JRC TECHNICAL REPORT

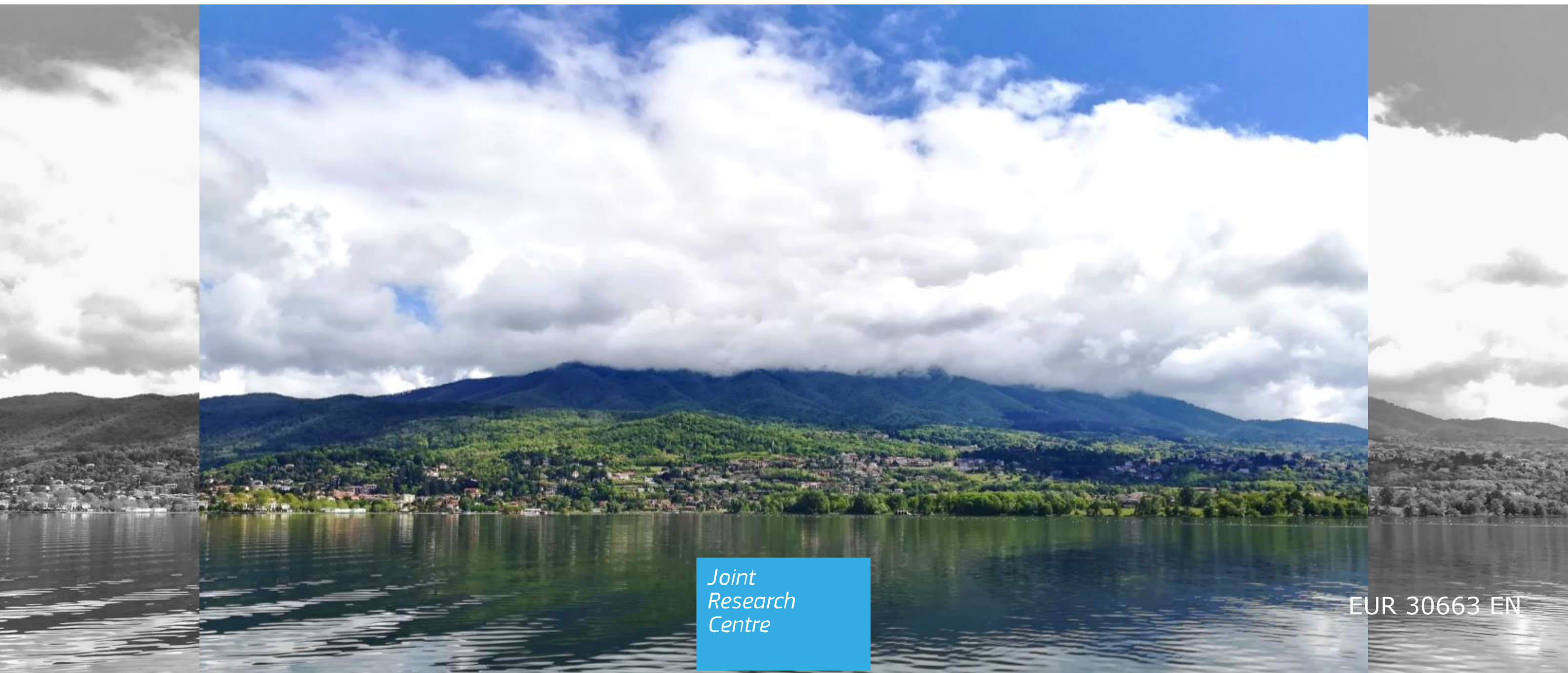
The LUISA Base Map 2018

A geospatial data fusion approach to increase the detail of European land use/land cover data

Cristian Pigaiani

Filipe Batista e Silva

2021



EUR 30663 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Carlo Lavalle
Address: European Commission Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
Email: carlo.lavalle@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC124621

EUR 30663 EN

PDF ISBN 978-92-76-34207-6 ISSN 1831-9424 doi:10.2760/503006

Luxembourg: Publications Office of the European Union, 2021

How to cite this report: Pigaiani, C and Batista e Silva, F, *The LUISA Base Map 2018 - A geospatial data fusion approach to increase the detail of European land use/land cover data*, EUR 30663 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-34207-6, doi:10.2760/503006, JRC124621.

Contents

Abstract

The LUISA Base Map 2018 is a land use/land cover map of Europe compatible with the CORINE Land Cover nomenclature, but offering a considerable higher spatial and thematic detail. It was produced by employing an automated, reproducible and structured geographical data fusion approach, integrating land use data from diverse, trusted, off-the-shelf geospatial data sources. It is a key input to the European Commission Joint Research Centre LUISA territorial model. Because it is publicly available, it can be used in many other applications requiring fine spatial and/or thematic detail of land use/land cover consistently for Europe. This technical report describes the background and specifications of the LUISA Base Map 2018, documents the materials and methods employed in its production, and discussed its main strengths and limitations.

Acknowledgements

The authors of this report would like to thank: the GISCO team at Eurostat for the management and release of a license of the TomTom Multinet used in this project; the GHSL team from JRC.E.1, and in particular Christina Corban, for granting us early access to the GHSL built-up layer; and Eva Ivits, from the European Environment Agency, for the healthy discussions and collaboration concerning the mapping and applications of European land use and land cover data.

Executive summary

Detailed and up-to-date information on the spatial distribution of land uses is essential for monitoring and managing the conditions of the physical and human environments, to support urban and regional planning, for disaster risk assessment and many other applications. Since the 1990's, the main authoritative source of land use and land cover data for Europe has been the CORINE Land Cover (CLC) map, counting now five editions (1990, 2000, 2006, 2012, 2018), and is currently managed by Copernicus, the European Union's Earth Observation programme. However, CLC has a rather limited spatial resolution, making it unsuitable for many applications. In recent years, the Copernicus Land Monitoring Service has invested in new and more detailed land use and land cover products focusing specifically on urban areas (e.g. Urban Atlas), or layers focusing specific land cover categories such as water, or forested areas (i.e. High Resolution Layers). At the same time, new private and crowd-sourced databases have emerged, often containing land use-relevant information.

With the objective to fill the existing gap, the European Commission Joint Research Centre devised (in the early 2010's) an approach to integrate land-use relevant information from the increasingly available sources of geospatial data into a single, consistent and more detailed land use and land cover map for Europe, yet fully compatible with the CLC nomenclature. This gave rise to the 'LUISA Base Maps', used as starting point for the land use change simulations by JRC's LUISA territorial modelling platform, among other applications. In practice, the LUISA Base Map corresponds to a modified and enhanced version of the original CLC map, delivering higher spatial and thematic detail. The approaches developed to construct earlier editions of the LUISA Base Map have been described in two peer-reviewed publications, gathering altogether nearly 100 citations, according to Google Scholar at the time of writing.

Currently, the approach consists of a series of structured, automated and reproducible geospatial data fusion processes whereby highly detailed land use information from trusted, off-the-shelf datasets are integrated in a sequential manner, having the CORINE Land Cover as starting point. The main sources used to generate the 2018 edition of the LUISA Base Map include the CORINE Land Cover Change Maps, the Copernicus High Resolution Layers (forest, water, wetlands, and imperviousness layers), the Copernicus Urban Atlas and Coastal Zones land use maps, the Global Human Settlement Layer from the Joint Research Centre, as well as the TomTom Multinet and OpenStreeMap. The use of European-wide Earth Observation data as input and a uniform and automated methodology yields high comparability of the result across countries. Moreover, the 2012 version of the LUISA base map was reproduced using reference data from 2012, but taking advantage of the improved production methods herein described. The final products are publicly available in two resolutions: 50 metres (native resolution) and 100 metres (generated using a custom resampling method from the original 50 metres version).

While the LUISA Base Maps do not have a single, direct EU policy dimension per se, they provide information to support a broad set of policy domains and can be used as element in other analytical and modelling tools that address policy questions more directly. In effect, the LUISA Base Maps contribute to the knowledge base of the European Commission's Knowledge Centre for Territorial Policies, together with other key reference data sets such as the JRC-GEOSTAT Population Grid 2018 and the Annual Regional Database of the European Commission (ARDECO).

1 Introduction

Detailed and up-to-date information on the spatial distribution of land use and land cover is essential for monitoring the conditions of the physical and human environment and their interactions, for natural resource management, to support urban and regional planning, for disaster risk assessment among other applications.

The CORINE Land Cover (CLC)[1] is a pan-European digital map of land use and land cover (LULC), currently covering 39 European countries. The first version was produced for the year 1990 and it has been updated regularly since then, with subsequent editions for 2000, 2006, 2012 and 2018. The CLC is meant for temporal and cross-country comparisons, thanks to its consistent specifications and harmonized methodology. It includes 44 mutually exclusive LULC categories, organized in a hierarchical structure. For its European-wide coverage, temporal frequency and overall consistency, CLC has become a well-known and widely used source of LULC data for numerous studies and assessments concerning land use changes, landscape, ecosystem services, natural resources and hazards (Feranec et al., 2016).

However, CLC has its own drawbacks, of which one the most notable is the coarse spatial resolution of 25 hectares of Minimum Mapping Unit (MMU). The high MMU does not impact all LULC classes equally. LULC categories which occur in smaller and more dispersed patches of land tend to be more affected by the spatial generalization rules, and end up unmapped and integrated in the surrounding dominant LULC patches, leading to potentially strong land accounting biases. For instance, arable land may be overestimated generally at the expense of urban fabric areas. The limited spatial resolution of CLC prevents its use in applications that require higher spatial detail (Batista e Silva, Lavalle, & Koomen, 2013).

To address this and other limitations of the CLC, the European Commission Joint Research Centre (JRC) has proposed and developed over the past years a data fusion approach to increase the detail of CLC. The approach is based on the enhancement of CLC by means of a structured integration of more detailed LULC information available from multiple off-the-shelf, trusted geospatial data sources. This approach has been already applied to enhance the CLC versions 2006 (Batista e Silva, Lavalle, & Koomen, 2013) and 2012 (Rosina et al., 2020). The resulting maps have been used for two main purposes: a) as base LULC map for the LUISA model simulations (Batista e Silva, Lavalle, Baranzelli, et al., 2013; Jacobs-Crisioni et al., 2017; Lavalle et al., 2016), and, b) as input to the production of detailed population distribution maps (Batista e Silva et al., 2020; Batista e Silva, Gallego, et al., 2013).

This report documents the enhancement of the CLC 2018, resulting in the herein called "LUISA Base map 2018". The production of the LUISA Base Map 2018 was motivated by the release of the CLC 2018, in mid-2019, and to ensure that the LUISA model and other assessments carried at the JRC remained up to date.

As documented in Table 1, the now 3 versions of the LUISA Base Maps represent also 3 stages of evolution of this line of products. The version 2006 was essentially a first, successful proof-of-concept, focused on the improvement of the spatial detail. In the version 2012, we introduced novel improvements to the methodology, to obtain additional thematic detail within the artificial surfaces. The version 2018 integrates all previous improvements, and focuses on streamlining the production to allow higher automation and reproducibility. This was achieved by programming the entire workflow, using standard and open source libraries. During the development, the methodology was further fine-tuned for increased internal consistency of the product specifications. Besides, the native working cell size was reduced from 100 m to 50 m, for increased overall precision, while maintaining the MMU at 1 hectare for artificial surfaces and 5 hectares for all remaining non-artificial surfaces as in the previous versions.

