land

MDPI

Article

Thematic Comparison between ESA WorldCover 2020 Land Cover Product and a National Land Use Land Cover Map

Diogo Duarte [1,*], Cidália Fonte [1,2], Hugo Costa [3,4] and Mário Caetano [3,4]

[1] Institute for Systems Engineering and Computers at Coimbra (INESC Coimbra), DEEC, University of Coimbra, Polo 2, 3030-290 Coimbra, Portugal
[2] Department of Mathematics, University of Coimbra, Apartado 3008, EC Santa Cruz, 3001-501 Coimbra, Portugal
[3] Direção-Geral do Território, Rua da Artilharia Um, 107, 1099-052 Lisbon, Portugal
[4] NOVA Information Management School (NOVA IMS), Campus de Campolide, Nova University Lisbon, 1070-312 Lisbon, Portugal
* Correspondence: diogovad@inescc.pt

Abstract: This work presents a comparison between a global and a national land cover map, namely the ESA WorldCover 2020 (WC20) and the Portuguese use/land cover map (Carta de Uso e Ocupação do Solo 2018) (COS18). Such a comparison is relevant given the current amount of publicly available LULC products (either national or global) where such comparative studies enable a better understanding regarding different sets of LULC information and their production, focus and characteristics, especially when comparing authoritative maps built by national mapping agencies and global land cover focused products. Moreover, this comparison is also aimed at complementing the global validation report released with the WC20 product, which focused on global and continental level accuracy assessments, with no additional information for specific countries. The maps were compared by following a framework composed by four steps: (1) class nomenclature harmonization, (2) computing cross-tabulation matrices between WC20 and the Portuguese map, (3) determining the area occupied by each harmonized class in each data source, and (4) visual comparison between the maps to illustrate their differences focusing on Portuguese landscape details. Some of the differences were due to the different minimum mapping unit ofCOS18 and WC20, different nomenclatures and focuses on either land use or land cover. Overall, the results show that while WC20 detail is able to distinguish small occurrences of artificial surfaces and grasslands within an urban environment, WC20 is often not able to distinguish sparse/individual trees from the neighboring cover, which is a common occurrence in the Portuguese landscape. While selecting a map, users should be aware that differences between maps can have a range of causes, such as scale, temporal reference, nomenclature and errors.

Keywords: land use land cover; ESA WorldCover; Portuguese authoritative thematic data; Portugal; map comparison

Citation: Duarte, D.; Fonte, C.; Costa, H.; Caetano, M. Thematic Comparison between ESA WorldCover 2020 Land Cover Product and a National Land Use Land Cover Map. *Land* **2023**, *12*, 490. https://doi.org/10.3390/land12020490

Academic Editors: Eufemia Tarantino, Enrico Borgogno-Mondino, Marco Scaioni and Alessandra Capolupo

Received: 15 December 2022
Revised: 7 February 2023
Accepted: 13 February 2023
Published: 16 February 2023

1. Introduction

Land use/land cover (LULC) information is necessary for many research areas, such as climate change monitoring [1], deforestation monitoring [2], global biomass estimation [3] and assessing several United Nations Sustainable Development Goals [4] among others [5,6]. The production of such LULC information is currently driven by both national and international initiatives and institutes, each considering a different set of aims and objectives which are then reflected in the final product. In general, global maps are often focused on the land cover component and with fewer classes when compared to more local to national maps, often produced by national mapping agencies which usually have dedicated products for land use considering a larger set of thematic classes.

Many efforts have been made to globally map LULC. Finer Resolution Observation and Monitoring—Global Land Cover (FROM-GLC) [7] was the first global map produced with a 10 m spatial resolution, having 2017 as the reference year and produced with Landsat and Sentinel-2 imagery. ESRI 2020 Land Cover is entirely based on Sentinel-2 imagery with 10 m spatial resolution and maps 10 LULC classes [8], already having a 2021 version of the product available. Previous products exist, but with lower spatial resolutions given that these products were mainly based on Landsat data. GlobeLand30 is one of these products [9], mapping 10 LULC classes at a 30 m spatial resolution, using satellites Landsat HJ-1 (China Environment and Disaster Reduction Satellite) and GD-1 (China High Resolution Satellite). More recently, in October 2021, the ESA WorldCover2020 (WC20) product [10] was released by the European Space Agency and aims at mapping 11 land cover classes at a 10 m resolution based on both Sentinel-1 and Sentinel-2 imagery, for the reference year 2020. Together with this product, an user [8] and a validation report [11] were also released. According to the validation report, the classes tree cover (TC), cropland (CL) and permanent water bodies (PW) were the best mapped with WC20, whereas the classes shrubland (SL), bare/sparse vegetation (BSV), snow and ice (SI) and moss and lichen (ML) were the ones presenting the worst results.

While global and continent-wide accuracy assessments are usually performed for the products as a whole, there is no information regarding their thematic accuracy for individual countries. This is more relevant given the high resolution of the product, which may stimulate the use of global and continental products in regional applications covering a geographical fraction of the Earth's surface much smaller than that of the original product. While focused on global scale mapping, LULC maps may not be able to correctly map underrepresented landscapes occurring in restricted regions of the world [12]. For example, [13,14] have indicated Southern Europe as one of the regions presenting a lower thematic accuracy than the ones reported for the globe/continent when considering several previously released LULC products.

For a long period of time, national mapping institutes have been providing national LULC maps. Such maps are often tailored for a given national extent, taking into account the intra-national landscapes [15] which aim at delivering products with more thematic detail, often focusing on the land use and considering a larger set of classes compared to global products. For example, in the Portuguese case, the National Mapping Agency releases a thematic map every 5 years denominated "Carta de Uso e Ocupação do Solo"; the 2018 version (COS18) for continental Portugal contains 83 classes in a hierarchical approach with four levels [16].

Currently, users in need of LULC information have a large amount of publicly available maps at their disposal, either national maps released by national mapping institutes or continental/global maps released by international organizations. These maps present major differences, given the different objectives and needs by each of the map producers. Varying MMU, class nomenclatures and focus (either on land cover or land use) are among the most significant differences.

This paper reports the results of a thematic comparison between a national (COS18) and a global (WC20) LULC dataset to illustrate the potential differences that users can find while selecting a LULC map. The comparison is based on the computation of the overall proportion of agreement between the two datasets, each of the datasets classes' areal coverage and a visual comparison. The study area is mainland Portugal and three smaller study areas within Portugal and having different landscape characteristics. In this way, these experiments provide an extra set of information regarding the WC20 dataset's thematic contents and how it relates with an authoritative thematic map. Moreover, the framework used here can also be applied in different countries and aid potential users in assessing the overall similarities and discrepancies between global and national LULC products.

2. Study Areas and Data

The study area is continental Portugal, located in the Iberian Peninsula. Three regions were also selected to perform the comparison, which are representative of the different landscapes present within continental Portugal, and also of differences regarding population density, urbanization degree and climate [17] (Figure 1). Areas A and C represent two different landscapes. Region A has high population density and a fine-grained mosaic of land use in a humid climate, while region C is characterized by large plains and farms, and hence has more homogenous land use patterns within a drier climate. Region B contains the metropolitan area of Lisbon, the Portuguese capital, and has a higher urbanization and population density.

