

## Horizon 2020

Space Call - Earth Observation: EO-3-2016: Evolution of Copernicus services

Grant Agreement No. 730008

# ECoLaSS

## Evolution of Copernicus Land Services based on Sentinel data



## D15.2

### “D45.1b – Prototype Report: New LC/LU Products (Issue 2)”

Issue/Rev.: 2.0

Date Issued: 28.12.2019

submitted by:



in collaboration with the consortium partners:



submitted to:



European Commission – Research Executive Agency

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 730008.

## CONSORTIUM PARTNERS

No.	PARTICIPANT ORGANISATION NAME	SHORT NAME	CITY, COUNTRY
1	GAF AG	GAF	Munich, Germany
2	Systèmes d'Information à Référence Spatiale SAS	SIRS	Villeneuve d'Ascq, France
3	JOANNEUM RESEARCH Forschungsgesellschaft mbH	JR	Graz, Austria
4	Université catholique de Louvain, Earth and Life Institute (ELI)	UCL	Louvain-la-Neuve, Belgium
5	German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Wessling	DLR	Wessling, Germany

### CONTACT:

GAF AG

Arnulfstr. 199 – D-80634 München – Germany




Phone: ++49 (0)89 121528 0 – FAX: ++49 (0)89 121528 79

E-mail: [copernicus@gaf.de](mailto:copernicus@gaf.de) – Internet: [www.gaf.de](http://www.gaf.de)

### DISCLAIMER:

The contents of this document are the copyright of GAF AG and Partners. It is released by GAF AG on the condition that it will not be copied in whole, in section or otherwise reproduced (whether by photographic, reprographic or any other method) and that the contents thereof shall not be divulged to any other person other than of the addressed (save to the other authorised officers of their organisation having a need to know such contents, for the purpose of which disclosure is made by GAF AG) without prior consent of GAF AG. Nevertheless single or common copyrights of the contributing partners remain unconditionally unaffected.

### DOCUMENT RELEASE SHEET

	NAME, FUNCTION	DATE	SIGNATURE
Author(s):	Alice Lhernould (SIRS) Katharina Schwab (GAF) Yoann Courmont (SIRS) Tatiana De-Wyse (SIRS) Antoine Masse (SIRS) Sophie Villerot (SIRS)	20.10.2019	
Review:	Sophie Villerot (SIRS)	28.12.2019	
Approval:	Eva Sevillano Marco (GAF)	28.12.2019	
Acceptance:	Massimo Ciscato (REA)		
Distribution:	Public		

### DISSEMINATION LEVEL

DISSEMINATION LEVEL		
PU	Public	X
CO	Confidential: only for members of the consortium (including the Commission Services)	

### DOCUMENT STATUS SHEET

ISSUE/REV	DATE	PAGE(S)	DESCRIPTION / CHANGES
1.0	02.08.2018	25	First version of WP45 Deliverable on the Prototype of a New Land Cover Product
2.0	28.12.2019	59	Final version of WP45 Deliverable on the Prototype of a New Land Cover Product. Major changes from sections 3-6, with final results from phase 2 implementation

### APPLICABLE DOCUMENTS

ID	DOCUMENT NAME / ISSUE DATE
AD01	Horizon 2020 Work Programme 2016 – 2017, 5 iii. Leadership in Enabling and Industrial Technologies – Space. Call: EO-3-2016: Evolution of Copernicus services. Issued: 13.10.2015
AD02	Guidance Document: Research Needs Of Copernicus Operational Services. Final Version issued: 30.10.2015

ID	DOCUMENT NAME / ISSUE DATE
AD03	Proposal: Evolution of Copernicus Land Services based on Sentinel data. Proposal acronym: ECoLaSS, Proposal number: 730008. Submitted: 03.03.2016
AD04	Grant Agreement – ECoLaSS. Grant Agreement number: 730008 – ECoLaSS – H2020-EO-2016/H2020-EO-2016, Issued: 18.10.2016
AD05	D22.1b - EO and other Data Requirements Report (Issue 2), Issued: 21.02.2019
AD06	D6.2: D31.1b - Methods Compendium: Sentinel-1/2/3 Integration Strategies, (Issue 2), Issued: December 2019
AD07	D7.2: D32.1b- Methods Compendium: Time Series Preparation, (Issue 2), Issued: 14.05.2019
AD08	D8.2: D33.1b - Methods Compendium: Time Series Analysis for Thematic Classification (Issue 2), Issued: December 2019
AD09	D11.2: D41.1b - Prototype Report: Time Series derived Indicators and Variables (Issue 2), Issued: December 2019
AD10	D13.2: D43.1b - Prototype Report: Improved Permanent Grassland (Issue 2), Issued: December 2019
AD11	D14.2: D44.1b - Prototype Report: Crop Area and Crop Status/Parameters (Issue 2), Issued: December 2019

## EXECUTIVE SUMMARY

The Horizon 2020 (H2020) project, “Evolution of Copernicus Land Services based on Sentinel data” (ECoLaSS) addresses the H2020 Work Programme 5 iii. Leadership in Enabling and Industrial technologies - Space, specifically the Topic EO-3-2016: Evolution of Copernicus services. ECoLaSS is being conducted from 2017–2019 and aims at developing and prototypically demonstrating selected innovative products and methods as candidates for future next-generation operational Copernicus Land Monitoring Service (CLMS) products of the pan-European and Global Components. ECoLaSS assesses the operational readiness of such candidate products and eventually suggests some of these for implementation. This shall enable the key CLMS stakeholders (i.e. mainly the Entrusted European Entities (EEE) EEA and JRC) to take informed decisions on potential procurement as (part of) the next generation of Copernicus Land services from 2020 onwards.

To achieve this goal, ECoLaSS makes full use of dense time series of Sentinel-2 and Sentinel-3 optical data as well as Sentinel-1 Synthetic Aperture Radar (SAR) data, complemented by Medium-Resolution (MR) Sentinel-3 optical data if needed and feasible. Rapidly evolving scientific developments as well as user requirements are continuously analyzed in a close stakeholder interaction process, targeting a future pan-European roll-out of new/improved CLMS products, and assessing the potential transferability to global applications.

This Deliverable **D15.2: “D.45.1b – Prototype Report: New Land Cover/Land Use (LC/LU) Products (Issue 2)”** constitutes the final report of the Work Package (WP) 45, which is centred around the creation of new LC/LU products, that will mainly be based on Sentinel data, optical and SAR, and will present an increased spatial resolution to fit all actual and future CLMS production requirements, in particular the new CORINE Land Cover Plus (CLC+) Backbone product. The applied methodologies build directly on the methods and processing lines developed in Task 3, especially WP33 “Time Series Analysis for Thematic Classification” [AD08], but also on the WP41 “Time Series-derived Indicators and Variables” [AD09] and the WP44 “Crop Area and Crop Status/Parameters” [AD11].

After the production of a cropland type HRL in the first phase, another development has been conducted: the setting of aggregation rules to combine individual HRL into a spatially and temporally coherent product. The creation of agri-environmental intensification indicator has been developed in the final report of WP 43 “Improved Permanent Grassland” [AD10], with the creation of an intensification layer, related to the count of mowing events.

This document is structured as follows: Section 1 of the document details the objectives of this WP45. Section 2 presents the backgrounds of previous LC/LU products and the summary of their limitations. Section 3 is a short presentation of the demonstration sites and section 4 describes all the applied time series analysis methods, based on the testing from Task 3. Section 5 lays out the prototype implementation approaches, from the integration of auxiliary data in the Sentinel workflow, the experimental set-up to the final validation of the prototype.

The ECoLaSS project follows a two-phased approach of two times 18 months duration. The first deliverable constitutes was issued at month 18, presenting preliminary results. In the second 18-month project cycle, a second issue of this deliverable is published, containing all relevant updates and final results.

## Table of Contents

<b>1</b>	<b>INTRODUCTION .....</b>	<b>1</b>
<b>2</b>	<b>BACKGROUND AND SUMMARY OF REQUIREMENTS .....</b>	<b>2</b>
<b>3</b>	<b>DEMONSTRATION SITES.....</b>	<b>3</b>
3.1	ECoLASS DEMONSTRATION SITES .....	3
3.2	DEMO-SITE SOUTH-WEST .....	4
3.3	DEMO-SITE CENTRAL .....	6
3.4	DEMO-SITE WEST .....	7
<b>4</b>	<b>PROTOTYPE IMPLEMENTATION: NEW LAND COVER.....</b>	<b>8</b>
4.1	DATA AND PROCESSING SETUP .....	8
4.1.1	Input Data and Data Integration.....	8
4.1.2	Pre-processing .....	11
4.1.3	Experimental Setup .....	12
4.2	CLASSIFICATION RESULTS AND VALIDATION .....	26
4.2.1	NLC Raster products for South West and Central demo sites .....	26
4.2.2	NLC Vector products for South West and Central demo sites.....	28
4.3	PROTOTYPE SPECIFICATIONS .....	33
<b>5</b>	<b>PROTOTYPE IMPLEMENTATION: HRL COMBINED LAYER.....</b>	<b>39</b>
5.1	DATA AND PROCESSING SETUP .....	40
5.2	EXPERIMENTAL SETUP.....	41
5.3	RESULTS .....	41
5.3.1	Demo-Site South-West .....	41
5.3.2	Demo-Site Central .....	44
5.3.3	Demo-Site West.....	50
5.3.4	Conclusion .....	54
5.4	PROTOTYPE SPECIFICATIONS .....	55
<b>6</b>	<b>CONCLUSION AND OUTLOOK .....</b>	<b>58</b>
<b>7</b>	<b>REFERENCES .....</b>	<b>59</b>

## List of Figures and Tables

Figure 3-1: - Biogeographic Regions of Europe (2015) and European ECoLaSS Demonstration Sites .....	3
Figure 3-2: - South-West Prototype Site, draped over CLC dataset (2018).....	5
Figure 3-3: - Central demonstration site, draped over CLC 2018.....	6
Figure 3-4: - West demonstration site draped over CLC 2018.....	7
Figure 4-1: Experimental Setup for the production of the raster classification.....	13
Figure 4-2: Hardbone over the South West demonstration site.....	15
Figure 4-3: Hardbone over the Central demonstration site.....	16
Figure 4-4: Zoom on the softbone over the Central demonstration site .....	17
Figure 4-5: Original distribution of bioregion on the South West Demosite .....	18
Figure 4-6: Simplified bioregion and spatial distribution of the samples on the South West Demosite ....	18
Figure 4-7: Original distribution of bioregion on the Central Demosite .....	19
Figure 4-8: Simplified bioregion and spatial distribution of the samples on the Central Demosite .....	19
Figure 4-9: – Prototype NLC_2018_010m_CE_03035_prototype_v01 over the Central site. ....	27
Figure 4-10: - Prototype NLC_2018_010m_SW_03035_prototype_v01 over the South-West site. ....	27
Figure 4-11: – Prototype NLC_2018_010m_SW_03035 vector version over the South-West site. ....	31
Figure 4-12: – Prototype NLC_2018_010m_CE_03035 vector version over the Central.....	32
Figure 4-13: – Illustration of Softbone issue on specific object that are not in ITT CLC+ nomenclature....	32
Figure 5-1: - Figure extracted from the ITT EEA/DIS/R0/19/012 - CLC with a 1 km raster/grid superimposed (top) illustrating the difference between encoding a particular unit as raster pixel (centre) or a grid cell (bottom).....	39
Figure 5-2: – Overview of the HRL combined layer over the <i>South-West</i> demonstration site.....	42
Figure 5-3: – Left: No Layer in Mountainous Areas over the Pyrenees, Spain. Right: No Layer over agricultural areas near Toulouse, France. ....	43
Figure 5-4: – Example of overlaps between IMD and TCM (blue) over Toulouse, France.....	43
Figure 5-5: – Example of overlaps between GRA and TCM (green) over the <i>South-West</i> demosite.....	44
Figure 5-6: – Overview of the HRL combined layer prototype over the <i>Central</i> demosite. ....	46
Figure 5-7: – No Layer in Mountainous Areas over the Zugspitze, Germany .....	46
Figure 5-8: – No Layer over swamps in Southern Bavaria, Germany.....	47
Figure 5-9: – No Layer over a quarry in Southern Bavaria, Germany .....	47
Figure 5-10: – No Layer over vineyards in the Rhine Valley near Mannheim, Germany .....	47
Figure 5-11: – Example of overlaps between IMD and TCM (blue) over Munich, Germany. ....	48
Figure 5-12: – Example of overlaps between IMD and TCM (yellow) alongside roads.....	48
Figure 5-13: – Example of overlaps between WAW and TCM (yellow) at the border between water bodies and tree covered areas at Bodensee, Germany.. ....	49
Figure 5-14: – Examples of overlaps between GRA and CRM (violet).....	49
Figure 5-15: – Overview of the HRL combined layer prototype over the <i>West</i> demosite.....	51
Figure 5-16: – Inconsistencies due to different crop masks in the <i>West</i> site .....	52
Figure 5-17: – Distribution of overlaps between GRA and CRM in the <i>West</i> site .....	53
Figure 5-18: – Example of overlaps between IMD and TCM (blue) over Brussels, Belgium. ....	54
Table 3-1: - Description of the selected Prototype Sites.....	4
Table 4-1: - Number of S-2 images used in the composite input dataset for each tile over the South-West site .....	9
Table 4-2: - Number of S-2 images used in the composite input dataset for each tile for the Central site .	9
Table 4-3: - Typology matching between the various Copernicus datasets finally used for both demonstration sites: CLC2028, RPZ, OSM, HRL.....	10
Table 4-4: CLC2018 lasses distribution in the ECoLaSS test and demosites (phase 2).....	21

Table 4-5: - List of rules to populate the vector layer created by the fusion of hard and soft bones, using the raster classification. ....	23
Table 4-6: - Adapted typology for the vector product. ....	24
Table 4-7: – Final nomenclature for the New Land Cover Prototype ....	25
Table 4-8: – Final nomenclature for the New Land Cover Prototype in vector format ....	25
Table 4-9: Independently established confusion matrix for the Central demonstration site – Plausibility analysis ....	26
Table 4-10: Independently established confusion matrix for the South-West site – Plausibility analysis .	26
Table 4-11: Independently established confusion matrix for the Central demonstration site vector product .....	29
Table 4-12: Independently established confusion matrix for the South West demonstration site vector product .....	30
Table 4-13: – Detailed specifications for primary 10m NLC status layer .....	34
Table 4-14: – Palette used for primary 10m NLC status layer .....	35
Table 4-15: – Detailed specifications for primary 10m NLCV status layer .....	36
Table 4-16: Palette used for primary 10m NLCV status layer .....	37
Table 4-17: Colour palette for the probability layers .....	38
Table 5-1: – Input layers and the referring code number .....	41
Table 5-2: – Distribution of pixels within layer combination in the <i>South-West</i> site.....	42
Table 5-3: – Distribution of pixels within layer combination in the <i>Central</i> site.....	45
Table 5-4: – Distribution of pixels within layer combination in the <i>West</i> site .....	50
Table 5-5: – Detailed specifications for primary 10m NLC status layer .....	56
Table 5-6: – Palette used for the 20m HRL combined layers .....	57



## Abbreviations

AOI	Area Of Interest
BRI	Brightness Index
CLMS	Copernicus Land Monitoring Service
CLC	CORINE Land Cover
CLC+	CORINE Land Cover Plus
CORINE	Coordination of Information on the Environment
CT	Classification Tree
ECoLaSS	Evolution of Copernicus Land Services based on Sentinel data
EAGLE	EIONET Action Group for Land monitoring in Europe
EEA	European Environment Agency
EEA-39	39 European Countries
EEE	Entrusted European Entities
EIONET	European Environment Information and Observation Network
EO	Earth Observations
ETM+	Enhanced Thematic Mapper Plus
EU	European Union
HR	High Resolution
HRL	High Resolution Layers
IMD	Imperviousness Degree
INSPIRE	Infrastructure for Spatial Information in the European Community
ITT	Invitation To Tender
JRC	Joint Research Centre
LAEA	Lambert Azimuthal Equal-Area (projection)
LC	Land Cover
LPIS	Land Parcel Identification System
LSMSS	Large-Scale Mean Shift Segmentation
LU	Land Use
LZW	Lempel-Ziv-Welch
MMU	Minimal Mapping Unit
MSS	Minimum Segment Size
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-InfraRed
NLC	New Land Cover
OA	Overall Accuracy
OSM	Open Street Map
PA	Producer Accuracy
RF	Random Forest
S-1	Sentinel-1
S-2	Sentinel-2
S-3	Sentinel-3
SAR	Synthetic Aperture Radar
SWIR	Short Wave InfraRed
TCD	Tree Cover Density
TIFF	Tag Image File Format
TM	Thematic Mapper
UA	User Accuracy
XML	Extensible Markup Language
WAW	Water And Wetness
WP	Work Package

# 1 Introduction

The Horizon 2020 (H2020) project, “Evolution of Copernicus Land Services based on Sentinel data” (ECoLaSS) addresses the H2020 Work Programme 5 iii. Leadership in Enabling and Industrial technologies - Space, specifically the Topic EO-3-2016: Evolution of Copernicus services. ECoLaSS is being conducted from 2017–2019 and aims at developing and prototypically demonstrating selected innovative products and methods as candidates for future next-generation operational Copernicus Land Monitoring Service (CLMS) products of the pan-European and Global Components. ECoLaSS assesses the operational readiness of such candidate products and eventually suggests some of these for implementation. This shall enable the key CLMS stakeholders (i.e. mainly the Entrusted European Entities (EEE) EEA and JRC) to take informed decisions on potential procurement as (part of) the next generation of Copernicus Land services from 2020 onwards.

To achieve this goal, ECoLaSS makes full use of dense time series of High-Resolution (HR) Sentinel-2 optical and Sentinel-1 Synthetic Aperture Radar (SAR) data, complemented by Medium-Resolution (MR) Sentinel-3 optical data if needed and feasible. Rapidly evolving scientific developments as well as user requirements are continuously analysed in a close stakeholder interaction process, targeting a future pan-European roll-out of new/improved CLMS products, and assessing the potential transferability to global applications.

This Deliverable **D15.2: “D45.1b – Prototype Report: New LC/LU Products (Issue 2)”** aims to demonstrate the feasibility of future new LC/LU products, exploiting the full potential of the Sentinel constellation, while remaining coherent with previous pan-European LC/LU mapping such as the CORINE Land Cover (CLC) time series. This work is part of WP 45 of Task 4: “Thematic Proof-of-Concept/Prototype on Continental/Global Scale”. This report is accompanied by the Deliverable **D15.4: “P.45.2b - Data Sets of New LC/LU Products”**. This report serves as documentation of the prototype datasets.

In the ECoLaSS project a prototype is defined as a prototypic/thematic proof-of-concept implementation of an improved or newly defined potential future Copernicus Land layer, building on the methods and processing lines developed in Task 3. The consortium has selected representative demonstration sites both in Europe and Africa, covering various bio-geographic regions and biomes. All prototype products and services are being prototypically implemented in a selection of these sites in the frame of the Task 4 WPs. In ECoLaSS, proofs-of-concept/prototype demonstrations are carried out with respect to five topics of relevance: (i) Time series derived indicators and variables, (ii) Incremental Updates of HR Layers, (iii) Improved permanent grassland identification, (iv) Crop area and crop status / parameters monitoring, and (v) New LC/LU products. This deliverable focuses on the prototype **New LC/LU Product** as part of WP45.

After methods had been tested by the Task 3 WPs (cf. Deliverables D31.1b, D32.1b, D33.1b, D34.1b, D35.1b) in all test sites and algorithms are described, the demonstration activities of Task 4 have set up the developed processing lines in demonstration sites and derive the final prototype versions. This deliverable focusses on the new LC/LU products envisioned in the framework of the upcoming CLC+, in particular the CLC+ Backbone, whose Invitation To Tender (ITT) requirements have been finally released, and the future CLC+ Core, which is expected to be populated with other CLMS datasets, such as the HR Layers (HRLs).

This report comprises a chapter on the background to the New Land Cover Product Prototype and associated requirements (Chapter 2); a description of the Demonstration Sites where the prototype is implemented (Chapter 3); an overview of the methodologies carried over from the testing and benchmarking in Task 3 (Chapter 4); followed by a Chapter on the prototype implementation itself including a description of the dataset (Chapter 5); and a summary and outlook (Chapter 6).

## 2 Background and Summary of Requirements

This section discusses the implementation of the final version of the CLC+ products, as detailed in the recent ITT that came out in summer 2019. This ITT is the result of several workshop dedicated to the user requirement clarification steps, which have taken various shapes over the course of this project.

The CLC database is rightfully considered as the flagship LC/LU for Europe, produced by the European Environment Agency (EEA) and belonging to the CLMS portfolio. Through more than 25 years, it has provided consistent and coherent LC and LU information to stakeholders and decision-makers.

However, this consistency enforcement is based on satellite parameters dating back to 1990 – hence a minimal mapping unit (MMU) of 25ha for the status layers. Change layers have been introduced later and exhibit a MMU of 5ha, an vast improvement for users but leading to an irreconcilable disparity between change and status layers.

Through multiple consultation of users and stakeholders, the following subsets for the next generation of CLC+ product have been designed:

- CLC+ Backbone;
- CLC+ Core;
- CLC+ Instance;
- CLC+ Legacy.

The CLC+ Backbone, composed of raster and vector products, whose production call for tender has been released during the summer 2019, is based on the latest state-of-the-art LC products, whose details can be found in WP33 report [AD08].

Various products have been successfully created to map country and even continent(s) at high or medium spatial resolution and several examples based on Sentinel-2 (S-2) time series can be mentioned:

- At a national scale, the CESBIO created maps of French territory in 2017 based on Sentinel-2 with 17 land cover classes, at a 10m resolution, followed in 2018 by a refined version with 23 classes.
- The project S-2 for Global Land Cover (S2GLC), funded by the ESA, has demonstrated the feasibility of pan-European LC mapping using only S-2 time series for a given year.

The previously known limitations of some of the pioneer national or global automated LC mapping projects could be summarized as too low spatial and temporal resolutions, as well as some inconsistencies between the different datasets – however, those limitations are currently being lifted by the throughout use of the Sentinel constellations.

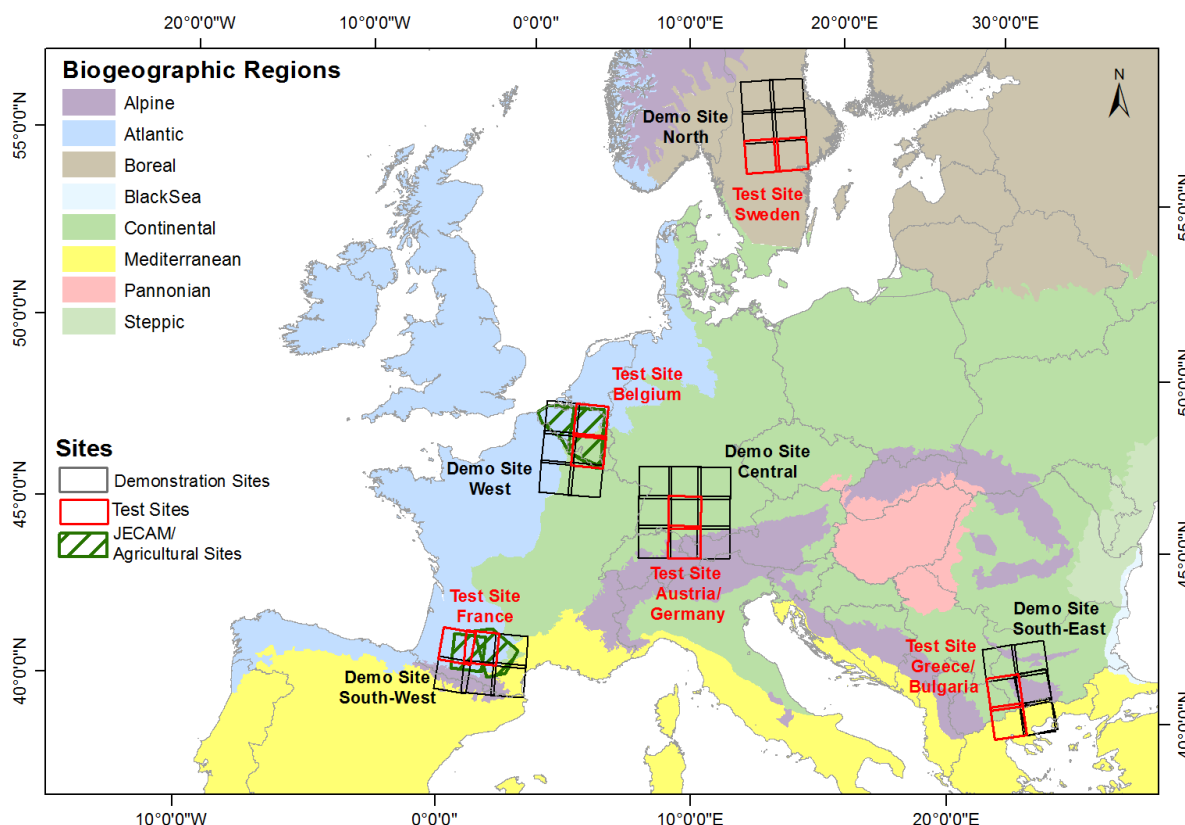
All those summarized issues still call for the emergence of new LC products, which should exhibit new properties to increase their spatial and temporal consistency, all the while enforcing a coherence with the previous datasets, constituting highly valuable time series of information on the biodiversity and LC changes in the landscape. The creation of a pan-European HR LC layer will be obtained by merging together all the currently available layers, in addition to a new agricultural layer. This merge constitutes an opportunity to also enforce a logical consistency, this time between the various current and upcoming thematic products, which are being produced independently, without requiring post-processing to ensure the spatial and temporal coherence of the whole set of data, produced in the framework of Copernicus.

