



EvoLand
LAND MONITORING EVOLUTION

D3.2 EvoLand Prototypes Report - Final version

Work Package 3 – Prototype Demonstration

Authors:

Fabian Berndt, Andreas Langner, Fahad Jahangir, André Stumpf, Linda Moser, Anna Grabatin-Homolka (GAF), Wanda De Keersmaecker, Kasper Bonte, Ruben Van De Kerchove, Tim Ng, Nick Gutkin, Mathilde De Vroey, Astrid Vannoppen (VITO), Alexandre Pennec, Laurine Meunier, Hoa Vu, Yoann Courmont, Eve Poitevin, Christophe Fatras, Iris Lucas, Loic Faucqueur, Justine Huge (CLS), Martin Puhm (JR)

Date: 29.10.2025



**Funded by
the European Union**

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Full Title	Evolution of the Copernicus Land Service portfolio (integrating novel EO data and latest Machine Learning algorithms to continuously monitor the status, dynamics and biomass of the land surface)		
Project number	101082130	Acronym	EVOLAND
Start date	01.01.2023	Duration	36 months
Granting authority	European Health and Digital Executive Agency (HaDEA)		
Project Coordinator	Vlaamse Instelling voor Technologisch Onderzoek N.V. (VITO)		
Date of delivery	Contractual	M33	Actual M34
Type	R - Document, report	Dissemination level	SEN - Sensitive
Lead beneficiary	GAF		
Lead author	Fabian Berndt	Email	Fabian.berndt@gaf.de
Other authors	Andreas Langner, Fahad Jahangir, André Stumpf, Linda Moser, Anna Grabatin-Homolka (GAF), Wanda De Keersmaecker, Kasper Bonte, Ruben Van De Kerchove, Tim Ng, Nick Gutkin, Mathilde De Vroey, Astrid Vannoppen (VITO), Alexandre Pennec, Laurine Meunier, Hoa Vu, Yoann Courmont, Eve Poitevin, Christophe Fatras, Iris Lucas, Loic Fauqueur, Justine Hugué (CLS), Martin Puhm (JR)		
Reviewer(s)	Tim Ng (VITO)		
Keywords	Prototypes, demonstrators, implementation, Phase 1, Phase 2		

Document Revision History				
Version	Issue date	Stage	Changes	Contributor
1.0	31.08.2024	Interim Version (D3.1)		GAF, VITO, CLS
2.0	29.10.2025	Final Version (D3.2)	Integration of final prototype results	GAF, VITO, CLS, JR

Disclaimer

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Copyright message

© **EvoLand consortium, 2025**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgment of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged

Acronyms

Acronym	Description
AD	Applicable Document
AE	AutoEncoders
AGB	Above Ground Biomass
AI	Artificial Intelligence
ALOS	Advanced Land Observing Satellite
AOI	Area Of Interest
AP	Attribute Profiles
AWEIsh	Automatic Water Extraction Index with shadow
BEAST	Bayesian Estimator of Abrupt change, Seasonal change, and Trend
BFAST	Breaks for Additive Season and Trend
C	Prototype demonstrator (i.e., C1, C2, ..., C11)
CAP	Common Agricultural Policy
CCI	Climate Change Initiative
CDSE	Copernicus Data Space Ecosystem
CGLOPS	The Global Land Component of the Copernicus Land Service
CGLS	Copernicus Global Land Service
CLMS	Copernicus Land Monitoring Service
CNES	Centre national d'études spatiales
CTY	Crop Type
DAP	Differential Attribute Profiles
DEM	Digital Elevation Model
DOY	day of the year
DSM	Digital Surface Model
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EEA	European Environment Agency
EFFIS	European Forest Fire Information System
EO	Earth Observation
EoS	end of season
ESA	European Space Agency
EU	European Union
EWMA	Exponentially Weighted Moving Average
FAO	Food and Agriculture Organization
FAPAR	Fraction of Absorbed Photosynthetic Active Radiation
FCH	Forest Canopy Height
FNEWSs	Remote sensing based National Detection System for Forest Damages (for Germany)
GCOS	Global Climate Observing System
GDMP	Gross Dry Matter Productivity
GEDI	Global Ecosystem Dynamics Investigation
GEOBIA	Object-Based Image Analysis
GFMAP	General Framework for Mapping
GPP	Gross Primary Production
GRD	Ground Range Detected
GSHHG	Global Self-consistent, Hierarchical, High-resolution Geography
GSW	Global Surface Water
HRL	High-Resolution Layer
ICOS	Integrated Carbon Observation System
IMD	Imperviousness Density
IT	Information Technology
LCFM	Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring service

LPIS	Land-Parcel Identification System
LR	Low Resolution
LSC	Land Surface Categories
LUCAS	land Use / Cover Area Frame Survey
LUE	Light Use Efficiency
MAE	Masked Autoencoders
MLP	Multi-Layer Perception
MMDC	Multi-Modal DataCube
MMU	Minimum Mapping Unit
NASA	National Aeronautics and Space Administration
MBWI	Multi-Band Water Index
NDBI	Normalized Difference Built-up Index
NDSI	Normalized Difference Snow Index
NDTI	Normalized Difference Tillage Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NFI	National Forest Inventory
NISAR	NASA-ISRO SAR Mission
NRT	Near-Real Time
NVLCC	Non-Vegetated Land Cover Characteristics
OA	Overall Accuracy
OLS	Ordinary Least Squares
PALSAR	Phased Array L-band Synthetic Aperture Radar
PA	Producer Accuracy
PSUs	Primary Sampling Unit
RMSE	Root Mean Square Error
S1	Sentinel-1
S-2	Sentinel-2
SAR	Synthetic Aperture Radar
SCL	Scene Classification Layer
SDG	Sustainable Development Goal
SLF	Small Landscape Features
SoS	start of season
SRTM	Shuttle Radar Topography Mission
SSUs	Secondary Sampling Units
SWF	Small Woody Features
SWI	Snow Water Index
SWIR	Short-wave infrared
T	Task
TCCM	Tree Cover Change Mapping
TCD	Tree Cover Density
TCW	Tasseled Cap Wetness
TRL	Technical Readiness Level
UA	User Accuracy
UDFs	'user-defined functions'
VH	vertical-horizontal polarization
VHR	Very High Resolution
VLCC	Vegetated Land Cover Characteristics
VV	vertical polarization
WaW	Water and Wetness
WP	Work Package
WVM	Woody Vegetation Mask

Table of Content

List of Tables.....	7
List of Figures.....	10
Applicable documents	16
Executive Summary.....	17
1 Introduction.....	19
1.1 About EvoLand	19
1.2 Purpose and Objectives of WP 3 Prototype Demonstration.....	19
1.3 Link to WP1 Requirements Analysis.....	21
1.4 Link to WP2 Method Development	21
1.5 Link to WP4 Operationality Assessment.....	22
1.6 Document Structure.....	22
2 Prototype Overview	24
2.1 Keywords and Definitions.....	24
2.2 Demonstration Sites	26
2.3 Developed methods transferred to prototypes.....	28
2.4 Code and Data Deliverables.....	30
3 Forest Prototypes	32
3.1 C1 – Continuous Forest Monitoring	32
3.2 C2 – Forest Disturbances.....	50
3.3 C3 – Forest Biomass.....	59
3.4 C12 – Tree Types.....	81
4 Agriculture prototypes	92
4.1 C4 – Cover Crop Type.....	92
4.2 C5 – Grassland and Cropland GPP	100
4.3 C6 – Small Landscape Features	109
5 Water prototypes.....	128
5.1 C7 – Improved Water Bodies.....	128
6 Urban prototypes.....	149
6.1 C8 – Continuous Imperviousness Monitoring.....	149
6.2 C9 – Automated Land Use Mapping of Urban Dynamics	169
7 General Land Cover prototypes.....	187
7.1 C10 – Continuous Mapping of Land Surface Categories.....	187

7.2 C11 – On-demand Land Cover Mapping.....	195
8 Summary and Conclusions.....	202
9 Outlook and Way Forward	207
References.....	208

List of Tables

Table 1. Document structure.....	22
Table 2. Keywords and their definition in the EvoLand project.....	24
Table 3. Prototype sites and time frame covered by reference years, by prototype, for Phase 1 and Phase 2 of the EvoLand project.....	26
Table 4. Methods developed in EvoLand that have been transferred to prototype implementation.....	28
Table 5. Technical and thematic specifications for the C1 prototype.....	32
Table 6. List of datasets used for the C1 prototype development.....	34
Table 7. Confusion matrix of C1 for the Central European demonstration site.....	42
Table 8. Confusion matrix of C1 for the Spanish demonstration site.....	42
Table 9. Confusion matrix of C1 for the Swedish demonstration site.....	42
Table 10. Confusion matrix of C1 for the Central European demonstration site, excluding validation samples with change observed in 2021.....	45
Table 11. Average detection delay in months by demonstration site.....	46
Table 12. Confusion matrix of the TCDM data for the Central European demonstration site.....	48
Table 13. Technical and thematic specifications for the C2 prototype.....	50
Table 14. List of datasets used for the C2 prototype development.....	51
Table 15. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Central European demonstration site.....	55
Table 16. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Spanish demonstration site.....	56
Table 17. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Spanish demonstration site – after exclusion of the samples which were omitted already in C1.....	56
Table 18. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Central European demonstration site – after exclusion of the samples which were omitted already in C1.....	57
Table 19. Technical and thematic specifications for the C3 prototype.....	59
Table 20. List of datasets used for the C3 prototype development.....	61
Table 21. List of datasets used for the C12 prototype development.....	83
Table 22. Classification scheme of the Tree Types Prototype.....	84
Table 23. Test set accuracies for the Tree Types classification.....	86
Table 24. Technical and thematic specifications for the C4 prototype.....	93
Table 25. Input and ancillary data for the C4 cover crop type map.....	95
Table 26. Confusion matrix and performance metrics of the 2020 cover crop type map of Belgium, validated with the LPIS dataset.....	98
Table 27. Confusion matrix and performance metrics of the 2019 cover crop type map of Belgium, validated with the LPIS dataset.....	98
Table 28. Technical and thematic specifications for the C5 prototype.....	100
Table 29. list of datasets used for the prototype development.....	102
Table 30. Overview of different parameters used in the current CLMS GPP model.	103

Table 31. Validation metrics for the first version of the cropland and grassland GPP prototype. The prototype was validation using the ICOS GPP in-situ data (for cropland and grassland separately) and with the CGLS 300m dataset (for cropland and grassland together).	108
Table 32. Technical and thematic specifications for the C6 prototype.	109
Table 33. C6 prototype validation sample distribution per test site.	122
Table 34. C6 Thematic accuracy for Spain test site for 2020.	122
Table 35. C6 Thematic accuracy for Spain test site for 2022.	122
Table 36. C6 Thematic Accuracy for Eastern Europe Test Site for 2020.	123
Table 37. C6 Thematic Accuracy for Eastern Europe Test Site for 2022.	123
Table 38: C6 Thematic Accuracy for Belgium Test site for 2020.	123
Table 39. C6 Thematic Accuracy for Belgium Test site for 2022.	124
Table 40. C6 Thematic accuracy for Change layers (all sites).	124
Table 41. Technical thematic specifications for the C7 prototype.	130
Table 42. EO Ancillary data for water bodies mapping prototype.	131
Table 43. Confusion matrix for the Belgium test site.	144
Table 44. Confusion matrix for the Spain test site.	144
Table 45. Confusion matrix for the Ivory Coast test site.	144
Table 46. Confusion matrix for the Sweden test site.	145
Table 47. Confusion matrix for the Sweden test site (Summer).	146
Table 48. Confusion matrix for the Sweden test site (Winter).	147
Table 49. Technical thematic specifications for the candidate C8.	149
Table 50. list of input datasets used for the C8 prototype development.	152
Table 51. list of ancillary datasets used for the C8 prototype development.	152
Table 52. A summary of regression line parameters for the scatterplots derived from the plausibility analysis as reference data.	161
Table 53. Summary of the internal accuracy assessment extracted from the confusion matrices of the Imperviousness Degree product with the threshold set to 1% (OA= Overall Accuracy; UA = User Accuracy; PA = Producer Accuracy). Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 10% and red cells fall more than 10% short.	162
Table 54. Summary of the internal accuracy assessment extracted from the confusion matrices of the Imperviousness Degree product with the threshold set to 30% (OA= Overall Accuracy; UA = User Accuracy; PA = Producer Accuracy). Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 10% and red cells fall more than 10% short.	163
Table 55. C8 Thematic accuracy for the change layer (Tile 31UFS; Belgium test site) based on the weighted confusion matrix.	166
Table 56. C8 Thematic accuracy for the change layer (Tile 31UFS; Belgium test site) based on the unweighted confusion matrix.	166
Table 57. Prototype specifications.	169
Table 58. EO and ancillary datasets description.	172
Table 59. Intermediate omission rates.	175
Table 60. Change classes.	177

Table 61. Change / no change confusion matrix for the Southern France site for 2020-2021.....	179
Table 62. Change / no change confusion matrix for the Southern France site for 2021-2022.....	179
Table 63. Change / no change confusion matrix for the Belgium site for 2020-2021.....	179
Table 64. Change / no change confusion matrix for the Belgium site for 2021-2022.....	179
Table 65. Change / no change confusion matrix for the central Europe site for 2020-2021.....	180
Table 66. Change / no change confusion matrix for the central Europe site for 2021-2022.....	180
Table 67. Change / no change confusion matrix for the Côte d'Ivoire site for 2020-2021.....	180
Table 68. Global change / no change confusion matrices for all sites.....	181
Table 69. Global change category confusion matrices for all sites.....	184
Table 70. Global 2020-2021 classes confusion matrices for all sites.....	184
Table 71. Technical thematic specifications for the C10 prototype.....	187
Table 72. List of datasets used for the C10 prototype development.....	189
Table 73. Class legend for the C10 prototype development.....	189
Table 74. Overview model benchmark.....	192
Table 75. Confusion matrix of the LSC model over the test data.....	193
Table 76. Technical thematic specifications for the C11 prototype.....	196
Table 77. List of datasets used for the C11 prototype development.....	197
Table 78. Class legend for the C11 prototype forest management type classification.....	197
Table 79. List of input data for the ML model of the C11 prototype.....	198

List of Figures

Figure 1. Demonstration sites for the EvoLand project.....	26
Figure 2. Overview of a forest in the Central European demonstration site (near Steinach, Germany) affected by a bark beetle infestation, being cleared step by step to avoid a further spread of the pest, beginning in summer 2019. Monitoring period 2019-2021 (Tile 32UPA, base imagery: S-2).....	37
Figure 3. Middle: Forest clearing because of a bark beetle infestation starting in the summer 2019. Monitoring period 2019-2021 (Germany, tile 32UPA). Top & bottom: Tree cover density (TCD) from the HRL Vegetated Land Cover Characteristics (VLCC), depicting the situation before (2018) and after the event (2021).....	38
Figure 4. Forest affected by a storm event in 2019. Left: Tree cover density (TCD) from the HRL Vegetated Land Cover Characteristics (VLCC), depicting the situation before (2018, upper left) and after the event (2020, lower left). Upper right: C1 for the monitoring period 2019-2021 (Central site, tile 32UNA). Lower right: Google Earth imagery from 08/23/2019.....	39
Figure 5. C1 result for an AOI in French Guiana, showing correct detections of tree cover disturbances (yellow boxes), surrounded by massive false detections.....	40
Figure 6. Location of the sample points for the Central European (top), the Spanish (lower left) and the Swedish demonstration site (lower right).	42
Figure 7. Time-series of monthly composites of Tasseled-Cap Wetness (TCW) data, analysed for point A of Figure 8 (Central European site).....	43
Figure 8. Theoretical examples visualizing the temporal offset for the detection of disturbance events occurring at different points in time during the monitoring window for a certain year (here April to September, like applied in the Central site) and with differing intensities of signal changes. Blue represents normal spectral behaviour, whereas the shift towards red represents signal changes with varying intensities.....	45
Figure 9. C1 for an area in the Spanish demonstration site, which suffered from several subsequent fire events. The green delineations show the extent and the recorded year and months of the single fire events, as provided by the European Forest Fire Information System (EFFIS).....	47
Figure 10. Top left: S-2 acquisition from 17/04/2019; Top right: C1 detecting a disturbance caused by a bark beetle infestation starting in summer 2019. Monitoring period 2019-2021 (Germany, tile 32UPA); Mid-left: C2 hard classification of disturbance agents; Mid-right: C2 class probabilities for the disturbance agent classes: bark beetle (red), clear cut (green), and salvage logging(blue); Lower-left & -right: Google Earth imagery before the pest and at a later stage, showing the progressing tree dieback (brownish colour).....	54
Figure 11. A subset of the Central demonstration site, showing large areas of salvage logging to prevent the spread of a bark beetle infestation. Upper left: S-2 acquisition after the logging activities; Upper right: Result of C1;	

Lower left: Disturbance agent classes; Lower right: Disturbance agent probabilities as RGB composite.....	54
Figure 12. Training accuracy for the 2021 AGB in the Sweden test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.....	63
Figure 13. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2021 (Phase 1 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.-09.2021. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	64
Figure 14. Plot showing the predicted versus the actual biomass for the validation samples of the 2021 Sweden test site (Phase 1).	65
Figure 15. Plot showing the predicted versus the actual biomass for the validation samples of the 2020 Sweden test site (Phase 2).	66
Figure 16. Plot showing the predicted versus the actual biomass for the validation samples of the 2022 Sweden test site (Phase 2).	66
Figure 17. ESA CCI Biomass dataset of 2021– Swedish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	67
Figure 18. C3 AGB dataset of 2021 – Swedish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	67
Figure 19. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2020 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2020. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	68
Figure 20. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2022 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2022. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	69
Figure 21. Training accuracy for the 2021 FCH for the Sweden test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.....	70
Figure 22. Prediction for the FCH of the Sweden test site for the reference year 2021 (Phase 1 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.-09.2021. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	71
Figure 23. Plot showing the predicted vs. the actual FCH for the validation samples of the 2021 Sweden test site (Phase 1).	72
Figure 24. Plot showing the predicted versus the actual tree height values for the validation samples of the 2020 Sweden test site (Phase 2).	72
Figure 25. Plot showing the predicted versus the actual tree height values for the validation samples of the 2022 Sweden test site (Phase 2).	73
Figure 26. Prediction for the FCH of the Sweden test site for the reference year 2020 (Phase 2 result). The arrow indicates the position of a close-up (lower	

left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2020. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.....	74
Figure 27. Prediction for the FCH of the Sweden test site for the reference year 2022 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2022. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.....	75
Figure 28. Training accuracy for the AGB model in the test site Spain, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.....	76
Figure 29. Prediction for the AGB in the Spain test site for the reference year 2021 with a close-up to a higher AGB forest area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	77
Figure 30. Plot showing the predicted vs. the actual biomass for the validation samples of the Spain test site.....	77
Figure 31. ESA CCI Biomass dataset of 2021 of tree canopy height – Spanish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.....	78
Figure 32. C3 AGB dataset of 2021 of tree canopy height – Spanish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.	78
Figure 33. Training accuracy for the FCH model in the Spain test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.....	78
Figure 34. Prediction for the FCH in the Spain test site with a close-up to an area with higher FCH values. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.....	79
Figure 35. Plot showing the predicted vs. the actual FCH for the validation samples of the Spain test site.....	79
Figure 36. AOI for the Tree Types mapping, Central European demonstration site extended northwards.....	82
Figure 37. Distribution of training samples per tree type class in the dataset.....	85
Figure 38. Schematic representation of a TempCNN (Pelletier et al., 2019).	85
Figure 39. TempCNN test set confusion matrix for tree type classification (counts per class, synthetic based on recall and support).	87
Figure 40. Tree Types mapping - final result for the study area.....	89
Figure 41. Detailed zoom-in of Tree Types mapping in the demonstration site (left) compared with true-colour satellite imagery (right).	90
Figure 42. Hierarchical cover crop typology with cover crops grouped by family and then by morphological or phenological similarities.	93
Figure 43. Workflow for the cover crop type candidate.....	96
Figure 44. Locations of the 2020 cover crop type validation samples across Flanders, Belgium.....	98
Figure 45. Cover crop type map of Belgium (2020).....	99
Figure 46. GPP pipeline for grassland and cropland as integrated in OpenEO.....	105

Figure 47. Snapshot (zoomed) of GPP estimation of 31UFS tile for 2022.	105
Figure 48. Scatterplot of GPP target (ICOS tower validation data) and modelled GPP for cropland (left panel) and grassland (right panel), using the GPP model as presented in equation 1.	106
Figure 49. Scatterplot of GPP target (ICOS tower validation data) and modelled GPP for cropland (left panel) and grassland (right panel) using the original GDMP model.	107
Figure 50. Bias in function of month of year, using the GPP model as presented in equation 1.	107
Figure 51. Bias in function of month of year, using the original GDMP model.	108
Figure 52. C6 SLF workflow.	113
Figure 53. VHR imagery (Pleiades) and Woody vegetation probability map as result of the classification step.	115
Figure 54. Grass strip between crop parcels (background: Pléiades).	116
Figure 55. Grass strip on S-2 time series (from left to right: March, April, May, August).	116
Figure 56. NDVI standard deviation between March and August 2020. Arrows indicate grass margin (down) and access track (up).	117
Figure 57. <i>Frangi filter applied on the indices standard deviation in 2020 (from left to right: Google Satellite as background, NDVI standard deviation between March and August 2020, Frangi filter > 0.003 threshold confirmed on both standard deviation of NDVI and NDWI between March and August 2020).</i>	118
Figure 58. C6 SLF results illustrating mapping of a grass margin in Belgium (background: Pléiades).	118
Figure 59. Pond detection (background: Google Satellite).	119
Figure 60. SLF Change identification workflow.	120
Figure 61. C6 prototype results over Belgium site (background; Pléiades).	121
Figure 62. C6 prototype results over Spain site (background; Pléiades).	121
Figure 63. C6 prototype results over Romania site (background; Pléiades).	121
Figure 64. CGLOPS Water bodies production workflow.	129
Figure 65. C7 prototype Workflow during phase 1.	131
Figure 66. Value sampling on multiple test sites.	132
Figure 67. Solar Zenith Angle on Belgium and Spain test sites.	133
Figure 68. Example of shadow impact (left) and estimate (right) in Southern Spain.	134
Figure 69. Example of raw shadow estimation with an opening and dilation of one pixel (in red) over tile 30SWH for January 21st.	135
Figure 70 -Concept of permanent/temporal water pixel classification.	136
Figure 71. Example of permanent mask (purple) derived from GSW occurrences above 80%.	137
Figure 72. Example of permanent mask (purple) derived from a whole year water body occurrence.	137
Figure 73. Rolling map updating for three consecutive months.	138
Figure 74. Example of monthly synthesis taking as an example one pixel in time.	139

Figure 75. Example of permanent (dark blue) and temporary (light blue) water pixels classification (right) from the raw NRT water mask (left), using the rolling map (middle) on Cuevas de Almanzora Lake (Southern Spain) on September 10th, 2022.....	139
Figure 76. C7 prototype workflow during phase 2.....	140
Figure 77. C7 prototype results (summer above, winter below) in Spain.....	140
Figure 78. C7 prototype results (summer above, winter below) in Belgium.....	141
Figure 79. C7 prototype results (humid season above, dry below) in Ivory Coast.....	142
Figure 80. C7 prototype results (summer above, winter below) in Sweden.....	143
Figure 81. Commissions examples on all sites.....	145
Figure 82. Commissions linked to cloud shadow in Ivory Coast (red points).....	146
Figure 83. Functional architecture of the classification process (Source: Broceliande Manual - User Guide for Broceliande - François Merciol – Jérôme Moré).....	155
Figure 84. Temporal synthesis of built-up area classification. Panel shows monthly binary built-up/non-built-up masks for tile 31UFS from January to June 2021 (first three rows, left to right). The integrated mask (bottom left), derived from the logical aggregation of all six-monthly masks, is compared against ESRI basemap imagery (bottom right). The final composite mask is more accurate because the aggregation method reduces inconsistencies. For instance, it excludes the agricultural field (red box) that was falsely classified in May and June 2021 at the bottom....	156
Figure 85. Example from tile 31UFS showing the combined binary mask (left) and the result after filtering (right).....	157
Figure 86. Scatterplots for the 3 test sites for the Imperviousness degree product.	160
Figure 87. Example of omission errors.....	165
Figure 88. Improvement brought by the 5m super resolution. Left is the HRL IMD 2018 with a spatial resolution of 10 meters and right the C8 prototype with a spatial resolution of 5 meters for the reference year 2020.....	167
Figure 89. General workflow of C9 Prototype.....	172
Figure 90. 31TFJ Tile (Sentinel-2 at level 3A) - July 2020.....	173
Figure 91. 31TFJ Tile (Sentinel-2 at level 3A) - July 2021.....	173
Figure 92. Resulting image of the PCA on the difference between the two images above.....	173
Figure 93. Histogram of the heatmap (Figure 92 - Resulting image of the PCA on the difference between the two images above) of tile 31TFJ (method 1 thresholding (dotted blue and dotted red) - method 2 thresholding (solid blue & dotted green)).....	175
Figure 94. Example of improved geometry detected by the chain - red: initial geometry, green: improved geometry – OSM background map.....	176
Figure 95. Unmasked clouds and changes detected (in red) – source: Sentinel-2 at level 3A – tile 32UMU – 15 th July 2020 (1) and 15 th July of 2021 (2).....	182
Figure 96. Parking area and changes detected (in red) – source: Sentinel-2 at level 3A – tile 31TFJ – 15 th July 2020 (1) and 15 th July of 2021 (2).....	182

Figure 97. Parking area on the same location and year that Figure 96 – source: Google Earth.	182
Figure 98. Greenhouse's reflectance effects and changes detected (in red) – source: Sentinel-2 at level 3A – tile 31TFJ – 15 th July 2020 (1) and 15 th July of 2021 (2).	183
Figure 99. Rural commission and changes detected (in red) – source: Sentinel-2 at level 3A – tile 32UMU – 15 th July 2020 (1) and 15 th July of 2021 (2).	183
Figure 100. Example of annotations generated over a patch located in a cropland area. The upper row shows the Copernicus Sentinel-2 data for each month and the bottom row shows the associated annotations.	191
Figure 101. Example of LSC annotations over the validation dataset.	192
Figure 102. Example of LSC maps over tile 34TES (Romania, Hungary). Each column represents a month (Jan – Dec) and each row represents a year (2020-2022). The corresponding Sentinel-2 imagery is shown together with the maps. The time series of LSC maps show fluctuations between herbaceous and bare cover over the cropland area, while the tree class remains stable over time.	194
Figure 103. Example of seasonal changes in snow and ice cover over tile 33VWJ (Sweden).	194
Figure 104. Example of seasonal confusion between the tree and shrub class within tile 30SUH (Spain).	195
Figure 105. Overview of the openEO pipeline.	199
Figure 106. Example of forest management type classification in Côte d'Ivoire. (A) annual median Sentinel-2 composite, (B) annual median Sentinel-1 composite, and (C) forest management classification.	200
Figure 107. Forest management layer processed in openEO (left) and locally processed forest management layer (right) over an area of 1024x1024 pixels (10240 by 10240 m). About 0.068% of the pixels are different.	200

Applicable documents

ID	Applicable document
[AD1]	Grant Agreement
[AD2]	D1.2 Service requirements report – final version
[AD3]	D1.4 EO and in-situ data requirements report – final version
[AD4]	D1.6 Infrastructure requirements report – final version
[AD5]	D2.11 Training dataset for AI purposes – final version
[AD6]	D2.15 EvoLand infrastructure - final version
[AD7]	D2.17 Methods Compendium – final version
[AD8]	D2.19 Test Datasets – final version
[AD9]	D2.21 Source Code – final version
[AD10]	D2.4 Novel EO Datasets – final version
[AD11]	D2.9 Existing and novel in-situ and training datasets – final version
[AD12]	D3.4 EvoLand prototypes source code – final version
[AD13]	D3.6 EvoLand prototypes datasets – final version
[AD14]	D3.8 Prototype Validation Plan – final version

Executive Summary

The document “D3.2 EvoLand Prototypes Report - final version” describes all the prototypes developed in the Horizon Europe project EvoLand constituting Work Package (WP) 3. The EvoLand project aims at developing prototypes for candidate services based on satellite image time series data, targeting the evolution of the Copernicus Land Monitoring Service (CLMS) portfolio for the years 2027 and onwards. Therefore, recent algorithms in the field of Machine Learning and Deep Learning are tested, as well as novel Earth Observation (EO) data and infrastructure applied, to continuously monitor the status and dynamics of the Earth at high spatial and temporal resolution. In total twelve prototypes thematically placed in the fields of forest, agriculture, water, urban and general land cover are core outcomes of the project.

This report presents each prototype in terms of Introduction & Objectives covering also the specifications of each prototype, followed by the State of Play including the connection to the CLMS portfolio and the requirements collected in WP1 (Requirements Analysis). Thereafter the demonstration sites, main input and ancillary data are described, as well as the workflow including the methods transferred from WP2 (Methods Development). The results and verification section is followed by conclusions and an outlook beyond the EvoLand project. In addition to this report and delivered at the same time, two further deliverables constitute the source code of the developed prototypes (delivered on [EvoLand Github](#), including a supporting document “D3.4 EvoLand prototypes source code - final version”) and the datasets of the prototype’s outcomes (delivered on the S3 EvoLand bucket, including a supporting document “D3.6 EvoLand prototypes datasets - final version”).

Prior to the prototype development a requirement analysis was undertaken, where users and key stakeholders were interviewed related to their needs with respect to different user domains and policy areas (e.g., the European Green Deal, LULUCF, Ecosystems and Biodiversity, Forest Management, Agriculture and Farm to Farm, Soils, Urban Monitoring, Water Management and Others) [AD2]. The outcomes of the requirement analysis were considered for the prototype developments where possible. In addition, seven different tasks of state-of-the-art method development have been undertaken and tested based on smaller test sites (i.e., with the size of 1-2 Sentinel tiles or sub-tiles) and have been assessed regarding their potential to be transferred to the prototype development [AD7] to be then rolled-out onto larger demonstration sites. The prototype conceptualisation phase had started in parallel with these method developments, but technical developments were mainly initialized after the method handover (i.e., at M25, corresponding to end of February 2025) and were finalized until M30 (i.e., end of September 2025).

Phase 1 of the prototyping part of the EvoLand project has concluded with implementation of all the eleven foreseen prototypes. Further improvements have been implemented in Phase 2 in order to work towards demonstrating the potential to be rolled-out for pan-European implementation. The target of

demonstration in minimum three prototype sites was already achieved by three prototypes within Phase 1. After Phase 2, all prototypes have achieved this target, except two, whereof one was only introduced in Phase 2 as a reaction to evolving user needs. The target of covering three reference years was achieved by four prototypes in Phase 1 and at the end of the project all prototypes fulfil the target, except the one introduced in Phase 2 and one further prototype, affected by the lack of adequate in-situ data.

Main topical conclusions across most prototypes, individually presented in the respective chapters, are that in many cases EO data needs would be suitable to fulfil the requirements for a future pan-European processing, however, more efficient pre-processing lines are needed to tackle data volumes from dense time series of several reference years. Improved cloud-screening is one major issue to be tackled. The majority of prototypes have some limitations due to the lack of (multitemporal or up-to-date) reference data.

Following the prototype implementation an evaluation of the potential for operational roll-out for each candidate prototype, both for prototypes that are new to the CLMS portfolio as well as for prototypes targeting the improvement of existing CLMS products was conducted. A benchmarking along different criteria, split into three overarching groups of science & standards, impact & relevance and operational feasibility, carried out as part of WP4 (Operationality Assessment), shows the fitness as well as possible limitations for operational roll-out. This report therefore directly provides input to “D4.6 Report on candidates for operational roll-out”.

1 Introduction

1.1 About EvoLand

The Evolution of the Copernicus Land Monitoring Service (CLMS) portfolio integrating novel Earth Observation data and latest Machine Learning algorithms to continuously monitor the status, dynamics and biomass of land surface (EvoLand) project (Project no: 101082130) [AD1] develops innovative methods, algorithms and candidate CLMS prototypes monitoring the status and changes of land use/land cover and various land surface characteristics at high spatial and temporal resolution. Under the new Copernicus 2.0 programme, CLMS will continue to provide products and services for the 2021-2027 period, thus building time series that enable for long term change monitoring. Additionally, harmonisation and further integration of the products across components is envisaged as well as a more efficient workflows by further automation, modularity of the software and data products, and the increased use of Artificial Intelligence (AI) to provide real and near real time data processing, to respond to existing and emerging European policy needs.

In short EvoLand will:

- provide tangible proof of the evolution potential of the CLMS in terms of improved information content, quality and timeliness;
- produce up to eleven operationally benchmarked and qualified CLMS candidates monitoring the status and changes of land use/land cover and various land surface characteristics at high resolution;
- develop new innovative methods and algorithms (data fusion, continuous monitoring, AI, biomass mapping, novel infrastructure, EO and in-situ data); and
- analyse policy, data and infrastructure requirements for the prototype services.

1.2 Purpose and Objectives of WP 3 Prototype Demonstration

The goal of WP3 on “Prototype Demonstration” is the implementation of twelve prototypes for future improved and new CLMS candidate products in the demonstration sites. This Work Package is shaped by the “Requirements Analysis” carried out in WP1 and builds up on the WP2 on “Method Development”, and is strongly linked to WP4 on “Operationality Assessment”. WP3 encompasses the following Tasks (T) and prototype demonstrators (short-named with C1 – C12):

- **T3.1 Forest:** containing the prototypes Continuous Forest Monitoring (C1), Forest Disturbances Mapping (C2), Biomass Mapping (C3) and Tree Types Mapping (C12)

- **T3.2 Agriculture:** containing the prototypes Cover Crop Type (C4), Grassland and Cropland GPP (C5), and Small Landscape Features (C6)
- **T3.3 Water:** Improved Water Bodies (C7)
- **T3.4 Urban:** Continuous Imperviousness Monitoring (C8) and Automated Land Use Mapping of Urban Dynamics (C9)
- **T3.5 General Land Cover:** Continuous Mapping of Land Surface Characteristics (C10) and on-Demand Land Cover Mapping (C11)

From the **EvoLand Specific Objectives**, depicted in the Grant Agreement [AD1], two of them are particularly relevant for WP3, i.e. the prototyping phase: (i) “Implement algorithms and processing chains for eleven (+1) next-generation Copernicus Land Monitoring Service candidates focusing on the status and changes of land use/land cover characteristics across various thematic domains. These new services will be demonstrated and validated over several prototype sites, both in and outside Europe”; and (ii) “Provide efficient, scalable and open-source candidate service processing lines based on generic algorithm development and a modular architecture, demonstrated in cloud environments”. Whereas the first of these two objectives is the core of WP3 development and output, the second one refers to efficient processing lines to generate those candidate services.

