Open land cover maps - validation from the user perspective

Kovačić, Leona

Leona Kovačić ◆ DIPLOMSKI RAD ◆2025.

SVEUČILIŠTE U ZAGREBU
GEODETSKI FAKULTET

Leona Kovačić

OPEN LAND COVER MAPS – VALIDATION FROM THE USER PERSPECTIVE

Diplomski rad

Zagreb, 2025.

SVEUČILIŠTE U ZAGREBU
GEODETSKI FAKULTET

Leona Kovačić

OPEN LAND COVER MAPS – VALIDATION FROM THE USER PERSPECTIVE

Diplomski rad

Zagreb, 2025.

SVEUČILIŠTE U ZAGREBU
GEODETSKI FAKULTET

Na temelju članka 19. Etičkog kodeksa Sveučilišta u Zagrebu i Odluke br. 1_349_11 Fakultetskog vijeća Geodetskog fakulteta Sveučilišta u Zagrebu, od 26.10.2017. godine (klasa: 643-03/16-07/03), uređena je obaveza davanja „Izjave o izvornosti“ diplomskog rada koji se vrednuju na diplomskom studiju geodezije i geoinformatike, a u svrhu potvrđivanja da je rad izvorni rezultat rada studenata te da taj rad ne sadržava druge izvore osim onih koji su u njima navedeni.

IZJAVLJUJEM

Ja, **Leona Kovačić**, (JMBAG: 0083227087), rođena dana 13.11.2000. u Splitu, izjavljujem da je moj diplomski rad izvorni rezultat mojeg rada te da se u izradi tog rada nisam koristio drugim izvorima osim onih koji su u njemu navedeni.

U Zagrebu, dana ____________

____

Potpis studenta / studentice

I. AUTOR	
Ime i prezime:	Leona Kovačić
Datum i mjesto rođenja:	13. studenog 2000., Split, Republika Hrvatska
II. DIPLOMSKI RAD	
Naslov:	Open land cover maps - Validation from the user perspective
Broj stranica:	61
Broj tablica:	7
Broj slika:	25
Broj bibliografskih podataka:	20 + 7 URL
Ustanova i mjesto gdje je rad izrađen:	Geodetski fakultet Sveučilišta u Zagrebu
Mentor:	izv. prof. dr. sc. Mario Miler
Komentor:	izv. prof. dr. sc. Martina Baučić
Voditelj:	dr. sc. Dino Dobrinić
III. OCJENA I OBRANA	
Datum zadavanja teme:	15.01.2024.
Datum obrane rada:	21.02.2025.
Sastav povjerenstva pred kojim je branjen diplomski rad:	izv. prof. dr. sc. Mario Miler
	doc. prof. dr. sc. Luka Rumora
	prof. dr. sc. Damir Medak

Veliko hvala mom mentoru prof. Mariju Mileru zajedno s voditeljima na prenesenom znanju i ukazanom povjerenju. Iskrenu zahvalnost želim izraziti svojoj komentorici prof. Martini Baučić na strpljivom vodstvu, nesebičnoj podršci i ohrabrujućoj motivaciji, ne samo tijekom izrada ovog rada, već tijekom cjelokupnog studiranja. Zahvaljujem svim profesorima i asistentima Fakulteta građevinarstva, arhitekture i geodezije u Splitu te Geodetskog fakulteta u Zagrebu. Čast mi je bila studirati na ovim institucijama.

Konačno, posebno zahvaljujem svojoj obitelji i prijateljima na bezuvjetnoj ljubavi, strpljenju i vjeri u mene.

Otvorene karte zemljišnog pokrova - metode validacije iz korisničke perspektive
***Sažetak*:**

Validacija podataka od ključne je važnosti kako bi se osiguralo da podaci o zemljišnom pokrovu odgovaraju području primjene. Ovaj rad korisnicima omogućuje donošenje informiranih odluka temeljenih na prikladnosti otvorenih karata zemljišnog pokrova za njihove specifične potrebe. Proces validacije, koji uključuje pragmatične korake i kvantitativne izračune, korisnicima pruža alate za procjenu točnosti podataka i nesigurnosti koje mogu utjecati na njihove analize. Pragmatični koraci nude korisnicima jasan pristup za procjenu podataka, uključujući proučavanje izvještaja o validaciji od strane proizvođača podataka i vizualnu inspekciju rezultata usporedbom s podacima veće kvalitete. S druge strane, kvantitativni izračuni, uključujući izračun površine pokrivene svakom klasom zemljišta i evaluaciju pomoću metodologije oblika, omogućuju objektivnu i mjerljivu procjenu. Ovi izračuni pomažu korisnicima razumjeti ne samo točnost podataka, već i širenje nesigurnosti od ulaznih podataka do rezultata.

Ključne riječi: *zemljišni pokrov; validacija; kvaliteta podataka; GIS; ESA WorldCover; GLC_FCS30D; ESRI Sentinel-2 Land Cover*

Open land cover maps - validation from the user perspective

***Abstract*:**

The ability to effectively validate and assess land cover data is essential for ensuring its relevance and reliability in real-world applications. By providing users with a structured, practical approach to data validation, this work empowers them to make well-informed decisions based on the suitability of open land cover maps for their specific needs. The validation process, which includes both pragmatic steps and quantitative calculations, offers users the tools to evaluate the data's accuracy and the uncertainties that may affect their analyses. Pragmatic steps provide users with a clear approach to evaluating the data based on their context, such as assessing the validation report by data producer or visual inspection of the results. Meanwhile, the quantitative calculations allow for a more objective, measurable assessment, which helps to pinpoint areas of uncertainty and discrepancies in the data. These calculations, including the areas covered by each land cover class and the evaluation using shape metrics, help users understand not only the accuracy of the data but also the propagation of uncertainties from input data sets to results.

Keywords: *land cover; validation; data quality; GIS; ESA WorldCover; GLC_FCS30D; ESRI Sentinel-2 Land Cover*

CONTENT

INTRODUCTION 1
1.1 Overview of global land cover maps 3
1.1.1 COPERNICUS GLOBAL LAND SERVICES LAND COVER MAP AT 100 m (CGLS-LC100) 6
1.1.2 ESA WorldCover 7
1.1.3 CORINE Land Cover 9
1.1.4 Dynamic World 11
1.1.5 ESRI Sentinel-2 Land Cover 13
1.1.6 GLOBAL LAND COVER WITH FINE CLASSIFICATION SYSTEM AT 30 m (GLC_FCS30D) 14
1.2 Main application areas of land cover maps 17
1.2.1 Agricultural monitoring 18
1.2.2 Land degradation 20
1.2.3 Sustainable urban development 21
1.2.4 Water resources management 22
1.2.5 Climate change adaptation 23
MATERIALS AND METHODS 25
2.1 Validation and uncertainty estimation 25
2.1.1 ESA WorldCover 2021 validation method example 27
2.2 Proposal of the validation method 29
2.2.1 Validation of input data 30
2.2.2 Validation of GIS operations 32
2.2.3 Uncertainties in area calculations 35
2.3 Method testing on the study case 40
2.3.1 Study area 41
2.3.2 Visual inspection 43
2.3.3 Area calculation 48
2.3.4 Validation of GIS overlay 49
RESULTS 50
3.1 Results of area calculation 50
3.2 Results of Shape Index calculation 54
DISCUSION 55
CONCLUSION 60
REFERENCES 62

LIST OF URLS 64
FIGURES 65
TABLES 66
ANNEX 67

INTRODUCTION

Land cover refers to the Earth’s terrestrial surface covered by natural (grass, trees, water, ice, bare ground, soil) and human structures. Modifications of such cover means replacement of one cover type by another, for example, with urbanization of coastal zones thanks to growth of tourism and hospitality activities, natural coastal areas change from grass and soil to urban structures and buildings. Changes in the urbanization sprawl affect features of land and its inherent properties and often lead to irreversible alterations of the environment. These modifications influence various aspects of the environment including biodiversity, soil health, water runoff, material and energy cycles. Such changes are primarily driven by human activities aimed at achieving specific goals, such as economic growth or resource management. Human activities exploiting the Earth’s surface led to alterations in land cover, which affect its physical and biological characteristics. Therefore, land use changes, caused by how humans utilize the land, cause changes in land cover (Das, 2014).

While closely related, the terms land use and land cover refer to quite different dimensions of the Earth's surface. Land use refers to the various activity humans perform on land to extract resources or obtain benefits, such as agriculture, construction or forestry. It involves the management and manipulation of land for specific purposes. Land cover commonly describes the physical characteristics of the Earth's surface defined as the vegetation (natural or planted) or man-made constructions (buildings, roads etc.) such as water, ice, bare rock, grass and sand. In short, land use indicates how people are using the land, whereas land cover indicates the physical land type (Das 2014). Although, land use class influences land cover type, changes in one do not imply change in other or vice versa. In other words, having the same land cover in two different locations does not imply the same land use in both of those. Similarly, same land uses do not imply identical land covers. Directive INSPIRE (2007/2/EC) of the European Parliament defines land cover as physical and biological cover of the Earth’s surface including artificial surfaces, agricultural areas, forests, seminatural areas, wetlands, water bodies. Land use is defined as territory characterized according to its current and future planned functional dimension or socio-economic purpose (e.g., residential, industrial, commercial, agricultural, forestry, recreational).

The Food and Agriculture Organization (FAO) of the United Nations plays a crucial role in providing global insights into agriculture, forestry, food security and land use/land cover. The FAO is responsible for collecting and sharing valuable information related to these areas. As defined by the FAO, land use refers to the activities, arrangements and resources that people apply to a particular land cover type in order to produce, modify or sustain it. This definition of land use creates a clear connection between land cover and human activities, highlighting how people's actions shape and interact with the environment around them. The Land Cover Classification System (LCCS), developed by the FAO, was created to provide a consistent framework for classifying and mapping land cover. Its main goal was to overcome the limitations of predefined land cover categories, which often don't fit well with real-world situations and make accurate classification challenging. By introducing a more standardized approach, the LCCS allows for better mapping and classification of land cover in practice and is often used by today's land cover products (URL 1).

Frequent updates to global land use and land cover (LULC) datasets are crucial for gaining insights into the condition, trends and pressures exerted by human activities on biodiversity, ecological and anthropogenic processes. These datasets serve as a foundation for monitoring the dynamics of the environment and understanding the broader implications of land use changes. In general, land cover changes can be grouped into three categories: periodic changes caused by phenological variability, trend changes driven by natural behavior (such as vegetation growth) and abrupt changes caused by natural or human disturbances (such as deforestation or urban expansion) (Zhang et al., 2024).

Land cover is a complex concept, making it a highly comprehensive product to evaluate. Nowadays, there is a wide diversity of land cover products available, thanks to advancements in remote sensing technologies, satellite imagery and geographic information systems (GIS). These products often vary in spatial resolution, classification systems and temporal coverage, offering different levels of detail to suit diverse research and management needs. They provide valuable insights into how land cover is changing over time and can be used for purposes like environmental monitoring, land management and policy planning. Land cover maps are not just datasets, they are tools for decision making and shaping policies worldwide. The problems facing our and future generations do not know political borders. From the pressing issues of climate change to the alarming pollution affecting our health, Earth is inevitably changing every day and humankind must respond with integrated solutions. The importance of land cover products is clear, but a key question that arises is: with so many open-access options available, how can users choose the right product for their needs?

Selection of appropriate land cover has to satisfy the need to successfully detect and track impervious (built-up) areas. This task raises the question of how to choose a suitable land cover product for specific purposes generally, not just for determining urban areas. Land cover products are, as said, very useful and powerful tools for numerous applications and industry sectors, from decision makers to land planning policies, therefore there ought to be a method users can conduct to be sure which source is best suited for their needs. In general, validation methods are given and carried out by producers or third-party validation organizations. These validations are usually challenging to understand for the end user, leaving him in doubt when it comes to land cover products selection. Beside validation done by the data producers, users should report about validation and uncertainty estimation for the results of performed analysis that uses open land cover maps. Thus, limitation of the results should be clear to policy decision makers.

Main objective of this thesis is to provide a simple yet effective method for users to evaluate the appropriateness of open land cover data for their specific needs. Moreover, once the data is selected, understanding the level of uncertainty in the results is crucial for ensuring the reliability of any analysis or decision-making process. Task in this work is defined as detection of built areas in the narrow coastal zones of Croatia. Urbanization of the coastline is an irreversible process influencing both life on land in the water. Therefore, it is of utmost importance to track and determine the changes of the environment caused by anthropogenic actions for policy makers to be able to make long term spatial planning decisions.

1.1 Overview of global land cover maps

Nowadays, there are numerous sources of land cover data. Due to the development of fine-resolution remote sensing techniques and easier access to aerial and satellite imagery, the availability of land cover (LC) data has been increasing in the last ten years. Land cover data are available at local, national and global level. The open access culture has also contributed to free access to spatial information and unrestricted use of data, including LC datasets. These LC datasets are mainly used to create land cover maps which are continuously expanding. One of the key reasons for the growing number of land cover products is the increasing availability of high-resolution (10–30 m) data from various Earth Observation (EO) satellite missions, which are now accessible for free. Additionally, advancements in hardware and software are significantly enhancing processing capabilities. Therefore, the primary challenge has shifted from issues related to obtaining data (such as cost or lack of availability) to fine-tuning the different aspects of the classification process (Gilić et al., 2023).

D. García-Álvarez et al. (2022) have classified LC data producers into four main groups: individual users and small actors, research projects, governmental and other organizations and citizens producing LC information through Volunteering Geographic Information (VGI) initiatives. Main characteristic of local and small-scale producers is individually created maps for specific areas and points in time. This allows users to obtain very specific datasets that match their specific purpose usually connected to small projects within the community. Therefore, these types of products tend to remain unavailable for a larger public and are usually provided without the necessary technical information and general metadata. Nationally or internationally funded research projects usually generate LC datasets in collaboration with universities and research institutions. However, the limited duration of these projects often affects the continuity of their mapping efforts, and once the project concludes, the datasets are typically not updated or improved. Governmental and other large organizations are major producers of land cover data. The primary goal in these cases is to provide information about the areas under the organization's jurisdiction or those impacted by its policies and decisions. This data serves as a valuable resource for the policymaking process and is often part of broader cartographic initiatives by national, regional, and sometimes international organizations. Since these projects are typically part of official mapping efforts conducted by governments and large organizations, they are often supported by substantial long-term funding. This backing increases the likelihood that the databases will be regularly updated and improved. Moreover, LC data produced by global organizations or governments are usually supported with validation information and metadata in addition to offering highly detailed and accurate information. Volunteered Geographic Information (VGI) refers to geographic data collected, shared and validated by individuals or groups, typically outside of traditional professional or governmental frameworks. This data is often generated through crowdsourcing and similar practices. The best-known example of this production is OpenStreetMap.

Land cover maps for national and regional areas are usually only available for highly developed countries which can afford to invest in the production of spatial information and in research programs such as the European Union, Australia, the United States and China. Many international institutions and organizations require comprehensive and consistent global data to support their activities. Global datasets are also essential for research communities focused on studying the Earth. On the other hand, national governments and organizations need large volumes of data to inform policy making at the national level. Additionally, numerous other institutions, associations, professionals, and researchers need highly detailed data for their work. Supra-national LC maps have been developed by the European institutions to assist policymaking and environmental monitoring in Europe. In other continents, supra-national LC maps are usually developed within the context of different projects funded by international institutions, such as the Food and Agriculture Organization (FAO) and various different US and European institutions. The latter include the European Space Agency (ESA) and the Joint Research Centre (JRC) of the European Commission, which have been actively involved in the production of supra-national LC maps for many developing areas with important biodiversity values (García-Álvarez et al., 2022). Through the Copernicus program, the EU has developed coherent and consistent LC mapping products specifically designed to monitor LC dynamics in targeted areas. These products are highly detailed both spatially and thematically, and have been created to address the needs of their intended user community or to provide essential data supporting various policies.

Among dozens of freely available land cover products, here are reviewed the most relevant ones currently available:

- Copernicus Global Land Services Land Cover Map at 100 m (CGLS-LC100; Buchhorn, Lesiv, et al., 2020),
- European Space Agency (ESA) WorldCover (Zanaga et al., 2021),
- CORINE Land Cover (Büttner et al., 2021),
- Dynamic World (Brown et al., 2022),
- Esri Sentinel-2 Land Cover (Karra et al., 2021),
- Global Land Cover with Fine Classification System at 30m (Zhang et al., 2024).

Beside general land use cover maps recent trends show interest for thematic land use cover maps. Thematic Land Use Cover (LUC) datasets map parts of the Earth's surface as a specific land cover, considering not just its extent but also its intensity of distribution. One of the most common features mapped by thematic LUC products are vegetation covers. They were first class among land covers to be specifically studied when thematic LUC maps were introduced. Interest in this area has continued due to the valuable insights these datasets offer for tracking forests, climate change and other related issues focusing on areas of special biodiversity or environmental value. Thematic LUC maps typically depict land covers with more precision than general LUC maps. Some maps offer data on the percentage of the study area covered by a particular land cover type. In other instances, they define the boundaries of a specific land cover in great detail and with high accuracy. The fact that thematic LUC maps focus on a single, specific cover normally means they are more accurate than general LUC maps. Thematic agricultural LUC datasets usually show the extent of croplands and pasturelands or the cover fraction per unit of analysis, i.e., per pixel. Built-up areas are also commonly included in thematic LUC maps. Like vegetation datasets, they provide a time series of data essential for change detection. Lastly, thematic LUC maps focusing specifically on water and other covers display water occurrence information beside other themes. There are fewer thematic maps available for water compared to other groups of thematic maps, but these maps often include information on changes over time, just like other more represented thematic map groups (García-Álvarez et al., 2022).

1.1.1 COPERNICUS GLOBAL LAND SERVICES LAND COVER MAP AT 100 m (CGLS-LC100)

As part of the European Copernicus service which provides systematic global monitoring of the Earth's land surface, the Copernicus Global Land Service released yearly global LC maps at 100 m resolution available from 2015 to 2019 (Buchhorn et al., 2020). Copernicus' Global Land Cover product aims to provide dynamic LC layers that allow creating custom LC maps that fit the needs of different map users. The first Copernicus Global Land Service Land Cover Map at 100 m (CGLS-LC100) was derived from 100 m time-series of the vegetation instrument on board of the PROBA satellite (PROBA-V short for Project for On-Board Autonomy-Vegetation), a database of high-quality LC reference sites and several additional datasets. All input datasets were conducted manually at 10 m spatial resolution which allowed the generation of fraction layers since input data for classification have approximately 100 m spatial resolution. Overall, 141,000 unique 100 x 100 m training locations were used to train a single random forest classifier to generate the Copernicus Global Land Cover layers (Buchhorn et al., 2020).

The processing in this tiling grid, with UTM projection, ensures high quality and facilitates the continuity with Sentinel-2 observations. In the training phase, a random forest (RF) classification algorithm was applied to determine optimal classification parameters for different ecoregions (so-called biome clusters) so that regional differences are taken into account, and then each classification model was used to generate region-wide pre-products, including discrete classification and fraction layers (Buchhorn et al., 2020). Then, the local adaptive classification modelling first divided the globe into a lot of regions and then trained the corresponding local classifiers using the regional training samples, and the global land-cover map was spatially mosaicked by a lot of regional land-cover classification results. For generating final discrete land cover maps, expert rules applied on auxiliary datasets OpenStreetMap (OSM) data, 90-m CopernicsDEM digital surface model, and others that are listed in Table 1) were used (Buchhorn et al., 2020). The available global annual maps include a discrete classification with 23 classes, following UN-FAO Land Cover Classification System (LCCS) and related surface area statistics (km^2) and versatile cover fractions (%) for the 10 base classes. FAO's Land Cover Classification System (LCCS) are: 6 closed forest classes, 6 open forest classes, shrubland, herbaceous vegetation, herbaceous wetland, cropland, moss and lichen, bare ground/sparse vegetation, permanent snow and ice, built-up, permanent water cover and open sea.

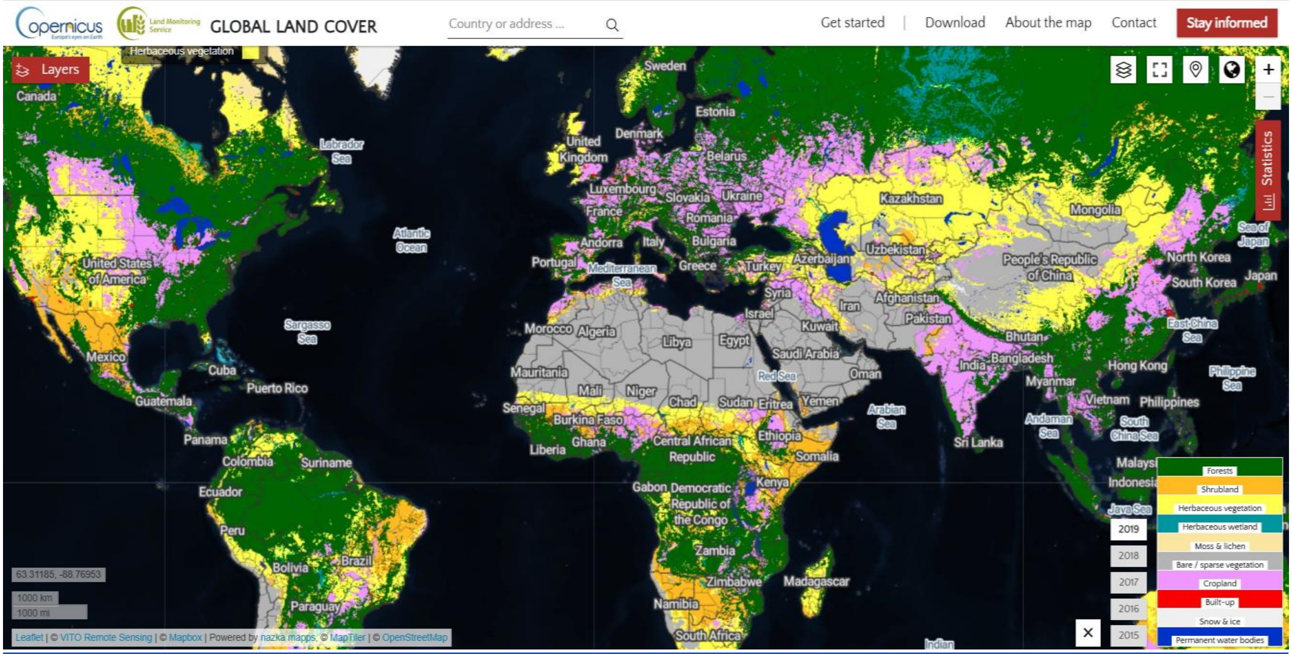


Figure 1.1.1.1 Global land cover layers viewer for 2019 https://lcviewer.vito.be/

1.1.2 ESA WorldCover

The ESA WorldCover global 10-m land cover product is currently available only for 2020 and 2021. It is produced by the European Space Agency, based on Sentinel-2 and Sentinel-1 constellations. The discrete classification map provides 11 classes and is defined using the Land Cover Classification System (LCCS) developed by the Food and Agriculture Organization (FAO). Therefore, land cover is classified into 11 classes: tree cover, shrubland, grassland, cropland, built-up, bare/sparse vegetation, snow and ice, permanent water bodies, herbaceous wetland, mangroves, moss and lichen. As reported by the ESA WorldCover product user manual (Zanaga et al., 2021) land cover classification uses the same algorithm that was used for CGLS-LC100 with the main difference of Sentinel multispectral images being used instead of PROBA-V scenes. When it comes to temporal information, the product represents land cover data for the years 2020 or 2021, covering the period from January 1 to December 31. The processing is based on Sentinel-1 and Sentinel-2 data for the respective reference year.