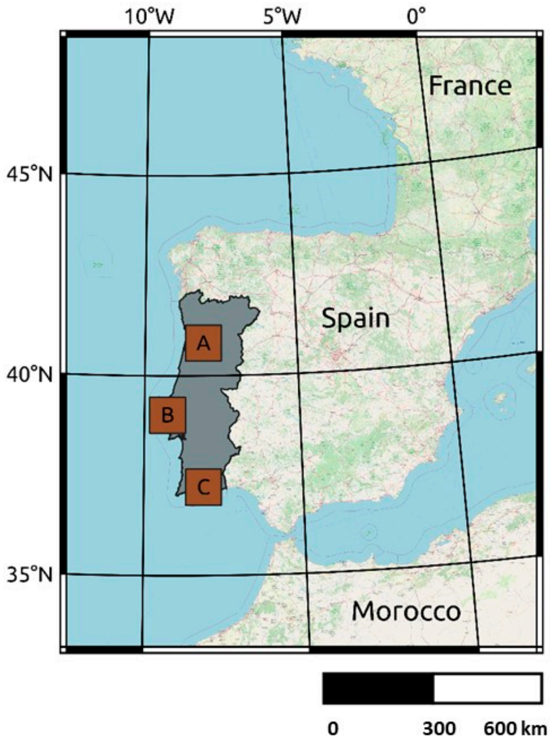


Figure 1. Continental Portugal area in gray with the several regions A, B and C in brown. Grid in WGS84 Geographical Coordinates.

ESA WorldCover 2020 is a global land cover raster map produced using image recognition methods and Sentinel-1 and 2 data (both radar and optical imagery) as input, mapping 11 land cover classes at a 10 m resolution. The statistical accuracy assessment used the Copernicus Global Land Service Validation data [18], which is based on probability sampling. The validation data contain more than 20,000 primary sampling units (PSUs) corresponding to 100 m $\times$ 100 m square spatial units. Each of these PSUs was divided into 10 $\times$ 10 m blocks denominated secondary sampling units (SSUs). Regarding the statistical accuracy assessment for Europe (EU), the overall accuracy was 76.8% $\pm$ 0.2%, computed with a total of 3118 PSUs and their corresponding SSUs.

COS2018 is a vector land use/land cover map with a MMU of 1 hectare, manually produced by photointerpretation of orthophotomaps of decimeter spatial resolution and auxiliary information [19]. The COS18 thematic nomenclature is hierarchical with four levels. An example of its detail and land use focus can be seen in Table 1, showing the several subclasses in all four levels of the cropland main class. The remainder of the nomenclature can be seen in the original COS18 technical specifications document due to the large number of classes and sub-classes, 83 in total (DGT, 2019; page 15). As can be seen, level 4 has fine thematic details, where olive tree stands are distinguished from olive tree stands with temporary croplands within the trees. While land cover may be extrapolated from each of these classes, there is often a problem with, for example, heterogenous areas of tree cover and croplands (e.g., classes 2.3.1.x) which are represented in a single class. The technical specifications indicate 85% as the minimum overall accuracy of the product.

Table 1. Example of class subdivision of class 2 (Croplands) in COS18. The remainder of the extensive nomenclature can be seen on page 15 in DGT,2019.

Level 1	Level 2	Level 3	Level 4
2—Cropland	2.1—Temporary cropland	2.1.1—Temporary croplands	2.1.1.1—Temporary croplands (arable and non-arable land)
			2.1.1.2—Rice fields
	2.2—Permanent cropland	2.2.1—Vineyards	2.2.1.1—Vineyards
		2.2.2—Fruit trees	2.2.2.1—Fruit trees
		2.2.3—Olive trees	2.2.3.1—Olive trees
	2.3—Heterogenous cropland	2.3.1—Temporary croplands	2.3.1.1—Temporary croplands associated with vineyards
			2.3.1.2—Temporary croplands associated with fruit trees
			2.3.1.3—Temporary croplands associated with olive trees
		2.3.2—Complex cropland landscape	2.3.2—Complex cropland landscape
		2.3.3—Cropland with natural spaces	2.3.3.1—Cropland with natural spaces
	2.4—Plant nurseries	2.4.1—Plant nurseries	2.4.1.1—Plant nurseries

3. Thematic Comparison Framework

The thematic comparison framework detailed in this section was followed to compare two LULC products, in this case, WC20 with COS18. First, COS18 was downloaded from the official repository and was transformed to the raster data model with a 10 × 10 m^2 spatial resolution. Then, WC20 was downloaded from the official web application and transformed to the system ETRS89/PT-TM06 (EPSG: 3763) with the pixels coinciding with the generated COS18 product.

The thematic comparison framework followed to perform the thematic comparison between WC20 and COS18 is composed of four main steps:

(1) Class nomenclature harmonization. The comparison betweenWC20 and COS18 requires nomenclature harmonization between the products. Class harmonization is frequently a difficult task, as it may involve different class definitions aimed at describing different landscapes and often with a focus on either land use or land cover.
(2) A visual analysis of both maps was made for the whole country and for each of the regions A, B and C.
(3) The harmonized class area percentage was calculated for both datasets and each of the considered regions.
(4) A cross-tabulation matrix (also known as a confusion matrix when one of the maps is used as the reference) was generated for continental Portugal and for each of the regions A, B and C. This allows the computation of the overall proportion of agreement (OPA) between both maps, and the marginal proportions of agreement (MPA) per class, relative to the WC20 and COS18 classes (corresponding to the user's and producer's accuracy when considering one of the maps as the reference) [20–22]. Equation (1) was used to compute the OPA, where p_{ii} represents the number of pixels classified with the same class in WC20 and COS18 and n is the number of classes. Equations (2) and (3) were used to compute the MPA for each class at the WC20 and COS18, respectively, where p_{ij} represents the number of pixels classified with class i in WC20 and class j in COS18.

$$OPA = \frac{\sum_{i=1}^{n} p_{ii}}{Total\ number\ of\ pixels} \quad (1)$$

$$MPA(i_{WC20}) = \frac{\sum_{j=1}^{n} p_{ij}}{Number\ of\ pixels\ of\ class\ i \in WC20}. \quad (2)$$

$$MPA(j_{cos18}) = \frac{\sum_{i=1}^{n} p_{ij}}{Number\ of\ pixels\ of\ class\ j \in COS18}. \quad (3)$$

4. Results

The results regarding the thematic comparison between WC20 and COS18 are shown in this section. Hence, the section is divided into four sub-sections, each concerning a given step of the thematic comparison framework.

4.1. Class Nomenclature Harmonization

WC20 often uses the 10% minimum cover of specific characteristics (e.g., tree coverage) for an area to be assigned to a given class, while COS18 indicates a different minimum area coverage regarding each class (e.g., 25% on Grasslands and 10% on Forest areas), which will be indicated below in the detailed class nomenclature description. In this case, the nomenclature of COS18 was adapted to match that of WC20 but only considered the classes that exist in the Portuguese territory, which excludes the classes mangroves, moss and lichen and snow and ice. This was performed by consulting the user manual for both products [8,16], resulting in a nomenclature of eight classes as indicated in the left column of Table 2, mapped to the corresponding classes in WC20 and COS18.

Table 2. Nomenclature harmonization between WC20 and COS18. On the left, the nomenclature used in this study to perform the comparison.

