<b>Total</b>	<b>U_Accuracy</b>
<b>C_1</b>	36	0	5	19	1	0	0	61	0.590
<b>C_2</b>	0	70	2	10	1	4	0	87	0.804
<b>C_3</b>	0	2	113	13	0	6	0	134	0.843
<b>C_4</b>	1	3	8	150	6	0	0	168	0.892
<b>C_5</b>	1	0	4	2	11	0	0	18	0.611
<b>C_6</b>	0	0	3	1	0	23	0	27	0.851
<b>C_7</b>	0	0	1	1	2	1	0	5	0
<b>Total</b>	38	75	136	196	21	34	0	500	
<b>P_Accuracy</b>	0.947	0.933	0.831	0.765	0.524	0.676	0		<b>0.668</b>

Table 4.12 – ESRILC confusion matrix recalculated

<b>S2GLC</b>									
<b>Class Value</b>	<b>C_1</b>	<b>C_2</b>	<b>C_3</b>	<b>C_4</b>	<b>C_5</b>	<b>C_6</b>	<b>C_7</b>	<b>Total</b>	<b>U_Accuracy</b>
<b>C_1</b>	16	0	0	0	1	0	0	17	0.941
<b>C_2</b>	0	32	40	18	6	3	0	99	0.323
<b>C_3</b>	0	2	131	9	0	3	0	145	0.903
<b>C_4</b>	4	5	16	123	3	0	0	151	0.815
<b>C_5</b>	0	9	12	1	19	4	0	45	0.422
<b>C_6</b>	2	0	1	8	0	26	0	37	0.703
<b>C_7</b>	0	0	1	0	0	5	0	6	0
<b>Total</b>	22	48	201	159	29	41	0	500	
<b>P_Accuracy</b>	0.727	0.668	0.652	0.774	0.655	0.634	0		<b>0.587</b>

Table 4.13 – S2GLC confusion matrix recalculated

The results of the resampling approach demonstrate that the accuracy values remain around the same number as registered in the independent sampling analysis. This allows for a further comprehension on the possibility of an existence of a bias. However, seen that, for the same number of points, and same type of

photointerpretation analysis, the overall accuracy values do not differ significantly, it is possible to generally confirm that there are no major bias associated with the accuracy points in the independent sampling analysis.

Observing the matrices, it is possible to note that some trends persisted, such as the fact that the Bare/Sparsely Vegetated class is still generally the class with the lowest accuracy values, however with a higher average than before, reaching approximately 61%.

<b>Overall Accuracy</b>	<b>COSSim</b>	<b>ELC10</b>	<b>Worldcover</b>	<b>ESRILC</b>	<b>S2GLC</b>
Recalculated OA	81%	56%	76%	67%	59%

Table 4.14 – Recalculated accuracies after assessment for the resampling approach

Table 4.14 states the resampling approach recalculated overall accuracies for all five maps. The most significant change in accuracy levels was seen for the ESRILC map, which previously shown to have a 59% overall accuracy for Continental Portugal, however, when recalculating it, this value increased eight percentual points to 67% for the same area.

### **4.3 Finding Hot Spots**

With the accuracy assessment tools put into practice, and with the photointerpretation done for each point of each map, it is now possible to gather the number of points in which their original classification does not match the photointerpretation one. For each accuracy assessment points' layer, a copy was made. Then, the attribute table was opened and then all the rows in which the “Classified” nomenclature code did not match the “GrndTruth” (*Ground Truth*) nomenclature code, were chosen. The remaining rows were deleted for each map. This leaves the result of all the points that were misclassified for each map (Figures 4.9 to 4.13). Taking not that this is done for the accuracy assessment points of the independent sampling analysis, where the points are differently located for each map. This process was not repeated for the resampling strategy, where all maps have the same location of accuracy assessment points, because the overall accuracies' results proved to not be largely significantly different.