[1] https://land.copernicus.eu/pan-european/corine-land-cover

The objective of this technical report is to provide a detailed account of the materials and methods employed in the production of the LUISA Base Map 2018. In chapter 2 we will cover the product specifications, and describe the input data and each step of the workflow. In chapter 3 we make considerations concerning the data quality and future work.

Table 1. The three versions of the LUISA Base Map and their development stages.

	LUISA Base Map		
	Version 2006	**Version 2012**	**Version 2018**
Development stage	Proof-of-concept	Improvement	Consolidation and streamlining
Underlying CLC version	2006	2012	2018
Production period	2010-2012	2016-2017	2020
Main differences *vis-à-vis* CLC	- Smaller MMU; - Improved classification of urban fabric classes.	- Same as LUISA Base Map 2006; - Higher thematic detail for artificial surfaces.	- Same as LUISA Base Map 2012; - Higher internal consistency; - Increased cell size; - Streamlined production.
Pixel size	100 m	100 m	50 m and 100 m
Reference	Batista e Silva et al. (2013)	Rosina et al. (2020)	Pigaiani and Batista e Silva (2021)
Data access	Upon request	Public*†	Public‡
Main usages	LUISA model, population downscaling		

* URL (original version 2012): https://figshare.com/articles/dataset/Increasing_the_detail_of_land_use_cover_data_for_Europe_by_combining_heterogeneous_data_sets/6210392

† The version 2012 of the LUISA Base Map has been reproduced in 2021 using reference data from 2012, but taking advantage of the improved production methods described in this report. URL: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUISA/EUROPE/Basemaps/

‡ URL: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUISA/EUROPE/Basemaps/2018/VER2021-03-24/

2 Data and methods

2.1 Overview and product specifications

The aim of the work is to create a harmonised and ready-to-use, enhanced version of CORINE Land Cover map, herein referred as LUISA Base Map, for the whole of Europe, taking advantage of the increasing wealth of land use-relevant data available from multiple emerging geospatial data sources.

The workflow comprises 9 main steps that integrate suitable data from different sources and formats, including categorical raster data, interval raster data and vector polygon data, and following a set of established rules in order to produce more detailed cartographic representation of land use and land cover (Figure 1). In practice, the CLC gets further modified at each step sequentially, resulting in a series of intermediate and partially 'refined' land use maps (herein after referred generically as 'CLCr').

The CLC LULC nomenclature is structured in 3 nested hierarchical levels, with increasing thematic detail, and each class is mutually exclusive. For example:

- 1XX – Artificial surfaces
  - 11X – Urban fabric
    - 111 – Continuous urban fabric
    - 112 – Discontinuous urban fabric
- (…)
- 3XX – Forest and semi-natural areas
  - 31X – Forest
    - 311 – Broad-leaved forest
    - 312 – Coniferous forest
    - 313 – Mixed forest
- (…)

All steps involved the creation of the semantic correspondence between the information included in the input datasets and the CLC nomenclature, ensuring a compatibility between the final LUISA Base map LULC nomenclature and that of CLC down to its third hierarchical level. There are two important differences, however. In the LUISA Base Map, the Artificial surfaces group (1XX) is broken down into a forth level (1XXX), while the thematic detail of the wetlands group (4XX) is not preserved due to the integration of wetland data that cannot be easily sub-classified. The full nomenclature of the CLC Base Map can be consulted in Annex 1.

After the entire data fusion process is completed, the LUISA Base Map provides a higher thematic detail (17 artificial land use classes compared to 11 in the original CLC) and a higher spatial detail (MMU of 1 ha for artificial surfaces and 5 ha for non-artificial surfaces compared to 25 ha in the original CLC).

It is important to point out that, in order to limit the loss of information, the MMU specifications were applied at the level of each input dataset before its actual integration in the map: Land use and land cover patches, once extracted from input data sources, were filtered in accordance with the MMU rules and only then merged with the target map.

However, each step that adds new land use patches to the prior intermediate land use map creates as well additional fragmentation of the previously depicted land use patches, therefore potentially decreasing the MMU below the desired specifications. To address the over fragmentation of the final result, the last step of the workflow applies a custom majority filter to smooth land use patches and achieve a closer to expected MMU.

Several open source software technologies were used for the automation of the whole process. Python was the programming language chosen to develop the project while GDAL and OGR libraries were used for manipulating geospatial raster and vector data. NumPy and SciPy libraries provided an assortment of routines for fast operations on large arrays and POSTGIS was used as main geoprocessing engine for vector datasets that couldn't be easily manipulated otherwise. The modularity and automation of the process allows for reproducibility, easier maintenance, revision or update of the LUISA Base Map.

Table 2 lists the main product specifications.

Table 2. Main specifications of the LUISA Base Map 2018.

Product features	Digital thematic map
Geographic projection / Reference system	ETRS 1989 Lambert Azimuthal Equal-Area projection
Reference year	2018
Pixel size	50 m (native version), 100m (resampled version)
Spatial resolution	MMU: 1 ha for artificial surfaces and 5 ha for non-artificial surfaces
Number of thematic classes	46 (17 artificial surfaces classes and 29 non artificial surfaces classes)
Coverage	Pan European (39 countries): Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom.
Delivery format	Raster (tiff file)

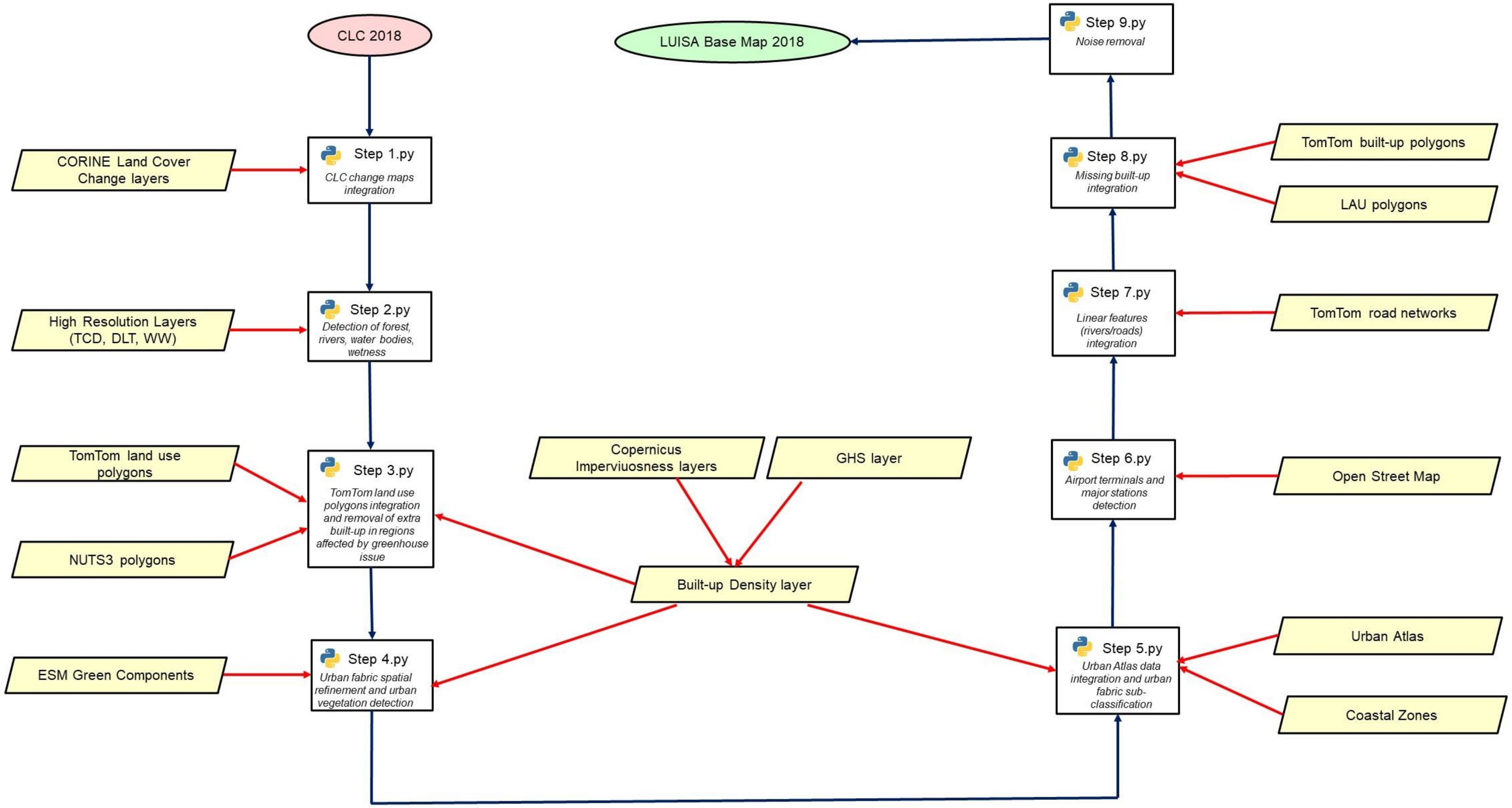


Figure 1. Simplified data processing flowchart of the production of the LUISA Base Map 2018.

2.2 Input data

CORINE Land Cover Change (CLCC) layer

The CLC change map or layer[2] is an important complementary product made publicly available along with every release of the core, 'status' CLC layer. While the 'status' layer represents the situation of land use and land cover for the reference date of the release, the 'change' layer records the land use and land cover changes occurred between the reference dates of the most recent release and the previous releases. The layer is derived from satellite imagery by the direct mapping of changes based on image-to-image comparison and is more detailed (5 ha) than CLC status layer (25 ha). Only patches from the 2012-2018 change layer were included in the LUISA Base map[3].

High Resolution Layers (HRLs)

Copernicus is the European system for monitoring the Earth and is coordinated and managed by the European Commission. As part of Copernicus framework, *High resolution layers*[4] (HRLs) provide information on specific land cover characteristic and are produced from satellite imagery through a combination of automatic processing and interactive rule based classification.

Compared to the previous versions, for the 2018 reference year the HRL products have increased in resolution to 10 meters using as the main source the Sentinel satellites. Furthermore, the HRL 2018 Imperviousness contains two sub-products: the *Imperviousness Density* layer (IMD) that provides the percentage of sealed surface and the new *Impervious Built-Up* binary layer (IBU) representing built-up areas.

The Global Human Settlement Layer (GHSL)

The *Global Human Settlement Layer*[5] (GHSL) produces global spatial information about the human presence on the planet and is produced by the Joint Research Centre (JRC) with the support of DG for Regional and Urban Policy (DG REGIO). It uses heterogeneous data including satellite imagery, census data and volunteered geographic information and is available at 10 m spatial resolution. Along with the Copernicus Imperviousness Layers the GHSL was processed to produce the LUISA intermediate *Built-Up Density layer* at 50 m spatial resolution.