The prototypes are demonstrated in at least three **demonstration sites**, and for three required **reference years (i.e. 2020–2022)**. There are only three exceptions, which are the C3 Prototype which was rolled out in four demonstration sites, but with regards to the quality only two were finally considered, secondly the C4 prototype due to a lack of in-situ data and third the C12 prototype, which was introduced at a later project stage and therefore was only rolled out for one AOI (which was nevertheless extended from 10 to 25 Sentinel-2 tiles). The outcomes of this WP are reports, open-source code and datasets:

- This document, “D3.2 EvoLand prototypes report - final version” **describes the prototypes in detail**, consisting of information on the objectives, methodology, results, validation and conclusions, as well as the way ahead beyond the EvoLand project.
- D3.4 consists of **source codes** available in different [GitHub repositories](#) and an accompanying document “D3.4 EvoLand prototypes source code - final version” [AD12].
- D3.6 consists of the resulting **datasets** (provided as cloud-optimised GeoTIFFs), accompanied by INSPIRE-conform **metadata**, both uploaded onto the **EvoLand S3 CREO bucket**, and an additional accompanying document “D3.6 EvoLand prototypes datasets - final version” [AD13].
- The document “D3.8 Prototype Validation Plan - final version” [AD14] describes the validation approaches that were chosen for the prototypes.

1.3 Link to WP1 Requirements Analysis

WP1 collects stakeholder and key end-user requirements for an enhanced CLMS portfolio in the Copernicus 2.0 period to 2027 and beyond, capturing key requirements how to respond to current and upcoming EU policy reporting objectives whilst building on latest technological developments. This is performed through an analysis of existing studies and continuous screening of latest political & technological developments, complemented by meetings and interviews with key stakeholders (i.e., the EEA and the JRC) and end users, as well as questions and answers by stakeholders and users in response to dedicated EvoLand webinars [AD2]. Moreover, requirements related to future EO and in-situ data needs [AD3] as well as processing infrastructure [AD4] were gathered. These requirements were consolidated to clearly define the prototype topical and technical specifications, input and ancillary data, as well as infrastructure and operational requirements for the prototypes and envisaged future CLMS products. The requirements gathered in [AD2] are addressed for each prototype in the sub-section on the State of Play.

1.4 Link to WP2 Method Development

In WP2, algorithms and methods required for the candidate CLMS prototypes were developed, and hence this WP is naturally closely interlinked with WP3 on prototype demonstration. The method development was done by using existing and novel EO data, in-situ data and infrastructure on smaller test sites being located within the larger prototype sites.

One part of WP2 reviewed and compiled existing and upcoming EO datasets as well as all in-situ datasets for model development, calibration and validation, which provided a relevant basis for all prototypes. Additionally, a novel, modular, scalable and open-source Information Technology (IT) infrastructure was established, which has been prototypically used for some of the prototypes, see the main WP2 deliverable “Methods Compendium” [AD7] and accompanying [AD5] and [AD6].

The second part of WP2 developed new algorithms for T2.4 weakly supervised learning (exploiting different single-date and time-series based Deep Learning models), T2.5 improved spatial and temporal resolution, T2.6 continuous monitoring (i.e. time series-based change detection and continuous land cover mapping) and T2.7 biomass mapping (i.e., above-ground woody biomass and tree canopy height mapping). These methods and benchmarking results are described in the main WP2 deliverable “Methods Compendium” [AD7].

Whereas constitutes the reports and results to describe the methods [AD7], the source code of the testing is delivered as well [AD8], alongside the testing datasets [AD9].

1.5 Link to WP4 Operationality Assessment

The prototype results are directly feeding into WP4 with respect to their operationality assessment. First, they have been presented to the main stakeholders (i.e., the European Environment Agency (EEA) and the Joint Research Center (JRC)). The stakeholder interaction and feedback on the prototype presentation is described as part of Task T4.1, with the deliverable “D4.3 Stakeholder consultation report – final version”.

Task T4.2 had its main focus on assessing the prototype candidates related to their potential for operational roll-out. A benchmarking along different criteria, split into three overarching groups of science & standards, impact & relevance and operational feasibility is carried out. The aim is to identify the prototypes which are most urgently and topically needed, technologically and scientifically most advanced, provide most significant added-value to the CLMS portfolio, and are most cost-efficient, etc., to qualify for being suggested as most promising candidates for future implementation into the CLMS operational product portfolio from 2027 onwards.

Further information on methods transferred to prototypes is provided in section 2.4.

1.6 Document Structure

This deliverable is the **final version** (D3.2) of the **Prototype Demonstration Report** for the EvoLand project. Table 1 presents the structure of the document, highlighting the different prototype implementation themes and methods tested in the project.

Table 1. Document structure.

Theme	Prototypes (Phase 1)	Chapter
Forest	C1 - Continuous Forest Monitoring	3.1
	C2 - Forest Disturbances Mapping	3.2
	C3 - Biomass Mapping	3.3
	C12 – Tree Types Mapping	3.4
Agriculture	C4 - Cover Crop Type Mapping	4.1
	C5 - Cropland/Grassland GPP Monitoring	4.2
	C6 - Small Landscape Features	4.3
Water	C7 - Improved Water Bodies Mapping	5
Urban	C8 - Continuous Imperviousness Monitoring	6.1
	C9 - Automated Land Use Mapping of Urban Dynamics	6.2
General Land Cover	C10 - Continuous Mapping of Land Surface Characteristics	7.1
	C11 - On-demand Land Cover Mapping	7.2

The structure within each prototype chapter is the following:

1. Introduction & Objectives
2. State of Play (referring to the CLMS portfolio, the requirements collected as part of WP1.1, and the innovation potential)
3. Selection of Prototypes sites
4. Input and Ancillary Data & Pre-processing
5. Methodology and Workflow, connecting to the methods developed in WP2
6. Results and Verification
7. Conclusions
8. Outlook beyond the EvoLand project, connecting to WP4 on the assessment for the operational potential of the prototypes

2 Prototype Overview

The following chapters provide overarching information on the prototype development, which starts with keywords, definitions and terminologies applied in the EvoLand project's method development and prototyping (Section 2.1). An overview of the demonstration sites and timescale (reference years) is provided (section 2.2), followed by a list and short description of the main methodologies transferred from WP2 to the WP3, i.e., the prototype development (section 2.3). To conclude this chapter, the code and data deliveries of the prototypes are explained (section 2.4). The 12 prototypes are presented in detail in the chapters 3–7.

2.1 Keywords and Definitions

The following table presents an overview of frequently used definitions, keywords and terminology within the EvoLand project, which is vital for correct understanding of the described contents and hence included here:

Table 2. Keywords and their definition in the EvoLand project.

keyword	Definition in EvoLand
Agents of disturbance	Agents are events/drivers that cause forest disturbance. There are fast agents such as wildfire or storms, or agents with slow onset such as bark beetle infestation or drought. Moreover, there are human induced clearings (i.e., to clear up the forest after a forest disturbance), or clear cuts (forest management)
Bio-geographic regions	Areas characterized by specific combinations of biological and geographical features.
Change	Any temporary and/or permanent changes on LC, being visible on the ground and/or detectable by EO methods. Specification in the frame of forest monitoring: Change in the context of Forest candidates means no seasonal, phenological change (since this is not in the interest of users) but structural/spectral change. i.e. loss of healthy tree cover / complete loss of whole trees (clear cuts)
Continuous monitoring	Ongoing observation and assessment of changes or developments over time.
Continuous monitoring/ Continuous mapping	Provision of consecutive and regular updates of the status of a certain land cover comprising collected information of a certain reference period (providing higher frequency than that in current services where it is meaningful and feasible) Update frequency varies for the different candidates due to different data sources and thematic characteristics
Disturbance	Specific understanding in the frame of forest monitoring: Tree cover canopy disturbance is seen as specific part of change: a descriptive analysis of spectrally observable and verifiable parameters that clearly depict a negative change (=loss) in the tree cover canopy caused by certain drivers. Disturbances include a quantitative aspect: • Loss of tree cover per area and time And disturbances include a qualitative aspect: • in assigning agents/drivers to changed areas • in assigning severity to changed areas

Fine-tuning	The action of marginally updating the parameters of a pre-trained Machine Learning model with the goal of using it for a different task or with different that from the setting on which it was previously trained.
Forest	Forest in this context means any “tree covered area”; independent from any national/subnational forest definition
Forest Canopy Height	Understood as “tree cover canopy height” instead of forest canopy height to align with C1 + C2 where the definition for forest mapping is “tree cover” (not “forest”). This also allows for applying a forest definition and thus offers flexibility to the user.
Embeddings	(Data) embeddings are representations of data (pixels, images, time-series) in a vector space that is different from the original one in at least one dimension (spatial, spectral, temporal). These representations are usually obtained by Machine Learning to fulfil one or several properties (compression, class separability, noise reduction, etc.).
Hyperspectral imaging	Technology capable of capturing and analysing a wide range of wavelengths in the electromagnetic spectrum.
In-situ data	Data collected directly from the location or environment being studied, as opposed to remotely sensed data.
Land surface categories (LSC)	Categories of direct observable or objectively measured surface properties, e.g. presence of water or photosynthetically active vegetation, applicable to each individual observation (pixel). Categorisation can be based on a scale-independent, discrete, hierarchical, and dichotomous scheme of classes based on directly and instantaneously observable bio-geo-physical properties of the planetary surface
Latent space / Latent variable	In a Machine Learning or statistical context, a latent space is a representation space for variables that are not observed or that are internal to the system. In Deep Learning in general and in AutoEncoders in particular, the output of internal layers that may be used for downstream tasks are examples of latent variables.
Near-Real Time (NRT)	NRT in the context of forest disturbances = fastest possible product generation on newest analysable data, i.e., the maximum feasible time interval for data reception, processing and product generation on single scene (or temporal composite, e.g. monthly, bi-weekly) Example: In the case of Sentinel-2: data theoretically every 5 days; realistic implementation in this project (considering cloud cover & processing) = monthly (max. 15 days). For forest: winter months might have to be treated differently (snow cover, leaf-off season)
Pretext (learning) task	A learning task for a Machine Learning algorithm whose objective function is easy to compute (e.g. no need for labelled data) but that allows pre-training the model to be reused or fine-tuned for other tasks.
Scalable infrastructure	Infrastructure capable of adapting and expanding to accommodate changing needs or increased demands.
Weakly Supervised Learning	Machine learning approach that requires less annotated data compared to traditional methods.

2.2 Demonstration Sites

The prototypes are implemented in eight representative demonstration sites, mainly located in the EEA-38 territory (six in continental Europe and one in French Guiana) plus one in Africa (Côte d'Ivoire), covering various bio-geographical regions and landscapes. Whereas the testing part of WP2 was carried out on smaller sub-sites located within the prototype sites, the full extent of the prototype sites is displayed in Figure 1. An exception has been made for the C6 prototype, which needs very high resolution (VHR) data as input, which are not available for a full prototype site and the C12 prototype which was rolled out on an extended version of the Central European site, covering 25 instead of ten S-2 tiles.

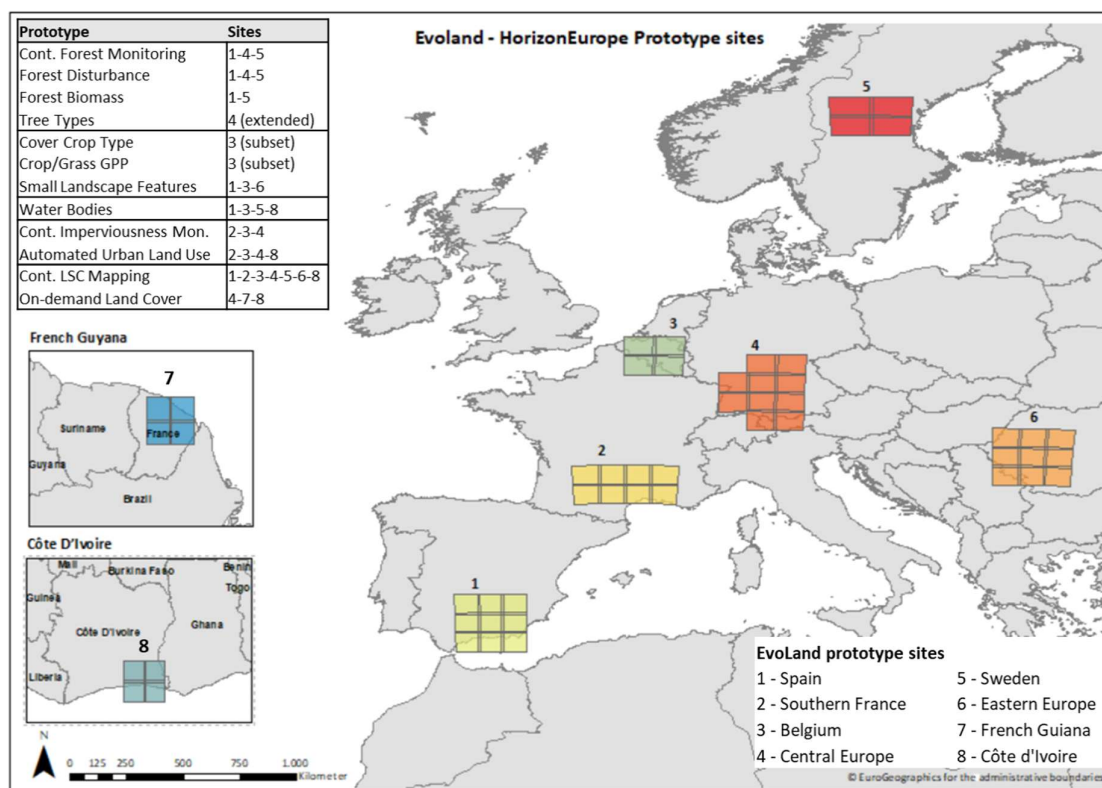


Figure 1. Demonstration sites for the EvoLand project.

Table 3 gives an overview on which sites have been chosen for which prototype within project Phase 1 and 2:

Table 3. Prototype sites and time frame covered by reference years, by prototype, for Phase 1 and Phase 2 of the EvoLand project.

PROTOTYPES		SITES and TIME FRAME			
		Phase 1		Phase 2	
No.	Prototype name/s (Phase 1)	Prototypes sites	time frame covered for	Prototypes sites planned	time frame planned for

			prototype output years		prototype output years
C1	Continuous Forest Monitoring	Spain (1), Central (4)	2019-2021 (Central); 2020-2022 (Spain);	Spain (1), Central (4), Sweden (5), testing only: French Guyana (7)	2019-2021 (Central); 2020-2022 (Spain, Sweden)
C2	Forest Disturbances Mapping	Spain (1), Central (4)	2019-2021 (Central); 2020-2022 (Spain);	Spain (1), Central (4), Sweden (5)	2019-2021 (Central); 2020-2022 (Spain, Sweden)
C3	Biomass Mapping	Spain (1), Sweden (5), testing only: French Guyana (7)	2021	Sweden (5), testing only: Central (4), French Guyana (7)	2020-2022
C4	Cover Crop Type Mapping	Belgium (3) – only 1 tile	2021-2022	Belgium (3) subset	2018-2021
C5	Cropland/Grassland GPP Monitoring	Belgium (3) – only 1 tile	2018 (dry year), 2021 (normal/wet year)	Belgium (3) subset	2017-present
C6	Small Landscape Features	Spain (1), Belgium (3), Eastern Europe (6) - (smaller subset of tiles)	2022	Spain (1), Belgium (3), Eastern Europe (6)	2021-2022
C7	Improved Water Bodies Mapping	Spain (1), Belgium (3), Sweden (5), Ivory coast (8)	2022 (2 dates per year -dry/humid season)	Spain (1), Belgium (3), Sweden (5), Ivory coast (8)	2020-2022
C8	Continuous Imperviousness Monitoring	Southern France (2), Belgium (3), Central Europe (4)	2020-2022 (every 6 months)	Southern France (2), Belgium (3), Central Europe (4)	2020-2022
C9	Automated Land Use Mapping of Urban Dynamics	Southern France (2), Belgium (3), Central Europe (4)	2020-2021	Southern France (2), Belgium (3), Central Europe (4), Ivory coast (8)	2020-2021 + 2021-2022
C10	Continuous mapping of Land Surface Categories	Belgium (3), Sweden (5)	2020 - 2022	Spain (1), Southern France (2), Belgium (3), Central Europe (4), Sweden (5), Eastern Europe (6), Cote d'Ivoire (8)	2020-2022
C11	On-demand Land Cover Mapping	Belgium (3) – only one tile	Single scene	Central Europe (4), French Guyana (7), Ivory coast (8)	Single scene
C12	Tree Types Mapping	N/A	N/A	Central Europe (4) extended to 25 tiles	2021

2.3 Developed methods transferred to prototypes

Table 4 provides an overview of the methods developed in WP2 that have been transferred to prototype developments as part of WP3 in both phases of the project. The methods are described in detail in the “Methods Compendium - final version” [AD7].

Table 4. Methods developed in EvoLand that have been transferred to prototype implementation.

PROTOTYPES		METHODS	
Nr.	Prototype name/s (Phase 1)	WP2 Task methods used (Phase 1)	WP2 Task methods used (Phase 2)
C1	Continuous Forest Monitoring	T2.6 Time Series Modelling	T2.4 Weakly supervised learning; T2.6 Time Series Modelling;
C2	Forest Disturbances Mapping	T2.6 Time Series Modelling	T2.6 Time Series Modelling; T2.6 Forest Disturbance Classification
C3	Biomass Mapping	T 2.7 Biomass mapping	T 2.7 Biomass mapping
C4	Cover Crop Type Mapping	T2.3 Novel infrastructure	T2.2: Novel In-Situ and Training Data; T2.4 Weakly supervised learning
C5	Cropland/Grassland GPP Monitoring	T2.2 Novel In-Situ and Training Data; T2.3 Novel Infrastructure and Data Access; T2.5 Improved resolution; T2.7 Biomass mapping	T2.1 Novel EO Datasets; T2.7 Biomass mapping
C6	Small Landscape Features	T2.1 Novel EO Datasets; T2.5 Improved spatial resolution	T2.1 Novel EO Datasets; T2.5 Improved spatial resolution
C7	Improved Water Bodies Mapping	T2.5 Improved spatial resolution	T2.5 Improved spatial resolution
C8	Continuous Imperviousness Monitoring	T2.5 – Improved spatial resolution	T2.5 Improved spatial resolution
C9	Automated Land Use Mapping of Urban Dynamics	T2.5 Improved spatial resolution	T2.5 Improved spatial resolution
C10	Continuous mapping of Land Surface Categories	T2.6.3. Continuous mapping of LSC; T2.3 Novel Infrastructure and Data Access	T2.6.3. Continuous mapping of LSC; T2.2 Novel In-situ data/training data; T2.5 Improved spatial; temporal; spectral resolution
C11	On-demand Land Cover Mapping	T2.3 Novel Infrastructure and Data Access	T2.2 Novel In-Situ and Training Data; T2.4 Weakly supervised learning
C12	Tree Types	N/A	N/A

Related to the **C1 Continuous Forest Monitoring** prototype, methods from Tasks T2.4 on Weakly Supervise Learning and T2.6 on Continuous Monitoring were tested and benchmarked 16[AD7]. Within CLMS, forest change detection is a key part of the High-Resolution Layer (HRL) Vegetated Land Cover Characteristics (VLCC) and the pan-tropical forest component of the Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring Service (LCFM). Therefore, methods from T2.6 continuous time series modelling (e.g. BFAST cusum, BFAST mosum, EWMA, Kalman filter/CuSum, BEAST, etc.) were benchmarked and transferred to the C1 forest prototype on “Continuous Forest Monitoring” in project Phase 1 [AD7].; these

methods were further refined in Phase 2, tested towards their application in a tropical site, and compared with methods based on SAR-time series. Regarding weakly supervised learning methods (T2.4), for the forest case the SSL4ForestChange has been tested in Phase 1 and 2, however, was rated not the most promising approach for a prototypic implementation of a continuous forest change detection product, hence was not taken over to the prototyping activities. Testing for the Multi-Modal DataCube (MMDC) approach, AliSE and PROSAIL as input to forest prototypes has been initiated, and was evaluated in WP2 Phase 2 for suitability to be taken up by the prototypes, but were also not taken over to the prototyping.

For **C2 Forest Disturbance Mapping**, developments are undertaken as part of T2.6 on Continuous Monitoring, with a sub-topic of 'Forest Disturbance Classification'. The customized algorithm has been applied for forest disturbance classification in Phase 1, and was further improved in Phase 2 of the project in terms of input features and regional calibration.

C3 Biomass Mapping developments were taken over from the woody biomass and tree canopy height components of the Task T2.7 on Biomass Mapping [AD7]. In Phase 2 of WP2 further testing was done to increase the spatial resolution of Sentinel-2 input bands, as well as to apply Deep Learning models. A combination of parametric and non-parametric models at uncertainty levels was tested, as well as the inclusion of a tropical demonstration site. Methods for above-ground woody biomass were intended to be improved including L-band SAR data of the upcoming NASA-ISRO SAR (NISAR) Mission, however, due to a delayed launch the data was not available within the EvoLand testing and prototyping phases.

C4 Cover Crop Type Mapping saw its main innovation in implementing the workflow in openEO under T2.3 Novel infrastructure. In phase two, T2.2: Novel In-Situ and Training Data and T2.4 Weakly supervised learning will be assessed. We expect that C5 will make undergo significant improvement during the second development phase.

For **C5 Cropland/Grassland Gross Primary Productivity (GPP) Monitoring** the main method supporting this prototype is T2.7 Biomass mapping. In addition, T2.2 Novel In-Situ and Training Data where Integrated Carbon Observation System (ICOS) data were tested, T2.3 Novel Infrastructure and Data Access with the migration of the workflow to openEO and T2.5 Improved resolution where Sentinel-2 data was implemented also contributed to C4. In the second phase, T2.1 Novel EO Datasets will be further explored and T2.7 Biomass mapping will continue to be developed.

For **C6 Small Landscape Features**, the main developments came from the EO data under T2.1 Novel EO Datasets and T2.5 Improved spatial resolution. Being in early development in phase 1, the phase 2 provided the opportunity to further investigate the EO novel datasets under T2.1 Novel EO Datasets and T2.5 Improved spatial resolution to improve detection and classification of new features.

Concerning prototype **C7 Improved Water Bodies Mapping**, the outputs from task 2.5 (Improved Spatial Resolution) were explored during Phase 1. It showed very promising results to get 5m resolved water bodies but imply high calculation time that might prevent it in an operational environment. For Phase 2, the use of super resolution algorithm from T2.5 in an operational environment will be explored, focusing on the automatic determination of the AOI in which the algorithm would be deployed.

C8 Continuous Imperviousness Monitoring saw its main innovation in the implementation of super-resolution under T2.5 Improved resolution. For phase 2, the super resolution module was applied on Sentinel-2 L3A instead of Sentinel-2 L2A in phase 1.

For **C9 Mapping Urban Dynamics**, the chain and method are new. In phase 1, we tested the use of T2.5 Spatial Super resolution, but in the end, the type of input images (Sentinel-2 L3A) was not suitable, so we did not use it in the final chain. For phase 2, finally T2.5 Spatial Super resolution was used due to a Super-Resolution module development for Sentinel-2 L3A.

C10 Continuous mapping of Land Surface Categories (LSC) is a completely novel method. Its main development took place under 2.6.3. Continuous mapping of LSC and was implemented in openEO under T2.3 Novel Infrastructure and Data Access. The T2.6.3. Continuous mapping of LSC will continue and drive further developments in phase two, while T2.2 Novel In-situ data/training data and 2.5 Improved spatial, temporal, spectral resolution will be assessed.

C11 On-demand Land Cover Mapping is a new method without any pre-existing operational services. T2.3 Novel Infrastructure and Data Access was fundamental for its development as the novel prototype was entirely developed on openEO. In the second phase T2.2 Novel In-Situ and Training Data and T2.4 Weakly supervised learning methods will be tested and will lead to further improvement of the prototype.

The **C12 Tree Types Mapping** prototype is a special case, as it was only introduced at a later project phase, i.e. in Phase 2, as a reaction to evolving stakeholder requirements. Therefore, there is no direct link to the methods development.

2.4 Code and Data Deliverables

This document, “D3.2 EvoLand prototypes report - final version” is jointly delivered **with open source code** on the **EvoLand Github** (<https://github.com/EvoLand-Land-Monitoring-Evolution>) where a repository has been set up for each prototype’s source code. The details on the provided code are described as part of a “Readme” file on Github, as well as in an accompanying deliverable document “D3.4 EvoLand prototypes source code - final version” [AD12]. The **datasets** that the prototypes consist of are delivered on the **EvoLand S3 EvoLand bucket** in the form of cloud-optimised GeoTIFF files, accompanied with STAC metadata. All prototypes

are made available on a specific [Results Portal](#)¹. An additional accompanying document “D3.6 EvoLand prototypes datasets - final version”.

¹ Results Portal will be publicly accessible from November 2025.

3 Forest Prototypes

The Task T3.1 Forest encompasses the development of four prototypes: Continuous Forest Monitoring (C1), Forest Disturbances Mapping (C2) and Biomass Mapping (C3) were initiated in Phase 1, and are each described further in the three following sub-sections 3.1, 3.2 and 3.3, respectively. In addition, a Tree Types Mapping (C12) prototype (3.4) has been introduced in Phase 2 of the project in coordination with the stakeholders, as a reaction to their evolving needs and requirements. These four prototypes target an evolution of the CLMS portfolio in the forest domain, in particular the HRL VLCC, but as well the pan-tropical forest monitoring part of the LCFM and the HRL Vegetation Phenology and Productivity (VPP).

3.1 C1 – Continuous Forest Monitoring

3.1.1 Introduction & Objectives

The candidate prototype C1 - Continuous Forest Monitoring aims at extending the portfolio and scope of the Forestry products, mainly the current HRL Vegetated Land Cover Characteristics (VLCC) Forest products, by improving the update frequency of the forest mask in order to depict the change dynamics within tree covered areas and also to serve as an input to the forest disturbance agent classifier (C2). This information can be used to improve forest management and planning, support forest-related decision-making processes, or reporting on forest-relevant targets in EU policies. Moreover, EvoLand developments and prototypes may also serve to improve the CLMS tropical forest monitoring products, currently produced in the framework of the Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring service (LCFM).

Table 5. Technical and thematic specifications for the C1 prototype.

Prototype C1	
Prototype	C1 – Continuous Forest Monitoring
Thematic specifications	Time of detection of tree cover canopy disturbance
Prototype sites	1 – Spain 4 – Central Europe 5 – Sweden 7 – French Guiana (testing only)
Temporal frequency	Monthly
Spatial resolution	10m
MMU	Pixel-based
Envisaged accuracy	not defined

3.1.2 State of Play

Whereas the current HRL VLCC provides yearly status layers of tree cover at 10m spatial resolution, and 3-yearly change products at 20m spatial resolution, in

EvoLand, the prototype C1 is focussing on a continuous monthly tree cover change detection, depicting natural forest disturbances as well as human-made tree removal. Products are provided on pixel level (10m spatial resolution), and in monthly intervals. There are multiple well tested methodologies deriving timely high frequency forest disturbances which are also used for near-real-time (NRT) forest monitoring in the tropics, but until now there is no operational pan-European product showing forest dynamics in such a high temporal frequency. The operational potential of C1 is described as high, since such a product on a monthly update basis does currently not exist in the CLMS portfolio, and it can fit in line with a future evolution of the forest portfolio.

The requirements gathered as part of WP1.1 contained that the product allows users to apply different forest definitions. This is facilitated by working on pixel-based tree-cover level, hence users can apply their own Minimum Mapping Unit (MMU) and threshold values, tailored to their particular forest definition. The requirement of a product across all of Europe at a 10m resolution is covered by the specifications and is prototypically applied in the demonstration sites. The annual update of the product has been requested, which is exceeded by the C1 prototype since updates are provided at a higher cadence, i.e., monthly. The requirement of a product being available with a maximum latency of 1 year is targeted in WP4 operability assessment and is defined as “a product for a given reference year should be produced within the following year”.

3.1.3 Selection of Prototypes sites

In Phase 1 of the project two prototype sites were analysed: Central Europe (covering an area of 10 Sentinel-2 (S-2) tiles) and Spain (9 S-2 tiles). In Phase 2, those two demonstration sites were reproduced and additionally the demonstration site in Sweden (4 S-2 tiles) was processed and a test was conducted in the French Guiana site (4 S-2 tiles), as well. The temporal coverage of the demonstration sites differs for Central Europe (training with 2016-2018 and roll out for 2019-2021) and the three remaining sites, Spain, Sweden and French Guiana (training with 2017-2019 and roll out for 2020-2022).

3.1.4 Input and Ancillary Data & Pre-processing

The main input data originates from S-2 (optical) time series, while VLCC forest masks serve as ancillary input data.

Reference data were generated from multiple data sources: Multitemporal optical VHR imagery (e.g. VHR IMAGE2021, GoogleEarth timelines, BingMaps) were used for interpreting samples for the verification of detected forest change/loss, supported by S-2 time series, in a multitemporal manner along the timeline. Further existing products were used in addition, e.g. the Remote sensing based National Detection System for Forest Damages (FNEWSs) for Germany containing reference data for forest disturbances (e.g. bark beetle and wind throw), or the European Forest Fire Information System (EFFIS) containing near real-time and historical information on forest fires.

One of the ancillary data which were also compared to the EWMA results stems from the Tree Cover Disturbance Monitoring (TCDM). The TCDM-radar tool is an iterative unsupervised change detection approach based on Sentinel-1 radar data at high spatial (10m) and temporal resolution (bi-weekly since 2018) (Langner et Carboni, 2021; Langner et al., in preparation).

Furthermore, the Global Forest Change (Hansen et. al, 2013) and the European Forest Disturbance Atlas (EFDA) (Viana-Soto & Senf, 2024), together with the TCDM were used as input for the stratification of the disturbance stratum for the validation sampling.

A list of data sets is shown in Table 6 below:

Table 6. List of datasets used for the C1 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	The full time series of S-2 L2A data (with a max. cloud cover of 85%) was used for the calculation of monthly median features, for the period from 2016–2022	10 m	Revisit frequency of 5 days, compressed into monthly composites
HRL VLCC DLT	HRL VLCC Dominant leaf type layer for the derivation of a tree cover mask	10m	yearly
TCDM	Tree Cover Disturbance Monitoring	10m	Bi-weekly
FNEWS	Remote sensing based National Detection System for Forest Damages (FNEWSs) for Germany	N/A	Reference years 2020, 2021
EFFIS	European Forest Fire Information System (EFFIS) containing near real-time and historical information on forest fires	N/A	dynamic
VHR imagery optical	Several sources of VHR optical imagery (e.g. VHR IMAGE2021, GoogleEarth timelines, BingMaps)	~ 0.5m-4m	Irregularly spaced
Global Forest Change	Global forest changes including the year of change	30m	yearly
EFDA	European Forest Disturbance mapping providing several layers, including the “latest disturbance”, flagging areas of change including the respective year of the change	30m	yearly

3.1.5 Methodology and Workflow

3.1.5.1 EWMA processing

After a benchmarking of several unsupervised methods to detect forest changes and disturbances from a time series, performed as part of the WP2 (Method development), the ‘Exponentially Weighted Moving Average’ (EWMA) method has been rated the most promising. Initially published by Brooks et al. (2014) for Landsat time series data, it was tailored not only to detect major changes in a forest (such as clear cuts) but also for more subtle or minor forest degradation and

thinning. In EvoLand this method is tested towards detecting forest disturbances (i.e., windthrow, forest fire, insect infestations, but also clear cut and thinning occurrences) over the selected EvoLand prototypes sites and was therefore adjusted to S-2 time series. As actual input for the EWMA, monthly median tasselled cap wetness features were calculated from the cloud masked S-2 time series.

In EvoLand Phase 1 the cloud masking was utilizing the sen2cor Sentinel-2 Scene Classification Layer (SCL). The low quality of the cloud masks derived from the SCL turned out to be an issue, due to the high sensitivity of the approach towards remaining cloud artefacts. This resulted in large commissions of detected tree cover disturbances in the Phase 1 results, which raised the need for improved cloud masking. In Phase 2 the cloud masks were therefore derived by utilizing a pre-trained deep learning model (Aybar, 2024), which was trained on the CloudSEN12 dataset. This dataset contains ~50.000 image patches, including S-2 level1C and level-2A data, together with annotations for thick and thin clouds and cloud shadows, manually assigned both by experts and based on results of state-of-the-art cloud detection algorithms (Aybar et al., 2022).

The training of the model for the tree cover disturbance mapping was performed on a 3-year time-span of history (i.e. fitting a deterministic harmonic time series model considering seasonal components). The actual analysis period covers the three years following the training period, where, after eliminating seasonal fluctuations, residuals are derived in relation to the historical model and treated as normally distributed.

Continuous monitoring can be performed with this method, since it allows the computation of change maps for each satellite scene (updated from previous maps, after training of historical period). Hence, the analysis period could range from a single new incoming satellite scene (i.e. NRT) until up to a full year or several years of analysis. In the case of EvoLand the training was performed on the years 2016–2018 for the Central European demonstration site (2017–2019 for Spain, Sweden & French Guiana) and the analysis period was set from 2019–2021 (2020–2022 for Spain, Sweden & French Guiana).