Misclassified Points - COSSim

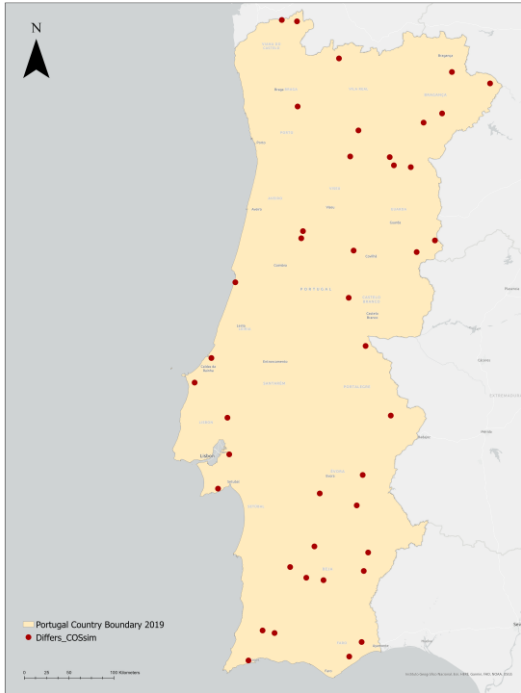


Figure 4.9 – Misclassified points for COSSim accuracy assessment photointerpretation

Misclassified Points - ELC10

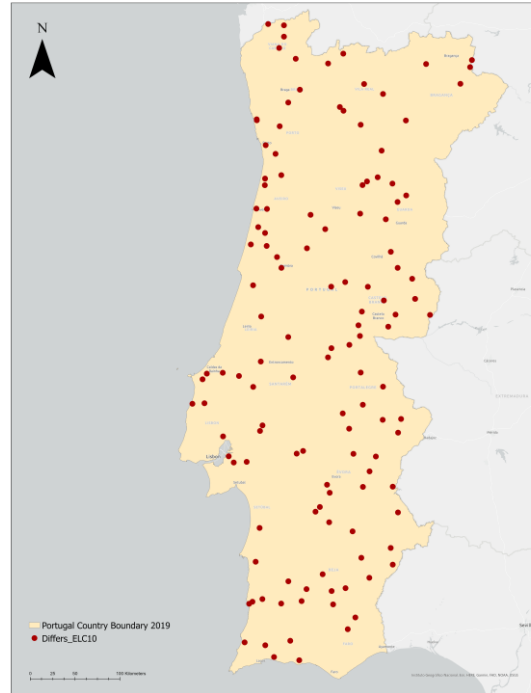


Figure 4.10 – Misclassified points for ELC10 accuracy assessment photointerpretation

Misclassified Points - Worldcover

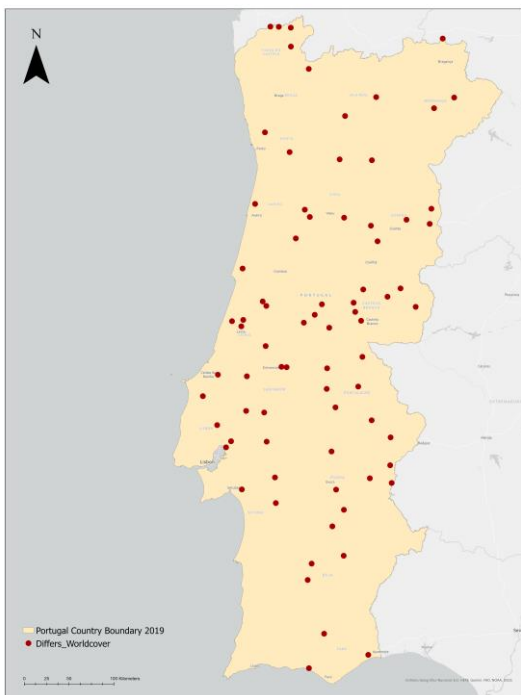


Figure 4.11 – Misclassified points for Worldcover accuracy assessment photointerpretation