TomTom Multinet

TomTom Multinet[6] is a proprietary vector database that comprises road network datasets, administrative and built-up areas, land use land cover polygons and point of interests (POIs). It has a thematic detail higher or equivalent to the CLC and for this reason a fair correspondence between the two schemas could be established. In terms of spatial detail, TomTom land use polygon includes many patches smaller than 25 Ha. For the 2018 release,

---

[2] http://land.copernicus.eu/pan-european/corine-land-cover/lcc-2012-2018

[3] The inclusion of earlier change layers (e.g. 2006-2012 and 2000-2006) would lead to the integration of outdated information, such as land use polygons which do not represent the reality on the ground in 2018.

[4] http://land.copernicus.eu/pan-european/high-resolution-layers

[5] http://ghsl.jrc.ec.europa.eu

[6] http://developer.tomtom.com/products/maps-api

in particular, additional material was processed which resulted in increased coverage throughout Europe.

Urban Atlas (UA)

The European Urban Atlas is part of the Copernicus Land Monitoring Services suite of cartographic products. The full dataset of *Urban Atlas 2018*[7] covers 788 FUAs including the whole of the European Union countries, EFTA countries, Wester Balkans, Turkey and UK. The nomenclature is the same as for the 2012 version and remains overall compatible with the CORINE Land Cover nomenclature. As for the spatial resolution, it includes 17 artificial surfaces classes with MMU of 0.25 ha and 10 rural classes (agriculture, forest, wetlands and water) with MMU of 1 ha. UA is derived mainly from using Earth Observation (EO) data backed by other reference data, such as Commercial Off-The-Shelf (COTS) or *Open Street Map* (OSM) navigation data and topographic maps.

Coastal Zones (CZ)

As Urban Atlas, the *Coastal Zones 2018*[8] product is part the Copernicus Land Monitoring Services and provides detailed LULC information extracted from Very High Resolution (VHR) satellite data and other available data. The set of maps cover all European coastal territory within a 10 km inland buffer zone. Its nomenclature is based on the Mapping and Assessment of Ecosystems and their Services (MAES) framework. Notwithstanding, it offers a high degree of compatibility with other LULC European products such as CLC and Urban Atlas, despite some adaptations to reflect better the peculiar features of coastal zones. It includes 71 distinct thematic classes of which 20 are artificial surfaces and 51 are rural and natural areas with a MMU of 0.5ha and a Minimum Mapping Width of 10 m.

Open Street Map (OSM)

Open Street Map[9] is a free, crowd-sourced, and world-wide renowned geographic database focused on the mapping of transport networks, but also administrative boundaries, biophysical features such as LULC, buildings, points of interest, among other features. It is built from voluntary contributors who digitize information from satellite imagery and local expert knowledge. Other data is gathered from out-of-copyrights maps, public domain databases or donations of proprietary databases. From OSM, we were mostly interested in specific transport features, such as train stations and airport terminals to help refine the thematic detail of CLC's transportation land use classes.

European Settlement Layer (ESM) – Green component

The European Settlement Map is a JRC mapping product focused on the detailed mapping of built-up areas. The ESM contains a 'green component' which we used as input to the LUISA Base Map 2018. In the ESM, the vegetation is firstly detected using a pseudo NDVI index and is then differentiated in two green layers depending on their location: vegetation in urban areas (BU Area-Green NDVIx) and vegetation outside of urban areas (NBU Area – Green NDVIx). The result is further enhanced through the integration of other datasets (buildings from OSM, streets, railways and inland water from TomTom, Street Area and Urban Green from Urban Atlas). The main advantages of the green component of the ESM is the European coverage and the very high spatial resolution (2.5m).

---

[7] http://land.copernicus.eu/local/urban-atlas/urban-atlas-2018

[8] http://land.copernicus.eu/local/coastal-zones/coastal-zones-2018

[9] http://www.openstreetmap.org

The most recent and stable version available of *ESM Green Components* at the time of LUISA Base Map production was ESM 2012 – Release 2017.

Table 3. Data sources used in the production of the LUISA Base Map 2018.

Dataset	Source	Reference year / version	Affected LULC Classes
CORINE Land Cover (CLC) and CLC Change (CLCC) maps	Copernicus	CLC: 2018; CLCC: 2012-2018 / V2018_20	All LULC classes
High Resolution Layers: Tree Cover Density, Dominant Leaf Type, Water & Wetness	Copernicus	2018	Forests, wetlands, waters
Coastal Zones	Copernicus	2018	Artificial surfaces classes
Copernicus Imperviousness Layers	Copernicus	2018	Urban fabric, sport and leisure facilities
Global Human Settlement Layer	JRC	2018 (version SN CNN)	Urban fabric, sport and leisure facilities
ESM Green Components	JRC	2012 Release 2017	Urban fabric, urban vegetation
Urban Atlas	Copernicus	2018	All LULC classes
TomTom land use	TomTom Multinet	Version 2018	Industrial/commercial/ transport units, green urban
TomTom built up			Urban fabric
TomTom network			Road and rail network
Open Street Map polygons	OSM Foundation	2020	Train stations, airport terminals

2.3 Detailed workflow

Step 1 – Integration of CLC change maps

The first step of the map refinement involved the integration of *CLC change maps*, affecting all land use and land cover classes. Each feature in the CLC change map corresponds to a polygon with the information on the LULC conversion between two reported years, i.e. the LULC class in the previous reported year (source LULC) and LULC class in the current reported year (destination). The polygons included in the 2012-2018 change vector layer were initially converted to raster format using the "centroid method", keeping the information of the destination (i.e. 2018) LULC class, and then overlaid on the original CLC 2018 map.

After this process, the spatial detail of the refined map increased from 25 ha (MMU of CLC map) to 5 ha (MMU of CLC change layers) in the affected areas.

Step 2 - Integration of forest, water and wetlands information from Copernicus High Resolution Layers

In this step, the *Copernicus High Resolution Layers* (HRLs) were combined with the refined map. At this stage, three specific layers were used and processed, namely: HRL *Tree Cover Density,* HRL *Dominant Leaf Type* and HRL *Water & Wetness Layer*.

HRL *Tree Cover Density* shows the level of tree cover density in a range from 0% to 100% at 10 m spatial resolution. The layer was firstly resampled to 50 m cell size using the "average" resampling method and then a 50% threshold was applied to obtain a boolean dataset of cells where the forest class was dominant.

HRL *Dominant Leaf Type* provides land cover classification of forested areas at 10 m spatial resolution using 3 thematic classes with the following categorical values: 0 for all non-tree covered areas, 1 for broadleaved forest and 2 for coniferous forest. After the resampling to 50 m spatial resolution by considering the most frequent leaf type occurring in the target cell, it was combined with the boolean dataset previously created to detect the areas of *broadleaved forest* (3110) and *coniferous forest* (3120).

A third class, *Mixed forest*, was distinguished through a focal analysis, by applying a 3 X 3 moving window of 50 m cells to the *Dominant Leaf Type* layer with values 1 and 2 for broadleaved and coniferous forest, respectively. The value of each cell in the output was calculated based on the average of the *Dominant Leaf Type* layer cells within the specified window. Resulting cells with values ranging from 1.3 to 1.7 were deemed to represent areas with significant presence of both forest types, and were thus classified as *Mixed forest* (3130).

HRL *Water & Wetness* is a thematic product showing the occurrence of water and wet surfaces at 10 m resolution. The location of water bodies and wetland is also present in the *Urban Atlas*. A *Water Density layer* (WD) and a *Wetness Density layer* (WETD) at 50 m spatial resolution were derived by combining *HRL Water & Wetness* and *Urban Atlas.* Cells where wetness density was dominant (WETD > 50%) were classified as *Wetlands* (4000). Because these layers do not contain the same thematic breakdown as in CLC, the description of wetlands in the resulting, refined map remained at the level 1 of the CLC nomenclature. In this case, the improvement of the spatial detail was done at the expense of the thematic detail. Attempts to distinguish inland (41X) from coastal (42X) wetlands would inevitably introduce buffer artefacts in the final map and were therefore discarded.

Generic water bodies were extracted from *HRL Water Density* (WD > 50%) and differentiated as *Water courses* (5110) and *Water bodies* (5120) by applying a sequence of morphological algorithms. However, the shape-based distinction between river and lakes could be less effective in complex aquatic systems leading to occasional inconsistencies in the classification.

For all the above considered LULC classes (forest types, wetland and water areas), only the detected patches of at least 20 contiguous cells were integrated in the target map, as per the set MMU specifications (5 ha for non-artificial classes).

Finally, to preserve the CLC thematic detail, *Green Urban Areas* (141), *Fruit Trees* (222), *Olive Groves* (223) and *Agro Foresty Areas* (244) weren't allowed to be updated by forest areas extracted from HRLs.

Step 3 – Integration of TomTom land use

The results of the previous step were upgraded by adding information from the *TomTom Multinet land use* dataset. This is a polygon layer featuring industrial, commercial, transport units and green urban areas.

A look-up table was constructed to establish the relationship between TomTom and CLC nomenclatures (see Annex 2).

Given the presence of false positives especially among TomTom land use polygons classified as *industrial or commercial units* (121), a specific procedure was developed to identify and exclude them. We used a zonal statistics function to measure the amount of built-up in each industrial or commercial units polygons from TomTom, and excluded those whose built-up share was under 50% of the polygon area.

The selected polygons were finally rasterized and their LULC class associations were transferred to CLCr map as follows:

1. Boolean rasterization of vector polygons at 10 m spatial resolution using the 'all touched' method;
2. Aggregation to 50 m resolution, with the count of 10 m cell within each 50 m cell;
3. Selection of 50 m cells with a count of 13 or more (corresponding to more than 50% of the 50 m pixel covered by the original polygon).

Generation of the Built-up Density Layer (BD)

The *Built-up Density layer* (BD) indicates the share of built-up area per pixel in a range from 0-100%. It is a key input data for several tasks of the map refinement (e.g., classification of urban fabric, detection of urban vegetation, refinement of urban fabric shapes). The following procedure was carried out to obtain the raster layer at 50 m spatial resolution.

First, the *HRL Imperviousness Density* was combined with the *HRL Impervious Built-up* to produce a layer at 10 m spatial resolution providing the percentage of sealed area for built-up areas only[10].

The output of the previous operation was then modified to improve the built-up detection in the following way: to limit the overestimation of urban fabric, built-up areas overlapping CZ cells of class *Greenhouses* (212) were filtered out. Furthermore, for regions known to have a high concentration of greenhouses (e.g., Andalusia, Liguria) and not entirely covered by CZ dataset, built-up was removed also in correspondence of CLC cells of *Non-irrigated arable land* (211), *Permanently irrigated land* (212), *Fruit trees and berry plantations* (222). Besides, the *GHS 2018 SN CNN* layer was used as a complementary source to integrate the sparse and isolated built-up in rural areas with low detection by the 2018 Copernicus Imperviousness layers. GHS layer, before being exploited, was pre-processed by filtering out the extra built-up along roads through the use of TomTom network dataset.