Used Classes	WorldCover2020 (WC20)	COS18
Built Up (BU)	50. Built Up	1. Built-up, excluding: 1.6.1.1 Golf courses 1.7.1.1 Public gardens and playgrounds 1.5.1.1 Open-pit mining 1.5.1.2 Open-pit mining—rock 1.5.2.1 Landfill 1.5.2.2 Junkyards and waste dump deposits
Cropland (CL)	40. Cropland	2. Agriculture, excluding: 2.2.1.1 Vineyards 2.2.2.1 Fruit trees 2.2.3.1 Olive trees
Grassland (GL)	30. Grassland	3. Herbaceous vegetation 1.6.1.1 Golf courses 1.7.1.1 Public gardens and playgrounds
Tree Cover (TC)	10. Tree cover	4. Agroforestry 5. Forestry 2.2.2.1 Fruit trees 2.2.3.1 Olive trees
Shrublands (SL)	20. Shrublands	6. Shrublands 2.2.1. Vineyards
Bare/Sparse Vegetation (BSV)	60. Bare/Sparse vegetation	7. Open spaces with little or no vegetation 1.5.1.1 Open-pit mining 1.5.1.2 Open-pit mining—rock 1.5.2.1 Landfill 1.5.2.2 Junkyards and waste dump deposits
Wetlands (WL)	90. Herbaceous wetlands	8. Wetlands
Permanent Water Bodies (PW)	80. Permanent water bodies	9. Water bodies

Below is a detailed analysis of the nomenclatures' similarities and differences for each class and dataset present in Table 2, and how they were considered in this study:

1. Built up (BU): WC20 does not consider golf courses, green spaces, sport facilities, waste dump deposits, open-pit mining sites and landfills in the built-up class. Hence, these COS18 classes were assigned to the appropriate land cover classes (e.g., COS18 Public gardens and playgrounds were assigned to class GL, while open-pit mining areas were assigned to class BSV).
2. Cropland (CL): WC20 considers CL as areas of land covered with cropland that is sowed/planted and harvestable at least once a year. COS18 considers in this class not only annual crops but also plantations (e.g., fruit trees) and heterogeneous areas which have both permanent and annual crops (e.g., fruit trees where underneath the canopy there are croplands). Hence, the plantations were mapped to the land cover class that is most appropriate for each type of plantation (e.g., vineyards were assigned to class SL while fruit trees were assigned to class TC). The same approach was used for the heterogeneous areas included in the COS18 class "Agriculture", which were mapped to the most similar WC20 class. It should be noted that COS18 also classifies tree and plants nurseries in the class "Agriculture", which also includes green houses. However, there is no distinction between covered and not covered nurseries hence they were all assigned to the class CL.
3. Grassland (GL): WC20 classifies areas which are predominantly covered by natural herbaceous plants (at least 10%) in this class, irrespective of whether this is natural or was achieved by human intervention. Likewise, COS18 also indicates this in their user manual. However, in the case of COS18 the area covered by grassland must be larger than 25% of the mapped area. As mentioned above, golf courses, gardens and playgrounds were also included in this class.
4. Tree cover (TC): WC20 classifies any area with more than 10% tree cover in this class, even if other classes might be present underneath the canopy. As mentioned above, this also includes plantations, such as fruit tree plantations. This is different in COS2018, which classifies tree plantations into the agricultural class. Hence, as mentioned in Table 2, these tree plantations sub-classes were mapped to the TC. Another detail regarding COS18 is the class 'Agroforestry areas', which consists of areas where there are annual crops or grasslands mixed with forest species, the latter having a minimum degree of cover of 10%. This is very specific to the Portuguese forest landscape regarding, for example, cork oak stands. Given this, these COS18 classes were classified in the TC class.
5. Shrublands (SL). WC20 and COS18 have similar definitions for this class, mainly composed of natural shrubs. Shrubs are defined as woody perennial plants with persistent and woody steams not taller than 5 m [8]. However, this description also includes vineyards, which has its own subclass in COS18. Therefore, for this thematic comparison, the COS18 vineyards class was mapped to the SL class. Another difficulty is that COS18 requires at least 25% of a given area be covered with shrublands for it to be classified into that class, while WC20 only requires at least 10% coverage. These types of differences cannot be overcome when harmonizing both nomenclatures, and therefore they may cause some differences between both maps.
6. Bare/Sparse vegetation (BSV). Areas such as land with exposed soil, sand, rocks and regions that are less than 10% covered by vegetation are classified in this class in WC20. In COS18, for an area to be classified as bare/sparse vegetation it cannot be covered with more than 25% vegetation. COS18 classifies several sub-classes such as open-pit mining in the artificial surface class, which was mapped to the BSV to match with the WC20 class description.
7. Wetlands (WL). For this class WC20 classifies mainly wetlands, 10% or more of which are permanently or regularly flooded by water. WC20 excludes sediment with no vegetation, swamp forests and mangroves. In COS18, the corresponding class may also include sediment with no vegetation.

8. Permanent water bodies (PW). For this class WC20 requires that the region should be covered by water for at least 9 months in a year, while COS18 describes it as any surface of water and does not specify any required time period for this characteristic.

4.2. Visual Analysis of WC20 and COS18

Figure 2 shows the WC20 and the COS18 mapped to the same nomenclature for continental Portugal, with regions A, B and C also highlighted. The differences were evident over the country, especially between the areas occupied by classes GL and TC. It was also evident that most of the country was mapped as TC or CL in COS18, while in WC20 most of the country was mapped as GL.