## 3 Demonstration Sites

All prototypes are implemented in selected representative demonstration sites, which cover various biogeographic regions and biomes. The New Land Cover Product, based on CLC+ Backbone specifications, is implemented over the South-West and the Central demonstration sites, and the combined HRL layer over the West, South-West and Central demo-sites.

### 3.1 ECoLaSS Demonstration Sites

The selected larger prototype sites (60,000/90,000 km<sup>2</sup> per prototype site) contain the 5 test sites from Task 3. These prototype sites are relevant in Task 4 for demonstrating the proposed candidates for a Copernicus Land Service Evolution roll-out on a larger scale. As shown in , the pre-selected prototype sites cover the Atlantic and Continental zones (Source EEA: <http://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3#tab-gis-data>) of the member and associated states of EEA-39. The selected prototype sites are located in the North of Europe, in the Alpine/Central region, in the West, in the South-West and in the South-East of Europe. All prototype products and services will be prototypically implemented in one or more prototype sites in project phase 1, and in three prototype sites in phase 2.



**Figure 3-1: - Biogeographic Regions of Europe (2015) and European ECoLaSS Demonstration Sites**  
 (Map: © European Environment Agency; administrative boundaries: ©EuroGeographics)

A short description of the different prototype sites is given in the following Table 3-1 below:

**Table 3-1: - Description of the selected Prototype Sites**

Location	Biogeographical region(s)	Countries	Distribution of CORINE land cover classes 2018 (Level 1) per prototype site *
Northern Europe	Boreal	Sweden	Artificial areas: 1.90%, Agricultural areas: 11.87%, Forest and semi-natural areas: 69.01%, Wetlands: 3.25%, Waterbodies: 13.94%
Alpine / Central Europe	Continental, Alpine	Germany, Austria, Switzerland, Italy and France	Artificial areas: 9.03%, Agricultural areas: 44.55%, Forest and semi-natural areas: 44.65%, Wetlands: 0.23%, Waterbodies: 1.55%
West Europe	Atlantic, Continental	Belgium, France, Luxembourg, Netherlands	Artificial areas: 13.47%, Agricultural areas: 63.08%, Forest and semi-natural areas: 21.43%, Wetlands: 0.39%, Waterbodies: 1.61%
South-East Europe	Mediterranean, Continental, Alpine	Serbia, Macedonia, Greece, Bulgaria	Artificial areas: 3.34%, Agricultural areas: 34.87%, Forest and semi-natural areas: 56.67%, Wetlands: 0.17%, Waterbodies: 4.93%
South-West Europe	Atlantic, Mediterranean, Alpine	France, Spain	Artificial areas: 3.59%, Agricultural areas: 48.05%, Forest and semi-natural areas: 47.06%, Wetlands: 0.10%, Waterbodies: 1.18%

### 3.2 Demo-Site South-West

In the frame of the Task 4 where all prototype products and services are implemented, the South-West demonstration site has been chosen as the primary demonstration site for the improved HRL Imperviousness (IMP) as well as a potential future Copernicus New Land Cover Layer Product at high resolution. The demonstration site contains the test site “France” that has been studied in Task 3.

The South-West demonstration site is used as a prototypic / thematic proof-of-concept implementation of the outcome of the methods and processing lines developed in Task 3. The demonstration site covers the south of France, small parts of Spain and the Principality of Andorra, in the shape of 6 Sentinel-2 tiles. Landscapes are divided between different biogeographic regions such as Mediterranean, Alpine and Atlantic. Three Sentinel-2 tiles are dominated by mountain landscapes, a mix of bare soils and natural grasslands, due to the presence of the Pyrénées. The 31TCJ tile is dominated by a strong proportion of impervious soil, due to the presence of a major French city, Toulouse. The plains surrounding this urban area are mainly composed of croplands mixed with grassland and an increasing amount of forest with the proximity of the coastal region. In general, the Mediterranean area in the East is a patchwork of cropland, dry grassland and vineyard. We also can notice a small portion of the Landes forest in the Nord-West of the demonstration site. Major landscapes can be visualized on this Figure 3-2 of the CLC 2018 product over the area where Andorra is wrongly coded as waterbody, since it does not belong to the EU.



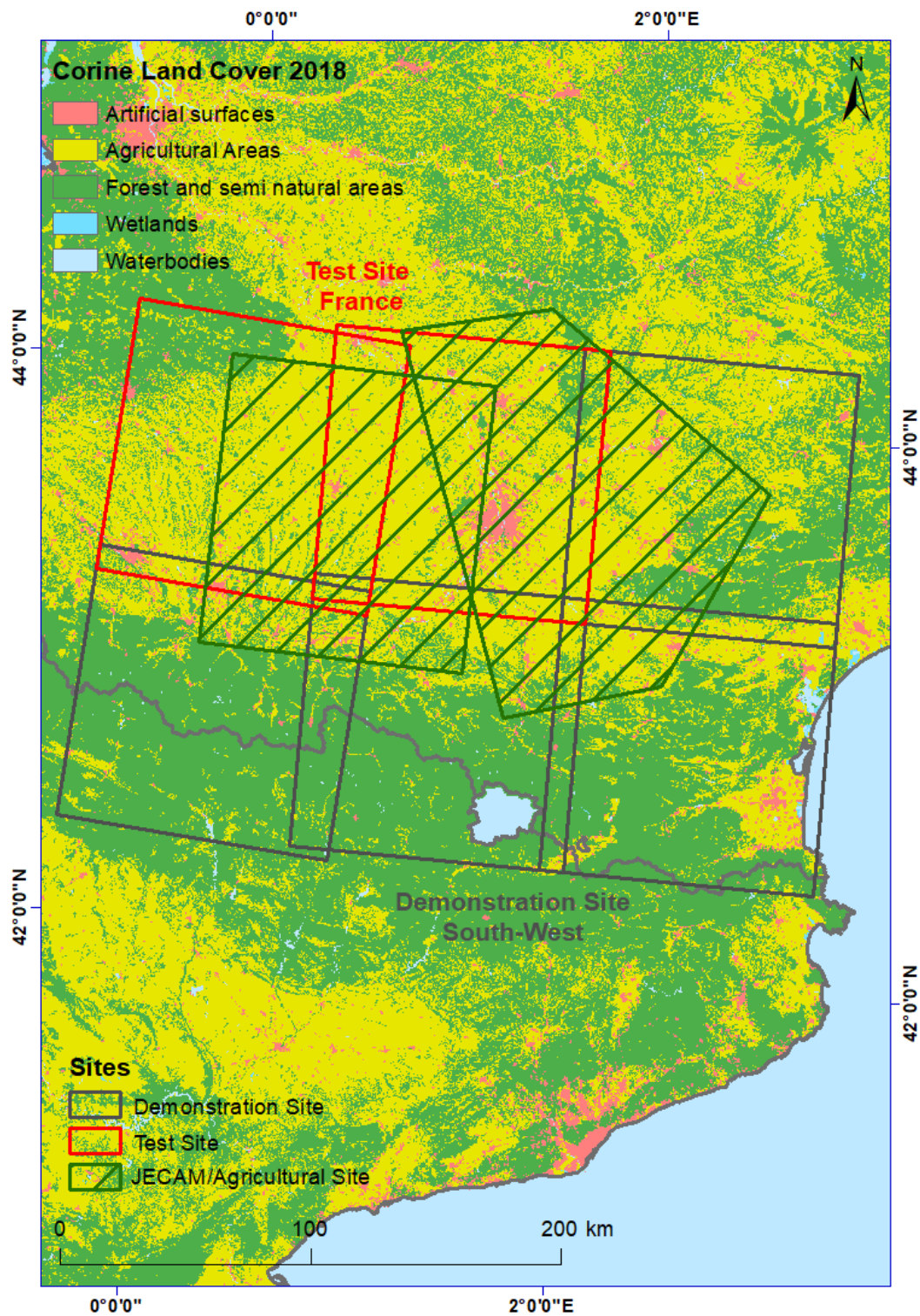


Figure 3-2: - South-West Prototype Site, draped over CLC dataset (2018)

### 3.3 Demo-site Central

The Central demonstration site, that is depicted in Figure 3-3, covers part of Germany, Austria, Switzerland, France, Italy and Liechtenstein. All prototypes are implemented over this area of interest. The landscape is covered with various types of croplands and grasslands. The Alpine Foreland is characterised by an annual precipitation rate which is significantly increasing towards the Alps, thus favouring grassland and limiting cropland use. Some tile present mountain-specific vegetation, such as forests, grasslands and wetlands, with dwarf pine and shrub areas above the tree line. The larger Alpine valleys are typically dominated by built-up areas and traffic infrastructure, with intermixed grassland and arable land areas.

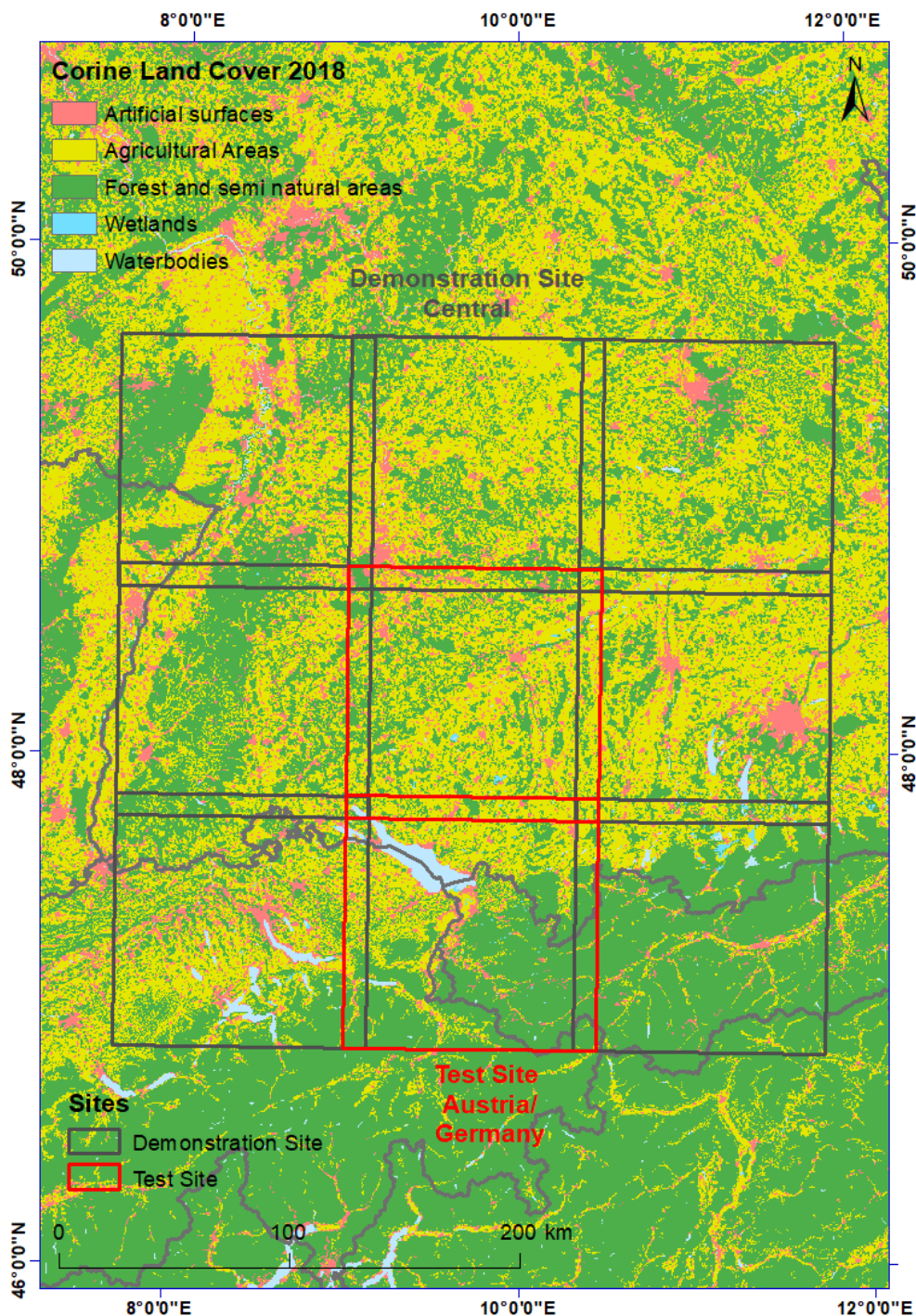


Figure 3-3: - Central demonstration site, draped over CLC 2018



### 3.4 Demo-site West

The demonstration site “West” is also a primary demonstration site for prototype developments of Task 4, related to the potential future Copernicus Land High Resolution Layer on Agricultural (AGRI), the combined HRL layer (NLC), various phenological products, and as well for improvements of the HRL Grassland (GRA) and Forest (FOR). It contains the test site “Belgium” where methodological developments were studied in Task 3. A map of the selected demonstration site West is depicted in Figure 3-4. As summarized in Table 3-1, half of the surface of the demonstration site is used for agricultural purposes, a quarter is covered by water bodies, and the last quarter is divided between forests and semi-natural surfaces (13.15%), artificial areas (7.81%) and wetlands (0.25%).

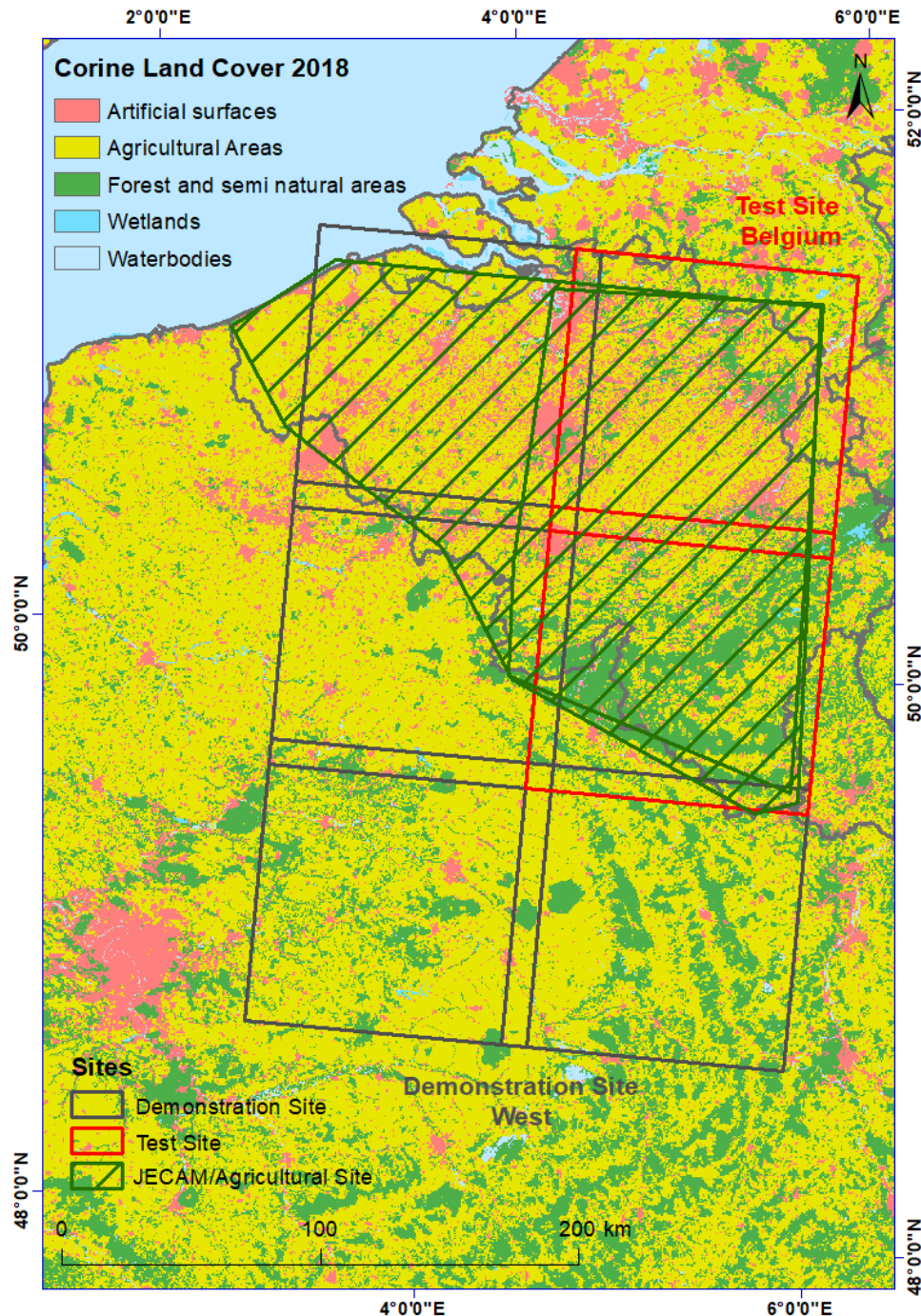


Figure 3-4: - West demonstration site draped over CLC 2018



## 4 Prototype Implementation: New Land Cover

This chapter shows the implementation for the New Land Cover prototype, composed of two products following the new CLC+ requirements at much as possible, all the while remaining in the framework of ECoLaSS. The two products are:

- A raster layer, with a basic typology, at a 10m spatial resolution and with a minimum mapping unit (MMU) of 10m by 10m (NLC\_2018\_010m\_SW\_03035\_prototype\_v01);
- A vector layer, based on a geometric skeleton, populated with the previous LC information, at a 10m resolution and with a MMU of 0.5ha (NLCV\_2018\_010m\_SW\_03035\_prototype\_v01).

The detailed methodologies explored in Task 3 to fine-tune the prototyping are described at length in the last version of the deliverable for WP 33 [AD08].

The section 4.1 tackles the data integration and the resulting processing setup. It should be noted that due to the release of the ITT during summer 2019, most of the results of phase 1, based on various preliminary reports following CLC+ workshops, has not been kept for clarity. The following section 0 presents the results and the validation and accuracy assessment of the and cover prototype and lastly, the dataset and its metadata, referring to the deliverable “**45.2b – Data Sets of New LC/LU Products (Issue 2)**” are described in the section 4.2.2.

### 4.1 Data and Processing Setup

Various pre-processing algorithms have been tested in the deliverable WP32 [AD07]. Since, even though the S-2 twin satellites are outperforming expectations in term of revisit time and spectral quality, as stated in WP22 [AD05], the pre-processing offered by ESA, Sen2Cor, remain weak to identify clouds and atmospheric veils. Over mountainous areas, this is particularly critical. Decision has been made to use the MAJA algorithm to treat S-2 L1C data over both demonstration sites, due to its robustness and use of time series to remove the perturbation due to the atmosphere.

The integration of the EO input data as well as ancillary data are described in section 4.1.1. The pre-processing of the Sentinel input data is explained in section 4.1.2 and the experimental set-up for preparing and performing the classification is detailed in section 4.1.3.

#### 4.1.1 Input Data and Data Integration

As explained in the deliverable WP 33 [AD08], focus has been put on optical data for both the raster classification, and the production of a segmentation, as expected from the CLC+ requirements. Complementary information on roughness and wetness coming from S-1 time series and on the texture from VHR images are outside the scope of the project and will therefore not be integrated into the vector layer.

Only S-2 time series from 2018 have been used for the soft bones and raster classification production, while the hard bones are based on ancillary data for 2018 and 2019.

##### 4.1.1.1 Image data

The raw input dataset for the creation of this prototype over the South-West demonstration site is composed by S-2 data from 2018/01/01 to 2018/11/12. When the cloud presence was too important (cloud cover over 90% of the usable image), images were rejected by MAJA and not integrated into the processing chain. The Table 4-1 shows the number of images used by S-2 tiles over the all demonstration site.

**Table 4-1: - Number of S-2 images used in the composite input dataset for each tile over the South-West site**

	30TYP	30TYN	31TCJ	31TCH	31TDJ	31TDH	total
<b>S-2 images</b>	65	86	81	88	66	84	470

Interpolated data, as explained in the deliverable of WP33 [AD08], are generated at fixed dates each month and based on the images before and after this artificial date. Weight are applied to the values of each pixel of the time series, taking in account the image right before and right after, but also the influence of further past and future dates – smoothing therefore any anomaly, such as cloudy pixels that slipped past the detection algorithm.

The raw datasets for the Central demonstration site are listed in the Table 4-2 and has also been processed using the MAJA algorithm for better consistency.

**Table 4-2: - Number of S-2 images used in the composite input dataset for each tile for the Central site**

	32TNT	32UNU	32UMV	32UMU	32TMT	32TPT	32UNV	32UPV	total
<b>S-2 images</b>	115	138	94	96	96	110	178	118	849

#### 4.1.1.2 Reference data

Several reference datasets were required for the creation of the hardbone. Those datasets were then also used for the calibration and validation of the classification model.

The hardbone is built upon:

- OSM data over France, Andorra and Spain for the South-West site and France, Germany, Austria, Liechtenstein and Switzerland for the Central site, where lines of road and railways have been selected, depending on their importance as well as rivers and water bodies information;
- EU-Hydro beta version: canals and rivers have been integrated when available in polygon form.

The data used for the calibration and validation of the classification model are:

- High-Resolution Layers 2015 that have been validated, to collect and generate samples – the Imperviousness product was used to create urban samples, the TCD layer for forest samples, WAW product for water bodies samples and the Grassland mask for grassland samples;
- CLC2018 was used to create samples Woody Shrubs, Rocks, Sand, Snow & Ice classes. These samples have also been verified by visual interpretation.

Due to the mismatch with numerous natural classes, the use of Urban Atlas 2012 data has been discarded at the demonstration site scale. The LUCAS 2018 dataset has not been integrated either, due to the lack of points for rare classes such as the evergreen deciduous trees, as discussed in deliverable of WP33 [AD08]. The spatial resolution of each point has also proved difficult to integrate at the S-2 scale.

Further detailed of each class matching is listed in Table 4-3.