The final result was resampled to 50 m spatial resolution by applying the average resampling method.

Step 4 – Urban fabric spatial refinement

In CLC, while the overall accounting of urban fabric is underestimated, the depicted urban fabric polygons themselves tend to be exaggerated due to generalization rules. To reduce the impact of this issue on the refined map, the ESM Green layer was employed to refine the shape of urban fabric patches. Moreover, a new class named *Urban vegetation* (1130) was introduced, comprising green areas detected within urban fabric patches in CLC map.

Among the ESM 2012 – release 2017 products, two layers providing green density with values ranging from 0% to 100% are available at 10 m spatial resolution: *Built-up Area Green NDVIx* and *No Built-up Area Green NDVIx*. These were first merged to obtain a single

[10] Built-up areas are a subset of all sealed or impervious surfaces. Sealed surfaces include, for instance, roads and other physical structures which are not buildings.

Green Density layer which was later resampled to 50 m spatial resolution by the average resampling method.

Then, cells of classes 111 or 112 with high share of green areas (> 50%) and scoring 0% as per the Built-up Density layer were detected and subsequently classified according to the following criteria:

- As class 1130 – *Urban vegetation* if the 1-pixel vicinity of green patches are surrounded by 80% of any of these classes altogether: 111, 112, 121, 122, 141, 511, 512, 513;
- As the nearest non-artificial class elsewhere.

Conversely, for cells of class 111 or 112 located at the edges of the respective patches were converted to the nearest non-artificial class if the built-up coverage was 0%, even if the density of green was <50%. In this case, the urban fabric cells are always replaced by the nearest non-artificial LULC classes in the CLCr.

It is worth mentioning that UA data, due to its fine grained representation of urban land use land cover, superseded in many cases "generic urban vegetation" areas with the more detailed class *Green Urban Areas* (see step 5). For this reason in the final product the presence of generic urban vegetation is more significant in small cities, villages and settlements not covered by UA data.

Step 5 – Integration of Urban Atlas and Coastal Zones data and urban fabric classification

In step 5, UA and CZ datasets were integrated in the refined CLC map and subsequently a custom algorithm was applied to differentiate CLC urban fabric classes into 4 distinct classes, according to building density provided by *Built-Up Density Layer*.

The legend of UA 2018 is analogous to that of CLC but more focused on urban areas with a higher number of artificial LULC classes. On the contrary, the thematic detail for non-artificial classes is reduced to the second or first level of CLC. The conversion of UA pixels into CLC equivalent pixels was done through the use of conversion rules defined in a decision matrix. The table in Annex 3 shows the matrix of decision rules where the rows represents the CLC map and the columns the UA map with its respective original nomenclature. While for some UA classes there was a direct semantic correspondence with CLC (*Industrial, commercial, public, military and private units* 1210, *Port Areas* 1230, *Airports* 1240, *Construction sites* 1330, *Green urban areas* 1410, *Sport and leisure facilities* 1420*, Pastures* 2300*, Orchards* 2500), for other classes the more detailed geometry was integrated and the missing thematic detail was derived through a proximity analysis (*Arable land* 2100, *Permanent crops* 2200, *Complex and mixed cultivation patterns* 2400, *Forests* 3100, *Herbaceous vegetation associations* 2100, *Open spaces with little or no vegetation* 3300). At this stage, the UA classes *Wetland* 4000 and *Water bodies* 5000 were already treated along with HRL Water & Wetness and incorporated in the refined map (see step 2). Lastly, UA railway roads classes were aggregated at third level of CLC and patches of forest class were reclassified as *Urban Vegetation* (1130) when located within built-up areas for consistency with step 4.

As for the *Coastal Zones* dataset, only artificial surface classes were added to the map and classified by establishing a direct correspondence between the CZ and CLC nomenclatures (see Annex 4). Conversely, polygons of non-artificial surface classes, frequently characterized by a larger areal extent (often spanning beyond the coastal zona buffer areas), were not integrated in order to avoid unpleasant boarder artefacts between the areas covered and uncovered by the CZ map.

Importantly, cells of urban fabric classes from both UA and CZ datasets were not directly overlapped to the map. Regardless of their sub-classification, they were assigned an arbitrary very low value of urban fabric density (<10%) and added to the *Built-up Density* layer (BD) only in those cells with null detection by both 2018 Copernicus Imperviousness layers and GHS layer.

After increasing the detection and spatial detail of urban fabric thanks to input from UA and CZ, the next task relates to the improvement of the thematic detail of urban fabric. Actually in CLC map urban fabric is represented by only two classes, *Continuous urban fabric* (111) and *Discontinuous urban fabric* (121), which are not sufficient to account for variety of built-up densities from sparse settlements to high density urbanized areas. To address this issue, four classes were systematically derived depending on the values provided by *Built-up Density* layer (BD):

- 1111 – *High density urban fabric*: (BD > 50% AND AREA >= 5 ha);
- 1121 – *Medium density urban fabric*: (30% < BD <= 50% AND AREA >= 5 ha;
- 1122 – *Low density urban fabric*: (10% < BD <= 30% and AREA >= 5 ha);
- 1123 – *Isolated or very low density urban fabric*: (BD <= 10%).

To impose MMU criteria, patches smaller than 20 pixels of 50 m (5 ha) were sequentially filtered: Class 1111 was formed by contiguous groups of minimum 20 high density pixels while smaller patches were downgraded to medium density. Class 1121 included contiguous groups of minimum 20 medium density pixels while smaller groups were converted to low density. Finally, contiguous groups of minimum 20 low density pixels were classified as 1122 class while the remainder areas formed class 1123. Once detected, patches of the four urban fabric classes were further homogenized and smoothed through the application of a custom majority filter.

Finally, class *Sport And leisure facilities* (142) was decomposed into *Sport and leisure green* (1421) and *Sport and leisure built-up* (1422) in accordance to the following conditions:

- 1421 – *Sport and leisure green*: (CLCr_cell = 142 AND BD = 0%);
- 1422 – *Sport and leisure built-up*: (CLCr_cell = 142 AND BD > 0%).

Step 6 – Integration of transport features from OpenStreetMap

Two additional classes *Airport Terminals* (1242) and *Major stations* (1222) were derived by incorporating polygon data from OpenStreetMap.

First, polygons referring to terminal building (OSM tag aeroway=terminal) and train stations (OSM tag building=station) were selected. Next, the polygons were burned into the raster band using the same methodology developed for the rasterization of TomTom land use polygons (see step 3). Finally, in order to limit the incorporation of any false positive in the OSM database, the derived cells representing airport terminals and major stations were only allowed to replace pixels nested in the existing transportation LULC classes in CLCr:

- 1242 – *Airport terminals*: (CLCr_cell = 124 and OSM_cell = airport terminal);
- 1241 – *Airport areas*: (CLCr_cell = 124 and OSM_cell = no airport terminal);
- 1222 – *Major stations*: (CLCr_cell = 122 and OSM_cell = major station);
- 1221 – *Road/rail networks and associated land*: (CLCr_cell = 122 and OSM_cell = no major station).

Step 7 – Linear features integration

The map was then improved by adding road networks and ensuring connectivity of linear features in view of their distinct function and importance in structuring and fragmenting the landscape.

In accordance with the classification provided by TomTom Functional Road Class (FRC), roads classified as Motorways (FRC=0), Major Roads of High Importance (FRC=1) and Other Major Roads (FRC=2) were selected from the database and removed of tunnel segments using the available attributes. The intersected areas representing roads and railways within tunnels were excluded since they did not establish a recognisable landscape feature above ground. Next, a variable buffer zone depending on the FRC (5 m for FRC=0 and 3 m for FRC=1 or FRC=2) was created for the remaining linear features and subsequently rasterized at 10 m spatial resolution through the "all touched" method. The resulting layer was resampled to 50 m spatial resolution by calculating the percentage of roads coverage per target pixel to produce a *Road Density layer* (RD) with values ranging from 0%-100%. The refined map was then updated with class *Road/rail networks and associated land* (1121) according to these rules:

- 1221 – *Road/rail networks and associated land*:

  ((CLCr_cell != 11XX, 121) AND (RD > 30%)) OR

  ((CLCr_cell = 11XX, 121) AND (RD > 30%) AND (RD > BD))

At this stage, rivers were already added to the map after the integration of UA data along with *HRL Water and Wetness* (step 2) but due to the MMU rules (5 ha for non-artificial surfaces) the connectivity of these linear features was not sufficient. To detect and integrate the missing linear elements, *Water Density Layer* (WD) produced in step 2 was used as shown below:

- 5110 – *Water courses*:

  ((CLCr_cell != 11XX) AND (WD > 10%)) OR

  ((CLCr_cell = 11XX) AND (WD > 10%) AND (WD > BD))

Therefore, by adding cells in which roads/rivers account for as little as 30%/10% of the cell (compared to more restrictive thresholds of within-pixel cover set for integrating other datasets as TomTom land use polygon layer and OSM) it was possible to force the spatial contiguity of linear features. However, transport networks or water courses are only allowed to replace urban fabric classes (11XX) if their pixel density exceeds that of built-up.

Step 8 – TomTom built-up integration

The aim of this step was to detect and minimize omission of urban fabric still prevalent in the refined LULC map.

To detect missing urban fabric, we compared the total area of urban fabric (TOT_11XX_{LAU}) with the number of residents in 2018 for all Local Administrative Units (LAU) in Europe (POP_{LAU}), and calculated the respective population density of built-up areas per LAU ($POP\_DENSITY_{LAU}$). Moreover, we calculated the average (μ) and the standard deviation (σ) of the LAU population densities for each country. The following two cumulative criteria were applied to identify areas (i.e. LAUs) with possible built-up omissions:

1. ($POP_{LAU} > 0$) AND ($TOT\_11XX_{LAU} = 0$ pixels)
2. ($POP_{LAU} > 0$) AND ($POP\_DENSITY_{LAU} > (\mu_{COUNTRY} + 2\sigma_{COUNTRY})$) AND ($TOT\_11XX_{LAU} < 12.5$ ha)

The identified LAUs were enriched by adding built-up polygons from TomTom dataset. Without built-up density information on these patches, such areas were deemed 1122 *Low density urban fabric*[11].

The overall intervention was marginal since only 1.6% of the almost 100 thousand LAUs were affected. The use of TomTom built-up data is limited to these areas because, contrary to other Earth Observation data sources used in this project, its consistency and quality cannot be assured.

Step 9 – Final smoothing

The combination of many and heterogeneous datasets during the CLC refinement steps leads inevitably to fragmentation of LULC patches and a lowering of the actual MMU in the final map. In step 9 we apply a custom majority filter to mitigate this issue.

The filter consists of a moving window of 3X3 pixels of 50 m and the removal of LULC patches of less than 4 contiguous pixels (< 1 ha)[12]. In addition, patches of the class *Isolated or very low density urban fabric* (1123) with an areal extent less than 5 ha and completely surrounded by a high density urban fabric were recoded as class 1111.

It should be underlined that this intervention is carried to achieve smoother LULC patches and overall less 'noisy' appearance of the final product, and does not fully enforce the MMU specification, especially that of 5 ha for non-artificial surfaces. As mentioned earlier, the MMU was only forced at the level of pre-processing of input data, but not after their integration in the LULC map.