3.1.5.2 TCDM processing

Input data to the TCDM-radar approach are the Sentinel-1 (S1) Ground Range Detected (GRD) scenes in vertical, vertical (VV) and vertical, horizontal (VH) polarization. The scenes are processed using the Sentinel-1 Toolbox to generate a calibrated, ortho-corrected product. TCDM-radar can cover the period January 2018 until present with a lag time of up to half a month, and due to the cloud-penetrating characteristics of the radar data it is able to deliver indications of potential disturbance events continuously every two weeks. The following paragraph summarizes the major processing steps of the TCDM-radar approach:

First, a delta analysis step is performed between the time features (median of VV and VH bands) of a 2-month analysis and an earlier 2-month reference window,

which are one year apart in order to account for seasonality. These delta results therefore integrate over all changes that occurred during that year and are combined to maximise the preliminary common delta result. After that step the analysis and reference windows are shifted by 2 weeks, followed by another preliminary delta analysis over the new time windows. In this way the whole analysis period is processed, providing intermediate level-1 change results. A temporal filtering step analyses all consecutive Level-1 change results. Only after a consecutive confirmation per pixel-location a potential disturbance signal is kept. After these intermediate results over the full analysis period, a second delta analysis step (Level 2) is performed over the subsequent Level-1 change results, delivering Level-2 change results showing only the activity data that occurred during single two-week periods covering the whole analysis period. Descending and ascending orbits are handled separately and combined afterwards. Eventually, a disturbance density related filtering step analyses the spatial proximity of disturbance detections, allowing to separate potential disturbance values that are situated in close vicinity from random noise, which is missing any specific spatial patterns.

3.1.6 Results and Verification

3.1.6.1 Results

The expected results are a detection of a disturbance in the tree cover canopy, which might be characterised by a subtle or strong signal, on a monthly basis. Every pixel can only be detected once during the analysis period; hence the starting date of a forest disturbance is derived. Figure 2 depicts an example of the C1 prototype results for an area in Germany, affected by a bark beetle infestation. To prevent a further spread of the pest, large areas of spruce forest have been cleared completely, starting in 2019. The period between the years 2019 and 2021 was monitored with the EWMA approach for the Central European Site and shows the progressing tree cover loss very well with the monthly-resolved time of detection of tree cover disturbances.

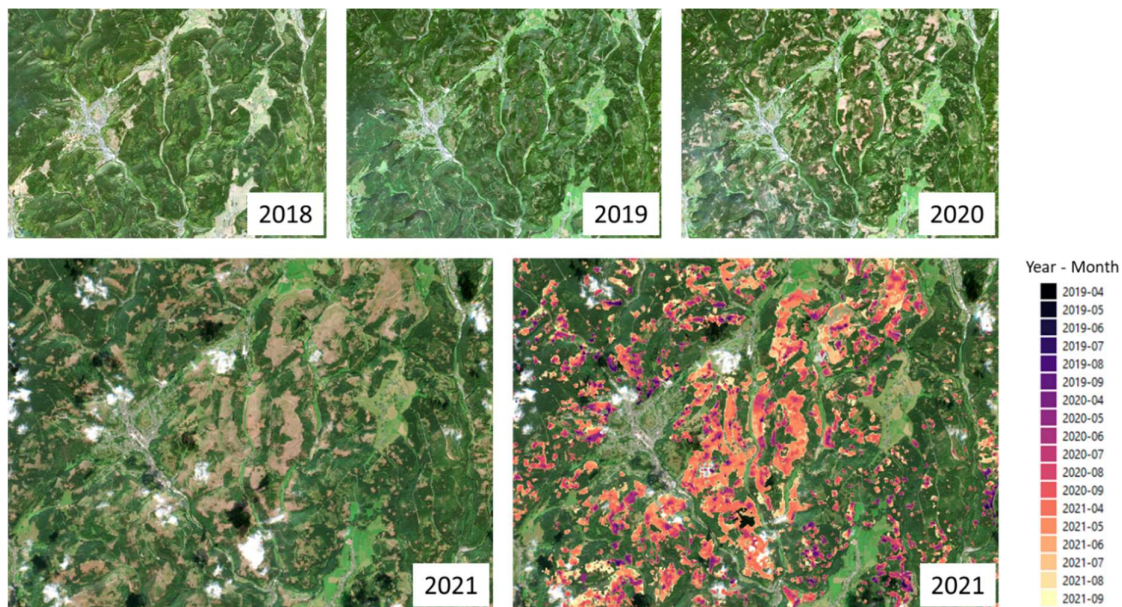


Figure 2. Overview of a forest in the Central European demonstration site (near Steinach, Germany) affected by a bark beetle infestation, being cleared step by step to avoid a further spread of the pest, beginning in summer 2019. Monitoring period 2019-2021 (Tile 32UPA, base imagery: S-2).

Figure 3 depicts a detailed view on a region of bark beetle infestation, located centrally in Figure 2. On the left and right side of the result the tree cover density (TCD) from the HRL VLCC is displayed, depicting the situation before (2018) and after the event (2021). It can be noted that the 'disturbed' area as indicated by the EWMA approach corresponds to the tree cover loss when comparing the 2018 and 2021 Tree Cover Density results, as the affected areas have been cleared.

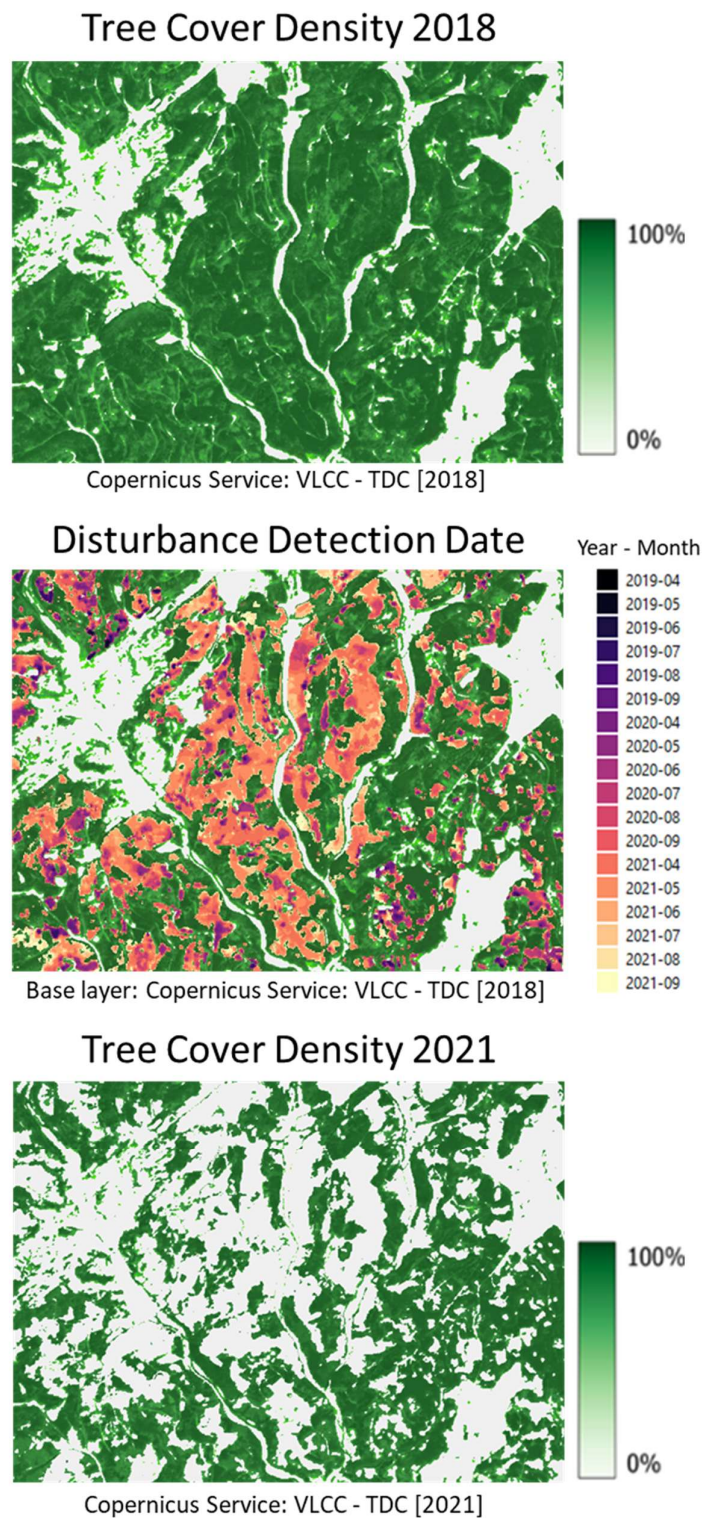


Figure 3. Middle: Forest clearing because of a bark beetle infestation starting in the summer 2019. Monitoring period 2019-2021 (Germany, tile 32UPA). Top & bottom: Tree cover density (TCD) from the HRL Vegetated Land Cover Characteristics (VLCC), depicting the situation before (2018) and after the event (2021).

Another example (Figure 4) shows a wind throw in summer 2019 in the Central demonstration site (tile 32UNA). The damage from the storm was detected around the time of August – September 2019.

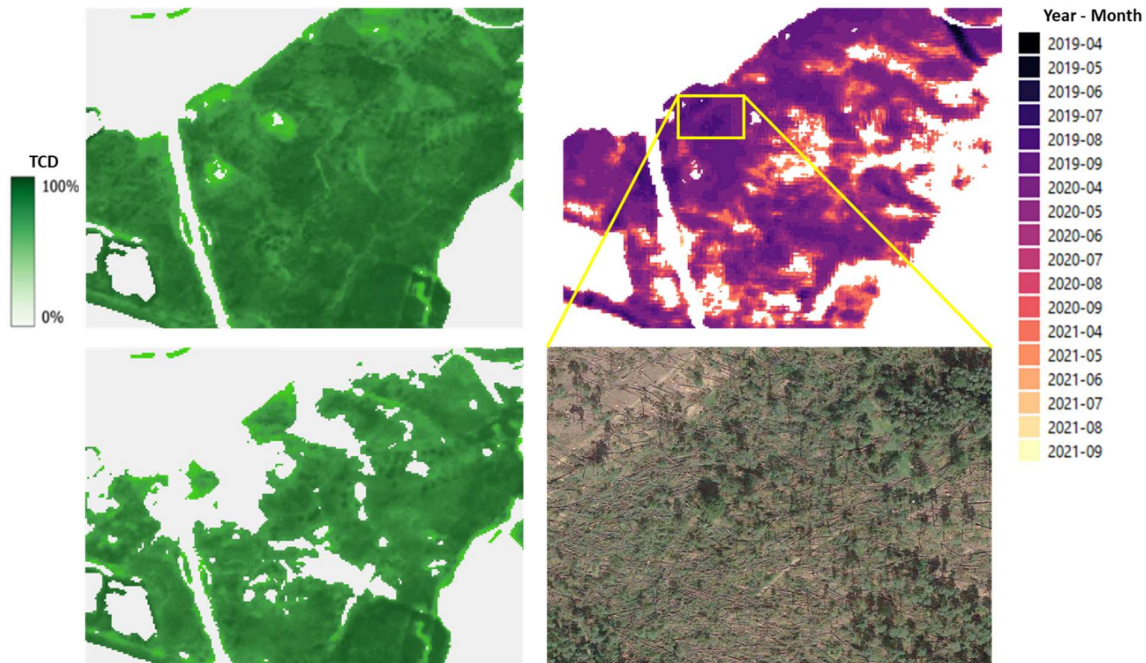


Figure 4. Forest affected by a storm event in 2019. Left: Tree cover density (TCD) from the HRL Vegetated Land Cover Characteristics (VLCC), depicting the situation before (2018, upper left) and after the event (2020, lower left). Upper right: C1 for the monitoring period 2019-2021 (Central site, tile 32UNA). Lower right: Google Earth imagery from 08/23/2019.

As mentioned in section 3.1.3, a test to transfer the methodology to tropical regions was conducted in the EvoLand demonstration site in French Guiana. This test revealed that a detection approach solely based on optical input data reaches its limits in the tropics, due to the very high cloud cover throughout the year. Even the improved cloud masking did not prevent massive false detections, which prevented the derivation of a meaningful forest disturbance prototype in this demonstration site. Figure 5 shows the EWMA result for a subset of the French Guiana demonstration site with some correctly detected disturbances (marked with yellow boxes), surrounded by large patches of false detections, characterized by mostly the same detection date for large areas.

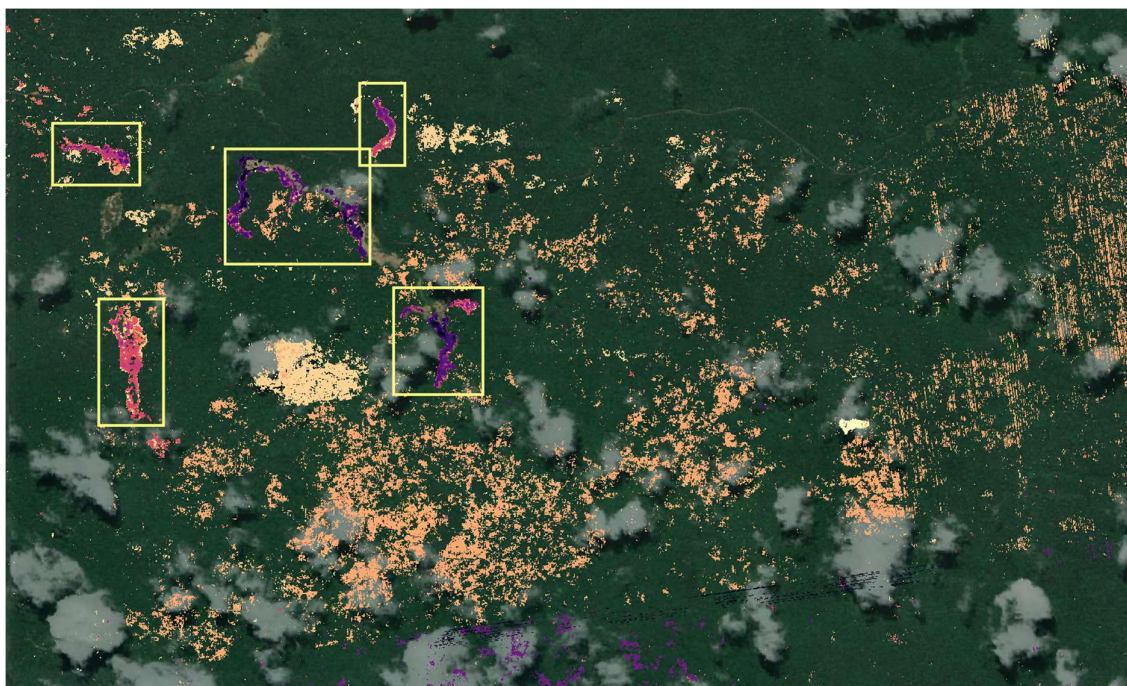


Figure 5. C1 result for an AOI in French Guiana, showing correct detections of tree cover disturbances (yellow boxes), surrounded by massive false detections.

3.1.6.2 Validation and Analysis

In Phase 1 an internal quality control of the optical-based EWMA and the radar-based TCDM-radar results regarding their ability to detect deforestation (loss of tree coverage beyond MMU of forest definition) as well as forest degradation (loss of tree cover less than MMU of forest definition as well as loss of chlorophyll content due to insect infestation) was undertaken. Back then the reference data had been collected manually for areas that were known to have suffered from disturbances and the layers were evaluated with regards to their ability to correctly detect those disturbance events.

In phase two the quality control on the final results has been conducted with a sound validation approach, based on systematic sampling and a stratification, resulting in weighted accuracies, better representing the actual quality of the outcomes. A detailed description of the validation approach is provided in [\[AD14\]](#), nevertheless there were some small differences implemented in the finally applied approach. First of all, 1000 sample points have been created with a systematic grid with regular spacing and were interpreted by experts. In this course the interpreters evaluated, if a tree cover disturbance had happened at the specific sample location and if so, when it happened and what was the cause of the disturbance (needed to evaluate C2). Thereafter, it was analysed, if there are enough samples for disturbed areas to allow for a statistically sound analysis (for that, at least 30 samples for disturbed areas were expected). As this was not the case in all demonstration sites, a subsequent stratification has been applied, based on the EWMA result and further ancillary layers, to create a disturbance stratum

with a 100m buffer (to account for potential omission errors likely appearing near disturbed areas), as described in [\[AD14\]](#) (section 3.3.2.2.1). The layers used for the stratification were the EWMA result, Global Forest Change, European Forest Disturbance Atlas, TCDM and a forest mask derived from HRL VLCC. So, in difference to the described approach, LUCA data, as well as the TCPC (HRL Tree Cover Presence Change) have not been included. The derived disturbance stratum has then been sampled, again with a systematic grid with regular spacing, and those additional sample locations have been interpreted, collecting the same information as done for the initial samples. After this additional stratified sampling, the number of samples covering actual disturbances was sufficient in all demonstration sites. For the Central European site, the years 2019-2021 and for the Spanish and the Swedish site, the years 2020-2022 were analysed.

In the accuracy analysis, the samples outside of the applied forest masks have been disregarded, as the C1 prototypes are considering only those areas for the disturbance detection. The resulting total numbers of samples used in the evaluation were:

- **653** out of 1348 interpreted in the Central European site
- **731** out of 1303 interpreted in the Spanish site
- **775** out of 1064 interpreted in the Swedish site

The sample locations per demonstration site are shown in Figure 6.



Figure 6. Location of the sample points for the Central European (top), the Spanish (lower left) and the Swedish demonstration site (lower right).

In this validation, the User and Producer Accuracy as well as the Overall Accuracy (OA) with 95% Confidence Intervals were calculated. To account for the stratified creation of the additional samples, area weights were calculated for those samples, while the initial samples were integrated without the application of area weights (i.e. weighting factor = 1).

The following tables show the results of the validation for the EvoLand demonstration sites in Central Europe (Table 7), Spain (Table 8) and Sweden (Table 9), where:

- OA: Overall Accuracy
- PA: Producer Accuracy
- OE: Omission Error
- UA: User Accuracy
- CE: Commission Error
- CI 95%: 95% Confidence Interval

Table 7. Confusion matrix of C1 for the Central European demonstration site.

C1 Central European site, 2019-2021						
C	Reference					
	undisturbed	disturbed	Total			
	undisturbed	516.74	13.62	530.36	UA:	CI 95%: CE:
	disturbed	4.08	11.32	15.40	73.5%	22.0% 26.5%
	Total	520.82	24.94	545.76		
		PA:	45.4%	OA:		96.8%
		CI 95%:	19.5%	CI 95%:		1.5%
		OE:	54.6%			

Table 8. Confusion matrix of C1 for the Spanish demonstration site.

C1 Spain, 2020-2022						
C	Reference					
	undisturbed	disturbed	Total			
	undisturbed	524.52	7.63	532.15	UA:	CI 95%: CE:
	disturbed	11.00	8.73	19.73	44.3%	21.9% 55.7%
	Total	535.52	16.37	551.88		
		PA:	53.4%	OA:		96.6%
		CI 95%:	24.2%	CI 95%:		1.5%
		OE:	46.6%			

Table 9. Confusion matrix of C1 for the Swedish demonstration site.

C1 Sweden, 2020-2022						
C	Reference					
	undisturbed	disturbed	Total			
undisturbed	597.67	10.36	608.03	UA:	CI 95%: CE:	

disturbed	2.79	8.83	11.62	76.0%	24.6%	24.0%
Total	600.46	19.19	619.65			
	PA:	46.0%		OA:	97.9%	
	CI 95%:	22.3%		CI 95%:	1.1%	
	OE:	54.0%				

The validation results show that the User Accuracy (UA) for the disturbed class in the Central European demonstration site is quite good with 73.5%. Evaluating the UA for the Phase 1 results in the Central site with the same validation points reveals a clear improvement in Phase 2. While the UA was only 43%, it could be increased by 30.5%. We can attribute this increase in UA to the improved cloud masking as described in 3.1.5.1. While Phase 1 results were suffering from frequent commission errors triggered by the limited accuracy of SCL-based cloud delineation, this problem could be mitigated by the improved cloud masking utilizing CloudSEN12. A time series plot of the monthly Tasseled-Cap Wetness (TCW) data of Phase 1 is displayed for a specific point (yellow cross) in Figure 7, clearly indicating an outlier with a much higher TCW value compared to the observations before and after. The outlier occurred in May 2021 (see yellow and light green coloured patch in upper middle detail figure). An inspection of the input data showed that this outlier was caused by an omitted cloud, which was resulting in widespread disturbance detections with the same detection date (see orange coloured patches in the lower left close-up), where no disturbance could be verified. While such isolated outliers can be identified rather easily in historical analyses, systems aiming at continuous monitoring and near real-time mapping often struggle. As such omitted clouds were appearing frequently this had a significant impact on the Phase 1 results for the Central site causing the need for an improved cloud detection, which was implemented in the second project phase.

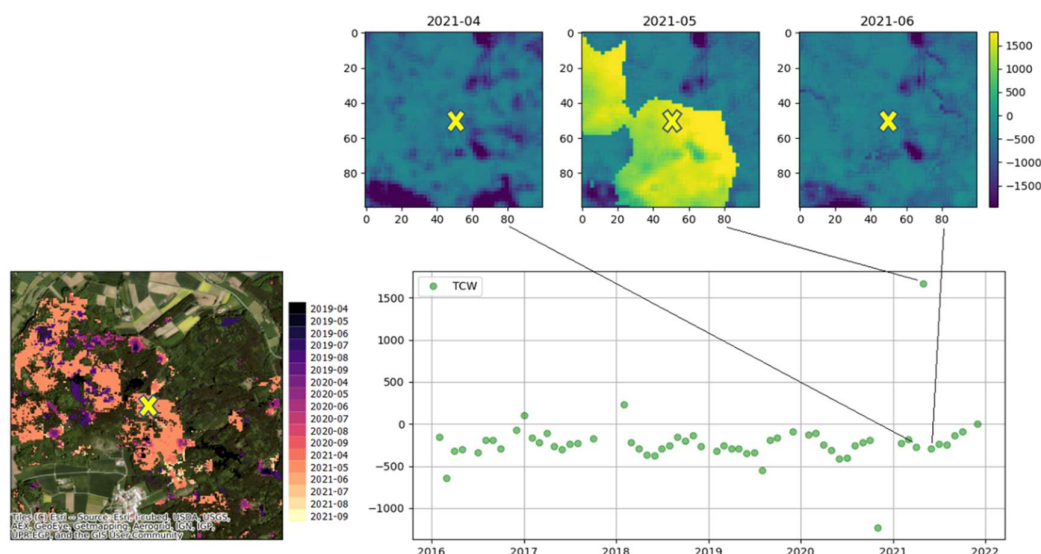


Figure 7. Time-series of monthly composites of Tasseled-Cap Wetness (TCW) data, analysed for point A of Figure 8 (Central European site).

The Producer Accuracy (PA) for C1 in the Central site on the other hand is rather low, which is due to two independent reasons.

First of all, the parameterization of the disturbance detection was done in a conservative way. While the methodology is theoretically very sensitive to slight signal changes, the sensitivity needed to be tuned down intentionally. Despite elaborate cloud masking and monthly median aggregation, the used input signal is rarely free of artefacts. Also, the used harmonic time series model is only an approximation of the true seasonal dynamics. Both signal artefacts and modelling errors result in non-random model residuals which can result in a large number of commissions in areas of actually healthy tree cover if the sensitivity is set too high.

Secondly, a certain delay in detecting disturbance events based on structural break points in the time series (i.e. unsupervised) is inherent to the method. A change is not necessarily detected with one single, but with repeated observations of a break in the time series. Lower sensitivity settings also favour larger detection delays. This is especially true for disturbances, which are not causing a sudden loss of tree cover (like e.g. fire events), but instead evolve over time, like for example bark beetle infestations. Figure 8 outlines different disturbance conditions and their influence on the time of detection using the implemented methodology. Examples A and B show how the intra-monthly timing of a disturbance might influence the detection delay. In example A, the disturbance event happens early in the month and several acquisitions show the changed signal. In example B, the change happens towards the end of the month and maybe only one or even no acquisition with the changed signal is input to the median feature, causing an increased detection delay. Furthermore, as mentioned above the intensity of the signal change varies for different types of disturbances and can be very subtle and only evolve over a longer period (example C). In all of the three examples (A, B & C), the disturbance event happens/starts in the same month, i.e. May 2019, but is detected in July (A), one month later in case B and another month later in case C. As the monitoring window per year is limited due to the necessity to exclude leaf-off season/winter months (depending on the region of interest), the time (i.e. month) of the year in which a disturbance happens has the largest potential for delaying the detection of the specific disturbance event. Example D in Figure 8 shows a disturbance event towards the end of July 2019 with only subtle representation in the S-2 signal, which prevents its detection still within 2019. As the monitoring period of the subsequent year only starts in April 2020, this would be the earliest point in time where this exemplary event could be mapped in C1, i.e. with a minimum delay of 9 months. Of course, extended periods of no usable ground observations due to persistent cloud and/or snow cover can also generally add to the detection delay.

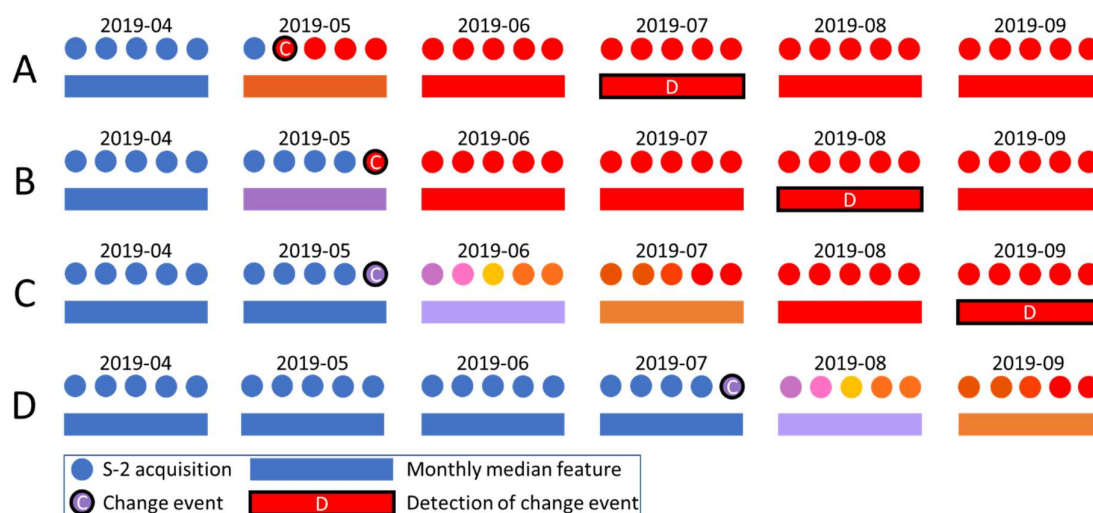


Figure 8. Theoretical examples visualizing the temporal offset for the detection of disturbance events occurring at different points in time during the monitoring window for a certain year (here April to September, like applied in the Central site) and with differing intensities of signal changes. Blue represents normal spectral behaviour, whereas the shift towards red represents signal changes with varying intensities.

The impact of the delayed detection on the PA gets tangible when removing the validation samples which are representing changes in the year 2021 (i.e. at the end of the multi-year monitoring period) from the evaluation. The result is shown in Table 10 where the PA has increased by 16.3%, reaching ~62%, while the UA is still above 70%.

Table 10. Confusion matrix of C1 for the Central European demonstration site, excluding validation samples with change observed in 2021.

EWMA 2019-2021 Phase 2						
Excluding validation points with change observed in 2021						
C	Reference					
		undisturbed	disturbed	Total		
	undisturbed	515,74	6,08	521,82	UA:	CI 95%: CE:
	disturbed	4,08	9,78	13,86	70,6%	24,0% 29,4%
	Total	519,82	15,86	535,68		
		PA:	61,7%			OA: 98,1%
		CI 95%:	23,9%			CI 95%: 1,2%
		OE:	38,3%			

Looking at the demonstration site in Spain, the results are quite different. The User accuracy for the Spanish demonstration site is significantly lower compared to the Central site, reaching only 44.3%. Nevertheless, an evaluation of the Phase 1 results with the same validation samples also shows an improvement here, as the UA was only 21.6%. The increase shows on the one hand again the positive impact of the improved cloud masking, significantly reducing commission errors, and on the other hand reflects a more conservative parameter setting. The results in Phase 1

already revealed, that the approach does not cope very well with the specific conditions in the demonstration site, being characterized by olive groves and Mediterranean forests. Especially the soil management in olive groves, with the removal of the understory, exposing the bare soil, is problematic as this is causing a significant break in the signal of the input data which is not related to tree cover disturbances. Furthermore, drought effects are having an influence on the received signal, leading to false detections of tree cover disturbances. Therefore, the parameterization was tuned towards the reduction of such commission errors, which has a negative effect on the PA and a trade-off between detection and overestimation of changes had to be made. Nevertheless, the result in the Spanish site still shows the highest PA of all three test sites, reaching 53.4%. Removing samples representing changes late in the monitoring period, i.e. from the second half of 2022, also shows a slight increase in the PA to 56.2%, while the UA remains stable. This is again pointing to the effect of the delayed detection, however with a lower magnitude compared to the central site.

Last but not least, the demonstration site in Sweden shows the highest Overall Accuracy (OA). It confirms the good UA, even exceeding the one in the Central European demonstration site, with 76%. The PA is also slightly higher with 46% in comparison (but still lower than for the Spanish site). When doing a similar evaluation as mentioned for the Central site, by excluding the validation samples with changes from the last year in the monitoring period, the same effect can be observed. While the UA still stays above 70%, the PA increases significantly to 54.6%. This confirms again the impact of the delayed disturbance detection.

A comparison of the average delay in the detection of the disturbance events in all demonstration sites (see Table 11) shows an increasing trend from south to north, which could point to an effect of generally higher cloud cover in the more northern areas. Further, it can be assumed that wildfires, being a frequent driver for tree cover disturbances in the Spanish site, are detected closer to the event, as the structural break in the signal is very significant in this case.

Table 11. Average detection delay in months by demonstration site.

Spain	Central Europe	Sweden
5,19	6,33	9,26

Figure 9 shows an example from the Spanish demonstration site, where several subsequent fire events happened in the same region. The C1 prototype nicely detected the different points in time, where the single disturbance events happened. Detected dates are very close the temporal information given by European Forest Fire Information System, shown with green delineations and dates in the figure, and even label the same months for large parts. This is an example for the timelier detection of tree cover disturbances with very significant signal changes in the S-2 data.

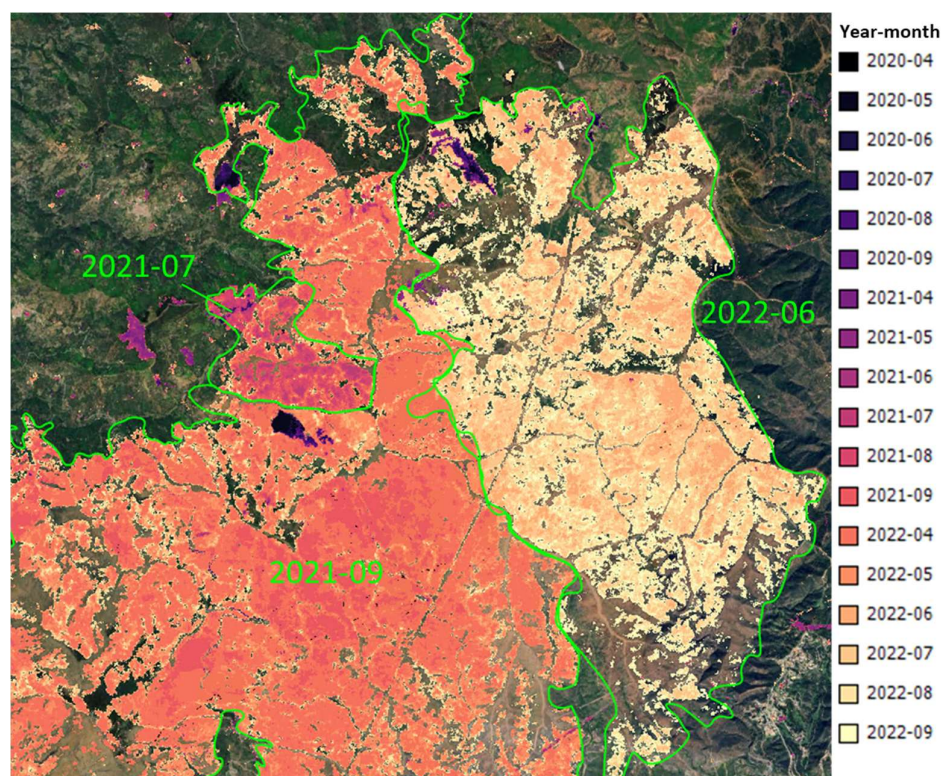


Figure 9. C1 for an area in the Spanish demonstration site, which suffered from several subsequent fire events. The green delineations show the extent and the recorded year and months of the single fire events, as provided by the European Forest Fire Information System (EFFIS).

The comparison with TCDM-radar data for the Central demonstration site (see Table 12) revealed a lower error of commission for the EWMA results. On the other hand, the PA and therefore the omission error are on the same level.

Table 12. Confusion matrix of the TCDM data for the Central European demonstration site.

TCDM Central European site, 2019-2021						
C	Reference					
	undisturbed	disturbed	Total			
	undisturbed	501,13	12,26	513,39	UA:	CI 95%: CE:
	disturbed	5,05	10,05	15,11	66,5%	23,8% 33,5%
	Total	506,18	22,31	528,49		
				PA:	45,1%	OA: 96,7%
				CI 95%:	20,6%	CI 95% 1,5%
				OE:	54,9%	

The rather low PA of TCDM is likely related to the fact that radar-based approaches can only detect changes if physical structures are altered. For example, the removal of tree cover leads to a change in the backscatter signal from diffuse (closed canopy cover) to specular of a clear-cut area and the delta between the signals before and after the disturbance event is used to monitor the event. In case no physical structure altered, no change can be detected. Most difficult to monitor are therefore degradation events that do not directly result in changes of the forest canopy coverage such as insect infestations (e.g. by bark beetles) as long as no salvage or clear-cut logging took place. Optical-based approaches such as EWMA on the other hand are sensitive towards changes in the spectral response that can be caused by clearing operations exposing non-photosynthetic vegetation/ soil or by insect infestations and fire events that lead to a drop of chlorophyll content.