**Table 4-3: - Typology matching between the various Copernicus datasets finally used for both demonstration sites: CLC2028, RPZ, OSM, HRL.**

Last version available of the CLC+ nomenclature	Adopted nomenclature on SW and Central test sites	Matching typology in Riparian Zone/Coastal Zone/ Natura 2000	Matching typology in CLC	Matching typology in other ancillary databases
1. Sealed (Buildings and flat sealed surfaces)	Sealed Areas	1.1.1.1 Continuous urban fabric	1.1.1 Continuous urban fabric	HRL 2015 IMP status layer IMD > 5%
		1.1.1.2 Dense urban fabric		
		-	1.1.2 Discontinuous urban fabric	
		1.1.1.3 Industrial or commercial units	1.2.1 Industrial or commercial units	
		1.2.1.1 Road network and associated land	1.2.2 Road and rail networks and associated land	
		1.2.1.3 Port areas	1.2.3 Port areas	
2. Woody needle leaved trees	Coniferous	3.2 Coniferous forest	3.1.2 Coniferous forest	HRL 2015 FOR>30% DLT status layer
3. Woody Broadleaved Deciduous	Broadleaves	3.1.3.1 Other natural & semi natural broadleaved forest	3.1.1 Broad-leaved forest	HRL 2015 FOR>30% DLT status layer
		3.1.5.1 Highly artificial broadleaved plantations		
4. Woody Broadleaved evergreen	Evergreen	3.1.4.1 Broadleaved evergreen forest	3.2.3 Sclerophyllous vegetation	-
5. Woody shrub	Clear cuts/Woody shrub	5 Heathland and scrub	3.2.4 Transitional woodland-shrub	-
			3.2.2 Moors and Heathlands	
6. Permanent herbaceous (grassland)	Grasslands	4 Grassland	3.2.1 Natural grasslands	HRL 2015 GRA status layer
			2.3.1 Pastures	
7. Periodically herbaceous (arable land)	Annual Crops	2.1.1.1 Non-irrigated arable land	2.1.1 Non-irrigated arable land	-
		2.1.3.1 Irrigated arable land and rice fields	2.1.2 Permanently irrigated land	

		2.1.4.1 Complex patterns of irrigated and non-irrigated arable land		
8. Lichens and mosses	Lichens and Mosses	7.2.1.2 Unexploited peat bog	4.1.2 Peatbogs	-
9. Sparsely vegetated	<i>Determined using zonal statistics in the final fusion</i>	6.1.1.1 Sparsely vegetated areas	3.3.3 Sparsely vegetated areas	-
10. Non-vegetated (Bare rocks, scree, sand, lichen, permanent bare soils)	Sand	6.2.1.1 Beaches	3.3.1 Beaches	-
		6.2.1.2 Dunes		
		6.2.1.3 River banks		
	Rocks	6.2.2.1 Bare rocks and rock debris	3.3.2 Bare rocks	
	Burnt areas	6.2.2.2 Burnt areas (except burnt forest)	3.3.4 Burnt areas	
11. Water	Water	9 Rivers and lakes	5.1.1 Water courses	HRL WaW 2015
			5.1.2 Water bodies	
		8 Lagoons, coastal wetlands and estuaries	5.2.1 Coastal lagoons	
			5.2.2 Estuaries	
			5.2.3 Sea and ocean	
12. Snow and Ice	Snow and Ice	6.2.2.3 Glaciers and perpetual snow	3.3.5 Glaciers and Perpetual Snow	-

#### 4.1.2 Pre-processing

The ECoLaSS South-West demonstration site in France and Spain is comprised of the footprints of six adjacent S-2 tiles (30TYN, 30TYP, 31TCH, 31TDH, 31TCJ and 31TDJ) for which Sentinel-2 data were processed as explained in the deliverables of the WP32 [AD07].

The Central demonstration site, covering part of France, Germany, Austria, Liechtenstein and Austria, is composed of 9 S-2 tiles, 32TNT, 32UNU, 32UMV, 32UMU, 32TMT, 32TPT, 32UNV, 32UPV. Instead of using Sen2Cor preprocessed images, whose cloud masks are known to have room for improvement, it has been decided to use MAJA pre-processed time series, whose cloud masks fare better (but are not perfect either), whose summarized details are given below.

The methods to obtain spatio-temporally consistent optical images with top of atmosphere reflectance values are:

- Atmospheric correction;
- Topographic normalization;
- Clouds, cloud shadow and snow masking, based on the previous dates available, in order to harness the maximum amount of information from the time series.

#### ATMOSPHERIC CORRECTION

The Copernicus Sentinel-2 data, processed at level L2A by CNES for THEIA Land data centre are available for download and corrected atmospheric effects, including adjacency effects. These atmospheric

corrections include the light absorption by air molecules and the light scattering by molecules and aerosols (Hagolle et al., 2015).

Several models may be used to perform atmospheric corrections. In the case of the MAJA software, the MACCS processor is the model used. It pre-computes look-up tables using an accurate radiative transfer code (with successive orders of scattering), that simulates the light propagation through the atmosphere. The MACCS/MAJA method combines different approaches to obtain robust estimates of aerosol optical thickness.

### **TOPOGRAPHIC NORMALISATION**

A topographic correction is necessary if the sites are characterized by mountainous terrain as it is for the South-West Demonstration site. The topography can significantly influence the radiometric properties of the signal received from the satellite (see Wulder & Franklin, 2012). This effect is caused by the different lighting angles resulting from the topography (see Gallaun et al., 2007). The aim of a topographical correction is to compensate for the differences in reflectance intensity between the areas with varying slope, exposure and inclination and to obtain the radiation values that the sensor would have in the case of a flat surface.

The Sentinel-2 data using the MAJA software and available for download are corrected from the topographic effects at the online THEIA data center.

### **CLOUD, CLOUD SHADOW AND SNOW MASKING**

The MAJA cloud detection method (Hagolle et al., 2010) is based on a number of threshold tests including the cirrus band (B10). Additionally, multi-temporal tests are carried out to detect clouds by measuring a steep increase of the blue surface reflectance. Finally, the correlation of the pixel neighbourhood with previous images is calculated to avoid over detections based on the assumption that two different clouds at the same location on successive dates will not have the same shape. If a large correlation is observed, the pixel is excluded from the cloud mask.

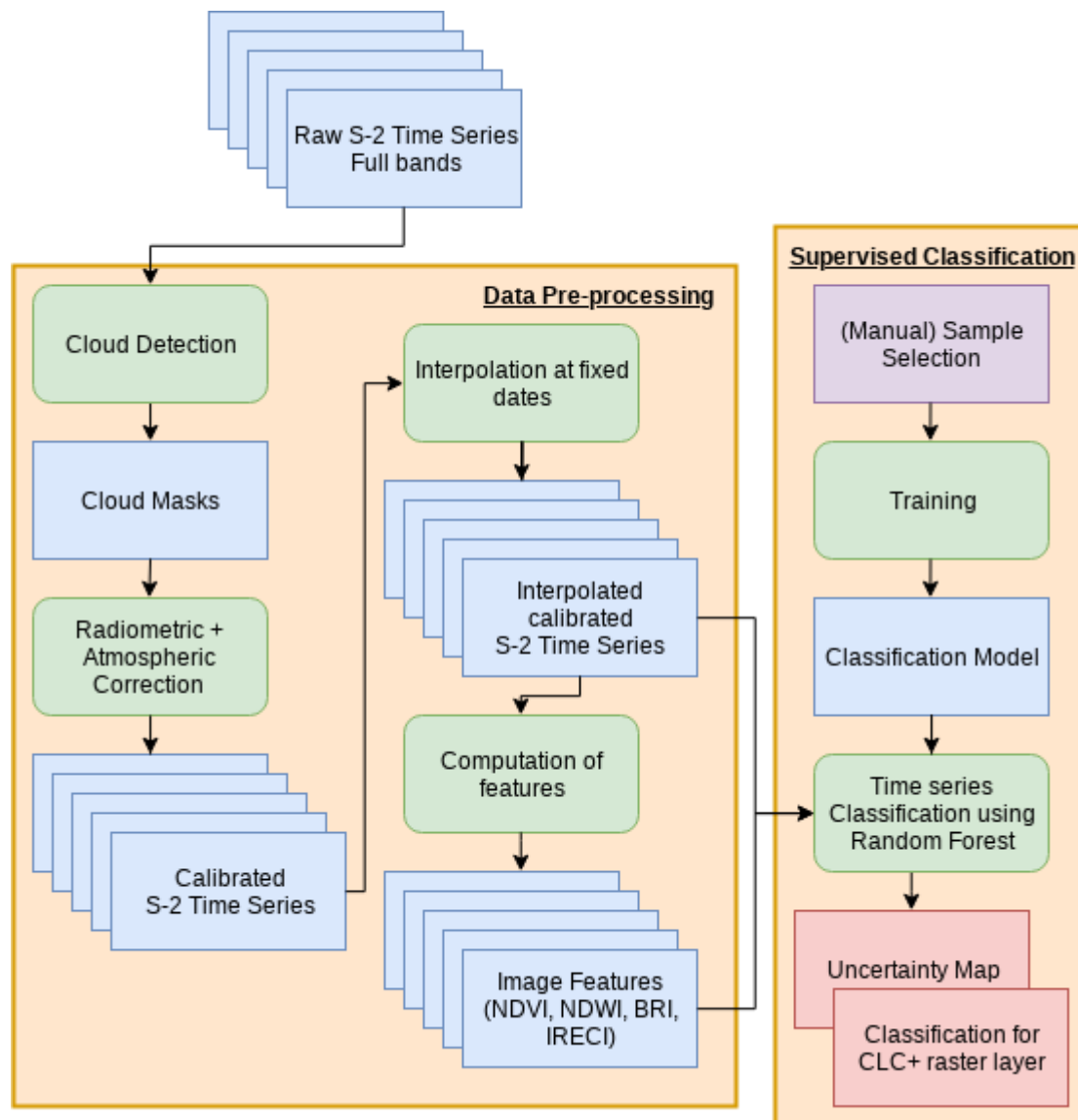
## **4.1.3 Experimental Setup**

The developed processing chain is able to process a large amount of input data within a reasonable amount of time to provide the classification results for the raster and vector products.

The workflow for the production of the New Land Cover Prototype (raster product) is listed hereafter:

1. The images processing, is subdivided in:
  - a. Applying the cloud mask on selected image data;
  - b. Computing the interpolated images at fixed dates based on each image of the time series;
  - c. Stacking those “artificial” images into multiband raster at the sample spatial resolution (10m);
  - d. Computing the spectral indices (NDVI, NDWI ad BRI);
  - e. Computing temporal features of those spectral indices (maximum, mean, minimum, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> percentiles and standard deviation);
2. The manual selection of samples, based on the biogeographical distribution of the classes over the demonstration sites;
3. The production of an independent pixel based classification:
  - a. Generating the sampling based on reference database (50% of the samples for calibration and 50% of the samples for validation);
  - b. Launching the RF classification;
  - c. Calculating automatically the associated confusion matrix;
4. Applying a majority filter to harmonize the results;

5. Application of potential binary thematic masks, such as IMP layers, in order to reduce the confusion between bare soils or croplands and urban areas.



**Figure 4-1: Experimental Setup for the production of the raster classification**

The results are obtained with one model over all tiles for each demonstration site, therefore eliminating the need to mosaic individual classifications (one per tile, as done in phase 1), leading to a smoother final raster result.

The workflow for the production of the vector product is listed hereafter:

1. The generation of the hard bones based on ancillary data:
  - a. Integration of OSM railways and roads networks;
  - b. Integration of EU-Hydro polygon datasets with the contribution of OSM water bodies;
  - c. Clean-up of this last addition for lakes and other non-linear water objects using another OSM class;
  - d. Final visual inspection and manual clean-up if need be.
2. The generation of the soft bones based on temporal phenological behaviour leading to a fast and visually satisfying segmentation of the landscape:

- a. Using the interpolated images and NDVI computed at the image processing step, just before the RF classification, a K-means classification is applied;
  - b. The resulting raster is then vectorized and only the geometry is kept.
3. The aggregation of the results with the softbone and the hardbone into a skeleton vector layer;
4. The computation of zonal statistics between the raster product and this skeleton vector layer in order to determine the distribution of each class per segment;
5. The final attribution of the segment to one of the 18 classes designed for the vector product following rules detailed in the CLC+ ITT.

#### 4.1.3.1 Image Processing

Finally, for this second phase, “artificially” images are produced, as described in the section 3.1.2, in particular the Weighted Average Compositing (WAC). This method is used here over the 2018 time series of S-2A and B: cloud masks are applied on the images, and an interpolation at fixed dates is computed in order to fill gaps due mainly to cloud presence, but also to saturated pixels or anomalies on the sensor – but also gaps left by the swath of the twin satellites. Indices and temporal features are then extracted from those images and fed to the model.

Several indices are also computed on those gap-filled images:

- NDVI (Normalized Difference Vegetation Index),
- NDWI (Normalized difference water index)
- BRI (brightness)

#### 4.1.3.2 Hard bones

The hardbone can be generated separately from the image processing. It consists on, first, collecting data that were previously listed in section 4.1.1.2, then, extracting only relevant information following some geometric rules. The methodology description has been detailed in the deliverable of WP33 [AD08] and is reproduced here for readability.

The focus here are the following databases:

- Open Street Map (OSM) for road and railway shapes;
- EU-Hydro and WISE- for rivers and canals.

Those linear networks are meant to be represented in those hard bones layer as polygons. However, a discrimination between elements wider than 20m and narrower than 20m is introduced in the ITT. This kind of information, such as an average width, cannot be retrieved from the ancillary databases selected. Therefore, this action would be derived from remote sensing data – and for finer accuracy, would be derived from VHR images, not S-1 nor S-2 time series.

This work has been deemed to be outside the scope of ECOLaSS – since it not only calls for a heavy use of VHR images over both demonstration sites but also for a heavy manual work to assess this average width along roads, railways and rivers.

In order to stay as close as possible to the CLC+ requirements, selection among what is expected to be the widest road and railway kinds has been done in OSM, integrating in the hard bones layer:

- OSM railways class called ‘rails’;
- OSM roads classes, namely ‘motorway’, ‘motorway\_link’, ‘trunk’, ‘trunk\_link’, ‘primary’ and ‘primary\_link’.

Data from EU-Hydro has been integrated, as well as main rivers from OSM. The WISE dataset, once downloaded, turned out to be corrupted for the year 2016. A 2012 dataset is available, but falls off the



range of the ancillary data in term of temporal fit. The retained classes integrated into the hard bones layer are:

- OSM class named 'water\_a' to retrieve water bodies in the shape of polygons;
- EU-Hydro classes named 'canals\_p' and 'rivers\_net\_p' for polygons only.

However, the class 'water\_a' also contains lakes – the OSM class 'waterways', which is only composed of purely linear objects, is used to remove all non-linear water objects. To complement this clean-up, the EU-Hydro 'InLandWater' is also involved in this intersection: elements falling outside of it are removed, leaving only polygons of linear water networks.

Therefore, methodological choices could be summarized as:

- OSM has been chosen over the South-West site (FR and ES) and the Central site (FR, DE, CH, AU) regarding the integration of roads and railways;
- OSM coupled with Eu-Hydro over both sites for water bodies and rivers.

Polygons are integrated into the hard bones layer in their current form, removed if no close linear network section exit in the vicinity or fused when possible if they fall below the expected MMU of 0.5ha.

Those resulting polygons are expected to be further subdivided in the second step using automated image segmentation.

The final result can be seen on Figure 4-2 for the South-West demonstration site and Figure 4-3 for the Central one.

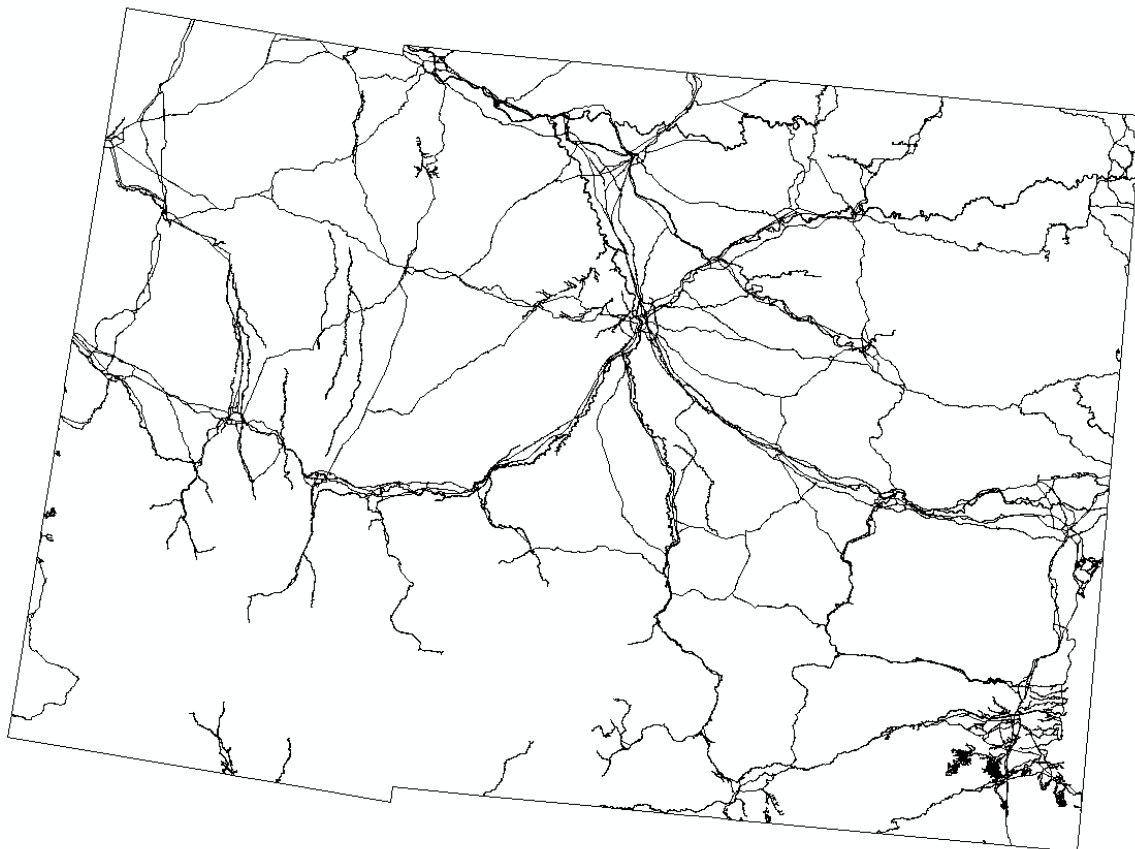
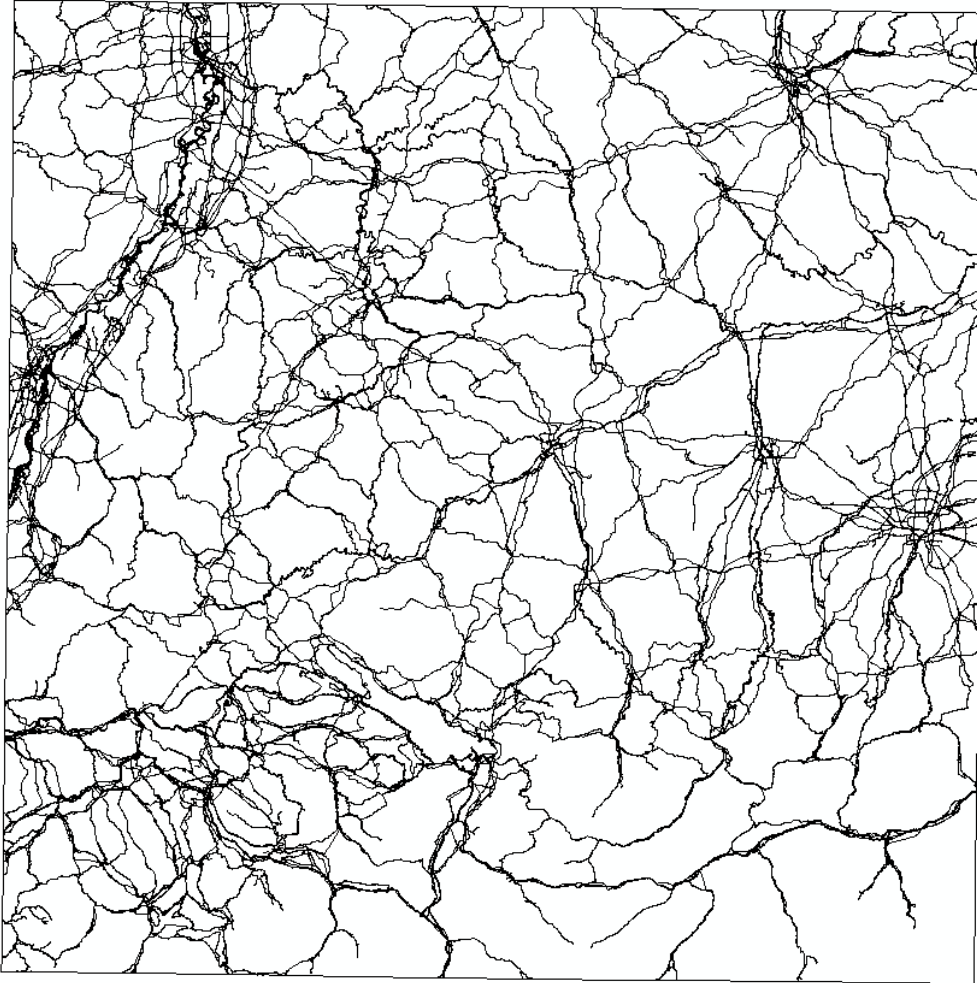


Figure 4-2: Hardbone over the South West demonstration site





**Figure 4-3: Hardbone over the Central demonstration site**

#### 4.1.3.3 Soft bones

The soft bones creation should enable the discrimination between objects in the landscape that behave temporally and spectrally differently during the year. This is a key and yet difficult step where balance between too few and too many segments has to be found.

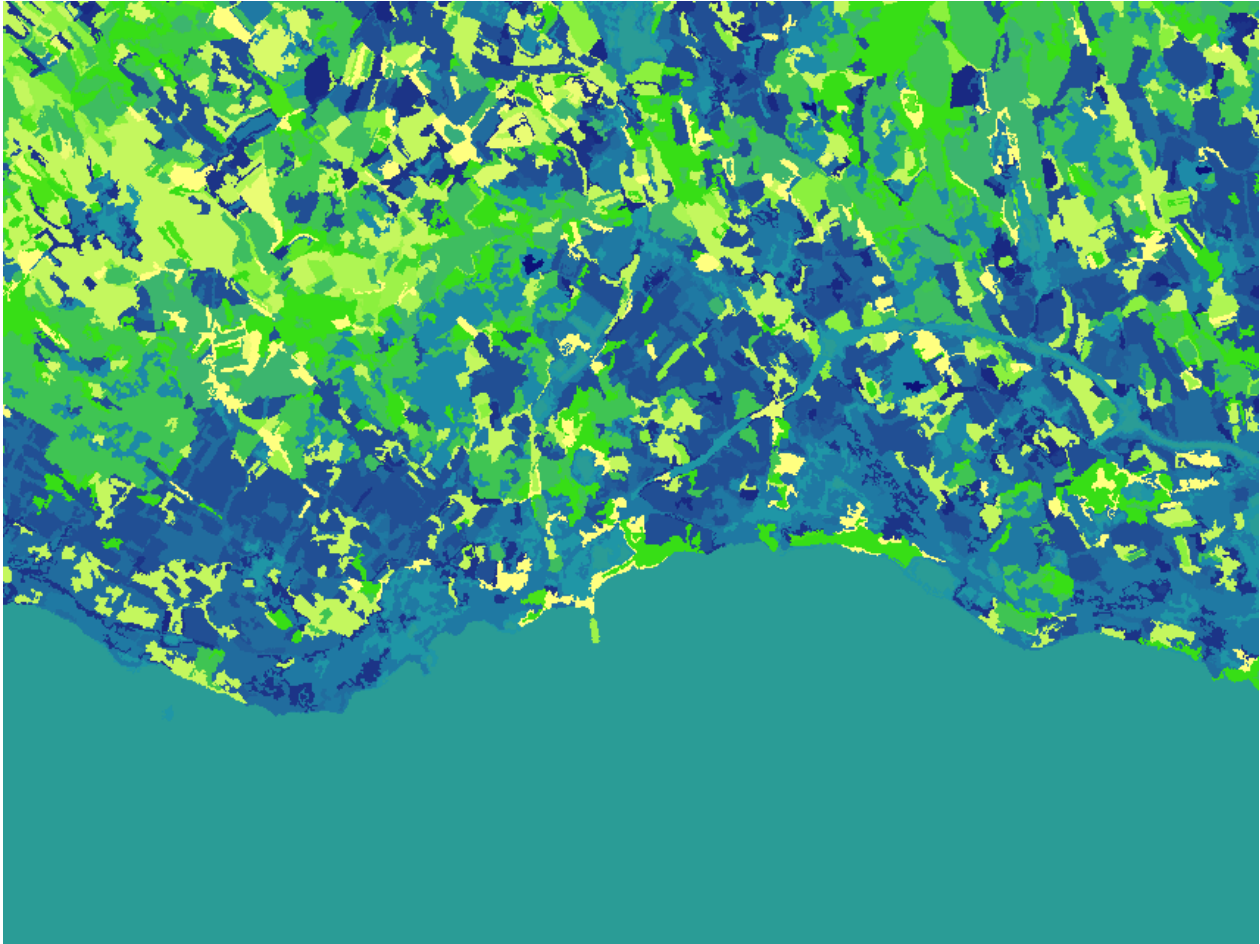
After benchmarking visual results and taking into account the reproducibility as well as the operability of such task, it has been decided in the deliverable of WP33 [AD08] to use the methodology leading to the creation of the phenological layer, which is reproduced here for readability.