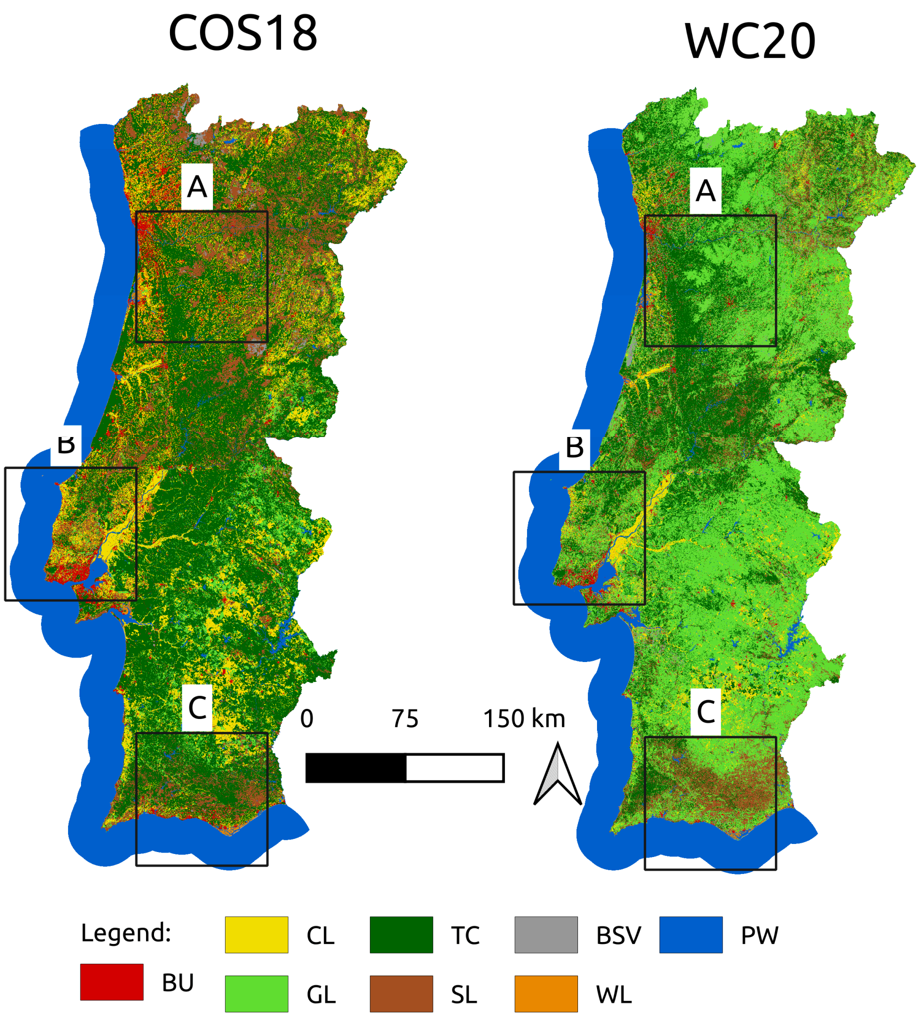


Figure 2. COS18 mapped to the WC20 nomenclature (see Section 4.1) to enable a direct comparison of both products for continental Portugal. Regions A, B and C are identified.

Figure 3 shows examples of the difference between the two products in more detail with a true color image (TCI) from Sentinel-2. Row (a) shows the regions mapped as CL in COS18 that were mapped as GL in WC20. The effect of the COS18 MMU can also be easily spotted in these images, as COS18 showed more homogenous patches while WC20 showed finer details. For example, regarding the BU class in row (b) of Figure 3, WC20 was able to identify several smaller urban green areas, which were generalized and mapped as BU in COS18 to match the MMU of 1 ha. The confusion between CL and GL was also seen in row (b) of Figure 3. Clear differences between the mapping of GL and SL are shown in row (c) of Figure 3.

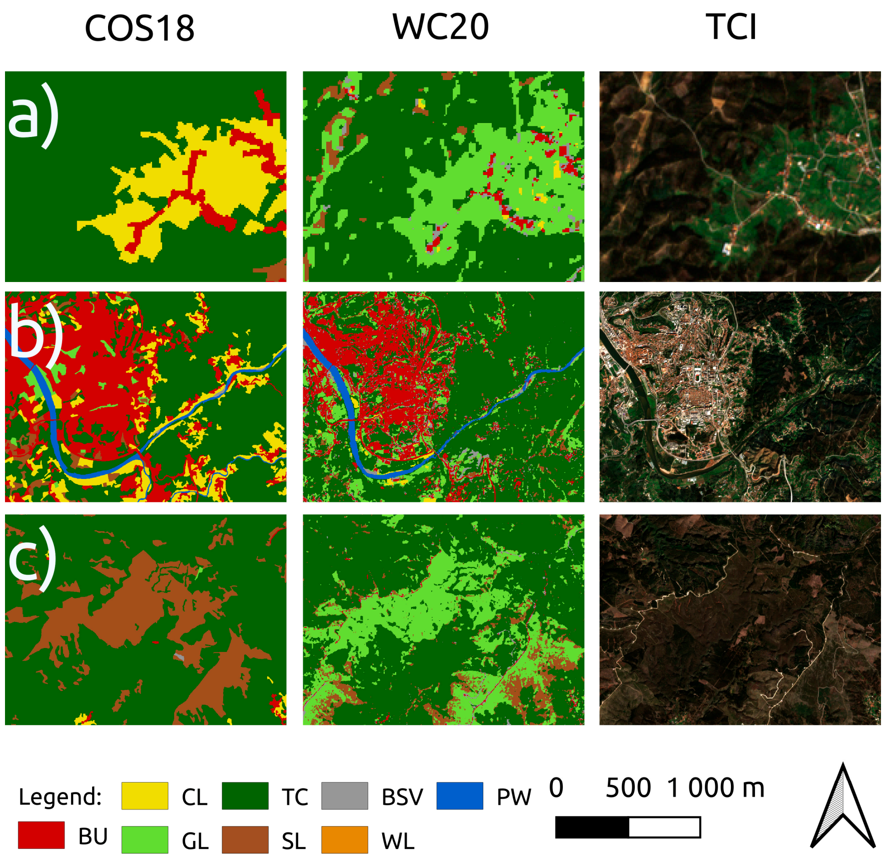


Figure 3. Details depicting three different areas (rows (**a**–**c**)) regarding COS18 (mapped to the WC20 nomenclature, see Section 4.1), WC20 and TCI (true color image, Sentinel-2).

Figure 4 shows an example of the class 'Agroforestry areas' in COS18, which were classified as TC in this study (see Section 4.1). In the center of the shown patch there was also an area assigned to class GL. Such 'Agroforestry areas' in COS18 were composed of sparse trees (minimum 10%), even if there was another LULC class between them or underneath the canopy. While COS18 mapped the homogeneous area as a whole into the 'Agroforestry areas' (in this study as TC) given its land use focus, the WC20 mapped most of the area as GL. However, the small green circles present in the TCI image were mostly related to oak trees which WC20 was not able to identify.

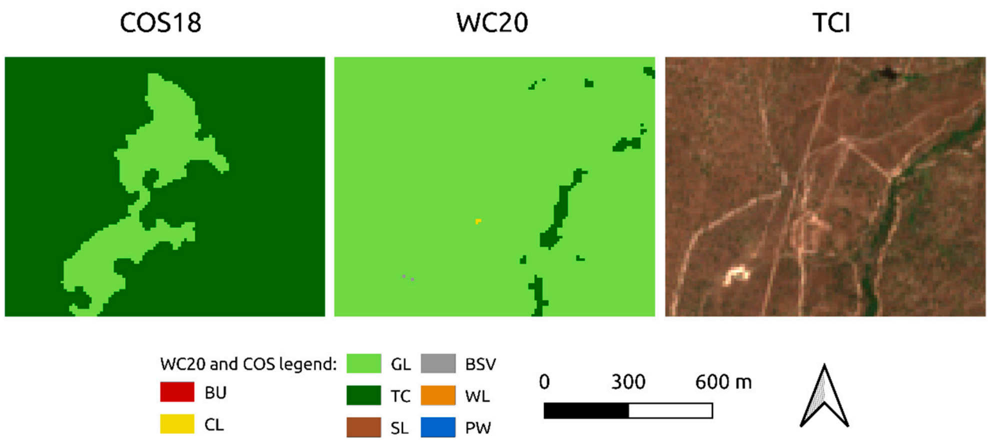
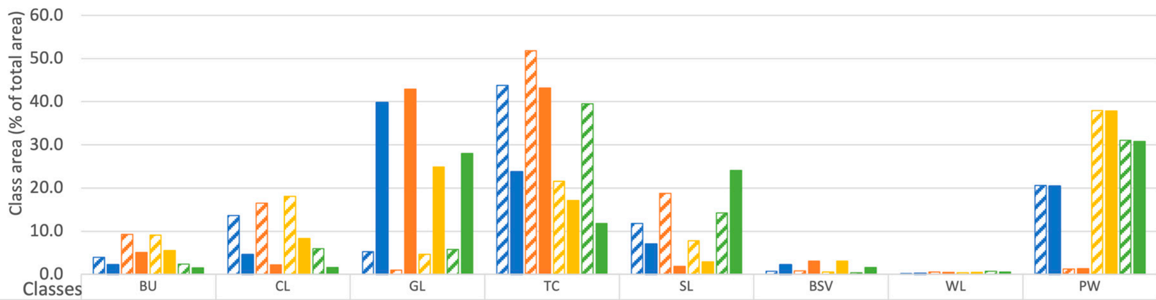


Figure 4. Example of 'Agroforestry areas' region with a central area covered in GL in COS18 and in WC20. TCI (true color image, Sentinel-2).