The radar approach also can experience difficulties when detecting windthrows – depending on the severity of those events – as the radar sensor might normally be negatively impacted by the tangle of fallen trees that can simulate some kind of diffuse backscatter signal and therefore lead to higher omissions. The TCDM-radar approach will be published in a paper (in preparation) which shows for a similar test site (in that case whole Germany) higher accuracy figures (PA 74.2% and UA 92.0%). These figures, however, are not directly comparable because the paper deals with a much longer analysis period (2018-2024) and omissions become proportionally less over a longer time span. Furthermore, a shift of one pixel (10 meters) was allowed in the paper in order to account for radar displacement. The most important difference in the analysis of TCDM relates to above-described characteristic of radar sensors that cannot detect changes in the signal when there are no changes in the physical structures. Therefore, drops in the chlorophyll content due to red or grey attacks caused by bark beetles were not considered as degradation event in that paper, while they are included in the evaluation for this report as optical approaches are potentially able to detect them and the validation approach was designed for that.

3.1.7 Conclusions

In conclusion, the C1 prototype, applying the EWMA method, can capture subtle disturbances (of small spectral magnitude), but there is a trade-off between the sensitivity for the detection of tree cover disturbances on a small spectral magnitude and over-detection, leading to commission errors. Therefore, it is necessary to tune the parameter setting according to biogeographical preconditions on a regional basis, whereas it has to be noted, that a negative impact on the accuracies by some specific conditions cannot be compensated. This was found for example in Spain, where drought effects and management practices in olive groves contributed to a higher rate of commission errors, compared to the other two demonstration sites.

With regards to the near real-time potential of the approach it has to be considered that there is general delay in the detection of disturbance events, which can vary significantly for different regions and due to various reasons, as explained in section 3.1.6. At this stage, a monthly resolved yearly product is therefore the most reasonable outcome.

The EWMA approach is of unsupervised nature, hence it does not require annotated training data representing forest disturbances. Instead, large amounts of EO input data in terms of dense Sentinel-2 time series over a training period of a minimum of three full years are required before the analysis on a monthly scale can be initiated for the subsequent years. This is connected to extensive (but fully automated) pre-processing efforts for years of dense time series, including cloud masking that is recommended to be used instead of the standard cloud masks derived from the SCL. Nevertheless, the approach is robust as it is feasible to process large data amounts while still being computationally effective, and in addition it builds up on an open-source tool.

3.1.8 Outlook beyond the EvoLand project

The results of the second phase of the EvoLand project showed significant improvements with regards to the reduction of commission errors in the detection of tree cover disturbances in C1. Nevertheless, there is the need to further improve the methodology towards a timelier detection of the events. However, the observed detection delays have partly been a direct cause of the methodological choices made to increase the robustness of the approach with respect to false positive detections; these development goals are often contradicting. In phase 2, we could show that recent advances in cloud masking have a direct and significant effect on the user accuracies. Unfortunately, the improved cloud masks have been available only very late in the development phase. The potential of monthly versus single-scene modelling on an even higher temporal scale (e.g. bi-weekly or scene-based) is therefore still a future topic and will need additional investigation. The comparison and synergies with SAR-based NRT monitoring results (i.e., TCDM) will have to be further analysed in this regard. Difficulties of highly heterogenous

regions will need to be tackled when improving time series modelling tools towards pan-European roll-out.

3.2 C2 – Forest Disturbances

3.2.1 Introduction & Objectives

The candidate prototype C2-Forest Disturbance Mapping aims at deriving the cause of a forest disturbance event, also called driver or agent in this report. Forest disturbances are defined as areas with a negative change of vital tree cover or a significant loss of tree cover density. Disturbances are hence part of ‘tree cover change’ but focus only on loss or degradation and not on gain of tree cover. The following disturbance agents are defined in the EvoLand project: (i) insect infestation (focus on bark beetle), (ii) clear cut and thinning activities, (iii) healthy forest, events of (iv) subsequent human-induced forest clearing to prevent further damage or spread of pest (i.e. salvage logging), (v) wildfire and vi) wind throw/storm damage.. The specifications of the C2 prototype are depicted in Table 13:

Table 13. Technical and thematic specifications for the C2 prototype.

Prototype C2	
Prototype	C2 – Forest Disturbances
Thematic specifications	Forest disturbance agent classification: (i) insect infestation (e.g., bark beetle) (ii) clear cut and logging/thinning (iii) healthy (iv) salvage logging (v) wildfire (vi) wind throw/storm damage
Prototype sites	1 – Spain 4 – Central Europe 5 – Sweden
Temporal frequency	monthly
Spatial resolution	10m
MMU	Pixel-based
Envisaged accuracy	N/A

3.2.2 State of Play

The prototype C2 is focussing on a continuous probability estimation of the most likely disturbance cause, using the location and timing of the EvoLand C1 prototype as input, i.e. monthly forest disturbance detections. Products are provided on pixel level (10m spatial resolution). The innovation potential of the C2 prototype is high, since it is not yet part of the CLMS portfolio.

The requirements gathered in WP1.1 covered that the service should systematically produce an annual product highlighting change. Another requirement targets that the service should support an additional product at higher cadence for

assessing specific events. Both is possible, as the timing and location from C1 allow for the agent classification at the time of the disturbance on a monthly composite level, whereas those results can also be aggregated, showing the most likely event during the monitoring period. The availability of the product on a 10m spatial resolution scale across Europe is prototypically deployed in the demonstration sites. The requirement that the product should at a minimum identify changes due to: droughts, infestations, wildfires and storms is mostly covered – except drought as an agent, since training the classifier towards detecting drought-induced changes would require a much longer time series analysis and further ancillary data. The requirement that the product should distinguish between forest management and forest disturbances is partially covered by the fact that also human-induced clearings, clear cuts or selective logging events are captured, however, the information if that human-induced action has been conducted due to forest management cannot be derived from EO data without further reference data. The desire to have a product available worldwide at a resolution of 200-300m is currently not in the scope of the project.

3.2.3 Selection of Prototypes sites

In Phase 2 of the project three prototype sites were analysed: Central Europe (covering an area of 10 S-2 tiles), Spain (9 S-2 tiles) and Sweden (4 S-2 tiles), for which S-2 time series data within an analysis period 2019–2021 (Germany) and 2020–2022 (Spain & Sweden) were classified.

3.2.4 Input and Ancillary Data & Pre-processing

The main input data originates from S-2 (optical) time series, as well as areas and timing of detected forest disturbances in the EvoLand C1 prototype, derived with the EWMA method.

Likewise, as for the C1 prototype, reference data was generated from multiple data sources: Multitemporal optical VHR imagery (e.g. VHR_IMAGE_2021, GoogleEarth timelines, BingMaps) were used for evaluating samples for verification of detected forest disturbance agents, supported by Sentinel-2 time series, in a multitemporal manner along the timeline. Further existing products were used in addition, e.g., the Remote sensing based National Detection System for Forest Damages (FNEWSs) for Germany containing reference data for forest disturbances (e.g. bark beetle and wind throw), and the European Forest Fire Information System (EFFIS) containing near real-time and historical information on forest fires. A list of datasets is shown in Table 14 below:

Table 14. List of datasets used for the C2 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Sentinel-2 L2A use for the calculation of monthly mean features, for the full time series of data up to 85% cloud cover (2016–2022)	10 m	Monthly composites

TCDM	Tree Cover Disturbance Monitoring	10m	Bi-weekly
FNEWS	Remote sensing based National Detection System for Forest Damages (FNEWSs) for Germany	N/A	Reference years 2020, 2021
EFFIS	European Forest Fire Information System (EFFIS) containing near real-time and historical information on forest fires	N/A	dynamic
VHR imagery	optical Several sources of VHR optical imagery (e.g. VHR IMAGE2021, GoogleEarth timelines, BingMaps)	~ 0.5m	Irregularly spaced

3.2.5 Methodology and Workflow

The method for classifying forest disturbances is based on (i) the outputs of monthly detected tree cover changes/losses from the C1 prototype, and (ii) spectral features, derived from Sentinel-2, which are analysed by a LightGBM classifier. The result is a probability layer, providing probabilities between 0-100% for each of the defined disturbance agent classes on a pixel-level. The algorithm was chosen, as it performed best in a benchmarking process conducted in WP2 and described in the final version of the Methods Compendium [AD7]. Despite of the fact that reference data needs to be generated for different classes, it is an advantage that the complexity of the required reference data is relatively low. Only annotated single pixels distributed randomly within disturbed patches are needed, in comparison to other approaches that require to carefully delineate polygons or even fully annotate whole image chips.

Disturbances caused by different agents have different spectral signatures for each Sentinel-2 band, which sometimes also overlap, hence it makes it difficult to distinguish some classes on a per-scene basis (for example a wind throw or beginning bark beetle infestation, where in both cases the optical signal might still depict the green needles of a tree). Therefore, for an NRT product it was chosen to present probabilities for each class (delivered in addition in EvoLand, but for a delivery to potential users this would be recommended instead of a hard classification).

In general, disturbance events occur on different scales (both in time and space). Temporally, wind throw, wildfire or logging events are characterized as fast change events (however, clearing/logging events could also range over a longer time span), whereas an insect infestation is a slowly developing change. In terms of area, disturbances can have different extents from very local affecting just a few trees (e.g. start of a bark beetle attack), to large and widespread (e.g. advanced bark beetle infestation or extensive wind throw). It must also be noted that different drivers can jointly cause forest disturbances (e.g. trees infested with bark beetles are prone to be affected by wind throw in case of a storm; or drought-affected trees are more prone to wind throw or to bark beetle infestation than healthy ones). The above-mentioned approach to deliver probability instead of hard classification

results will in particular take care of such cases of combined drivers, as it provides the probabilities for each class.

3.2.6 Results and Verification

The expected result is a classification of the most likely forest disturbance agent for every pixel that has been detected as tree cover canopy change/loss by the C1 prototype (see section 3.1.6). Figure 10 shows an example for a forest affected by a bark beetle infestation starting in summer 2019 in Germany. It shows the disturbed areas, as delivered by the C1 prototype (upper right), the C2 agent classification (mid-left), as well as the probabilities for three selected classes, as an RGB composite. The class of bark beetle (in red) is dominant and could be verified by the use of ancillary data. The blue patch in the north, however, is correctly depicted as clear cut. It has been found that this area is not affected by bark beetles, but was cleared afterwards for probably another reason.

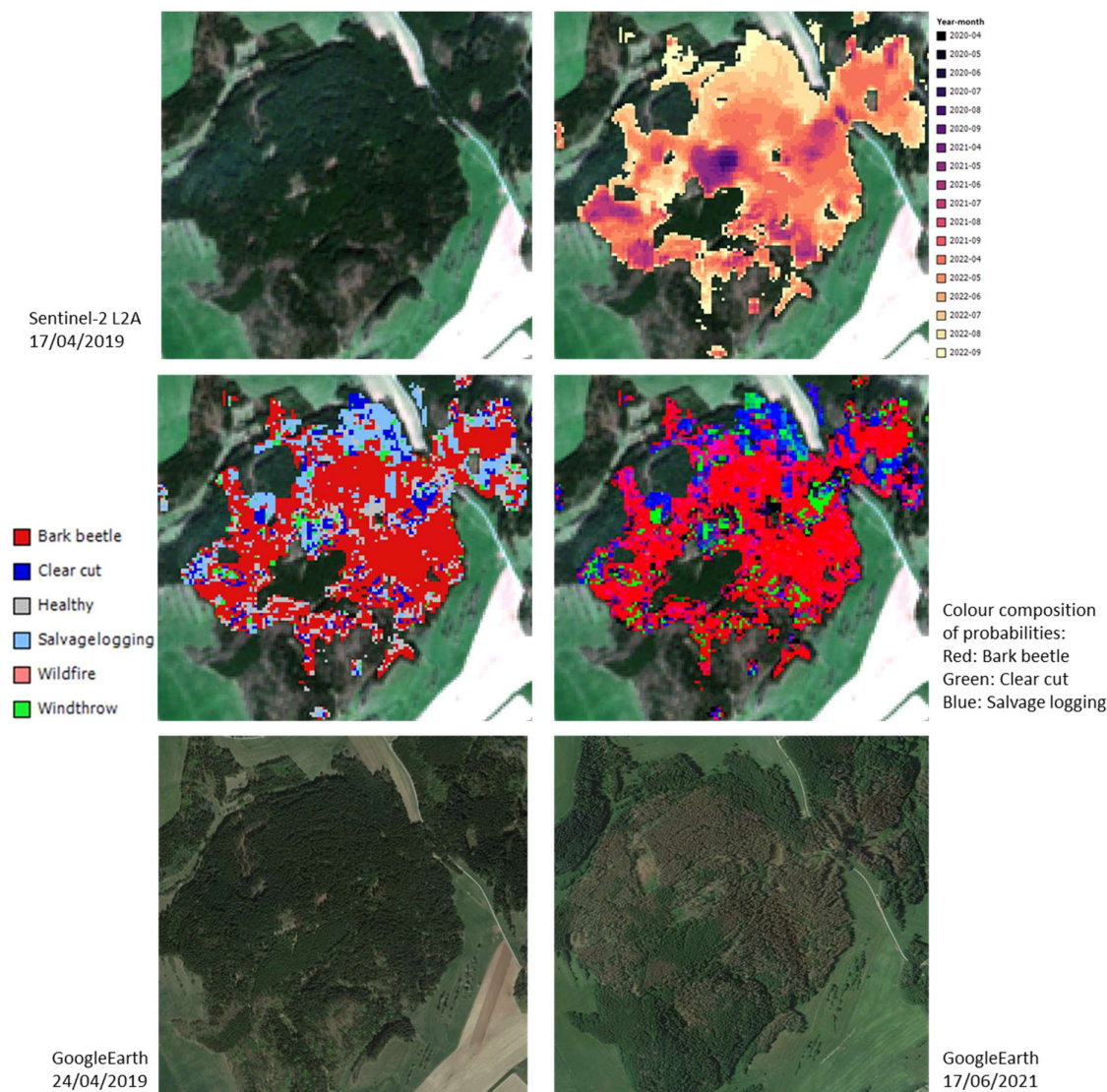


Figure 10. Top left: S-2 acquisition from 17/04/2019; Top right: C1 detecting a disturbance caused by a bark beetle infestation starting in summer 2019. Monitoring period 2019-2021 (Germany, tile 32UPA); Mid-left: C2 hard classification of disturbance agents; Mid-right: C2 class probabilities for the disturbance agent classes: bark beetle (red), clear cut (green), and salvage logging (blue); Lower-left & -right: Google Earth imagery before the pest and at a later stage, showing the progressing tree dieback (brownish colour).

Figure 11 shows the area, affected by large salvage logging activities (Central European site, S-2 tile 32UPA), which was already shown in section 3.1.6.1. The disturbance event started in late summer 2019, with the clearings reaching their peak in 2021, as detected with EWMA (upper right). The image on the lower images shows the classified disturbance agents (left) as well as their class probabilities (right). The class probability for salvage logging is displayed in the blue band, and is very dominant in this area, whereas the probability for bark beetle is displayed in the red band and only dominates a few smaller areas.

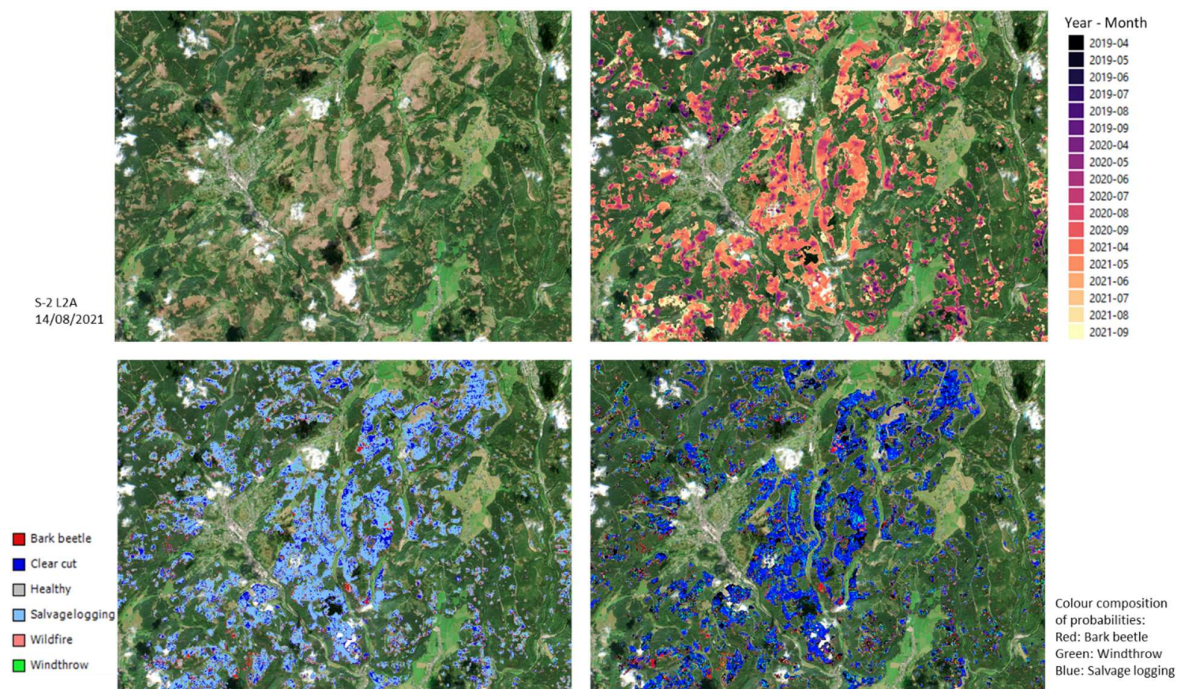


Figure 11. A subset of the Central demonstration site, showing large areas of salvage logging to prevent the spread of a bark beetle infestation. Upper left: S-2 acquisition after the logging activities; Upper right: Result of C1; Lower left: Disturbance agent classes; Lower right: Disturbance agent probabilities as RGB composite.

The ability of the LightGBM Classifier to correctly identify the underlying disturbance driver of the disturbance events detected with the C1 product was analysed in an internal quality control. For this purpose, the samples from Phase 1 which were manually selected samples for areas with known disturbances and drivers have been reused in combination with the samples from the C1 validation, verified as disturbance events. Those samples were used to do a very basic assessment of the ability of the approach to detect the correct driver. It has to be noted that this is a plausibility analysis only and not a statistically sound validation approach as done for C1. This is due to missing reference data which does not allow

for stratified validation sampling for disturbance agents. The collected reference data includes 219 sample points over the Central European and 213 sample points over the Spanish test site. For the site in Sweden only the samples from the C1 validation were available summing up to 39 verified changes, all being identified as clear cuts. An analysis of the plausibility of the different agents was therefore not possible for the Swedish demonstration site.

As there is no reference data for salvage logging, this class was aggregated with the clear-cut class for the analysis. Besides that, windthrow and bark beetle infestations were analysed for the years 2019-2021 for the Central European site. For the Spanish demonstration site clear cut/salvage logging and fire events were analysed for the years 2020-2022. The evaluation considers the hard (thematic) classification representing the class (agent) with the highest probability per pixel location.

For the Central European site, the C2 classification shows an overall accuracy (OA) of only 40.6% (see Table 15). In particular, the omission error for windthrow events (98%) and bark beetle infestations (88.9%) have a negative impact on the OA. Both classes show significant confusions with clear cuts, which could partially be due to timely clearing of affected areas obscuring the actual driver of the disturbance. Further, especially for the bark beetle class there are several samples that have not been detected as disturbed in C1 which might be due to the rather subtle change in the EO data signal. This can cause an omission of bark beetle events in the C1 prototype due to a certain detection delay, as described in section 3.1.6.2. As C2 focuses on the areas mapped as disturbed in C1, such omissions directly transfer into omissions in the disturbance classification as well. Clear cuts in comparison show the highest PA with slightly above 43% and the UA is on the same level with ~44%. The low UA for the clear-cut class is mostly due to the already mentioned confusions with the windthrow class, while the low PA is mostly caused by the omission of the events in the C1 prototype.

Table 15. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Central European demonstration site.

		Reference							
		None	Bark beetle	Clear cut	Windthrow	Total	UA:	CI 95%:	CE:
C2 Class	None	46	21	43	10	120	38.3%	8.7%	61.7%
	Bark beetle	0	4	4	1	9	44.4%	32.5%	55.6%
	Clear cut	0	11	38	37	86	44.2%	10.5%	55.8%
	Windthrow	0	0	3	1	4	25.0%	42.4%	75.0%
	Total	46	36	88	49	219			
PA:		100.0%	11.1%	43.2%	2.0%		OA: 40.6%		
CI 95%:		0.0%	10.3%	10.3%	4.0%		CI 95%: 6.5%		
OE:		0.0%	88.9%	56.8%	98.0%				

Table 16 shows the analysis for the Spanish demonstration site. Again, it has to be considered that this is not a stratified validation approach, but a plausibility analysis with selected samples of known disturbance events, as there was no adequate

reference data available to go beyond. For the Spanish site, the model in general performs better which might be related to the reduced number of only two observed drivers, i.e. clear cuts (including selective logging) and wildfires. The OA for C2 in the Spanish demonstration site is 62%. The analysis shows that the PA for wildfires is quite good with slightly above 65% and the UA even reaches 85.5%. This good representation of the wildfire class is on the one hand due to the clear break in the signal of the S-2 time series, allowing for a prompt detection and thus a lower omission for those events in C1. On the other hand, the spectral signature of burnt areas enables a more reliable classification in C2. The clear-cut class also reaches a good UA with 73.7% in the Spanish site, while the PA is low with only 34.1%. This is again mostly due to the omission of the disturbance events already in C1.

Table 16. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Spanish demonstration site.

		Reference				UA:	CI 95%:	CE:
		None	Clear cut	Wildfire	Total			
C2 Class	None	57	47	16	120	47.5%	8.9%	52.5%
	Clear cut	1	28	9	38	73.7%	14.0%	26.3%
	Wildfire	1	7	47	55	85.5%	9.3%	14.5%
	Total	59	82	72	213			
PA:		96.6%	34.1%	65.3%			OA:	62.0%
CI 95%:		4.6%	10.3%	11.0%			CI 95%:	6.5%
OE:		3.4%	65.9%	34.7%				

The exclusion of all sample points from the analysis in the Spanish demonstration site, where the disturbance events were omitted in the C1 prototype already, increases the producer accuracies for both clear cuts and wildfires significantly, showing that the general ability of the agent classification approach to detect the correct driver for disturbed areas is acceptable for the clear cut class and quite good for the wildfire class (see Table 17).

Table 17. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Spanish demonstration site – after exclusion of the samples which were omitted already in C1.

		Reference (without samples omitted in C1)				UA:	CI 95%:	CE:
		None	Clear cut	Wildfire	Total			
C2 Class	None	1	10	1	12	8,3%	15,6%	91,7%
	Clear cut	1	28	9	38	73,7%	14,0%	26,3%
	Wildfire	1	7	47	55	85,5%	9,3%	14,5%
	Total	3	45	57	105			
PA:		33,3%	62,2%	82,5%			OA:	72,4%
CI 95%:		53,3%	14,2%	9,9%			CI 95%:	8,6%
OE:		66,7%	37,8%	17,5%				

Doing the same analysis with the exclusion of the samples omitted in C1 for the Central European demonstration site, shows that the PA for the bark beetle and

clear-cut class improve. While the PA for the bark beetle class is still on a very low level, the PA for the clear-cut class reaches more than 61%. Nevertheless, it has to be noted that the frequent confusions of the clear-cut class with the other two classes is obviously the most significant issue in the Central site.

Table 18. Confusion matrix for the plausibility analysis of the disturbance agent classification for the Central European demonstration site – after exclusion of the samples which were omitted already in C1.

		Reference					UA:	CI 95%:	CE:
		None	Bark beetle	Clear cut	Windthrow	Total			
C2 Class	None	1	3	17	8	29	3,4%	6,6%	96,6%
	Bark beetle	0	4	4	1	9	44,4%	32,5%	55,6%
	Clear cut	0	11	38	37	86	44,2%	10,5%	55,8%
	Windthrow	0	0	3	1	4	25,0%	42,4%	75,0%
	Total	1	18	62	47	128			
PA:		100,0%	22,2%	61,3%	2,1%		OA:		34,4%
CI 95%:		0,0%	19,2%	12,1%	4,1%		CI 95%:		8,2%
OE:		0,0%	77,8%	38,7%	97,9%				

As already mentioned above there were not enough disturbance agent samples to do a plausibility analysis for the Swedish demonstration site. 39 verified clear cut samples were available and after excluding the ones omitted by C1 already, the number of samples was reduced to only 19. Eight out of those 19 points were classified correctly as clear cuts, while 6 samples were assigned to the bark beetle class in C2. Further three samples were confused with the wildfire class, rather pointing to classification noise. In the end this lack of reference data does not allow for any qualitative statement on the C2 result for the Sweden site.

3.2.7 Conclusions

The Disturbance Classifier has shown to be able to detect patches of the main disturbance agents correctly whereas it is not reliable enough on a pixel-basis. In any case, it is recommended to analyse the hard classification together with the probabilities to get a feeling for the overall classification and the ability of the classifier to separate specific classes from each other. When considering the hard classification, most of the confusion occurs because of bark beetle infestation and wind throw being classified as clear cuts in the Central European site. As mentioned in section 3.2.6, a possible explanation is the timely clearing of areas affected by pests or windthrow, as this is a common practice to avoid further damage by hindering a spread of the pest and in case of a windthrow by removing a source for potential infestations. Further, it might be important to feed the trees into the wood supply chain as fast as possible to reduce the economic damage. Such timely clearings do have the potential to obscure the initial cause of the disturbance, i.e. bark beetle or windthrow, as those two agents typically have a longer delay in their detection due to more subtle signal changes. Cleared areas on the other hand are characterized by a clear spectral signature, allowing for a faster

detection and thus being more dominant in the monitoring and classification process.

From a methodological and ancillary data point of view, regional differences in terms of the occurrence of disturbance agents have to be considered, hence regional calibration is of utmost importance. Which classes of disturbance agents occur in different regions, and if they can be reliably distinguished or should better be joined into fewer classes still remains subject to future testing. The method works in principle, but a large number of samples/reference data are needed, which might be very difficult to retrieve on a European scale. Furthermore, it has to be considered that in order to derive the C2 product, activity data from the C1 product is required. Thus, the reliability of the C2 product not only depends on the ability of the LightGBM classifier itself to correctly categorize the potential disturbance driver but is largely influenced by the performance of the continuous forest monitoring approach of C1.

From a technical point of view, the approach is robust and computationally effective, being capable to process large amounts of data. Nevertheless, it is demanding with regards to data download and time series pre-processing, which is time-consuming.

3.2.8 Outlook beyond the EvoLand project

As already described in [\[AD7\]](#) the absence of adequate ground truth data for both training of the classifier and the validation of the results turned out to be one of the biggest obstacles for a European-wide implementation of a disturbance classification product. Looking beyond the EvoLand project, it will be inevitable to gather such reference data. In this regards it will be important to have data available for all over Europe, as the agents for tree cover disturbances can vary across the different biogeographical regions of Europe. Further, it will be important to improve the Continuous Forest Monitoring approach as described in 3.1.8, as it is input to the C2 classification and issues in C1 also limit the potential of the disturbance agent classification.

3.3 C3 – Forest Biomass

3.3.1 Introduction & Objectives

The candidate prototype C3-Biomass consists of two main products: The Forest Canopy Height (FCH) and the Above-Ground Biomass (AGB), targeting woody biomass (Table 19). Forest canopy height can also be called tree cover canopy height. This also allows for applying a forest definition and thus offers flexibility to the user.

Table 19. Technical and thematic specifications for the C3 prototype.

Prototype C3	
Prototype	C3 - Biomass Mapping
Thematic specifications	Above-ground woody biomass (AGB) in t/ha; Forest canopy height (FCH) in meters
Prototype sites	1 – Spain 5 – Sweden 7 – French Guiana (testing only)
Temporal frequency	Annual
Spatial resolution	10m
MMU	pixel-based (0.25ha and 0.5ha tested in WP2)
Envisaged accuracy	N/A

3.3.2 State of Play

Biomass mapping is part of research projects for many years, as well as operationally performed in the Climate Change Initiative (CCI) Biomass, but not below 100m spatial resolution globally. Within CLMS there is high innovation potential, since such a product is not yet part of the CLMS forest portfolio. World biomass being a future expected product in the CLMS portfolio, Tree or Forest Canopy Height is not yet included, despite of some developments going on at the moment in the scientific community towards large-scale tree canopy height products.

3.3.3 Selection of Prototypes sites

The prototype test sites that have been analysed for Phase 1 are located in Sweden (with an area of 4 Sentinel-2 tiles) and Spain (with an area of 9 Sentinel-2 tiles). For Phase 2 the same site in Sweden was rolled out for two additional years and there were tests for the Central European demonstration site as well as the tropical site in French Guiana.

3.3.4 Input and Ancillary Data & Pre-processing

This chapter describes the input and reference data, used for AGB and FCH mapping for the different test sites in Sweden, Germany, Spain and French Guiana

that were tested in Phase 1 and/or 2. Reference data for model training and validation as well as model parameters were slightly modified between the different test sites and roll-outs for different years, as described below.

To prepare multi-sensor datasets for biomass mapping, input variables from Sentinel-1 radar, Sentinel-2 optical imagery, GEDI LiDAR, and topographic data were systematically pre-processed and harmonized into a consistent feature stack.

Sentinel-2 (Optical):

In Phase 1 the input data for the AGB, as well as the Forest Canopy Height mapping in Sweden and Spain were Sentinel-2 (S-2) L2A Median composites at 10m spatial resolution. While the data for the Sweden test site covers the period from 1st of June through to 15th September 2021, the S-2 data for the Spanish test site started from 1st of April till 30th of September of the reference year 2021. Those composites were derived using bands 2-11. Furthermore, texture features were derived from the green, red, and NIR S-2 bands with 10m spatial resolution using a 3x3 standard-deviation filter ['S-2_B3STD', 'S-2_B4STD', 'S-2_B8STD'].

In Phase 2 optical imagery was processed into cloud-free seasonal medoid composites to minimize atmospheric contamination. This was done for all demonstration sites except for Spain. For Sweden this covers the period from 1st of June through to 15th September 2021. While the whole half-year period was processed at once for Sweden, separate medoids were generated for the leaf-on season (April-September) for the second (Q2) and third (Q3) quarter of the year in case of the Central European site, thereby capturing key phenological dynamics. Input bands included visible (B02, B03, B04), red-edge (B05, B06, B07), near-infrared (B08, B8A), and shortwave infrared (B11, B12), providing a broad spectral basis to describe vegetation condition and biomass variability.

Sentinel-1 (Radar):

In Phase 2 interferometric Wide (IW) Ground Range Detected (GRD) mosaics in VH and VV polarizations of the Sentinel-1 sensor were aggregated on a monthly basis for the vegetation period (May–September). These were transformed into seasonal median composites by first converting input rasters (dB×100) into linear scale, stacking across months, and computing the pixel-wise median to suppress speckle and outliers. The aggregated results were then converted back to dB×100 and stored per year, tile, and polarization, yielding stable radar features sensitive to canopy structure and moisture conditions.

ALOS PALSAR:

Additional input data were L-band ALOS PALSAR mosaics in HH and HV polarization for the year 2021 at 25 m spatial resolution. ALOS PALSAR data were only used during Phase 1 for the demonstration sites in Spain and Sweden.

NFI:

The reference data for the model training and validation of the Sweden site in both Phases was point data from the National Forest Inventory (NFI) of Sweden, where all samples from the reference year 2021 were extracted. This resulted in 286 samples, split in two sets, used for training (90%) and validation (10%). From the 90% samples that have been used for training data, the algorithm randomly selects 15% for an internal validation. The NFI data includes both, information on above ground biomass and mean tree height, used accordingly for the two prototypes.

GEDI (LiDAR):

GEDI Level-2B data covering the period from April 1st 2021, to September 30th 2021 were extracted over the area of interest only for Germany and Spain using only strong beams (BEAM0000, BEAM0001, BEAM0010, BEAM0011). From this dataset above ground biomass (AGB) was extracted. For each footprint solar elevation and sensitivity values were retrieved. Invalid or low-quality returns were removed with sensitivity thresholds (>0.95) applied. To further reduce atmospheric noise, only night-time acquisitions were retained, ensuring higher fidelity canopy height and structure measurements.

Topography (DSM):

For all sites and Phases, topographic features were derived from the Copernicus Digital Surface Model (DSM). The DSM provided elevation (m) and the slope (%) was calculated directly from it. These terrain descriptors help in correcting elevation effects and topographic influences on forest growth, biomass distribution, and remote sensing backscatter/reflectance patterns.

Integrated Feature Stack:

The final harmonized dataset combines radar, optical, LiDAR and terrain information, enabling robust modelling of aboveground biomass by leveraging complementary sensitivities of different sensors to forest structure, composition, and environmental gradients. To reduce noise, the data were pre-processed using a Gaussian filter with a 5x5 filter size and a sigma value of 0.7.

A list of data sets it provided in Table 20:

Table 20. List of datasets used for the C3 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	L2A Median composites Derived texture features	10m	01.06.21 – 15.09.21
Sentinel-1	Monthly mosaics Temporal features	10m	2020, 2022
ALOS PALSAR	Mosaics with HH and HV polarisation	25m	2021
National Forest Inventory (NFI) Sweden	Point samples filtered for the year 2021 with information on biomass and mean tree height	Point data	2021
Global Ecosystem Dynamics Investigation (GEDI)	Level 4A: footprint level above ground biomass	25m diameter	2021

DSM	Topographic features from the Copernicus Digital Surface Model (DSM)	10m	2011-2015
-----	--	-----	-----------

3.3.5 Methodology and Workflow

For the biomass estimation a supervised machine learning framework was implemented using the harmonized multi-sensor feature stack derived from Sentinel-2 optical, Sentinel-1 radar (Phase 2) imagery, GEDI LiDAR metrics (Germany and Spain), NFI (Sweden), and DSM-based topographic variables.