Figure 2 shows an extract of the various intermediate steps of the CLC refinement process, from 1 to 9. Figure 3 highlights just the original CLC and final LUISA Base Map 2018. The differences are notorious visually.

Resampling to 100 m spatial resolution

The native version of LUISA Base map 2018 is delivered at 50 m resolution which, compared to the 2012 version (produced at 100 m resolution), represents a substantial improvement. However, applications like the LUISA model are optimized to work with 100 m LULC data. For this and other applications, we created a 100 m distribution of the LUISA Base Map based on a custom categorical resampling method. This method was developed with a two-fold purpose: to preserve as much as possible the identified built-up areas and the spatial contiguity of linear features (i.e. roads and rivers) that characterizes the original map at 50 m resolution. The resampling method works as follows:

1. For each target 100 m cells, an array of occurrences of classes in the underlying 4 source 50 m pixels was calculated.

[11] The rationale for this assumption is: the likelihood of a built-up area being captured by EO is higher for denser urban fabric. Hence, such areas are likely to show overall low built-up density.

[12] Queen's contiguity.

2. If a class is predominant (frequency > 50%) then it was transferred to the target pixel (equivalent of applying a majority filter only to the 4 source pixels covered by the target pixel).

3. If no class is predominant:
   3.1. For target pixels with ambiguous classification of the kind 50% (4 pixels of two different classes):
      3.1.1. If one of the classes is artificial (1XXX) or river (5110) and the other not (2XXX, 3XXX, 4XXX and 5XXX except 5110), the artificial class or river (1XXX, 5110) was transferred to the target pixel.
      3.1.2. If both classes belong to either artificial or river LULC classes (1XXX, 5110), the selected class depends on a priority list (see Annex 5).
      3.1.3. If both classes were not artificial (2XXX, 3XXX, 4XXX, 5XXX-except 5110) then an extended majority filter was applied (i.e. wider window of 4 x 4 cells of 50 m), the class with highest frequency (mode) is picked.
   3.2. For target pixels with ambiguous classification of the kind 25% (4 pixels of 4 different classes), the majority filter was applied as in point 3.1.3

Figure 4 shows an extract of the LUISA Base map at the native, 50 m resolution, and the resampled 100 m resolution.

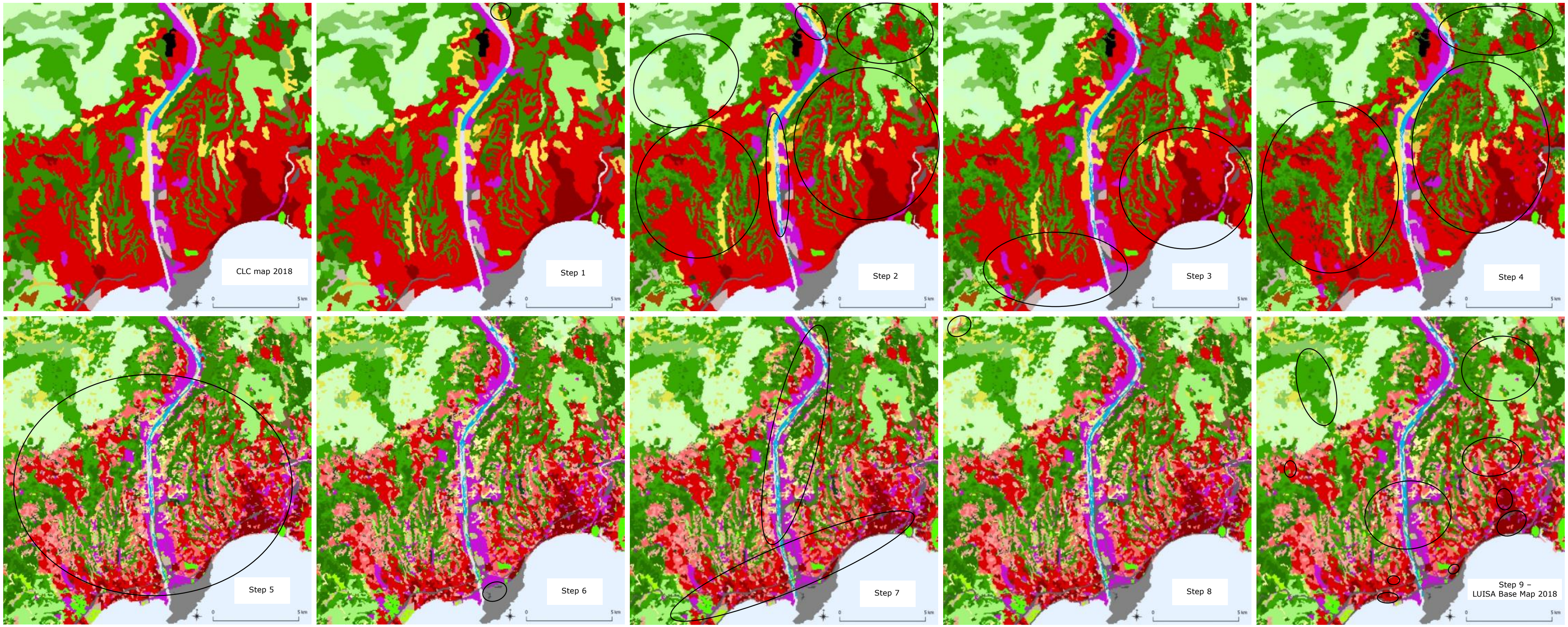


Figure 2. Sequential improvement in spatial/thematic detail for Nice, in France, after each step of the refinement process. Circled areas show the main enhancements compared to the previous step.

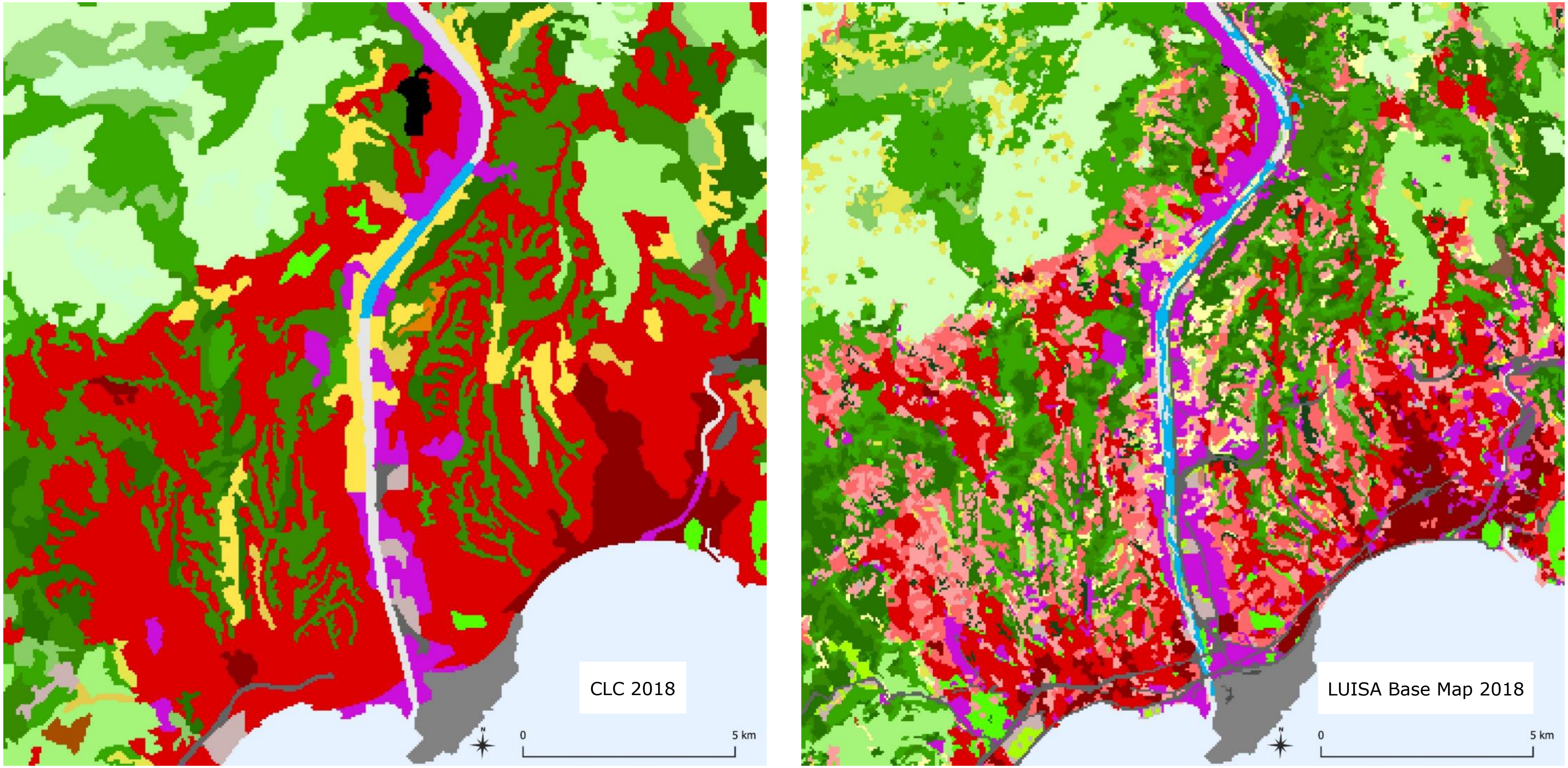


Figure 3. Difference between the original CLC 2018 and the LUISA Base Map 2018.