Misclassified Points - ESRILC

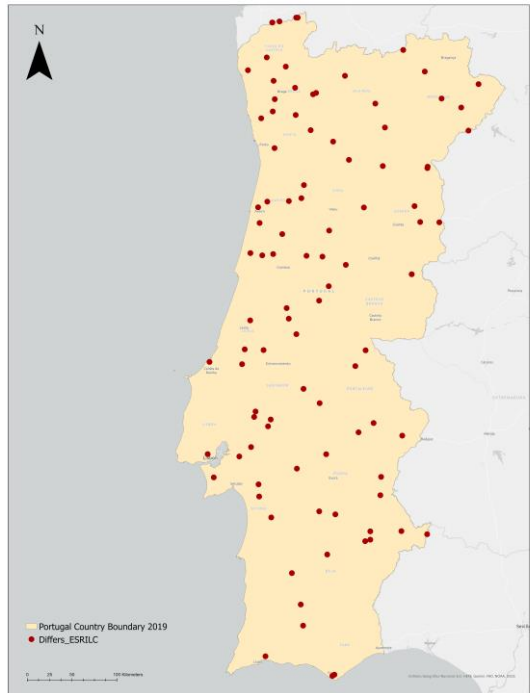


Figure 4.12 – Misclassified points for ESRILC accuracy assessment photointerpretation

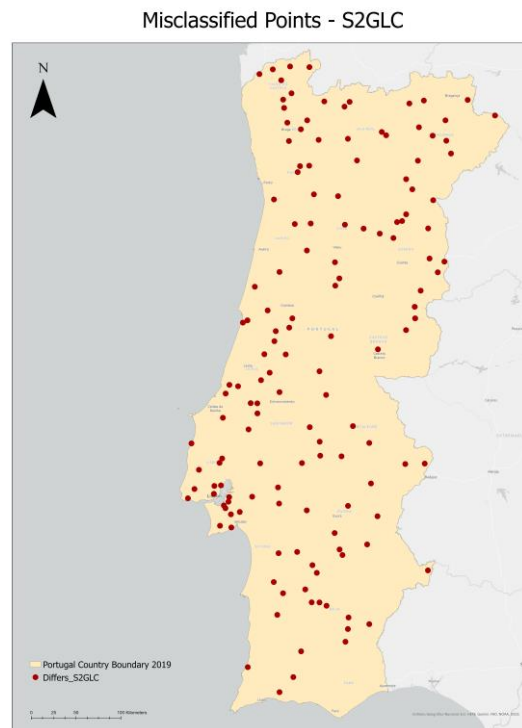


Figure 4.13 – Misclassified points for S2GLC accuracy assessment photointerpretation

In Figures 4.9 to 4.13, it is noticeable that some maps, when subjected to the accuracy assessment photointerpretation, have a much smaller number of points which were misclassified, this can be related to the overall accuracy scores, the higher the score the least number of misclassified points.

This is visually noticeable when observing the COSsim map and the Worldcover map, both these have an apparent smaller number of points. It is also hardly possible to identify very specific areas that registered false classification samples, this is, the points that represent a misclassification are distributed, in general, all throughout the map, without focus on any specific area. Nonetheless, this does not mean that some areas more affected do not exist.

To understand where are located some possible areas that have a larger amount of misclassified points, and some areas that had better classification, a hot spot approach was chosen. This was done using the ArcGIS Pro tool of “Find Hot Spots”, which given a set of features, in this case point features, it utilizes the Getis-Ord  $G_i^*$  statistic to find statistically significant hot spots and cold spots. The tool was run 5 times, one

for each map's layer of misclassified points, set with the following parameters, a Bin Size of 20km and a Neighborhood Size of 25km. Justifying the parameters, the distance chosen for both the Bin Size and the Neighborhood Size was based on a distance that was allowed to create a sensible amount of detail throughout the map, however a smaller size in these parameters would leave very large chunks of the map out of the analysis. Moreover, a larger size would generalize the areas in a way where defining a statistically significant cold or hot spot would be more prone to error.