A Random Forest regression model was selected due to its robustness against multicollinearity, the capacity to model complex nonlinear relationships, and the well-documented reliability in remote sensing applications. The model was calibrated using GEDI-derived AGB values as reference data, while radar, optical, and terrain variables served as predictor features. This design enabled the integration of complementary sensor information—capturing vegetation structure, spectral properties, and topographic context—to generate spatially continuous biomass estimates across the study area.

To address potential imbalances in the distribution of AGB values, the reference dataset was stratified into three bins: low biomass ($AGB < 100$ t/ha), medium biomass ($100 \leq AGB \leq 300$ t/ha), and high biomass ($AGB > 300$ t/ha). From each stratum, 1,000 samples were randomly drawn, ensuring balanced representation in the final training dataset and improving the model's ability to generalize across the biomass gradient.

The Random Forest was trained with carefully selected hyperparameters:

- 1,000 trees ($n_{estimators}$) to ensure model stability
- Maximum depth of 20 (max_depth) to balance complexity and overfitting risk
- Minimum of 5 samples per leaf ($min_samples_leaf$) and 2 samples per node split ($min_samples_split$) to constrain model growth
- A maximum of 7 features considered at each split ($max_features$) to enhance diversity among trees

The predictor variables included DEM-derived topographic features (surface elevation and slope) and Sentinel-2 spectral bands representing seasonal conditions, specifically composites (medians as well as percentiles in Phase 1 and medoids in Phase 2). These composites covered half-year data for Spain (2021 only Phase 1) and Sweden (2021 Phase 1 and 2020/22 in Phase 2). For the Central Europe test site, the data were split into a second (Q2) and a third (Q3) quarter of 2022.

To ensure model reliability and robustness, the training–evaluation process was repeated 26 times, each using a different random seed (random states 70–95). For each run, the dataset was split into an 80% training and 20% hold-out test set,

providing repeated assessments of model performance and stability across varying initial conditions.

The C3 product is delivered at 10m spatial resolution with a Minimum Mapping Unit (MMU) of 0.5 ha and at an annual temporal resolution.

3.3.6 Results and Verification

This chapter shows the outcome and validation results for the test sites in Sweden and in Spain stemming from Phase 1 and 2. Results for Central Europe and French Guiana are not included in this report, as it turned out that the implemented approach was not transferable to those regions and accuracies did not reach a satisfactory level, as further explained below. Biomass estimations are defined as t/ha of dry biomass, and Forest Canopy Height is presented in meters.

3.3.6.1 Test Site Sweden

Above Ground Biomass

The training accuracy (R^2 Score: 0.53) for the AGB model for the year 2021 (Phase 1) is displayed in Figure 12. The scatterplot of the points in comparison to the 1:1 line displayed in red colour within that figure points towards a possible saturation at around 120t/ha.

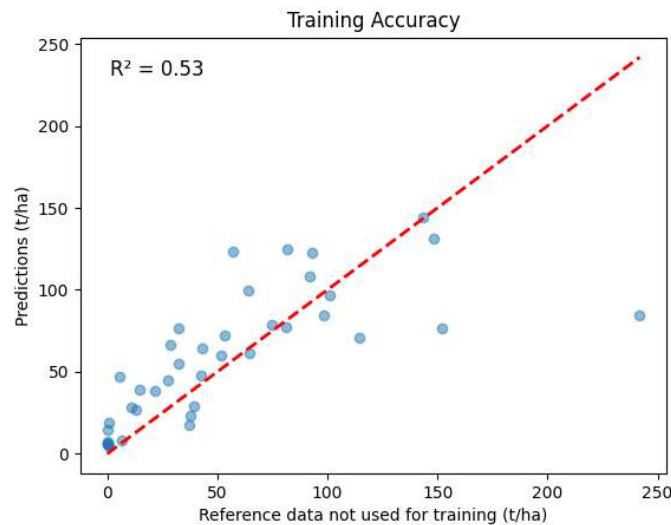


Figure 12. Training accuracy for the 2021 AGB in the Sweden test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.

Figure 13 shows the result for the 2021 AGB over the Sweden test site. While the western side of the study site is dominated by areas of lower AGB because of sparser tree coverage due to marshlands, the eastern side shows by more densely forested areas. Regions of high biomass above 150 t/ha mainly correspond to needle leaf forests according to the VLCC data of 2021.

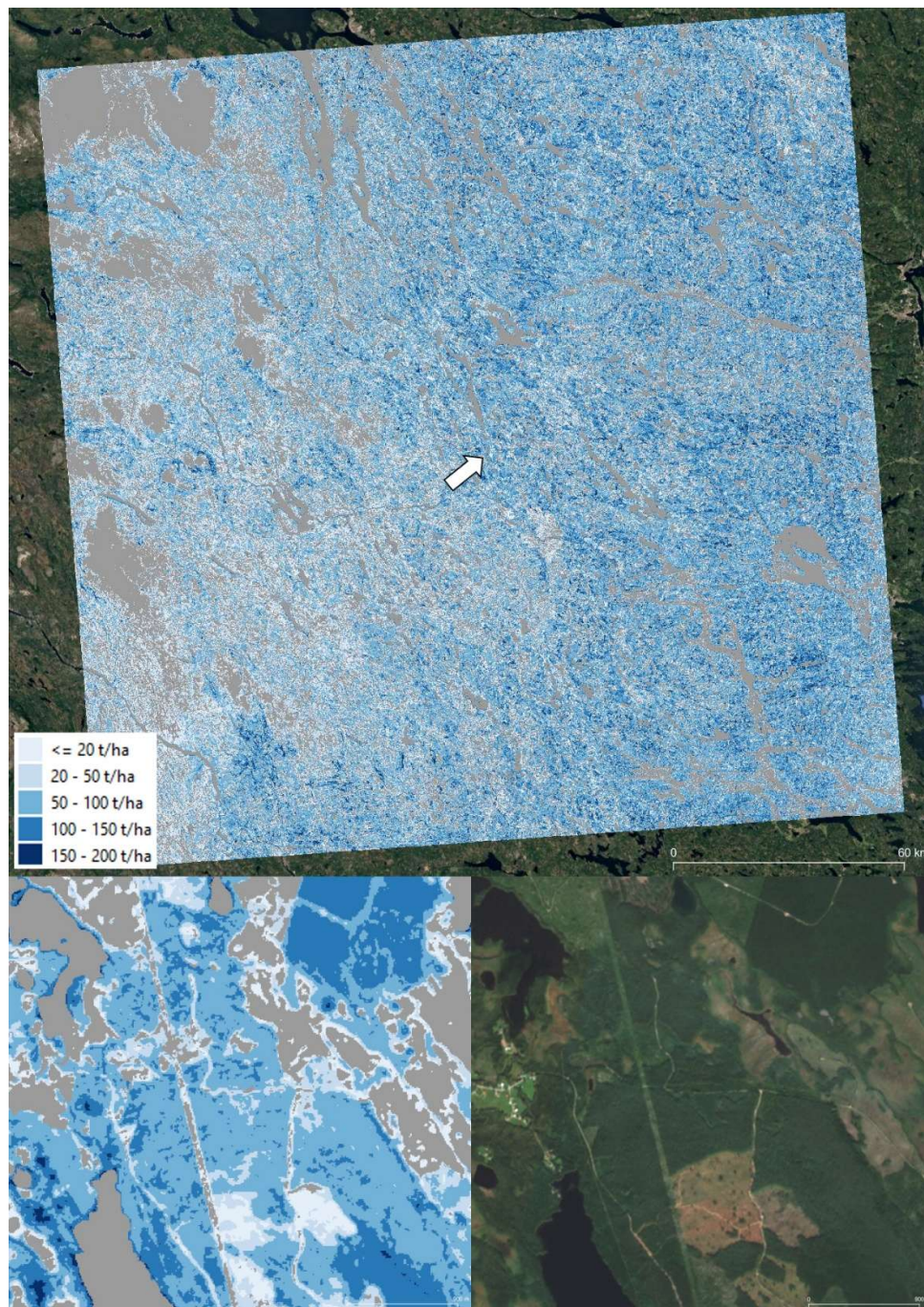


Figure 13. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2021 (Phase 1 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.-09.2021. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

Due to the impact of cloud cover during the summer season, a Sentinel-2 composite covering two months had to be derived as reference scene. The AGB result (Figure 13 on the lower left) shows an intermediate result of the clear-cut

activities (southern part of that figure) as compared to the Sentinel-2 reference composite (on the lower right) because the clear-cut activities occurred between July (no impact visible on the corresponding monthly S-2 composite) and August (full impact visible on the corresponding S-2 composite) of that year.

The validation of the 2021 result (Phase 1) resulted in a mean absolute error (MAE) of 21.90, a root mean squared error of 27.89 and a value of 0.51 for R^2 . The actual versus the predicted biomass for the validation samples is presented in the plot in Figure 14. Based on this restricted number of validation sample plots the model might overestimate AGB at lower biomass values and underestimate AGB at higher biomass values.

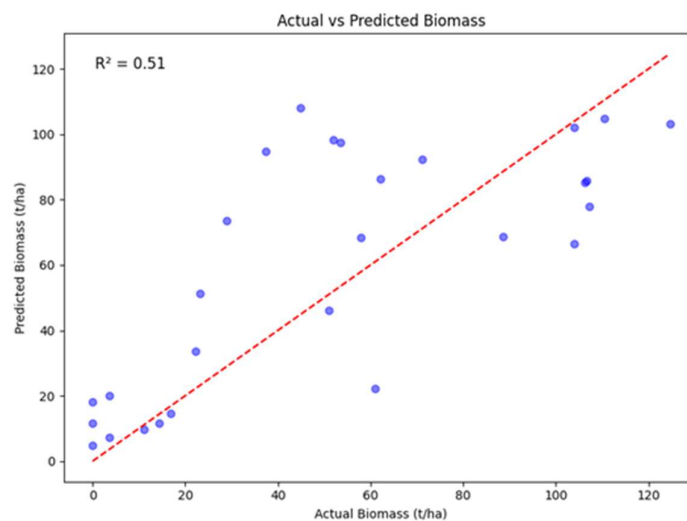


Figure 14. Plot showing the predicted versus the actual biomass for the validation samples of the 2021 Sweden test site (Phase 1).

The validation results for the years 2020 and 2022 of Phase 2 showed mean absolute errors (MAE) of 23.12 and 29.58, root mean squared errors of 31.50 and 40.01, and R^2 values of 0.71 and 0.52 respectively (Figure 15 & Figure 16). Similar to the 2021 result the scatterplots indicate an underestimation of AGB at higher biomass values most probably due to model saturation.

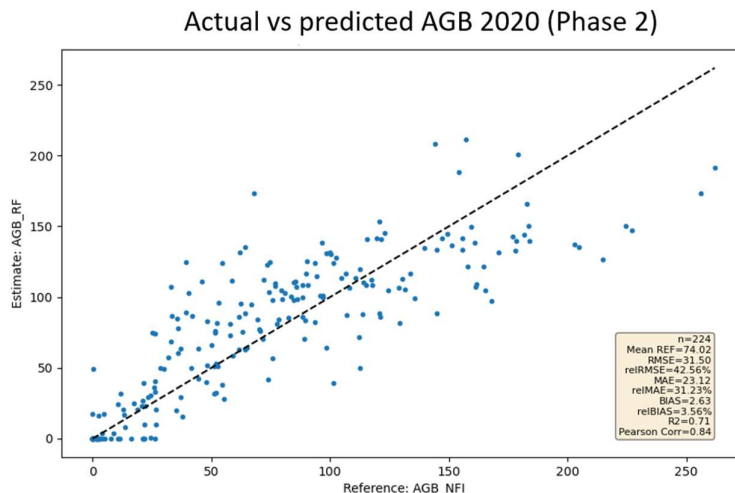


Figure 15. Plot showing the predicted versus the actual biomass for the validation samples of the 2020 Sweden test site (Phase 2).

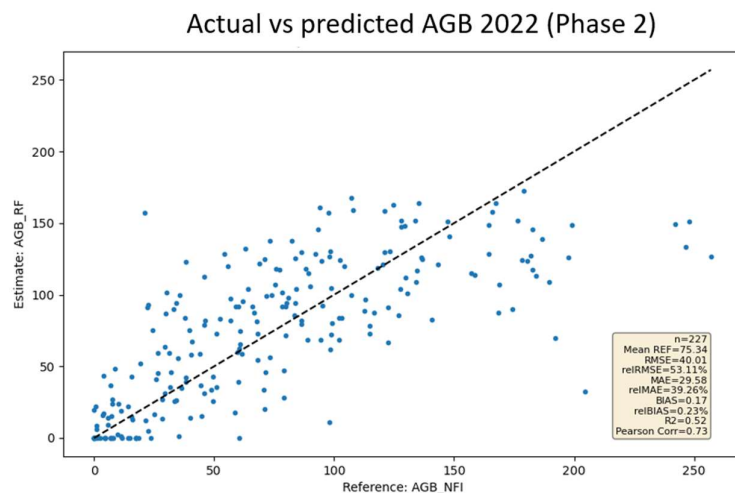


Figure 16. Plot showing the predicted versus the actual biomass for the validation samples of the 2022 Sweden test site (Phase 2).

Eventually, the derived AGB estimates of 2021 were visually compared to the European Space Agency (ESA) CCI Biomass dataset of 2021 (Figure 17), which comes at a spatial resolution of 100m. While the overall pattern generally fits well between the two datasets, the newly derived AGB dataset of C3 shows much more spatial details due to its higher resolution data (Figure 18).

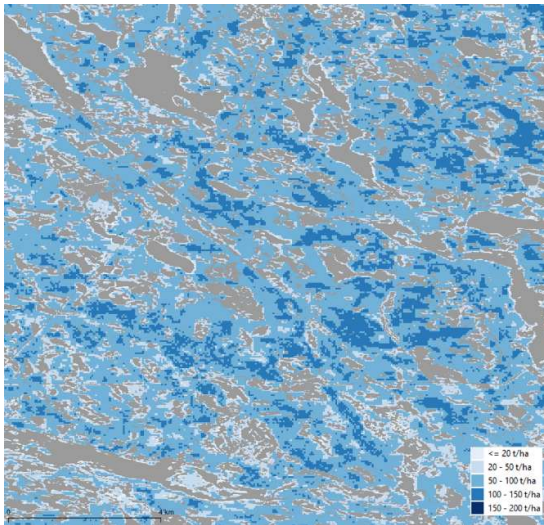


Figure 17. ESA CCI Biomass dataset of 2021– Swedish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

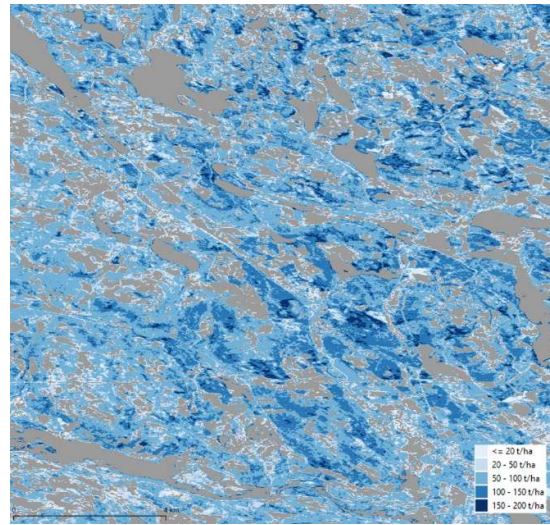


Figure 18. C3 AGB dataset of 2021 – Swedish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

In Phase 2 a slightly modified approach was used to derive AGB results over the Sweden test site for the year 2020 (Figure 19) and 2022 (Figure 20). In contrast to the 2021 result of Phase 1 there has been no bilateral filtering applied. Reason for this was to analyse to what extent the unfiltered results allow depicting small-scale changes in AGB over the different years. However, in most of the cases bilateral filtered results, that still keep fine-scale spatial details, are preferable because the main applicability is the extraction of ABG values over polygons instead of single pixel results.

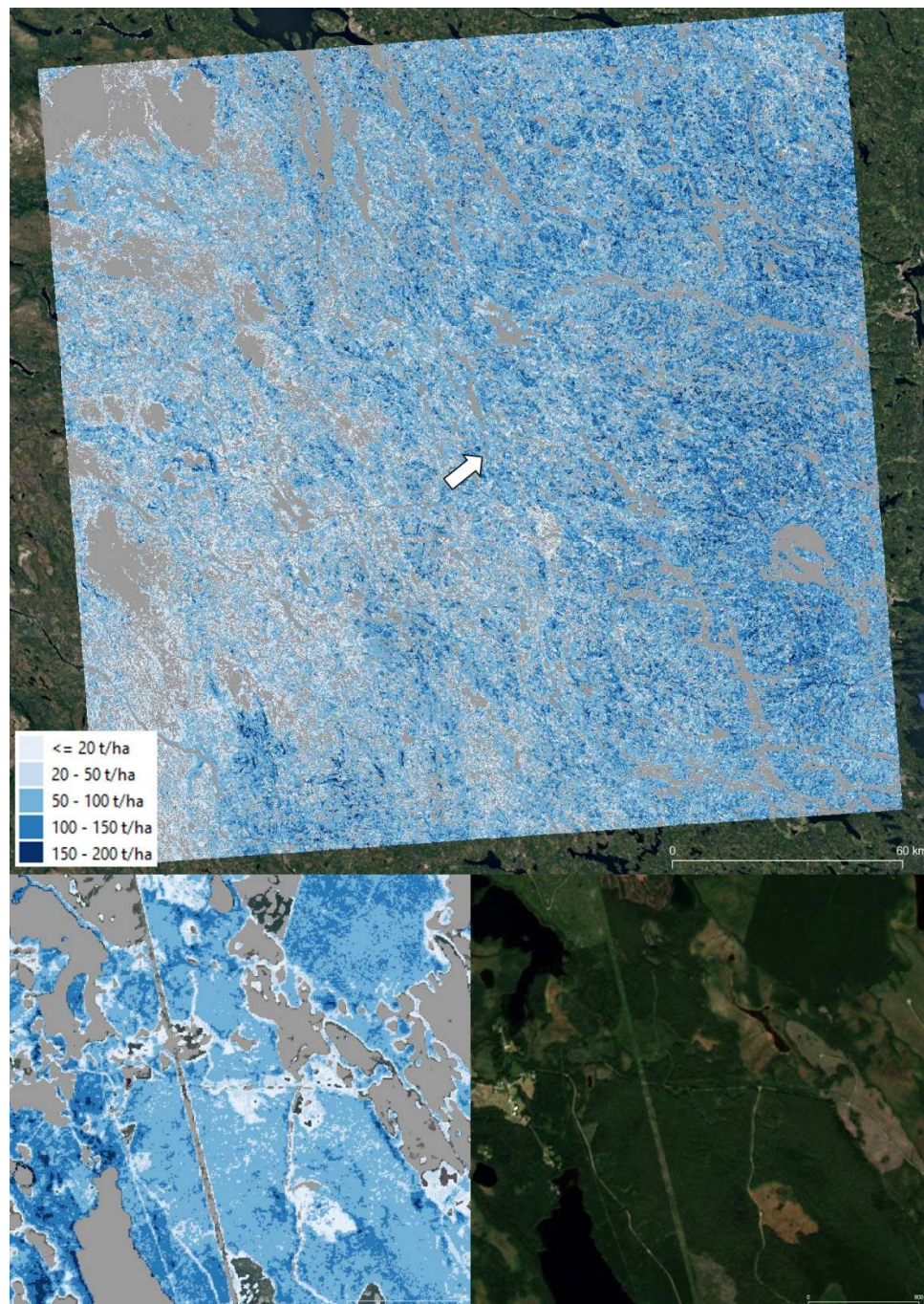


Figure 19. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2020 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2020. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

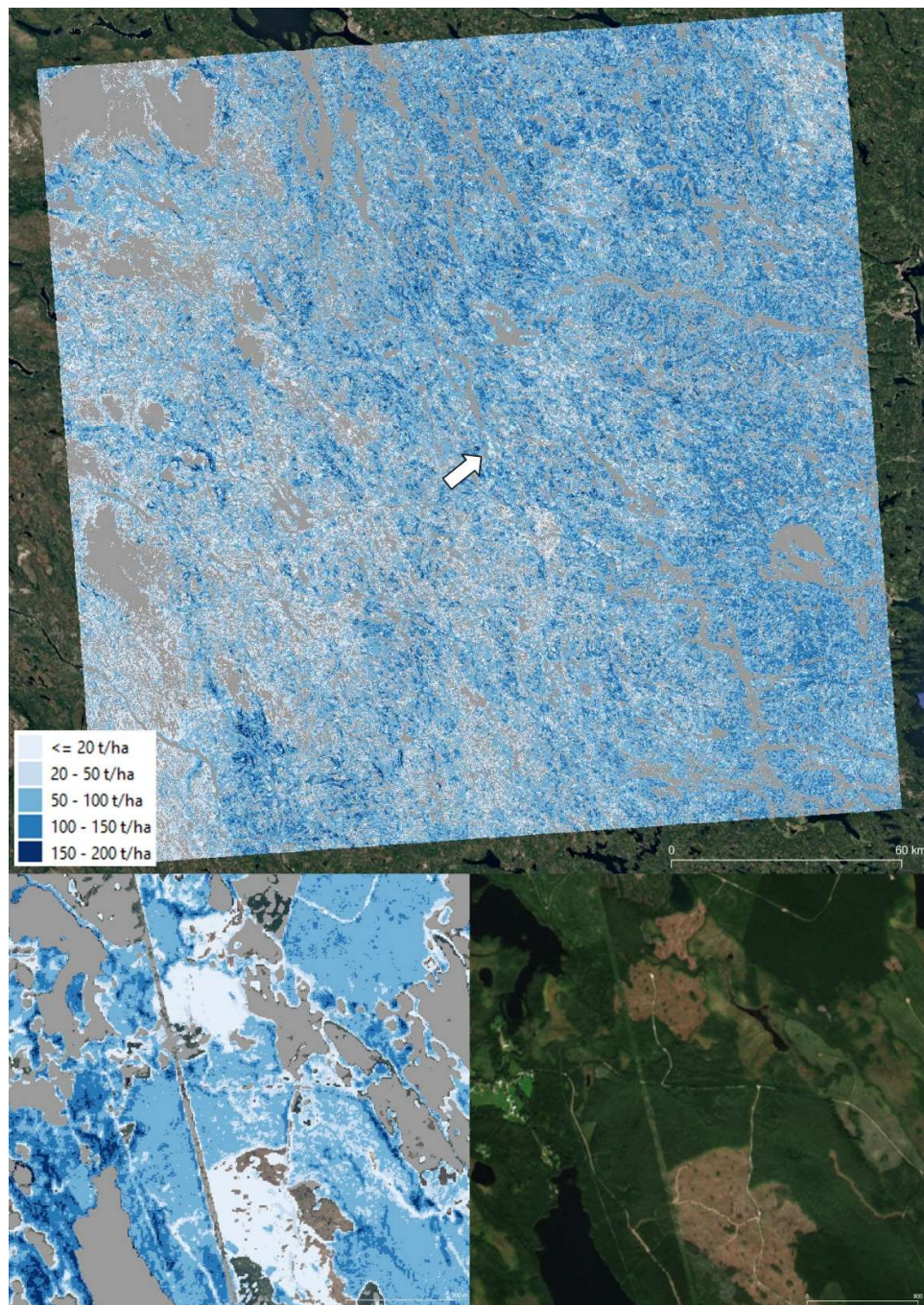


Figure 20. Prediction of the AGB (t/ha) for the Sweden test site for the reference year 2022 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2022. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

The clearing activities visible in the 2020 and 2022 results (images on the lower right of Figure 19 and Figure 20) are well depicted with the lowest biomass class in the corresponding AGB subsets (Figure 19 and Figure 20 lower left closeups).

Figure 21 shows the training accuracy (R^2 Score: 0.47) for the FCH data for the year 2021 (Phase 1) over the test site in Sweden. According to the scatterplot there might be an overestimation of FCH for smaller trees while larger trees might get underestimated in height.

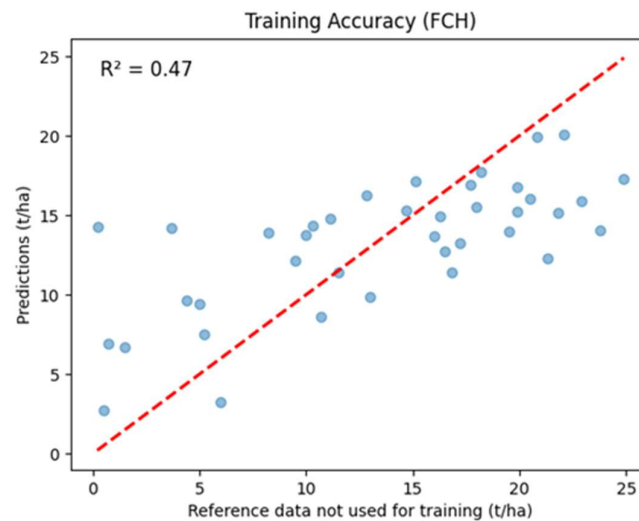


Figure 21. Training accuracy for the 2021 FCH for the Sweden test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.

Figure 22 shows the prediction for the 2021 Forest Canopy Height over the test site in Sweden. According to the VLCC data of 2021, regions with higher FCH values mainly correspond to needle leaf forests.

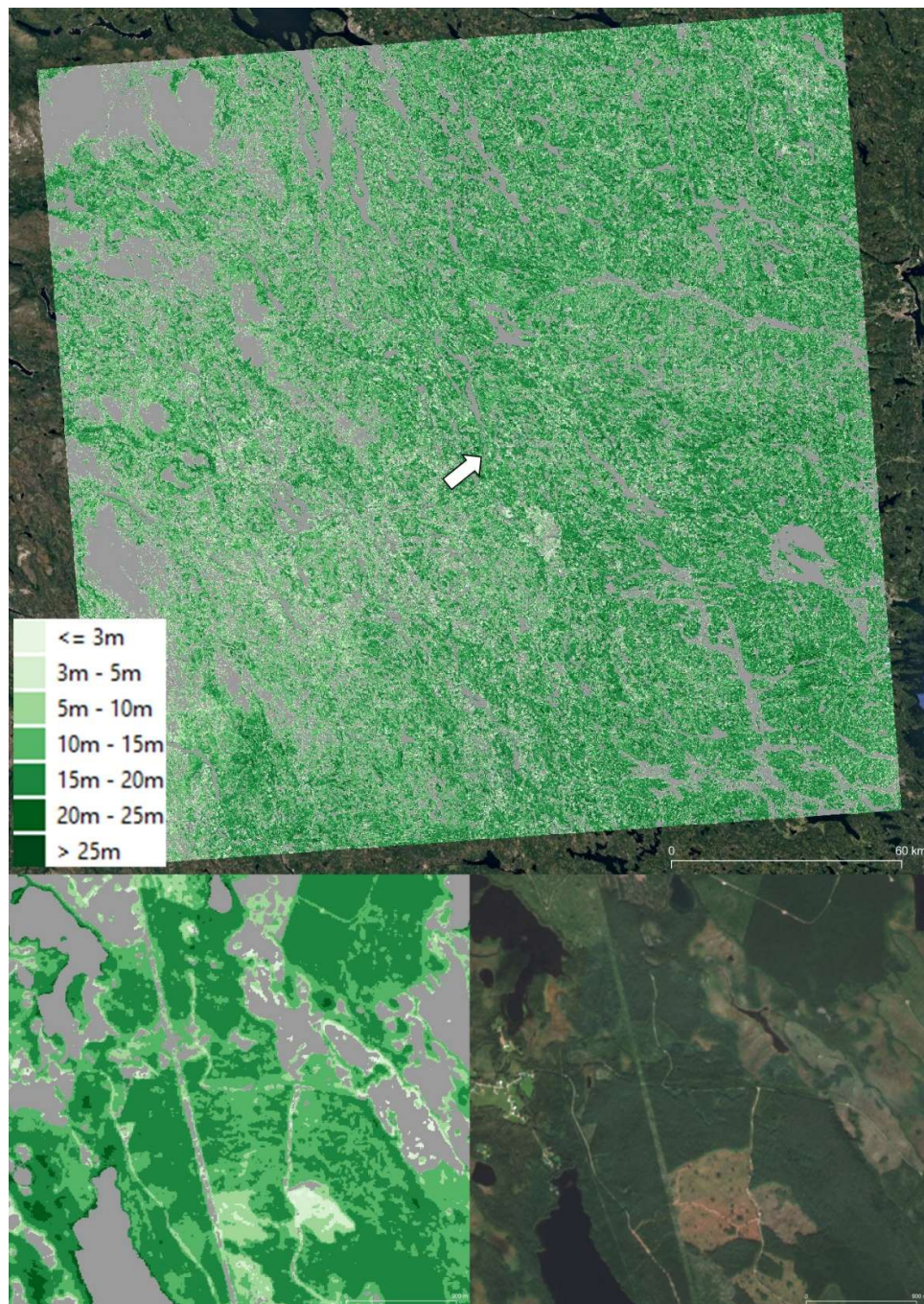


Figure 22. Prediction for the FCH of the Sweden test site for the reference year 2021 (Phase 1 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.-09.2021. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

Similar to the AGB result of 2021 above, the 2021 FCH result (Figure 22 on the lower left) shows an intermediate result of the clear-cut activities (southern part) as compared to the S-2 reference composite (lower right) because the clear-cut activities occurred between July (no impact visible on the corresponding monthly

S-2 composite) and August (full impact visible on the corresponding S-2 composite) of that year,

The validation of the 2021 FCH product (Phase 1) resulted in a mean absolute error (MAE) of 2.43, a root mean squared error of 3.11 and a value of 0.73 for R^2 . The actual versus the predicted FCH for the validation samples is presented in the plot in Figure 23.

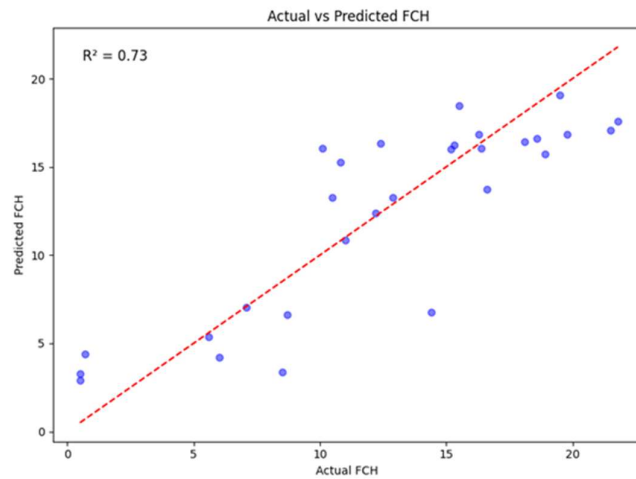


Figure 23. Plot showing the predicted vs. the actual FCH for the validation samples of the 2021 Sweden test site (Phase 1).

The validation results for the years 2020 and 2022 of Phase 2 showed mean absolute errors (MAE) of 2.52 and 3.49, root mean squared errors of 3.38 and 4.51, and R^2 values of 0.76 and 0.56 respectively (Figure 24 & Figure 25). Compared to the 2021 result the scatterplots indicate an underestimation of FCH at taller tree heights probably due to model saturation.

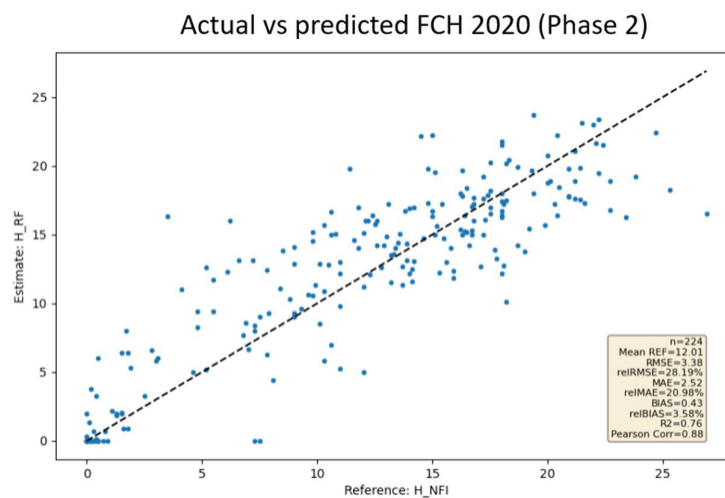


Figure 24. Plot showing the predicted versus the actual tree height values for the validation samples of the 2020 Sweden test site (Phase 2).

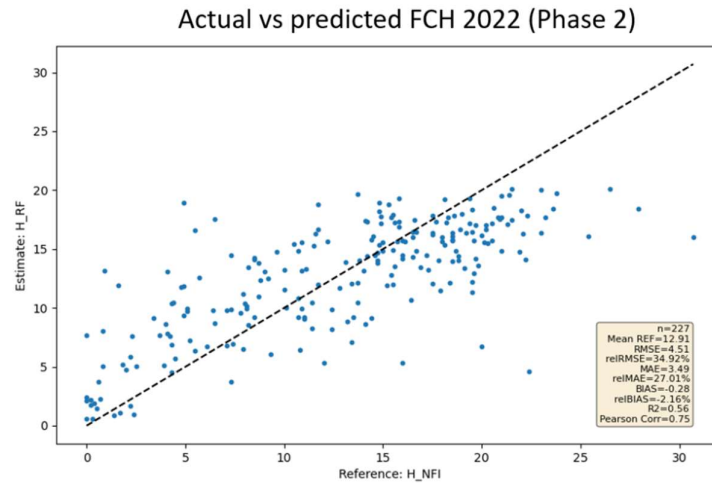


Figure 25. Plot showing the predicted versus the actual tree height values for the validation samples of the 2022 Sweden test site (Phase 2).

In Phase 2 a slightly modified approach was used to derive FCH results over the Sweden test site for the year 2020 (Figure 26) and 2022 (Figure 27). Similar to the Phase 2 AGB results, there has been no bilateral filtering applied for Phase 2 FCH, as well.