The methodology for producing the WorldCover land cover products starts with data preprocessing. Sentinel images are filtered for cloud coverage (Sentinel-2) and terrain corrected (Sentinel-1). Auxiliary data used in the process are altitude and slope are extracted from the Copernicus Global 30 m Digital Elevation Model (DEM). The processing includes information about the biome of the specific location, as well as meteorological and ecological features. Based on these, the yearly mean and standard deviation are calculated for 7 variables from the datasets: reference evapotranspiration, snow water equivalent, minimum temperature, maximum temperature, vapor pressure, vapor pressure deficit, and wind speed (Zanaga et al., 2021). The classification workflow is divided into a training and prediction phase. In the training phase, training points are randomly sampled for each training location. The training points selected are used for training different models (scenarios) with a gradient boosting decision tree algorithm (CatBoost). Three models are trained in total, trained during the training phase and used to predict the labels of the input data together with the class probabilities (Zanaga et al., 2021).

The validation results showed that the overall accuracy of the WorldCover product is 74.4 ±0.1% for 2020 and 76.7±0.5% for 2021. In terms of land cover types, tree cover and snow/ice, cropland, water body, and bare/sparse vegetation classes had high accuracies, while shrubs, herbaceous wetland, and moss/lichen classes were mapped with lower accuracies. Overall accuracy is 77%, with the highest accuracy of 80.7% for Asia and the lowest accuracy of 67.5% for Oceania (Zanaga et al., 2021). The ESA WorldCover products are provided per 3 x 3 degree tile, 2651 in total. Since the WorldCover maps for 2020 and 2021 were generated with different algorithm versions changes between the maps include both changes in real land cover and changes due to the used different algorithms.

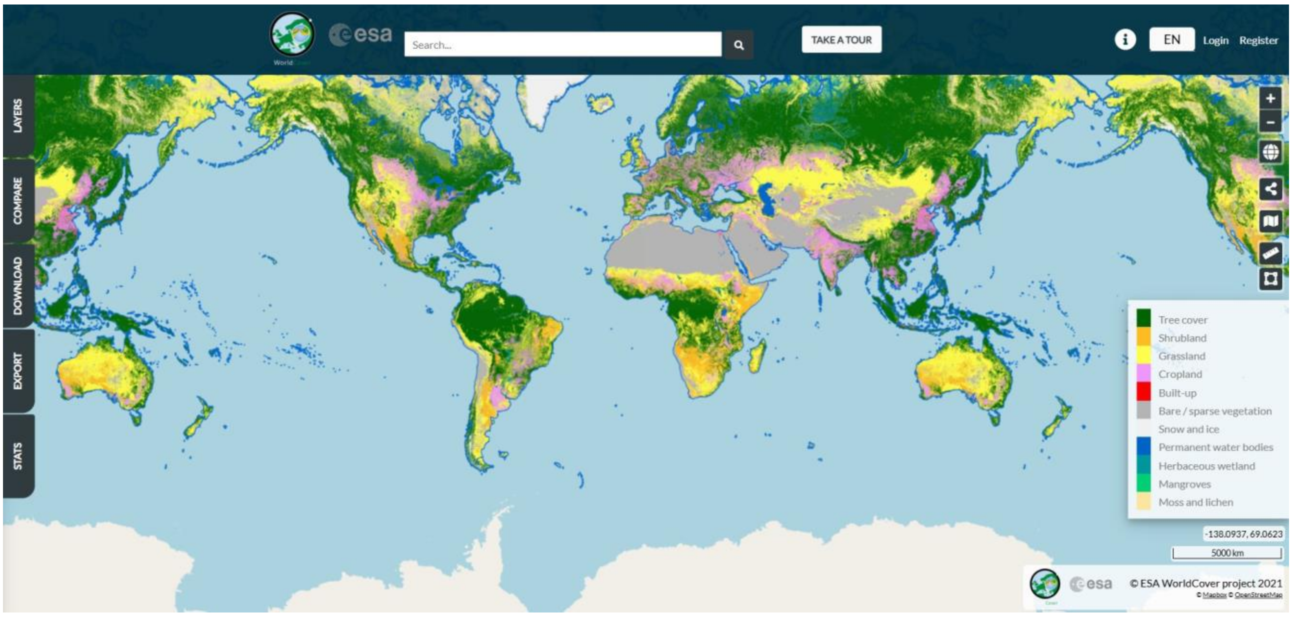


Figure 1.1.2.1 ESA WorldCover viewer for 2021 [*https://esa-worldcover.org/en*](https://esa-worldcover.org/en)

1.1.3 CORINE Land Cover

Copernicus is the European Union's Earth Observation Program. It offers information services based on satellite Earth observation and in situ (non-space) data. These information services are freely and openly accessible to their users through six thematic Copernicus services (atmosphere monitoring, marine environment monitoring, land monitoring, climate change, emergency management and security). The Copernicus Land Monitoring Service (CLMS) provides geographical information on land cover and its changes, land use, vegetation state, water cycle and earth surface energy variables to a broad range of users in Europe and across the world in the field of environmental terrestrial applications (Büttner et al., 2021).

Of all the CLMS datasets, CORINE Land Cover (CLC) is by far the best known. It is one of the oldest and most successful programs on land monitoring, offering very high levels of accuracy and detail over pan-European areas. The term CORINE stands for 'Coordination of information on the environment' and it was a European Commission program from 1985 dedicated to different environmental issues. The reference year of the first CLC inventory was 1990 (CLC1990), and the first update was created in 2000. Further inventories followed with an update cycle of 6 years with the last update in 2018. Through 40-year-long history the basic technical parameters of CLC have not changed: it has got 44 classes in five main land cover groups: artificial surfaces, agriculture, forests and semi natural areas, wetlands and water and secondly, the geometric detail consist of 25 hectare minimum mapping unit (MMU) and 100 meter minimum mapping width (MMW).

The first CORINE Land Cover project in 1990 included 12 of that time EU member countries and 10 countries of Central and Eastern Europe (preparing for the EU accession). A single date Landsat MSS and later Landsat TM satellite imagery taken in 1985-1996 in 1:100.000 color print format was used as the basis of photointerpretation. Mapping technology entailed photo-interpretation by drawing manually on a plastic overlay covering a 1:100000 scale printout of a satellite image map. In-situ (ancillary) data were mainly topographic maps and black-and-white aerial photographs as hardcopy. Drawings on the plastic overlay had to be digitized to create the final database. The lack of ortho-correction and the deformation of the plastic often caused geometric distortion of the resulting land cover data (Büttner et al. 2021). 28 years later the fifth and latest CLC inventory for the reference year 2018 was produced under the Copernicus program. All European Economic Area (EEA) countries and the UK were mapped, which includes transoceanic French, Portuguese and Spanish possessions such as French Guiana, Azores and Canary Islands. CLC 2018 is based on multitemporal Sentinel-2 (and optionally Landsat-8) satellite images. The so-called 'change mapping first' approach has been applied that consists of three steps: revision of last CLC (to remove mistakes and inconsistencies), creation of CLC-Change by image-to-image comparison and creation of new CLC by combining revised CLC and CLC-Change in a GIS operation. The evolution of CORINE Land Cover is shown in Table 1.1.3.1.

Future enhancement of CLC potentially addresses both geometric resolution (change of Minimum Mapping Unit) and thematic resolution (level-4/5 classes or attribution of polygons), as well as others (e.g., improvement of nomenclature). Another way of future development for CORINE Land Cover is in the direction of global coverage. The European Environment Agency (EEA) has determined to develop and design a conceptual strategy and associated technical specifications for a new series of products which should meet the current and future requirements for monitoring and reporting obligations. This process of development is nominally called the "2nd generation CORINE Land Cover (CLC)" and referred to as “CLC+” (Büttner et al., 2021). The “CLC+” implies products dedicated to improve land cover and use monitoring for the decade to come and become a new European baseline. Optimally, providing a true 100 m spatial resolution (1ha MMU) raster CLC product. True 100 m resolution means that the MMU is decreased to 1 ha. Currently available 100 m raster CLC product has actually 25 ha resolution (MMU), as it is created by transforming the 25 ha MMU vector product to a 100 m resolution raster.

Table 1.1.3.1 The evolution of CORINE Land Cover (Büttner et al., 2021)

	CLC1990	**CLC2000**	**CLC2006**	**CLC2012**	**CLC2018**
Satellite data	Landsat-5 MSS/TM single date	Landsat-7 ETM single date	SPOT-4/5 and IRS P6 LISS III dual date	IRS LISS III and RapidEye dual date	ESA Sentinel-2 dual date Landsat 8
Time consistency	1986-1996	2000 +/- 1 year	2006+/- 1 year	2011-2012	2017[1] (2018)
Geometric accuracy, satellite image	≤ 50 m	≤ 25 m	≤ 25 m	≤ 25 m	≤ 10 m
Min. mapping unit/width	25 ha/ 100m	25 ha/ 100m	25 ha/ 100m	25 ha/ 100m	25 ha/ 100m
Geometric accuracy, CLC	100 m	better than 100 m	better than 100 m	better than 100 m	better than 100 m
Targeted thematic accuracy ≥ 85%	not checked, probably not reached	reached	reached	reached	reached
CLC-Change	not implemented	non-standard change mapping methodology	standard change mapping methodology	standard change mapping methodology	standard change mapping methodology
Production time	10 years	4 years	3 years	2 years	1,5 year
Documentation	incomplete metadata	standard metadata	standard metadata	standard metadata	standard metadata
Access to the data	unclear dissemination policy	dissemination policy agreed from the start	free access for all users	free access for all users	free access for all users
Number of countries	28	39	39	39	39

Figure 1.1.3.1 CORINE Land Cover viewer for 2018 https://land.copernicus.eu/en/map-viewer

1.1.4 Dynamic World

Dynamic World is near real-time (NRT) global 10 m land use land cover mapping. This dataset is produced for the Dynamic World Project by Google in partnership with National Geographic Society and the World Resources Institute and it is available on Google Earth Engine. Dynamic World is based on Sentinel-2 images. These images are processed and filtered before further use. Annotations from Sentinel-2 Level 2A Surface Reflectance images are paired with masked and adjusted Sentinel-2 Level 1C Top of Atmosphere images to create the training data for the model. Cloud and shadow masking is done in three steps: first, using the Sentinel-2 Cloud Probability (S2C) product, then adjusting for over-masking with the Cloud Displacement Index (CDI), and finally, removing shadows based on the sun's position using a method called directional distance transform (DDT) (Brown et al., 2022).

Dynamic World modeling approach involves innovative training methodologies which set it apart from other land cover products. It relies on semi-supervised deep learning and requires spatially dense annotations. To collect a diverse set of training and evaluation data, the world was divided into three regions: the Western Hemisphere (160°W to 20°W), Eastern Hemisphere-1 (20°W to 100°E), and Eastern Hemisphere-2 (100°E to 160°W). At each sample location, an initial selection of Sentinel-2 images from 2019 was performed. The scenes were further filtered to remove images with many masked pixels. Individual tiles of 510 × 510 pixels, centered on the sample sites, were then extracted from random dates in 2019. The tiles were sampled in the UTM projection of the source image, with one tile selected corresponding to a single Sentinel-2 ID number and date. Annotation was performed by drawing vector polygons encompassing a single land cover class and then assigning a land cover class to that polygon.

To label a large dataset of Sentinel-2 scenes, producers worked with both expert and non-expert groups of annotators. The first group included 25 annotators with previous photo-interpretation and/or remote sensing experience. The expert group labeled approximately 4000 image tiles, which were then used to train and measure the performance and accuracy of a second non-expert group of 45 additional annotators who labeled a second set of approximately 20000 image tiles. For the validation set, 409 tiles were used and each of them was annotated by three experts and one non-expert. The final training dataset was produced after applying augmentations for the deep learning algorithm, namely rotations and contrasting (Brown et al. 2022). Fully Convolutional Neural Network (FCNN) is a deep learning model specifically used in training datasets of Dynamic World. It is applied to ensure the transfer of the supervised label data to a system that could be applied globally.

To generate Dynamic World NRT products, producers apply the normalization of the raw Sentinel-2 L1C imagery. This output is then masked using a cloud mask derived from the unnormalized L1C image. Creation of these images is triggered automatically when new Sentinel-2 L1C and S2C images are available. The NRT collection is continuously updated with new results. For a full Sentinel-2 tile (roughly 100 km x 100 km), predictions are completed within approximately 45 minutes. In total, Dynamic World evaluates approximately 12000 Sentinel-2 scenes per day. A new Dynamic World land cover image is processed approximately every 14.4 s (Brown et al., 2022).

The classification schema or "taxonomy" for Dynamic World consists of nine classes (forest land, grassland, cropland, wetland, built-up and other) to ensure easier application for estimating carbon stocks and greenhouse gas emissions. Unlike single-pixel labels, which are usually defined in terms of percent cover thresholds, the Dynamic World taxonomy was applied using dense polygon-based annotations such that land cover/use labels are applied to areas of relatively homogenous cover types with similar colors and textures (Brown et al., 2022).

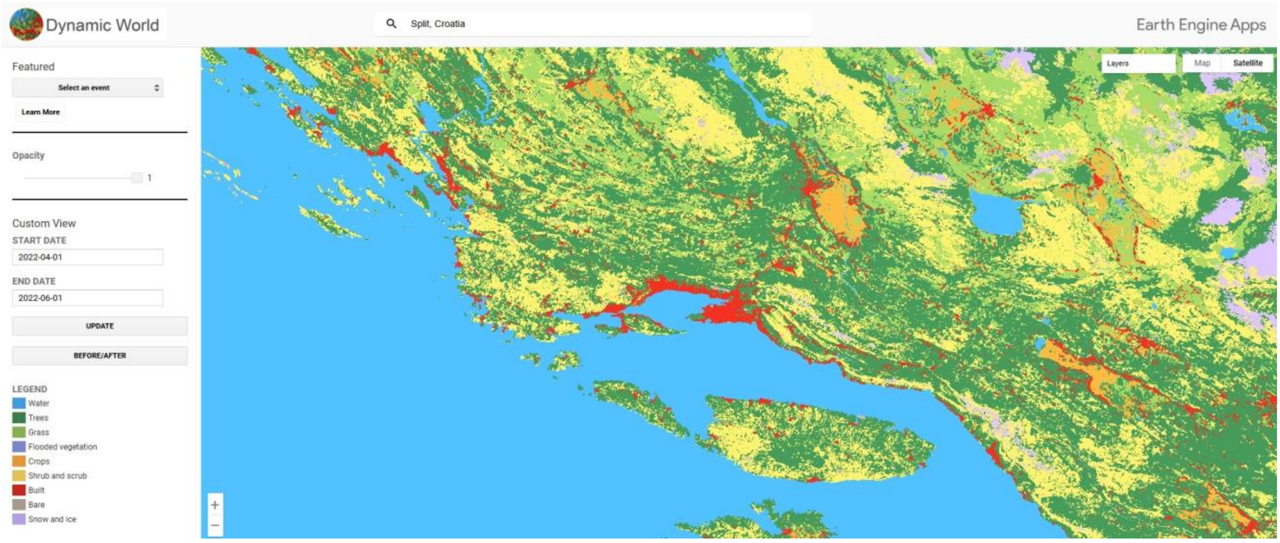


Figure 1.1.4.1 Dynamic World Land Cover viewer https://dynamicworld.app/explore/

1.1.5 ESRI Sentinel-2 Land Cover

Sentinel-2 land cover time series of the world is a 10 m resolution global map product produced by Impact Observatory and Esri hosted on the Microsoft Planetary Computer. Land use land cover (LULC) is derived from ESA Sentinel-2 imagery at 10 m resolution. Map is annually derived from 2017 to 2023. Each year is generated with Impact Observatory's deep learning AI land classification model, trained using billions of human-labeled image pixels from the National Geographic Society. The global maps are produced by applying this model to the Sentinel-2 Level-2A image collection on Microsoft's Planetary Computer, processing over 400000 Earth observations per year. The underlying deep learning model uses 6-bands of Sentinel-2 L2A surface reflectance data: visible blue, green, red, near infrared, and two shortwave infrared bands. To create the final map, the model is run on multiple dates of imagery throughout the year, and the outputs are composited into a final representative map for each year. The algorithm generates LULC output that provides a 9-class map of the surface, including: water, trees, flooded vegetation, crops, rangeland, built area, bare ground, snow/ice and clouds (no land cover information due to cloud coverage) (Karra et al., 2021).

As stated by Karra et al., no time-series product is available that maps global LULC at 10m Sentinel-2 resolution, meaning it can deliver the timeliest analysis of land use for the past year and the time series to detect change in fine resolution. Esri Sentinel-2 Land Cover maps are available to the public from ArcGIS Living Atlas of the World in GeoTIFF files in UTM projection and 6 degrees of longitude wide separated for each year.

Esri (Environmental Systems Research Institute) is a global leader in geographic information system (GIS) technology, providing software, tools and solutions for spatial data analysis, mapping and visualization, with prime product being ArcGIS. Esri's partner Impact Observatory is a mission-driven technology company with experience in providing advanced algorithms for AI and ML.

Figure 1.1.5.1 Esri Land Cover viewer [*https://livingatlas.arcgis.com/landcoverexplorer*](https://livingatlas.arcgis.com/landcoverexplorer)

1.1.6 GLOBAL LAND COVER WITH FINE CLASSIFICATION SYSTEM AT 30 m (GLC_FCS30D)

GLC_FCS30D is a global 30 m land-cover dynamics monitoring dataset produced by Aerospace Information Research Institute of the Chinese Academy of Sciences. GLC_FCS30-2020 serves as a benchmark dataset for generating training samples and identifying land-cover information. It is produced combining the 2019-2020 time series Landsat surface reflectance data, Sentinel-1 SAR data, DEM terrain elevation data, global thematic auxiliary dataset and prior knowledge dataset. Unlike traditional land-cover datasets, which group land types into broad categories (e.g., forest, urban, water), the GLC_FCS30D uses a fine classification system containing 35 land cover subcategories. This system includes a greater number of detailed land cover classes, providing a more detailed view of land types and their distribution globally.

GLC_FCS30D has been developed using continuous change detection and all available time-series Landsat imagery based on the Google Earth Engine platform. Researchers developed a novel method that integrates the benefits of continuous change detection algorithms, local adaptive modeling and Landsat satellite imagery to effectively identify landscape changes over time. They then applied temporal-consistency optimization algorithms to classify the areas of change and ensure the accuracy of the data. Producers firstly identified the temporally stable pixels and the time points of abrupt changes for changed pixels. To accurately determine the land cover of the changed pixels in time-series monitoring, producers derived spatiotemporally stable training samples, updated the changed pixels using multitemporal classifications and finally minimized the cumulative error caused by independent classifications (Zhang et al. 2024). The flowchart of the proposed method combining the continuous change-detection (CCD) algorithm and a local adaptive updating algorithm is shown in Figure 2.6.1. The GLC_FCS30D dataset was evaluated using over 84000 validation samples from across the globe and found to be highly accurate, particularly in identifying forests and croplands.

Validation was conducted through independent interpretation by trained interpreters based on high-resolution aerial photography, multitemporal Landsat images and other relevant ancillary datasets. In 2020, product achieved an overall accuracy of 80.88 % (±0.27 %) for the basic classification system (10 major land-cover types) and 73.04 % (±0.30 %) for the LCCS (Land Cover Classification System) level-1 validation system (17 LCCS land-cover types) (Zhang et al., 2024).

GLC_FCS30D has significant advantages over other global land-cover datasets in terms of land-cover type diversity: it contains 35 discrete land cover types, among which forests and wetlands are subdivided into 10 and 7 land cover subcategories. With imagery from 1985 till present day, GLC_FCS30D has been able to capture specific location temporal changes of Earth’s surface such as enlargements of Amazon rain-forest and enlargement of urbanization in the Yangtze River Delta area which are important, not only at the regional level, but also globally (Zhang et al. 2024).

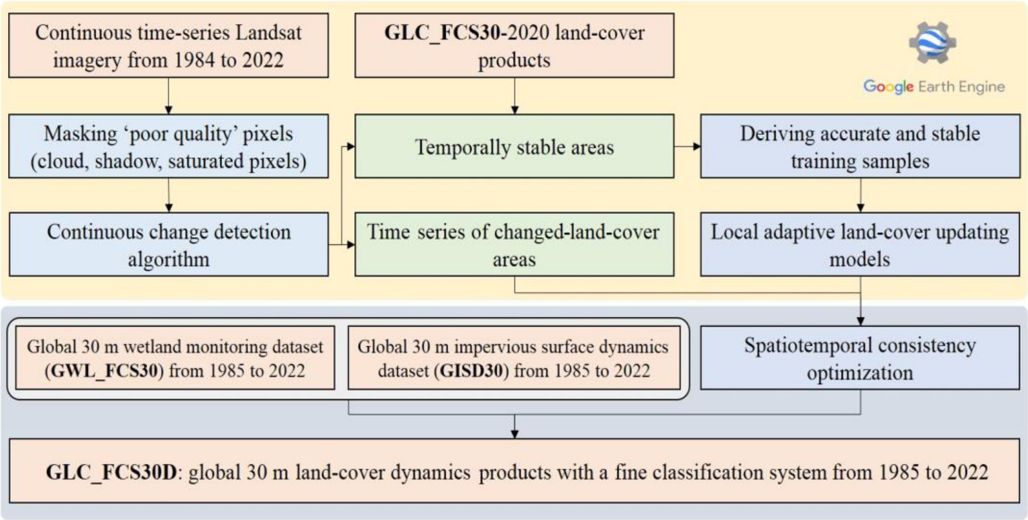


Figure 1.1.6.1 GLC_FCS30D workflow (Liu and Zhang, 2024)

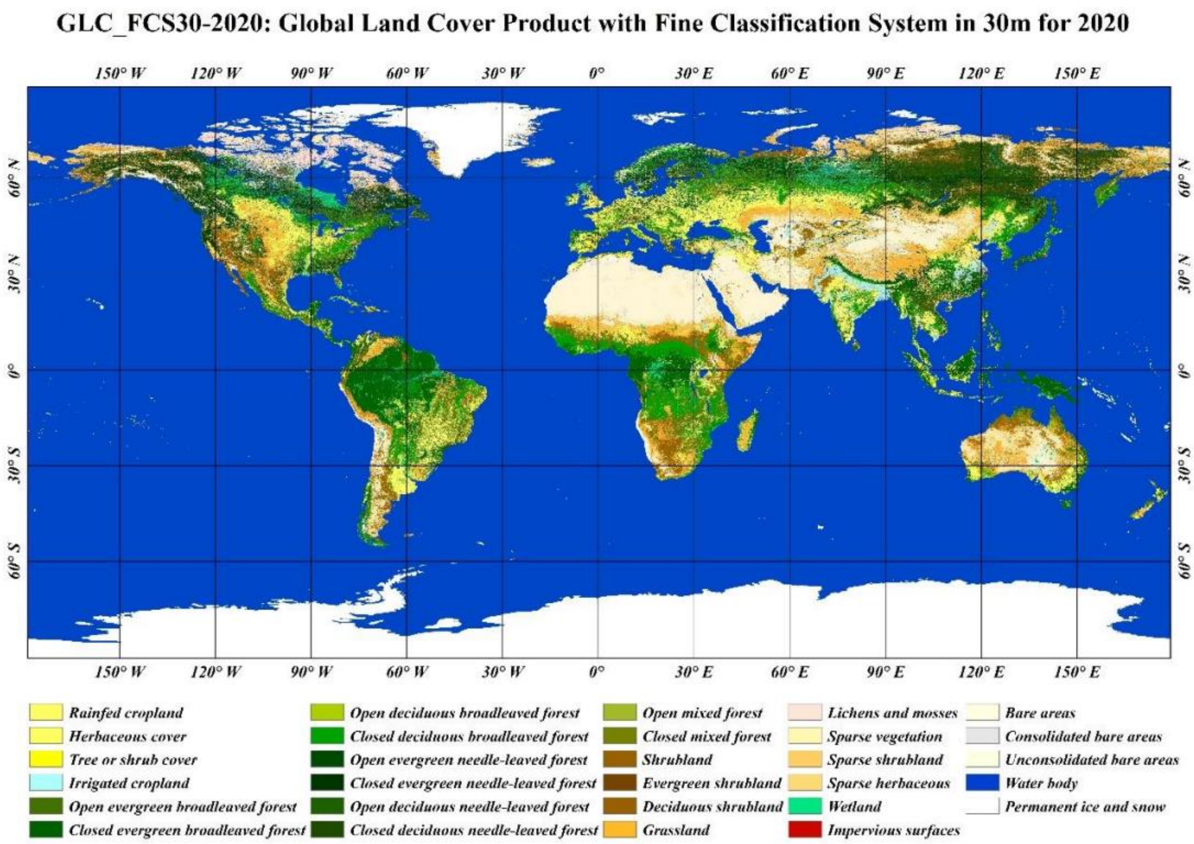


Figure 1.1.6.2 Overview of GLC_FCS30D with classes (Liu and Zhang, 2024)

1.2 Main application areas of land cover maps

Nowadays, ecosystems and biodiversity are facing numerous challenges, many of which are driven by human activities. Rising temperatures and more frequent extreme weather events are disrupting ecosystems worldwide. Species are forced to adapt, migrate or face extinction. Another burning problem is large-scale logging and land conversion for agriculture that are leading to habitat loss for countless species. Forests, especially tropical ones, play a crucial role in regulating climate, maintaining biodiversity and providing resources for many species. Conversion of natural landscapes into urban areas, agricultural fields or industrial zones reduces the extent and quality of habitats available for wildlife. This also contributes to climate change by reducing carbon storage capacity in forests and wetlands. Understanding our planet’s biodiversity is essential for the well-being of all life on Earth. Through Earth observations, we are not only able to collect spatial data about our planet and its physiognomy, but we are able to make informed environmental actions and recognize climate trends that will influence the future of our world. The policies and initiatives we as humans make today will shape the Earth of future generations. Therefore, making the right decision today is the crucial goal but also a risk for which we must equip ourselves with the most effective technology, data and software to help us with the decision-making process.