Figures 4.14 to 4.18 represent the 20 by 20 km squares in which there were misclassified points, and its colors represent if it is an area that is prone to being a cold spot, a hot spot, or statistically insignificant.

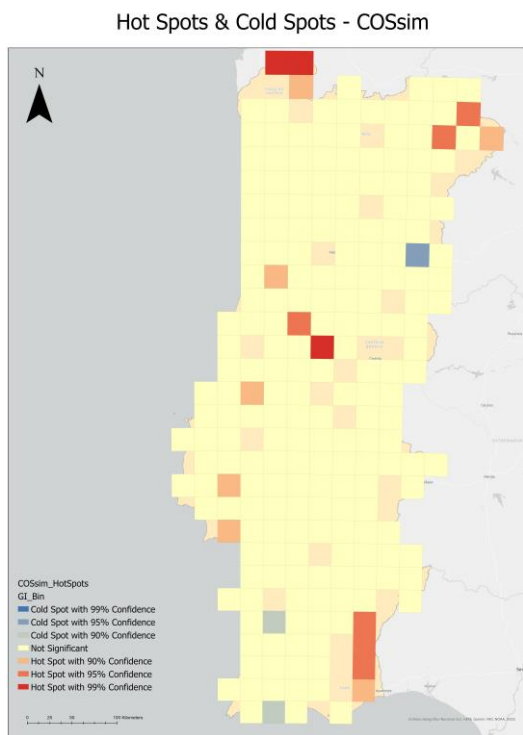


Figure 4.14 – Hot and cold spots of misclassification of sample points in COSSim

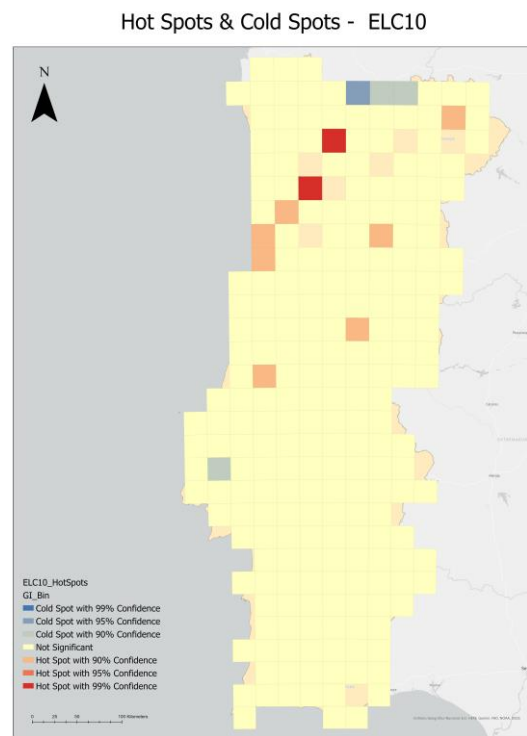


Figure 4.15 – Hot and cold spots of misclassification of sample points in ELC10

Hot Spots & Cold Spots - Worldcover

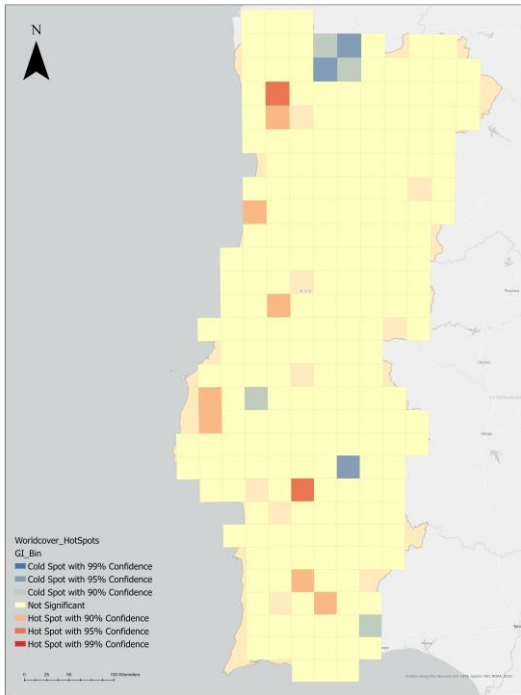


Figure 4.16 – Hot and cold spots of misclassification of sample points in Worldcover