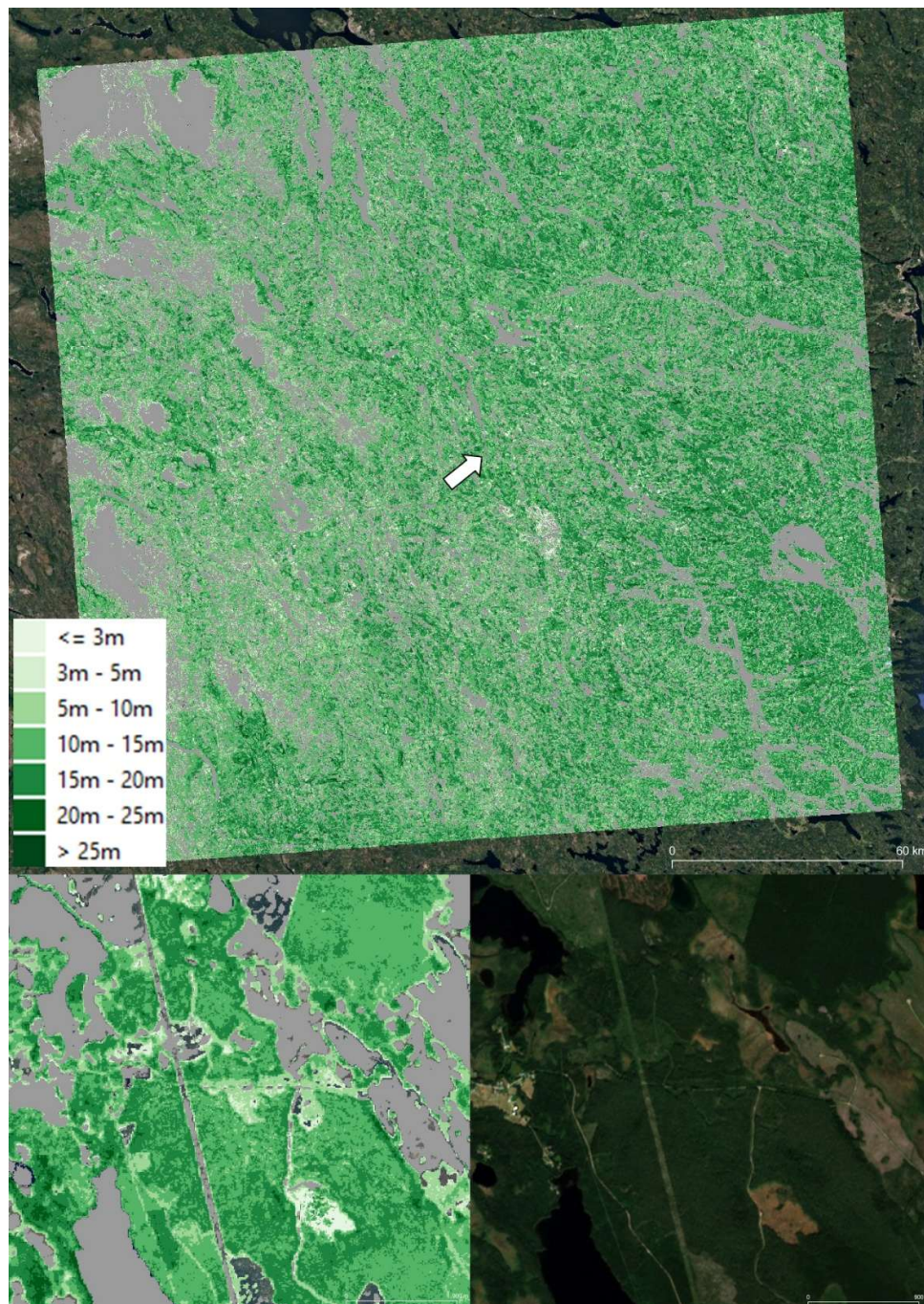


Figure 26. Prediction for the FCH of the Sweden test site for the reference year 2020 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2020. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

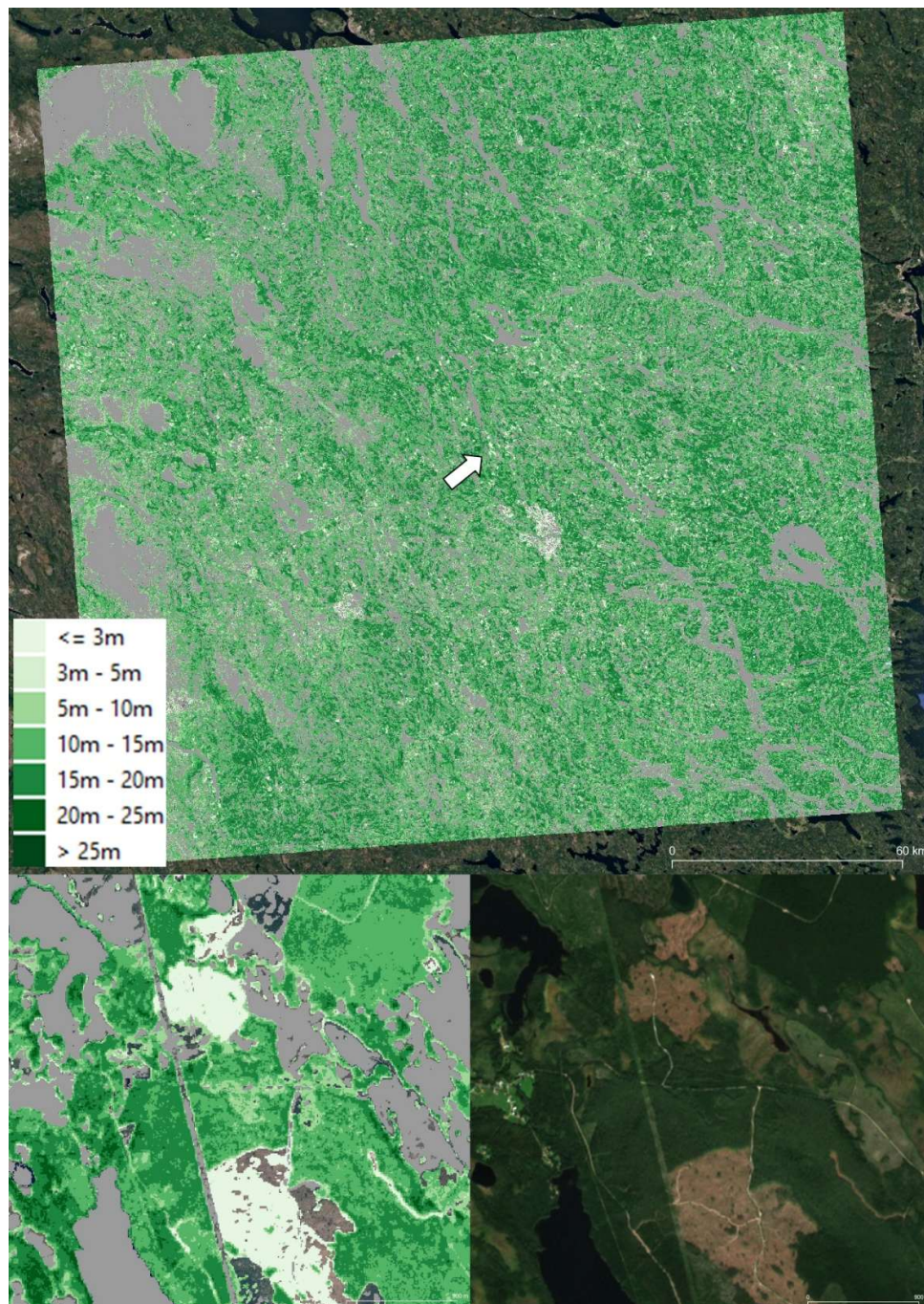


Figure 27. Prediction for the FCH of the Sweden test site for the reference year 2022 (Phase 2 result). The arrow indicates the position of a close-up (lower left figure) and the corresponding location depicted by a Sentinel-2 composite of 08.2022. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

The clearing activities visible in the 2020 and 2022 results (lower right images in Figure 26 and Figure 27) are well depicted with the lowest tree height class in the corresponding FCH subsets (lower left closeups in Figure 26 and Figure 27).

3.3.6.2 Test Site Spain

Above Ground Biomass

The training accuracy (R^2 Score: 0.54) for the AGB model in Spain is displayed in Figure 28. Similar to the data for Sweden the model for Spain seems to underestimate the AGB predictions with respect to the reference data not used for training as some indications of saturation can be observed.

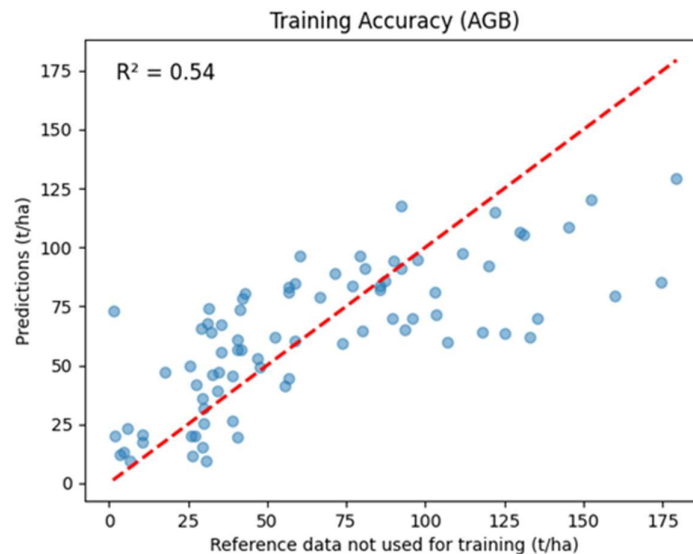


Figure 28. Training accuracy for the AGB model in the test site Spain, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.

Figure 29 shows the result for the AGB over the Spain test site. While the central part of the study area is dominated by lower AGB values below 50 t/ha that mainly cover olive tree plantations, the forests/ forest plantations over the mountainous regions towards the North, East and South show higher AGB values of greater 50 or even above 100 t/ha. There are no AGB values exceeding the class 100-150 t/ha.

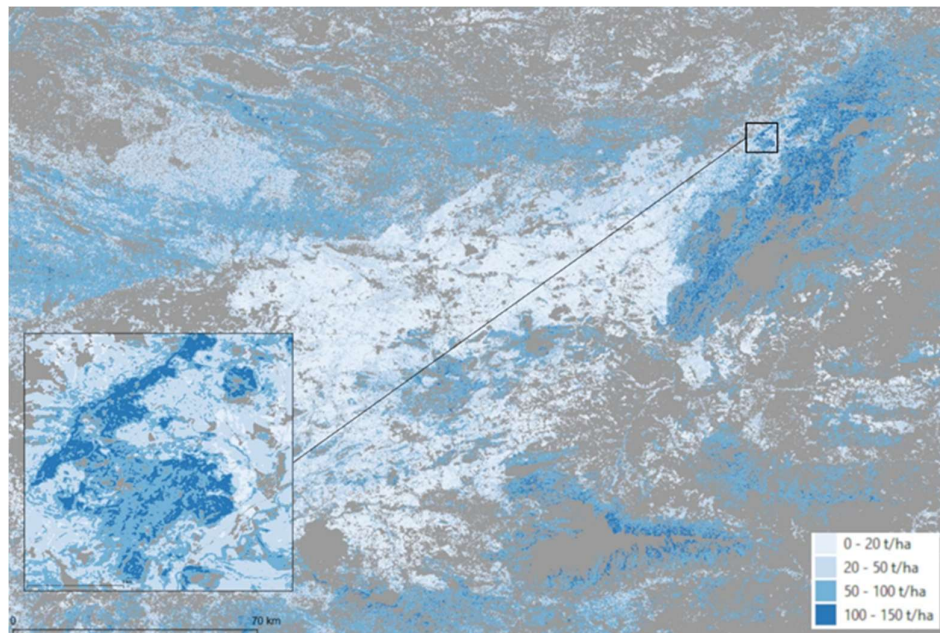


Figure 29. Prediction for the AGB in the Spain test site for the reference year 2021 with a close-up to a higher AGB forest area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

The validation resulted in a mean absolute error (MAE) of 16.51, a root mean squared error of 20.55 and a value of 0.68 for R^2 . The actual versus the predicted biomass for the validation samples is presented in the plot in Figure 30. Based on these validation sample plots there might be a slight overestimation of AGB at lower biomass values and an underestimation of AGB at higher biomass values. This result is similar to the Sweden test site.

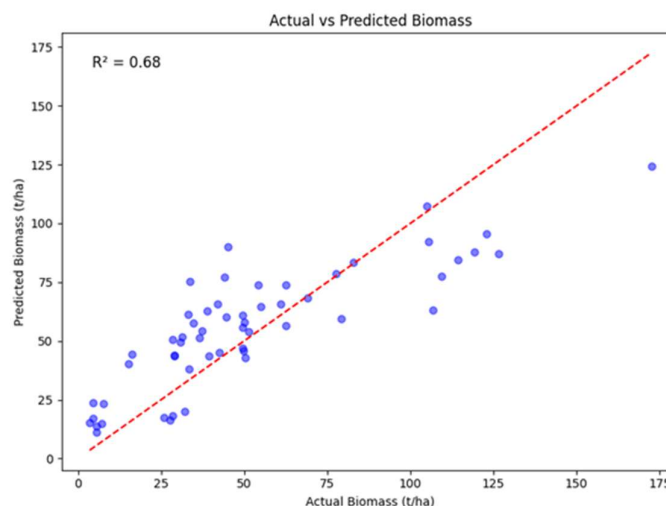


Figure 30. Plot showing the predicted vs. the actual biomass for the validation samples of the Spain test site.

The so derived AGB estimates were – similar to the Swedish test site – compared to the ESA CCI Biomass dataset of 2021 (Figure 31). While the C3 ABC data of Figure 32 show no very high AGB class (150-200 t/ha), the CCI Biomass dataset includes

this class. This supports the hypothesis stated above that the C3 AGB result tends to underestimate areas with higher biomass. In addition, the comparison of both figures highlights that areas with lower AGB values within the CCI Biomass dataset show intermediate AGB values in the C3 product, which seems to indicate that the model tends to overestimate areas of low biomass.

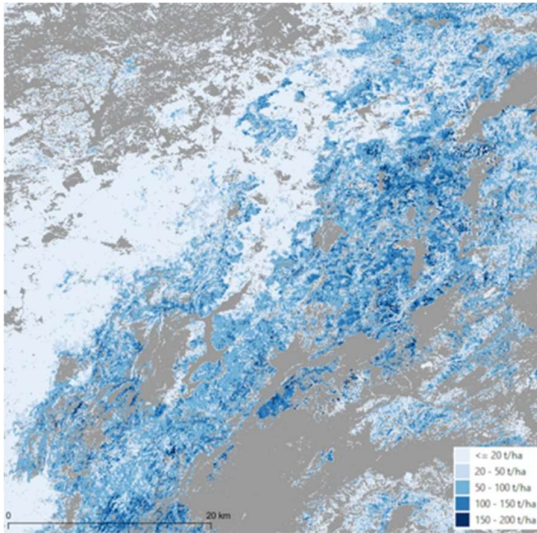


Figure 31. ESA CCI Biomass dataset of 2021 of tree canopy height – Spanish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

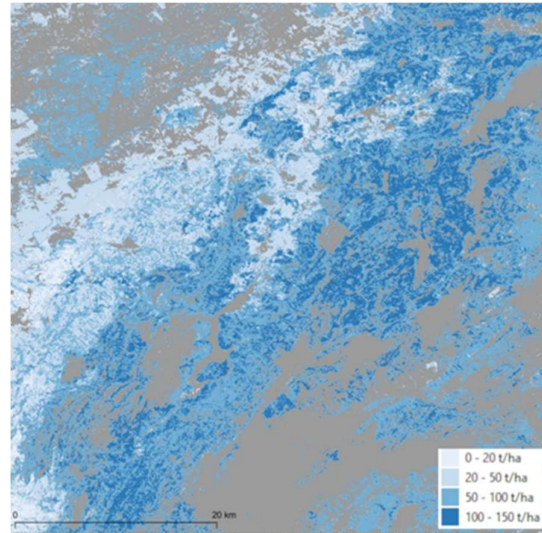


Figure 32. C3 AGB dataset of 2021 of tree canopy height – Spanish study area. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

Forest Canopy Height

Figure 33 shows the training accuracy (R^2 Score: 0.51) for the FCH over the test site in Spain. Similar to the test site in Sweden, the scatterplot points towards an overestimation of FCH for smaller trees while larger trees might get underestimated in height.

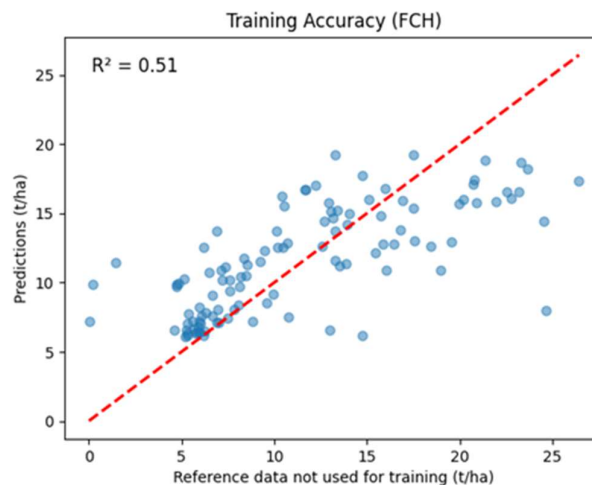


Figure 33. Training accuracy for the FCH model in the Spain test site, applying the 15% of samples that have been set aside for internal validation as part of the training. Red dotted line is the 1:1 line.

Figure 34 shows the prediction for the Forest Canopy Height over the test site in Spain. According to the freely available VHR data (Google Earth and ESRI World Imagery), regions with higher FCH values mainly correspond to needle leaf forests.

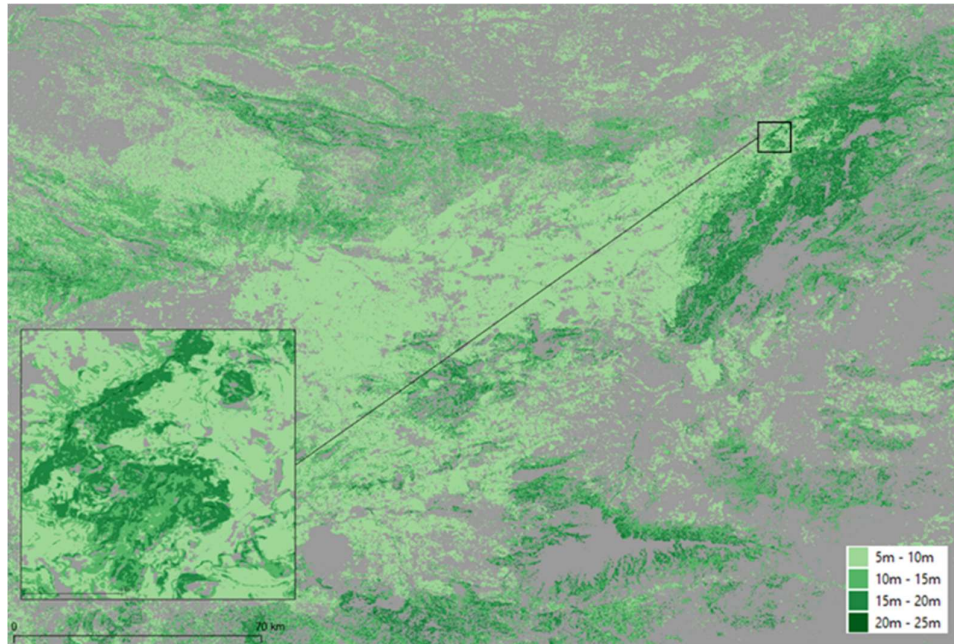


Figure 34. Prediction for the FCH in the Spain test site with a close-up to an area with higher FCH values. Grey coloured areas refer to non-forested regions according to the VLCC map of 2021.

The validation of the FCH product resulted in a mean absolute error (MAE) of 2.83, a root mean squared error of 3.65 and a value of 0.47 for R^2 . The actual versus the predicted FCH for the validation samples is presented in the plot in Figure 35.

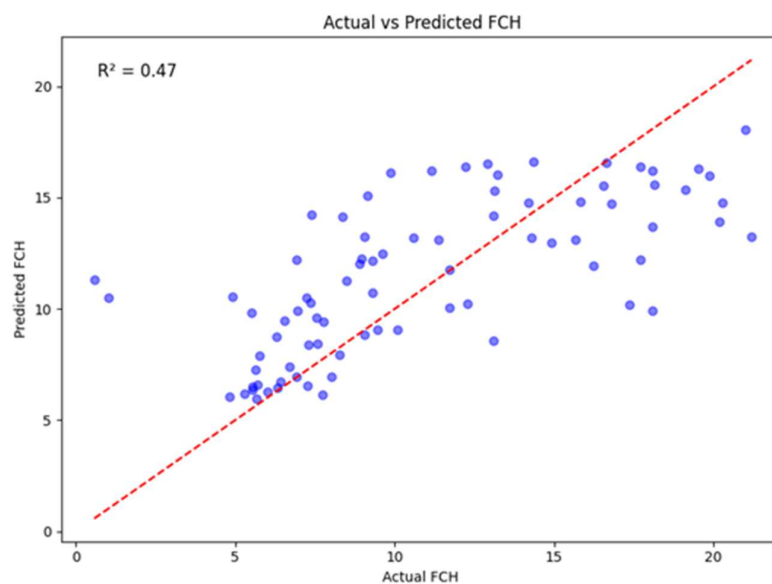


Figure 35. Plot showing the predicted vs. the actual FCH for the validation samples of the Spain test site.

3.3.7 Conclusions

The models developed in Phase 1 for above-ground biomass (AGB) and forest canopy height (FCH), based on optical and SAR time series, DEM-derived information, and GEDI LiDAR data, were able to reproduce land cover structures visible in HR and VHR reference data. Nevertheless, the tested methods remained at an early stage of development and were not yet operational. Initial results indicated systematic limitations: saturation effects at higher biomass and canopy height values led to underestimation, while AGB tended to be overestimated at lower biomass levels. These findings underscored the need for further refinement and validation before any European-wide application could be recommended. The addition of FCH to an AGB product was also identified as feasible, as both can be derived through comparable workflows with limited additional effort.

In Phase 2, reference and input datasets were adapted and expanded. Spain was not reprocessed, whereas the results for Sweden were extended with two additional reference years using S-2 optical and S-1 radar data, instead of the S-2 and ALOS PALSAR combination used in Phase 1. NFI data continued to serve as reference. Results for Sweden, consistent with earlier findings for Spain, confirmed saturation effects at higher biomass and canopy height levels. Furthermore, two additional sites were introduced: The Central European site and the tropical demonstration site in French Guiana. Both were analysed with S-2 and S-1 data for roll-out and GEDI data as reference. However, model performance at these sites was very poor (e.g., $R^2 = 0.14$ for Central Europe), suggesting that models are performing well in medium- to lower-biomass regions, but cannot be directly transferred to higher-biomass conditions.

Finally, and most importantly, it should be noted that the anticipated availability of L-band NISAR radar data did not materialize, as the sensor was launched only recently and first data was not available in time to be considered for this work. As this novel mission was initially planned to deliver the main input there was the need to find alternative approaches to best possible estimate the biomass. It has to be concluded that this worked quite well in some test regions, but the implementation of a robust and transferrable approach was not possible without this data source.

3.3.8 Outlook beyond the EvoLand project

Addressing the saturation effects observed in the settings of both Phases will be key to improving the robustness of AGB and FCH models. Upcoming/newly started satellite missions are expected to play a major role in this regard. The NASA-ISRO SAR (NISAR) mission, operating at L-band, is anticipated to provide greater sensitivity to forest structure, particularly at intermediate to high biomass levels. The ESA Biomass mission, with its P-band observations, will enable deeper canopy penetration and reduce saturation, offering improved retrievals in tropical and boreal regions. Combining these datasets with optical time series and LiDAR reference information such as GEDI will strengthen large-scale monitoring

capacities. In parallel, advances in multi-sensor fusion, machine learning methods, and systematic validation against field and inventory data will be essential for developing operational products to support sustainable forest management and international climate policy frameworks.

3.4 C12 – Tree Types

3.4.1 Introduction & Objectives

Mapping tree types is an essential step for understanding forest composition, monitoring forest health, and supporting biodiversity assessments. The focus of the C12 Tree Types prototype is the extension of the current CLMS Tree Cover & Forest portfolio to provide further information on the dominant tree species or genus (i.e. tree types). The target nomenclature and user requirements detailed in EEA (2025) and Bartaloš (2025).

Therefore, the objective of this task was to develop and test a methodology for mapping tree type in the Central European demonstration site, which was extended northwards, covering 25 Sentinel-2 tiles, in this case. The primary aim was to assess the potential of deep learning approaches, specifically the TempCNN model, to classify dominant tree species groups using Sentinel-2 time series, combined with in-situ reference data from the Global Biodiversity Information Facility (GBIF) and the state-wide forest inventory data for the state of Lower Saxony, Germany.

3.4.2 State of Play

Accurate and up-to-date tree species mapping is a critical component for effective forest management, biodiversity conservation, and climate change research. Sentinel-2, with its high temporal and spectral resolution, provides a suitable data source for species-level classification. Previous studies have demonstrated the feasibility of differentiating between major coniferous and broadleaved groups, but classification of species-level classes remains challenging due to spectral similarity and seasonal variation. Following early attempts on tree species mapping at local level (e.g. Persson et al., 2018) recent studies have shown the potential of time-series analyses (Landsat, Sentinel2) for tree species mapping at national to continental scales (Blickensdörfer et al., 2024; Hermosilla et al., 2022) if National Forest Inventory data are available for calibration and validation. For a comprehensive review of the state-of-the-art we refer to review in EEA (2025).

3.4.3 Selection of Prototypes sites

For the Tree Types prototype roll out the Central European demonstration site was selected, covering an initial area of 10 S-2 tiles, and extended northwards to leverage detailed national forest inventory data. The total study area spans 25 S-2 tiles, ensuring a heterogeneous forest landscape, being representative of Central European conditions.

The region includes both broadleaved and coniferous-dominated forests, with key species such as *Fagus sylvatica*, *Quercus* spp., *Betula*, *Populus*, *Picea abies*, and *Pinus* spp.. This selection allowed testing the methodology across diverse ecological and management contexts.



Figure 36. AOI for the Tree Types mapping, Central European demonstration site extended northwards.

3.4.4 Input and Ancillary Data & Pre-processing

The model was trained on a comprehensive dataset derived from two primary sources: the Global Biodiversity Information Facility (GBIF), which contributed with 35,810 samples, and the TreeSat dataset (Ahlsweide et al., 2023) which combines data from state and national forest inventories for the German state of Lower Saxony and provided 50,381 samples.

Table 21. List of datasets used for the C12 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Level-2A time series	10m	01.01.20 – 31.12.22
GBIF	the Global Biodiversity Information Facility	Point data	2021
TreeSat	state and national forest inventories	Point data	2021
HRL Dominant Leaf Type	HRL DLT for filtering of training data, the tree cover mask, and derivation of t the derivation of a tree cover mask	10m	yearly 2018-2021

Two main reference datasets were used:

- GBIF occurrences (35,810 samples): Derived from species occurrence records, providing broad coverage. The data has been retrieved from GBIF using the occurrences API² and filtered disregard records before the year 2000 or with known issues according the provided metadata.
- TreeSat national forest inventory (50,381 samples): Offering structured forest type information in Lower Saxony.

A crucial pre-processing step involved filtering the combined dataset to ensure high data quality and sample reliability. We removed all points which were not permanently mapped as tree cover in the HRL DLT 2018-2021 or estimated to have a canopy height <5m according to the canopy height map by Tolan et al. (2024). This process resulted in a refined training dataset of 53,750 samples.

Remote sensing data: Sentinel-2 time series for 2020–2022 were used for training, while predictions were performed for the reference year 2021. Only observations with a cloud coverage $\leq 80\%$ were considered for the processing. The three reference years are split into three partially overlapping time-series covering the respective reference year ± 3 months (i.e. October of the previous year to March of the following year).

Temporal resampling and interpolation: Valid observations were aggregated into equidistant 10-day bins using the mean value. Bins without valid observations were filled via linear interpolation over time.

Classification scheme: 13 Level-2 classes were defined, structured into three Level-1 groups:

² <https://techdocs.gbif.org/en/openapi/v1/occurrence>

- Broadleaved Evergreen (11–14)
- Broadleaved Deciduous (21–26)
- Coniferous (31–33)

Table 22. Classification scheme of the Tree Types Prototype.

Level 1	Level 2	Class code
Broadleaved evergreen	Quercus (evergreen spp.)	11
	Olea europaea	12
	Eucalyptus	13
	Other Broadleaved evergreen	14
Broadleaved deciduous	Fagus	21
	Quercus (deciduous spp.)	22
	Betula	23
	Alnus	24
	Populus	25
	Other Broadleaved deciduous	26
Coniferous	Picea	31
	Pinus	32
	Other Coniferous	33

The chart below illustrates the distribution of training samples across the different tree type classes included in the dataset. Each bar represents the total number of samples available for a given class, derived from the filtered training dataset.

The largest number of samples is found in Other Broadleaved Deciduous (11,915), followed by Fagus (9,208) and Quercus (deciduous spp.) (7,511). Coniferous classes such as Pinus (7,223) and Picea (4,293) are also well represented. In contrast, classes like Populus (802) and Other Broadleaved Evergreen (859) have relatively fewer samples.

This imbalance in the number of training samples is an important consideration for model training and evaluation, as under-represented classes may be more challenging to classify accurately. Appropriate strategies such as class-balancing or weighted loss functions may be required to mitigate the effects of sample imbalance.

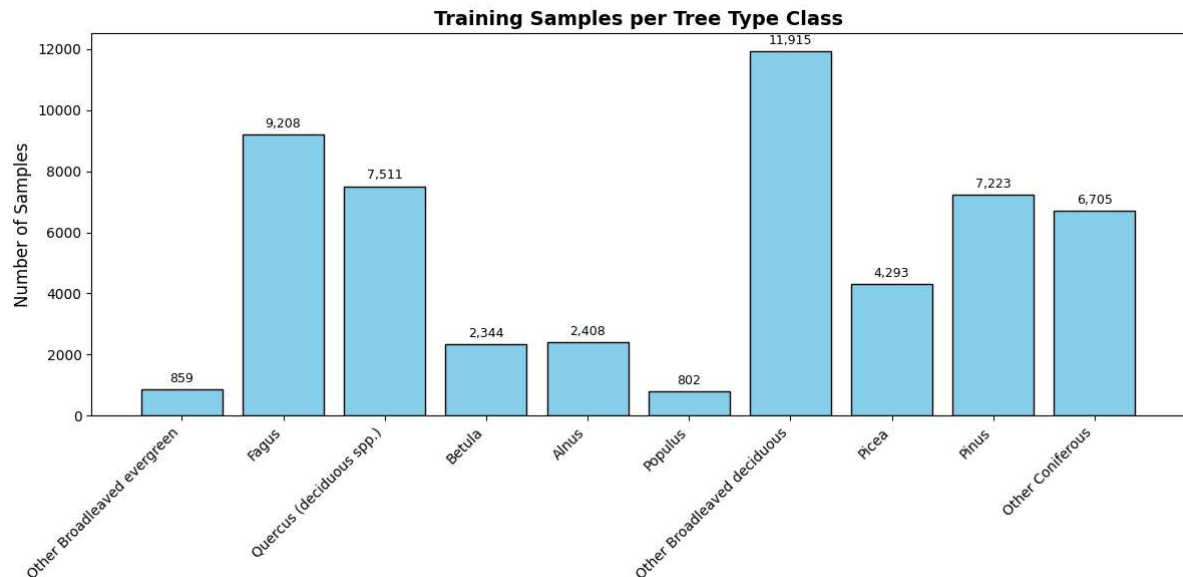


Figure 37. Distribution of training samples per tree type class in the dataset.

3.4.5 Methodology and Workflow

A deep learning approach was applied using a Temporal Convolutional Neural Network (TempCNN). TempCNN was employed for the classification task, as it is designed to analyse sequential data by applying convolutional filters across the temporal dimension.

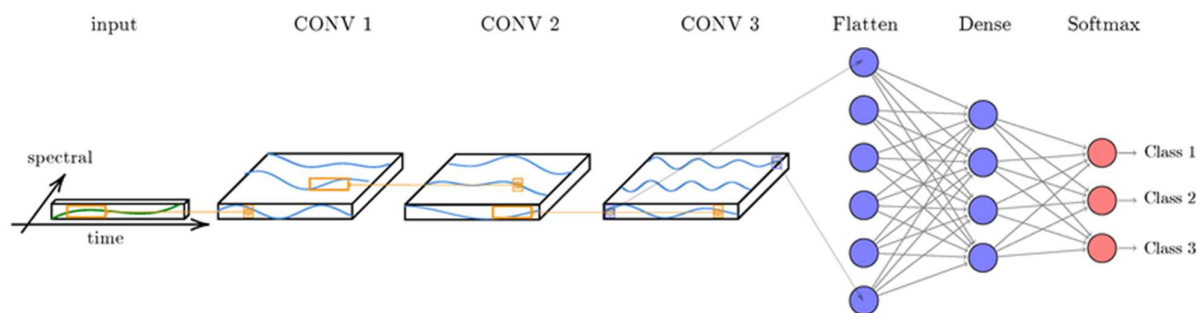


Figure 38. Schematic representation of a TempCNN (Pelletier et al., 2019).

The model was trained using the pre-processed, filtered samples, learning to recognize the unique spectral and temporal signatures of each tree type throughout the timeseries 2020-2022.

The workflow consisted of:

1. Data preparation: Extraction of Sentinel-2 time series per sample location; application of filtering and time-series interpolation.
2. Class definition: Mapping of reference data into 13 target Level-2 classes.

3. Model training and evaluation: TempCNN is trained with 80% of the available samples, 20% are retained as a test set for accuracy assessment.
4. Prediction and post-processing: Model outputs for the reference year 2021 are post-processed including spatial smoothing and harmonization with the HRL DLT layer to ensure consistency with the dominant leaf type information.