The Group on Earth Observations (GEO) is a global collaboration between over hundred governments, organizations and associates, dedicated to understanding Earth’s complexity as a whole. As a collaborative intergovernmental body, GEO is dedicated to co-producing user-driven Earth Intelligence solutions. GEO is based on collaboration of science, policy and community interests for the collective benefit of society. By collecting and sharing information, ranging from satellite images of forests to oceanic temperature readings and beyond, GEO provides a comprehensive view of our planet's well-being (URL 2).

The GEO has defined nine Social Benefit Areas (SBAs) in which Earth observations, including LULC data, provide useful evidence in support of policymaking. The SBAs are:

- biodiversity and ecosystem sustainability,
- disaster risk reduction,
- climate change,
- energy and mineral resource management,
- food security and sustainable agriculture,
- sustainable development goals,
- public health,

- sustainable urban development and
- water resources management.

Social Benefit Areas refer to regions or sectors in which specific social advantages or improvements can be achieved for a community. These areas are typically identified based on their potential to address key societal needs, improve the quality of life and contribute to overall well-being. SBAs are often used in the context of policy planning, development initiatives and social impact assessments. They can address the way land use or resources are managed to maximize social outcomes.

All application areas are interconnected and strive for the common goals. Selected application areas addressed in this thesis are all examples of usage of land cover maps targeting the general Social Benefit Areas. These applications, and numerous others, comprehend different themes concerning environmental sustainability and future development on local, national and global level.

1.2.1 Agricultural monitoring

Over the past years, food insecurity has steadily increased, now affecting more than 30% of the global population (URL 2). Climate change, poverty, conflict, population growth and disease have all intensified the severity of food insecurity. To tackle this challenge, decision-makers need improved data on crop production and land use to monitor and enhance food security. GEOGLAM (Global Agricultural Monitoring), a GEO initiative, delivers valuable information on land conditions and agricultural production at local, national and global levels. One of many projects under the GEOSLAM cap is Copernicus4GEOGLAM. It is a collaborative initiative that combines the capabilities of Copernicus, managed by the EC Joint Research Centre, with the objectives of GEOGLAM. This initiative helps regions adapt to climate change by offering reliable data that can inform sustainable agricultural practices and reduce the impact of climate variability on food systems. Data used for the project are field-survey and Earth observation data. This project aims at producing baseline information allowing countries in Africa to improve their agricultural monitoring systems. Examples of national implementation are the following three countries: Tanzania, Uganda and Kenya. For each country the in-season and end season mapping of the crop type and crop mask is described. Satellite data used are Sentinel-2 and Sentinel-1 images over the area of interest between January and August of 2021. Data on crops and other land cover classes have been acquired in the field on 500 x 500 m segments. Various classification algorithms were tested, including supervised (maximum likelihood), TempCNN and Random Forest (RF) of which RF showed best results. Based on monthly synthesis Sentinel-2 images, precomputed features and ground truth from fieldwork (75% for training, 25% for validation), the RF classifier has been applied on all the tiles to produce the crop type map. The initial classification output contains 45 classes (of which 34 crop types). Crop mapping involves identifying the types, distribution and conditions of crops grown across the country. Some of the crop types for Kenya are: maize, millet, beans, rice, tea, wheat, sugarcane and potatoes. Kenya is a country of over 52 million inhabitants where 38.6% of people live below the poverty lines (UN World Food Program). Agriculture is the main economic driver but is very vulnerable to climate shocks, with arid and semi-arid regions making up 80 percent of the country's land. Rural communities that depend on crops (95%) are rain-fed.

GEOGLAM's crop mapping tracks how agricultural land is used over time, providing insights into shifting cropping patterns, expansion of farmland and changes in agricultural practices. GEOGLAM provides estimates of crop yields, which are critical for food security planning. Accurate yield forecasts help governments, aid organizations and farmers plan for potential food shortages and distribution needs. Based on information provided by Copernicus4GEOGLAM program, African farmers are able to grow and adapt their crops to changing environmental conditions. By integrating land cover data with other environmental data, GEOGLAM helps assess the impacts of climate change on crop yields and land quality, allowing for targeted interventions to restore degraded land or protect crops from emerging threats. The use of satellite data in crop mapping helps improve decision-making, making agriculture more resilient and ensuring that land resources are used efficiently and responsibly. Figure below shows the result of Copernicus4GEOGLAM as map of crop types distribution in Kenya in-season and at the end of season.

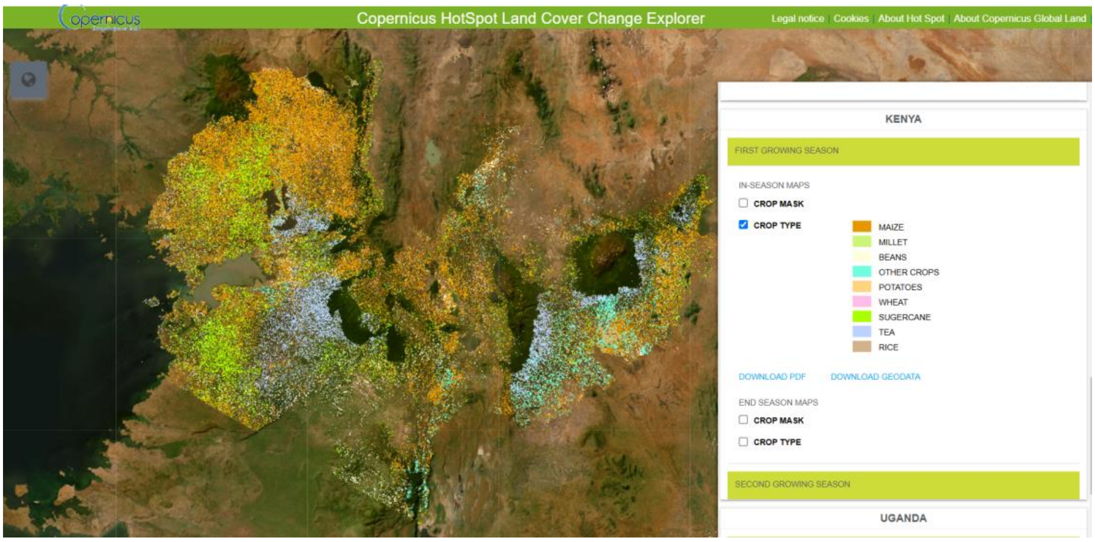


Figure 1.2.1.1 Copernicus4GEOGLAM viewer of crop types in Kenya https://hsm.land.copernicus.eu/

1.2.2 Land degradation

Land degradation is defined as the reduction or loss of the biological or economic productivity and complexity of rain fed cropland, irrigated cropland, or range, pasture, forest and woodlands resulting from a combination of pressures, including land use and management practices. It leads to a decrease in the land's ability to support crops, livestock, or native vegetation. Currently, 25% of the world's land is degraded (URL 2). Addressing land degradation requires preventive mapping and measuring land degradation at the national level across all land types. Without this information, investments, policies, and interventions will be ineffective. Therefore, GEO has been working on ensuring policymakers, organizations and communities have the knowledge needed for sustainable land management. They have developed the Land Degradation Neutrality Initiative that aims to improve land degradation measurement and mapping while also contributing to land degradation neutrality (LDN). LDN is defined as a state whereby the amount and quality of land resources necessary to support ecosystem functions and services remain stable. LDN is measured through SDG 15 Life on Land, specifically indicator 15.3.1, proportion of land that is degraded over total land area (URL 3). Goal 15 is defined by the United Nations Convention to Combat Desertification (UNCCD) as to protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification and halt and reverse land degradation and biodiversity loss (URL 3). LDN targets are designed to guide countries in their efforts to prevent further land degradation, restore degraded lands, and maintain or improve the land's productivity.

SDG indicator 15.3.1 is a binary (degraded/not degraded) quantification based on the analysis of available data for three sub-indicators to be validated and reported by national authorities. The sub-indicators are: trends in land cover, land productivity and carbon stocks. The assessment and quantification of land degradation is challenging for any single indicator to comprehensively represent the land's condition. Although individual sub-indicators are necessary, they are not sufficient on their own. These sub-indicators capture changes in various ways: for instance, land cover or productivity trends can highlight relatively rapid changes, while shifts in carbon stocks reflect slower changes, indicating potential trends or nearing critical thresholds. National data on the three sub-indicators is collected through existing sources (e.g., databases, maps, reports). Regional and global data sources used for each of the sub-indicators are the following. Land cover and land cover change data are available in the ESA-CCI-LC, containing annual land cover area data at 300 m spatial resolution for the period from 1992 to present or SEEA-MODIS containing annual land cover area data at 500 m spatial resolution for the period 2001-2019. The datasets for calculating land productivity data represented as vegetation indices and their derived products are: MODIS data products, averaged at 250 m pixel resolution, integrated over each calendar year since 2000 and Copernicus Global Land Service products, averaged at 1 km pixel resolution and integrated over each calendar year since 1998. And lastly, soil organic carbon stock data reflect the balance between organic matter gains and they are available in the Harmonized World Soil Database (HWSD). The One Out, All Out principle is applied taking into account changes in the sub-indicators which are depicted as: positive or improving, negative or declining or stable or unchanging. If one of the sub-indicators is negative for a particular land unit, then it would be considered as degraded subject to validation by national authorities. By looking at variations in sub-indicators along with local factors like climate, soil, and land use, national authorities can identify which areas are degraded, calculate the total affected land, and report the results.

1.2.3 Sustainable urban development

Land cover maps are indispensable tools for urban development, helping to ensure that cities are built sustainably, efficiently and in a way that promotes the well-being of their inhabitants and the natural environment. Nowadays, urban areas are expanding uncontrollably and without sustainable planning. This growth tends to produce a range of environmental, social and infrastructural challenges. Common mission is to develop cities and human settlements that are inclusive, safe, resilient and sustainable while managing environmental, climate and disaster risk more effectively. Land cover maps analytically help urban planners and policymakers understand many aspects of city growth as well as the ecological, infrastructural and overall urban dynamic effects of these changes. Uncontrolled urban expansion consumes important rural land. This inefficient spread characterizes urban sprawl. Land cover maps provide much important data to track urban sprawl. Observing meaningful changes in land cover by comparing land cover maps from different time periods, enables us to see the large expansion of built-up areas and the conversion of agricultural or forested land into urban areas along with the overall increase in impervious surfaces.

One of many applications of land cover data within urban planning and development is Indicator on Land cover change. The LCC indicator 25 (CCI25) is a candidate indicator developed under the Integrated Monitoring and Assessment Programme of the Mediterranean Sea (IMAP) (URL 4) in support of Integrated Coastal Zone Management (ICZM) Protocol (URL 5). This indicator aims to support balanced allocation of uses and avoid urban sprawl by limiting linear extension of urban development including transport infrastructure along the coast and securing ecosystem health. It is directly linked to ecological objectives concerning coastal ecosystems and landscapes. Main objective of the LCC indicator is to inspect land use/land cover change of purpose to which land is profited by humans as an almost irreversible process. Maintaining the natural dynamics of coastal areas and preserving coastal ecosystems and landscapes is a common goal for which the LCC indicator provides an inventory of the urbanization pressures on coastal ecosystems. Due to climate change, an increase of risks in coastal zones is foreseen in the natural and built environment. The main spatial data sets needed for the LCC indicator 25 calculation are the following: land use/land cover data, coastline, elevation data, protected areas and administrative units. Among examined land cover maps, authors of the indicator concluded that the best data source for LC/LU data is Copernicus Global Land Cover and Change data (CGLS-LC100). Although, Copernicus Global Land Cover data, having a pixel of 1 ha, does not recognize small changes, but still classifies very well the built-up areas. On the other hand, ESA WorldCover Project data reveals very detailed land cover situations and best matches the aerial image due to superior spatial resolution of 10 m. As it was available only for the year 2020 at the time, its suitability for detecting changes cannot be assessed and with the future updates for coming years, ESA WorldCover data is the best candidate to be used for the LCC indicator 25 in future (Baučić et al., 2022a).

1.2.4 Water resources management

Water, which makes up over 70% of the Earth's surface, is fundamental to all aspects of life, whether for daily household use, agricultural production, industrial processes or sustaining ecosystems. With the growing demand for water and the rising unpredictability of climate patterns, it has become more urgent than ever to closely monitor and protect the world’s water resources. By monitoring extreme weather impacting water resources, such as prolonged droughts or unexpected flooding, we are able to track weather anomalies (URL 2).

Another key aspect to monitor is freshwater resources to ensure they are sustainably managed and kept clean for human consumption. Sustainable water management includes forecasting and control of quantity and quality of water, anticipating the needs of agricultural lands and predicting how climate change might alter freshwater availability. Land cover is never static; it has consistently been modified by natural and anthropogenic drivers. Studying the influence of land cover on hydrological processes such as change in surface roughness; soil structure and infiltration rate is crucial for water resources. Also, land cover changes effects on water quantity and quality. Land uses such as urban and agricultural areas have a tendency to diffuse contamination in freshwater systems. Measuring impact of agricultural activities and impact of urbanization on the quality of water means to quantify the toxic and hazardous waste in water. On the other hand, grass and forests reduce water runoff and soil erosion, but affect water quality in the form of present organic matter and sediments. Water on its way to groundwater or river channels is governed by a number of factors including geology, soils, topography and land use, and also extreme conditions such as floods and droughts. By implementing Integrated Water Resources Management (IWRM), countries and regions can better manage these challenges while ensuring that water resources are used in a way that benefits both people and the environment.

1.2.5 Climate change adaptation

Understanding how our world and climate has evolved provides valuable insights into shaping a more sustainable and prosperous future. Ofcourse, the future cannot be predicted, but by analyzing historical data and identifying trends, we can make informed projections about potential future scenarios. While these predictions are based on current patterns and assumptions, they provide valuable insights into possible outcomes and help guide decision-making. In an effort to predict land cover changes by 2050, Clark Labs at Clark University has teamed up with Esri. This collaboration leverages decades of satellite observation data from the European Space Agency (ESA) Climate Change Initiative (CCI). The result is a new series of predictive global land cover maps, now available for open use, which classify land cover by type and assess its vulnerability to human development. These maps offer a powerful tool for understanding potential future changes in land cover, enabling better planning and decision-making for sustainable development (URL 6).

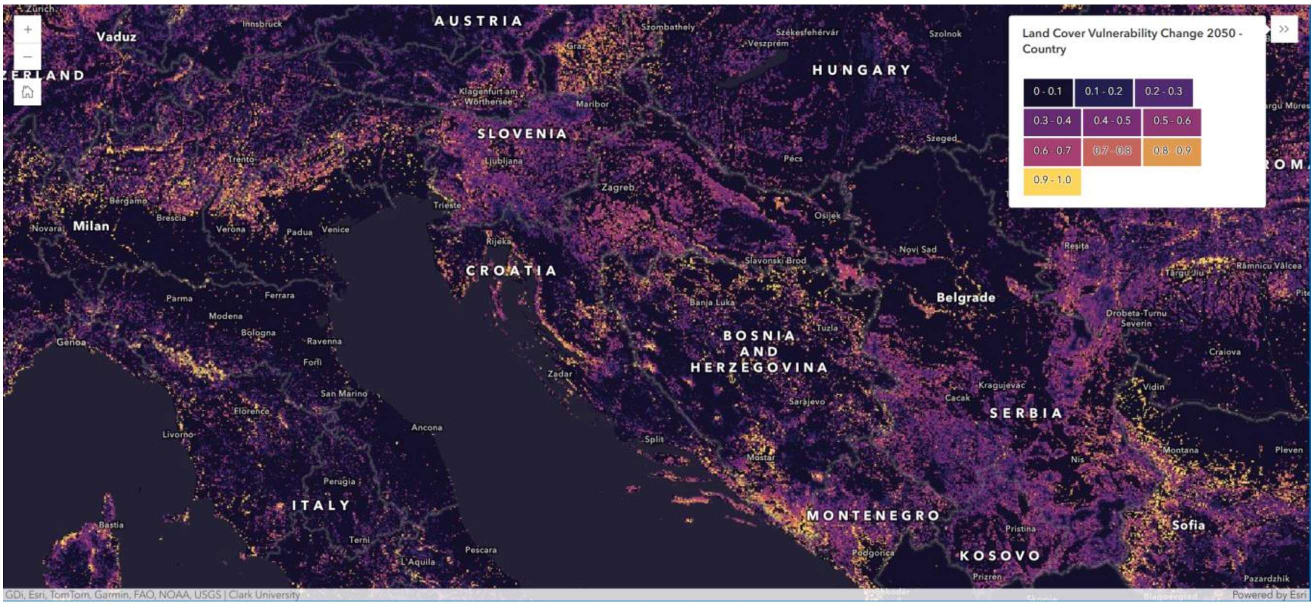


Figure 1.2.5.1 Land Cover Vulnerability Change 2050 prediction by Esri and Clark University

To identify the areas most vulnerable to change, land cover across each country was assessed from 2010 to 2018. This assessment helped create a vulnerability model that predicts where natural vegetation could be converted to agricultural and urban land by 2050. The model incorporates both quantitative and qualitative factors to understand the global pressures of land development. These factors include proximity to modified areas, infrastructure, population density, Gross Domestic Product (GDP) data, as well as geophysical and bioclimate information. By merging the forecasts for each country, the model generates global land cover and vulnerability maps with a resolution of 300 meters, offering a detailed look at future land changes. Specifically, decision-makers can project future land cover patterns, allowing them to better assess the impact risks imposed by development or industry in a given location (URL 6).

MATERIALS AND METHODS

2.1 Validation and uncertainty estimation

Validation can be defined as the process by which we assess how certain or reliable data or result is. This is done by comparing it against other data or information that we use as a reference and consider to be true (García-Álvarez et al., 2022). Ground truth data refers to information that is directly collected through observations, measurements or analysis on the ground, providing a reference to compare against remotely sensed data or automated classifications. Ground truth data helps determine the reliability of land cover classifications generated through remote sensing techniques. Field surveys are one of the most direct and reliable sources of ground truth data. This involves the collection of data through field visits to specific sites, where land cover types are identified and recorded manually or using handheld devices. The biggest downside of field surveys is cost and time-consummation. Another source of ground truth data is satellite imagery. High-resolution imagery allows for the identification of fine-scale features such as individual trees, buildings and roads, enabling precise classification and validation of land cover types. Lastly, existing validated land cover maps can be used as reference for validation of new maps. These maps may come from national or international mapping programs that have already undergone validation and quality assurance processes and therefore are reliable sources for ground truth data. Additionally, crowdsourced data, that involves the collection of ground truth data by a large group of individuals, can be a useful supplement to traditional ground truth methods. With the increasing production and use of LUC maps and models, more attention has been paid to the uncertainty and limitations of these data and analyses.

Uncertainty can be defined as the lack or the degree of certainty about any data or geospatial analysis due to the difference between reality and its representation through geospatial data or tools (García-Álvarez et al., 2022). Understanding the differences between land cover maps and real landscapes, along with their reliability, is essential. Only by doing so can we assess the accuracy of the information they provide and determine how reliable they are as a basis for policy decisions. A full validation exercise that characterizes all the uncertainties in a dataset or model is a complex task. It requires advanced expertise and a range of tools and strategies, each tackling different sources of uncertainty.

The main objective of any validation process is to ensure that the map correctly represents the actual land cover patterns on the ground. Various methods exist to evaluate how well a land cover map reflects the true distribution of land cover types. Below are the most common methods of validation used in land cover mapping:

- The Cross-Tabulation Matrix
- Accuracy assessment statistics
- Kappa Coefficient
- Areal and spatial agreement metrics

Cross-tabulation is a core analysis method that explores the spatial relationship between two datasets, either raster or vector. It combines the datasets spatially, generating a map or table that shows how the values in one dataset correspond to those in the other. This helps us understand whether the two datasets share the same values at a given location and, if not, how they are related to other values. A cross-tabulation matrix is an effective tool for validating land use and land cover maps. It compares the map’s classification results with reference data, enabling the identification of discrepancies and the evaluation of classification accuracy. The agreement between the reference data and the LUC map being evaluated can be assessed using the cross-tabulation matrix only if they correspond to the same date. By cross-referencing various land cover categories, the matrix reveals areas of agreement and disagreement, offering insights into the map's precision and reliability. A cross-tabulation matrix is also referred to in the literature as the confusion or error matrix. Cross tabulation is usually the first step in any validation, as the matrix that provides information regarding the spatial agreement between the LUC map being validated and the reference dataset. When considering the validation of uncertainty of a LUC map series with two or more-time dates, the cross-tabulation matrix is the optimal tool that provides most information about the land cover change between two LUC maps. The cross-tabulation matrix provides valuable information, which can easily be summarized using other tools and metrics (García-Álvarez et al., 2022).

It's important to remember that the accuracy of a land use and land cover map is not typically consistent across the entire mapped area, and significant spatial variations may occur. The larger the area being mapped, the more likely it is that accuracy will vary between different regions. The cross-tabulation matrix does not provide information about these spatial differences (García-Álvarez et al., 2022).