Hot Spots & Cold Spots - ESRILC

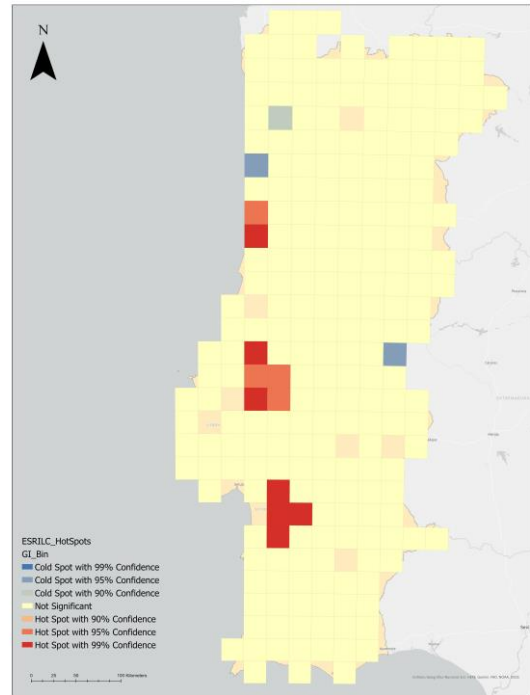


Figure 4.17 – Hot and cold spots of misclassification of sample points in ESRILC

Hot Spots & Cold Spots - S2GLC

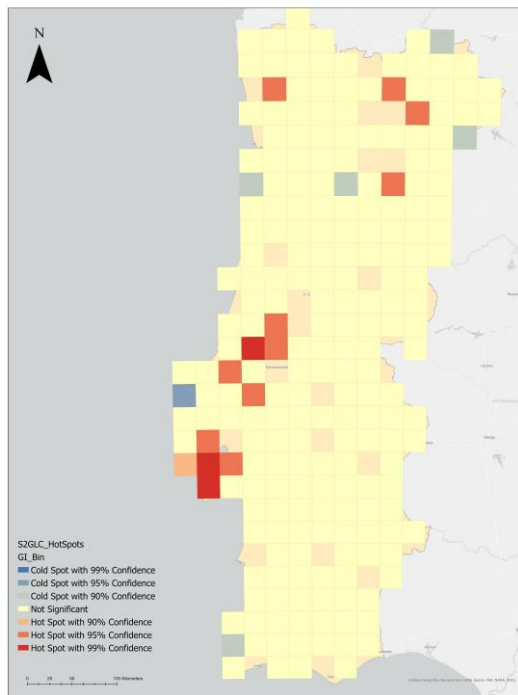


Figure 4.18 – Hot and cold spots of misclassification of sample points in S2GLC

The hot and cold spots show different realities for each map. A hot spot is defined as an area, point or location in which there is a bigger concentration of point features for the defined distances. A cold spot represents the opposite, an area, point or location in which there is no concentration of point features, or this concentration is more insignificant than the average.

For the maps represented in Figures 4.14 to 4.18, it is possible to observe that some of the 20 by 20 km squares do not possess any type of classification, this is due to the lack of points of misclassification in that area. Without the point features present, it is not possible to conclude upon the existence of neither a hot or cold spot, or even its statistical insignificance.