4.3. Class Area Comparison

The class area percentages for both datasets in each of the classified regions were computed to enable a quantitative comparison of the differences between both products, for continental Portugal and regions A, B and C. Figure 5 shows the percentages of class area coverage per class in both datasets for the three study areas, and Figure 6 shows the differences of class area percentages between COS18 and WC20.



Classes	BU	CL	GL	TC	SL	BSV	WL	PW
COS18 (PT Region)	4.0	13.7	5.2	43.7	11.8	0.7	0.2	20.6
WC20 (PT Region)	2.2	4.5	39.8	23.7	7.0	2.2	0.2	20.4
COS18 (Region A)	9.3	16.5	1.0	51.8	18.7	0.8	0.6	1.2
WC20 (Region A)	5.1	2.2	42.9	43.2	1.9	3.0	0.4	1.4
COS18 (Region B)	9.1	18.0	4.7	21.6	7.8	0.6	0.4	37.9
WC20 (Region B)	5.5	8.3	24.8	17.1	2.9	3.0	0.4	37.8
COS18 (Region C)	2.4	6.0	5.8	39.5	14.2	0.4	0.7	31.0
WC20 (Region C)	1.5	1.6	28.0	11.8	24.1	1.6	0.6	30.8

Figure 5. Percentage of class area coverage for WC20 and COS18 for continental Portugal (PT Region) and regions A, B and C.

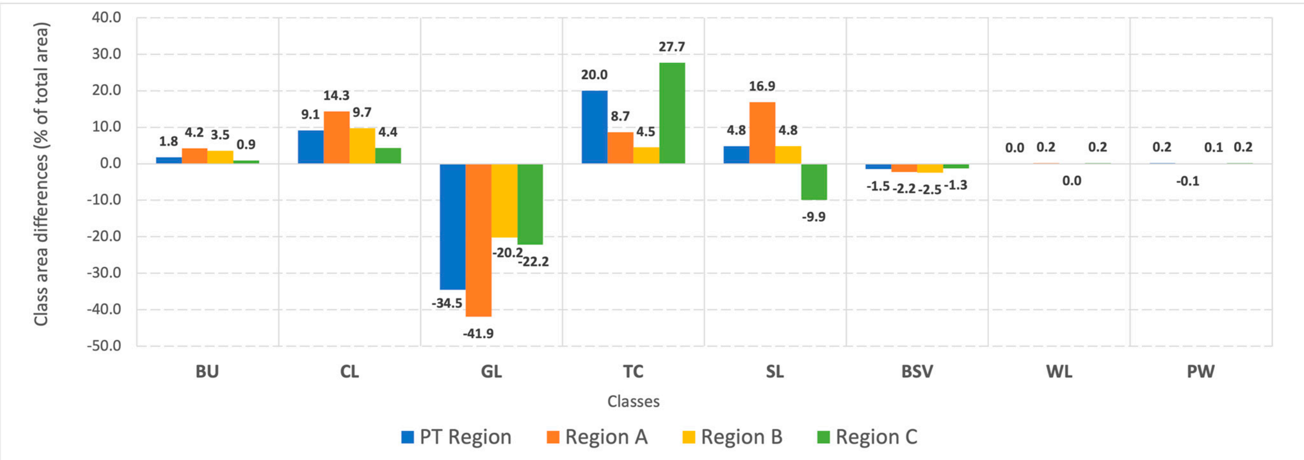


Figure 6. Differences of class area percentages between COS18 and WC20 ($COS18_{area\%} - WC20_{area\%}$) for all the study sites.

The results confirm the large area differences identified visually for some classes. The class GL presents the largest percentage of area difference between COS18 and WC20, ranging from −20.2% (for region B) to −41.9% (for region A), which confirms that GL covers a much larger area in WC20 than in COS18. The same tendency was observed for the class BSV, which also occupied larger areas in WC20 than in COS18 but with much smaller differences, varying between −1.3% (for region C) and −2.5% (for region B). The opposite tendency was observed for classes TC, CL, SL and BU. For the class TC, the differences ranged from 4.5% to 27.7%, with COS18 showing larger areas when compared to WC20. The class CL also occupied a larger area in COS18 than in WC20. Even though the differences ranged only between 4.4% and 14.3% of the total area, since this class occupies relatively small areas, these differences may correspond to several times the whole area mapped in WC20 when compared with COS18. The class SL had a higher coverage in the WC20 when compared to COS18 for region C (−9.9%), but lower for the other study areas, with the largest difference corresponding to region A (16.9%). Regarding the class BU, the differences varied between 0.9% for region C and 4.2% for region A; this class occupies relatively small areas (Figure 5) and in some cases it was close to double of the coverage in COS18 when compared with WC20. Regarding classes WL and PW, little deviations were observed in areas between both datasets in all study areas.

4.4. Proportion of Agreement between WC20 and COS18

Table 3 shows the cross-tabulation matrix obtained for continental Portugal, where the matrix cells are expressed in km^2. The matrix showed that there is a large amount of confusion between several classes. For example, the class GL in WC20 was classified in COS18 as all other classes, with most of it classified as TC but also as (in decreasing order) CL, SL, GL, BU, BSV, PW and WL. Other examples of large classification differences are: (1) the class TC in COS18 was classified in WC20 in almost equal parts as TC or GL; (2) most of the SL regions in COS18 were classified as GL in WC20, but also as TC to a large extent; and (3) most of the CL in COS18 was classified as GL in WC20.

Table 3. Cross-tabulation matrix obtained for continental Portugal. The p_{ij} values in the cells represent the area, in km^2, of class i in WC20 and class j in COS18.

		COS18							
	Classes	**BU**	**CL**	**GL**	**TC**	**SL**	**BSV**	**WL**	**PW**
WC20	BU	1971	189	30	147	58	26	1	4
	CL	64	3743	324	620	236	3	1	15
	GL	1038	8890	5013	20,402	7759	509	18	196
	TC	514	1543	180	21,763	2065	26	10	65
	SL	55	277	140	4569	2590	16	12	11
	BSV	730	354	79	705	297	178	8	42
	WL	1	55	11	1	2	1	123	35
	PW	8	8	6	25	7	27	93	22,358

The results obtained for the OPA are shown in Table 4, while the MPA_{WC20} and MPA_{COS18} per class for the four regions under analysis are depicted in Figure 7.

Regarding the OPA, a value of only 52% was obtained for continental Portugal, while for regions, A, B and C it was equal to 45%, 68% and 56%, respectively. These low values show that there was no class agreement in about half of the extent of these two datasets for continental Portugal when considering the nomenclature harmonization performed for the study. In addition, there was also a difference in the OPA per region that achieved 23% for the classified regions (namely between regions A and B). However, this was expected to happen due to the different proportions of area of the several classes in the four classified regions. For example, regions B and C include a large proportion of PW, which is the class with largest agreement between both datasets (the class with higher MPA_{WC20} and MPA_{COS18}). It can even be seen that there was a positive correlation between the OPA of the study areas and the proportion of PW in each region (higher values for region B, then region C, then PT region and finally region A).