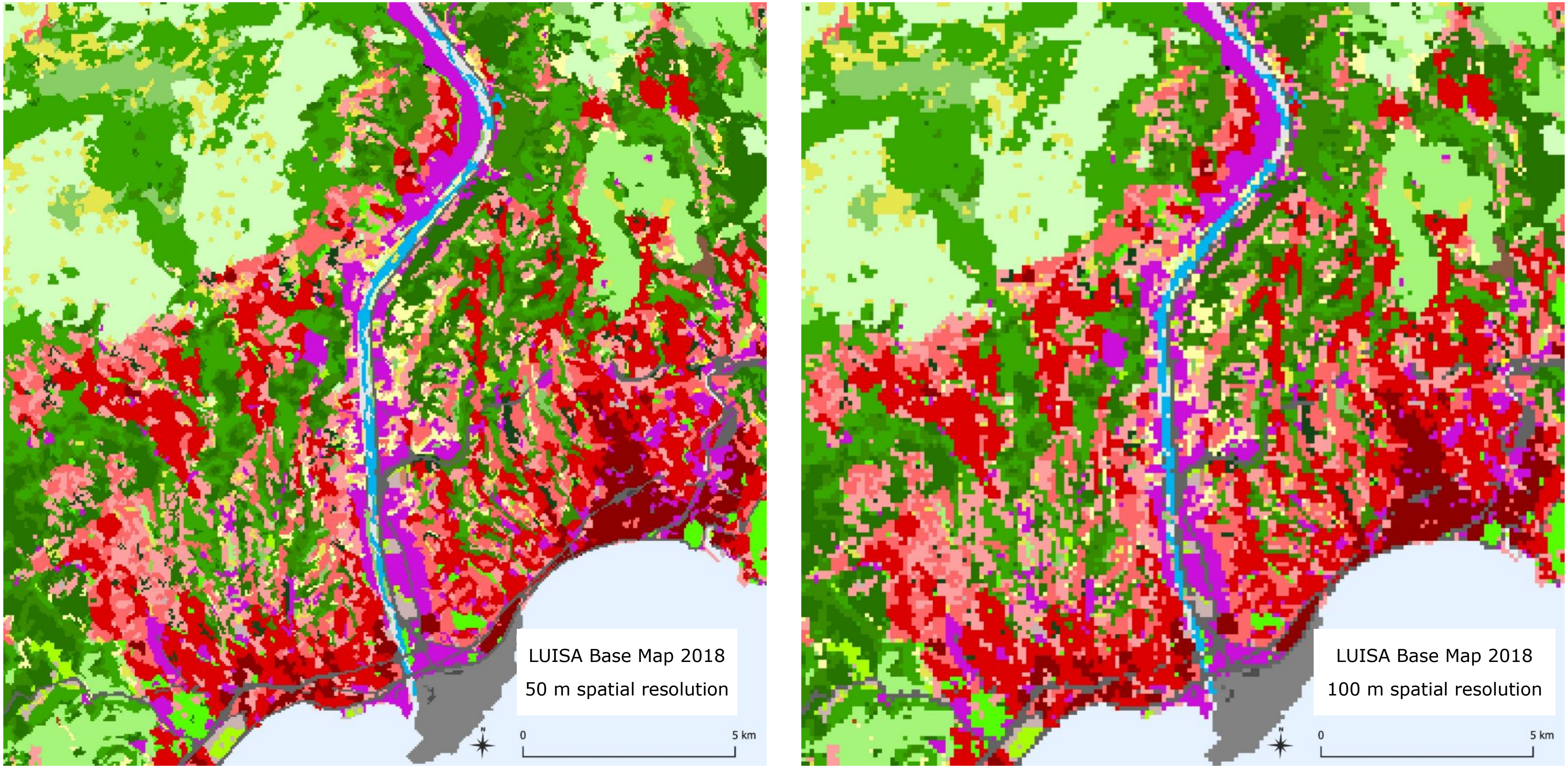


Figure 4. LUISA Base Map 2018 at 50 m spatial resolution (left) and the resampled version at 100m spatial resolution (right).

3 Discussion and conclusions

The LUISA Base Map 2018 is a detailed land use and land cover map constructed by integrating information from multiple off-the-shelf, trusted datasets, including the CORINE Land Cover maps, Copernicus Earth Observation products, TomTom Multinet and OpenStreetMap. The approach used to construct the LUISA Base Map 2018 results from the consolidation of previously documented developments, and applies now a streamlined, fully automated and reproducible chain of data fusion algorithms.

Compared to the CORINE Land Cover, the LUISA Base Map has significantly higher spatial and thematic resolutions, allowing more detailed assessments. It is publicly available at 50 and 100 metre cell sizes, covering most of the European continent, including the EU27, EFTA countries, the UK and the Western Balkans.

During the development of the LUISA Base map, the results were inspected carefully at each step to make sure the algorithms produced the expected outputs. The quality of the final product is intrinsically linked to the quality of all its inputs. Products from the Copernicus Land Monitoring Service (CLCC, HRL, UA, CZ) and JRC (GHSL) are generally validated by their producers, but any errors present in the input data inevitably propagate to the LUISA Base Map. The final accuracy of the LUISA Base Map 2018 has not been systematically and formally (i.e. statistically) assessed yet, but the extensive quality assessments documented for the versions 2006 and 2012 support the overall approach adopted.

A proper validation is also challenging. To be meaningful, the validation of a LULC map must be carried ideally at the spatial and thematic resolution of its specifications. There are no other EU wide LULC products with similar specifications to the LUISA Base Map. And the EU-wide, mono-thematic layers (e.g. focusing on specific land covers) produced by Copernicus that could be used in the validation were used themselves in the production of the LUISA Base Map, thus invalidating their use as independent validation datasets.

The data integration was performed in view of preserving the detail of each individual input, while being coherent with the general framework of a LULC map, and in particular with the CLC nomenclature. This makes the LUISA Base Map more than just the sum of its inputs. As such, we believe that the risks related to the unknown final accuracy of the LUISA Base Map are outweighed by its overall benefits.

The LUISA Base Map has its own limitations too. Worth mentioning is the inconsistency of detail arising from the integration of layers such as the UA and CZ which have a limited geographical extent (selection of urban areas and the coastal zones, respectively); as well as from the integration of the TomTom built-up polygons in specific municipalities. Confronted with the trade-off between maximum product's consistency and maximum detail wherever possible, we opted for the latter. Another limitation is the higher focus on the thematic detail of artificial surfaces, in view of applications such as population modelling and urban analyses. The thematic detail of non-artificial surfaces has not been enhanced, and it has even been reduced in the case of wetland areas.

The fully automated methodology allowed us to reproduce the LUISA Base Map 2012. For this purpose, we used the same data sources used for the 2018 version, but reference data from 2012. The employment of the revised and improved methodology herein described to reproduce the LUISA Base Map 2012 yielded a more comparable version to the 2018. Full comparability is, however, not achievable due to differences in the methods used to generate the input data for the two reference years. This is the case of, for example, data from the Copernicus land monitoring service (i.e. High Resolution Layers) and the JRC (GHSL).

The JRC plans to continue maintaining the approach and the code, and to release any revisions to the LUISA Base Maps, should any corrections and improvements be deemed necessary.

The LUISA Base Map contributes to the knowledge base of the European Commission's Knowledge Centre for Territorial Policies[13], together with other key reference datasets such as the JRC-GEOSTAT Population Grid 2018[14] (Batista e Silva et al., 2021) and the Annual Regional Database of the European Commission (ARDECO)[15].

[13] https://knowledge4policy.ec.europa.eu/territorial

[14] https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat

[15] https://knowledge4policy.ec.europa.eu/territorial/ardeco-database_en

References

Batista e Silva, F., Dijkstra, L., & Poelman, H. (2021). *The JRC-GEOSTAT population grid 2018*. Publications office of the European Union. https://doi.org/in press

Batista e Silva, F., Freire, S., Schiavina, M., Rosina, K., Marín-Herrera, M. A., Ziemba, L., Craglia, M., Koomen, E., & Lavalle, C. (2020). Uncovering temporal changes in Europe's population density patterns using a data fusion approach. *Nature Communications*, *11*(1), 4631. https://doi.org/10.1038/s41467-020-18344-5

Batista e Silva, F., Gallego, J., & Lavalle, C. (2013). A high-resolution population grid map for Europe. *Journal of Maps*, *9*(1), 16–28. https://doi.org/10.1080/17445647.2013.764830

Batista e Silva, F., Lavalle, C., Baranzelli, C., Perpiña Castillo, C., Jacobs-Crisioni, C., Barranco, R., Zulian, G., Maes, J., Vandecasteele, I., Ustaoglu, E., Barbosa, A., & Mubareka, S. (2013). *Direct and indirect land use impacts of the EU cohesion policy - Assessment with the land use modelling platform*. https://doi.org/10.2788/60631

Batista e Silva, F., Lavalle, C., & Koomen, E. (2013). A procedure to obtain a refined European land use/cover map. *Journal of Land Use Science*, *8*(3), 255–283. https://doi.org/10.1080/1747423X.2012.667450

Feranec, J., Soukup, T., Hazeu, G., & Jaffrain, G. (2016). *European Landscape Dynamics - CORINE Land Cover Data*. CRC Press - Taylor & Francis Group.

Jacobs-Crisioni, C., Diogo, V., Perpiña Castillo, C., Baranzelli, C., Batista e Silva, F., Rosina, K., Kavalov, B., & Lavalle, C. (2017). *The LUISA Territorial Reference Scenario: A technical description*. Publications office of the European Union. https://doi.org/10.2760/902121

Lavalle, C., Batista e Silva, F., Baranzelli, C., Jacobs-Crisioni, C., Vandecasteele, I., Barbosa, A. L., Maes, J., Zulian, G., Perpiña Castillo, C., Barranco, R., & Vallecillo, S. (2016). Land Use and Scenario Modeling for Integrated Sustainability Assessment. In J. Feranec, T. Soukup, G. Hazeu, & G. Jaffrain (Eds.), *European Landscape Dynamics - CORINE Land Cover Data* (1st Editio, pp. 237–262). CRC Press - Taylor & Francis Group.

Rosina, K., Batista e Silva, F., Vizcaino, P., Marín Herrera, M., Freire, S., & Schiavina, M. (2020). Increasing the detail of European land use/cover data by combining heterogeneous data sets. *International Journal of Digital Earth*, *13*(5), 602–626. https://doi.org/10.1080/17538947.2018.1550119

List of abbreviations and definitions

BD	Built-up Density layer
BU	Built-up
CLC	CORINE Land Cover map
CLCC	CORINE Land Cover Change map
CLCr	Intermediate step of the refinement of the CLC map
DG REGIO	Directorate-General for Regional and Urban Policy
EFTA	European Free Trade Association
EO	Earth Observation
ESM	European Settlement Map
ETRS	European Terrestrial Reference System 1989
FUA	Functional Urban Area
GDAL	Geospatial Data Abstraction Library
GHSL	Global Human Settlement Layer
HRL	High Resolution Layer
IBU	Impervious Built-Up layer
IMD	Imperviousness Density layer
JRC	Joint Research Centre
LAEA	Lambert azimuthal equal-area projection
LAU	Local Administrative Unit
LUISA	Land Use-based Integrated Sustainability Assessment
MMU	Minimum Mapping Unit
NBU	Non Built-up
OGR	OpenGIS Simple Features Reference Implementation
OSM	Open Street Map
POI	Point of Interest
RB	River Binary layer
RD	Road Density layer
UA	Urban Atlas
WD	Water Density layer
WETD	Wetness Density layer

List of figures

List of tables

Annexes

Annex 1. LUISA Base Map 2018 nomenclature

The table below provides the LUISA Base Map expanded legend. The first three levels (level 1, level 2, level 3) correspond to the original CLC legend except for the Urban vegetation class (113). The fourth level includes the additional breakdown of the *Artificial Surfaces* group, from 11 to 17 land use classes. The *Wetlands* is kept at level 1 of CLC. Darkened areas in the table represent differences vis-à-vis the original CLC nomenclature.