Accuracy assessment statistics are standard metrics that measure the similarity between two georeferenced datasets. These statistics, derived from the cross-tabulation matrix, provide detailed insights into the matrix's contents. Key information includes overall accuracy, producer’s accuracy, and user’s accuracy. Typically presented alongside the cross-tabulation matrix, they offer additional information, such as category area adjusted for error levels and confidence intervals. Overall accuracy is one of the most commonly used metrics in validation showing the agreement between two maps expressed in values of between 0 and 100. As the name says, it provides accuracy over map as a whole and does not provide information about the accuracy at category level of the LUC map. Overall accuracy is highly correlated with the Kappa Index (Olofsson et al., 2014), which explains why both metrics provide similar information. A key difference is that Kappa accounts for the agreement expected by chance, whereas overall accuracy does not factor this in. The Kappa Index assesses the agreement between two sources of spatial data, corrected by the agreement that is expected by chance. They are typically used to compare the agreement between two maps and to compare one map with reference information (García-Álvarez et al., 2022). The main advantage of Kappa indices is that they provide a standard measure. Kappa agreement ranges between −1 and 1, where 1 means total agreement, −1 total disagreement and 0 random agreement. As a general rule, Kappas above 0.7–0.8 are considered good enough for validation. Kappas above 0.9 indicate very high agreement (García-Álvarez et al., 2022).

The areal and spatial agreement metrics are obtained from the cross-tabulation matrix and therefore do not provide any additional information. However, they are standard metrics that allow us to measure the agreement between two maps and summarize it in a single figure. Therefore, it could be said that they are similar to the user's and producer's accuracy metrics and to Kappa indices. These metrics can be used to study how similar a land cover map is to another reference map or to check the similarity between a two map for the same year.

The methods mentioned above do not employ fuzzy logic, but instead apply a binary logic. When calculating agreement, the two elements agree or don't agree. Partial agreements are not considered.

Another important term in validation are producer's and user's accuracy: The producer's accuracy is the likelihood that a specific land cover type in the reference data was correctly identified by the map, where on the other hand user's accuracy is the likelihood that a pixel classified as a specific land cover type by the map actually corresponds to that land cover type in the real world (Pontius et al., 2008).

2.1.1 ESA WorldCover 2021 validation method example

Validation of land cover products are usually given by the producers. The aim of any validation report is to assess the accuracy and reliability of the land cover map. The Product Validation Report of ESA

WorldCover 2021 is an example of how producers conduct validations of land cover products. This process for WorldCover 2021 was conducted independently by two independent teams.

The ESA WorldCover product has been independently validated by Wageningen University (statistical accuracy) and IIASA (spatial accuracy). The WorldCover 2021 v200 product reaches an overall accuracy of 76.7% (Tsendbazar et al., 2022). The Product Validation Report describes the validation results of the WorldCover 2021 product. To provide information on map quality, consistency and fitness for use, producers used two validation methods:

- Statistical accuracy assessment and
- Qualitative comparison with the WorldCover 2020.

The statistical accuracy assessment was designed to evaluate the product's accuracy using a rigorous statistical method, which is based on independent validation data selected using probability sampling. The validation dataset is based on random sampling, employing a global stratification. Globally, the validation dataset consists of 21 752 primary sampling units. The available primary sampling units per continent allow to carry out a statistical accuracy assessment for each continent with a high level of precision. Each primary sampling unit covering an area of 100 m × 100 m was divided into 10×10 small blocks called secondary sampling units. Therefore, the validation dataset is compatible for assessing land cover maps with 10 - 100 m resolutions. For the thematic representation, the generic land cover elements recorded at each secondary sampling unit include trees, shrubs, grass, crops, built-up areas, bare area, lichens/mosses, open water, snow and ice, and regularly flooded areas. The land cover elements were defined according to the United Nations Land Cover Classification System (LCCS).

To validate the WorldCover2020 product, the CGLS-LC validation data for the reference year 2019 was updated with validation data for 2020 by revisiting 2860 random validation points and points with a high possibility of change. For the reference year 2021, producers further updated the validation dataset with a total of 4410 validation sites. Next to revisiting a random subset of validation sites, they have also revisited the validation sites with high possibility of land cover change between 2020 and 2021 (Tsendbazar et al., 2022).

In terms of class-specific accuracies, tree, snow/ice, cropland, water body and bare vegetation classes had high accuracies. Grassland and built-up classes had moderate accuracies, while shrubs, wetlands and moss/lichen classes had relatively lower accuracies globally. In general, at the global scale, there was a slight underestimation of moss and lichen and bare vegetation class, while a slight overestimation of trees was observed when comparing against the validation dataset. Shrubs, grassland, cropland, built-up and permanent water bodies classes were mapped with balanced user's and producer's accuracies (Tsendbazar et al., 2022).

For qualitative comparison, the WorldCover 2021 and 2020 products are compared visually on selected locations. Visual comparison between the products confirmed the improved depiction of arid areas and the delineation of cropland, bare vegetation, grassland and lichen moss classes. The differences in the versions are both due to the improvements made in the classification algorithm as well as the natural dynamics between the two years.

Results of validation showed that the overall map accuracy of the WorldCover 2021 product was 76.7±0.5%. At the continental level, it had an overall accuracy above 72%. The overall accuracy was highest for Asia (82.1%) followed by South America, Europe, and Africa. The lower accuracy for Oceania could be due to high shrubland vs grassland and trees vs grassland confusions in the open woodlands of Australia. Similarly, high shrubland vs trees and grassland vs trees confusions in Siberian temperate tundra regions could explain the lower overall accuracy in Eurasia. In terms of class-specific accuracies, tree cover, snow/ice, cropland, water body, and bare/sparse vegetation classes had high accuracies. Grassland and built-up classes had moderate accuracies, while shrubs, wetlands, and moss/lichen classes had relatively lower accuracies globally.

2.2 Proposal of the validation method

To fully utilize open global and regional land cover maps, users must apply an appropriate validation method to access the following:

- appropriateness of open data for the intended use and
- estimation of the uncertainties of the results.

This work has the intention to propose simple but effective methods users could perform to access open data appropriateness for their need and to estimate the level of uncertainty in their results. This proposal is based on and further elaborates the method from the validation example given in Baučić et al. (2022b). The proposed steps can be grouped into:

- validation of input data (by studying validation reports provided by the data producer, by requirements analysis/data fitting intended use, by comparison with the data of higher quality using visual inspection and using quantitative methods such as area or pixel calculations),

- validation of GIS operations (examining uncertainties introduced by GIS operations – propagation of uncertainties from input data sets to results in the form of spatial, temporal and thematic errors).

The results include uncertainties, introduced by input data and by the calculation steps performed in QGIS. To correctly interpret and to understand limitations of the results, it is necessary to perform validation. The validation assessment should provide the level of reliability of data and results, highlighting any potential errors or biases in the mapping process. This helps users make informed decisions by understanding the accuracy and confidence associated with the land cover maps. Furthermore, validation ensures that the data meets the required standards for specific applications, allowing users to better understand the scope and limitations of the maps. This process ensures that the data is appropriate for its intended purpose and can be confidently used in decision-making processes. By verifying the accuracy and reliability of the results, validation provides users with the necessary insights to assess whether the land cover maps can support their intended use effectively.

2.2.1 Validation of input data

Land cover data chosen for this thesis are the following three:

- ESA WorldCover Project Land Cover,
- Esri Sentinel-2 Land Cover and
- GLC_FCS30D

all for the year 2021. The input data validation is based on validation reports provided by each data producer. For the ESA WorldCover 2021 overall accuracy is around 76%, with tree cover, snow/ice, cropland, permanent water body and bare/sparse vegetation classes being classified more accurately than shrubs, herbaceous wetland, and moss/lichen classes. For the GLS_FCS30D overall accuracy is greater than 80% for the basic classification system of 10 major land cover types. The accuracy is little lower for the LCCS level-1 validation system and it equals around 73% and around 68% for the LCCS level-2 system. For the Esri Sentinel-2 Land Cover overall accuracy of 85% across all nine classes is achieved. Across all regions, water, trees, crops and built area perform particularly well with user's accuracies above 80%, while other classes, notably grass, flooded vegetation, and bare ground, need further improvement.

Regarding the spatial extent, all three of selected land cover datasets have world coverage. When it comes to spatial resolution, ESA WorldCover and Esri Sentinel-2 Land Cover provide 10 m resolution. On the other hand, the GLC_FCS30D is a 30 m spatial resolution product. ESA

WorldCover is now existing for reference years 2020 and 2021. Esri Sentinel-2 Land Cover covers temporal scale from 2017 to 2023. The GLC_FCS30D temporal resolution before the year 2000 was every 5 years, while after 2000, it is updated annually. When it comes to digital availability, all of the three land cover data are open data and accessible in the form of GeoTIFF files.

Input datasets also use different classification systems. The most detailed dataset classification wise is the GLC_FCS30D which contains 35 land cover categories based on UN-LCCS. The land cover map is especially detailed when it comes to forests and low vegetation. ESA WorldCover is a 11-class product also based on UN-LCCS. The Esri Sentinel-2 Land Cover consists of 9 classes. The most important disparity between UN-LCCS and the Esri classification system is in defining green urban areas such as parks, yards and groves. In Esri Sentinel-2 Land Cover they will appear as built areas rather than trees or rangeland classes. Built-up areas in the UN-LCCS system do not include urban park and recreation areas. The lack of spatial detail in land use classification means that land features like trees, parks and groves are often generalized into broader human activity categories, such as built area. The described disparity of classification systems is not the only one that can emerge when comparing different land cover products, therefore studying classification schemes should be taken into account when performing actions on any land cover products. Land cover products can only be compared and analyzed together when their classification schemes align or in other words, maps share the same classes. To achieve this correspondence, if not prior, land cover raster ought to be reclassified to the mutually represented categories.

Based on validation information reported by producers, both ESA WorldCover and Esri Sentinel-2 Land Cover have superior spatial resolution of 10 m compared to GLC_FCS30D's 30 m resolution. On the other hand, GLC_FCS30D is superior in the context of variety and detail classification wise, offering 35 fine classes. Moreover, GLC_FCS30D is available for the period from 1985 to 2022, enabling monitoring and change detection over three and a half decades. Esri Sentinel-2 Land Cover has a temporal scale of 7 years, therefore being a prime to ESA WorldCovers's only two annual releases so far.

The next step in validation users can perform is visual inspection of data on selected locations, significant for intended use or calculations. Aerial photos from the reference year are used as higher quality data and details on the photos are the source of ground truth data for comparison. In this phase, users review the land cover data by focusing on specific locations that are particularly important for the intended use of the map or for further calculations. These locations are often selected based on their relevance to the research or analysis goals. For example, if the map is being used to study land use change in a specific urban area, a user may focus their inspection on the city and suburbs. The goal is to ensure that the land cover classification in the map aligns with the real-world features on the ground.

Final step is a comparison of land cover data by area based quantitative method. The calculation of the area in this thesis is based on the Manual for IMAP Candidate Common Indicator 25 (Gilic et al., 2024). It is important to emphasize that all comparisons of areas for the purpose of validating land cover data against each other or reference data, must refer to the same year. The area of each land cover class is calculated per reporting units for each land cover source selected for validation. Each land cover type (built-up, agricultural, forest and seminatural, wetlands and water bodies) is assigned a specific area value; based on the pixel size and the number of pixels assigned to each class the area of each class is calculated. Once the areas of different land cover classes are calculated for each reporting unit, the next step is to compare the calculated areas across the same zones or reporting units for different land cover maps. The goal of this comparison is to identify differences or discrepancies between the classified land cover data and the reference or actual land cover data. By comparing the areas of land cover classes, differences between the calculated results and reference data will be highlighted. Understanding these differences is crucial for the validation process. The results of this comparison provide insight into the accuracy of the land cover data and highlight areas where adjustments may be needed, giving the user inurement that the land cover data is accurate and reliable for decision-making for their intended use.

Through validation of input data, users can make a first assessment of the appropriateness of land cover data for their intended use. Regarding the requirements of the selected project, the user is able to choose the superior data source that suits them the best.

2.2.2 Validation of GIS operations

Calculations on land cover data include various GIS operations. In data preprocessing data coordinates are transformed and raster data is aligned and mosaiced. Also, vector data such as reporting units are created by buffering and overlapping functions. All of these operations have a certain influence on final results. The primarily data used for creating reporting units in this work is based on coastline. It is a referent line that represents the border between sea and land in land cover data. As sea level varies, the horizontal position and shape of the coastline itself is variable and consequently affecting reporting units. Therefore, it is a possible source of uncertainty.

With all GIS operation possible uncertainties are introduced into the calculations. GIS zonal function enables overlaying reporting units (vector) over land cover products (raster). Zonal function includes an algorithm for pixel inclusion/exclusion that has a risk of uncertainties. The problem emerges when some zones include more overlapping pixels than whole ones. That means that the significant area of the coastal strip is not covered by whole pixels but parts. This situation happens when the size of a pixel is big compared to the size of a zone, therefore is directly connected with spatial resolution. For example, using a 100 m resolution land cover in a 300 m wide zone will result in a significant area of the zone covered by parts of pixels. In that case, further analysis requires data of higher accuracy regarding spatial resolution. Also, the situation of overlapping pixels can occur when the shape of the zone is inconvenient (narrow and irregular shapes tend to have more overlapping pixels). Another origin of uncertainty comes from the fact that land cover class is assigned to pixels according to the majority of land cover identified. In that case, a small part of the pixel could be of another land class then assigned.

Figures below (2.2.2.1 and 2.2.2.2) illustrate overlay of the narrowest coastal strip with two different spatial resolution land cover data. The border area of the coastal strip is not covered by whole pixels but parts. For 10 m resolution land cover the vast majority of the pixels are completely inside the polygon of the strip, where 30 m resolution land cover has visibly more area covered by overlapping pixels. It can be assumed that results for the narrow strip, when using lower resolution land cover (in this case GLC_FCS30D), have more uncertainties than for higher resolution land cover (in this case ESA WorldCover). This should be taken into account when comparing these land covers. Finally, spatial resolution has proven to have a significant influence on results of data processing using GIS operations. Therefore, poorly chosen land cover dataset with unsuitable spatial resolution regarding the needed requirements, can be a source of uncertainties in the calculations.

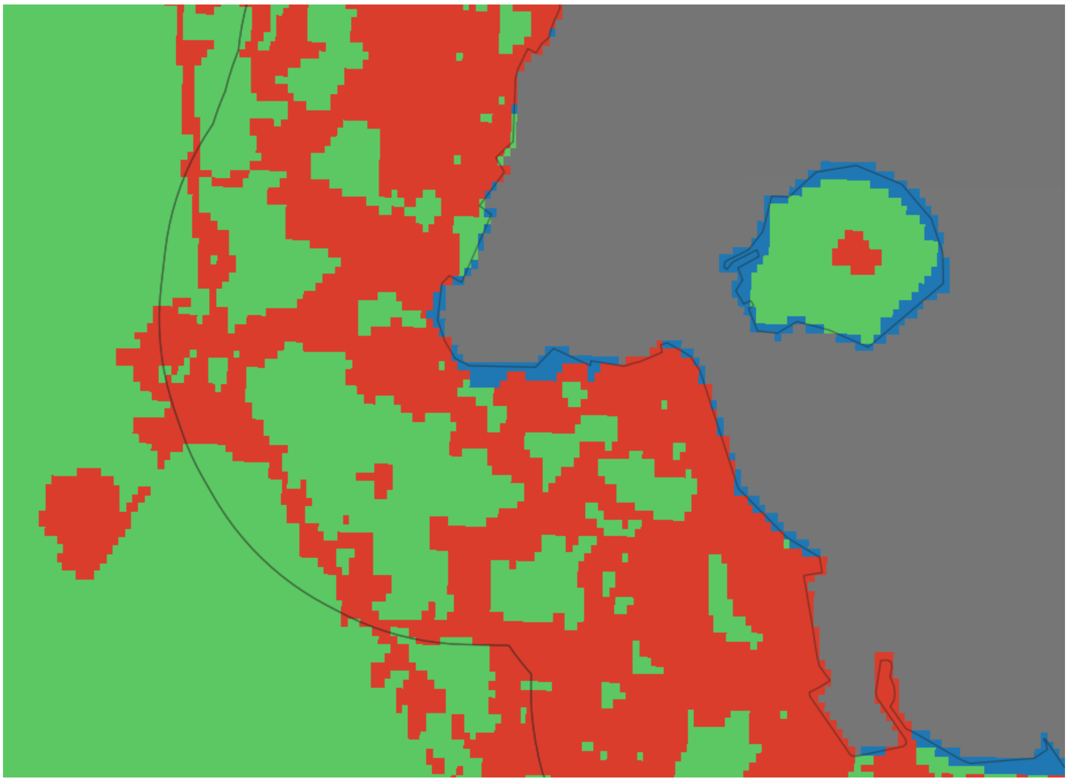


Figure 2.2.2.1 The narrowest coastal strip of 300 m width and the ESA WorldCover with spatial resolution of 10 m

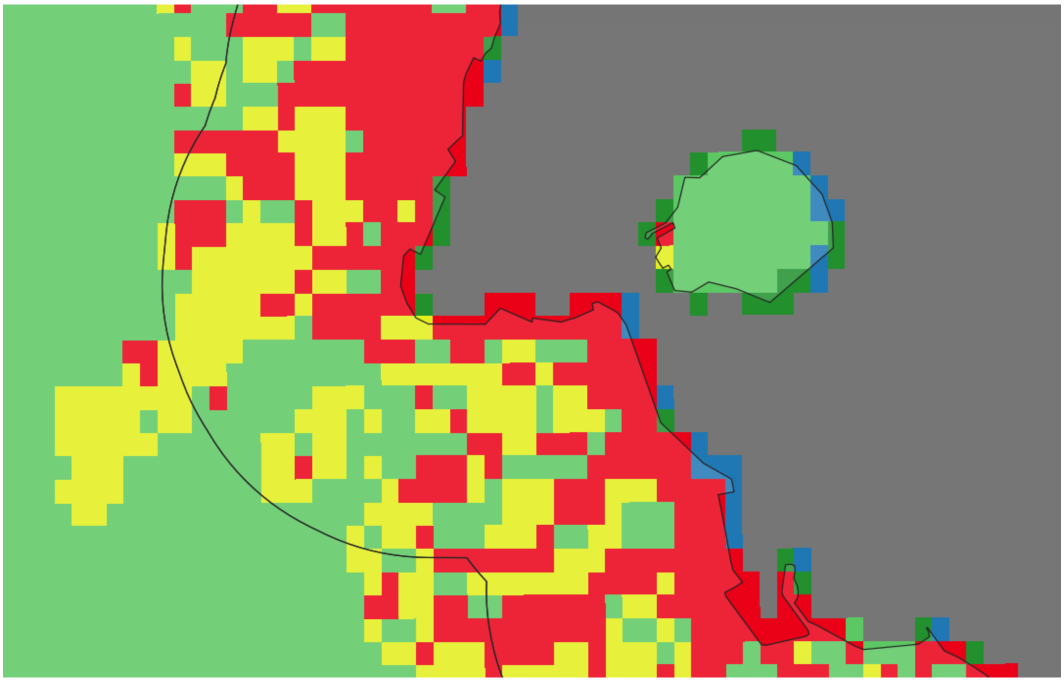
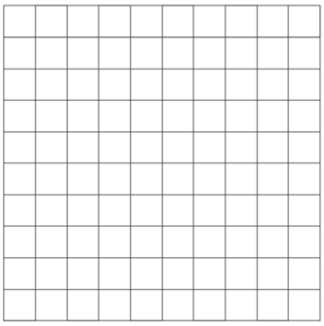


Figure 2.2.2.2 The narrowest coastal strip of 300 m width and the GLC_FCS30D with spatial resolution of 30 m

2.2.3 Uncertainties in area calculations

Evidently, the polygon’s geometry can influence the results of calculations computed in QGIS and introduce a certain level of uncertainties into the process. Generally speaking, the shape of a geometric object is a function of its morphology and it is a difficult parameter to quantify concisely. Within landscape ecology, numerous different indices have been developed to quantify specific spatial characteristics of shapes within landscapes. Same shape metrics can be applied to investigate how size and shape of polygon features condition area calculation accuracy (Petris et al., 2024). In the context of addressing the uncertainties regarding the overlapping pixels problem and to calculate the influence of a polygon's shape on uncertainties of results, quantitative measures were selected and tested. The simplest shape metric is the ratio of area to perimeter of the polygon. For long and narrow polygons, the perimeter is bigger compared to the polygon of the same area shaped more like a circle or square. Therefore, long and narrow polygons have larger borders (or perimeters) and are likely to have more overlapping pixels occurring on those borderlines. Ideal shape for a polygon to be covered with minimum overlapping pixels is naturally rectangular because of the pixel’s square nature.



$$\frac{A}{P} = \frac{a^2}{4a} = \frac{a}{4}$$

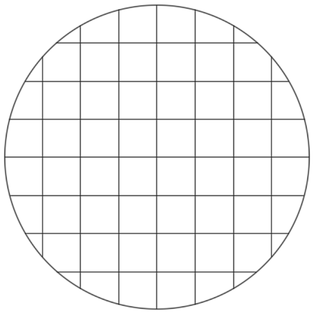
for $a = 1$ (unit square) $\frac{A}{P} = \frac{1}{4} = 0.25$ being the unit ratio

A - area of the square

P - perimeter of the square

a - side of the square

The next mathematical approximation of the polygon’s shape is a circle.



$$\frac{A}{C} = \frac{r^2\pi}{2r\pi} = \frac{r}{2}$$

for $r = 1$ (unit circle) $\frac{A}{C} = \frac{1}{2} = 0.5$ being the unit ratio

A - area of the circle

C - circumference (perimeter) of the circle

The perimeter - area ratio is the linear function of one parameter. Therefore, the obvious problem with this metric is that the ratio varies with the size of the shape. Another shape metric is the Shape Index (SI) (Forman, Gordon, 1986). It is an index used in landscape ecology to quantify the complexity or irregularity of the patch and to measure the compactness of a shape. It is commonly used to describe how elongated or fragmented a patch is compared to a simple geometric shape, such as a circle or square. A lower Shape Index indicates a more complex or irregular shape, while a higher index suggests a more compact or simple shape. The index is based on isoperimetric inequality, a mathematical principle that relates the perimeter (or boundary length) of a shape to its area. It states that for a given perimeter, the shape that encloses the largest area is a circle.

The Shape Index can be calculated using the formula:

$$\frac{P^2}{A} \geq 4\pi$$

P - the perimeter of the shape

A - the area of the shape

The equality holds if and only if the shape is a circle. Polygons that deviate significantly from circular shapes tend to have higher perimeters relative to their areas, indicating more complex or fragmented structures. The value of the Shape Index is 1 for the circle and closer to 0 for elongated and irregular shapes.

For the purpose of exploring the Shape Index as a measure of spatial convenience of the polygon for area calculation, the said index is calculated for all 220 Adriatic settlements in Croatia. The two settlements with the highest and the lowest Shape Index are determined - Škabrnja with the highest and Murter-Kornati with the lowest index. As previously said, the Shape Index is closer to 1 for more circular and regular shapes, and closer to 0 for elongated and irregular shapes. Therefore, it can be concluded that the Škabrnja with the SI = 0,755 has a relatively regular shape as the opposite of Murter-Kornati with the SI = 0,008 and highly irregular shape (Annex 4). The calculations are confirmed by visual inspection shown on the Figure below (2.2.3.1).