The phenological layers are generated based on the NDVI time series, where each pixel of the demonstration site is classified into an “activity” class, depending on the behaviour of its NDVI values during the year. The methodology is detailed in the report of W41 [AD09].

This classification allows the regrouping of coherent ensembles in the landscape, just like a segmentation on the time series would do, with a particular focus on the vegetation, matching the type of high level typology that is set for the raster classification. A majority filter of 0.5ha (corresponding to 50 S-2 pixels) is applied and the resulting raster is transformed into a vector layer. The geometry thus created is used as a soft bones layer.

The methodology, which relies on the differentiation of the various yearly dynamics of the vegetation cover of the landscape, as required for the construction of the soft bones, could be easily replicated from one year to the other, solving one of the issues raised by the CLC+ ITT. It should also be noted that the execution time of this type of segmentation is optimised in the creation of the input datasets needed, since they are also produced before the launch of the RF classifier.

Snapshot of final results can be seen on Figure 4-4 for the Central site.



**Figure 4-4: Zoom on the softbone over the Central demonstration site**

#### 4.1.3.4 Reference datasets generation

Splitting the learning model over bioregions creates artificial delineation between those various bioregions – leading to an inhomogeneous “look and feel” for the raster classification, and the same goes for a splitting over S-2 tiles.

However, for the classifier applied on each whole demonstration site to learn the spectral variations of a same class over various bioregions, decision has been made to select samples in equal quantity over those bioregions. This choice allows the classifier to learn all the spectral variations originating from those various kinds of landscapes.

Figure 4-5 and Figure 4-7 show the original distribution of bioregions on Central and South West demonstration sites. They have been simplified in order to keep 4 or 5 different regions where a selection of samples has been made.

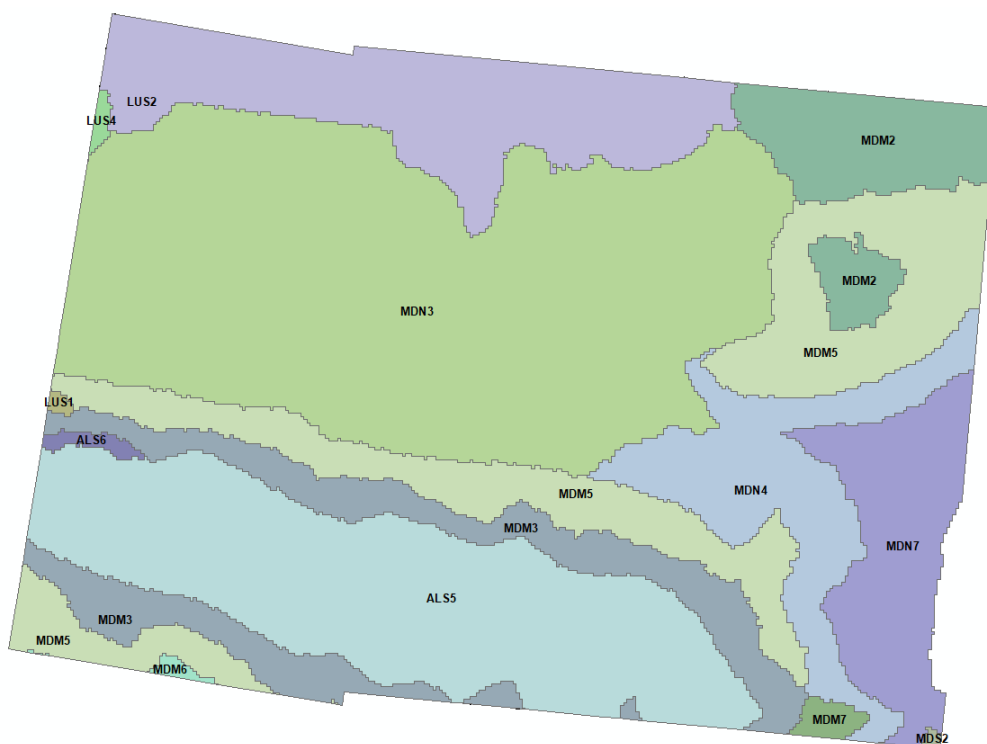


Figure 4-5: Original distribution of bioregion on the South West Demosite

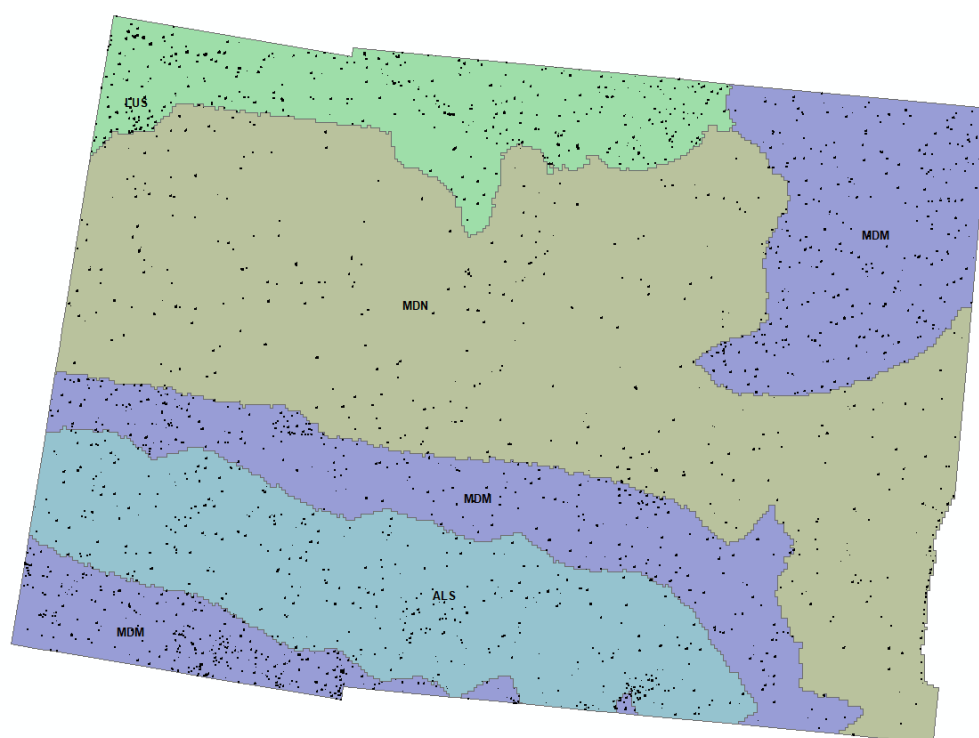


Figure 4-6: Simplified bioregion and spatial distribution of the samples on the South West Demosite

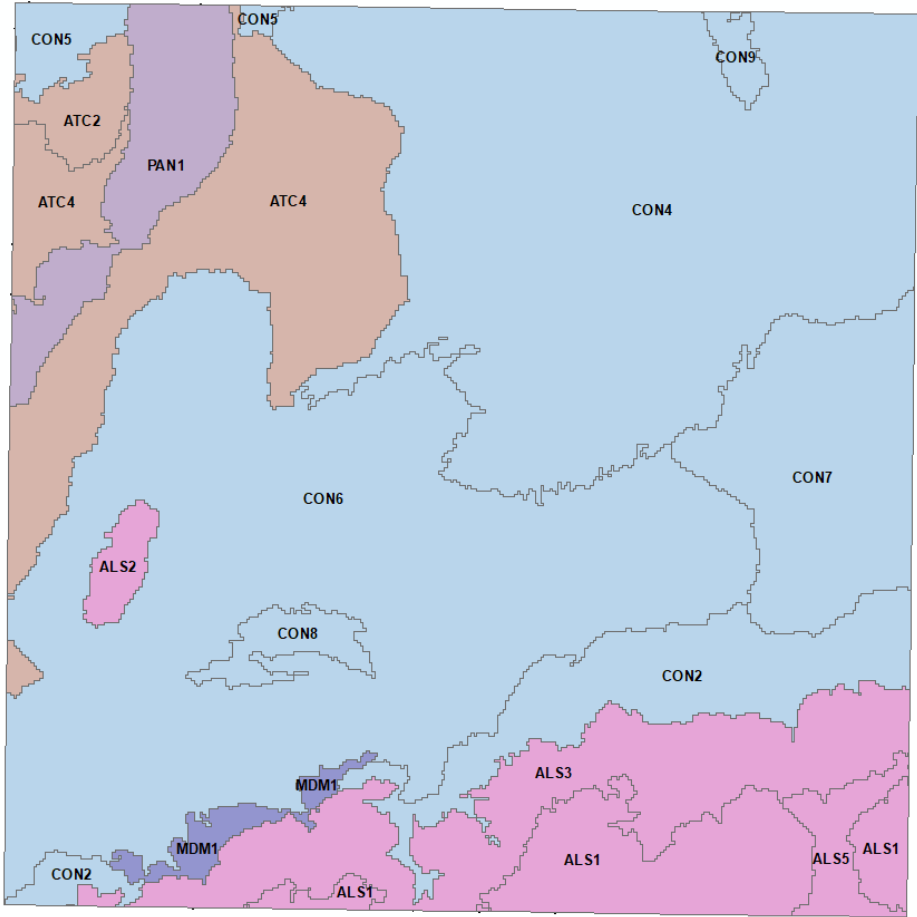


Figure 4-7: Original distribution of bioregion on the Central Demosite

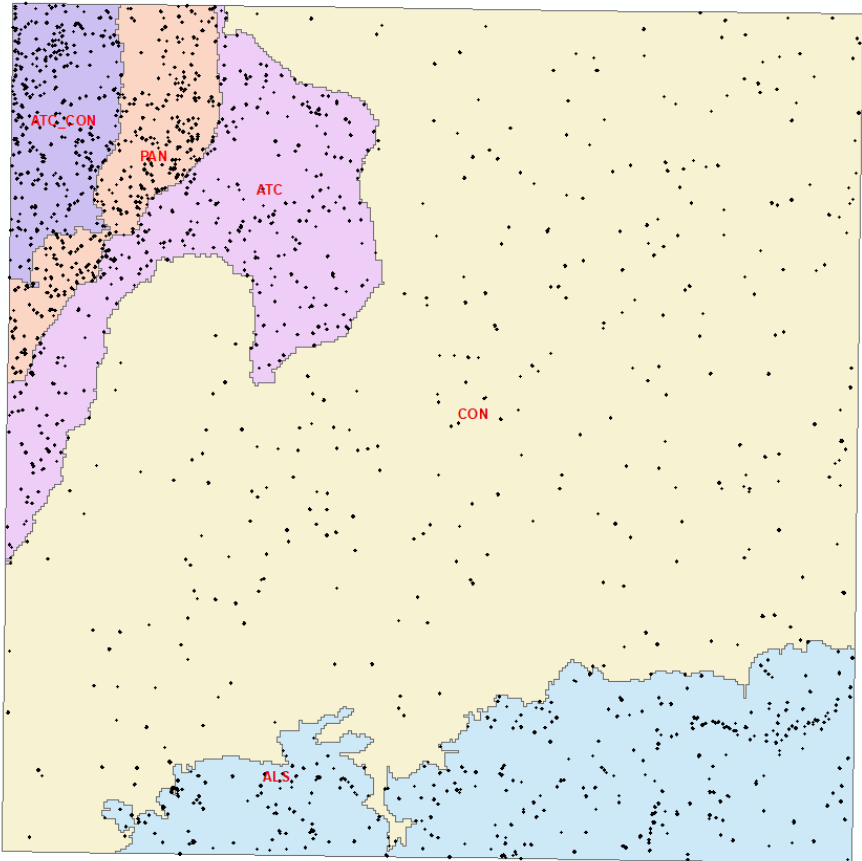


Figure 4-8: Simplified bioregion and spatial distribution of the samples on the Central Demosite

The creation of a dataset of samples based on the validated HRL 2015 products is required for the calibration and for the validation of the result of the classification. As input for the supervised machine learning algorithms, a clear set of training data is required. Those samples must be exempt from anomalies and must be a suitable statistical representation of the considered area. They must be representative of the whole study site by covering all the reflectance variations of the classes. They were selected thanks to a spatial distribution per bioregion.

To this end, reliable samples have been derived from the historical HRL 2015 IMD, Forest, Grassland and Water and CLC2018 data. An automated random selection has been processed on each bioregion.

It should be noted that:

- In the deliverable of WP33 [AD08], raster classifications done over the South-West test site did not take into account the class “bare soils” – too few samples were found in the CLC2018, however, they have been reintroduced over the demonstration site due to the presence of the Pyrénées, in to two classes: “sand” and “rocks”, in order to better separate their respective spectral signatures;
- No “sparsely vegetated” class has been used on raster classification – the problem being the availability of samples: samples from CLC2018 are provided at a 25ha scale, while no class related to sparsely vegetated can be found in the LUCAS databases, which could have provided circular samples of 3m radius, closer to the pixel level required by the raster classification. Too much confusion would have been created between the “bare soils”, “woody shrubs”, for example, and “sparsely vegetated” class. However, this class can easily be created in the vector product using the zonal statistics;
- No evergreen class has been considered over the South-West demonstration site – the samples taken from CLC2018 at a 25ha contains several classes, such as rocks, grasslands, woody shrubs or even broadleaved trees, in order to reach the MMU, leading to strong confusion between those classes;
- The lichens and mosses class could be based on the peat bogs class from CLC2018; however, this class is also considered as a wetland category and the land can be regularly flooded, at various dates of the year, leading to confusion with water bodies;
- The permanent crops class that was introduced over the test sites has been split between woody shrubs and arable lands.

Difficulties therefore remain to classify:

- Woody broadleaved evergreen trees – which are often merged with other classes to reach the 25ha MMU in the CLC2018 samples;
- Sparsely vegetated covers – since samples at 10m scale still needs to be properly created;
- Lichens and mosses, which can range from peat bogs to bare rocks covered in lichens including tundra environment.

The table below shows the classes distribution in the test and demosites in phase 2 for the CLC2018 categories.

**Table 4-4: CLC2018 classes distribution in the ECoLaSS test and demosites (phase 2)**

Classes		CLC 2018			
		SW Test site	SW Demo site	Central Test site	Central Demo site
1 – Sealed/Urban		1131	2159	1859	8365
2 – Needle leaved/Coniferous Forest		375	1801	1903	7002
3 – Broadleaved (Deciduous) Forest		2440	4433	1043	4760
4 – Broadleaved Evergreen		0	421	0	0
5 – Woody Shrubs		422	1227	388	1000
6 – Permanent Herbaceous/Grasslands		1013	1271	525	1235
7 – Periodically Herbaceous/Croplands		1890	4873	3975	16122
8 – Lichens & Mosses		0	0	28	113
9 – Sparsely Vegetated		0	526	0	388
10 – Non Vegetated	Sand	0	21	5	29
	Rocks	0	247	124	283
	Burnt Areas	0	3	0	0
11 - Water		146	288	104	441
12 – Snow & Ice		0	11	21	63

#### 4.1.3.5 Classification

The next step consists in a random forest algorithm to classify the different classes - according to the tests conducted in WP33 [AD08], taking the interpolated S-2A and B datasets at fixed dates (twice a month) as input. Spectral indices as well as temporal statistics are further added to the input data.

A majority filter is also applied after the classification - this presents two advantages:

- To harmonize the results, as it smooths the pixel-based classification and let appear a more natural landscape;
- To merge isolated pixels into greater ensembles.

Finally, once the classification has been launched and a result produced, the validation process can take place for the raster classification.

Two further steps are required to create the vector product.

#### 4.1.3.6 Fusion of hard bones and soft bones

Both hard bones and soft bones layers are vector layers. In order to merge both, a polygon union is executed, which will preserve all polygons from both layers.

Polygons will then be further subdivided, for example the road polygons from the hard bones will be further divided into the segmentation polygons.

It seems like a simple operation to launch; however, the amount of data is reaching the limits of most algorithm in terms of size. This fusion is indeed conducted at demonstration site scale, leading to millions of polygons being set in memory before being potentially subdivided.

#### 4.1.3.7 Fusion of the landscape objects with the raster product

The final fusion between the product generated by the fusion of hard and soft bones, called the landscape objects, and the raster classification has been described in the deliverable of WP33 [AD08], and is replicated here for clarity.

The attribution of land cover to polygons of the landscape objects from the pixel-based classification raster, is done using zonal statistics, which is a vector operation that list the number of pixels present in a given polygon for each class of the raster classification. For each polygon, percentage of LC coverage per class is then determined based on this pixel count, and attribution is made:

- To the dominant LC class;
- To the 3 most dominant LC classes;
- To a particular class given a percentage of LC present in the polygon.

For each class of the final vector product and their associated subclasses, rules can be written, as detailed in Table 4-5. The woody trees classes are defined with:

- The woody – needle leaved trees;
- The woody – broadleaved, deciduous trees;
- The woody – broadleaved, evergreen trees.

The vegetation classes from the raster classification are the following:

- The woody trees;
- The woody – shrubs;
- Permanent herbaceous;
- Periodically herbaceous;
- Lichens and mosses.



**Table 4-5: - List of rules to populate the vector layer created by the fusion of hard and soft bones, using the raster classification.**

Raster classes	Vector classes	Vector subclasses	Rules
1 - Sealed (buildings and flat sealed surfaces)	Built-up land	11. Very high sealing degree	Sealed > 80%
		12. High sealing degree	Sealed between 50 and 80%
2 - Woody – needle leaved trees	Woodland – needle leaved trees	21. Pure needle leaved	Needle leaved trees > 75% Woody trees >= 50%
		22. Dominantly needle leaved	Needle leaved trees between 50 and 75% Woody trees >= 50%
3 - Woody - broadleaved, deciduous trees	Woodland – broadleaved trees	31. Pure broadleaved	Broadleaved trees > 75% Woody trees >= 50%
4 - Woody – broadleaved, evergreen trees		311. Pure deciduous	Deciduous trees > 75% Woody trees >= 50%
		312. Pure evergreen	Evergreen trees > 75% Woody trees >= 50%
		32. Dominantly broadleaved	Broadleaved trees between 50 and 75% Woody trees >= 50%
5 - Woody – shrubs	Shrubland	-	Shrubs > 50%
6 - Permanent herbaceous (i.e. grasslands)	Permanent herbaceous land (i.e. grasslands)	51. Without trees	Permanent herbaceous > 50% Woody trees <= 10%
		52. With few trees	Permanent herbaceous > 50% Woody trees between 10 and 30%
		53. With many trees	Permanent herbaceous > 50% Woody trees between 30 and 50%
7 - Periodically herbaceous (i.e. arable land)	Periodically herbaceous land (i.e. arable land)	-	Periodically herbaceous >= 50%
8 - Lichens and mosses	Lichens and mosses land		Lichens and mosses >= 50%
9 - Sparsely vegetated	Partly vegetated land	81. Intermediate vegetation cover	Total of vegetation classes between 30 and 50% Non-vegetated >= 50%
		82. Low vegetation cover	Total of vegetation classes between 10 and 30% Non-vegetated >= 50%
10 - Non-vegetated (i.e. rock, screes, sand, lichen, permanent bare soil)	Non-vegetated land (i.e. rock, screes, sand, lichen, permanent bare soil)		Non-vegetated >= 90%
11 - Water	Water		Water >= 50%
12 - Snow and ice	Snow and ice		Snow and ice >= 50%



This will be translated in our cases as described in Table 4-6. It should be noted that some of the polygons do not possess a dominant class with more than 50% of the overall surface, due to the correctness of the segmentation. In order to minimize the impact of those polygons on the validation, a last vector subclass is added, called “others”. Further explanations are given in the following section on validation.

**Table 4-6: - Adapted typology for the vector product.**

Raster classes	Vector subclasses	Rules
<b>1 - Urban</b>	11. <i>Very high sealing degree</i>	Urban (1) > 80%
	12. <i>High sealing degree</i>	Urban (1) between 50 and 80%
<b>32 -Needle leaved trees</b>	21. <i>Pure needle leaved</i>	Needle leaved trees (32) > 75% Woody trees >= 50%
	22. <i>Dominantly needle leaved</i>	Needle leaved trees (32) between 50 and 75% Woody trees >= 50%
<b>31 – Broadleaved trees</b>	31. <i>Pure broadleaved</i>	Broadleaved trees (31) > 75% Woody trees >= 50%
	32. <i>Dominantly broadleaved</i>	Broadleaved leaved trees (32) between 50 and 75% Woody trees >= 50%
<b>30 - Woody – shrubs</b>	5. <i>Woody Shrubs</i>	Shrubs (30) > 50%
<b>23 - Permanent herbaceous (i.e. grasslands)</b>	51. <i>Without trees</i>	Permanent herbaceous (23) > 50% Woody trees <= 10%
	52. <i>With few trees</i>	Permanent herbaceous (23) > 50% Woody trees between 10 and 30%
	53. <i>With many trees</i>	Permanent herbaceous (23) > 50% Woody trees between 30 and 50%
<b>20 - Periodically herbaceous (i.e. arable land)</b>	7. <i>Periodically herbaceous</i>	Periodically herbaceous (20) >= 50%
<b>9 - Sparsely vegetated</b>	81. <i>Intermediate vegetation cover</i>	Total of vegetation classes between 30 and 50% Non-vegetated (33+34) >= 50%
	82. <i>Low vegetation cover</i>	Total of vegetation classes between 10 and 30% Non-vegetated (33+34) >= 50%
<b>33+34 - Non-vegetated (i.e. rock, screes, sand, lichen, permanent bare soil)</b>	10. <i>Non-vegetated</i>	Non-vegetated (33+34) >= 90%
<b>50 - Water</b>	50. <i>Water</i>	Water (50) >= 50%
<b>35 - Snow and ice</b>	35. <i>Snow and Ice</i>	Snow and ice (35) >= 50%
<b>N/A</b>	99 <i>Other</i>	Class with no single class dominance

Where:

- Woody Trees are the sum of the percentage for the classes:
  - o The woody – needle leaved trees (32);
  - o The woody – broadleaved, deciduous trees (31);
- Vegetation classes are then:
  - o The woody – needle leaved trees (32);
  - o The woody – broadleaved, deciduous trees (31);
  - o Permanent herbaceous – grasslands (23);
  - o Periodically herbaceous – croplands (20);
- Non vegetated class is the union of rocks (33) and sand (34) classes.

#### 4.1.3.8 Validation analysis procedure

A confusion matrix is automatically generated after the classification ran, based on the sampling data detailed at section 4.1.3.4, allowing the adjustment of the number of classes (by fusion, suppression or creation of new classes). Indeed, if the overall accuracy (OA) is lower than expected, or if the omission or commission rates are too high, or if the look and feel simply does not seem good enough, the classification is re-processed.

A final nomenclature of 10 classes for the raster product has been adapted and can be seen in Table 4-7.