3.4.6 Results and Verification

The model's performance was evaluated using standard classification metrics on both the test and validation datasets. Following is the model classification report:

Test Report Analysis:

Table 23. Test set accuracies for the Tree Types classification.

a	precision	recall	f1-score	support
Fagus	0.748	0.670	0.707	2078
Quercus	0.815	0.834	0.824	1670
Betula	0.643	0.456	0.533	498
Alnus	0.698	0.625	0.660	493
Populus	0.600	0.092	0.159	98
Broadleaved	0.683	0.750	0.715	2572
Picea	0.767	0.802	0.784	907
Pinus	0.836	0.924	0.878	1654
Coniferous	0.831	0.825	0.828	1493

accuracy			0.763	11463
macro avg	0.736	0.664	0.676	11463
weighted avg	0.760	0.763	0.758	11463

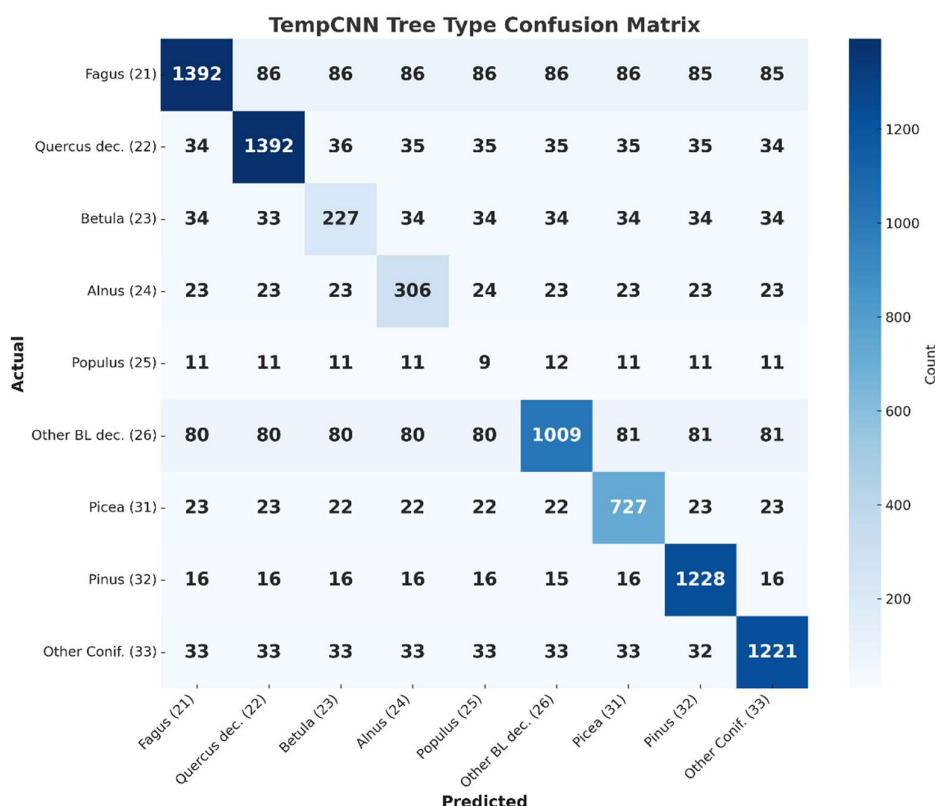


Figure 39. TempCNN test set confusion matrix for tree type classification (counts per class, synthetic based on recall and support).

The model achieved an overall accuracy of 76.3% on the test set. The weighted average f1-score was 0.758, indicating a strong performance across classes proportional to their sample size. Notably, classes such as Quercus (deciduous spp.) (Code 22) and Pinus (Code 32) performed exceptionally well, with f1-scores of 0.824 and 0.878 respectively. This suggests the model is highly effective at distinguishing these dominant species.

However, some classes showed lower performance. Populus (Code 25) had a very low f1-score of 0.159, likely due to the small number of samples in the test set (98). Similarly, Betula (Code 23) had a lower f1-score of 0.533. These results highlight the impact of sample size on model performance for less common tree types.

Classification performance:

- Overall accuracy on the test set: 76.3%
- Weighted F1-scores on test set: 0.758

Class-wise observations:

- High performance: Pinus (32, F1 = 0.878) and Quercus deciduous spp. (22, F1 = 0.824).

- Moderate performance: *Fagus* (21, F1 = 0.707), *Alnus* (24, F1 = 0.660), Other broadleaved deciduous (26, F1 = 0.715).
- Low performance: *Populus* (25, F1 = 0.159), with very low recall (0.092), indicating confusion with other deciduous classes.
- Conifers overall: Performed well, particularly *Pinus* and Other Coniferous, while *Picea* showed slightly lower performance.

The final tree type map for 2021 was harmonized with the HRL DLT layer. This post-processing step ensured that the predicted tree types were consistent with the dominant leaf type information across forests. Harmonization particularly improved the spatial coherence of minority classes and reduced misclassifications along forest edges

These results demonstrate that implemented approach is effective at distinguishing broad species groups and some dominant genera (e.g., *Fagus*, *Pinus*, *Quercus*), but struggles with minority classes such as *Populus*. Low accuracy on classes which are poorly represented in the training data is a commonly known issue for machine learning algorithms and in line with findings by other studies on tree type mapping.

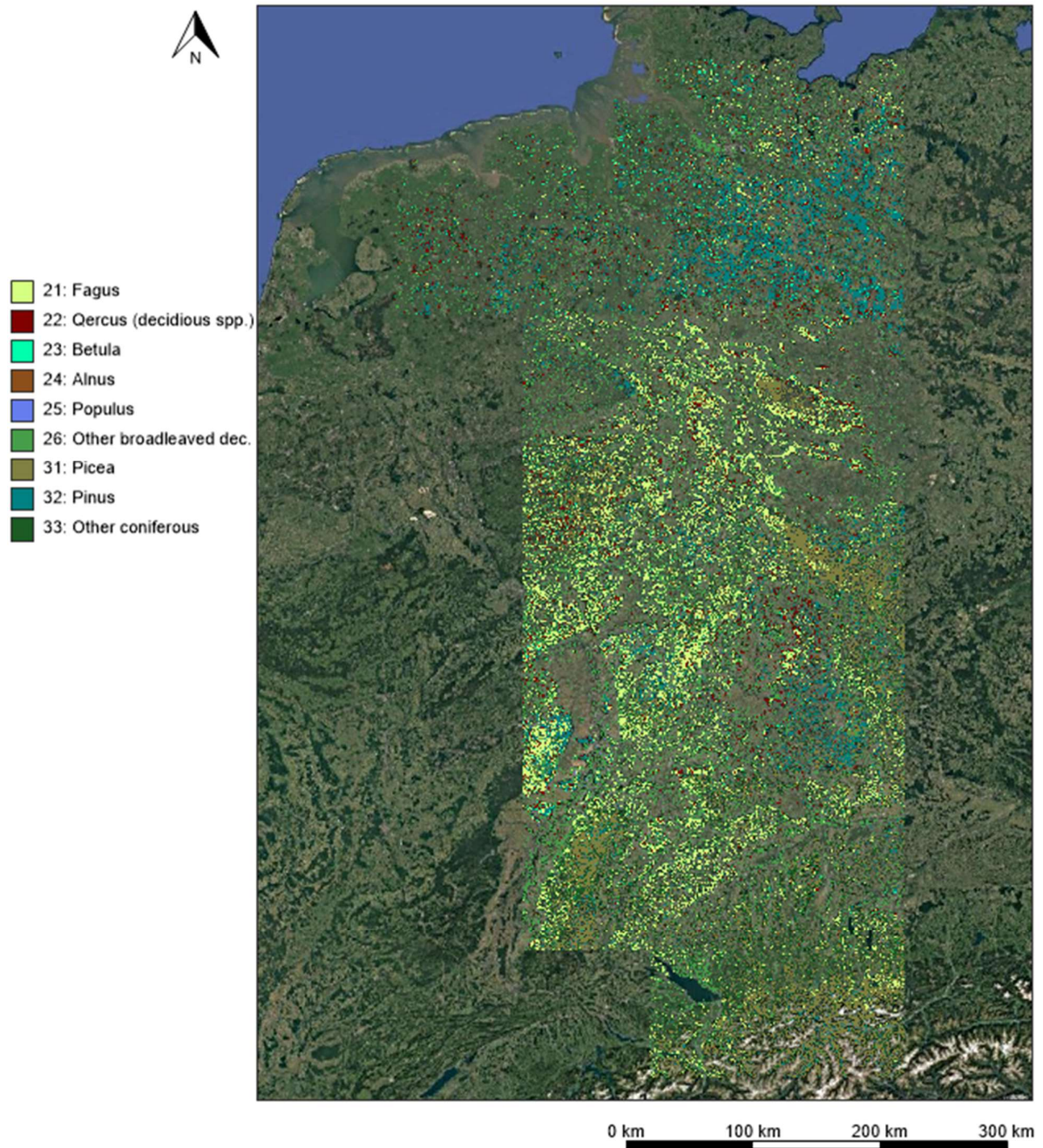


Figure 40. Tree Types mapping - final result for the study area.

Zoom-in analysis of the final map product illustrates that the classification successfully delineates small forest patches and minority species within the demonstration site. The remaining classification uncertainties are primarily concentrated in areas with mixed-species stands and highly fragmented canopy boundaries.

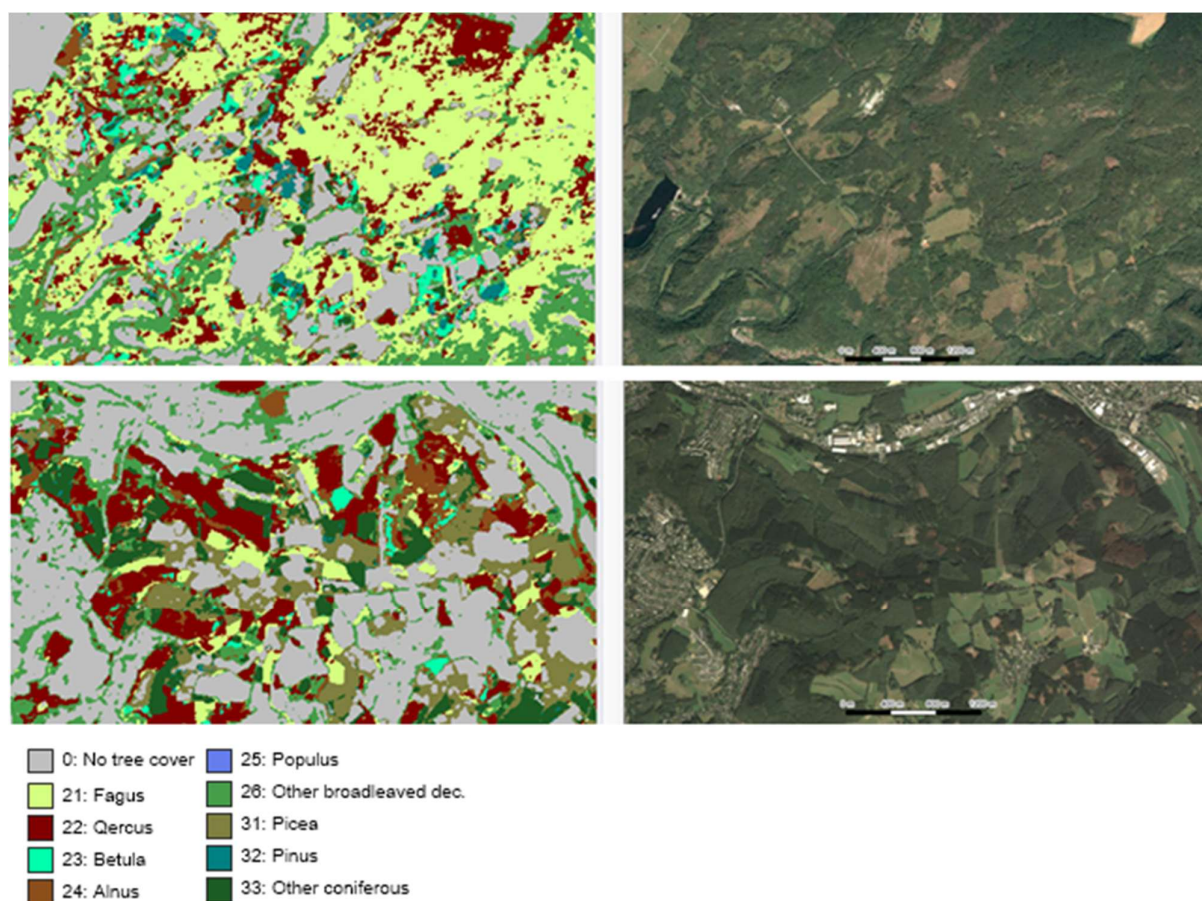


Figure 41. Detailed zoom-in of Tree Types mapping in the demonstration site (left) compared with true-colour satellite imagery (right).

3.4.7 Conclusions

The proposed approach showed strong potential for tree type classification with Sentinel-2 time series. The integration of diverse data sources, namely GBIF and TreeSat, combined with a rigorous filtering process, was a key factor in achieving an overall accuracy of 76.3% on the test set. While the model excels at classifying common species such as *Quercus* and *Pinus*, performance for less-represented classes is limited by the number of available training samples. Classification performance was particularly strong for dominant species with distinct phenological signatures, including *Pinus* and deciduous *Quercus*, whereas smaller classes which are often also mixed with other species suffered from confusion. The overall accuracy values of around 76% are therefore promising for operational applications in Central Europe, but class imbalance remains a significant challenge, especially for underrepresented species such as *Populus* and *Betula*.

3.4.8 Outlook beyond the EvoLand project

Looking beyond the current scope, several opportunities exist to further enhance the model's capabilities. Expanding the training dataset, particularly for

underrepresented classes, would likely improve classification performance for those species. Further efforts should focus on improving the handling of minority classes through advanced sampling strategies or data augmentation, and on integrating additional data sources such as Sentinel-1 SAR, LiDAR, or hyperspectral imagery to improve class separability. Another important step will be to assess the transferability of the trained model across different European regions in order to evaluate its generalization capacity. In addition, further exploitation of temporal dynamics, including phenology and multi-year time series, could enhance discrimination between species. Finally, the operational implementation of tree type mapping within European monitoring frameworks would provide valuable support for forestry, biodiversity conservation, and climate reporting. The successful application of this TempCNN-based approach can thus serve as a foundation for developing a scalable, automated system for tree species monitoring across broader geographical areas.

4 Agriculture prototypes

The following chapter presents a comprehensive overview of the three agricultural CLMS prototypes of Task 3.2, which include: i) T 3.2.1 Cover crop type mapping, ii) T3.2.2 Cropland/grassland primary production, and iii) T3.2.3 Small landscape features mapping.

4.1 C4 – Cover Crop Type

4.1.1 Introduction & Objectives

Cover crops (or catch crops) are cultivated primarily to enhance and protect the soil rather than for harvest. They are typically planted during intervals between main crop cycles when the soil might otherwise remain bare. Common cover crop types include grasses (e.g., rye, oats), legumes (e.g., clover, vetch), and brassicas (e.g., radishes, mustard).

Cover crops offer numerous benefits, including reducing soil erosion and nutrient leaching, increasing soil organic matter, providing natural fertilization that lessens the need for synthetic fertilizers, and enhancing biodiversity in agricultural systems. Recognized as a key practice in sustainable agriculture by the Common Agricultural Policy (CAP), the implementation of cover crops requires effective monitoring.

While various operational methods and services have been established for mapping and monitoring main crop seasons, few focus on cover crop mapping. EO-based cover crop mapping and monitoring present specific challenges, such as the scarcity of reliable reference data. Additionally, the higher likelihood of clouds and snow during the secondary growing season (often winter) complicates the use of optical image time series for crop mapping. Furthermore, many cover crop seed mixes exhibit similar temporal patterns, making them difficult to differentiate (Schulz et al., 2021).

The objective of this candidate is to develop a prototype service to map cover crop types at a 10-meter resolution on an annual basis, using novel collected in-situ data as well as a hierarchical set of inter-connected Machine Learning models based on Sentinel-1 and Sentinel-2 imagery. Table 24 shows the specifications of the C4 prototype:

Table 24. Technical and thematic specifications for the C4 prototype.

Prototype C4	
Prototype	Crop cover type mapping
Thematic specifications	Map cover crop types at 10m resolution utilizing novel collected in-situ data on cover crop types as well as hierarchical set of inter-connected Machine Learning models based on Sentinel-1 & 2.
Prototype sites	Belgium (3)
Temporal frequency	Yearly
Spatial resolution	10 meters
MMU	1 pixel
Envisaged accuracy	80%

Due to the complexity of achieving high accuracy and detailed thematic mapping, identifying the 10 most important cover crop types for each country (as stated in the requirements) may not be feasible. Instead, we aim to differentiate cover crop groups with the highest possible thematic detail while maintaining reasonable accuracy. Figure 42 illustrates a proposed hierarchical typology, where cover crop types are initially grouped by plant family (e.g., Leguminosae, Brassicaceae). These families are then categorized by morphological or phenological similarities into "grass-like" and "leaf-rich" groups, with a third group comprising mixes of green covers.

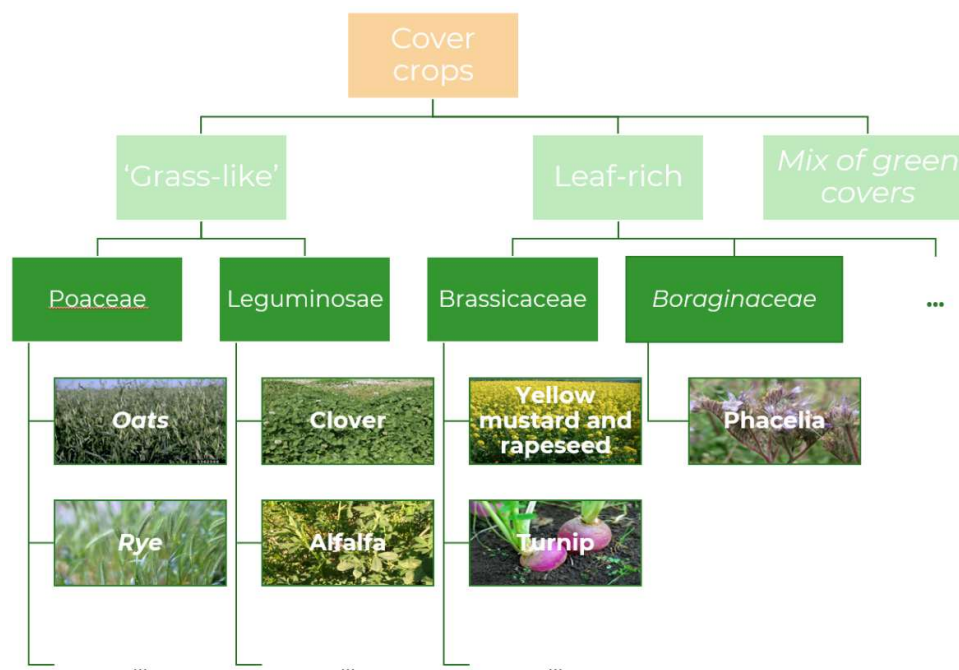


Figure 42. Hierarchical cover crop typology with cover crops grouped by family and then by morphological or phenological similarities.

4.1.2 State of Play

The new CLMS HRL Vegetated Land Cover Characteristics (VLCC) component already includes specific “secondary crop” layers, that provide parcel-based information on the presence and seasonality of cover crops. However, it does not differentiate between cover crop types. The goal of this candidate is to enhance the existing VLCC product by providing more detailed information on types or groups of cover crops.

The HRL VLCC products already meet the requirements for a cover crop presence flag and a minimum mapping unit of 1 hectare. Further requirements include that the product should at a minimum identify the 10 most important cover crops in each country or region, have an accuracy of at least 85% and target the highest possible precision. As explained in section 4.1.1., the thematic detail of the prototype needs to be adapted in function of the separability of cover crop types and groups, in order to reach the required accuracy and precision.

Additional requirements state that the product should cover the entire EU and be available at least annually. Although the current focus is on one test site and one growing season, the annual and EU-wide availability of input data (Sentinel-2 and HRL VLCC) makes upscaling theoretically feasible. However, the main challenges for method transfer will be the lack of reference data and regional differences in cover crop types.

4.1.3 Selection of Prototypes sites

For this prototype, the focus was set on demonstration site 3, located in Belgium. The choice was based on the availability of cover crop type information in the Flemish Land-Parcel Identification System (LPIS) dataset, as well as in a small in-situ dataset collected in the same region.

4.1.4 Input and Ancillary Data & Pre-processing

The HRL VLCC secondary crop type (CPSCT) product is the baseline for this prototype. It identifies parcels with a secondary crop season and assigns them to four seasonal categories but does not distinguish between cover crops and next-season winter crops. To flag winter crops and inconsistencies, and extract actual cover crop parcels, the CPSCT product is pre-processed through a rule-based procedure using four additional HRL VLCC products:

- Secondary crop emergence date (CPSCE)
- Secondary crop season duration (CPSCD)
- Main crop type of the next season (CTY)
- Main crop emergence date of the next season (CPMCE)

The resulting flagged CPSCT product is then polygonised to generate the final cover crop parcels for classification.

The classification of cover crop parcels in different types and groups is based on Sentinel-2 Normalized Difference Vegetation Index (NDVI) time series. The images are pre-processed and the NDVI time series are extracted using openeo. For a given year, all Sentinel-2 L2A images from July to May of the next year are collected. The images are masked using a dilated Scene Classification Layer. The NDVI is computed, then aggregated and interpolated to a 10-days interval time series. Finally, the NDVI time series are aggregated spatially, taking the mean value per parcel.

The reference dataset for training and validation is derived from farmer declarations in the Flemish LPIS, which includes data on “late crops”. Since the parcel boundaries of the LPIS dataset do not match those in VLCC, an intersection of both parcel delineations is used to build the reference dataset. The 189 different types of crops declared as “late crops” in the LPIS are reclassified according to the typology described above. Only parcels with a cover crop (both according to LPIS and VLCC) are retained.

To obtain a balanced dataset, a stratified random sampling design is applied. Stratification is implemented in two levels. First, samples are distributed equally across the three cover crop groups (“grass-like,” “leaf-rich,” and “mix of green covers”). Within each group, we sample 50% of each cover crop type family (e.g., *Poaceae*, *Leguminosae*) with the fewest parcels. Additional samples are then drawn from the larger crop family until the group reaches 2,000 parcels in total. The resulting cover crop reference dataset is finally split into training (75%) and validation (25%) subsets using stratified random sampling, with the cover crop groups as strata.

The input and ancillary data used for this prototype are listed in Table 25 and shown, together with the pre-processing steps, in the workflow Figure 43.

Table 25. Input and ancillary data for the C4 cover crop type map.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Sentinel-2 L2A spectral bands and NDVI	10 - 20m	Aggregated and interpolated to 10-days intervals
CLMS HRL VLCC	CPSCT – Secondary Crop Type (presence/absence) CPSCE – Secondary crop emergence date CPSCD – Secondary crop season duration CTY – Main crop type CPMCE – Main crop emergence date	10m	Yearly
LPIS Flanders	Farmer declarations on “Late crop” types	-	Yearly

4.1.5 Methodology and Workflow

The complete workflow for the cover crop type prototype is shown in Figure 29. This workflow can be divided in two main components, namely (1) model training and validation (yellow and purple the workflow) and (2) inference and post-processing (green in the workflow). Both components involve an image pre-processing and feature extraction step which is done in OpenEO, as described in previous section (blue frame in the workflow).

First, the cover crop type classification model is trained and validated using the LPIS-based reference dataset. A 4-fold cross-validation approach is applied to train a CatBoost³ model. The classification accuracy is assessed using the validation dataset, that was not used for training. The outputs of this component are a cover crop type classification model and performance metrics.

Secondly, the trained model is used to classify cover crop parcels extracted from the HRL VLCC CPSCT product, as described in previous section. The classified parcels are rasterized and combined with the flagged HRL VLCC CPSCT layer to generate the final cover crop type map. The final map legend includes the classified cover crop types as well as classes and flags from the HRL VLCC CPSCT product. The “non crop”, “no secondary crop” and the “HRL VLCC quality flag” are directly inherited from the CPSCT product, while the “winter crop flag” and inconsistency flag” are obtained through the rule-based pre-processing step.

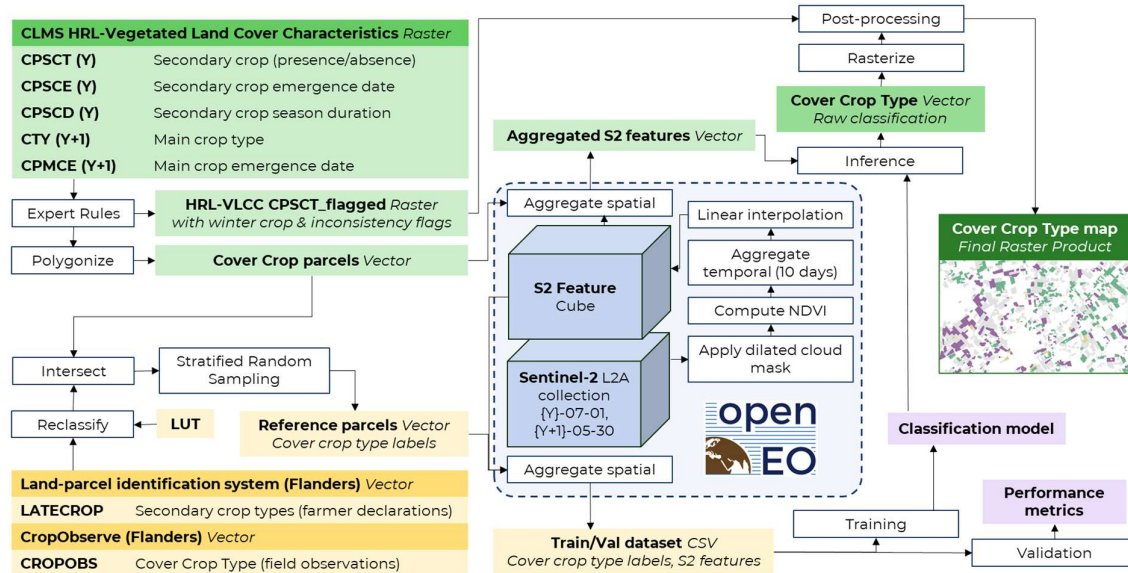


Figure 43. Workflow for the cover crop type candidate.

³ “CatBoost is a machine learning algorithm that uses gradient boosting on decision trees. It is available as an open source library” - [CatBoost](https://catboost.ai/)

4.1.6 Results and Verification

The separability of different levels of the hierarchical typology (Figure 42) was evaluated during the development phase of this prototype. To reach the required classification accuracy, the final cover crop type map is limited to two main classes: “grass-like” and a combined “leaf rich & seed mixes” class.

Several feature combinations were also tested during development, including Sentinel-2 spectral bands and NDVI time series and temporal statistics, as well as Sentinel-1 features. The best performances were obtained with 10-days interpolated NDVI times series, which was therefore retained for the final production workflow.

The workflow was initially developed on the 2020 growing season. The classification performances were evaluated using the validation dataset obtained by stratified random sampling for the Flemish LPIS, as described in section 4.1.4. The 1500 validation samples (500 “grass-like” and 1000 “leaf-rich & mix”) are distributed representatively across Flanders (Figure 44). No samples were available in Wallonia (the southern region of Belgium), so the prototype could not be evaluated there.

The confusion matrix and performance metrics of the 2020 cover crop type map of Belgium are shown in Table 26. The two groups of cover crops show a relatively good separability, with F1-scores of 0,73 and 0,85 for “grass-like” and “leaf-rich & mix”, respectively. The overall accuracy is 81%. Figure 45 shows the cover crop type map generated for the 2020 growing season over Belgium.

A new classification model was trained and validated for the 2019 growing season. The training and validation dataset was generated based on the 2019 LPIS dataset, following the same procedure as for 2020. The performance metrics (Table 27) are very close to those obtained in 2020, with an overall accuracy of 81%, showing a good transferability of the approach across growing seasons.

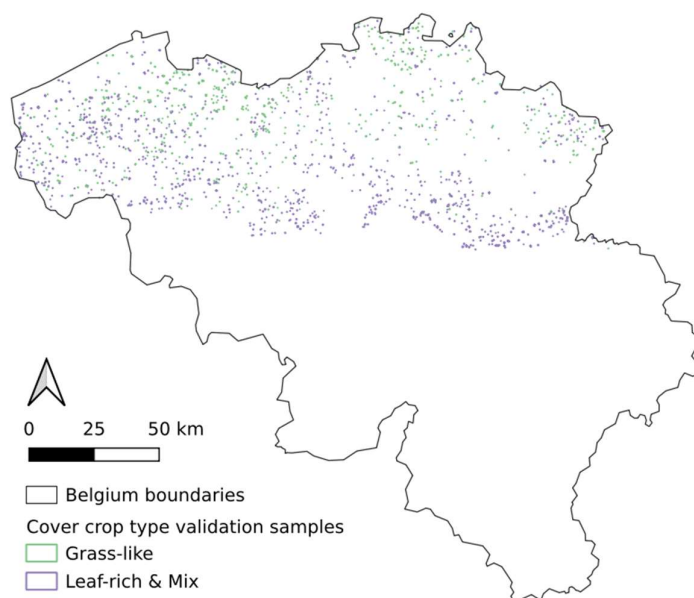


Figure 44. Locations of the 2020 cover crop type validation samples across Flanders, Belgium.

Table 26. Confusion matrix and performance metrics of the 2020 cover crop type map of Belgium., validated with the LPIS dataset.

		REFERENCE		Total	User Acc.	F1 score
		Grass-like	Leaf-rich & Mix			
MAP	Grass-like	397	186	583	68%	0,73
	Leaf-rich & Mix	103	814	917	89%	0,85
	Total	500	1000			
	Prod. Acc.	79%	81%			
		Overall Acc.	81%			

Table 27. Confusion matrix and performance metrics of the 2019 cover crop type map of Belgium, validated with the LPIS dataset.

		REFERENCE		Total	User Acc.	F1 score
		Grass-like	Leaf-rich & Mix			
MAP	Grass-like	371	161	532	70%	0,71
	Leaf-rich & Mix	129	839	968	87%	0,85
	Total	500	1000			
	Prod. Acc.	74%	84%			
		Overall Acc.	81%			

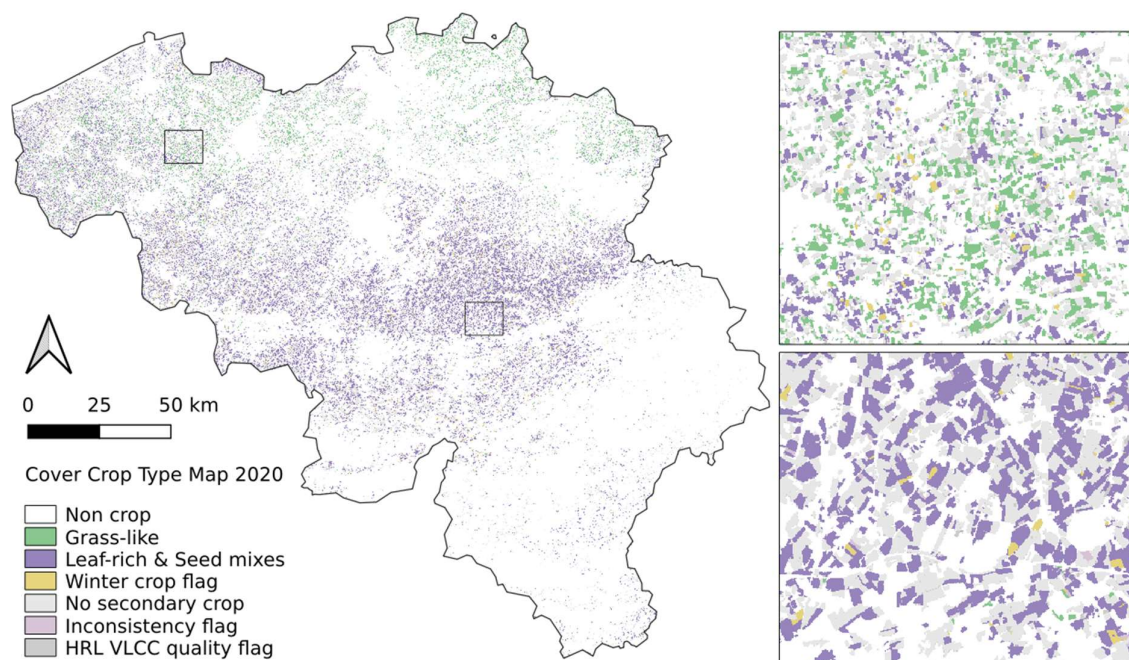


Figure 45. Cover crop type map of Belgium (2020).

4.1.7 Conclusions

The cover crop type prototype successfully differentiates two main groups of cover crops using Sentinel-2 NDVI time series, with the CLMS HRL VLCC secondary crop products as a baseline. An overall accuracy of 81% was achieved across two growing seasons in Belgium, demonstrating good temporal transferability and highlighting the potential of EO data for cover crop type mapping.

At the same time, several limitations currently hinder further development and broader application of this prototype. The most critical challenge is the lack of reliable reference data. LPIS datasets from Member States rarely include detailed cover crop information, and even when available, as in the Flemish LPIS used here, secondary crop declarations carry inherent uncertainties. This significantly constrains the development of robust classification models.

Another challenge lies in defining a cover crop typology that achieves sufficient separability. In this study, only two broad groups could be distinguished. Cover crops are often composed of seed mixes from multiple plant families (e.g., Poaceae, Legumes, Brassica), which complicates classification and introduces confusion. For example, in the Flemish LPIS, 45% of declared 'late crops' are listed as a 'Mix of non-flowerly green covers.'

4.1.8 Outlook beyond the EvoLand project

The Belgian test site results provide valuable insights and demonstrate that cover crop type mapping from EO data is feasible. While the current prototype remains at a relatively low readiness level, it establishes a foundation for further development. The method showed good temporal transferability within the test region, but broader application and scaling will depend on addressing the limitations identified.

Most importantly, the availability of reliable reference data will determine the potential for further operationalizing and scaling. Larger, higher-quality datasets would enable more detailed typologies, improved accuracy and the development of robust cover crop classification models. Additional field data and agronomic expertise will also be essential to ensure transferability across regions.