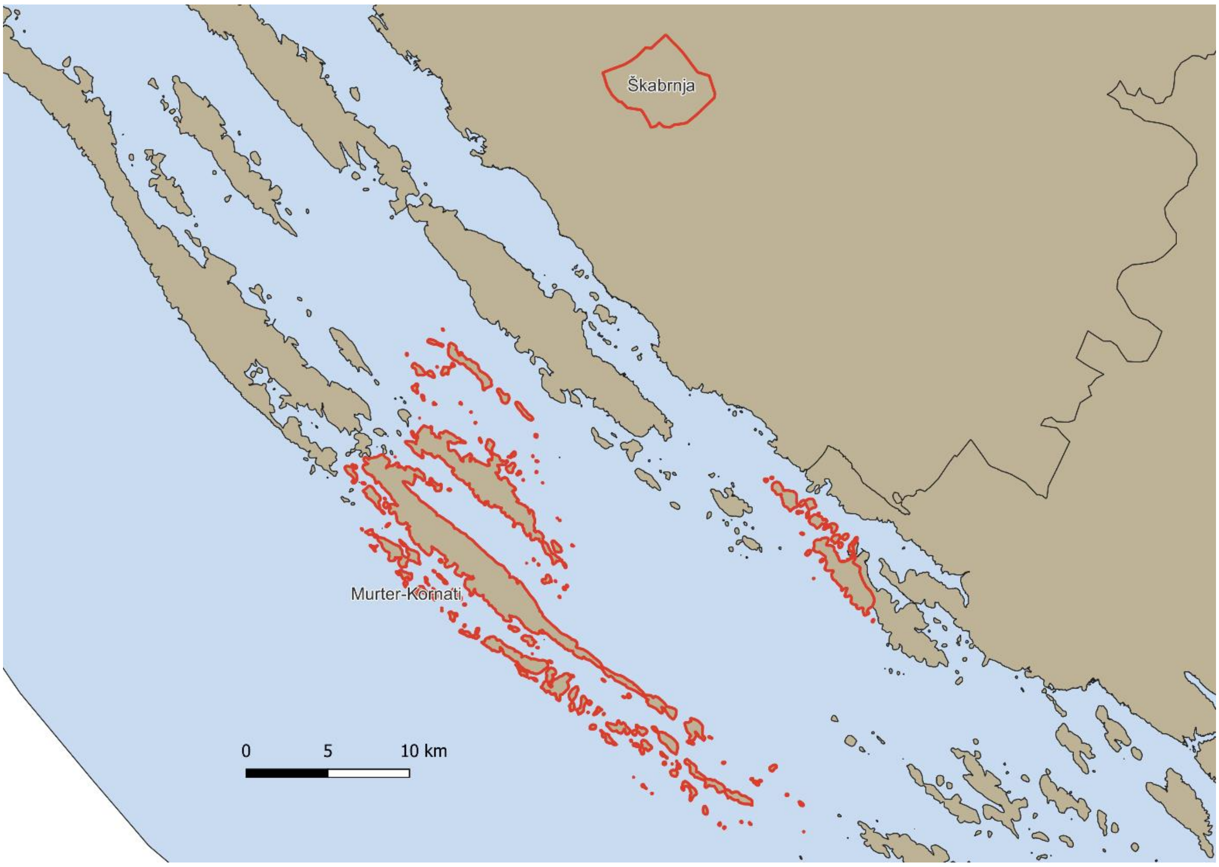


Figure 2.2.3.1 Škabrnja and Murter-Kornati settlements

Besides being able to quantify the shapes of polygons with numeric measure and determine extremes, it is crucial to investigate how the shape affects land cover calculations performed on chosen polygons with different spatial resolutions of land cover maps. The focus is on overlapping, bordering pixels –

those pixels that lie on the boundaries of a polygon and can often be considered as partial representations of the area. This exploration concentrates at two settlements, Škabrnja and Murter-Kornati, which are selected for their extreme Shape Indices. By comparing different spatial resolutions (10 m and 30 m), the aim is to investigate how the shape of the settlements influences the accuracy of land cover calculation at different scales. For this question to be answered, for two selected settlements (Škabrnja and Murter-Kornati) the share of bordering pixels in the total area covered with pixels is calculated both for pixel size of 10 m and 30 m. The former pixel sizes match the spatial resolutions of ESA WorldCover and Esri Sentinel-2 Land Cover (10 m) and GLC_FCS30D (30 m). For each settlement, the bordering pixels are extracted by identifying which pixels lie on the boundary of the polygon, but are not fully contained within it. This can be done by intersecting the boundaries of the polygon with the grid of pixels at both the 10 m and 30 m resolutions.

Next, the number of bordering pixels is counted for each resolution and their share, as a percentage of the total area covered by pixels within the settlement, is calculated. This gives insight into how the choice of spatial resolution influences the proportion of bordering pixels. At a finer resolution (10 m), each pixel represents a smaller area of the landscape. Therefore, more pixels are required to represent the boundary of a polygon, especially for complex or irregular shapes like settlements with jagged edges. At coarser resolution (30 m), the pixel size is larger, meaning fewer pixels are required to cover the same area. The boundaries of polygons are less detailed and complex edges are likely to be smoothed out into the pixel grid. This reduces the number of bordering pixels, which are found at the edges of the polygon. However, larger pixels cover a bigger area (30 m x 30 m = 900 m^2 per pixel) and as a result, the calculated area may overestimate or oversimplify the true land cover. The coarser resolution doesn't account for small details of the boundary and it could lead to more area being classified under a single polygon. A higher number of bordering pixels might be introduced at finer resolutions, while coarser resolutions could lead to fewer bordering pixels but at the expense of less detail and potential overestimation of the area due to the larger pixel size. Therefore, it is proposed to calculate the percentage of the bordering pixels in the total number of pixels. This method can shed light on the trade-off analysis between resolution, accuracy and shape complexity. Method is performed on two chosen settlements and results are shown in Table 2.2.3.1.

Table 2.2.3.1 Shape Index for two chosen settlements with share of bordering pixels at finer (10 m) and coarser (30 m) resolution

	Shape Index	Share of bordering pixels in 10 m resolution (%)	Share of bordering pixels in 30 m resolution (%)
Škabrnja	0,755	1,14%	3,36%
Murter-Kornati	0,008	5,56%	15,60%

For Škabrnja, the share of bordering pixels is 1.14% for the 10 m resolution, and increases to 3.36% for the 30 m resolution. The share of bordering pixels for Murter-Kornati is 5.56% at the 10 m resolution, but rises significantly to 15.60% at the 30 m resolution (Annex 4).

Evidently, Škabrnja performed better than Murter-Kornati, which aligns with expectations based on their Shape Indices. Since Škabrnja has a more regular and compact shape, it results in fewer bordering pixels, leading to less uncertainties in area calculations. In contrast, Murter-Kornati, with its more complex and irregular shape, exhibits a higher proportion of bordering pixels, which introduces more uncertainties to the area calculation, especially at coarser resolutions where bordering pixels are making up over 15% of total area.

Finally, it can be concluded that as the resolution decreases (and pixel size increases), the impact of bordering pixels along the boundary becomes more pronounced. Larger pixels cover a greater portion of the polygon's edge, resulting in a higher representation of bordering areas that do not fully belong to the polygon. This increases the uncertainty level introduced into the calculation. Therefore, at finer resolutions, the percentage of bordering pixels is reduced, leading to a more accurate representation of the polygon's boundaries and less uncertainties in area calculation. Figures below illustrate bordering pixels versus pixels fully contained by the polygon at finer and coarser resolution. The Shape Index proposes a method for quantifying uncertainties in area calculations, where a smaller SI corresponds to higher uncertainties in the area, while larger SI values indicate smaller uncertainties.

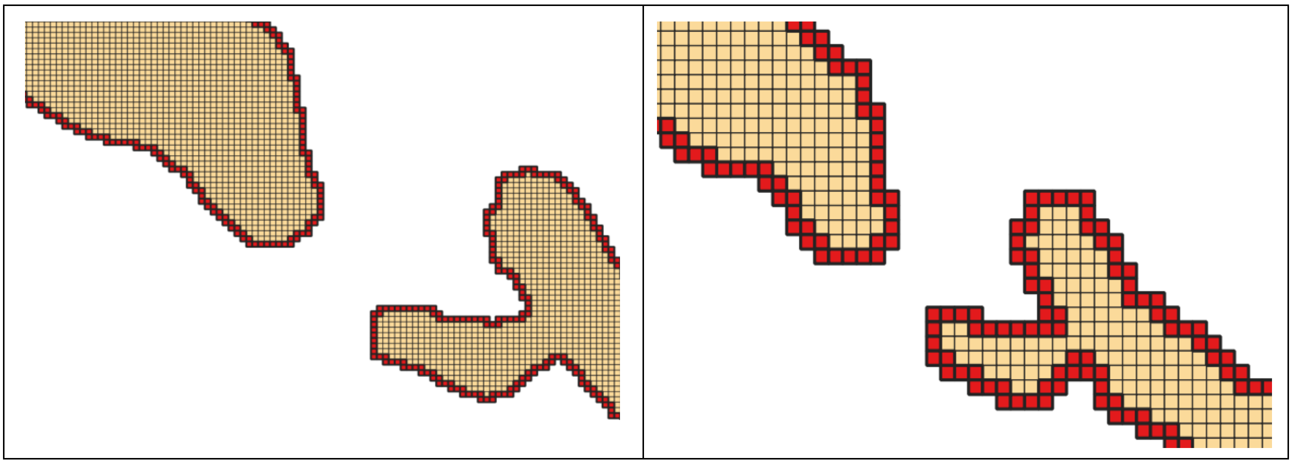


Figure 2.2.3.2 Bordering pixels (red) versus pixels inside the polygon (yellow) at 10 m (left) and 30 m (right) resolution

2.3 Method testing on the study case

For the purpose of testing out the proposal method, following processing and calculation are carried out on the selected area of Croatia's coastal zones. The focus of this work is to identify built-up areas along the defined coastal zones of Croatia. This task aims to assess the extent of urban development in the region of interest. The procedure is based on the Manual for IMAP Candidate Common Indicator 25 (Gilic et al., 2024). Main steps in calculation include: preparing the QGIS platform and selected plugins, downloading free and open-source land cover and other spatial data, preprocessing of the acquired data and finally, calculating and representing outputs. Calculating area, which is the most important result, represents a special problem due to complexity caused by Earth's curved surface. Curved surfaces cannot be projected on a flat device (paper or screen) without deformation. Therefore, all area calculations are based on the surface of the ellipsoid. This approach simplifies and increases accuracy of area calculation by avoiding problems connected with cartographic projection (Gilic at al., 2024). Although, ellipsoid also cannot perfectly represent Earth's surface, it is a closest mathematical approximation of the Earth. All measurements in this paper are done on the GRS 1980 ellipsoid (EPSG:7019).

For the purpose of testing the validation method, three following land cover data sources are used:

- ESA WorldCover Project Land Cover (ESA WCP),
- GLC_FCS30D and
- ESRI Sentinel-2 Land Cover

all for the referent year 2021. All three land cover sources are open and can be downloaded for free in the form of GeoTIFF raster tiles. Preparing LC data implies merging all tiles into one or also so-called mosaicking. When performing mosaicking of raster tiles virtual raster can be created. Virtual raster takes almost none of the storage space and is treated in QGIS as a regular raster, therefore is an optimal format choice. While the ESA WorldCover 2021 is delivered in 3 x 3 degree tiles, GLS_FCS30D is saved as 5 x 5 degree tiles. ESRI Sentinel-2 Land Cover follows the UTM system tiling, each tile having 6 degree longitude width. ESA WorldCover and ESRI Sentinel-2 Land Cover are both 10 m resolution datasets, whereas GLS_FCS30D has 30 m spatial resolution.

2.3.1 Study area

Area of interest is defined by the coastline of Croatia extracted from OpenStreetMap (OSM). Three coastal strips have been constructed as inland zones that are 0 - 300 m, 300 - 1000 m and 1 - 10 km away from coastline created as buffers from coastline. The coastal strips and their width are defined by ICZM Protocol (URL 5). Administrative units on the county level are downloaded from OSM and used in calculation. Finally, creating an intersection between administrative units and coastal strips produces a reporting unit. Reporting units are base units for calculation and processing of LC data. Reporting units are created for every county (7) and every coastal strip (3), resulting in a total of 21 reporting units.

Table 2.3.1.1 – Study area per county and coastal strip in km^2

	Area per coastal strip in km^2			
County	**300 m**	**1 km**	**10 km**	**Total area in km^2**
Dubrovnik-Neretva County	247,60	411,15	1061,54	1720,29
Istria County	117,53	172,01	1340,07	1629,61
Lika-Senj County	71,33	112,02	844,73	1028,08
Primorje-Gorski Kotar County	259,19	382,03	1282,28	1923,50
Split-Dalmatia County	257,38	375,53	1652,26	2285,18
Šibenik-Knin County	152,30	77,61	528,14	758,05
Zadar County	308,01	397,22	1363,62	2068,85
Ukupni zbroj	**1413,36**	**1927,57**	**8072,63**	**11413,56**

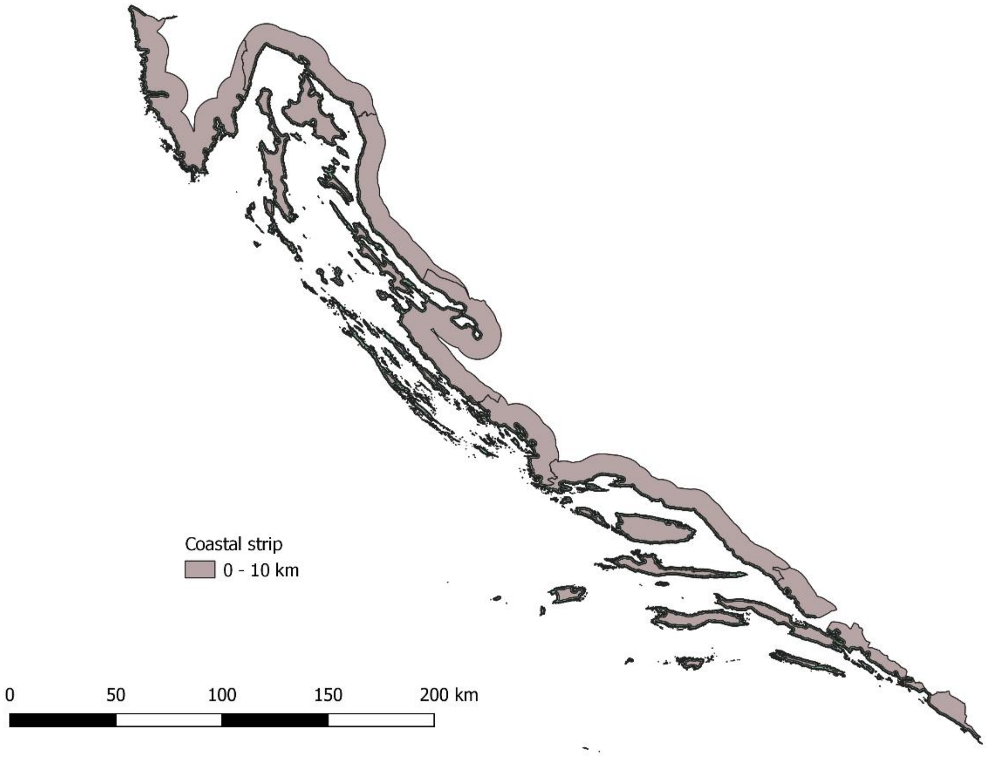


Figure 2.3.1.1 Study area

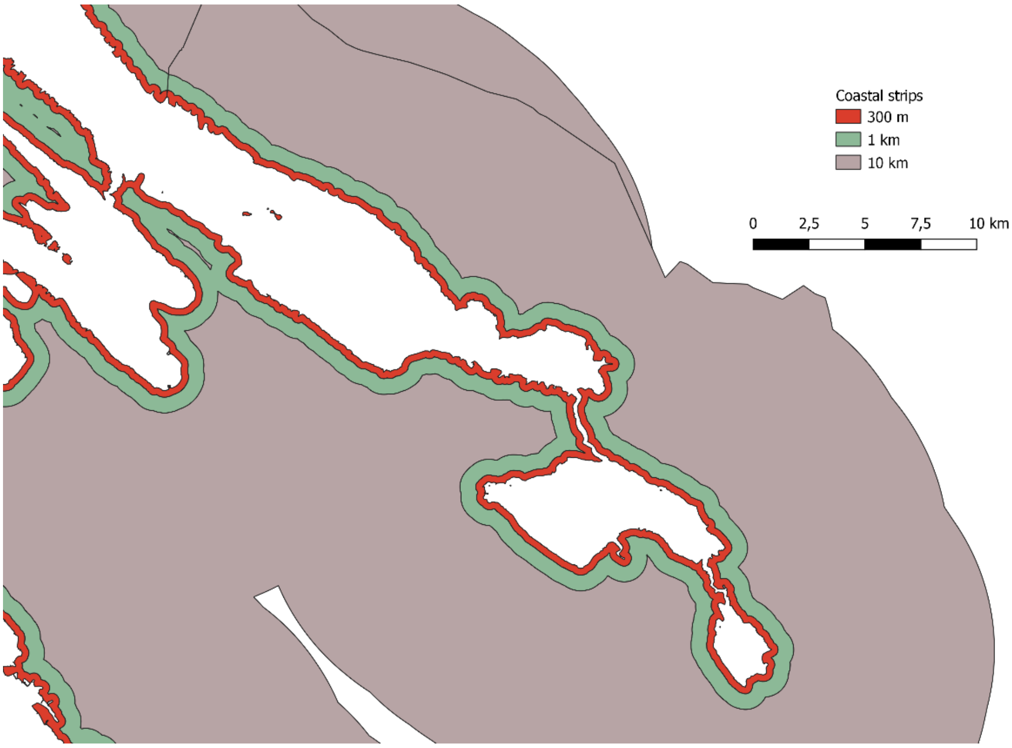
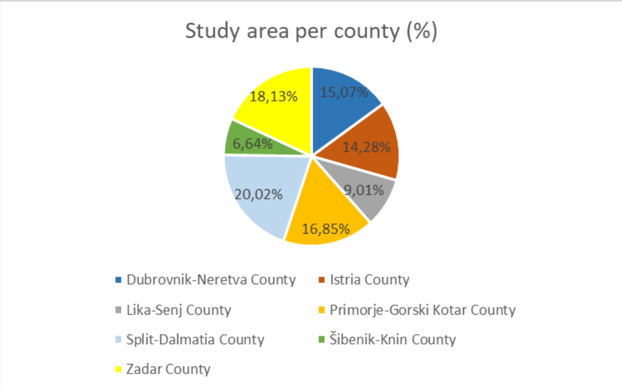


Figure 2.3.1.2 Study area with coastal strips



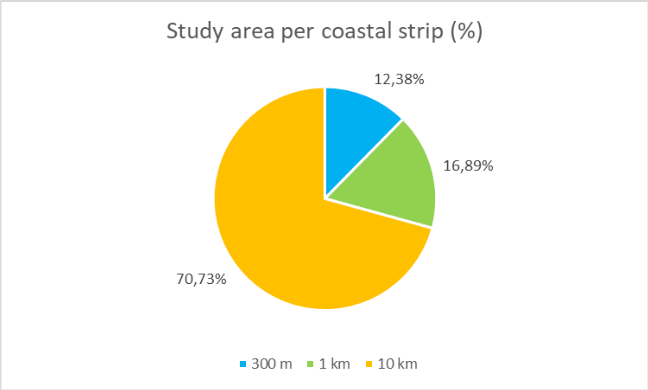


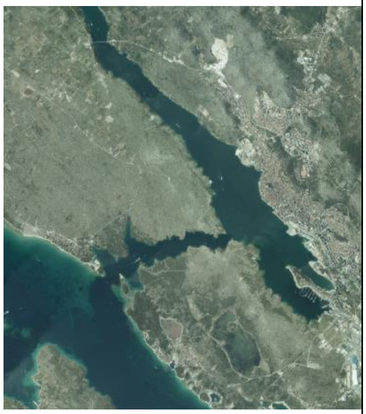
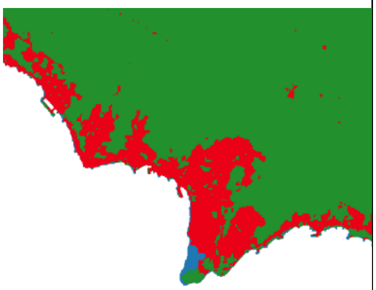
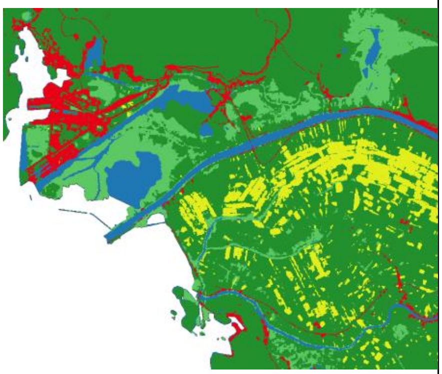
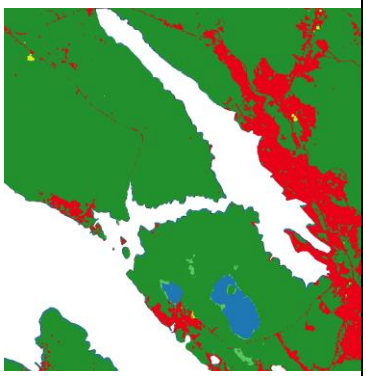
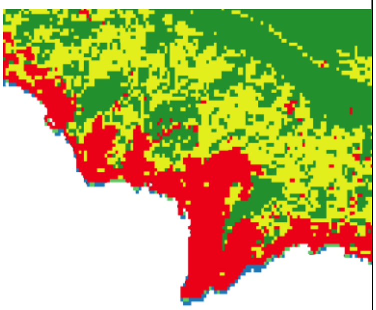
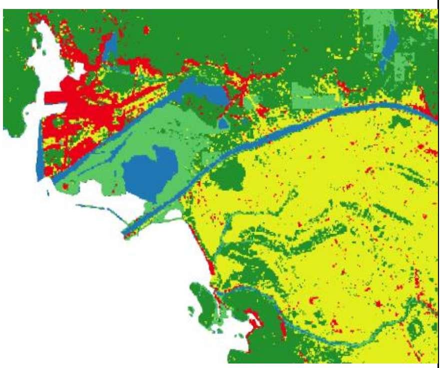
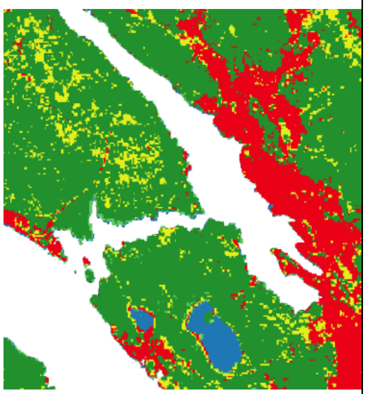
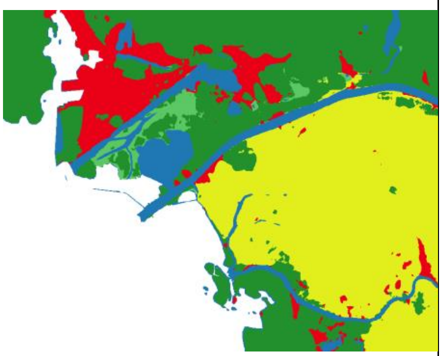
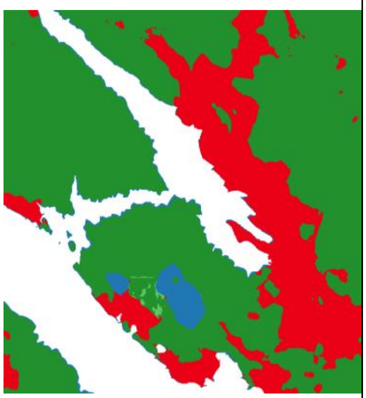
Figure 2.3.1.3 Study area per seven Adriatic county and per coastal strips in percentages