**Table 4-7: – Final nomenclature for the New Land Cover Prototype**

Class	Label
1	Urban
20	Cropland
23	Grassland
30	Shrubs
31	Broadleaved Trees
32	Needle-leaved trees
33	Non-vegetated
35	Snow & Ice
50	Water

The results reached a good quality, avoiding most of the confusions listed previously, by putting aside the most difficult classes to sample.

The final nomenclature for the vector product has also been adapted and can be found in Table 4-8.

**Table 4-8: – Final nomenclature for the New Land Cover Prototype in vector format**

Class	Label
11	Very high sealing degree
12	High sealing degree
21	Pure needle leaved
22	Dominantly needle leaved
31	Pure Broadleaved
32	Dominantly broadleaved
5	Woody – shrubs
51	Grasslands – Without trees
52	Grasslands – With few trees
53	Grasslands – With many trees
7	Arable Land
81	Sparsely Vegetated – Intermediate Vegetation Cover
82	Sparsely Vegetated – Low Vegetation Cover
10	Non Vegetated
50	Water
35	Snow and Ice
99	Other (no dominant class)

## 4.2 Classification Results and Validation

To be able to determine the global accuracy of the New Land Cover layer, several validations were conducted.

### 4.2.1 NLC Raster products for South West and Central demo sites

The results for the raster product - the Central demonstration site- are presented in Table 4-9.

**Table 4-9: Independently established confusion matrix for the Central demonstration site – Plausibility analysis**

CENTRAL demosite PLAUSI		REFERENCE									Total	User Accuracy	
		1 urban area	20 cropland	23 grassland	30 shrubs	31 broadleaved	32 needleleaved	33 rock	35 snow	50 water			
PRODUCT	1 urban area	78.24	6.10			1.02	1.02	1.02		1.02	88.40	88.51%	
	20 cropland		177.73	6.54							184.27	96.45%	
	23 grassland	3.74	14.95	230.74				3.74			253.16	91.14%	
	30 shrubs	2.96	1.97	6.91	9.87	5.92					27.65	35.71%	
	31 broadleaved	1.03		1.03		171.40	18.59				192.05	89.25%	
	32 needleleaved			1.11		1.11	192.17			1.11	195.51	98.30%	
	33 rock	2.79		0.93	0.93		1.86	31.66	0.93		39.12	80.95%	
	35 snow							0.26	2.56		2.82	90.91%	
	50 water							0.57		16.47	17.04	96.67%	
	Total	88.76	200.75	247.27	10.81	179.45	213.64	37.24	3.49	18.60	1000.00		
Producer Accuracy		88.14%	88.53%	93.32%	91.38%	95.51%	89.95%	85.03%	73.33%	88.56%		91.08%	Overall Accuracy
													0.89 Kappa

These results are quite promising, the OA is above 91%. However, some classes exhibit a slightly less sharpened accuracy. Omissions errors can be spotted on the snow class, that could be explained by confusion with the clouds. Commissions errors should be noticed on the shrubs with other kind of vegetations, in particular the grasslands class, which can be due to samples mixing both classes; rocks have a spectral signature close to urban areas and confusions can also arise between those two classes.

The results for the raster product - South-West site - are presented in Table 4-10.

**Table 4-10: Independently established confusion matrix for the South-West site – Plausibility analysis**

SOUTH WEST demosite PLAUSI		REFERENCE										Total	User Accuracy
		1 urban area	20 cropland	23 grassland	30 shrubs	31 broadleaved	32 needleleaved	33 rock	34 sand	35 snow	50 water		
PRODUCT	1 urban area	56.38	3.03		1.03							60.44	93.28%
	20 cropland	5.02	225.65	10.08	6.05	3.04		1.00				250.85	89.95%
	23 grassland	1.00	7.19	194.17	2.04	2.05						206.46	94.05%
	30 shrubs			8.07	95.97	4.04	6.06					114.13	84.09%
	31 broadleaved			2.05	3.06	202.08	5.03					212.23	95.22%
	32 needleleaved				2.02	6.00	92.14					100.16	91.99%
	33 rock		0.91					35.70		0.58		37.19	95.99%
	34 sand								0.15			0.15	100.00%
	35 snow									5.75		5.75	100.00%
	50 water			0.75							12.01	12.76	94.12%
Total		62.41	236.79	215.12	110.16	217.22	103.22	36.70	0.15	6.33	12.01	1000.13	
Producer Accuracy		90.34%	95.30%	90.26%	87.12%	93.03%	89.26%	97.27%	100.00%	90.91%	100.00%		91.99%
													0.90 Kappa

These results are also promising with an OA of almost 92% on the South-West site. Commissions errors are also present on the Shrubs class, for the same reason pointed earlier.

The prototype of the raster classification could certainly be improved with different approach such as various modification of the classes, a refinement of the samples used in the training as well as in the validation, the use of textural indices from VHR data or even the use of S-1 data.

The relative look and feel is shown below in the Figure 4- for the Central site and the Figure 4- for the South-West site. The different land cover classes are quite well identified: a real spatial "landscape coherence" can be seen, based on the delineation that could be seen between bioregions.



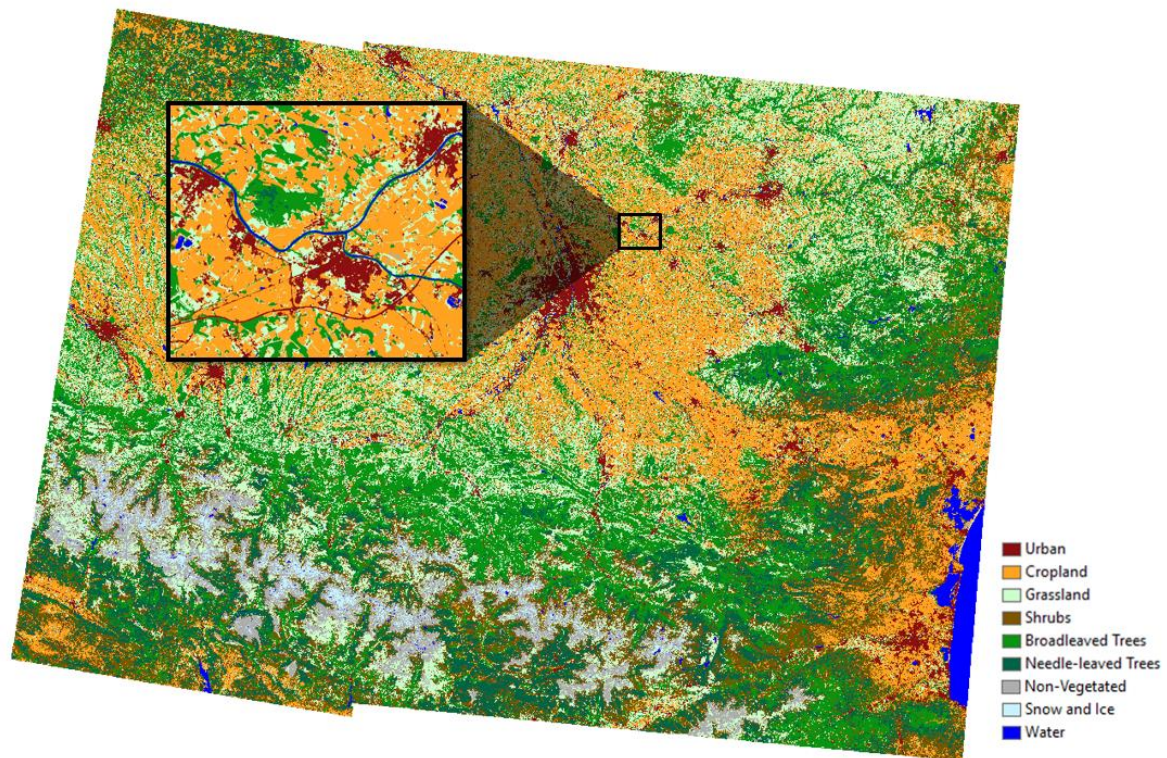


Figure 4-9: - Prototype NLC\_2018\_010m\_SW\_03035\_prototype\_v01 over the South-West site.

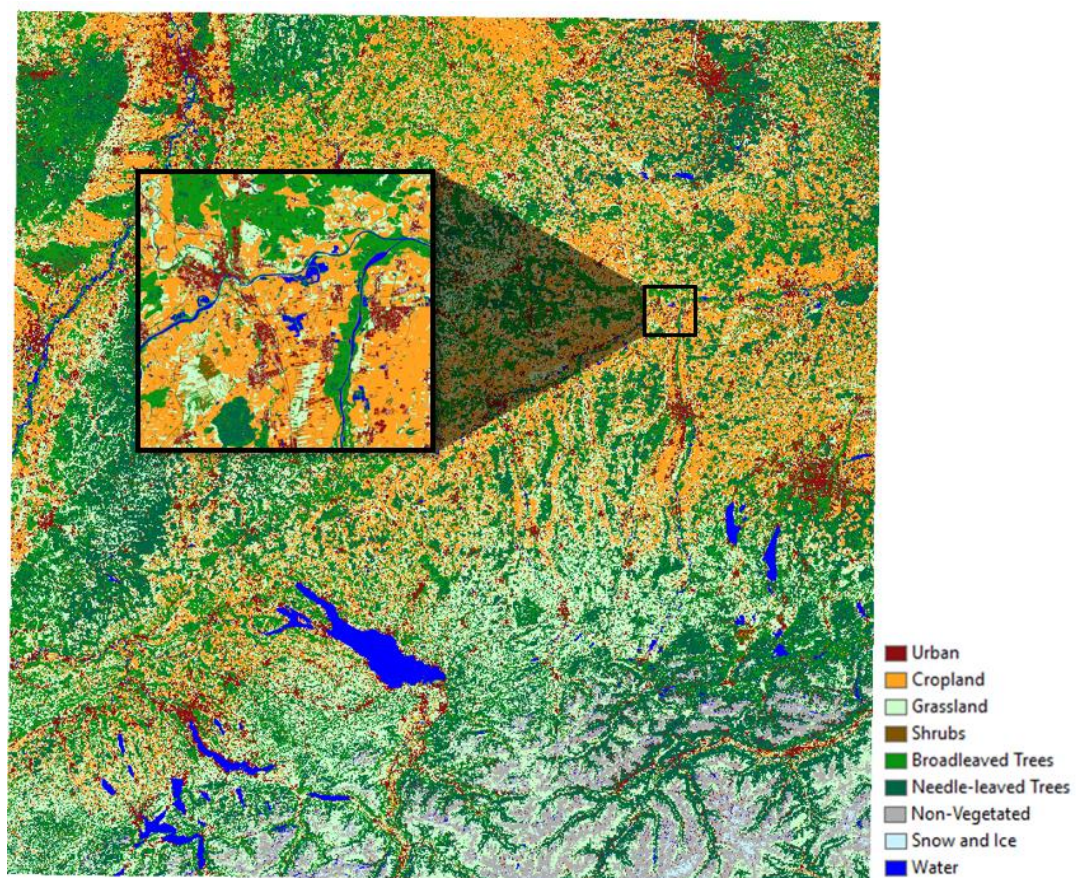


Figure 4-10: – Prototype NLC\_2018\_010m\_CE\_03035\_prototype\_v01 over the Central site.

#### 4.2.2 NLC Vector products for South West and Central demo sites

The results for the vector product - the Central demonstration site- are presented in Table 4-11.

These overall results are quite promising, the OA is above 88%. However, some classes exhibit a slightly less sharpened accuracy like low vegetation cover. Omissions errors can be spotted on low vegetation cover, that could be explained by vector product nomenclature and the choice to create an “other” class with non-dominant class. Commissions errors are noticed on the permanent herbaceous, intermediate vegetation cover and low vegetation cover classes also – this could be linked to the quality of the soft bones and the samples, which is even more obvious on object-level validation. An example of soft bones issue is illustrated in Figure 4-11.

The results for the vector product - the South West demonstration site- are presented in Table 4-12.

These results are also promising with an OA of almost 90% on the South-West site. Same conclusions as for Central site can be expressed as difficulties to validate classes at object level, in particular for vegetation or sparsely vegetated covers, can lower validation score.

The relative look and feel is shown below in the Figure 4-10 for the Central site and the Figure 4-9 for the South-West site. The different land cover classes are quite well identified: a real "landscape coherence" can be seen, as mentioned before – matching the delineation of the bioregions.

However, limitations of the segmentation based on S-2 only time series lead to problem for discrimination of mixed class where no dominant class are present.



**Table 4-11: Independently established confusion matrix for the Central demonstration site vector product**

CENTRAL VECTOR with class 99		REFERENCE																99 Other	Total	User Accuracy	Commission
		5 Woody Shrubs	7 Periodically Herbaceous	10 Non-Vegetated	11 Very High Sealing degree	12 High Sealing degree	21 Pure Needle Leaved	22 Dominantly Needle Leaved	31 Pure Broadleaved	32 Dominantly Broadleaved	35 Snow & Ice	50 Water	51 Permanent Herbaceous Without Trees	52 Permanent Herbaceous With Few Trees	53 Permanent Herbaceous With Many Trees	81 Intermediate Vegetation Cover	82 Low vegetation Cover				
PRODUCT	5 Woody Shrubs	4.88			0.99							0.99							6.86	71.21%	28.79%
	7 Periodically Herbaceous		182.81		1.02							3.11							186.94	97.79%	2.21%
	10 Non-Vegetated			8.95															8.95	100.00%	0.00%
	11 Very High Sealing degree				38.50														38.50	100.00%	0.00%
	12 High Sealing degree		5.23		6.01	27.45			1.02										39.71	69.13%	30.87%
	21 Pure Needle Leaved						125.70		2.14				0.93						128.77	97.61%	2.39%
	22 Dominantly Needle Leaved						16.19	44.84	2.07			0.57	0.93						64.60	69.41%	30.59%
	31 Pure Broadleaved						2.14		103.14	3.18									108.46	95.10%	4.90%
	32 Dominantly Broadleaved						5.48	10.48	3.10	36.47			3.74						59.26	0.00%	100.00%
	35 Snow & Ice																			0.00%	100.00%
	50 Water						1.11					16.80							17.91	93.80%	6.20%
	51 Permanent Herbaceous Without Trees	0.93	9.03	0.93									163.06						173.96	93.74%	6.26%
	52 Permanent Herbaceous With Few Trees	0.99			1.02		2.22		1.03				10.63	50.64	1.97				68.49	73.93%	26.07%
	53 Permanent Herbaceous With Many Trees						1.11		1.03						0.93	6.16			9.23	66.67%	33.33%
	81 Intermediate Vegetation Cover			0.93													0.93		1.86	50.00%	50.00%
	82 Low vegetation Cover			0.93									0.93						1.87	0.00%	100.00%
99 Other	0.99	8.49	19.90	10.10		11.97		6.15		3.49	1.68	21.87						84.63	0.00%	100.00%	
Total	7.79	205.56	31.65	57.62	27.45	165.92	55.32	119.68	39.64	3.49	19.05	206.20	51.57	8.12	0.93	0.00	0.00	1000.00			
Producer Accuracy	62.69%	88.93%	28.29%	66.81%	100.00%	75.76%	81.05%	86.18%	91.99%	0.00%	88.20%	79.08%	98.19%	75.78%	100.00%	0.00%	0.00%		81.03%	Overall Accuracy	
Omissions	37.31%	11.07%	71.71%	33.19%	0.00%	24.24%	18.95%	13.82%	8.01%	100.00%	11.80%	20.92%	1.81%	24.22%	0.00%	100.00%	100.00%		0.78	Kappa	

**Table 4-12: Independently established confusion matrix for the South West demonstration site vector product**

SW VECTOR without class 99		REFERENCE																		
		5 Woody Shrubs	7 Periodically Herbaceous	10 Non-Vegetated	11 Very High Sealing degree	12 High Sealing degree	21 Pure Needle Leaved	22 Dominantly Needle Leaved	31 Pure Broadleaved	32 Dominantly Broadleaved	35 Snow & Ice	50 Water	51 Permanent Herbaceous Without Trees	52 Permanent Herbaceous With Few Trees	53 Permanent Herbaceous With Many Trees	81 Intermediate Vegetation Cover	82 Low vegetation Cover	Total	User Accuracy	Commission
PRODUCT	5 Woody Shrubs	68.75		0.91			6.03					10.22						85.91	80.02%	19.98%
	7 Periodically Herbaceous	2.03	210.51	1.00	2.03							2.01						217.58	96.75%	3.25%
	10 Non-Vegetated			7.32														7.32	100.00%	0.00%
	11 Very High Sealing degree				24.06													24.06	100.00%	0.00%
	12 High Sealing degree		4.03	0.05	2.01	8.05						5.04						19.19	41.97%	58.03%
	21 Pure Needle Leaved						62.15											62.15	100.00%	0.00%
	22 Dominantly Needle Leaved						5.08	27.16	1.00									33.24	81.71%	18.29%
	31 Pure Broadleaved						1.03	2.03	149.75				2.06					154.86	96.70%	3.30%
	32 Dominantly Broadleaved	3.05					3.04	5.03	1.03	28.59			1.03					41.76	7.29%	92.71%
	35 Snow & Ice			0.91							1.15		0.58					2.64	43.57%	56.43%
	50 Water			0.15								10.51						10.66	98.59%	1.41%
	51 Permanent Herbaceous Without Trees	1.03	9.22										123.72	3.06				137.03	90.28%	9.72%
	52 Permanent Herbaceous With Few Trees		1.02						1.03					36.72				38.77	94.71%	5.29%
	53 Permanent Herbaceous With Many Trees													1.03	4.11			5.13	80.01%	19.99%
	81 Intermediate Vegetation Cover												2.06			0.00		2.06	0.00%	100.00%
	82 Low vegetation Cover	1.01		7.32							0.58		0.58				0.00	9.48	0.00%	100.00%
Total		75.86	224.79	17.67	28.09	8.05	71.30	40.25	152.80	28.59	1.73	10.51	147.28	40.81	4.11	0.00	0.00	851.84		
Producer Accuracy		90.62%	93.65%	41.42%	85.64%	100.00%	87.18%	67.49%	98.00%	100.00%	66.67%	100.00%	84.00%	89.98%	100.00%	0.00%	0.00%		89.52%	Overall Accuracy
Omissions		9.38%	6.35%	58.58%	14.36%	0.00%	12.82%	32.51%	2.00%	0.00%	33.33%	0.00%	16.00%	10.02%	0.00%	100.00%	100.00%		0.88	Kappa

Several issues could be identified with the vector products, that will still need to be addressed in the CLC+ Backbone production. On the one hand, the segmentation only based on S-2 has limitations that could be easily corrected with an incremental process, where a second segmentation could be undergone only on identified polygons that contain classes that are non-dominant (no class above the 50% presence threshold). Manual correction and VHR based segmentation (which are beyond the scope of this project) can also be investigated to help better discriminate objects containing multiple minor classes.

On the other hand, the generalization of this type of segmentation approach could be an issue. The number of phenological activity classes was fixed in the report of WP41 [AD09]. This ought to be adapted for different types of bioregions and, based on the current state of knowledge, this is still an open research topic that needs to be resolved before a full automated process for soft bones creation.

As mentioned in the deliverable of WP33 [AD08], since the phenological segmentation is based on a NDVI time series, some activity classes, such as the one for the water bodies, the urban areas or the bare soils, could be better discriminated by applying a mask (based on a NDVI threshold depending on the bioregions) and the connected components algorithm, which can further discriminate pixels based on other spectral (NDWI for water bodies) or textural (PanTex for urban) indices, splitting those polygons into smaller ones.

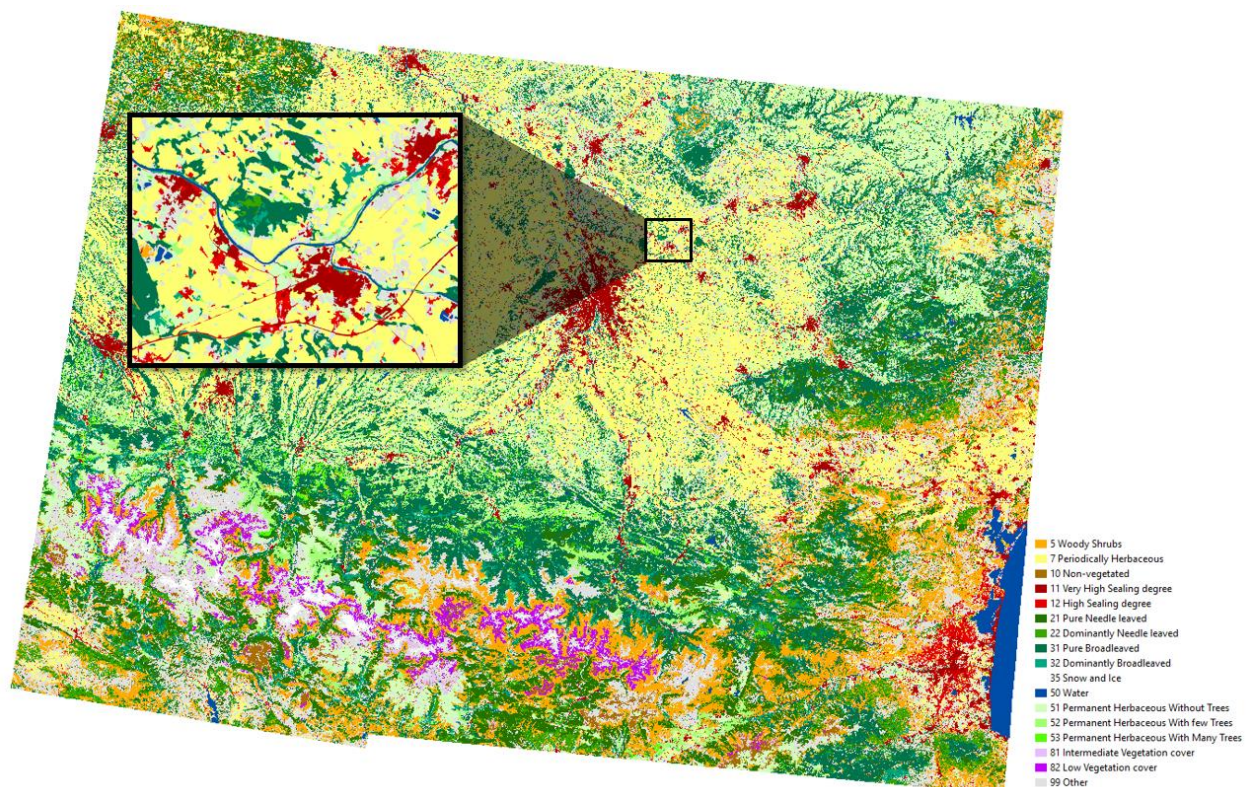


Figure 4-9: – Prototype NLC\_2018\_010m\_SW\_03035 vector version over the South-West site.