Table 4. Overall proportion of agreement (OPA) for continental Portugal and the three sub-regions A, B and C.

Continental Portugal	Region A	Region B	Region C
52%	45%	68%	56%

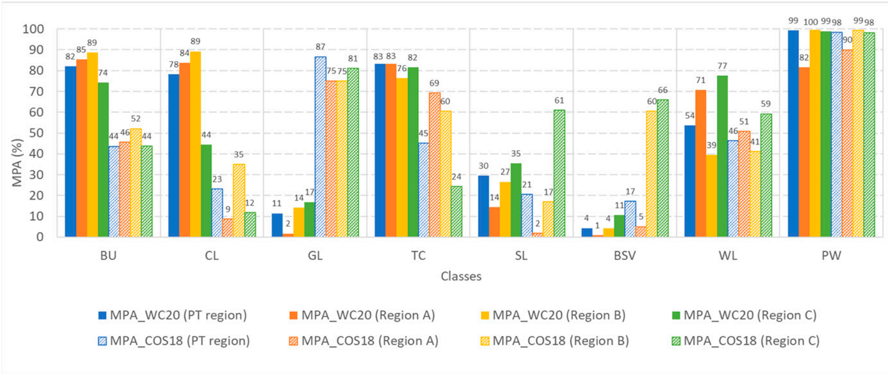


Figure 7. The MPA_{WC20} (solid fill) and the MPA_{COS18} (striped fill) for continental Portugal (PT) (blue color) and the three sub-regions A, B and C (represented, respectively, by the orange, yellow and green colors) plotted for each class.

The MPA_{WC20} (Figure 7), which enables the identification of classes' commissions in the WC20 when compared to COS18 (given by 100%—MPA_{WC20}), varied between 1% (for class BSV in region A) and 100% (for class PW in region B). The MPA_{WC20} for the regions A, B and C seemed to follow the behavior obtained for continental Portugal. However, larger differences were found for some classes, such as for class CL, in which the MPA_{WC20} in region C was only 44% while for Portugal (PT) it was 78%. This may be explained by regional differences regarding, for example, the types of crops and/or natural vegetation or their typical layout, or even the different topographic characteristics of the selected regions. Equally large differences were found in the class WL, where for region B, MPA_{WC20} was 39% while for region C it was 77% (a difference of 38%). This shows that, in region B, 61% (100% − 39%) of the regions classified as WL in WC20 where not mapped as such in COS18. Another aspect that stands out from Figure 7 are the very low values obtained for classes BSV, GL and SL, which means these classes have large commissions in WC20 when compared to COS18.

The MPA_{COS18} per class enables the identification of the classes' omissions in WC20 when compared with COS18 (given by a MPA_{COS18} of 100%). It varied between 2% (for class SL in region A) and 99% (for class PW in region B) (see Figure 7). Several classes closely followed the MPA_{COS18} obtained for continental Portugal, such as BU, GL and PW. However, larger regional differences were obtained for some classes, such as for the class CL, which presented a variation of MPA_{COS18} between 9% and 35%, where regions A and C presented lower values and region B presented a value 12% higher than the one obtained for the PT region. The class SL showed, for region C, a MPA_{COS18} of 61%, whereas for region A it had a value of only 2% and 17% for region B. The class TC also presented large MPA_{COS18} variations between regions, where region C was 24%, while region A was 69%. This shows that there is regional variation of the agreement between both datasets, which may be dependent on the characteristics of the regions and types of existing vegetation. For the PT region, the MPA_{COS18} for TC was only 45%, which means there was 55% omissions in this class when compared with COS18. This is particularly relevant because this class occupies a large proportion of the country. Large variations of the MPA_{COS18} per region were also observed for the class BSV, where similarly lower values were obtained for A and PT regions, while it presented 60% and 66%, respectively, for regions B and C. The very different orders of magnitude of MPA_{COS18} for some classes also stood out from Figure 7. Very high values were obtained for classes PW and GL, which means there were few omissions, and much lower values for classes CL and SL (except for region C), which correspond to large omissions when compared to COS18. MPA_{COS18} of the class BU are also relatively low (around 50% for the four regions), which corresponds to omissions of about 50% when compared with COS18.

When analyzing the differences between MPA_{WC20} and MPA_{COS18} per class, in both class CL and class BU, the MPA_{WC20} had much larger values than MPA_{COS18}, which indicates an underestimation for these classes in WC20, as they showed few commissions and large omissions when compared with COS18. The opposite occurred with the class GL, which had much larger values of MPA_{COS18} than MPA_{WC20} (for the PT region, at 87% and 11%, respectively), which correspond to a large overestimation of this class when compared with COS18.

4.5. Comparison of the Agreement between WC20 and COS18 and WC20 Accuracy Indices Reported for Europe

The WC20 OPA obtained in the analysis presented in this paper for continental Portugal (52%) is 24% lower than the overall accuracy obtained for EU (76%).

Table 5 shows a comparison between the per class MPA_{WC20} for Portugal with the user's accuracy (UA) obtained for the EU, as well as the per class comparison between the MPA_{COS18} and the producer's accuracy (PA).

Table 5. WC20 user's and producer's accuracy per class reported for Europe (UA (EU) and PA (EU), respectively), and the MPA for WC20 when compared to COS18 and vice-versa, obtained for continental Portugal (MPA_{WC20}(PT) and MPA_{COS18}(PT), respectively), along with the differences between the comparable values when COS18 is used as the reference.

Classes	UA (EU)	MPA_{WC20} (PT)	UA (EU)-MPA_{WC20} (PT)	PA (EU)	MPA_{COS18} (PT)	PA (EU)-MPA_{COS18} (PT)
BU	86	82	4	64	44	20
CL	90	78	12	81	23	58
GL	66	11	55	67	87	−20
TC	80	83	−3	94	45	49
SL	33	30	3	13	21	−8
BSV	35	4	31	28	17	11
WL	37	54	−17	41	46	−5
PW	96	99	−3	90	98	−8

The classes with higher agreement between MPA_{WC20} and UA were the classes TC, SL and PW (with differences of only 3%) and BU (with a difference of 4%), while the classes with larger differences were GL (a difference of 55%) and BSV (a difference of 31%). In addition, the classes with lower UA in the EU were SL, BSV and WL, while the classes with lower MPA_{WC20} in Portugal were BSL (with only 4%), GL and SL. Therefore, if the mapping of COS18 to the classified nomenclature is considered to reflect the ground truth for the considered classes, there were much larger commission errors for the PT in classes BSV and GL than for the EU, while the opposite occurred for the class WL (HW in WC20) (a difference of 17%).

When comparing the MPA_{COS18} with the PA obtained for the EU (Table 5), it can be seen that the classes with lower differences were classes WL (5%), SL (8%) and PW (8%), while the classes with larger differences were CL (58%) and TC (49%), and differences of 20% were obtained for classes BU (with PA > MPA_{COS18}) and GL (with PA < MPA_{COS18}).