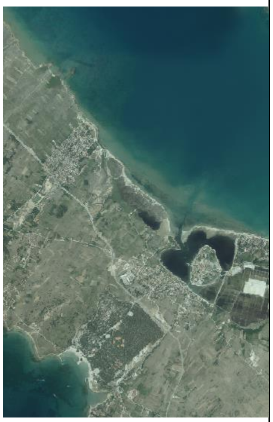
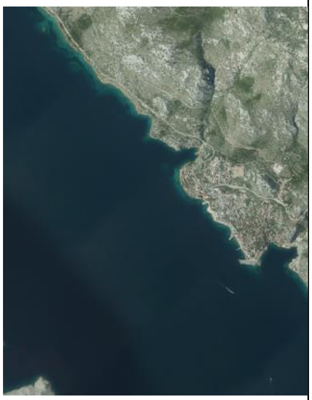
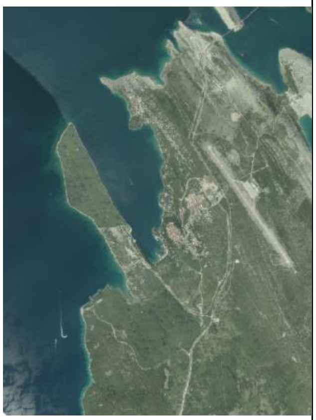
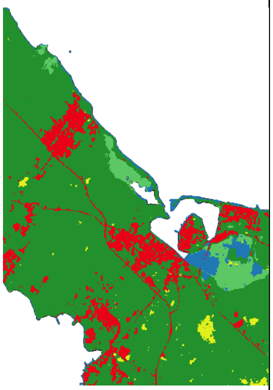
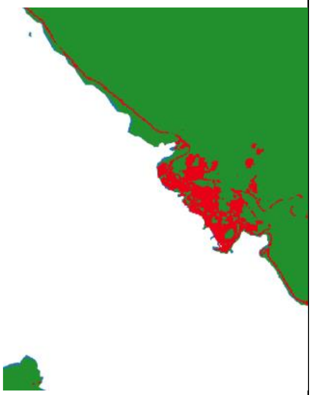
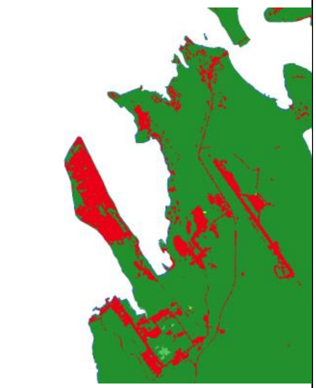
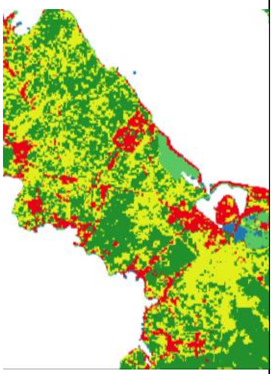
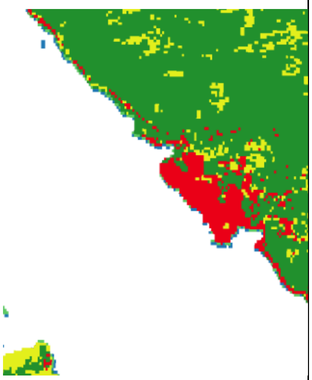
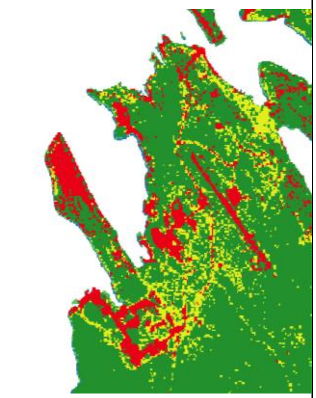
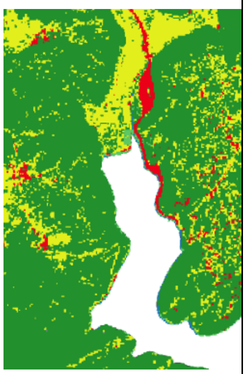
Study area consists of seven coastal counties, where the largest share refers to Split-Dalmatia County (over 20%) followed by Zadar County (over 18%). The smallest share refers to Šibenik-Knin County (over 6%). The narrowest coastal strip, which ranges from 0 to 300 meters in width, has the smallest width among coastal strips. However, it is important to note that although this strip is more than two times smaller in terms of width compared to the middle strip, which ranges from 300 to 1000 meters, it represents a roughly similar proportion of the total study area, whereas the middle strip represents only 16.89%, compared to 12.38% the narrowest strip represents (Figure 2.3.1.3). The middle coastal strip is significantly broader than the narrowest strip and thus expected to make up a larger portion of the total length of Croatia's coastline, but this is not a case due to indentedness of Croatia's coast (Figure 2.3.1.1).

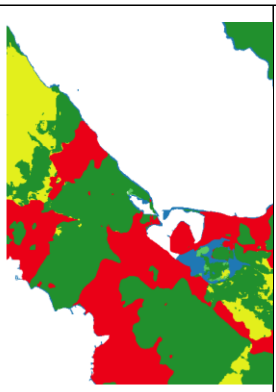
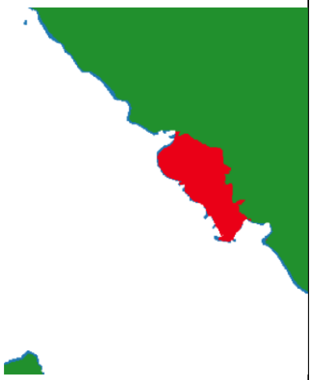
2.3.2 Visual inspection

Visual inspection of data is performed on the selected locations along Croatia's coastline. For each county one location is selected (Figure 2.3.2.1). Each tested land cover map is visually compared to an aerial image for 2021. Aerial images used in visual inspection are produced by the State Geodetic Administration of the Republic of Croatia. They are available as a part of The Digital Orthophoto Map (DOF), the official state orthophoto map produced at a scale of 1:5000 for the entire territory of the Republic of Croatia (URL 7).

Figure 2.3.2.1 Selected locations for visual inspection

Pl.	1. Dugi Rat, Split-Dalmatia County	2. Neretva Delta, Dubrovnik-Neretva County	3. Šibenik, Šibenik-Knin County
Aerial images			
ESA WorldCover			
GLC_FCS30D			
ESRI Sentinel-2 Land Cover			

Pl.	4. Nin, Zadar County	5. Karlobag, Lika-Senj County	6. Omišalj, Primorje-Gorski Kotar County	7. Raša bay, Istria County
Aerial images				
ESA WorldCover				
GLC_FCS30D				

ESRI Sentinel-2 Land Cover				

Validation by visual inspection revealed how significant the spatial resolution of the input data is. The ESA WorldCover with 10 m resolution always corresponds to the land cover on aerial images. GLC_FCS30D has a tendency to overestimate agricultural areas mostly where others assigned forest and seminatural. When it comes to urban areas, ESA WorldCover and GLC_FCS30D match each other better and the aerial image than Esri Sentinel-2 Land Cover. This is especially visible in location Omišalj where an airport is present. Both ESA WorldCover and GLC_FCS30D have successfully detected the airport runway as built-up, whereas on Esri Sentinel-2 Land Cover map the said runway is not recognizable.

Generally speaking, ESA WorldCover and Esri Sentinel-2 Land Cover respond highly differently to visual inspection. Although they share the same fine resolution (10 m), ESA WorldCover typically displays more detailed land cover features and better corresponds to aerial images, whereas Esri Sentinel-2 Land Cover has a tendency to group land cover types generalizing the area. ESA WorldCover's finer granularity and ability to preserve more subtle variations in land cover make it better aligned with high-resolution aerial images. As such, fine features like small bodies of water, urban infrastructure or roads can be better distinguished. Esri's approach relies on a more generalized classification scheme that simplifies the land cover into more aggregated classes. The land cover boundaries appear smoother and more continuous, potentially missing smaller or more complex features that are visible in ESA WorldCover. Therefore, the generalization in Esri Sentinel-2 Land Cover is suitable for applications where the primary concern is understanding overall land cover distribution on a larger scale, rather than focusing on the small details of each land cover type. Due to its higher level of detail, ESA WorldCover is more suitable for tasks that require fine-grained land cover classification, such as urban planning or monitoring. The high resolution and detailed classification can help identify smaller changes in land use, vegetation and other landscape features.

Overall, while both ESA WorldCover and Esri Sentinel-2 Land Cover use the same 10-meter resolution, ESA WorldCover offers more detailed and accurate representations of land cover, aligning more closely with aerial imagery. Esri Sentinel-2 Land Cover, by contrast, adopts a more generalized approach that is often less detailed but can still be useful for broader-scale land cover assessments. The choice between these two datasets ultimately depends on the specific needs of the user, such as whether detailed, fine-scale analysis or broader, aggregated land cover categories are required.

Further, both ESA WorldCover and GLC_FCS30D perform reliably in urban areas, with ESA WorldCover's finer resolution (10 m) providing a more accurate representation of built-up areas. Both datasets generally provide accurate representation of built-up areas, especially in regions with clearly defined infrastructure, such as cities or settlements. This makes them particularly useful for urban monitoring, where precise classification of built-up zones is critical. The higher resolution of ESA WorldCover allows it to capture smaller features and subtle variations in the landscape that may be missed by lower-resolution datasets like GLC_FCS30D. This makes ESA WorldCover particularly effective in areas with complex land cover types. While GLC_FCS30D responded well to visual inspection, due to its coarser resolution (30 m) it is more limited in its ability to capture detailed land cover variations, making the ESA WorldCover superior dataset by this inspection.

2.3.3 Area calculation

In processing LC data for calculating area two main functions are performed: clipping LC raster with reporting units and reclassification of the LC raster. Before clipping, the pixel size of the original LC raster is divided with 2, meaning that the output raster has spatial resolution 2 times higher than the resolution of the original raster and therefore contains 4 times more pixels. This is done to reduce errors in later calculations. Since not all land cover products use the same classification scheme, it is necessary to reclassify raster data. For output results to be comparable, the same classification rules need to be followed for different land cover data. Reclassification rules define how original classes of land cover data are transformed into five chosen categories. The selected categories used for this project are the following: built-up, agricultural, forest and semi-natural, wetlands and waterbodies. The chosen categories are based on requirements of ICZM Protocol (URL 5). Esri Sentinel-2 Land Cover classification system consists of 9 classes defined within the project, and ESA WorldCover's 10 classes are based on the UN-LCCS system. On the other hand, GLC_FCS30 contains 35 fine land cover types and subcategories.

Finally, the last step is to overlay raster LC data with the reporting units resulting in calculation of area of each pixel in the LC raster for corresponded LC class and sum for each reporting unit (Gilić, Baučić, 2024). Using a Raster calculator, areas of each LC class are calculated separately. After all processing that is done, all data and calculations are stored in each attribute table. Using Zonal statistics, areas of each LC class within each reporting unit are summed and exported in xlsx format.

2.3.4 Validation of GIS overlay

According to the proposed shape metrics method, users are able to calculate the Shape Index and use it to assess the compactness of a polygon that GIS operations are performed on, in the context of uncertainty introduced to calculations due to shape complexity. The index is calculated for polygons representing coastal strips of three widths inland of the coastline: 0 - 300 m, 300 - 1000 m and 1 - 10 km. The uncertainty level in area calculation performed on the coastal strips is based on the Shape Index. The Shape Index indicates complexity of the shape and thus area calculation uncertainty levels. The strip 0 - 300 m, containing all of the coastline including outlines of islands and rocks, is an extremely complex multipolygon and is expected to have Shape Index close to zero.

The figure below (2.3.4.1) shows coastal strips polygons with their visible complexity and thus area calculation complexity, particularly for the narrowest 0 - 300 m coastal strip in red.

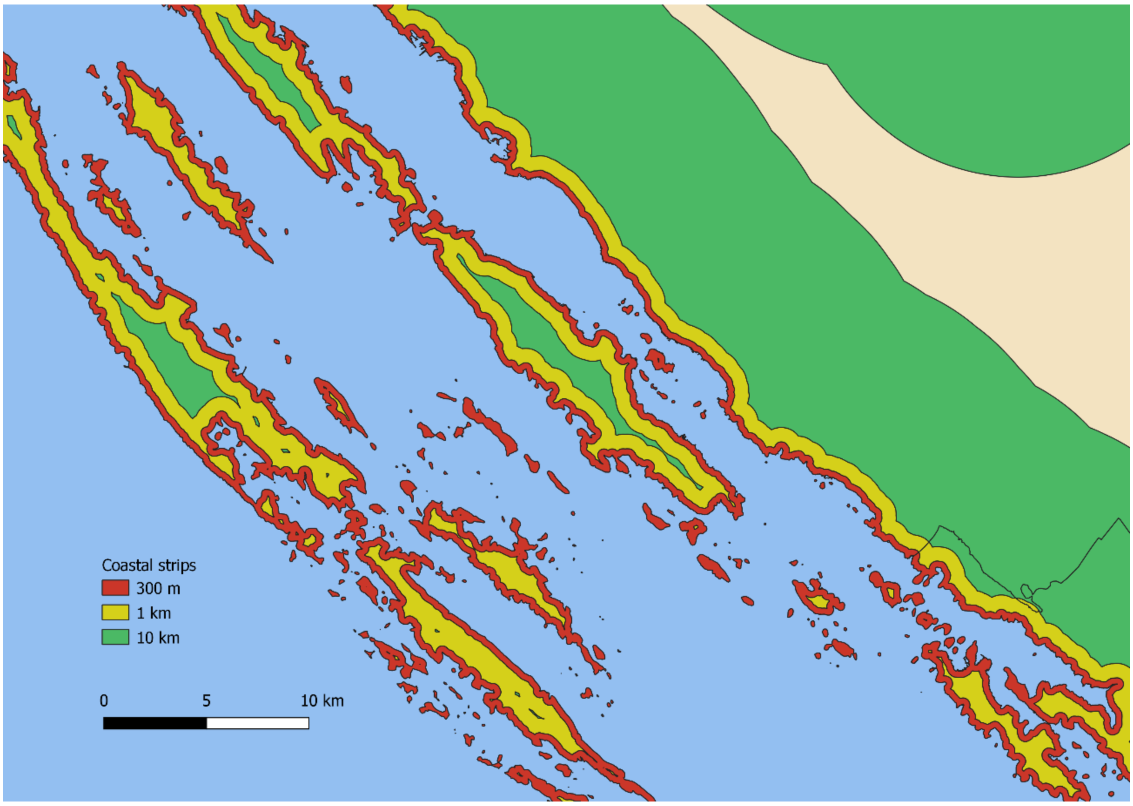


Figure 2.3.4.1 The cut out from map showing complexity of polygons representing coastal strips

RESULTS

3.1 Results of area calculation

Calculated areas of land cover classes on the coastal strip level are given in the following tables and graphs. A detailed overall statistic for each land cover dataset per reporting units (county and coastal strip) is given in the Annex. All calculations are refereeing on the year 2021.

Table 3.1.1 Land cover area in km^2 per coastal strips based on ESA WorldCover, 2021

Coastal strip	Built-up	Agricultural	Forest and seminatural	Wetlands	Water bodies	Total coastal zone
300 m	122,55	2,23	1244,75	3,27	40,56	1413,35
1 km	94,89	6,62	1817,77	4,04	4,25	1927,57
10 km	173,47	125,10	7677,72	34,53	61,81	8072,63
Total	**390,91**	**133,95**	**10740,24**	**41,84**	**106,62**	**11413,56**

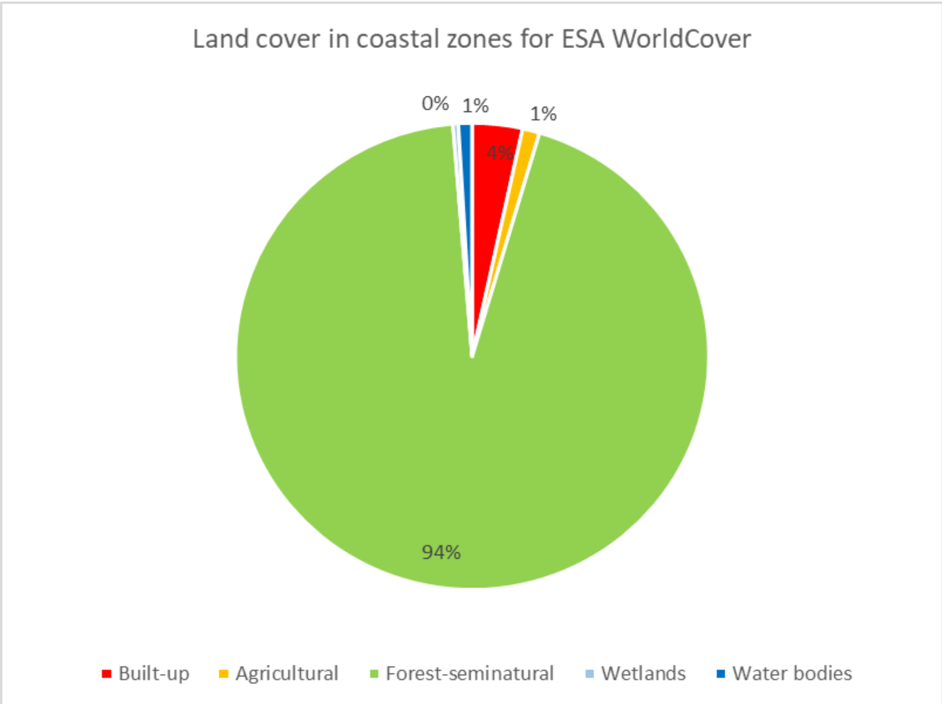


Figure 3.1.1 Land cover classes in study area based on ESA WorldCover, 2021

In the study area, forest and seminatural land dominates in the coastal zones with 94% followed by built-up with 4%. Agricultural areas occupy only 1% with the largest share in the 10 km strip. Wetlands and water bodies take out the smallest area around 1%. Table 3.1.1 provides detailed data per coastal strip for five land cover classes.

Table 3.1.2 Land cover area in km^2 per coastal strips based on GLC_FCS30D, 2021

Coastal strip	**Built-up**	**Agricultural**	**Forest and seminatural**	**Wetlands**	**Water bodies**	**Total coastal zone**
300 m	148,08	202,76	1021,59	27,27	13,66	1413,36
1 km	109,23	282,78	1529,38	3,01	3,03	1927,44
10 km	175,48	1425,77	6397,58	15,71	57,51	8072,05
Total	**432,79**	**1911,32**	**8948,54**	**45,99**	**74,21**	**11412,85**

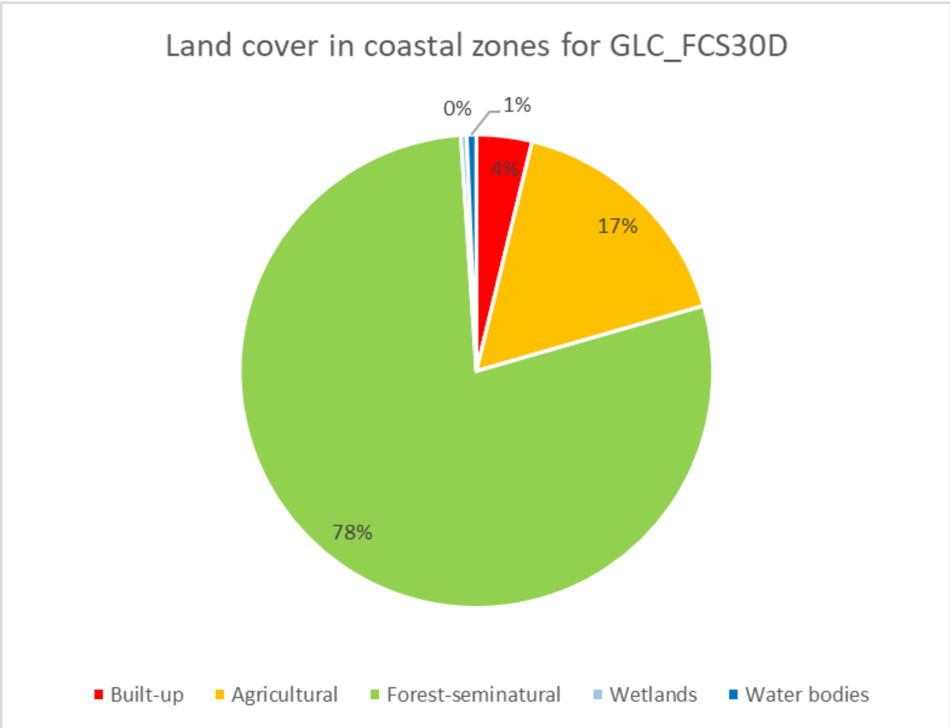


Figure 3.1.2 Land cover classes in study area based on GLC_FCS30D, 2021

In the study area, forest and seminatural land dominates in the coastal zones with 78% followed by agriculture with 17%. Built-up areas occupy 4%, the same as ESA WorldCover. Water bodies occupy 1% of the study area, followed by wetlands with under 1%. Table 3.1.2 provides detailed data per coastal strip for five land cover classes.

Table 3.1.3 Land cover area in km2 per coastal strips based on Esri Sentinel-2 Land Cover, 2021

Coastal strip	**Built-up**	**Agricultural**	**Forest and seminatural**	**Wetlands**	**Water bodies**	**Total coastal zone**
300 m	209,75	5,18	1108,09	0,75	88,87	1412,63
1 km	192,77	27,98	1699,80	0,98	5,10	1926,63
10 km	397,72	397,32	7195,50	10,52	67,35	8068,41
Total	**800,24**	**430,47**	**10003,39**	**12,25**	**161,32**	**11407,68**

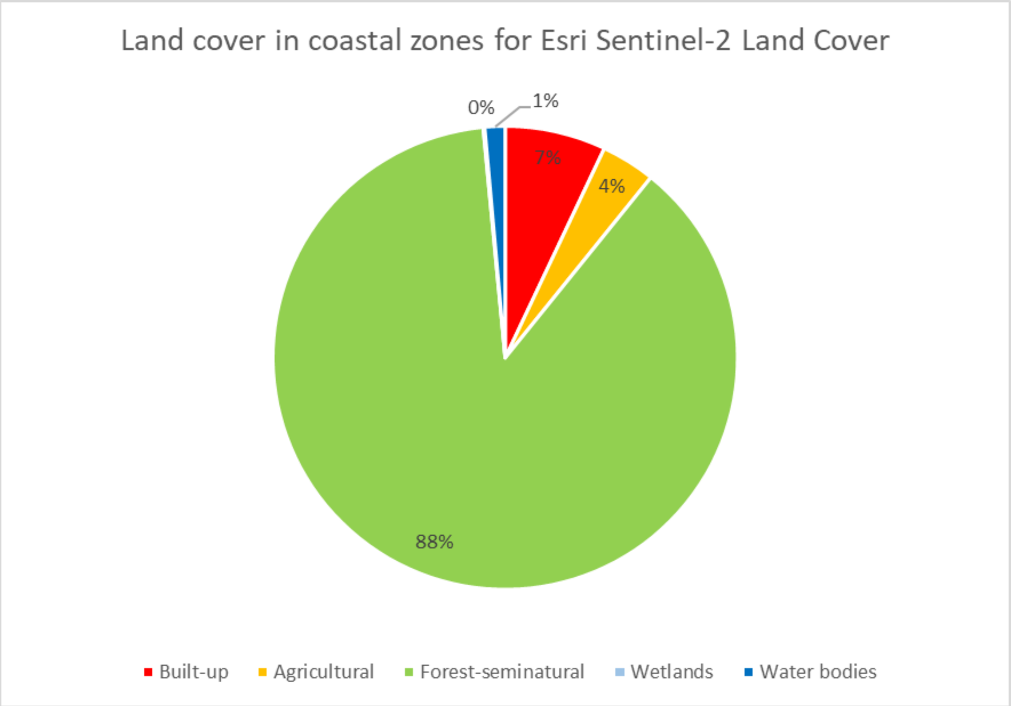


Figure 3.1.3 Land cover classes in study area coastal zones based on Esri Sentinel-2 Land Cover, 2021

In the study area, forest and seminatural land dominates in the coastal zones with 88% followed by built-up areas with 7%. Agriculture occupies 4%. Water bodies occupy 1% of the study area, followed by wetlands close to 0%. Esri Sentinel-2 Land Cover classifies more water bodies at the expense of wetlands. Table 3.1.3 provides detailed data per coastal strip for five land cover classes.