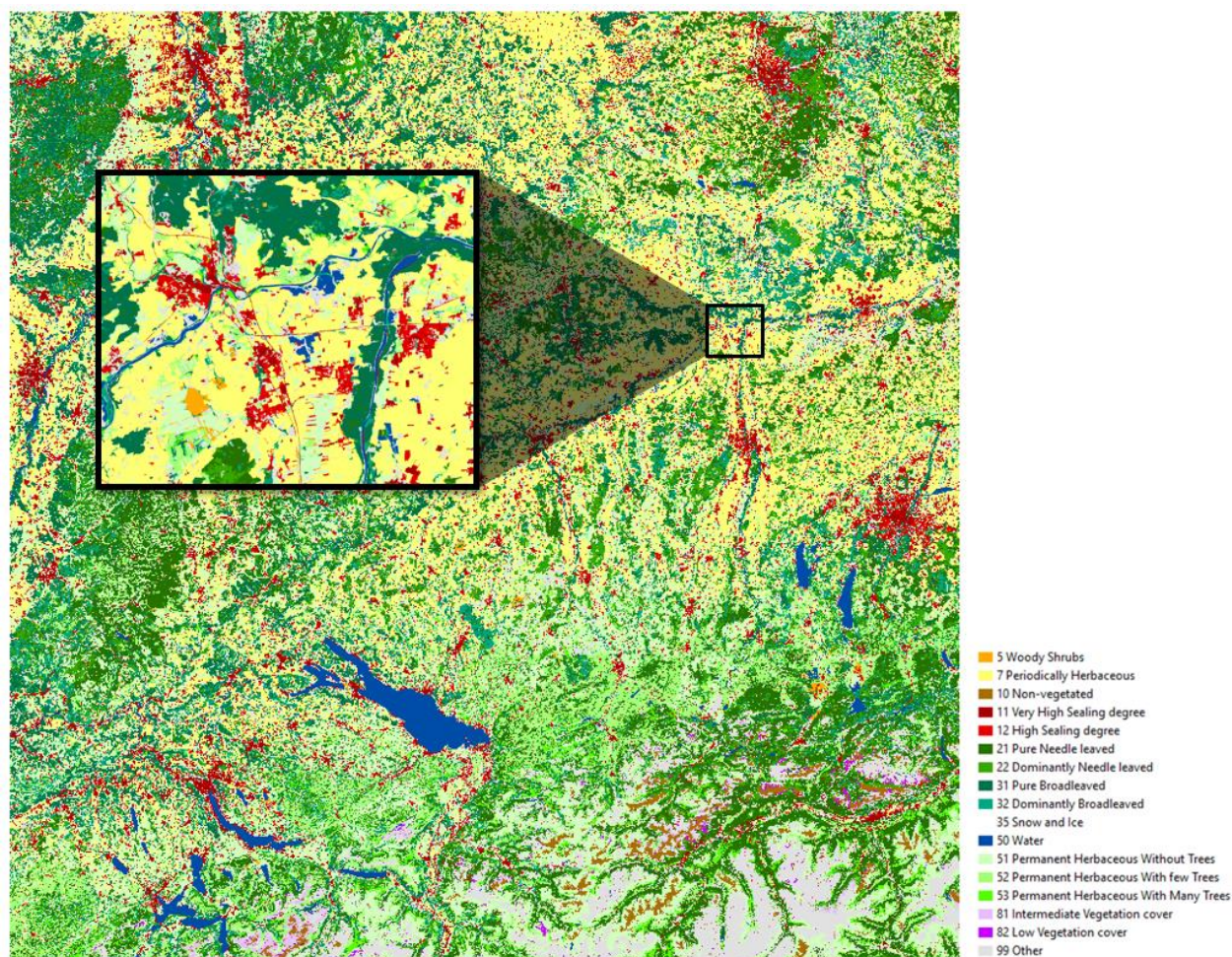


Figure 4-10: – Prototype NLC\_2018\_010m\_CE\_03035 vector version over the Central.

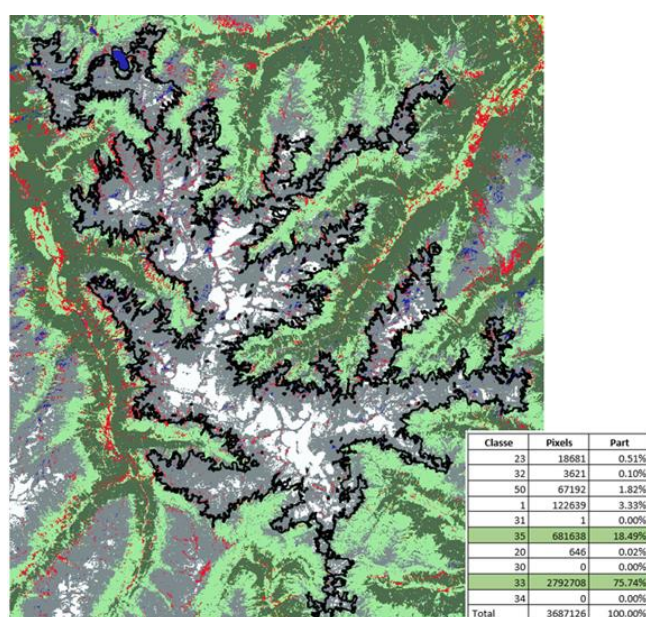


Figure 4-11: – Illustration of Softbone issue on specific object that are not in ITT CLC+ nomenclature.

In that case, a new segmentation could be envisaged to separate the two main classes: snow and non-vegetated classes

## 4.3 Prototype Specifications

This section provides a description of the dataset properties and metadata for the implemented prototypes, also referring to “P45.2b - Data Sets of New LC/LU Products”.

Within ECoLaSS, a standardised and harmonised product file naming convention for all prototypes has been developed which is oriented along the already existing naming convention of the CLMS High Resolution Layers. The naming convention consists of the following 7 descriptors:

**THEME      YEAR      RESOLUTION      EXTENT      EPSG      TYPE      VERSION**

as follows:

### THEME

3 letter abbreviation for main products (DLT (dominant leaf type), TCC (tree cover change), GRA (grassland), IMD (imperviousness degree), IMC (imperviousness change classified), CRT (crop type), CRM (crop mask) and NLC (new land cover products)).

### REFERENCE YEAR

2017 in four digits; change products in four digits (e.g. 1517)

### RESOLUTION

Four-digit (020m and 010m)

### EXTENT

2-digit code for demonstration-sites (CE (central), NO (north), WE (west), SW (southwest), SE (southeast), SA (South Africa), ML (Mali))

### EPSG

5-digit EPSG code (geodetic parameter dataset code by the European Petroleum Survey Group) “03035” for the European LAEA projection

### TYPE

prototype

### VERSION

3-digit code “v01”

### EXAMPLE:

“DLT\_2017\_010m\_NO\_03035\_prototype\_v01.tif” stands for: Dominant Leaf Type, 2017 reference year, 10m, Demonstration-site North, European projection (EPSG: 3035), prototype, version 01

“TCC\_1517\_020m\_NO\_03035\_prototype\_v01.tif” stands for: Tree Cover Change, 2015-2017 change period, 20m, Demonstration-site North, European projection (EPSG: 3035), prototype, version 01

The raster products of the ECoLaSS prototypes are delivered as GeoTIFF (\*.tif) with world file (\*.tfw), pyramids (\*.ovr), attribute table (\*.dbf) and statistics (\*.aux.xml), enabling an instant illustration and analysis of the products within Geographic Information System (GIS) software. Each product is accompanied with a product-specific colour table (\*.clr) and INSPIRE-compliant metadata in XML format.

Metadata are provided together with the products as INSPIRE-compliant XML files according to the EEA Metadata Standard for Geographic Information (EEA-MSGI). EEA-MSGI has been developed by EEA to



meet needs and demands for inter-operability of metadata. EEA's standard for metadata is a profile of the ISO 19115 standard for geographic metadata and contains more elements than the minimum required to comply with the INSPIRE metadata rules. Detailed conceptual specifications of the EEA-MSGI and other relevant information on metadata can be found at <http://www.eionet.europa.eu/gis>.

The following two prototypes files as part of **D15.4 – P45.2b – Data Sets of New LC/LU Products (Issue 2)** in the demosites Central and South-West were submitted, including the raster classification and the vector product:

- **NLC\_2018\_010m\_SW\_03035\_prototype\_v02.tif**
- **NLCV\_2018\_010m\_SW\_03035\_prototype\_v01.tif**
- **NLC\_2018\_010m\_CE\_03035\_prototype\_v01.tif**
- **NLCV\_2018\_010m\_CE\_03035\_prototype\_v01.tif**

The prototype specifications for the NLC layer are listed in Table 4-13.



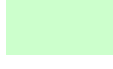




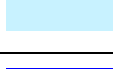

**Table 4-13: – Detailed specifications for primary 10m NLC status layer**

<b>Products</b>
New Land Cover 2018 – NLC_2018_10m
<b>Extent</b>
Demo site South-West
<b>Geometric resolution</b>
Pixel resolution 10m, grid to fully conform to the EEA Reference Grid.
<b>Coordinate Reference System</b>
European ETRS89 LAEA projection
<b>Geometric accuracy (positioning scale)</b>
Less than one pixel (According to ortho-rectified satellite image base delivered by Theia.)
<b>Thematic accuracy</b>
85% of overall accuracy
<b>Data type</b>
8bit unsigned Raster, compressed with LZW
<b>Minimum mapping unit (MMU)</b>
One pixel (10m*10m)
<b>Necessary attributes</b>
Raster value, count, class name

Raster coding (Thematic pixel values)	
New Land Cover 2018 – NLC_2018_10m	
1: urban 20: cropland 23: grassland 30: shrubs 31: broadleaved trees 32: needle-leaved trees 33: non-vegetated 35: snow & ice 50: water	
<b>Metadata</b> XML metadata files are to be produced according to INSPIRE metadata standards	
<b>Delivery format</b> GeoTIFF	

The associated color palette can be found in the Table 4-14.

**Table 4-14: – Palette used for primary 10m NLC status layer**

New Land Cover 2018 – NLC_2018_10m					
Class Code	Class Name	Red	Green	Blue	
1	Urban	140	15	18	
20	Cropland	255	163	33	
23	Grassland	204	255	205	
30	Shrubs	128	85	0	
31	Broadleaved trees	10	150	18	
32	Needle-leaved trees	0	100	70	
33	Non-Vegetated	174	174	174	
35	Snow & ice	200	242	255	
50	Water	0	0	255	

**Table 4-15: – Detailed specifications for primary 10m NLCV status layer**


















<b>Products</b>
New Land Cover 2018 – NLCV_2018_10m
<b>Extent</b>
Demo site South-West and Central
<b>Geometric resolution</b>
Pixel resolution 10m, grid to fully conform to the EEA Reference Grid.
<b>Coordinate Reference System</b>
European ETRS89 LAEA projection
<b>Geometric accuracy (positioning scale)</b>
Less than one pixel (According to ortho-rectified satellite image base delivered by Theia.)
<b>Thematic accuracy</b>
95% of overall accuracy
<b>Data type</b>
Vector data (gdb)
<b>Minimum mapping unit (MMU)</b>
One pixel (10m*10m)
<b>Necessary attributes</b>
FID, count of each raster class, Vclass (vector class), Shape_Length, Shape_Area

<b>Vector coding (Thematic pixel values)</b>
New Land Cover 2018 – NLCV_2018_10m 11: Very high sealing degree 12: High sealing degree 21: Pure needle leaved 22: Dominantly needle leaved 31: Pure Broadleaved 32: Dominantly broadleaved 5: Woody – shrubs 51: Grasslands – Without trees 52: Grasslands – With few trees 53: Grasslands – With many trees 7: Arable Land 81: Sparsely Vegetated – Intermediate Vegetation Cover 82: Sparsely Vegetated – Low Vegetation Cover 10: Non-Vegetated 50: Water 35: Snow and Ice 99: Other (no dominant class)
<b>Metadata</b>
XML metadata files are to be produced according to INSPIRE metadata standards
<b>Delivery format</b>
GDB

The associated color palette can be found in the Table 4-16.




**Table 4-16: Palette used for primary 10m NLCV status layer**

New Land Cover 2018 – NLCV_2018_10m					
Class Code	Class Name	Red	Green	Blue	
11	Very high sealing degree	168	0	0	
12	High sealing degree	230	0	0	
21	Pure needle leaved	38	115	0	
22	Dominantly needle leaved	56	168	0	
31	Pure Broadleaved	0	115	76	
32	Dominantly broadleaved	0	168	132	
5	Woody – shrubs	255	170	0	
51	Grasslands – Without trees	211	255	190	
52	Grasslands – With few trees	163	255	115	
53	Grasslands – With many trees	85	255	0	
7	Arable Land	255	255	115	
81	Sparsely Vegetated – Intermediate Vegetation Cover	232	190	255	
82	Sparsely Vegetated – Low Vegetation Cover	197	0	255	
10	Non-Vegetated	168	112	0	
50	Water	0	77	168	
35	Snow and Ice	255	255	255	
99	Other	225	225	225	



In addition to the prototype layers, probability layers are provided as a by-product to the raster classifications of NLC\_2018 products (CE & SW). The additional raster in the \*.tiff-file serves as one of the accuracy parameters that are described in detail in the WP33 final report [AD08] and range from 0 to 100%. The higher the percentage the higher the probability is that the respective pixel belongs to the depicted class. In this manner, the probability band depicts the error map at pixel level. Further, areas that are excluded by the referring mask get the value “101”. An overview of the probabilities’ colour palette is given in Table 4-17.

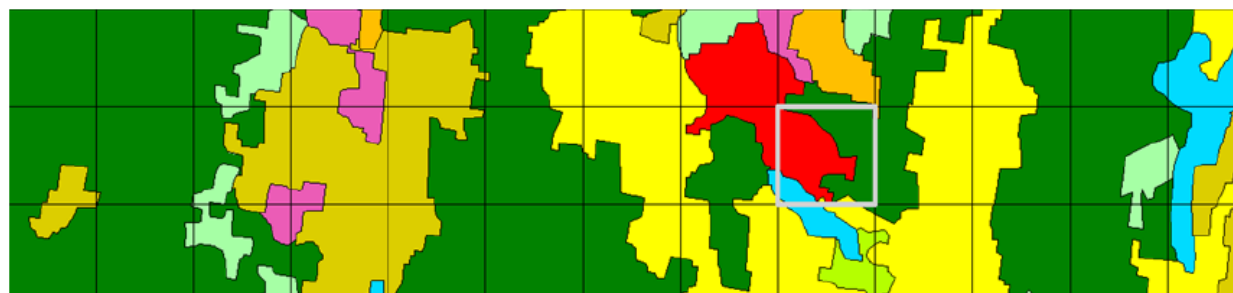
**Table 4-17: Colour palette for the probability layers**

Probabilities					
Class Code	Class Name	Red	Green	Blue	
0-100	Probabilities 0-100%	-	-	-	
101	Areas excluded by binary Mask	128	128	128	
255	Outside Area	0	0	0	

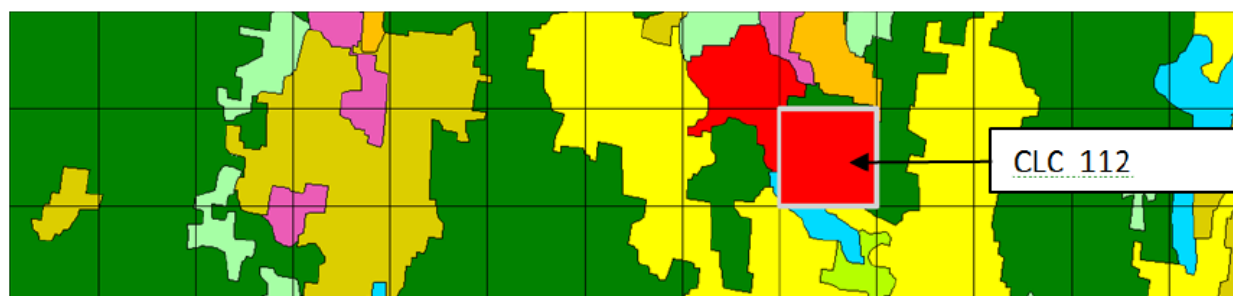
## 5 Prototype Implementation: HRL Combined Layer

In a draft section in this ITT, the next step after the creation of the CLC+ Backbone, is outlined – and called the CLC+ Core product.

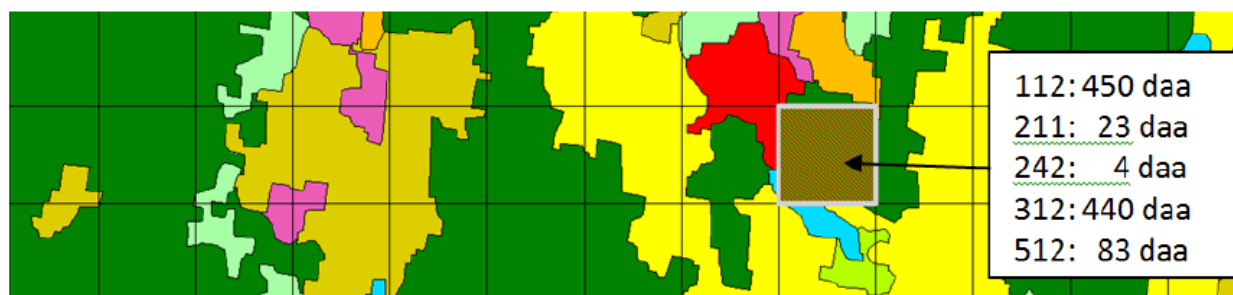
This new generation dataset will be presented in a new format, between raster and vector shapes: a multi-GRID information system. Where each pixel of a raster is classified, i.e. assigned a single particular LC, the grid cells can be characterized by how much it contains of each LC or other information. An example can be seen on Figure 5-1.



CLC with 1 km raster/grid superimposed



Raster representation: A single class is assigned to each pixel



Grid representation: The internal composition of the grid cell is kept as an attribute vector.

Figure 5-1: - Figure extracted from the ITT EEA/DIS/R0/19/012 - caption can be read as: " CLC with a 1 km raster/grid superimposed (top) illustrating the difference between encoding a particular unit as raster pixel (centre) or a grid cell (bottom). "daa" is a Norwegian unit: 10 daa = 1 ha." Reproduced from [AD08].

The creation of a pan-European HR LC layer will be obtained by superposing together all the currently available layers, in addition to the newly created agricultural layer in this project.

This first superposition step could also be an example of the content expected in this GRID format product, even though it is based solely on the HRLs, and at the resolution of those layers, 20m by 20m pixels.

It is expected that the grid size should be for example 10 by 10m to match the content of the datasets used to fill it – which will stay below the MMU of the grid, that could be 0.5ha or lower to match the CLC+ Backbone product specifications. At the moment, the CLC+ Core is expected to be populated with:

- CLC+ Backbone;
- HRLs (not only LC information, but also parameters such as tree density or sealing degree);
- Hot Spot monitoring (Urban Atlas, Natura 2000, Riparian Zones);
- Any other relevant datasets: CLC, LPIS, OSM ...

It should be noted that, to display such GRID interactive format, an online access to a server and its database is required – this cannot be provided as deliverable in the framework of this project.

The next step of the work presented here, which consists in merging the superimposed layers and can be presented as a raster, constitutes an opportunity to enforce a logical consistency between the current and upcoming thematic products, which are being produced independently, without requiring post-processing to ensure the spatial and temporal coherence.

No testing or benchmarking is required to produce this fused layer, which is does not require classification methodology, therefore details of the implementation are provided directly in this report, without being further detailed in the deliverable of WP33 [AD08]. The most important constraint for those new LC products lays in the consistency and continuity of those with the previous LC products.

In the following sections the Data and Processing Setup (5.1) as well as the Experimental Setup (5.2) and the final Results (5.3) of the HRL Combined Layer are described. The Prototype Specifications (5.4) round up this chapter.

## 5.1 Data and Processing Setup

For the prototype on the combination of HRLs 2015, which was implemented in the three different demonstration sites *South-West* (Figure 3-2), *Central* (Figure 3-3), and *West* (Figure 3-4), the following datasets were used:

- LOT1: IMD\_2015\_020m\_eu\_03035\_d04\_full
- LOT2: DLT\_2015\_020m\_eu\_03035\_V1\_1
- LOT3: GRA\_2015\_020m\_eu\_03035\_V1\_4
- LOT4: WAW\_2015\_020m\_eu\_03035\_d06\_full

Since the HRL 2015 on Small Woody Features (SWF) was still under production and not finalized yet it could not be included in the layer combination carried out within ECoLaSS. Detailed specifications, e.g. on the production of the HRLs 2015 or the technical characteristics, can be found on the CLMS homepage (<https://land.copernicus.eu/pan-european/high-resolution-layers>).

Apart from the HRL 2015 datasets the prototypes on Agriculture (AGRI), created within the first project phase of ECoLaSS were included in the analysis; namely CRM\_2017\_010m\_CE\_03035\_prototype\_v01 for the *Central* site, CRM\_2016\_010m\_WE\_03035\_prototype\_v01 for the French part of the *West* site and CRT\_2017\_010m\_WE\_03035\_prototype\_v01 for the Belgian part of the *West* site (all AD11). For *South-West* no AGRI prototype was available. In this site the combined layer only contains the four HRL 2015 layers listed above.

Due to the fact that all HRL 2015 layers have a spatial resolution of 20m the AGRI prototypes had to be resampled to 20m resolution before the layer combination could be done. Furthermore, the IMD layer was reclassified so that the continuous scale from 1-100% of imperviousness was summarized to one class. This led to a binary map with the classes 1 for impervious areas from 1-100% and 0 for all other land cover.

Even though the layer Dominant Leaf Type (DLT) was not reclassified, the final output of the HRL Combined Layer did only compile binary information (classes 0 and 1) regarding this layer instead of 3 classes. The correct naming for this layer thereby is Tree Cover Mask (TCM).

In the HRL Water and Wetness (WAW) pixels that were not classified as part of a water or wetness area were assigned to “NoData”, which is not the case in any other HRL 2015. In order to make a comparison possible these pixels were reclassified and assigned to the value 0.

For the *West* site some preparations were necessary for the AGRI-related layers. Since for the Belgian part of the site only a Crop Type Map (CRT) was available the actual Crop Mask (CRM) had to be derived first. In a second step the two separate prototypes had to be mosaicked to one layer. In the area where both initial Crop Masks were overlapping, all pixels assigned to cropland, in either one or ideally both of the masks, were classified as cropland in the mosaic as well.

## 5.2 Experimental Setup

For the HRL combined layer all 4-5 base layers described in section 5.1 are combined to one single layer containing the respective information in which layer the pixel has a value >0 and where not. The different layers are numbered as depicted in Table 5-1. Out of these layer codes the class combination is derived as follows: if the respective pixel shows an overlap between layer 1 and 2 the respective class combination value is 12. In case one of the pixel is assigned to 254 or 255 in one of the layers the pixel is marked as such and is not further analysed.

**Table 5-1: – Input layers and the referring code number**

Layer	Layer Code
WAW 2015	1
IMD 2015	2
GRA 2015	3
TCM 2015	4
CRM 2017; CRM 2016/2017	5

## 5.3 Results

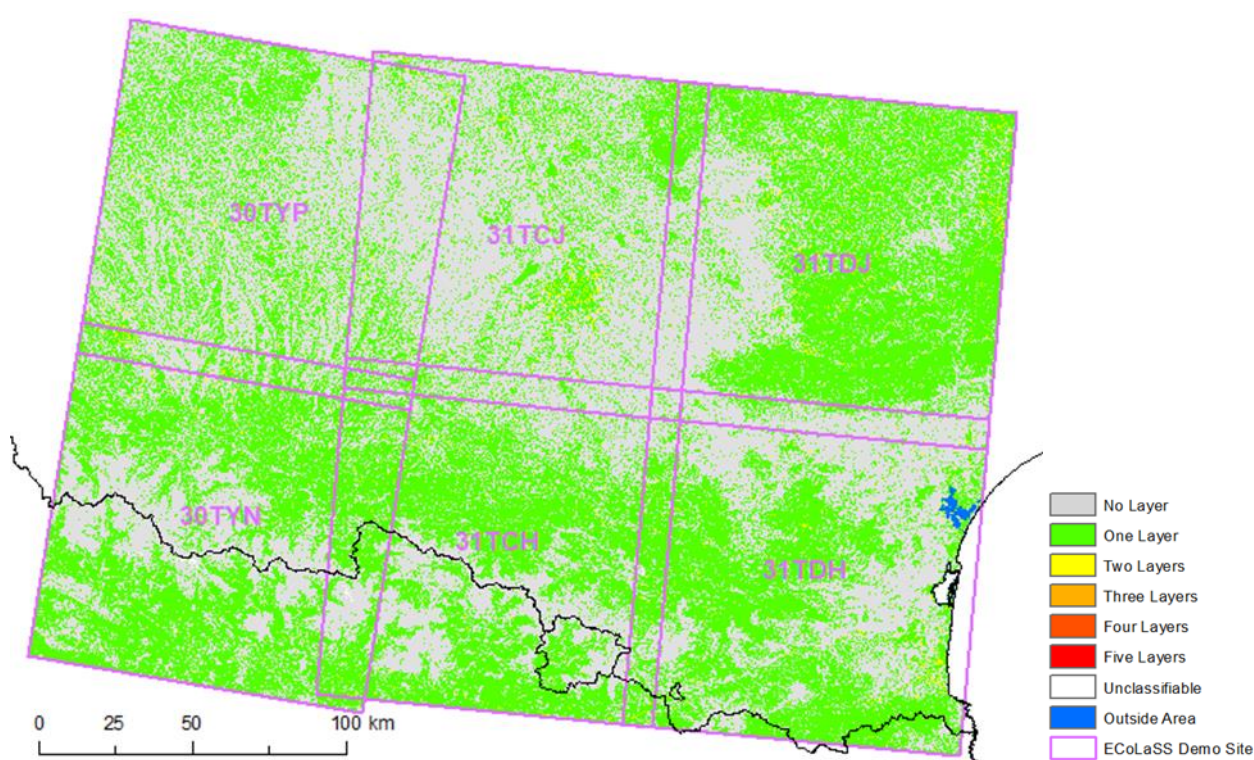
In the following the results for the layer combination for the demonstration sites *South-West* (5.3.1), *Central* (5.3.2), and *West* (5.3.3) are described. The chapter is being closed by an overarching conclusion (5.3.4).