Level 1	Level 1 Label	Level 2	Level 2 Label	Level 3	Level 3 Label	Level 4	Level 4 Label
1	Artificial surfaces	11	Urban fabric	111	Continuous urban fabric	1111	High density urban fabric
				112	Discontinuous urban fabric	1121	Medium density urban fabric
						1122	Low density urban fabric
						1123	Isolated or very low density urban fabric
				113	Urban vegetation	1130	Urban vegetation
		12	Industrial, commercial and transport units	121	Industrial or commercial units	1210	Industrial or commercial units
				122	Road and rail networks and associated land	1221	Road and rail networks and associated land
						1222	Major stations
				123	Port areas	1230	Port areas
				124	Airports	1241	Airport areas
						1242	Airport terminals
		13	Mine, dump and construction sites	131	Mineral extraction sites	1310	Mineral extraction sites
				132	Dump sites	1320	Dump sites
				133	Construction sites	1330	Construction sites
		14	Artificial, non-agricultural vegetated areas	141	Green urban areas	1410	Green urban areas
				142	Sport and leisure facilities	1421	Sport and leisure green
						1422	Sport and leisure built-up
2	Agricultural areas	21	Arable land	211	Non-irrigated arable land	2110	Non-irrigated arable land
				212	Permanently irrigated land	2120	Permanently irrigated land
				213	Rice fields	2130	Rice fields
		22	Permanent crops	221	Vineyards	2210	Vineyards
				222	Fruit trees and berry plantations	2220	Fruit trees and berry plantations
				223	Olive groves	2230	Olive groves
		23	Pastures	231	Pastures	2310	Pastures
		24	Heterogeneous agricultural areas	241	Annual crops associated with permanent crops	2410	Annual crops associated with permanent crops

CLC 1	Level 1 Label	CLC 2	Level 2 Label	CLC 3	Level 3 Label	CLC 4	Level 4 Label
				242	Complex cultivation patterns	2420	Complex cultivation patterns
				243	Land principally occupied by agriculture	2430	Land principally occupied by agriculture
				244	Agro-forestry areas	2440	Agro-forestry areas
3	Forest and seminatural areas	31	Forest and seminatural areas	311	Broad-leaved forest	3110	Broad-leaved forest
				312	Coniferous forest	3120	Coniferous forest
				313	Mixed forest	3130	Mixed forest
		32	Shrub and/or herbaceous vegetation associations	321	Natural grassland	3210	Natural grassland
				322	Moors and heathland	3220	Moors and heathland
				323	Sclerophyllous vegetation	3230	Sclerophyllous vegetation
				324	Transitional woodland shrub	3240	Transitional woodland shrub
				331	Beaches, dunes and sand plains	3310	Beaches, dunes and sand plains
				332	Bare rock	3320	Bare rock
		33	Open spaces with little or no vegetation	333	Sparsely vegetated areas	3330	Sparsely vegetated areas
				334	Burnt areas	3340	Burnt areas
				335	Glaciers and perpetual snow	3350	Glaciers and perpetual snow
4	Wetlands	40	Wetlands	400	Wetlands	4000	Wetlands
5	Water bodies	51	Inland waters	511	Water courses	5110	Water courses
				512	Water bodies	5120	Water bodies
				521	Coastal lagoons	5210	Coastal lagoons
		52	Marine waters	522	Estuaries	5220	Estuaries
				523	Sea and ocean	5230	Sea and ocean

Annex 2. Correspondence between TomTom and CLC nomenclature.

TomTom Multinet		2018 CORINE Land Cover Map	
Class	**Feature Name**	**CLC 3**	**Level 3 Label**
7110	Building (Area)	-	-
7170	Park/Garden	141	Green urban areas
7180	Island	-	-
9353	Company Ground	121	Industrial or commercial units
9715	Industrial Area	121	Industrial or commercial units
9720	Industrial Harbour Area	123	Port areas
9730	Freeport	121	Industrial or commercial units
9731	Abbey Ground	142	Sport and leisure facilities
9732	Airport Ground	124	Airports
9733	Amusement Park Ground	142	Sport and leisure facilities
9734	Arts Centre Ground	142	Sport and leisure facilities
9737	Castle (not to visit) Ground	142	Sport and leisure facilities
9738	Castle (to visit) Ground	142	Sport and leisure facilities
9739	Church Ground	142	Sport and leisure facilities
9740	City Hall Ground	121	Industrial or commercial units
9741	Courthouse Ground	121	Industrial or commercial units
9742	Fire Station Ground	121	Industrial or commercial units
9743	Fortress Ground	142	Sport and leisure facilities
9744	Golf Course Ground	142	Sport and leisure facilities
9745	Government Building Ground	121	Industrial or commercial units
9746	Hippodrome Ground	142	Sport and leisure facilities
9748	Hospital Ground	121	Industrial or commercial units
9750	Library Ground	-	-
9751	Light House Ground	-	-
9753	Monastery Ground	142	Sport and leisure facilities
9754	Museum Ground	142	Sport and leisure facilities
9756	Parking Area Ground	-	-
9757	Petrol Station Ground	-	-
9758	Place of Interest Building	-	-
9759	Monument Ground	142	Sport and leisure facilities
9760	Police Office Ground	121	Industrial or commercial units
9761	Prison Ground	121	Industrial or commercial units
9762	Railway Station Ground	122	Road and rail networks and associated land
9763	Recreational Area Ground	141	Green urban areas
9767	Sports Hall Ground	142	Sport and leisure facilities
9768	Stadium Ground	142	Sport and leisure facilities
9769	State Police Office Ground	121	Industrial or commercial units
9770	Theatre Ground	142	Sport and leisure facilities
9771	University or College Ground	121	Industrial or commercial units
9774	Water Mill Ground	-	-
9775	Zoo Ground	142	Sport and leisure facilities
9776	Runway	124	Airports
9777	Post Office Ground	121	Industrial or commercial units
9778	Windmill Ground	-	-
9780	Institution	121	Industrial or commercial units
9788	Cemetery Ground	-	-
9790	Shopping Center Ground	121	Industrial or commercial units
9791	School Ground	121	Industrial or commercial units

Annex 3. Matrix of decision rules for conversion of UA - CLC nomenclature

The following table shows the decision matrix where the rows represent the CLC map and the columns represent the UA map with its respective nomenclature.

<table>
<tr><th></th><th>1111 1121 1122 1123 1130</th><th>1210</th><th>1221 1222 1223</th><th>1230</th><th>1240</th><th>1310</th><th>1330</th><th>1340</th><th>1410</th><th>1420</th><th>2000</th><th>2100</th><th>2200</th><th>2300</th><th>2400</th><th>2500</th><th>3100</th><th>3200</th><th>3300</th><th>4000</th><th>5000</th></tr>
<tr><td>111</td><td rowspan="40">Integrated in Built-up Density layer and then processed altogether in step5</td><td>121</td><td>122</td><td>123</td><td>124</td><td>111</td><td>133</td><td>111</td><td>141</td><td>142</td><td>111</td><td rowspan="11">Closest CLC 21X value</td><td rowspan="14">Closest CLC 22X value</td><td>231</td><td rowspan="18">Closest CLC 24X value</td><td>222</td><td rowspan="22">Closest CLC 31X value</td><td rowspan="25">Closest CLC 32X value</td><td rowspan="29">Closest CLC 33X value</td><td rowspan="40">Integrated with HRL Water & Wetness, then treated altogether in step2</td><td rowspan="40">Integrated with HRL Water & Wetness, then treated altogether in step 2</td></tr>
<tr><td>112</td><td>121</td><td>122</td><td>123</td><td>124</td><td>112</td><td>133</td><td>112</td><td>141</td><td>142</td><td>112</td><td>231</td><td>222</td></tr>
<tr><td>121</td><td>121</td><td>122</td><td>123</td><td>124</td><td>121</td><td>133</td><td>121</td><td>141</td><td>142</td><td>121</td><td>231</td><td>222</td></tr>
<tr><td>122</td><td>121</td><td>122</td><td>123</td><td>124</td><td>122</td><td>133</td><td>122</td><td>141</td><td>142</td><td>122</td><td>231</td><td>222</td></tr>
<tr><td>123</td><td>121</td><td>122</td><td>123</td><td>124</td><td>123</td><td>133</td><td>123</td><td>141</td><td>142</td><td>123</td><td>231</td><td>222</td></tr>
<tr><td>124</td><td>121</td><td>122</td><td>123</td><td>124</td><td>124</td><td>133</td><td>124</td><td>141</td><td>142</td><td>124</td><td>231</td><td>222</td></tr>
<tr><td>131</td><td>121</td><td>122</td><td>123</td><td>124</td><td>131</td><td>133</td><td>131</td><td>141</td><td>142</td><td>131</td><td>231</td><td>222</td></tr>
<tr><td>132</td><td>121</td><td>122</td><td>123</td><td>124</td><td>132</td><td>133</td><td>132</td><td>141</td><td>142</td><td>132</td><td>231</td><td>222</td></tr>
<tr><td>133</td><td>121</td><td>122</td><td>123</td><td>124</td><td>133</td><td>133</td><td>133</td><td>141</td><td>142</td><td>133</td><td>231</td><td>222</td></tr>
<tr><td>141</td><td>121</td><td>122</td><td>123</td><td>124</td><td>141</td><td>133</td><td>141</td><td>141</td><td>142</td><td>141</td><td>231</td><td>222</td></tr>
<tr><td>142</td><td>121</td><td>122</td><td>123</td><td>124</td><td>142</td><td>133</td><td>142</td><td>141</td><td>142</td><td>142</td><td>231</td><td>222</td></tr>
<tr><td>211</td><td>121</td><td>122</td><td>123</td><td>124</td><td>211</td><td>133</td><td>211</td><td>141</td><td>142</td><td>211</td><td>211</td><td>231</td><td>222</td></tr>
<tr><td>212</td><td>121</td><td>122</td><td>123</td><td>124</td><td>212</td><td>133</td><td>212</td><td>141</td><td>142</td><td>212</td><td>212</td><td>231</td><td>222</td></tr>
<tr><td>213</td><td>121</td><td>122</td><td>123</td><td>124</td><td>213</td><td>133</td><td>213</td><td>141</td><td>142</td><td>213</td><td>213</td><td>231</td><td>222</td></tr>
<tr><td>221</td><td>121</td><td>122</td><td>123</td><td>124</td><td>221</td><td>133</td><td>221</td><td>141</td><td>142</td><td>221</td><td rowspan="26">Closest CLC 21X value</td><td>221</td><td>231</td><td>222</td></tr>
<tr><td>222</td><td>121</td><td>122</td><td>123</td><td>124</td><td>222</td><td>133</td><td>222</td><td>141</td><td>142</td><td>222</td><td>222</td><td>231</td><td>222</td></tr>
<tr><td>223</td><td>121</td><td>122</td><td>123</td><td>124</td><td>223</td><td>133</td><td>223</td><td>141</td><td>142</td><td>223</td><td>223</td><td>231</td><td>222</td></tr>
<tr><td>231</td><td>121</td><td>122</td><td>123</td><td>124</td><td>231</td><td>133</td><td>231</td><td>141</td><td>142</td><td>231</td><td rowspan="23">Closest CLC 22X value</td><td>231</td><td>222</td></tr>
<tr><td>241</td><td>121</td><td>122</td><td>123</td><td>124</td><td>241</td><td>133</td><td>241</td><td>141</td><td>142</td><td>241</td><td>231</td><td>241</td><td>222</td></tr>
<tr><td>242</td><td>121</td><td>122</td><td>123</td><td>124</td><td>242</td><td>133</td><td>242</td><td>141</td><td>142</td><td>242</td><td>231</td><td>242</td><td>222</td></tr>
<tr><td>243</td><td>121</td><td>122</td><td>123</td><td>124</td><td>243</td><td>133</td><td>243</td><td>141</td><td>142</td><td>243</td><td>231</td><td>243</td><td>222</td></tr>
<tr><td>244</td><td>121</td><td>122</td><td>123</td><td>124</td><td>244</td><td>133</td><td>244</td><td>141</td><td>142</td><td>244</td><td>231</td><td>244</td><td>222</td></tr>
<tr><td>311</td><td>121</td><td>122</td><td>123</td><td>124</td><td>311</td><td>133</td><td>311</td><td>141</td><td>142</td><td>311</td><td>231</td><td rowspan="18">Closest CLC 24X value</td><td>222</td><td>311</td></tr>
<tr><td>312</td><td>121</td><td>122</td><td>123</td><td>124</td><td>312</td><td>133</td><td>312</td><td>141</td><td>142</td><td>312</td><td>231</td><td>222</td><td>312</td></tr>
<tr><td>313</td><td>121</td><td>122</td><td>123</td><td>124</td><td>313</td><td>133</td><td>313</td><td>141</td><td>142</td><td>313</td><td>231</td><td>222</td><td>313</td></tr>
<tr><td>321</td><td>121</td><td>122</td><td>123</td><td>124</td><td>321</td><td>133</td><td>321</td><td>141</td><td>142</td><td>321</td><td>231</td><td>222</td><td rowspan="15">Closest CLC 31X value</td><td>321</td></tr>
<tr><td>322</td><td>121</td><td>122</td><td>123</td><td>124</td><td>322</td><td>133</td><td>322</td><td>141</td><td>142</td><td>322</td><td>231</td><td>222</td><td>322</td></tr>
<tr><td>323</td><td>121</td><td>122</td><td>123</td><td>124</td><td>323</td><td>133</td><td>323</td><td>141</td><td>142</td><td>323</td><td>231</td><td>222</td><td>323</td></tr>
<tr><td>324</td><td>121</td><td>122</td><td>123</td><td>124</td><td>324</td><td>133</td><td>324</td><td>141</td><td>142</td><td>324</td><td>231</td><td>222</td><td>324</td></tr>
<tr><td>331</td><td>121</td><td>122</td><td>123</td><td>124</td><td>331</td><td>133</td><td>331</td><td>141</td><td>142</td><td>331</td><td>231</td><td>222</td><td rowspan="11">Closest CLC 32X value</td><td>331</td></tr>
<tr><td>332</td><td>121</td><td>122</td><td>123</td><td>124</td><td>332</td><td>133</td><td>332</td><td>141</td><td>142</td><td>332</td><td>231</td><td>222</td><td>332</td></tr>
<tr><td>333</td><td>121</td><td>122</td><td>123</td><td>124</td><td>333</td><td>133</td><td>333</td><td>141</td><td>142</td><td>333</td><td>231</td><td>222</td><td>333</td></tr>
<tr><td>334</td><td>121</td><td>122</td><td>123</td><td>124</td><td>334</td><td>133</td><td>334</td><td>141</td><td>142</td><td>334</td><td>231</td><td>222</td><td>334</td></tr>
<tr><td>335</td><td>121</td><td>122</td><td>123</td><td>124</td><td>335</td><td>133</td><td>335</td><td>141</td><td>142</td><td>335</td><td>231</td><td>222</td><td>335</td></tr>
<tr><td>400*</td><td>121</td><td>122</td><td>123</td><td>124</td><td>400</td><td>133</td><td>400</td><td>141</td><td>142</td><td>400</td><td>231</td><td>222</td><td rowspan="6">Closest CLC 33X value</td></tr>
<tr><td>511</td><td>121</td><td>122</td><td>123</td><td>124</td><td>511</td><td>133</td><td>511</td><td>141</td><td>142</td><td>511</td><td>231</td><td>222</td></tr>
<tr><td>512</td><td>121</td><td>122</td><td>123</td><td>124</td><td>512</td><td>133</td><td>512</td><td>141</td><td>142</td><td>512</td><td>231</td><td>222</td></tr>
<tr><td>521</td><td>121</td><td>122</td><td>123</td><td>124</td><td>521</td><td>133</td><td>521</td><td>141</td><td>142</td><td>521</td><td>231</td><td>222</td></tr>
<tr><td>522</td><td>121</td><td>122</td><td>123</td><td>124</td><td>522</td><td>133</td><td>522</td><td>141</td><td>142</td><td>522</td><td>231</td><td>222</td></tr>
<tr><td>523</td><td>121</td><td>122</td><td>123</td><td>124</td><td>523</td><td>133</td><td>523</td><td>141</td><td>142</td><td>523</td><td>231</td><td>222</td></tr>
</table>