The map that contains the largest areas of hot spots is the ESRILC one, meaning that there exists a higher concentration of misclassified points, most specifically in the North Alentejo and Ribatejo regions. S2GLC also contains a large hot spot over the city of Lisbon and the Ribatejo region. In the Worldcover map, exists a large cold spot over the Gerês Mountains region. These findings aid in the identification of areas of bigger misclassification based on samplings tested.

## **5. DISCUSSIONS AND CONCLUSIONS**

This chapter serves to collect all the knowledge gathered throughout the present work, and structure it. It serves to discuss what has been done, what conclusions can be realized, as well as limitations of the work and what could be done for potential improvement, or continuation of this work in the future. Moreover, it is crucial to understand if the stated objectives and research questions in the beginning were answered throughout the research, and in what ways were they answered.

### **5.1 Results**

Throughout the work, results were obtained from the analysis and evaluations made. When observing the comparison of nomenclatures, it is possible to conclude and confirm that there exists a large variety of classes throughout the maps, in some cases, maps contain classes that are not even present in the study area of Continental Portugal, for this case classes were removed. The creation of a unified nomenclature resulted in a harmonization of the maps' classification that allows for comparisons between them.

All maps occupy the total length of the study area, but each one does it differently, each class occupies a different size for each map, due to different methods of classification or different algorithms applied to the map. When observing the distribution of each class size per map it was noted that for each class the maps follow similar trends, meaning some classes always characterize the largest portion(s) of the land, while others do the opposite, and occupy the smallest portion(s) of the study area.

When validating the maps' accuracies, through photointerpretation of sample, findings showed that maps like ELC10, ESRILC and S2GLC proved to have a much higher overall accuracy when classifying their full study area, be it Europe, or in the case of ESRILC, the World. Their accuracies dropped from 86% for all to 58%, 59% and 54% respectively. On the other hand, the maps of Worldcover and COSsim saw their overall accuracies increase after the accuracy assessment, from 81% and 75% respectively, to 84% and 78% respectively as well.

However, it is crucial to note that, the original accuracy values were measured for a different number of classes, for example, for the COSsim map, the 81% overall accuracy was measured for 13 classes. In the case of this work's recalculated overall accuracy, it was based on the unified nomenclature of 6 classes.

In the resampling approach, where the same location of a set of 500 points were considered for all maps, almost the exact same occurred. The maps ELC10, ESRILC and S2GLC dropped in their accuracies from 86% all to 56%, 67% and 59% respectively, while Worldcover increased from 75% to 76%, and COSsim remained at its 81% of overall accuracy.

Taking in consideration the fact that the accuracy sampling was done for a smaller amount of classes, and still, this set of maps, ELC10, ESRILC and S2GLC, provided significantly smaller values of accuracy, these maps proved to be very lacking in quality of classification for Continental Portugal.

Looking at the findings from the computation of hot and cold spots regarding the existence of misclassified points, it was found that there are, in its majority no very significant areas where hot and cold spots happened. However, the Ribatejo area seemed to persist on the hot spots, especially for the maps of ESRILC and S2GLC, which says that it is an area in which misclassification happened. Thus, this could also be an area where the map is least accurate, when comparing to other places within the study area. Other methods of processing and creating spatially distributed measures of accuracy, is through a process defined in [26], which discusses the use of geographically weighted models, for assessing spatial error variations.

## **5.2 Conclusions**

With the conclusion of this work, a comparison between a set of 5 maps was done, with the addition of an assessment on their validation. With the results in mind, it should be evaluated if the research questions proposed in the beginning were answered through the work.

When it comes to understanding if internationally developed maps for large regions, lack in accuracy when compared to national maps, they do, mostly only. In the case of

this research ESA Worldcover's map proved to sustain its accuracy levels when looking only into Continental Portugal, however it was the only one, the remaining international maps strongly proved the opposite. The COSsim map, the only national map, maintained its accuracy levels as well. The main intake to conclude upon this is regarding the fact that, the results of this study were able to demonstrate that there is a need for the creation of more national maps, due to these being able to provide higher accuracy scores. Other examples besides the Portuguese one rely on, for example, the French Theia land cover map [27], and the Austrian land information system of Austria (LISA) 2.0 [28], where both these maps invest on a more exact classification of these countries in specific.