Comparing different land cover datasets is a difficult task because each land cover source is developed using different methodologies, resolutions, classification systems and underlying data sources. The following Figure 3.1.4 compares three land cover datasets tested in this thesis based on the area calculation of five major land cover classes in coastal zones of study area.

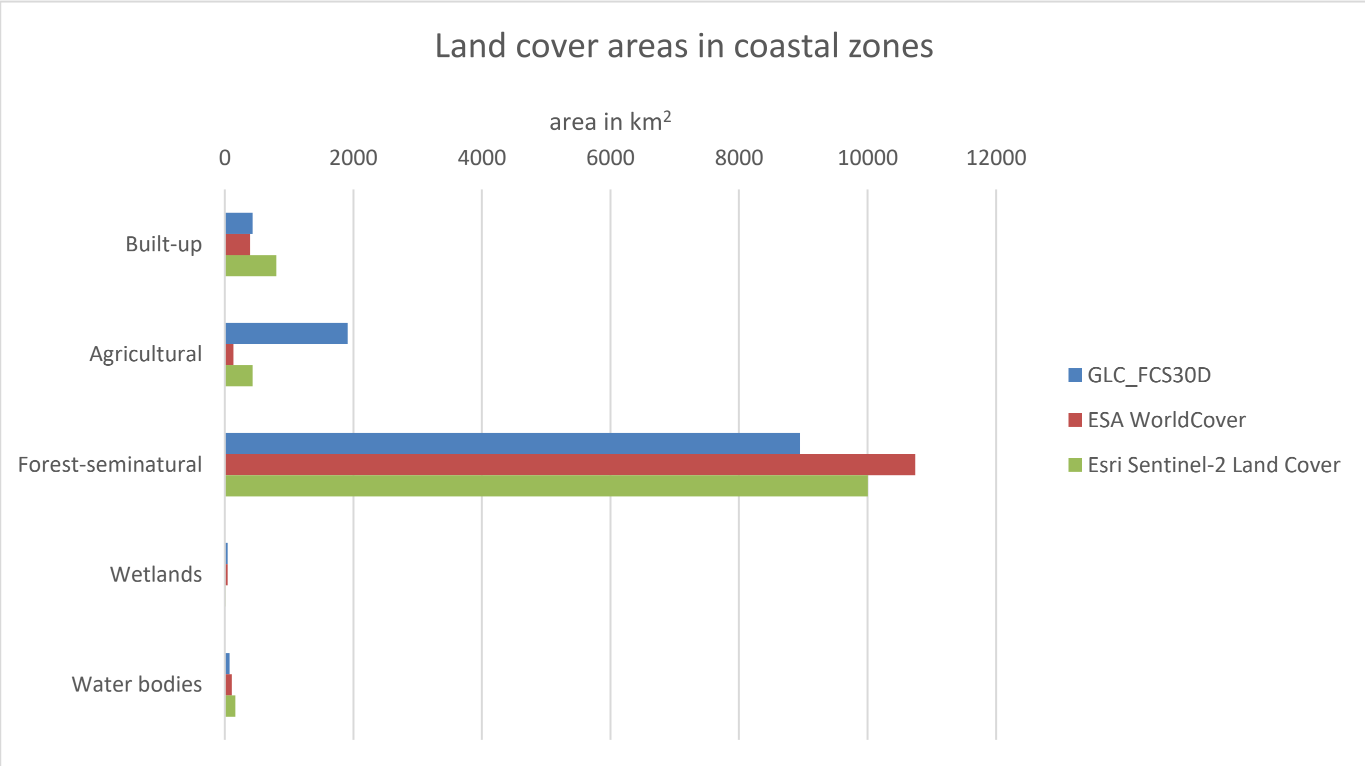


Figure 3.1.4 Land cover classes areas in km^2 in coastal zones in study area calculated for year 2021 from ESA WorldCover, GLC_FCS30D and Esri Sentinel-2 Land Cover

3.2 Results of Shape Index calculation

Another calculation conducted is the Shape Index. The Shape Index is calculated for each reporting unit in each coastal strip and given in the Annex 5. Average value of the index per coastal strip is given in Table 3.2.1.

The strip 0 - 300 m, containing all of the coastline including outlines of islands and rocks, is an extremely complex multipolygon as confirmed by low Shape Index for all reporting units in that strip (Annex 5), with average SI of 0.0015 (Table 3.2.1). As proposed by the method, this implicates high uncertainties in area calculation for that strip. Middle coastal strip 300 m - 1000 m has an average SI 0.0074 (Table 3.2.1) implying significant but still lower uncertainty levels in area calculation than for the narrowest strip. The last strip 1 - 10 km performed best with average SI of 0.0871 (Table 3.2.1) meaning area calculations in this strip contain lowest level of uncertainty among the three strips.

Table 3.2.1 Average Shape Index per coastal strip

Coastal strip	Average Shape Index
300 m	0,0015
1 km	0,0074
10 km	0,0871
Total average	0,0320

DISCUSION

Purpose and goal of this exploration is to give a user set of simple and useful steps they can perform themselves to validate land cover data in context of their intended purpose. These steps include both pragmatic approach and quantitative calculations. Quantitative calculations in this thesis are divided into two main tasks. First is to calculate area covered by each land cover class per reporting unit. The second task involves evaluating the shape of the reporting unit using proposed shape metrics for estimation of uncertainties in area calculation.

Based on area calculation comparison of three tested land cover sets on Figure 3.1.4, the largest absolute difference is expressed in agricultural areas, where, according to GLC_FCS30D, four times more agriculture was recorded than according to the Esri Sentinel-2 Land Cover, and fourteen times more than according to ESA WorldCover. On the other hand, both Esri Sentinel-2 Land Cover and ESA WorldCover report a larger area of forest and semi-natural land compared to GLC_FCS30D. This implies that in areas where GLC_FCS30D identifies agricultural land, both ESA and Esri classify the land as forest and seminatural. Such a trend is likely a consequence of GLC_FCS30D initial fine resolution classification of 35 land cover types and its ability to differentiate a greater number of land cover types. The fine-resolution classification used by GLC_FCS30D allows for more precise differentiation between agricultural and natural land. GLC_FCS30D distinguishes 10 different forest types and divides cropland into rainfed, irrigated, herbaceous and trees or shrubs cropland covers (Zhang et al., 2024). Understanding these differences is essential for accurate land use and land cover analysis, particularly in regions where agriculture and natural land are closely interspersed.

Another large difference is reflected in built-up areas, where Esri Sentinel-2 Land Cover records double the number of urban areas compared to both ESA WorldCover and GLC_FCS30D. The primary difference between Esri Sentinel-2 and ESA WorldCover/GLC_FCS30D lies in their approach to land use versus land cover. Due to Esri Sentinel-2 emphasis on land use, it includes more areas in the built-up type, considering spaces such as parks and residential yards as part of the urban landscape because of their function in human settlements, even if these areas contain natural features. In contrast, ESA WorldCover and GLC_FCS30D focus on land cover, classifying these areas based on physical characteristics like trees or grass, which results in their classification as forest and seminatural land types, rather than built-up. This difference in classification criteria explains why Esri Sentinel-2 Land Cover records double the number of urban areas compared to ESA WorldCover and GLC_FCS30D – a result of the varying definitions of what constitutes built-up land versus natural land.

Figure 4.1 illustrates how different definitions of urban areas affect the outcome result on example of city Split. As noticed, ESA WorldCover classifies urban greens – parks, yards and gardens as natural land, whereas Esri Sentinel-2 generalizes all those features under the built-up area. Esri Sentinel-2's focus on land use results in a more inclusive definition of urban area. Areas that Esri Sentinel-2 considers urban (like parks or yards) are majorly classified as natural by ESA WorldCover. This highlights the impact of the classification criteria on the mapping of urban versus natural areas in addition to pointing out the importance of correctly interpreting land cover and land use data. For users, depending on their specific objectives, especially when dealing with specific themes like urban planning and environmental monitoring, this information is crucial as it directly influences the accuracy and relevance of their analysis.

Figure 4.1 Split city according to ESA WorldCover (left) versus according to Esri Sentinel-2 Land Cover (right) classification

Based on Shape Index calculation (Annex 5), all reporting units report disadvantageous shape indices, pointing out that all strips have extremely complex shapes. The Shape Index is in correlation with coastal strip width; the index is lowest (meaning the most irregular shape) for narrowest coastal strip (0 - 300 m), moderate for middle strip (300 - 1000 m) and most convenient for the broadest strip (1 - 10 km). All reporting units in the narrowest coastal strips record the most inconvenient Shape Index with minimum SI = 0,0008 for Zadar County and maximum SI = 0,0033 for Lika-Senj County. This occurs due to the extremely indented coastline of Croatia, with numerous islands and bays that extend the coastline, creating irregular and elongated shapes. Therefore, as stated in the previous chapter, this is a strong indicator of presence of bordering pixels. The middle coastal strip has an inconvenient Shape Index, though not as problematic as the first strip. Lastly, the broadest coastal strip records a relatively convenient Shape Index. In other words, results of GIS overlay for the reporting units in the first coastal strip include more uncertainties than for the second and the third coastal strip.

Using the proposed method, the Shape Index assesses the compactness of a reporting unit in relation to the uncertainty introduced during GIS operations due to the appearance of bordering pixels. In addition to the validation performed by data producers, users should also report on the validation process and uncertainty estimation for the analyses they conduct using open land cover maps. This ensures that the limitations of the results are clear to policymakers.

A conclusion about uncertainty level introduced in area calculation by GIS operation of overlapping for the narrowest coastal strip is as follows; at finer spatial resolution (10 m), uncertainty level is medium, whereas at coarser resolution (30 m) uncertainty level is high. For the middle coastal strip, at finer spatial resolution (10 m), uncertainty level is small, whereas at coarser resolution (30 m) uncertainty level is medium. For the broadest coastal strip, both at finer spatial resolution (10 m) and coarser resolution (30 m) uncertainty level introduced by overlay is small.

The discussion about the validation of land cover data in the context of specific user needs is crucial for ensuring the accuracy and reliability of the results derived from land cover datasets. Open global and regional land cover maps are increasingly used in a variety of applications, from monitoring environmental changes to informing policy decisions. However, for these maps to be truly useful, users must be able to assess their appropriateness and the uncertainties inherent in the data. The approach outlined in this work emphasizes the importance of providing users with a clear, step-by-step method for validating the data. This is vital because land cover maps, while valuable, often come with a degree of uncertainty due to factors like varying data quality, resolution and the methods used in the process. By incorporating both pragmatic approaches and quantitative methods, users can gain a deeper understanding of how well the data fits their specific needs. The step-by-step method is illustrated in Figure 4.2.

All validated land cover datasets are of great quality and highly appropriate and valuable for a variety of needs. However, when it comes to urban detection, based on the steps of the proposed validation method, it can be concluded that the most appropriate land cover dataset out of the three tested is ESA WorldCover. The ESA dataset demonstrated superior performance in terms of spatial resolution and classification criteria, which made it the most suitable for the specific area and objectives of the study. Furthermore, the selected dataset closely corresponds to aerial imagery used as higher quality data, reinforcing its accuracy and relevance for the analysis.

Area calculations revealed that, among the tested datasets, ESA WorldCover is the most appropriate for tracking built-up areas and urban environments. The ESA dataset demonstrated superior accuracy in identifying urban areas, making it particularly effective for monitoring urban growth, land use changes and the expansion of built-up regions. Its higher spatial resolution and classification criteria consistent to land cover allowed for more precise measurements of urban areas, making the ESA dataset a valuable source for urban monitoring.

At ESA WorldCover's fine 10 m resolution, the percentage of bordering pixels is reduced, which results in a more accurate representation of the polygon's boundaries. This improvement leads to lower uncertainties in area calculations, as smaller pixel sizes allow for more precise outline of land cover types, especially in complex or irregularly shaped polygons like reporting units of study area. Fine resolution data ensure that the calculated area more closely matches the actual extent of polygons being analyzed. Therefore, the level of uncertainties introduced by GIS overlay into results with ESA WorldCover is small for all coastal strips tested.

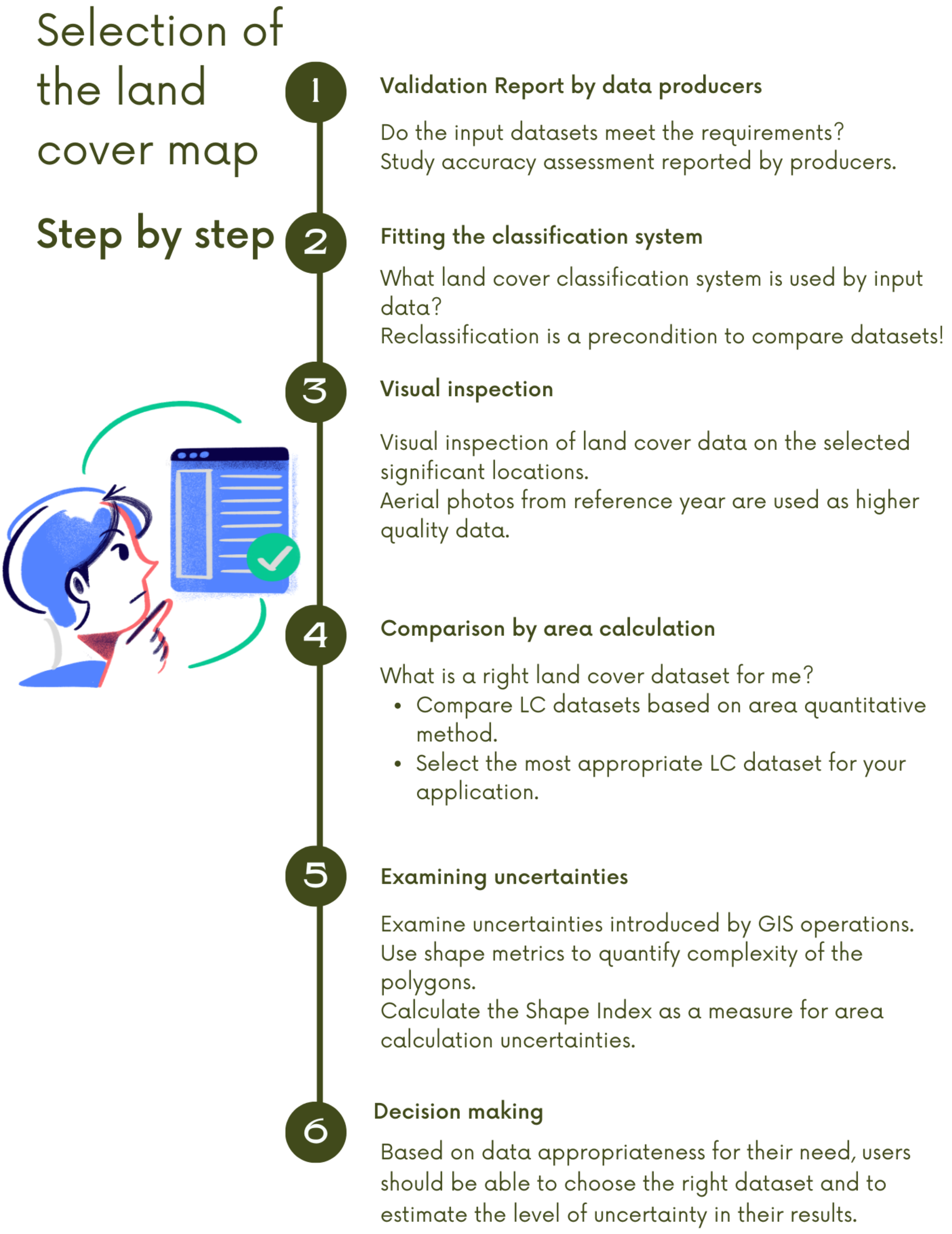


Figure 4.2 Proposed method for selecting land cover map step-by-step

CONCLUSION

The aim of this work is to propose a simple yet effective method that users can apply to assess the appropriateness of open land cover data for their specific needs and to estimate the level of uncertainty in their results. The proposed approach is designed to be accessible and practical for users with varying levels of expertise, while still providing insights into the quality and reliability of the data.

Firstly, users should focus on assessing the quality of the input data before performing any GIS analysis. It involves examining the accuracy, resolution and appropriateness of the open land cover maps for the intended purpose. Users should verify whether the spatial resolution and classification scheme align with the objectives of the analysis. Next, every step in a GIS workflow, from data input to processing and analysis, introduces the potential for errors or uncertainties. These uncertainties can stem from a variety of factors, including data inaccuracies, spatial resolution limitations or errors in the methodology used to perform the GIS operations. Understanding how these uncertainties propagate through GIS operations is essential for ensuring the reliability and accuracy of the results.

By offering a structured, user-friendly validation framework, this work enables users to assess open data appropriateness for their need and to estimate the level of uncertainty in their results. Based on the user's specific needs, this method will help users choose the land cover data source that best aligns with their objectives. The choice of source depends on factors such as the required spatial resolution, temporal coverage and the level of detail needed for the intended analysis. By assessing the suitability of the data through a step-by-step approach, users can ensure they are working with the most reliable and relevant information for their goals.

The first quantitative task—calculating the area covered by each land cover class per reporting unit for all datasets considered, allows users to quantify land cover distribution in relation to their specific area of interest. This step provides a clear, objective measure of land cover that can be used in decision process. The second task—evaluating the shape of the reporting unit through shape metrics adds an additional layer of validation by assessing the spatial configuration of land cover within each reporting unit. Shape metrics can reveal important insights into the spatial patterns of polygons, such as elongation or irregularity, which are essential for estimating level uncertainties in the results. The combination of these two tasks provides a comprehensive framework for validating land cover data from user perspective. Ultimately, the goal is to empower users to confidently interpret land cover data, while also understanding the limitations and uncertainties associated with it. By providing a clear method for validating data, this work ensures that users are better equipped to make informed decisions based on reliable, accurate information. Ensuring that users can make well-informed decisions based on reliable data is essential for understanding data quality and uncertainties, which contributes to more informed and effective decision-making. When users have access to validated, high-quality land cover data, they are better positioned to implement strategies for responsible decision-making that benefits both ecosystems and communities.

Selecting the right land cover data source becomes a pivotal decision point because it directly influences the effectiveness and accuracy of the analysis. The user's objective—whether it's monitoring land cover changes, supporting policy decisions or conducting ecological research—will determine which data source aligns best with their requirements in terms of spatial resolution, temporal coverage and data accuracy. By providing users with the tools to assess and compare multiple datasets, they can make informed decisions that enhance the accuracy and relevance of their analyses, thus improving the outcomes of their research and projects. Furthermore, this approach encourages transparency, as users will have a clearer understanding of how uncertainties may affect their results.

The way forward to validation of land cover maps from user perspective is adaptation of more detailed and comprehensive qualitative methods for determining uncertainties for other GIS operation (beyond just overlay analysis), which are frequently used when raster data (such as satellite imagery) is analyzed with vector data (such as reporting units). For example, examining uncertainty propagation in data processing and analysis for interpolation functions of continuous phenomena. Additionally, the future methods ought to develop new indices or ratios that could serve as measures of uncertainty.

REFERENCES

Baučić, M., Morić Španić, A., Gilić, F. (2022a): Upgraded LCC Indicator 25 proposal, Report, PAP/RAC, Split, Croatia

Baučić M., Morić Španić A., Gilić, F. (2022b): Validation of testing results for upgraded LCC Indicator 25 in pilot areas, Report, PAP/RAC, Split, Croatia

Brown C.F., Brumby S.P., Guzder-Williams B., Birch T., Hyde S.B., Mazzariello J., Czerwinski W., Pasquarella V.J., Haertel R., Ilyushchenko S., et al. (2022): Dynamic world, near real-time global 10 m land use land cover mapping, https://doi.org/10.3390/rs14164101

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

Büttner, G., Kosztra, B., Maucha, G., Pataki, R., Kleeschulte, S., Hazeu, G. W., Vittek, M., Schroder, C., & Littkopf, A. (2021): Copernicus Land Monitoring Service - CORINE Land Cover, User Manual, Copernicus Publications, https://land.copernicus.eu/user-corner/technical-library/clc-product-user-manual/view

Das, G.G. (2014): Land Use, Land Cover, and Food-Energy-Environment Trade-Off: Key Issues and Insights for Millennium Development Goals, Encyclopedia of Agriculture and Food Systems, Academic Press, 114-133, https://doi.org/10.1016/B978-0-444-52512-3.00081-4

De Petris, S., Sarvia, F., Borgogno-Mondino, E. (2024): About polygon area uncertainty in GIS and its implications on agro-forestry estimates, Ecological Informatics, Volume 81

Food and Agriculture Organization/United Nations Environmental Programmed (FAO/UNEP) (1999): Terminology for Integrated Resources Planning and Management

Forman, R.T.T. and Godron, M. (1986): Landscape Ecology, John Wiley and Sons Ltd., New York

García-Álvarez, D., Nanu, S., Camacho Olmedo, M.T., Paegelow, M., Mas, J.F. (2022): Land Use Cover Datasets and Validation Tools, Springer Cham, https://doi.org/10.1007/978-3-030-90998-7

Gilić, F., Baučić, M. (2024): Manual for calculating IMAP Common Indicator 25, Report, PAP/RAC, Split, Croatia

Gilić, F., Gašparović, M., Baučić, M. (2023): Current state and challenges in producing large-scale land cover maps: review based on recent land cover products, Geocarto International, 38(1), https://doi.org/10.1080/10106049.2023.2242693

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M. and Brumby, S. P. (2021): Global land use / land cover with Sentinel 2 and deep learning, IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, doi: 10.1109/IGARSS47720.2021.9553499

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock C.E., Wulder, M.A. (2014): Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment 148, 42–57

Pontius, R.G., Boersma, W., Castella, JC. et al. Comparing the input, output, and validation maps for several models of land change. Ann Reg Sci 42, 11–37 (2008). https://doi.org/10.1007/s00168-007-0138-2

Tsendbazar N., Xu, P., Herold, M., Lesiv M., Duerauer, M. (2022): WorldCover product validation report, https://worldcover2021.esa.int/data/docs/WorldCover_PVR_V2.0.pdf

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O., (2022): ESA WorldCover 10 m 2021 v200, https://doi.org/10.5281/zenodo.7254221

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J. (2021): GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery, Earth Syst. Sci. Data, 13, 2753–2776, https://doi.org/10.5194/essd-13-2753-2021

Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., and Liu, L. (2024): GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method, Earth Syst. Sci. Data, 16, 1353–1381, https://doi.org/10.5194/essd-16-1353-2024

LIST OF URLS

URL 1: https://www.fao.org/land-water/home/en/ (12.12.2024.)