5. Discussion

In the class nomenclature harmonization between the two datasets, several decisions were made in a bid to make the nomenclatures comparable. However, several issues persisted. For example, while the built-up class was the one of the less problematic regarding class harmonization, there may be, for example, non-vegetated playgrounds mapped in COS18 that will be erroneously mapped to GL, given that COS18 does not distinguish urban green areas from other types of playground areas. Moreover, COS18 contains several classes of built-up with different built-up densities, which combined with its MMU, is capturing green urban spaces in these classes, while WC20 will classify these as GL. The COS18 class 'Agroforestry areas', in spite of being classified in the TC class for the comparison, may contain large patches of GL between the trees in sparsely forested areas. The lack of harmonization between the nomenclatures is partly related to the focus of the maps (either land use or land cover) and the disparate MMU. These differences are then translated into both the computation of statistical metrics through the overall and marginal proportion of agreement and the comparison of each class area in each dataset [23].

Regarding the comparison metrics, the WC20 classes showing a better agreement with COS18 were classes WL and PW both presenting overall higher MPA derived from the cross-tabulation matrices. This agreement may be due to the lower ambiguity between land use and land cover for water-related classes. Nonetheless, the class WL in this study presenteda higher MPA when considering the accuracy results reported for WC20 in Europe. In both the BU and TC classes, the results seem to indicate an underestimation for these classes in WC20. This is also seen in the percentage of the area occupied by these classes in both datasets, where the COS18 coverage was close to twofold higher when considering classes BU and TC (Figure 6). These results differ from the ones reported for WC20 in Europe, especially in the TC class (see Section 4.4), where higher values of PA were reported for these classes. However, the large difference between the maps regarding TC may be due to a combined effect from the larger MMU of COS18 and its land use focus. Regarding class CL, it seems to have been underestimated in WC20, as shown by the MPA. The area coverage was also around three-fold less in WC20 when compared to COS18. Moreover, in Figure 3 the underestimation for the CL class by WC20 was also visible. The CL MPA between COS18 and WC20 also differed considerably when comparing with the reported PA for Europe, from 81% in the report to 23% in this study (Table 5). Most likely, COS18 includes areas of agricultural land use in class CL even if it is not evident from phenological spectral reflectance, which is a necessary condition for an automatic classification.

The classes showing the worst agreement between COS18 and WC20, when it comes to the statistical metrics, were classes GL, SL and BSV. The area coverage difference in class GL and the results of the computation of the proportion of agreement indicated an overestimation of this class in WC20. Moreover, from Figures 1 and 3, GL class overestimation is visible, in that the CL class was often confused with GL and sparse/individual trees were also classified as GL. Such results also differed from the ones reported by the WC20 validation report for that class, especially when considering UA, which for the European case was 66% and when compared with the proportion of agreement between WC20 and COS18 only 11% was obtained. Regarding the class SL, it presented a lower MPA and was often confused by WC20 as class GL. However, when looking at the results on the validation report for Europe (Table 5), even the PA for Portugal was improved. The class BSV already presented a low UA and PA for Europe which was also the case when comparing with the results of the proportion of agreement between WC20 and COS18. The noted confusions are common in automated image classification procedures due to similar spectral signatures of non-tree vegetated cover such as grasslands, croplands and shrublands, or due to the difficulty of detecting isolated trees in images with insufficient spatial resolution. In such cases, COS18 should be more accurate as it relies on visual interpretation of orthophotomaps.

Overall, the different minimum mapping unit between the products, focus on either land use or land cover, class nomenclature harmonization between a global and a national product are then translated into thematic differences which are not due to actual differences (on the ground) in land use or land cover. However, the magnitude of the differences reported here cannot be explained by the different characteristics of the products alone. For example, the overestimation of GL, wrongly classified as GL when actually there is TC or CL, can be seen in WC20. An underestimation of TC by WC20 was also identified, namely in the southern landscape which contained forests composed of sparse trees and which WC20 was often not able to identify. In fact, this type of landscape is common in the Iberian peninsula, having its own level-3 class in the CORINE Land Cover nomenclature [24] and even being classified in a dedicated classification procedure [15]. WC20, with its global coverage, fails to represent these types of landscape by often missing small evidences of tree cover and erroneously classifying all these landscape as GL. The results presented here are in line with the ones reported in previous studies, where southern and Mediterranean countries also presented lower accuracy values when compared with other regions using global or continental LULC products [13–15,25]. While falling short in detecting region-specific landscapes, WC20 in an urban setting is able, for example, to distinguish small green urban areas or small sets of CL which was not mapped in COS18 either due to its focus on land use or its MMU of 1 ha. This study presented an overall assessment of the products regarding their thematic content. However, for a more specific study within a given application, other approaches might be followed such as in [26] where a weight is given to each class according to user defined criteria which will then derive a score for each dataset.

6. Conclusions

This paper presented a thematic comparison between the global land cover product WC20 and the COS18 product, the latter being an official LULC map produced by the Portuguese National Mapping Agency. The comparison area was continental Portugal and three sub-regions with different landscape characteristics. The challenging nomenclature harmonization of the two data sources was performed to enable the comparison and to minimize the differences in the classes' definitions. Once both products were mapped to a common nomenclature, the comparison included: a visual assessment of the differences and a comparison of the proportion of class area in both products for the four regions under analysis and the computation of cross-tabulation matrices, from which the OPA and the MPA per class were computed for continental Portugal and for the three smaller regions.

The main conclusions that can be extracted from the results indicate that large differences were found for the classes BSV, SL, GL, TC and CL. However, the ones with larger impacts on the product's overall agreement are classes GL (largely overestimated in WC20 when compared with COS18) and TC (with large omissions when compared with COS18), due to the large proportion of the territory they occupy. There was also agreement between both products for some classes that are less dependent on land use and land cover concepts, e.g., the class PW.

Regarding the two classes more related to human activities, namely BU and CL, they both appeared to be underestimated in WC20 when compared to COS18. The BU class, while presenting a lack of mapped areas in WC20 when compared with COS18, it was certainly affected by the MMU of 1 ha of COS18, given the heterogeneity in land cover, especially in urbanized areas. Hence, WC20 presents more detail within urbanized areas even with a focus on land cover. The TC class omissions by WC20 were also impacted by the MMU of COS18, where WC20 can distinguish between canopies and the cover below due its resolution. However, WC20 was often not able to identify single or small patches of trees and instead classified only the land cover below the canopies. These types of difficulties may turn out to be especially relevant in the Portuguese landscape where there are large patches of land composed of sparse/individual trees. This landscape pattern is already classified in a dedicated class in the Corine Land Cover nomenclature, while the literature presents dedicated classification procedures to identify them in the Iberian Peninsula. Hence, we can expect it to perform similarly in countries with similar landscapes.

When comparing the obtained MPAs between WC20 and COS18 (considering COS18 as reference data) with the WC20 UA and PA reported for Europe, similar values were obtained for some classes, such as SL, and better values were obtained, for example, for PW and WL. However, very large differences were found for the classes showing less agreement with COS18, namely the classes GL and CL, which may indicate that there are different levels of accuracy for Portugal than for the whole of Europe, which was already noted in the literature. At the same time, the levels of accuracy were also impacted by the different characteristics of WC20 and COS18. Map comparisons have many limitations, which, for thedatasets used in the comparison, include class nomenclature harmonization given different country-specific landscape characteristics, vegetation types and species, and the land use nature of COS18; hence, many subclasses of COS18 were hard to attribute to a single cover. Another issue regarding the comparison is the necessary coordinate transformation of the WC20 product to the COS18 coordinate system, which may be introducing positional errors that may then be translated into thematic errors. Moreover, the transformation of the vector COS18 format to raster also introduces errors. However, given the large differences found, the authors believe that these last sources of error did not have a significant influence on the obtained results. The reference year was also different and while LULC differences are expected to exist between the two dates, these were considered negligible given the extent of the comparison.