Although large scale maps come to be very useful, it is also to note that when observing into specific areas it could happen that their accuracy is not as good, or plainly mediocre. The land cover of a large region will be much more heterogeneous than of a smaller region, making it ever so difficult to classify properly. This could severely impact the usage of the map and skew any results, analysis or observations made with said map. It is here when it is crucial to note the importance of each country or region developing their own land use land cover maps, to preserve quality. Nonetheless, the recently released Worldcover map, is a new generation map that, through these good results, shows that the classification methods have evolved. This poses a question that, although national maps proved to be better in accuracy levels, new generation world cover maps could also be able to provide the same or close high accuracies in a large-scale area.

### **5.3 Limitations and Future Work**

The limitations encountered, impacted the steady flow of the work. These are related to various reasons, either be it external or internal factors. Regarding the data used, it is important to note that mapping entities describe what each class in their nomenclature consists of, but there is always present a limitation of how precise one can be when taking in consideration all the characteristics of a class. This means that, according to various degrees of classification of let's say urban areas, there always exists a limit on what ends up being classified urban between maps. One could consider certain types of areas or infrastructures in that class, and another map could



not consider the same. Also, important to note that depending on the algorithm or way of processing used by the mapping entity in question, the results of the classification will vary as well. This will result in maps having varied sizes of classes due to what is processed as what. Comparing these maps created by different strategies and approaches is harder.

Another limitation is regarding the fact that the accuracy assessment done was not possible to be done on the original satellite imagery used. Due to the complexity of the interpreted ground truth of each map, the access to the actual compilation of imagery is not possible. The mapping entities provide open access only to the already classified maps. Thus, the photointerpretation of the samples (points) in the maps with the ground truth data was done using ArcGIS Pro imagery base map, issues that could arise from this is in the misclassification of areas susceptible to quick change, such as artificial land areas.

The photointerpretation method is delicate and can easily be manipulated by biases, for this exact reason it is important to not forget that human visual interpretation is not guaranteed to be always precise.

## BIBIOGRAPHIC REFERENCES

- [1] Hashimoto, S., Tadono, T., Onosato, M., Hori, M. and Moriyama, T., 2011, July. A framework of ontology-based knowledge information processing for change detection in remote sensing data. In *2011 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3927-3930). IEEE.
- [2] Congalton, Russell & Gu, Jianyu & Yadav, Kamini & Ozdogan, Mutlu. (2014). Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sensing*. 6. 12070. 10.3390/rs61212070.
- [3] Alshari, E.A. and Gawali, B.W., 2021. Development of classification system for LULC using remote sensing and GIS. *Global Transitions Proceedings*, 2(1), pp.8-17.
- [4] Arvor, D., Belgiu, M., Falomir, Z., Mougenot, I. and Durieux, L., 2019. Ontologies to interpret remote sensing images: why do we need them?. *GIScience & remote sensing*, 56(6), pp.911-939.
- [5] Hasmadi, M., Pakhriazad, H.Z. and Shahrin, M.F., 2009. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia: Malaysian Journal of Society and Space*, 5(1), pp.1-10.
- [6] Canters, F., 1997. Evaluating the uncertainty of area estimates derived from fuuy land-cover classification. *Photogrammetric Engineering & Remote Sensing*, 63(4), pp.403-414.
- [7] Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), pp.185-201.
- [8] Soulard, C.E. and Acevedo, W., 2013, December. Multi-temporal harmonization of independent land-use/land-cover datasets for the conterminous United States. In *AGU Fall Meeting Abstracts* (Vol. 2013, pp. B41E-0448).