### 5.3.1 Demo-Site South-West

In the *South-West* demonstration site the combined layer shows that more than 49% of the site’s area are assigned to one layer: approx. 35% belong to the TCM, about 11% belong to GRA, 2% to IMD and approx. 1% to WAW (see Table 5-2). As already described in section 5.1 for this demonstration site no agriculture related layer is available, so that the combined layer consists of four different layers.

**Table 5-2: – Distribution of pixels within layer combination in the *South-West* site.**

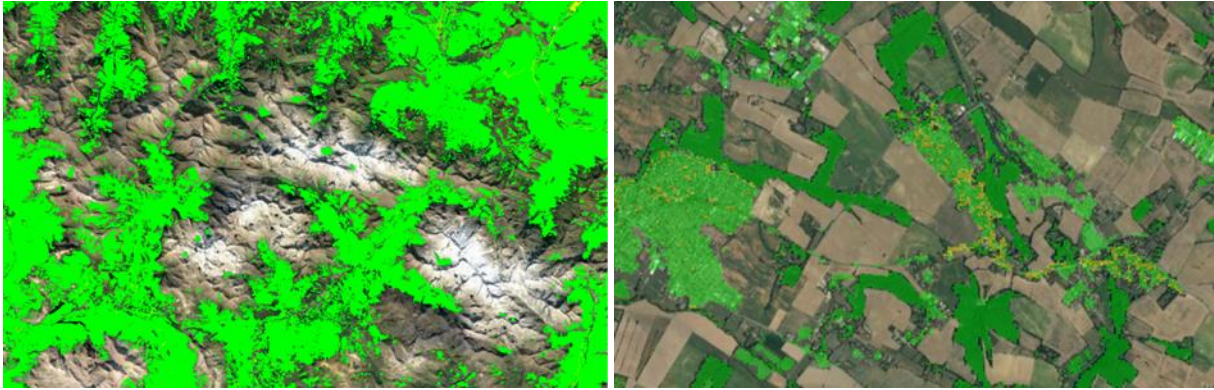
Layer Combination	Pixel Count	Percentage of overall pixel amount
0	76.504.794	49,41
4	55.153.096	35,62
3	16.954.726	10,95
2	3.160.848	2,04
1	1.399.986	0,90
34	658.622	0,43
24	487.820	0,32
255	181.576	0,12
14	156.537	0,10
13	75.889	0,05
254	46.542	0,03
23	39.165	0,03
12	5.422	0,00
234	2.436	0,00
134	1.529	0,00
124	864	0,00
123	46	0,00
1234	3	0,00
<b>SUM:</b>	<b>154.829.901</b>	<b>100,00</b>



**Figure 5-2: – Overview of the HRL combined layer over the *South-West* demonstration site.**

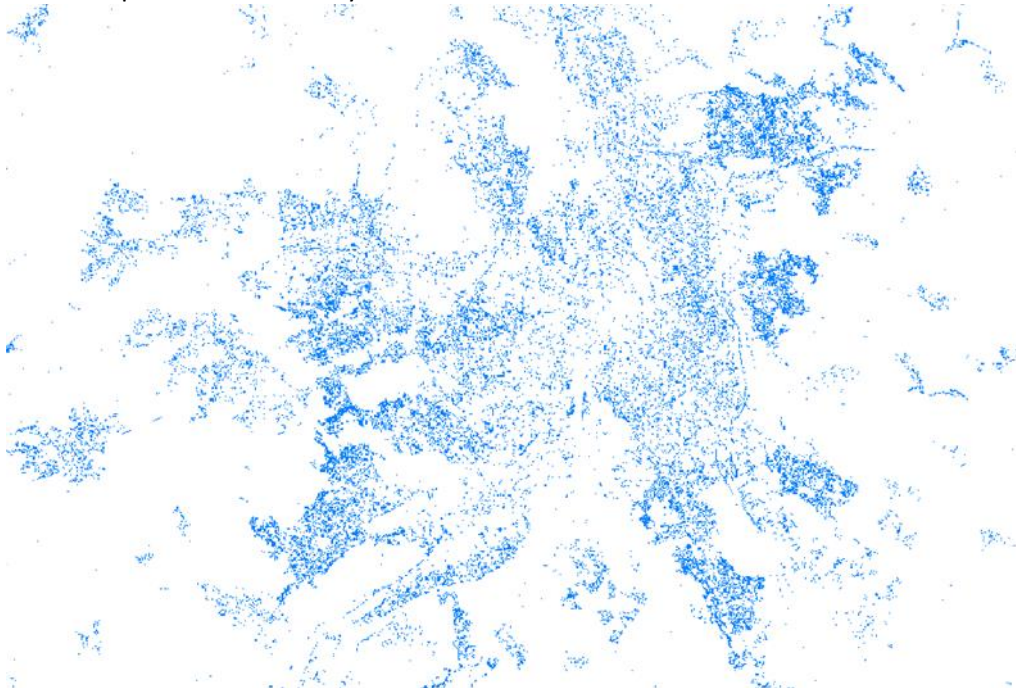
At the same time this is the main reason why nearly 50% of the overall area do not belong to any of the layers included. Apart from the agricultural areas, mountain ranges are partly not covered by any of the HRLs (see Figure 5-3).





**Figure 5-3: – Left: No Layer in Mountainous Areas over the Pyrenees, Spain (colourfree areas). Right: No Layer over agricultural areas near Toulouse, France. The green parts show areas with no overlaps.**  
 © Google. Image © 2018 DigitalGlobe.

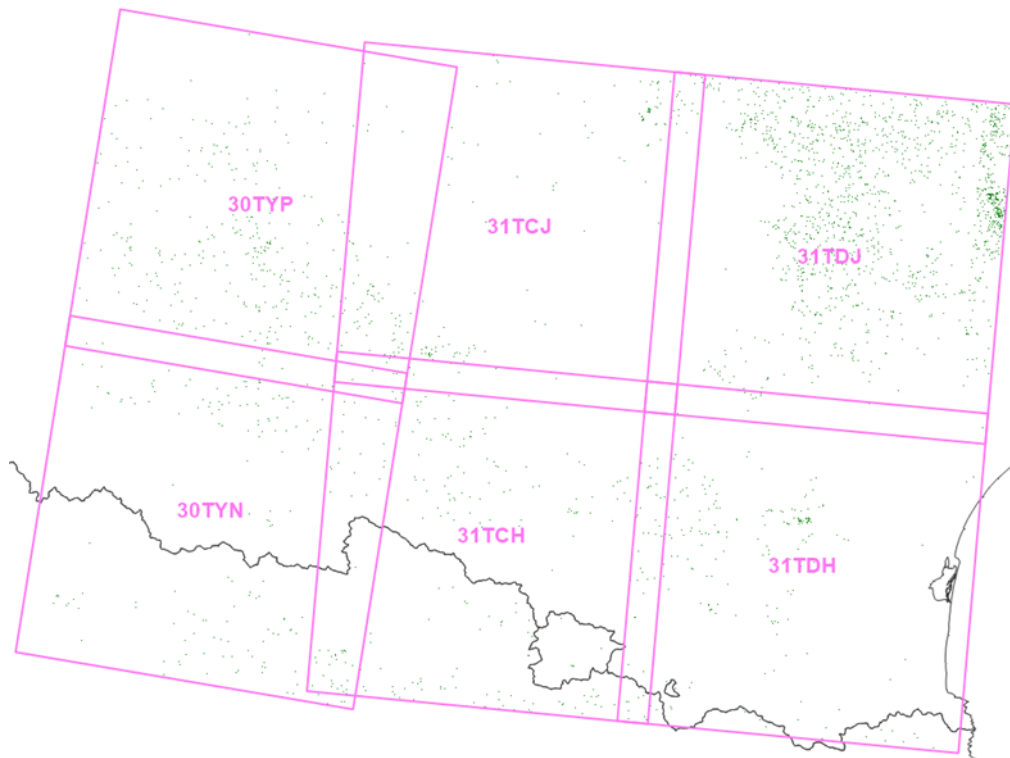
Over cities and villages the amount of overlaps between IMD and TCM increases as can be seen in Figure 5-4 where an example of Toulouse, France, is depicted. Due to the spatial resolution of the HRLs 2015 the rather dynamic changes between sealed areas and trees (like between backyards and houses or roads) cannot be captured satisfactorily.



**Figure 5-4: – Example of overlaps between IMD and TCM (blue) over Toulouse, France.**

What can be also observed in the *South-West* site are overlaps between GRA and TCM, especially in the North-Eastern part of the demonstration site (see Figure 5-5). This is caused by a mixed distribution of both layers in terms of land cover and frequent changes between the two. As overlaps mostly occur at borders between two or more layers and GRA and TCM alternate specifically in this part of the site (tile 31TDJ), this leads to an accumulation of overlapping pixels.





**Figure 5-5: – Example of overlaps between GRA and TCM (green) over the *South-West* demonstration site.**  
© EuroGeographics for the administrative boundaries

### 5.3.2 Demo-Site Central

In the demonstration site *Central* approx. 80% of the overall amount of pixels are clearly assigned to one layer (see Figure 5-6). Forest areas cover the biggest part of the demonstration site with approx. 39%, followed by agricultural areas with approx. 19%. Land cover related to grassland makes approx. 15% of the overall area whereas urban and water areas only cover about 6% and 2%, respectively. 0,26% of the overall pixel amount were unclassifiable, which means that either no satellite images were available or the images were not useful due to clouds, cloud shadows or snow (most of which occurring in the Alps). The amount of pixels in the outside area is negligible (see Table 5-3).

**Table 5-3: – Distribution of pixels within layer combination in the *Central* site**

Layer Combination	Pixel Count	Percentage of overall pixel amount
4	93.793.033	39,08
5	44.331.343	18,47
0	39.649.747	16,52
3	37.021.300	15,42
2	13.124.376	5,47
1	4.177.112	1,74
24	3.553.549	1,48
35	1.117.647	0,47
254	634.181	0,26
14	592.520	0,25
45	520.409	0,22
34	455.888	0,19
15	422.538	0,18
13	188.391	0,08
25	150.354	0,06
23	135.557	0,06
245	56.372	0,02
234	49.682	0,02
12	22.426	0,01
145	7.443	0,00
345	6.175	0,00
135	6.133	0,00
124	6.122	0,00
255	4.628	0,00
134	2.514	0,00
235	1.978	0,00
2345	1.000	0,00
125	421	0,00
123	167	0,00
1245	105	0,00
1345	68	0,00
1234	49	0,00
1235	2	0,00
12345	1	0,00
<b>SUM:</b>	<b>240.033.231</b>	<b>100,00</b>

16% of the overall area belong to no layer. There are mainly four reasons why this happens extensively in this site: mountainous areas with bare rock (see Figure 5-6, lower right), swamps or quarries, that are by definition not part of any of the HRLs, and errors in the CRM 2017 (e.g. omission errors in context to vineyards; see Figure 5-6, upper left). Figure 5-7 to Figure 5-10 show examples of the different areas that belong to no layer.

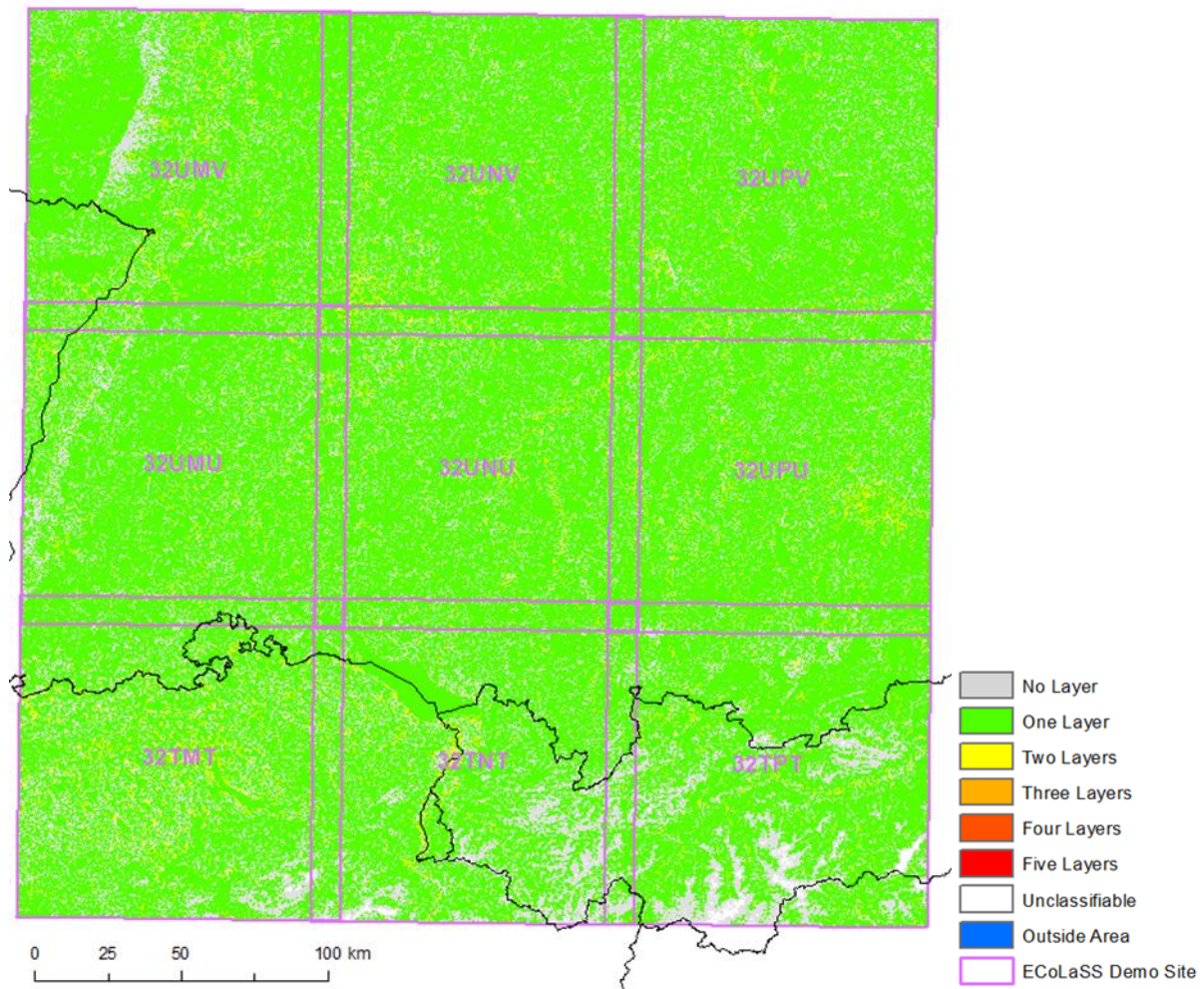


Figure 5-6: – Overview of the HRL combined layer prototype over the *Central* demonstration site.  
 © EuroGeographics for the administrative boundaries

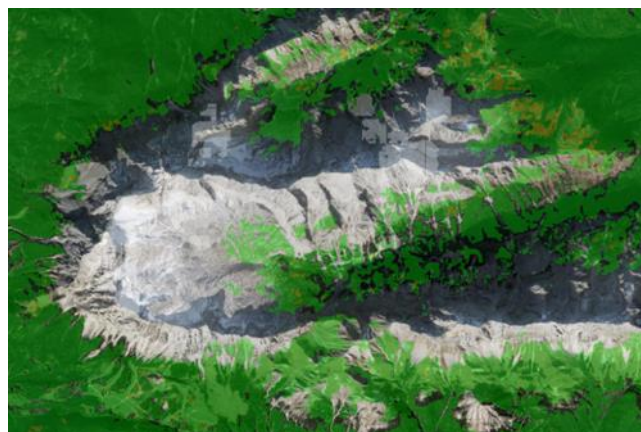


Figure 5-7: – No Layer in Mountainous Areas over the Zugspitze, Germany (colourfree areas). The green parts show areas with no overlaps whereas the yellow parts mark overlaps between 2 layers.  
 © Google. Image © 2018 DigitalGlobe.

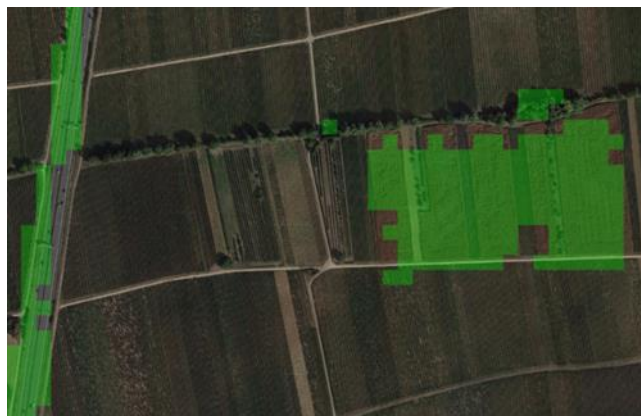




**Figure 5-8: – No Layer over swamps in Southern Bavaria, Germany (colourfree areas).** The green parts mark areas where the pixels have no overlap between different layers whereas the yellow and orange parts mark overlaps between 2 and 3 layers, respectively.  
© Google. Image © 2018 DigitalGlobe.



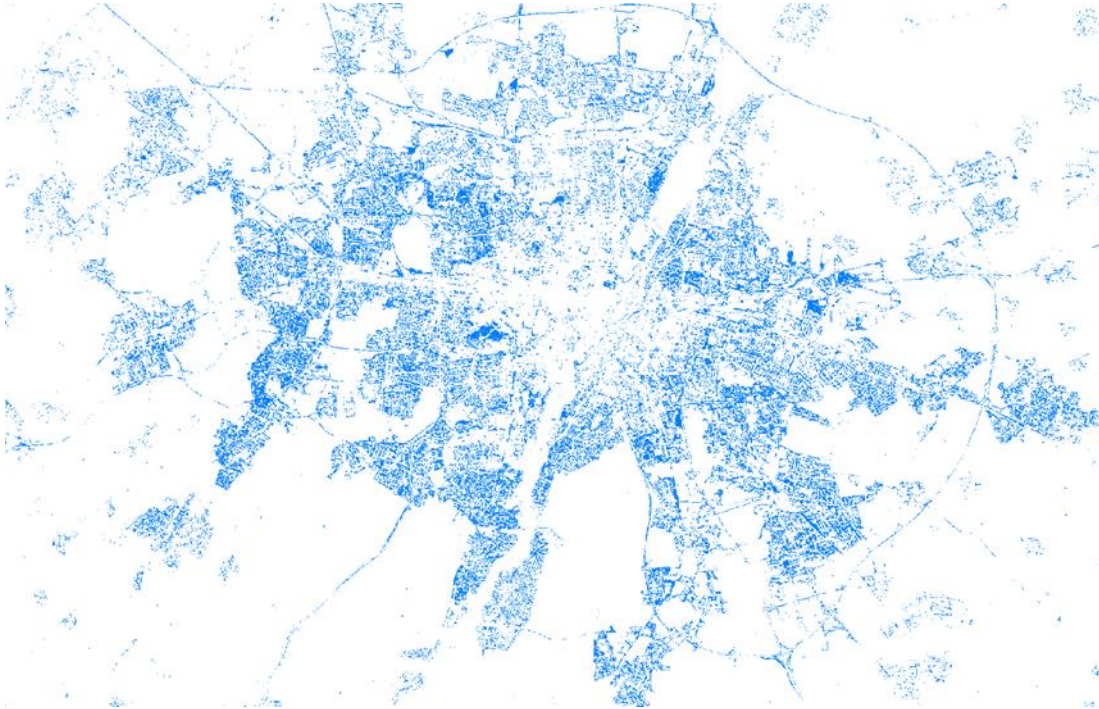
**Figure 5-9: – No Layer over a quarry in Southern Bavaria, Germany (colourfree areas).** The green parts mark areas where the pixels have no overlap with different layers whereas the yellow parts mark overlaps between 2 layers.  
© Google. Image © 2018 DigitalGlobe.



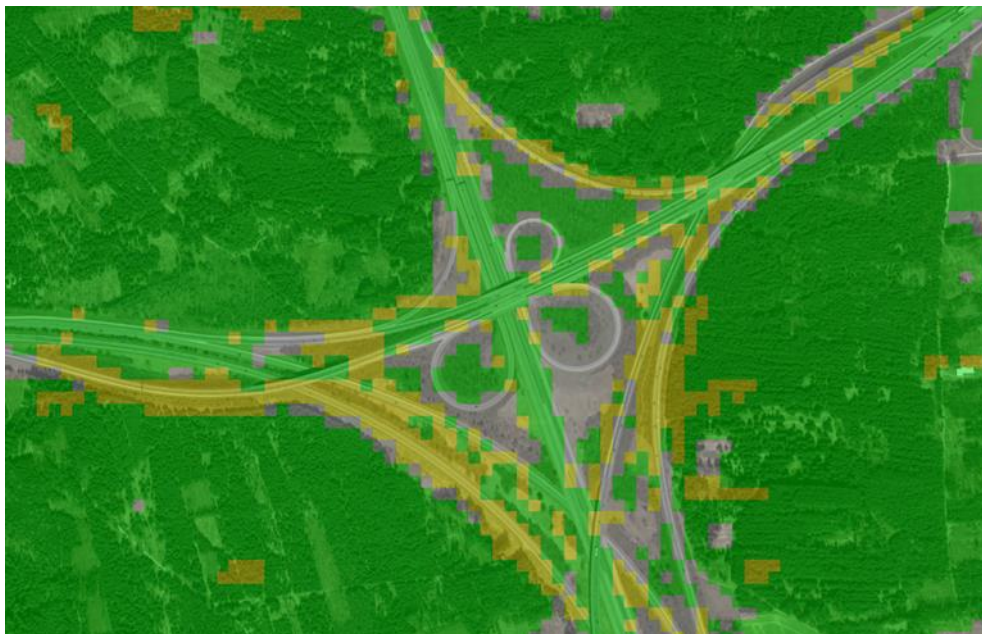
**Figure 5-10: – No Layer over vineyards in the Rhine Valley near Mannheim, Germany (color free areas).** The green parts mark areas where the pixels have no overlap with different layers.  
© Google. Image © 2018 DigitalGlobe.

Over cities and urban areas in general overlaps between the IMD 2015 and the TCM 2015 often occur (see Figure 5-11). One reason for that is the spatial resolution of the HRLs 2015 of 20m. Since changes between land cover classes within cities are rather small-scale, it is difficult to clearly assign one pixel to one class, e.g. over backyards or smaller public green spaces. Furthermore, when comparing the density

products of both the IMD and the TCM layer it gets obvious that within those overlap areas both layers show complementary densities. What can be also observed is the same kind of overlap alongside roads (see Figure 5-12).



**Figure 5-11: – Example of overlaps between IMD and TCM (blue) over Munich, Germany.**



**Figure 5-12: – Example of overlaps between IMD and TCM (yellow) alongside roads.**  
© Google. Image © 2019 DigitalGlobe.

The same effect as shown in Figure 5-12 occurs between the WAW and the TCM layer (see Figure 5-13). These overlaps were already subject of several plausibility checks during the production of the HRL FOR 2015 and can be explained by the fact that in most cases the trees actually extend over water bodies like rivers or lakes and lead to mixed pixels.