In spite of the challenging comparison between COS18 and WC20 due to different map characteristics such as different nomenclature, focus on either land use or land cover and MMU, map comparisons as performed in this study are useful given the decameter resolution of recent global LULC datasets that are only validated globally or continentally. This is even more relevant given the expected increase in both global and national LULC maps availability, produced by different map producers and with different applications in mind. Therefore, users should be aware of the characteristics and purposes of the maps to select which map to use according to their needs.

This analysis reported the results of a comparison between two products that may be used for applications that require a high level of detail. However, the large differences detected in some classes require the application of validation methodologies based on reference data that are not affected by the differences in the products specifications, such as the differences in classes definition and the application of an MMU. These differences highlight the importance of product comparison and validation, as when using different products for the same types of analysis, different conclusions may be drawn.

Author Contributions: Conceptualization, D.D., C.F., H.C. and M.C.; methodology, D.D. and C.F. software, D.D.; validation, C.F.; formal analysis, D.D., C.F., H.C. and M.C.; investigation, D.D. and C.F.; resources, D.D., C.F., H.C. and M.C.; data curation, D.D. and C.F.; writing—original draft preparation, D.D., C.F., H.C. and M.C.; writing—review and editing, D.D., C.F., H.C. and M.C.; visualization, D.D. and C.F.; supervision, C.F., H.C. and M.C.; project administration, D.D., C.F., H.C. and M.C.; funding acquisition, C.F., H.C. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by projects foRESTER (PCIF/SSI/0102/2017), SCAPEFIRE (PCIF/MOS/0046/2017) and FireLoc (PCIF/MPG/0128/2017), by Centro de Investigação em Gestão de Informação (MagIC) and grant UIDB/00308/2020, all funded by the Portuguese Foundation for Science and Technology (FCT). It was also supported by Compete2020 (POCI-05-5762-FSE-000368), funded by the European Social Fund.

Data Availability Statement: Publicly available datasets were analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bossard, M.; Feranec, J.; Otahel, J. *CORINE Land Cover Technical Guide: Addendum 2000*; European Environment Agency: Copenhagen, Denmark, 2000.
2. Buchhorn, M.; Lesiv, M.; Tsendbazar, N.E.; Herold, M.; Bertels, L.; Smets, B. Copernicus Global Land Cover Layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [CrossRef]
3. Chaaban, F.; El Khattabi, J.; Darwishe, H. Accuracy Assessment of ESA World Cover 2020 and ESRI 2020 Land Cover Maps for a Region in Syria. *J. Geovisualiz. Spat. Anal.* **2022**, *6*, 31. [CrossRef]
4. Comber, A.; Fisher, P.; Wadsworth, R. What is Land Cover? *Environ. Plan. B Plan. Des.* **2005**, *32*, 199–209. [CrossRef]
5. Costa, H.; Benevides, P.; Marcelino, F.; Caetano, M. Introducing Automatic Satellite Image Processing into Land Cover Mapping by Photo-Interpretation of Airborne Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *42*, 29–34. [CrossRef]
6. Costa, H.; Benevides, P.; Moreira, F.D.; Moraes, D.; Caetano, M. Spatially Stratified and Multi-Stage Approach for National Land Cover Mapping Based on Sentinel-2 Data and Expert Knowledge. *Remote Sens.* **2022**, *14*, 1865. [CrossRef]
7. Direcção Geral do Território. *Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 2018*; Direcção Geral do Território: Lisbon, Portugal, 2019.
8. ESA. *World Cover—Product User Manual v1.0*; ESA: Paris, France, 2021.
9. ESA. *World Cover—Product Validation Report v1.1*; ESA: Paris, France, 2021.
10. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef] [PubMed]

11. Jun, C.; Ban, Y.; Li, S. Open access to Earth land-cover map. *Nature* **2014**, *514*, 434. [CrossRef] [PubMed]
12. Kidane, M.; Bezie, A.; Kesete, N.; Tolessa, T. The impact of land use and land cover (LULC) dynamics on soil erosion and sediment yield in Ethiopia. *Heliyon* **2019**, *5*, e02981. [CrossRef] [PubMed]
13. Liu, H.; Gong, P.; Wang, J.; Wang, X.; Ning, G.; Xu, B. Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020—iMap World 1.0. *Remote Sens. Environ.* **2021**, *258*, 112364. [CrossRef]
14. Malinowski, R.; Lewiński, S.; Rybicki, M.; Gromny, E.; Jenerowicz, M.; Krupiński, M.; Nowakowski, A.; Wojtkowski, C.; Krupiński, M.; Krätzschmar, E.; et al. Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sens.* **2020**, *12*, 3523. [CrossRef]
15. Marlon, J.R.; Bartlein, P.J.; Carcaillet, C.; Gavin, D.G.; Harrison, S.P.; Higuera, P.E.; Joos, F.; Power, M.J.; Prentice, I.C. Climate and human influences on global biomass burning over the past two millennia. *Nat. Geosci.* **2008**, *1*, 697–702. [CrossRef]
16. McMahon, S.M.; Harrison, S.P.; Armbruster, W.S.; Bartlein, P.J.; Beale, C.M.; Edwards, M.E.; Kattge, J.; Midgley, G.; Morin, X.; Prentice, I.C. Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. *Trends Ecol. Evol.* **2011**, *26*, 249–259. [CrossRef]
17. Moreira, F. Overview of landscape research and assessment in Portugal. *Belgeo* **2004**, *2*, 329–336. [CrossRef]
18. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
19. Pontius, R.J., Jr.; Cheuk, M.L. A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 1–30. [CrossRef]
20. Stehman, S.V. Comparing thematic maps based on map value. *Int. J. Remote Sens.* **1999**, *20*, 2347–2366. [CrossRef]
21. Tucker, C.J.; Townshend, J.R.G. Strategies for monitoring tropical deforestation using satellite data. *Int. J. Remote Sens.* **2000**, *21*, 1461–1471. [CrossRef]
22. Desa, U.N. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: San Francisco, CA, USA, 2016; Available online: https://www.un.org/pga/wp-content/uploads/sites/3/2015/08/120815_outcome-document-of-Summit-for-adoption-of-the-post-2015-development-agenda.pdf (accessed on 2 December 2022).
23. Venter, Z.S.; Sydenham, M.A.K. Continental-Scale Land Cover Mapping at 10 m Resolution Over Europe (ELC10). *Remote Sens.* **2021**, *13*, 2301. [CrossRef]
24. Wickham, J.; Stehman, S.V.; Homer, C.G. Spatial patterns of the United States National Land Cover Dataset (NLCD) land-cover change thematic accuracy (2001–2011). *Int. J. Remote Sens.* **2018**, *39*, 1729–1743. [CrossRef] [PubMed]
25. Yang, Q.; Zhang, H.; Peng, W.; Lan, Y.; Luo, S.; Shao, J.; Chen, D.; Wang, G. Assessing climate impact on forest cover in areas undergoing substantial land cover change using Landsat imagery. *Sci. Total Environ.* **2019**, *659*, 732–745. [CrossRef] [PubMed]
26. Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. ESA WorldCover 10 m 2020 v100. Zenodo. Available online: https://doi.org/10.5281/ZENODO.5571936 (accessed on 2 December 2022). [CrossRef]