- [9] Yang, H., Li, S., Chen, J., Zhang, X. and Xu, S., 2017. The standardization and harmonization of land cover classification systems towards harmonized datasets: a review. *ISPRS International Journal of Geo-Information*, 6(5), p.154.
- [10] Herold, M., Woodcock, C.E., Di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J. and Schmullius, C.C., 2006. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp.1719-1727.
- [11] Gaur, S., Mittal, A., Bandyopadhyay, A., Holman, I. and Singh, R., 2020. Spatio-temporal analysis of land use and land cover change: a systematic model inter-comparison driven by integrated modelling techniques. *International Journal of Remote Sensing*, 41(23), pp.9229-9255.
- [12] Fonte, C.C., See, L., Laso-Bayas, J.C., Lesiv, M. and Fritz, S., 2020. Assessing The Accuracy Of Land Use Land Cover (Lulc) Maps Using Class Proportions In The Reference Data. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(3).
- [13] Foody, G.M., 2006. What is the difference between two maps? A remote senser's view. *Journal of Geographical Systems*, 8(2), pp.119-130.
- [14] Dingle Robertson, L. and King, D.J., 2011. Comparison of pixel-and object-based classification in land cover change mapping. *International Journal of Remote Sensing*, 32(6), pp.1505-1529.
- [15] Kassawmar, T., Eckert, S., Hurni, K., Zeleke, G. and Hurni, H., 2018. Reducing landscape heterogeneity for improved land use and land cover (LULC) classification across the large and complex Ethiopian highlands. *Geocarto international*, 33(1), pp.53-69.
- [16] Fonte, C.C., Patriarca, J. and Duarte, D., 2021. Effects of Geospatial Data Sources on the Identification and Characterization of Burnt Areas in Portugal. *ISPRS*

Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, pp.49-54.

[17] Fonte, C.C., See, L., Lesiv, M. and Fritz, S., 2019. A preliminary quality analysis of the climate change initiative land cover products for continental Portugal. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, pp.1213-1220.

[18] Rwanga, S.S. and Ndambuki, J.M., 2017. Accuracy assessment of land use/land cover classification using remote sensing and GIS. International Journal of Geosciences, 8(04), p.611.

[19] Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote sensing of environment, 80(1), pp.185-201.

[20] Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski, A., Wojtkowski, C., Krupiński, M., Krätzschmar, E. and Schauer, P., 2020. Automated production of a land cover/use map of Europe based on Sentinel-2 imagery. Remote Sensing, 12(21), p.3523.

[21] Page access for download link and files of COSsim 2018 map <https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/metadata/262dabc4-0217-44c1-8983-21f69d8670c7>

[22] Page access for download link and files of ELC 10 map <https://zenodo.org/record/4407051#.YhMevuj7Rdg>

[23] Page access for download link and files for ESA Worldcover map Wroldcover<https://worldcover2020.esa.int/download>

[24] Page access for download link and files for ESRILC map <https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac>

[25] Page access for download link and files for S2GLC map  
<https://s2glc.cbk.waw.pl/extension>

[26] Alexis J. Comber (2013): Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies, *Remote Sensing Letters*, 4:4, 373-380

[27] Page access for Theia Land Cover SEC <https://www.theia-land.fr/en/product/land-cover-map/>

[28] Page access for LISA 2.0 [https://land.copernicus.eu/eagle/files/eagle-related-projects/LISA\\_COPERNICUS\\_Banko\\_Bratislava\\_Nov\\_2015.pdf](https://land.copernicus.eu/eagle/files/eagle-related-projects/LISA_COPERNICUS_Banko_Bratislava_Nov_2015.pdf)

## ***LULC MAP COMPARISON***

***Comparison and validation of land use land cover maps derived from satellite imagery***

Fábio Miguel Domingues da Silva

2022

**LULC Map Comparison**  
*Comparison and validation of land use land cover maps derived from satellite imagery*

Fábio Miguel Domingues da Silva



Masters  
Program  
in **Geospatial  
Technologies**



Supported by:



Education and Culture

**ERASMUS MUNDUS**