*When the overlapping of the CLC map and the UA map was applied (step5), the original CLC wetlands classes of level 3 (4XX) were already aggregated to level 1 (400).

Annex 4. Correspondence between CZ nomenclature and CLC nomenclature

The following table shows the correspondence between Coastal Zones nomenclature and CORINE Land Cover map nomenclature. To avoid the introduction of buffer artefacts, only classes of artificial surfaces were considered.

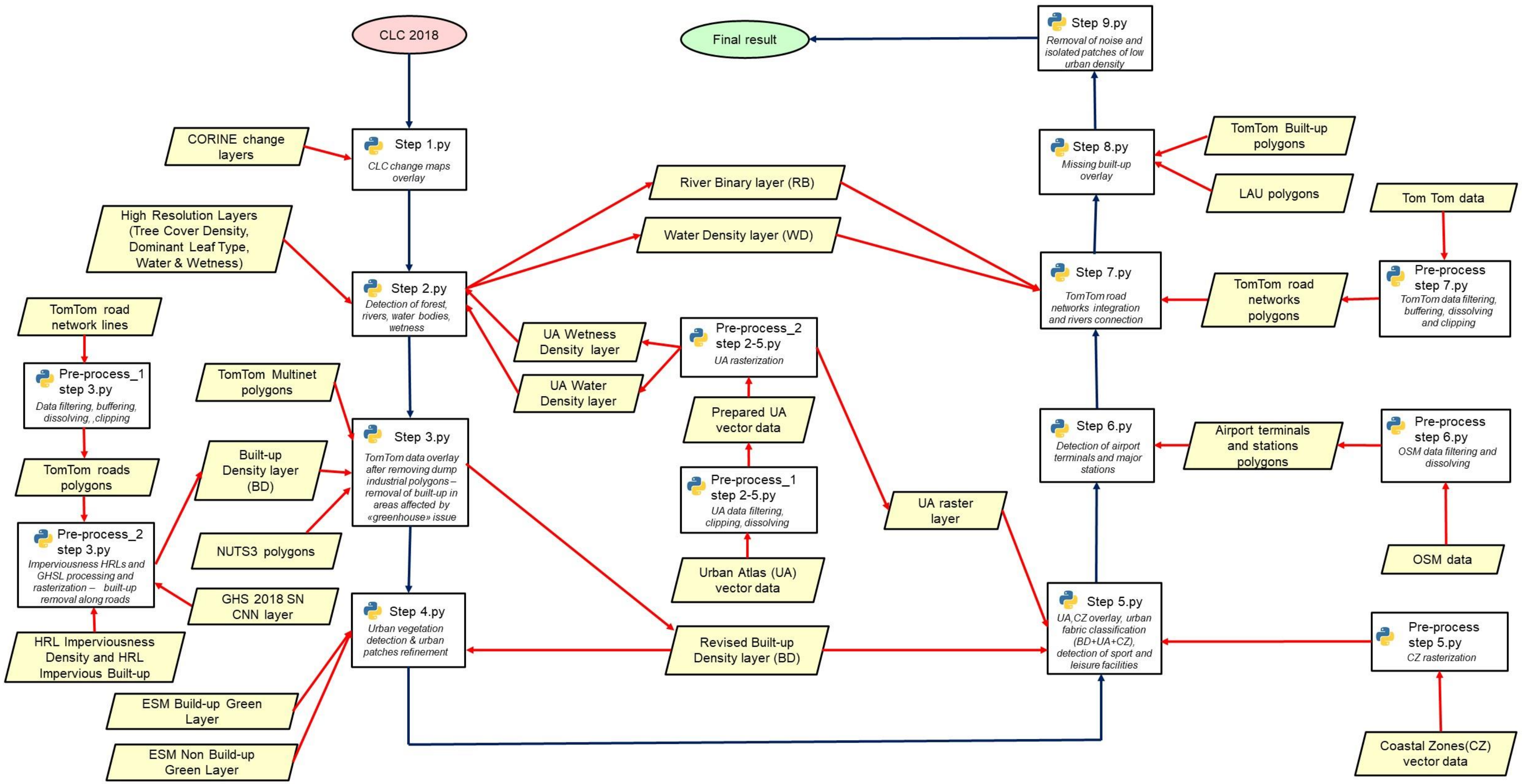
2018 Coastal Zones		2018 CORINE Land Cover Map	
CZ 4	**Level 4 Label**	**CLC 3**	**Level 3 Label**
1111	Continuous urban fabric (IMD ≥80%)	-	Integrated in Built-up Density layer
1112	Dense urban fabric (IMD ≥30-80%)	-	Integrated in Built-up Density layer
1113	Low density fabric (IMD <30%)	-	Integrated in Built-up Density layer
1121	Industrial, commercial, public and military units	121	Industrial or commercial units
1122	Nuclear energy plants and associated land	121	Industrial or commercial units
1210	Road networks and associated land	142	Road and rail networks
1220	Railways and associated land	124	Road and rail networks
1231	Cargo port	123	Port areas
1232	Passenger port	123	Port areas
1233	Fishing port	123	Port areas
1234	Naval port	123	Port areas
1235	Marinas	123	Port areas
1236	Local multi-functional harbours	123	Port areas
1237	Shipyards	123	Port areas
1240	Airports and associated land	124	Airports
1310	Mineral extraction sites	131	Mineral extraction sites
1320	Dump sites	142	Dump sites
1330	Construction sites	121	Construction sites
1340	Land without current use	-	-

Annex 5. Priority list used in the custom categorical resampling method

The table below provides the priority list for target pixels with ambiguous classification of the kind 50% and with both classes artificial (1XXX) or river (5110).

CLC 4	Level 4 label	Priority rank
1111	High density urban fabric	1
1121	Medium density urban fabric	2
1122	Low density urban fabric	3
1123	Isolated or very low density urban fabric	4
1124	Urban vegetation	11
1210	Industrial or commercial units	7
1221	Road and rail networks and associated land	5
1222	Major stations	12
1230	Port areas	13
1241	Airport areas	15
1242	Airport terminals	14
1310	Mineral extraction sites	16
1320	Dump sites	17
1330	Construction sites	18
1410	Green urban areas	8
1421	Sport and leisure green	9
1422	Sport and leisure built-up	10
5110	Water courses	6

Annex 6. Detailed data processing flowchart



KJ-NA-30663-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/503006

ISBN 978-92-76-34207-6