URL 2: https://earthobservations.org/solutions (12.12.2024.)

URL 3: https://www.fao.org/sustainable-development-goals-data-portal/data/indicators/1531-proportion-of-land-that-is-degraded-over-total-land-area/en (12.12.2024.)

URL 4: https://www.unep.org/unepmap/what-we-do/monitoring-and-assessments (1.12.2024.)

URL 5. https://www.unep.org/unepmap/who-we-are/contracting-parties/iczm-protocol (1.12.2024.)

URL 6: https://livingatlas.arcgis.com/landcover-2050/ (10.1.2025.)

URL 7. https://registri.nipp.hr/izvori/view.php?id=1513 (15.1.2025.)

FIGURES

Figure 1.1.1.1 Global land cover layers viewer for 2019 https://lcviewer.vito.be/ 7
Figure 1.1.2.1 ESA WorldCover viewer for 2021 https://esa-worldcover.org/en 8
Figure 1.1.3.1 CORINE Land Cover viewer for 2018 https://land.copernicus.eu/en/map-viewer.... 11
Figure 1.1.4.1 Dynamic World Land Cover viewer https://dynamicworld.app/explore/ 13
Figure 1.1.5.1 Esri Land Cover viewer https://livingatlas.arcgis.com/landcoverexplorer 14
Figure 1.1.6.1 GLC_FCS30D workflow (Liu and Zhang, 2024) 16
Figure 1.1.6.2 Overview of GLC_FCS30D with classes (Liu and Zhang, 2024)............................. 16
Figure 1.2.1.1 Copernicus4GEOGLAM viewer of crop types in Kenya https://hsm.land.copernicus.eu/ 19
Figure 1.2.5.1 Land Cover Vulnerability Change 2050 prediction by Esri and Clark University 23
Figure 2.2.2.1 The narrowest coastal strip of 300 m width and the ESA WorldCover with spatial resolution of 10 m 34
Figure 2.2.2.2 The narrowest coastal strip of 300 m width and the GLC_FCS30D with spatial resolution of 30 m 34
Figure 2.2.3.1 Škabrnja and Murter-Kornati settlements.. 37
Figure 2.2.3.2 Bordering pixels (red) versus pixels inside the polygon (yellow) at 10 m (left) and 30 m (right) resolution.. 40
Figure 2.3.1.1 Study area 42
Figure 2.3.1.2 Study area with coastal strips 42
Figure 2.3.1.3 Study area per seven Adriatic county and per coastal strips in percentages 43
Figure 2.3.2.1 Selected locations for visual inspection... 44
Figure 2.3.4.1 The cut out from map showing complexity of polygons representing coastal strips . 49
Figure 3.1.1 Land cover classes in study area based on ESA WorldCover, 2021 50
Figure 3.1.2 Land cover classes in study area based on GLC_FCS30D, 2021 51
Figure 3.1.3 Land cover classes in study area coastal zones based on Esri Sentinel-2 Land Cover, 2021.. 52
Figure 3.1.4 Land cover classes areas in km^2 in coastal zones in study area calculated for year 2021 from ESA WorldCover, GLC_FCS30D and Esri Sentinel-2 Land Cover... 53
Figure 4.1 Split city according to ESA WorldCover (left) versus according to Esri Sentinel-2 Land Cover (right) classification... 56
Figure 4.2 Proposed method for selecting land cover map step-by-step 59

TABLES

Table 1.1.3.1 The evolution of CORINE Land Cover 10

Table 2.2.3.1 Shape Index for two chosen settlements with share of overlapping pixels at finer (10 m) and coarser (30 m) resolution 39

Table 2.3.1.1 Study area per county and coastal strip in km^2 41

Table 3.1.1 Land cover area in km^2 per coastal strips based on ESA WorldCover 49

Table 3.1.2 Land cover area in km^2 per coastal strips based on GLC_FCS30D 50

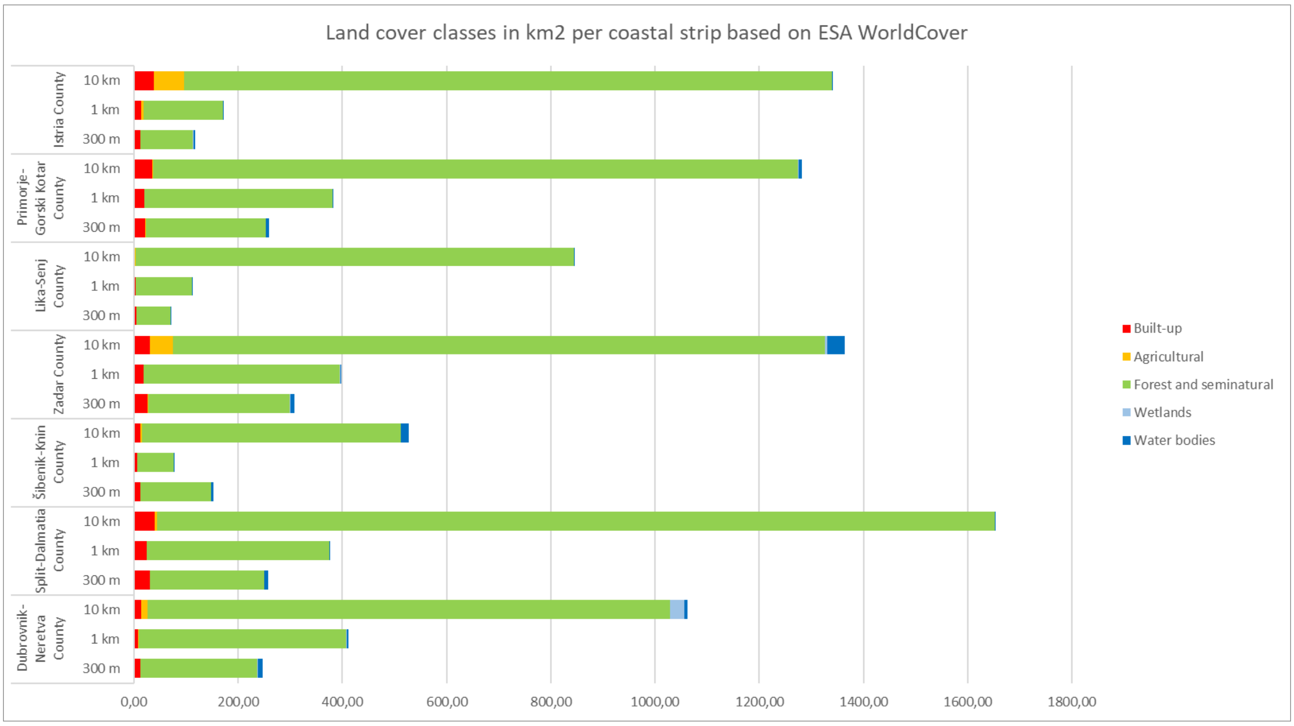
Table 3.1.3 Land cover area in km^2 per coastal strips based on Esri Sentinel-2 Land Cover 51

Table 3.2.1 Average Shape Index per coastal strip 53

ANNEX

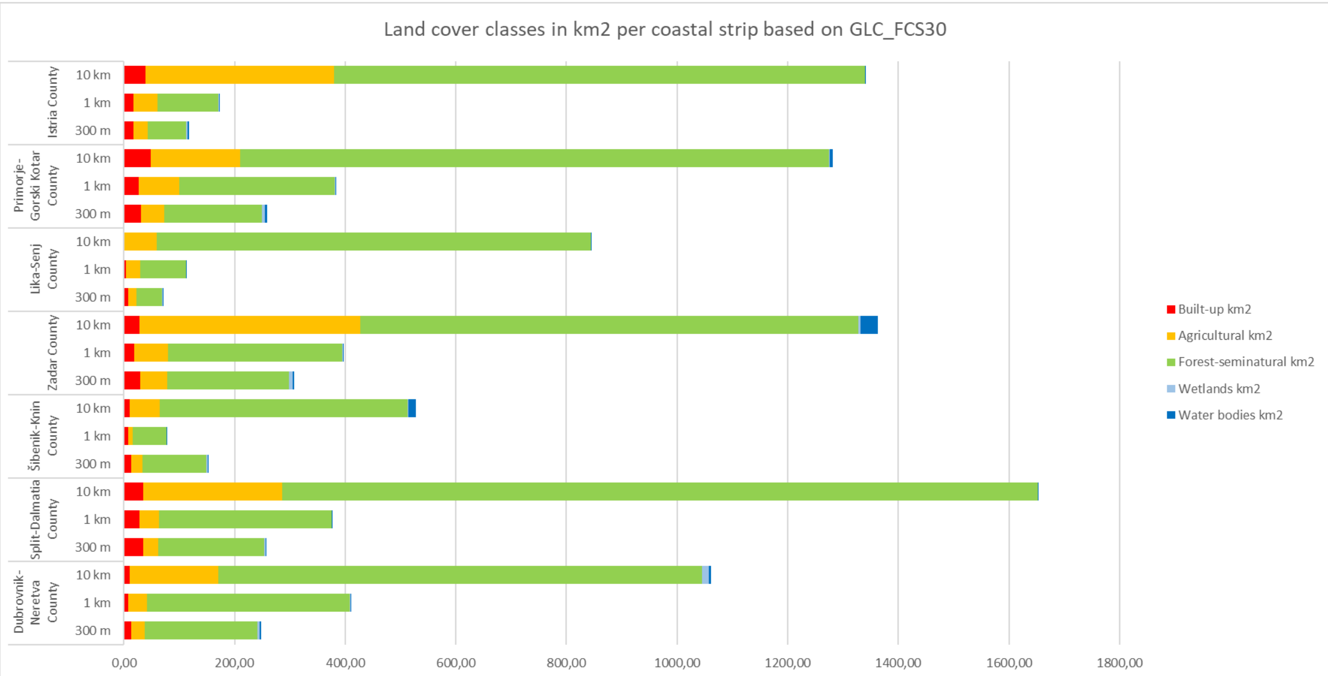
Annex 1 – Areas of land cover classes in km^2 per county based on ESA WorldCover, 2021

County	Coastal strip	Built-up	Agricultural	Forest and seminatural	Wetlands	Water bodies	Total coastal zone
Dubrovnik-Neretva County	300 m	12,71	0,16	224,24	1,11	9,38	247,60
	1 km	8,07	0,70	398,98	1,62	1,77	411,15
	10 km	14,29	12,56	1001,28	27,32	6,09	1061,54
Split-Dalmatia County	300 m	30,66	0,26	218,66	0,18	7,62	257,38
	1 km	24,40	0,89	350,14	0,01	0,10	375,53
	10 km	39,22	5,24	1607,19	0,05	0,56	1652,27
Šibenik-Knin County	300 m	12,58	0,05	135,12	0,16	4,39	152,30
	1 km	6,85	0,09	69,68	0,12	0,86	77,61
	10 km	12,83	3,04	496,39	0,45	15,42	528,14
Zadar County	300 m	26,84	0,51	271,39	0,99	8,27	308,01
	1 km	18,41	0,58	375,67	1,62	0,93	397,23
	10 km	31,27	43,95	1250,54	5,17	32,68	1363,61
Lika-Senj County	300 m	5,41	0,00	64,54	0,06	1,31	71,33
	1 km	2,86	0,04	109,12	0,00	0,00	112,02
	10 km	2,31	0,58	841,83	0,00	0,01	844,73
Primorje-Gorski Kotar County	300 m	22,41	0,40	230,03	0,26	6,10	259,20
	1 km	19,62	0,18	362,13	0,06	0,03	382,03
	10 km	35,69	1,31	1237,78	0,68	6,82	1282,28
Istria County	300 m	11,94	0,84	100,78	0,49	3,49	117,53
	1 km	14,68	4,14	152,04	0,60	0,54	172,01
	10 km	37,85	58,42	1242,71	0,85	0,23	1340,07



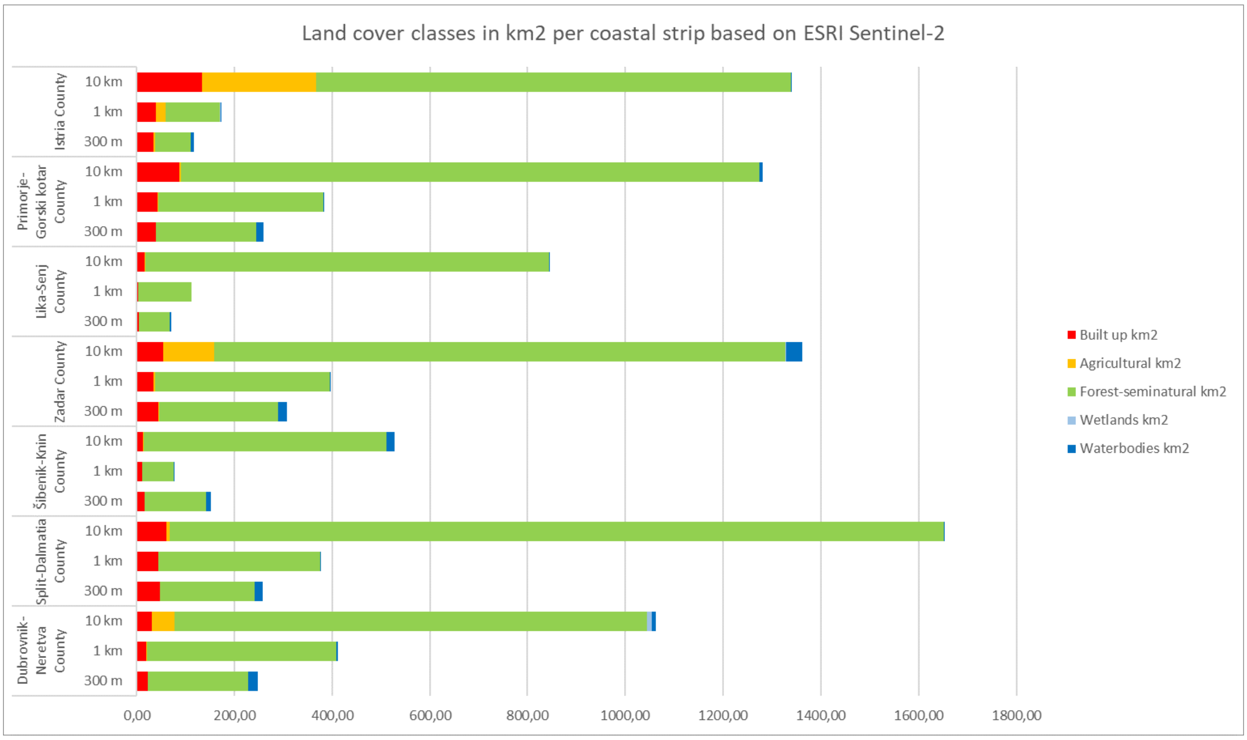
Annex 2 - Areas of land cover classes in km^2 per county based on GLC_FCS30D, 2021

County	Coastal strip	Built-up	Agricultural	Forest and seminatural	Wetlands	Water bodies	Total coastal zone
Dubrovnik-Neretva County	300 m	13,48	24,54	203,68	3,94	1,97	247,60
	1 km	7,43	34,15	367,11	1,07	1,38	411,15
	10 km	11,08	160,34	874,25	11,22	4,64	1061,54
Split-Dalmatia County	300 m	35,32	26,71	190,96	2,95	1,44	257,39
	1 km	27,92	35,10	312,43	0,02	0,05	375,52
	10 km	35,68	250,72	1365,59	0,08	0,22	1652,29
Šibenik-Knin County	300 m	13,69	19,40	116,02	2,44	0,75	152,30
	1 km	7,48	8,60	60,72	0,07	0,73	77,60
	10 km	10,09	55,33	447,86	0,41	14,45	528,14
Zadar County	300 m	29,63	48,45	220,73	6,56	2,64	308,01
	1 km	18,98	60,84	315,44	1,39	0,57	397,22
	10 km	28,70	399,34	900,28	3,52	31,78	1363,62
Lika-Senj County	300 m	7,58	15,66	45,57	1,58	0,93	71,32
	1 km	3,94	26,03	82,03	0,00	0,00	112,01
	10 km	1,73	58,30	784,62	0,00	0,01	844,65
Primorje-Gorski Kotar County	300 m	30,33	43,29	175,32	6,27	4,01	259,21
	1 km	26,56	74,30	281,07	0,03	0,01	381,96
	10 km	48,69	161,92	1064,72	0,32	6,43	1282,07
Istria County	300 m	18,05	24,72	69,31	3,53	1,93	117,54
	1 km	16,91	43,77	110,58	0,43	0,28	171,98
	10 km	39,51	339,82	960,25	0,17	0,00	1339,74



Annex 3 - Areas of land cover classes in km^2 per county based on Esri Sentinel-2 Land Cover, 2021

County	Coastal strip	Built-up	Agricultural	Forest and seminatural	Wetlands	Water bodies	Total coastal zone
Dubrovnik-Neretva County	300 m	23,03	0,52	204,05	0,24	19,83	247,66
	1 km	19,27	1,97	387,58	0,26	2,12	411,20
	10 km	30,55	46,71	967,23	9,32	8,11	1061,92
Split-Dalmatia County	300 m	48,16	0,02	192,87	0,02	16,21	257,27
	1 km	43,99	0,64	330,64	0,00	0,12	375,40
	10 km	61,70	6,70	1582,51	0,01	0,84	1651,75
Šibenik-Knin County	300 m	16,51	0,00	125,78	0,01	9,90	152,19
	1 km	10,63	0,00	65,84	0,08	1,01	77,56
	10 km	13,41	1,34	496,53	0,00	16,51	527,78
Zadar County	300 m	44,81	0,51	244,15	0,25	18,05	307,77
	1 km	33,86	3,46	358,08	0,39	1,11	396,90
	10 km	55,15	103,28	1169,41	1,17	33,58	1362,58
Lika-Senj County	300 m	4,51	0,01	63,54	0,00	3,20	71,27
	1 km	2,56	0,00	109,37	0,00	0,00	111,93
	10 km	15,79	1,61	826,41	0,00	0,00	843,81
Primorje-Gorski Kotar County	300 m	38,94	0,08	205,11	0,05	14,82	259,00
	1 km	43,17	1,12	337,36	0,00	0,08	381,74
	10 km	88,06	2,78	1182,86	0,01	7,58	1281,30
Istria County	300 m	33,80	4,04	72,59	0,17	6,87	117,47
	1 km	39,29	20,78	110,92	0,25	0,67	171,90
	10 km	133,07	234,90	970,56	0,00	0,73	1339,26



Annex 4 – The Shape Index calculations for Škabrnja and Murter-Kornati

	10 m resolution		30 m resolution	
	bordering	within	bordering	within
MURTER-KORNATI	45053	765773	14843	80319
total	810826		95162	
% bordering	5,56%		15,60%	
ŠKABRNJA	2579	224314	856	24642
total	226893		25498	
% bordering	1,14%		3,36%	

Annex 5 – The Shape Index calculations per county and coastal strip with perimeter and area

County	Coastal strip	Perimeter	Area	Shape Index
Zadar County	300 m	2227716,63	308008802,50	0,0008
Primorje-Gorski Kotar County	300 m	1907734,91	259194218,70	0,0009
Split-Dalmatia County	300 m	1846032,16	257382864,70	0,0009
Dubrovnik-Neretva County	300 m	1806956,66	247602594,20	0,0010
Šibenik-Knin County	300 m	1142549,47	152303480,80	0,0015
Istria County	300 m	856553,13	117534126,20	0,0020
Lika-Senj County	300 m	520270,34	71329072,51	0,0033
Dubrovnik-Neretva County	1 km	1263004,46	411152020,40	0,0032
Zadar County	1 km	1209028,69	397222738,50	0,0034
Split-Dalmatia County	1 km	1136131,09	375530654,90	0,0037
Primorje-Gorski Kotar County	1 km	1142218,76	382026567,10	0,0037
Istria County	1 km	528079,90	172006430,70	0,0078
Lika-Senj County	1 km	333883,45	112024362,90	0,0126
Šibenik-Knin County	1 km	237390,03	77610405,88	0,0173
Dubrovnik-Neretva County	10 km	620016,30	1061540372,00	0,0347
Primorje-Gorski Kotar County	10 km	577331,66	1282277153,00	0,0483
Split-Dalmatia County	10 km	631477,25	1652264775,00	0,0521
Zadar County	10 km	561860,14	1363619746,00	0,0543
Istria County	10 km	391920,27	1340067790,00	0,1096
Lika-Senj County	10 km	275978,12	844725945,80	0,1394
Šibenik-Knin County	10 km	196882,81	528136899,50	0,1712

Leona Kovačić

Datum rođenja: 13/11/2000

Državljanstvo: hrvatsko

Spol: Žensko

KONTAKT

Cetinska cesta 2
21310 Omiš, Hrvatska **(Kućna)**

leona.kovaciclk@gmail.com

(+385) 915818792

europass

OBRAZOVANJE I OSPOSOBLJAVANJE

2015 – 2019 Omiš, Hrvatska

Opća gimnazija Srednja škola Jure Kaštelan

2019 – 2022 Split, Hrvatska

Sveučilišni prvostupnik (baccalaureus) inženjer geodezije i geoinformatike Fakultet građevinarstva, arhitekture i geodezije u Splitu

Internetske stranice https://gradst.unist.hr/

2022 – TRENUTAČNO Zagreb, Hrvatska

Magistar inženjer geodezije i geoinfromatike Geodetski fakultet Sveučilišta u Zagrebu

Internetske stranice https://www.geof.unizg.hr/

POČASTI I NAGRADE

2020 Fakultet građevinarstva, arhitekture i geodezije u Splitu

Dekanova nagrada

2021 Fakultet građevinarstva, arhitekture i geodezije u Splitu

Dekanova nagrada

2022 Fakultet građevinarstva, arhitekture i geodezije u Splitu

Dekanova nagrada

2023 Geodetski fakultet Sveučilišta u Zagrebu

Nagrada fakulteta

KONFERENCIJE I SEMINARI

2020 Opatija

13. simpozij ovlaštenih inženjera geodezije

Leona Kovačić, Marino Kovačić, Frane Gilić, Martina Baučić, Danijela Jurić Kaćunić: Korištenje aerofotogrametrije u analizi stijenskih odrona grada Omiša

2023 Poreč

16. simpozij ovlaštenih inženjera geodezije

Martina Baučić, Frane Gilić, Leona Kovačić: E-CITIJENS – Sustav za upravljanje hitnim situacijama zasnovan na građanskom novinarstvu

RADNO ISKUSTVO

03/2024 – TRENUTAČNO Split, Hrvatska

Geodetski tehničar Geodetski zavod d.d. Split

obavljanje geodetskih poslova u uredu
terenska mjerenja
rad sa sudskim vještakom

2021 – 2022 Split, Hrvatska

Demonstrator Fakultet građevinarstva, arhitekture i geodezije u Splitu

kolegiji Matematička analiza, Vektorska analiza i Terenska mjerenja

JEZIČNE VJEŠTINE

MATERINSKI JEZIK/JEZICI: hrvatski

Drugi jezici:

engleski

Slušanje C1
Govorna produkcija C1
Čitanje C1
Govorna interakcija C1
Pisanje C1

DIGITALNE VJEŠTINE

Rad u GIS softverima i obrada prostornih podataka (ArcMap, QGIS, Agisoft Metashape) | Dobro poznavanje AutoCAD | Aktivno i svakodnevno korištenje SDGE-a, Zis-a i OSS-a | Rad u programima korištenim u polju fotogrametrije - Pix4D | Komunikacijski programi (Skype Zoom TeamViewer) | Osnovno poznavanje SketchUp-a | Odlično poznavanje MS Office