**Figure 5-13: – Example of overlaps between WAW and TCM (yellow) at the border between water bodies and tree covered areas at Bodensee, Germany.  
© Google. Image © 2018 DigitalGlobe.**

Another observation was that 15% of the pixels that show an overlap between one or more layers stand for conflicts between grassland and agriculture. The reason behind is that the distinction between agricultural grasslands and natural grasslands is not possible without further historic information (see Figure 5-14). Especially with the methods applied in the frame of the HRLs 2015. This also caused some confusion during the production of the CRM in the first project phase (see AD11) so that in this context the grassland classes had to be excluded from the CRM classification.



**Figure 5-14: – Examples of overlaps between GRA and CRM (violet); green parts are areas where no overlaps occur, whereas the yellow parts show other overlaps between two layers.  
© Google. Image © 2018 DigitalGlobe.**

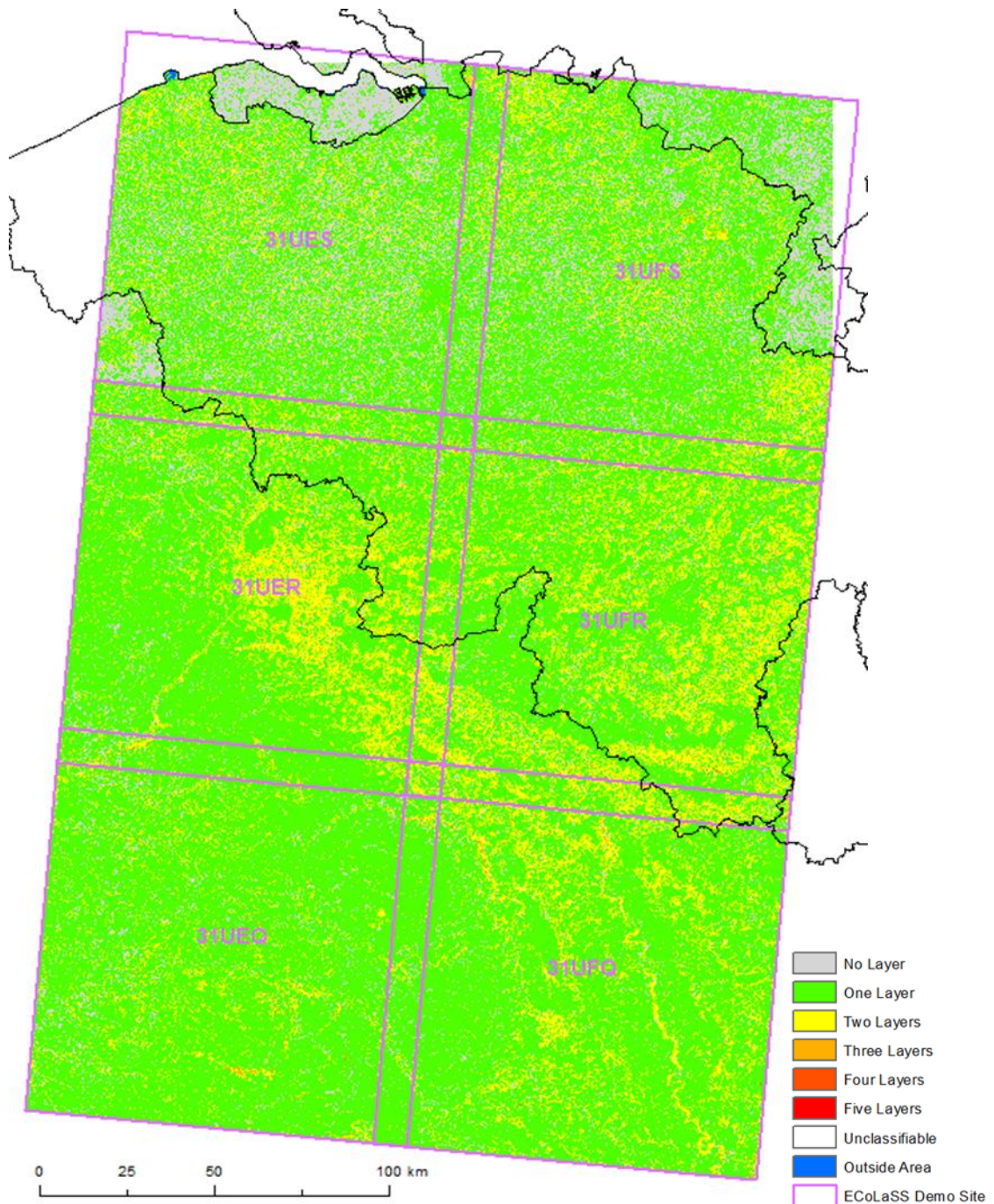
### 5.3.3 Demo-Site West

More than 72% of the amount of pixels in the West demonstration site are only part of one input layer whereas approx. 11% are not assigned to any (see Figure 5-15). Most of the area is cropland (~39%) followed by forests (~24%), impervious areas (~6%), grassland (~3%) and water/wetness (~1%).

**Table 5-4: – Distribution of pixels within layer combination in the West site**

Layer Combination	Pixel Count	Percentage of overall pixel amount
5	61.631.462	38,53
4	38.459.053	24,04
35	19.541.620	12,22
0	16.858.724	10,54
2	9.659.934	6,04
3	5.200.571	3,25
45	2.611.996	1,63
24	1.866.777	1,17
1	1.435.864	0,90
14	562.016	0,35
15	555.116	0,35
345	332.819	0,21
25	277.928	0,17
13	187.558	0,12
34	179.614	0,11
135	169.097	0,11
245	92.249	0,06
23	87.706	0,05
145	80.484	0,05
255	56.536	0,04
235	45.501	0,03
234	21.413	0,01
12	17.904	0,01
2345	12.917	0,01
254	5.058	0,00
124	3.547	0,00
1345	2.115	0,00
125	992	0,00
1245	361	0,00
123	103	0,00
1235	38	0,00
12345	11	0,00
1234	10	0,00
<b>SUM</b>	<b>159.957.094</b>	<b>100,00</b>

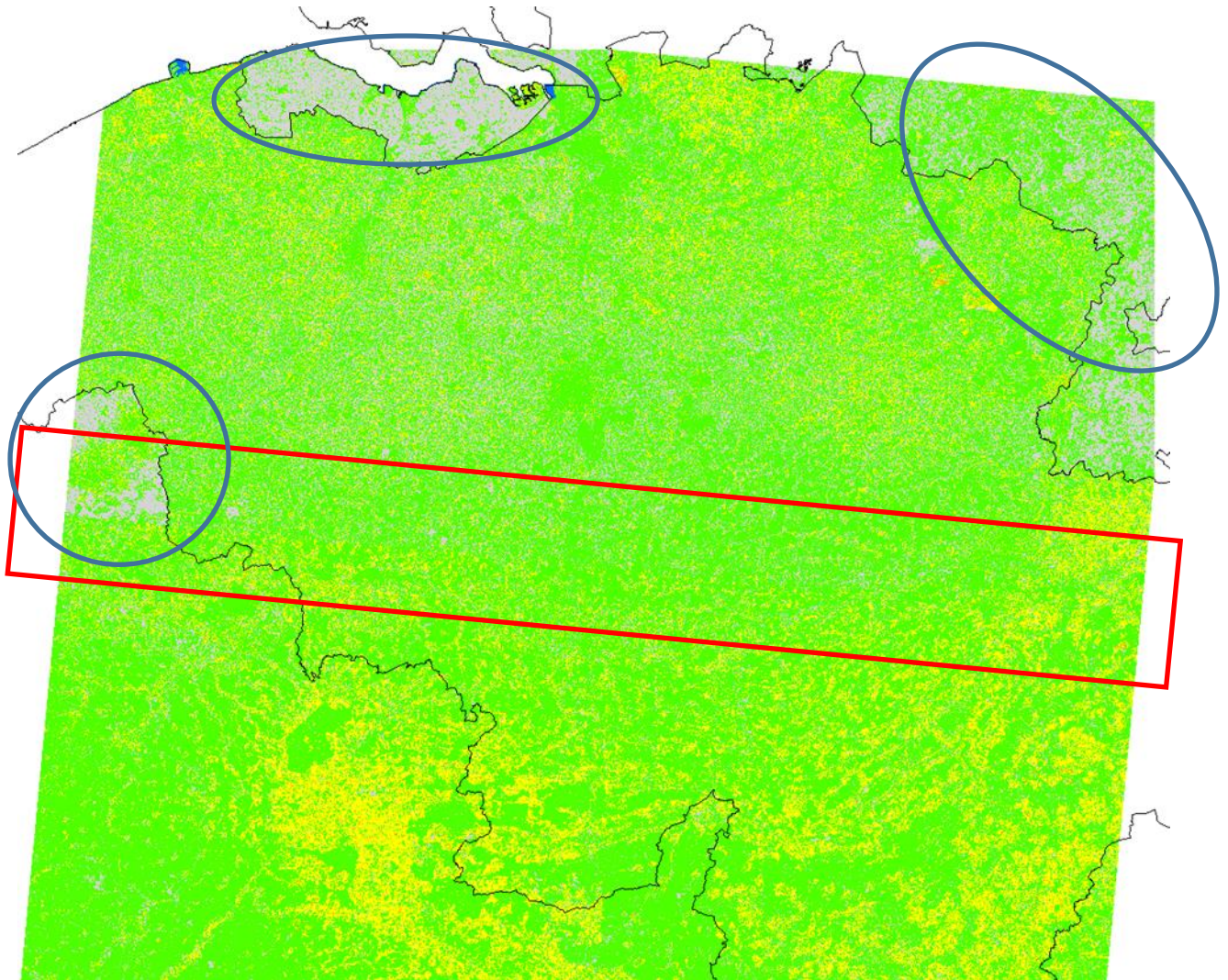




**Figure 5-15: – Overview of the HRL combined layer prototype over the *West* demonstration site.**  
 © EuroGeographics for the administrative boundaries

What catches the viewer’s eye is the obvious cut between the 31UES/31UFS tiles and the rest of the site (see Figure 5-15 and Figure 5-16). This effect is caused by the inclusion of two different crop masks (see 5.1) and thereby two different classification methods which are further explained in AD11. The fact that for the upper two tiles (31UES/31UFS) only the area of Belgium was mapped in the CRM explains the bigger areas with “No Layer” outside the country borders in France, the Netherlands, and Germany (see Figure 5-16, marked in blue).



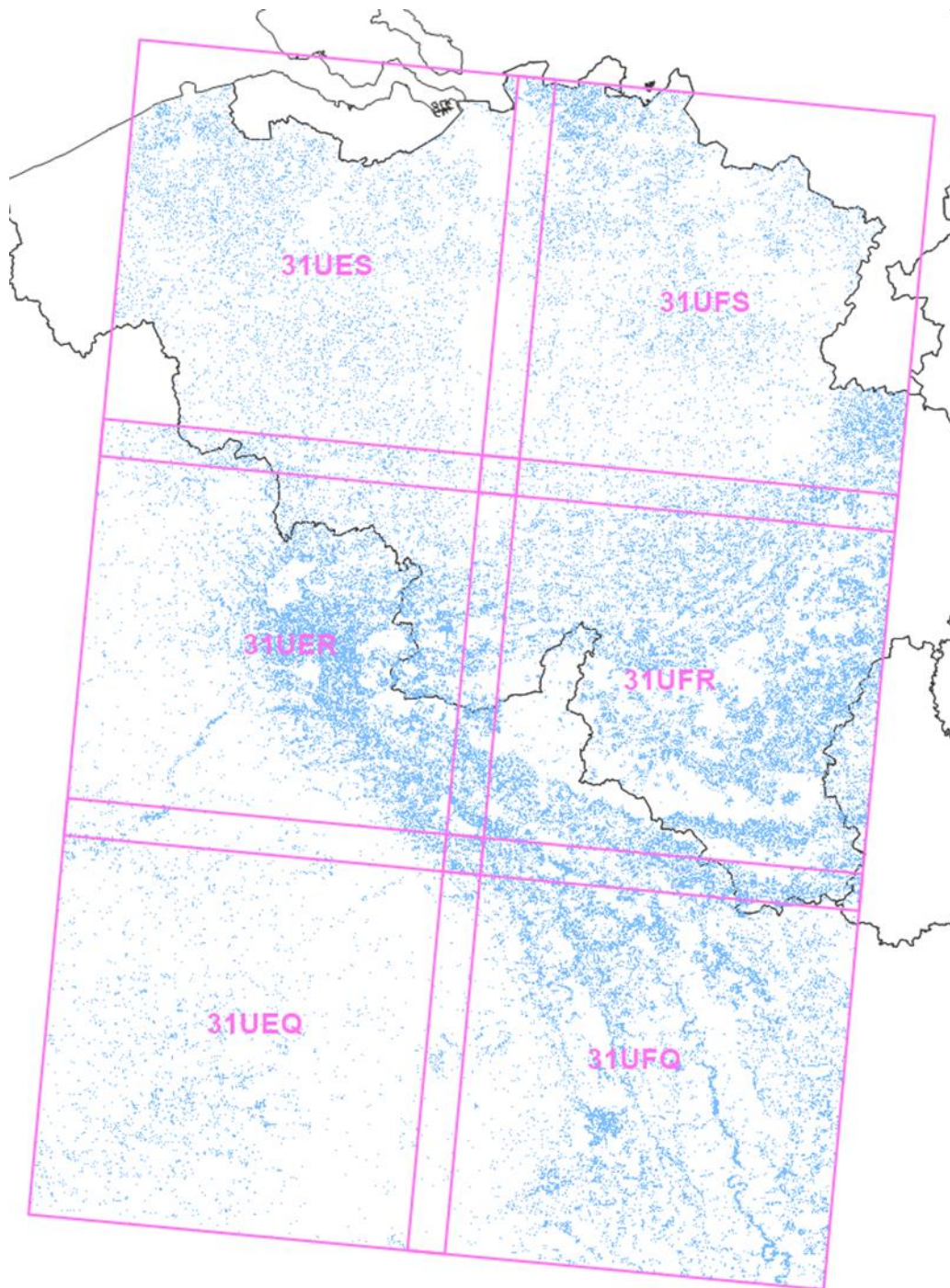


**Figure 5-16: – Inconsistencies due to different crop masks in the *West* site (red box), as well as limitations of mapping extent (blue circles).**

**© EuroGeographics for the administrative boundaries**

The different methods of crop mask production is also the reason for a generally higher frequency of overlaps between GRA and CRM in the lower 4 tiles of the site (31UER, 31UFR, 31UEQ, 31UFQ). Especially in the Ardennes, there are a lot of bigger areas with those overlaps occurring (see Figure 5-17). When compared to the distribution of forests and semi-natural areas shown in Figure 5-15 similarities are visible in the spatial distribution of those overlaps within the demonstration site. The reasons behind this effect are mostly the same as already described in section 5.3.2, caused by difficulties in the distinction between agricultural grasslands and non-agricultural grasslands.





**Figure 5-17: – Distribution of overlaps between GRA and CRM in the West site (in blue).**  
© EuroGeographics for the administrative boundaries

With 1.63% the class with overlaps of TCM and CRM is one of the biggest overlap categories. A further analysis of those areas revealed that the border effect, that could be observed in context to many other overlaps in all three demonstration sites, is mostly the reason in this case as well. Furthermore, in some regions where plantations (e.g., orchards) are part of the land cover the trees are classified as cropland and tree cover at the same time.

As already described in section 5.3.1 and 5.3.2 over cities or sealed areas in general overlaps between IMD and TCM accumulate due to class definition overlaps within the HRLs 2015. This effect can be also observed in the West demonstration site, e.g. over Brussels, Belgium (see Figure 5-18).





Figure 5-18: – Example of overlaps between IMD and TCM (blue) over Brussels, Belgium.

### 5.3.4 Conclusion

As an overarching result it can be stated that the HRLs of 2015 and the CRM 2017 are mostly compliant in the demonstration sites described before. Overlaps predominantly occur at borders between two layers, e.g. between TCM and WAW or TCM and CRM. This effect could already partially be observed during the production of the HRLs as well as during the first phase of ECoLaSS when the CRMs were developed. Furthermore, conflicts evolve out of overlapping layer definitions (as it happened in GRA and CRM) and not fully mature classification algorithms and methods in general. The approach for deriving the HRLs 2015 is based on only very few scenes from the vegetation period whereas for the ECoLaSS prototypes on cropland an approach using dense time series was implemented. Since in the currently ongoing update of the HRLs 2015 the methodological approach was changed towards the use of whole time series, it will be interesting to see what effects this has on the complementarity of the products compared to the CRM. Also the increase in spatial resolution from 20m to 10m is foreseen to increase the accuracy of each layer. Since the CRM was initially produced in 10m resolution as well, this might also decrease the overlaps with the other layers.

## 5.4 Prototype Specifications

This section provides a description of the dataset properties and metadata for the implemented prototypes, also referring to “*P45.2b - Data Sets of New LC/LU Products*”.

The raster products of the ECOLaSS prototypes are delivered as GeoTIFF (\*.tif) with world file (\*.tfw), pyramids (\*.ovr), attribute table (\*.dbf) and statistics (\*.aux.xml), enabling an instant illustration and analysis of the products within Geographic Information System (GIS) software. Each product is accompanied with a product-specific colour table (\*.clr) and INSPIRE-compliant metadata in XML format.

Metadata are provided together with the products as INSPIRE-compliant XML files according to the EEA Metadata Standard for Geographic Information (EEA-MSGI). EEA-MSGI has been developed by EEA to meet needs and demands for inter-operability of metadata. EEA’s standard for metadata is a profile of the ISO 19115 standard for geographic metadata and contains more elements than the minimum required to comply the INSPIRE metadata regulation. Detailed conceptual specifications on EEA-MSGI and other relevant information on metadata can be found at <http://www.eionet.europa.eu/gis>.

Within ECOLaSS, a standardised and harmonised product file naming convention for all prototypes has been developed which is oriented along the already existing naming convention of the CLMS High Resolution Layers. This file naming convention has been applied to all raster prototypes and associated reference files and is documented in the Deliverables of Task 4.

The following three prototypes files as part of **D15.4 – P45.2b – Data Sets of New LC/LU Products (Issue 2)** in the demosites Central, South-West and West were submitted:

- **NLC\_2015\_020m\_SW\_03035\_prototype\_v01.tif**
- **NLC\_2015\_020m\_WE\_03035\_prototype\_v01.tif**
- **NLC\_2015\_020m\_CE\_03035\_prototype\_v01.tif**

The naming convention consists of the following 7 descriptors:

**THEME      YEAR      RESOLUTION      EXTENT      EPSG      TYPE      VERSION**  
as follows:

### THEME

3 letter abbreviation for main products (DLT (dominant leaf type), TCC (tree cover change), GRA (grassland), IMD (imperviousness degree), IMC (imperviousness change classified), CRT (crop type), CRM (crop mask), NLC (new land cover products).

### REFERENCE YEAR

2017 in four digits; change/combination products in four digits (e.g. 1517)

### RESOLUTION

Four-digit (020m and 010m)

### EXTENT

2-digit code for demonstration-sites (CE (Central), NO (North), WE (West), SW (South-West), SE (South-East), SA (South Africa), ML (Mali)).

### EPSG

5-digit EPSG code (geodetic parameter dataset code by the European Petroleum Survey Group) “03035” for the European LAEA projection.

## TYPE

prototype

## VERSION

3-digit code, e.g. "v01"

### EXAMPLE:

"DLT\_2017\_010m\_NO\_03035\_prototype\_v01.tif" stands for: Dominant Leaf Type, 2017 reference year, 10m, Demonstration-site North, European projection (EPSG: 3035), prototype, version 01

"TCC\_1517\_020m\_NO\_03035\_prototype\_v01.tif" stands for: Tree Cover Change, 2015-2017 change period, 20m, Demonstration-site North, European projection (EPSG: 3035), prototype, version 01

The prototype specifications for the NLC layer are listed in Table 5-5.









**Table 5-5: – Detailed specifications for primary 10m NLC status layer**

<b>Products</b>
NLC_2015_020m_SW_03035_prototype_v01.tif NLC_2015_020m_CE_03035_prototype_v01.tif NLC_2015_020m_WE_03035_prototype_v01.tif
<b>Extent</b>
Demo site South-West, Central, and West
<b>Geometric resolution</b>
Pixel resolution 20m, grid to fully conform to the EEA Reference Grid.
<b>Coordinate Reference System</b>
European ETRS89 LAEA projection
<b>Geometric accuracy (positioning scale)</b>
Less than one pixel (According to ortho-rectified satellite image base delivered by Theia.)
<b>Thematic accuracy</b>
NA
<b>Data type</b>
16bit unsigned Raster, compressed with LZW
<b>Minimum mapping unit (MMU)</b>
One pixel (20m*20m)
<b>Necessary attributes</b>
Raster value, count, class name, number of layers, area km <sup>2</sup>

<b>Raster coding (Thematic pixel values)</b>	
HRL Combined Layer – NLC_2015_20m	
0: No Layer 1: One Layer 2: 2 Layers overlapping 3: 3 Layers overlapping 4: 4 Layers overlapping 5: 5 Layers overlapping 254: Unclassifiable 255: Outside Area	
<b>Metadata</b>	
XML metadata files are to be produced according to INSPIRE metadata standards	
<b>Delivery format</b>	
GeoTIFF	

The associated color palette can be found in the Table 5-6.

**Table 5-6: – Palette used for the 20m HRL combined layers**

New Land Cover 2017 – NLC_2017_10m					
Class Code	Class Name	Red	Green	Blue	
0	No Layer	204	204	204	
1	One Layer	85	255	0	
2	Two Layers	255	255	0	
3	Three Layers	255	170	0	
4	Four Layers	255	85	0	
5	Five Layers	255	0	0	
254	Unclassifiable	255	255	255	
255	Outside Area	0	112	255	

## 6 Conclusion and Outlook

In this second phase, the specifications for CLC+ Backbone have been clarified by the release of the ITT lead to the generation of both raster and vector products, while remaining in the frame of the ECoLaSS project.

Raster products, using automatic pixel-based classification, give excellent results, superior to requested accuracy of 90%. Conclusions on vector products with 18-classes nomenclature are less significant as improvement needs to be done on automatization of soft bones creation and validation to obtain the requested 95% threshold for the results.

As demonstrated here, a segmentation only based on S-2 retains flaws that could be easily corrected with either the integration of VHR data, or the redesign of the segmentation process, by adding further steps to ensure that all polygons have a dominant class (above the 50% threshold), for example by using another algorithm such as the connected components.

Regarding the combined HRL Layer, it can be stated that the HRLs of 2015 and the CRM 2017 are mostly compliant in the demonstration sites. Overlaps predominantly occur at borders between two layers, e.g. between TCM and WAW or TCM and CRM. This effect could already partially be observed during the production of the HRLs as well as during the first phase of ECoLaSS when the CRMs were developed. Furthermore, conflicts evolve out of overlapping layer definitions (as it happened in GRA and CRM) and not fully mature classification algorithms and methods in general.

The approach for deriving the HRLs 2015 is based on only very few scenes from the vegetation period whereas for the ECoLaSS prototypes on cropland an approach using dense time series was implemented. Since in the currently ongoing update of the HRLs 2015 the methodological approach was changed towards the use of whole time series, it will be interesting to see what effects this has on the complementarity of the products compared to the CRM. Also, the increase in spatial resolution from 20m to 10m is foreseen to improve the accuracy of each layer. Since the CRM was initially produced in 10m resolution as well, this might decrease the overlaps with the other layers.



## 7 References

- Breiman, L. (2001). Random Forests. *Machine Learning*, 4(1), 5-32.
- Hagolle, O., Huc, M., Villa Pascual, D., & Dedieu, G. (2010, August). A multi-temporal method for cloud detection, applied to FORMOSAT-2, Venµs, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, 8, 1747-1755.
- Hagolle, O., Huc, M., Villa Pascual, D., & Dedieu, G. (2015, March). A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, Landsat, VENµS and Sentinel-2 images. *Remote Sensing*, 7(3), 2668-2691.
- Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Shang, N., & Breiman, L. (1996). Distribution based trees are more accurate. *Ionosphere*, 33(2), 351.