Given the availability of Sentinel-2 imagery and the CLMS HRL VLCC products, the potential exists to generate cover crop type maps annually across the EU—provided sufficient reference data can be gathered. With continued investment in data acquisition and method development, this candidate service could evolve into a valuable operational product supporting sustainable agriculture monitoring.

4.2 C5 – Grassland and Cropland GPP

4.2.1 Introduction & Objectives

Information on the state and the evolution of agricultural productivity can be a relevant input for a wide variety of agricultural applications, including crop yield estimation, weather related impact on production and measuring carbon fluxes.

The current operational Copernicus CLMS Gross Primary Productivity (GPP) products at European level are based on images from medium resolution satellites (e.g., Sentinel-3, Modis) which is often not sufficient for monitoring productivity dynamics at field level. Furthermore, the parameters and algorithms used are not specifically adjusted to capture the productivity dynamics for agricultural land like cropland and grassland. Therefore, within the EvoLand project, the aim is to address the beforementioned limitations.

Table 28. Technical and thematic specifications for the C5 prototype

Prototype C5	
Prototype	Grassland and Cropland GPP
Thematic specifications	GPP estimation at 10 m resolution over cropland and grassland locations based on Sentinel-2
Prototype sites	Belgium (3)
Temporal frequency	10 m Dekad (10-daily)
Spatial resolution	10 meters
MMU	1 pixel

Envisaged accuracy	TBD
--------------------	-----

4.2.2 State of Play

The Copernicus CLMS Gross Primary Productivity (GPP) products at European level that is currently operational is at 300 m resolution and decadal scale⁴. There is also a 1 km GPP product available from 1999-2020⁵. Both products use the simplified Light Use Efficiency (LUE) concept as described by Monteith (1972) to model GPP. For the development of the prototype this same GPP model as used in the current GPP CLMS products was used. However, the model was adjusted to estimate GPP at 10m resolution and improve GPP estimations for grassland and cropland (see section 4.2.5 for further details). The current prototype addresses the following user requirements which were defined in WPI: the prototype allows to determine management practices and harvesting events of crops, has a resolution of 10m, has the highest possible precision, is updated every 10 days and can be generated for the entire EU territory.

4.2.3 Selection of Prototypes sites

For the development of this final prototype demonstration site 3, located in Belgium was selected. The current version of this prototype was produced at tile level, tile 31UFS (available in the Copernicus Data Space Ecosystem for 2018 and 2022).

4.2.4 Input and Ancillary Data & Pre-processing

As satellite data, Sentinel-2 derived fAPAR was used. The fAPAR was smoothed using a Whitakker smoother. To determine the amount of incoming radiation that plants can use for photosynthesis (R , equation 1) and temperature dependency of photosynthesis (E_t , equation 1) and the CO₂ fertilization effect (ECO_2 , equation 1) mean temperature and incoming solar radiation from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land were used. All these input data were retrieved for a bounding box of 30x30 m centred around the flux sites to cover partly the sensitivity of the flux measurements according to the wind direction for the calibration and validation of the GPP model. Furthermore, all data was aggregated to 10-daily scale to align with the requirement to produce a 10-daily GPP product. Currently, WorldCover⁶ was used to identify grass and cropland pixels.

21 ICOS towers located in crop or grassland were selected for the prototype GPP model calibration and validation. The ICOS towers have GPP data available

⁴ <https://land.copernicus.eu/en/products/vegetation/gross-dry-matter-productivity-v1-0-300m>

⁵ <https://land.copernicus.eu/en/products/vegetation/dry-matter-productivity-v2-0-1km>

⁶ <https://esa-worldcover.org/en>

between 2017-2022 (note, number of years varies from tower to tower). Table 1Table 29 provides a list of the datasets used in the prototype development.

Table 29. list of datasets used for the prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Sentinel-2 L2A used for fAPAR calculation, smoothed with Whitaker smoother	10 m	Aggregated and interpolated to 10-days intervals
ECMWF ERA5-Land	Re-analysis of meteorological data from ECMWF	+/-10 km	daily
ICOS fluxnet	GPP measurements from ICOS flux towers across Europe	/ (point)	30 minutes

4.2.5 Methodology and Workflow

The same Light Use Efficiency model as used for the CLMS GPP products was used to model GPP. However, the model was adjusted to improve resolution (from 300m to 10m) and performance in grassland and cropland. The adjustments applied are described below.

$$GPP = R * E_c * fAPAR * E_{LUEC} * E_t * E_{CO2} * E_{drought} * cf \text{ (equation 1)}^7$$

With (see

Table 30)

⁷ <https://land.copernicus.eu/en/technical-library/algorithm-theoretical-basis-document-dmp-gdmp-npp-gpp-version-1/@@download/file>

Table 30. Overview of different parameters used in the current CLMS GPP model.

Term	Meaning	Value	Unit
GPP	Gross Primary productivity , GPP=0.5 GDMP	0 – 32	gC/m ² /day
R	Total shortwave incoming radiation (0.2 μm – 3.0 μm)	0 – 320	GJ _T /haday
E _c	Fraction of PAR (0.4 μm – 0.7 μm)	0.48	J _p /J _T
FAPAR	PAR-fraction absorbed (PA) by green vegetation	0 – 1	J _{Ap} /J _p
E _{LUec}	Maximum light use efficiency (DM=Dry matter) under optimal conditions	Land use specific	kgDMGJ _{Ap}
E _t	temperature effect	0 - 1	-
E _{CO2}	Normalized CO ₂ fertilization effect, the fertilization effect in T and CO ₂ dependent (see ⁸)	> 0	-
E _{drought}	Drought stress factor	0-1	-
cf	conversion factor between dry matter per ha and gram carbon per m ²	0.05	

Temperature stress E_t

The complex T stress factor in the original 300m Copernicus GPP model was replaced by a simpler one.

$$E_t = \begin{cases} 0 & \text{if } T_{mean} < 273 \\ 1 & \text{if } T_{mean} \geq 273 \end{cases}$$

Drought stress E_{drought}

Drought stress factor terms is calculated as follows.

⁸ <https://land.copernicus.eu/en/technical-library/algorithm-theoretical-basis-document-dmp-gdmp-npp-gpp-version-1/@@download/file>

$$Edrought = \begin{cases} q(\partial - \partial^*)^2 + 1 & \text{if } \partial \leq \partial^* \\ 1 & \text{if } \partial > \partial^* \end{cases}$$

$$q = \frac{2 * (p_0 + p_1 \left(\frac{AET}{PET} \right)_{mean\ year} - 1)}{\partial^*}$$

With:

- $p_0=0.179$, $p_1=0.45$;
- ∂ =volumetric soil water content [0-7 cm] from ERA5-LAND, $\partial^*=0.45$ (0.3-0.6 evaluated), Volumetric soil water content from different layers tested, [0-7 cm] performed best
- for AET/PET the mean yearly value is taken

Compared to the currently operational 300 m GPP product in CLMS the presented prototype has following adjustments:

- the spatial detail of the GPP product is improved by using fAPAR derived from the Sentinel-2 constellation. This results in GPP estimation of 10 m resolution
- a year specific CO₂ fertilisation term (i.e. ϵ_{CO_2} , equation 1) is used
- the crop and grassland specific maximum light use efficiency (i.e. ϵ_{LUEc} , equation 1) is recalibrated using all the available ICOS tower measurements
- The temperature stress term is adjusted to a simple temperature stress factor
- A drought stress factor based on soil water content, actual and potential evapotranspiration is included

These adjustments were tested and developed in wp2. The entire pipeline for GPP estimation developed in WP2 (see [AD7]) was embedded in OpenEO⁹ (Figure 46), this eases scalability. Four input datasets form the basis for the GPP calculation: ERA5-Land which provides meteorological information, WorldCover which was used to localize grassland and cropland, Sentinel-2 which was used to derive fAPAR, and yearly CO₂ measurements which was used for the calculation of the CO₂ fertilization effect (i.e. ϵ_{CO_2} , equation 1). In the pipeline the theoretical maximum GDMP (i.e. GDMP max in Figure 46) based on meteorological conditions was first calculated. The GDMP max was first calculated on a daily scale and then aggregated to dekad scale to align with Sentinel-2 fAPAR data. In a second step in the pipeline the GDMP max was combined with fAPAR to derive the actual GDMP which was converted to GPP by applying a scaling factor. fAPAR was derived from Sentinel-2 and was cloud-masked, composited and interpolated on a decade scale.

⁹ <https://openeo.org/>

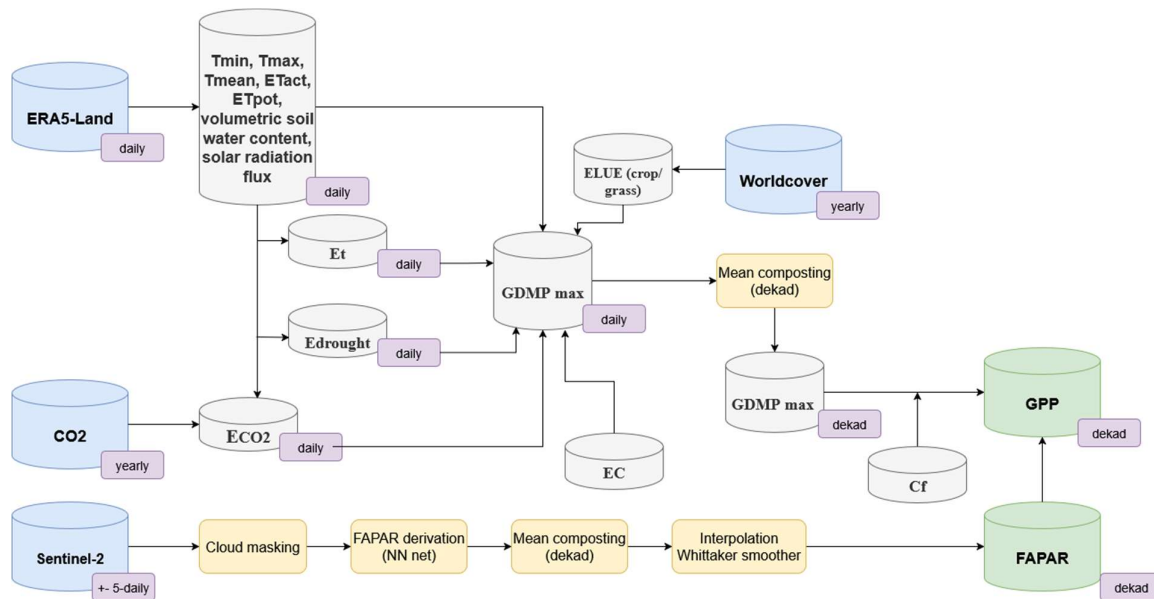


Figure 46. GPP pipeline for grassland and cropland as integrated in OpenEO.

4.2.6 Results and Verification

The prototype was tested on tile 31UFS for 2022 (see snapshot Figure 47). The prototype provides much more spatial detail compared to the coarse resolution 300m CGLOPS GPP product, furthermore also fewer temporal gaps compared to CGLOPS could be noticed.

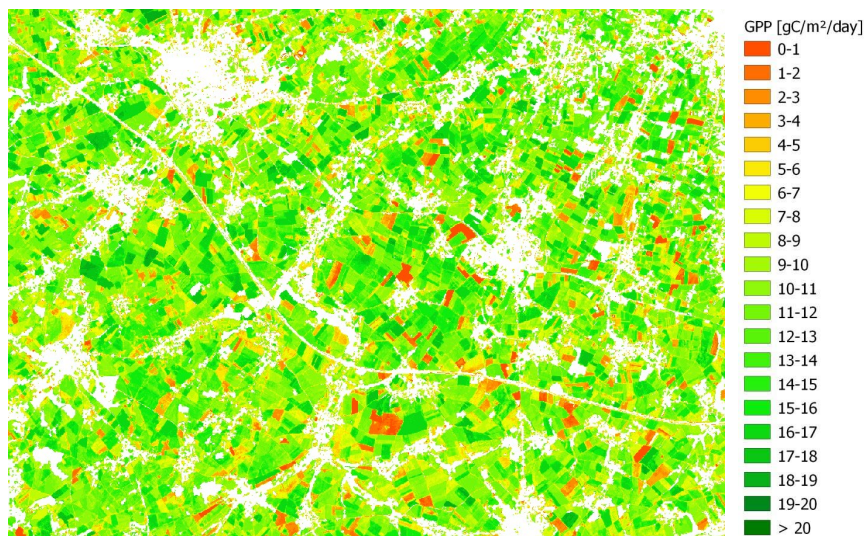


Figure 47. Snapshot (zoomed) of GPP estimation of 31UFS tile for 2022.

The current version of the prototype was validated using the ICOS tower data and the CGLOPS 300m GPP product (Table 31).

For the calibration of the maximum light use efficiency coefficient (ELUEC, equation 1) the ICOS tower data was split into a calibration (75%) and validation (25%). The results from the validation are given below for grass and cropland (Figure 48). To evaluate the results of the prototype the original GDMP model was also run for the validation towers, these results are presented in Figure 49. Overall, the precision was rather high, with especially for grassland more reliable results. Some deviations from the 1:1 line are still observed. For both grassland and cropland there is still an overestimation of the GPP by the model. In Figure 50 the bias is plotted in function of the month of the year; this figure indicates that the overestimation of the GPP by the model is mainly in the summer months. Though, in comparison with the original GDMP model the final prototype model clearly decreases the negative bias in springtime for cropland and reduces the positive bias in the summer months for both grassland and cropland (Figure 50 versus Figure 51).

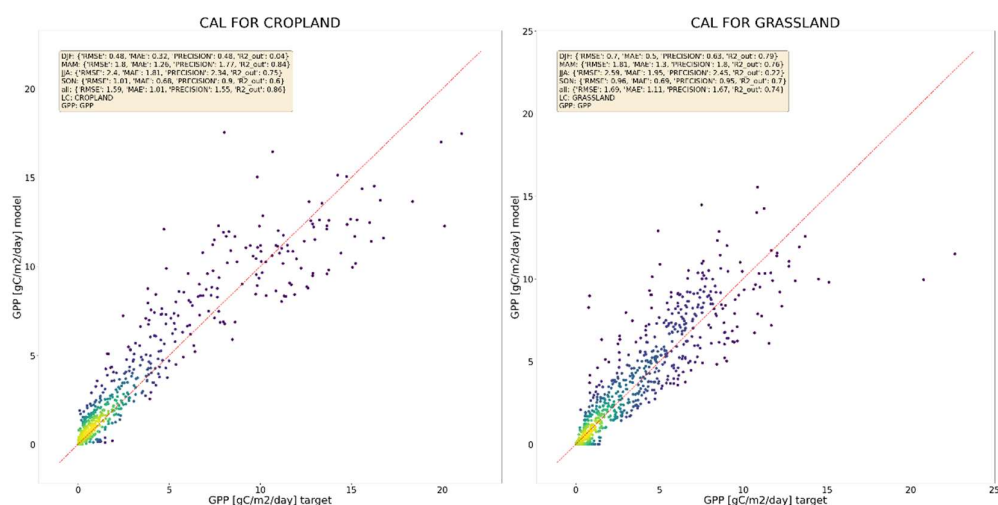


Figure 48. Scatterplot of GPP target (ICOS tower validation data) and modelled GPP for cropland (left panel) and grassland (right panel), using the GPP model as presented in equation 1.

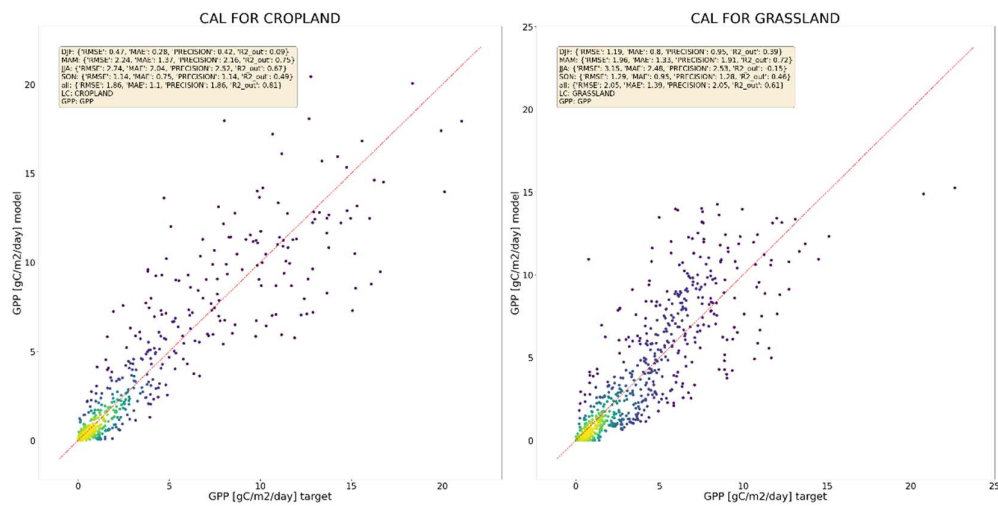


Figure 49. Scatterplot of GPP target (ICOS tower validation data) and modelled GPP for cropland (left panel) and grassland (right panel) using the original GDM model.

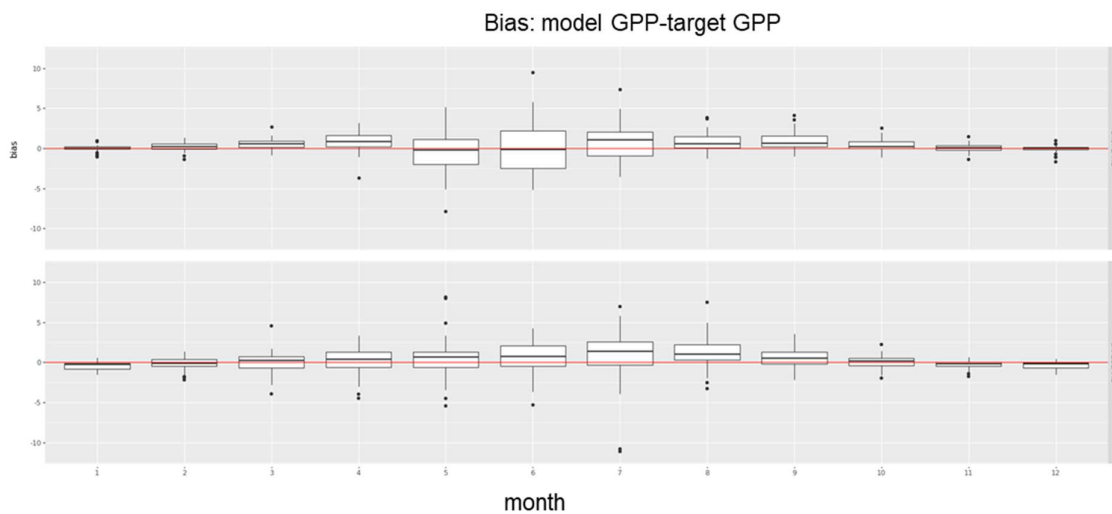


Figure 50. Bias in function of month of year, using the GPP model as presented in equation 1.

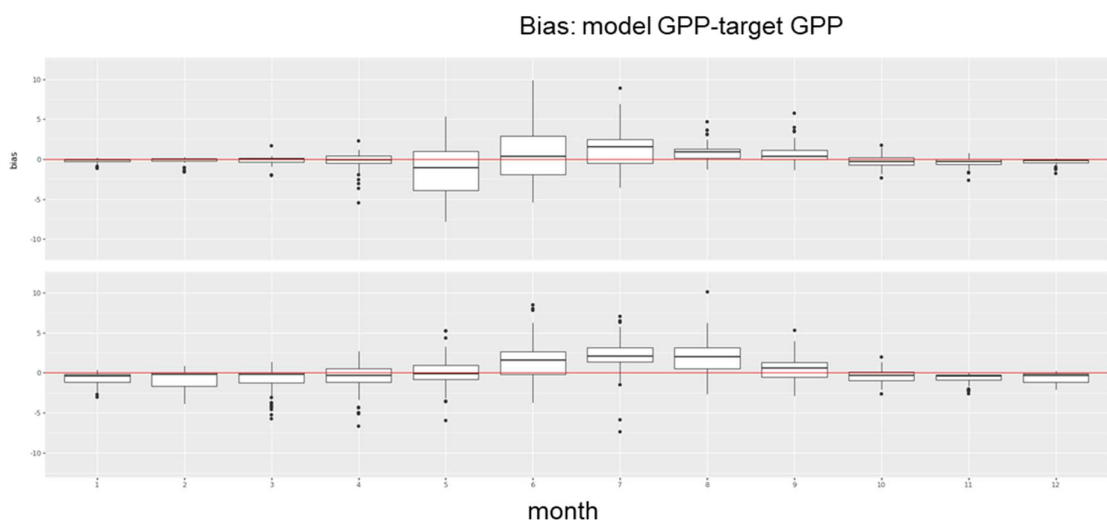


Figure 51. Bias in function of month of year, using the original GDMP model.

Table 31. Validation metrics for the first version of the cropland and grassland GPP prototype. The prototype was validation using the ICOS GPP in-situ data (for cropland and grassland separately) and with the CGLS 300m dataset (for cropland and grassland together).

Validation dataset	Validation metrics	cropland	grassland
ICOS GPP tower data	RMSE	1.59	1.69
	MAE	1.01	1.11
	Precision	1.55	1.67
	R ²	0.86	0.74

4.2.7 Conclusions

Phase 1 of the product already demonstrated some important advancements on the development of a high-resolution cropland and grassland specific GPP product. The resolution was increased from 300m to 10m which allows to observe within field GPP variations, which was one of the user requirements of this prototype. Furthermore, the existing LUE method used for GPP estimation was adjusted by calibrating the LUE factor specifically for crop and grassland locations. A validation of the calibration proved the precision was rather high, with especially for cropland more reliable results.

As a next step, the methodological advancements were implemented in OpenEO to ensure scalability of the approach. The output of the prototype was as a first test already produced for one tile located in Belgium (31UFS) and one calendar year (2022) to showcase the potential to interested stakeholders.

Overall, the results of the final prototype show an improvement compared to the original model.

4.2.8 Outlook beyond the EvoLand project

The results of this final prototype are very valuable to be able to evolve the current operational 300m GPP CLMS service to a higher resolution with a better model accuracy. The prototype demonstrates that including a drought stress factor (based on soil water content and actual and potential evapotranspiration) and simplifying the temperature stress factor improves the GPP estimates considerably. A next step will be to evaluate if these model improvements also result in more reliable GPP estimates in other Land Cover types (besides grass and cropland evaluated here). In addition, it should be evaluated if there is a way to further decrease the positive bias (i.e. overestimation of the model) in the summer months that are still observed in the final version of this prototype.

4.3 C6 – Small Landscape Features

4.3.1 Introduction & Objectives

Small Landscape Features (SLF) represent stable landscape elements that provides numerous ecological and socio-cultural functions, which can be grouped in 4 main categories:

- Soil and water conservation: Specific features such as grass margins, hedgerows or ditches slow runoff from fields, filtering pesticides and other potential pollutants before they reach surface waters.
- Climate protection and adaptation: SLF have significant roles in flood and landslide mitigation, carbon storage and sequestration.
- SLF support biological diversity by facilitating the free movement of species between habitats patches, thus providing important habitats and ecosystem services.
- Cultural identity: SLF compose and structure rural landscapes. They reflect the local agricultural history and provide a large variety of traditional landscapes trough Europe, thus providing an important cultural value.

Mapping and monitoring those features can thus have application in biodiversity monitoring, land management (biological corridor between habitats), soil and water conservation and climate adaptation.

Objectives of this prototype is to build on existing CLMS product Small Woody Features (SWF) which maps hedgerows and patches of tree across Europe, by including additional classes of features through the analysis of new EO datasets. The specifications of the C6 prototype are stated in next table.

Table 32. Technical and thematic specifications for the C6 prototype.

Prototype C6	
Prototype	C6 Small Landscape Features
Thematic specifications	Mapping of woody vegetated features outside forest and stable landscape features within agricultural areas (grass margin, stone walls, ditches)

Prototype sites	1 – Spain 3 – Belgium 6 – Eastern Europe
Temporal frequency	Annual
Spatial resolution	5m
MMU	TBD
Envisaged accuracy	HRL SWF 2018 as basis: 80% UA/PA

4.3.2 State of Play

From 2015 onwards, and with an update every 3 years, small woody features are mapped and monitored within CLMS portfolio, through HRL SWF products. It is the first and only HRL product to rely on VHR imagery to provide final layer at 5m spatial resolution including hedgerows and patches of trees corresponding to certain geometric specifications, outside of forested areas (which were masked out in the product).

Woody features not responding to the geometrical specifications (e.g. bigger or too small) could however be included in the final product for ensuring continuity (when connected to a “valid” SWF) and avoid artificial splits of features.

Starting with the 2018 update, the forest mask (derived from CLMS products and corresponding to the Food and Agriculture Organization (FAO) forest definition) allowing to exclude forested area from the product is now provided together with the SWF layers, and linear and patchy features are combined in one unique SWF class. Those 2 changes aimed to facilitate use of the products by end users, and a better understanding of relation with HRL Forest products.

Since the 2021 update, the SWF portfolio also includes a new layer: The Woody Vegetation Layer that maps all woody vegetation without any filtering (mask or geometrical filtering), allowing end users to apply their own definition of SWF.

User requirements collected by WP1 and related to the C6-SLF prototype mainly focus on adding new landscape features to the products (grass margins, terraces, hedges, stone walls, ponds) and improving resolutions of the product (both spatial and temporal to quickly identify changes). This prototype aims to demonstrate new approaches (e.g. use of new EO data) can help fulfil those requirements, by adding new classes of mapped features and increase the frequency between two updates by mapping those elements for the years 2020 and 2022.

It is worth noting that landscape features targeted by this prototype (e.g. grass margins, ditches, stone walls and ponds) are until now explicitly excluded from existing SWF products.

4.3.3 Selection of Prototypes sites

For the Phase 2 of the project, the implementation was carried out in 3 test sites, namely Spain, Belgium and Eastern Europe. C6 prototype has been deployed over

7 Sentinel 2 tiles during phase 2 of EvoLand project (30SWG, 30SWH, 31UES, 31UFS, 34TFQ, 34TFR, 34TFS), for the reference years 2020 and 2022.

The location of the test sites is presented in section 2.2.

4.3.4 Input and Ancillary Data & Pre-processing

4.3.4.1 Input data

For the realisation of this prototype, two EO datasets have been considered as input data: VHR imagery (Pleiades) and Sentinel-2 imagery.

While Pleiades images are already part of the Copernicus VHR coverage updated every 3 years, only multi-spectral imagery is available in this context. For the scope of EvoLand prototype tests, a complete access to pan-sharpened Pleiades imagery is possible, allowing to detect features below the 5m spatial resolution of the existing SWF product. Pan-sharpening process is performed by the Centre national d'études spatiales (CNES). To limit computational resources needs link with increased number of pixels to process and avoid salt and pepper effect that is common with very high-resolution dataset, Pleiades imagery at 50cm were resampled at 1m spatial resolution. Landscape details are still accurately discernible, while still being consistent with the required improvements of spatial resolution.

Sentinel 2 data are also included as input EO data. While the native spatial resolution of S-2 data is not compatible with the mapping of small landscape features, use of super-resolution algorithms developed within WP2 allow to obtain S-2 data at 5m spatial resolution. If this doesn't improve the current resolution of existing products, the dense time-series of S-2 dataset will provide valuable information not available when using single date acquisition of VHR dataset.

Therefore, S-2 time series were collected at L1C format over the area of interest, over a time window relevant to identify vegetation activity (April-September) for each reference year. Atmospheric correction was performed using MAJA method. Finally, following the outcomes of the WP2, 5m super-resolved images from Sentinel-2 L2A products (bands B02, B03, B04, B05, B06, B07, B08, B8A, B11 and B12) have been generated using Single Image Super-Resolution model, developed in WP2.

4.3.4.2 Ancillary data

Ancillary data were used for two different objectives:

- Collect in-situ data to train algorithms used for supervised classification of woody features
- Produce exclusion mask to outline relevant features

To train our classification algorithm aiming at mapping woody features, we extract training samples from existing CLMS products, namely HRL TCD 2018 and HRL SWF 2018. Both layers are clipped over the AOI extent. HRL TCD is converted to a

binary layer by applying a threshold of 50%, to keep only area with high density. Indeed, as HRL TCD is derived from Sentinel 2 analysis and is 10m spatial resolution, lower density pixels are mixed pixels containing woody vegetation plus another land cover class. By discarding area of low tree density, only “pure” woody information is kept as training sample. This binary layer is converted to polygons, and a random subset is selected. A random subset is also performed over polygons from HRL SWF to complete the training dataset.

CLCplus is used to derive an “agricultural mask” where the mapping of new features will focus. In the near future, this agricultural mask could be derived from the upcoming CTY and GRA layers which are part of the HRL VLCC product suite.

Digital Height model, if presenting a sub-metric spatial resolution, is an important source of information allowing to discriminate ground level structures (grass margins and ditches) from stone walls and low height hedges. During Phase 2 of EvoLand project, only the Belgian test site was covered by height data sufficiently recent and spatially accurate to be consistent with the C6 goals. DHV II dataset over Flanders and Elevation data derived from LIDAR acquisitions over Wallonie were combined and used to derive DHM over the corresponding Sentinel 2 tiles of the AOI. Open Street Map transportation network is used as input to filter out roads from the C6 results.

4.3.5 Methodology and Workflow

Existing HRL SWF products results from the analysis of 2-4m spatial resolution imagery, through a supervised or semi-supervised classification approach based on Machine Learning.

While this approach can be applied successfully to more detailed VHR imagery such as pan-sharpened Pleiades for mapping small woody features, it encounters limitations when applied to new landscape features such as grass margins, stone walls or ditches. Indeed, supervised classification relies, as indicated by its name, on training an algorithm to identify the location of the to-be-mapped features. The training consists of providing to a classification model a pair of datasets (EO imagery + reference/training) where it can learn the correspondence between location of features of interest and their spectral/textural signatures and then predict over the remaining of the AOI where features are located. Geo-datasets about woody vegetation exist at global, continental or national level, with diverse technical specifications (ex: several CLMS products contains a “woody vegetation” class and thus could be used for this purpose). Using them for training an algorithm to map woody vegetation over a new set of EO imagery is effective and already proven. To be able to apply the same approach for newly added features, it would require an accurate, representative across Europe and geo-referenced dataset (derived from other EO data analysis or from field work) containing numerous examples of grass margins, ditches and stone walls. To our knowledge and from investigation performed during the phase1 of WP3, such dataset does not exist. The supervised approach thus cannot be used to identify these features, and this is why the methods used for this prototype are split into two approaches:

1. Supervised approach on VHR dataset to identify woody features
2. Unsupervised approach based on signal thresholding and object extraction, applied to S-2 super-resolution, to identify other features

The workflow in Figure 52 shown below illustrates the different steps, which will be detailed in following sections.

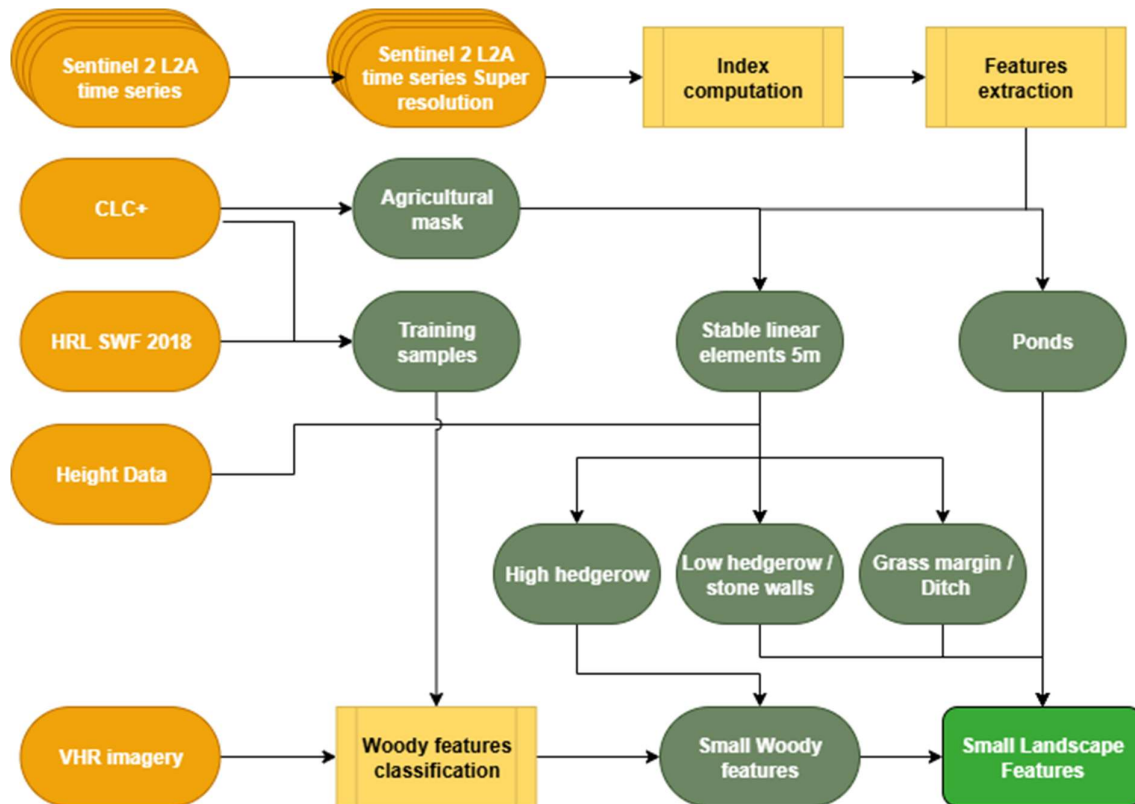


Figure 52. C6 SLF workflow.

4.3.5.1 Supervised approach

Woody features detection relies on a supervised classification of VHR imagery. It is therefore essential to have a reference dataset for calibration and training data (hereafter referred to as reference dataset). The heterogeneous coverage of SWF across the different test sites needs to be considered, therefore a sufficiently high number of reference samples were used to extract representative reference data to train accurately the classification algorithm.

HRL SWF 2018 was primarily use as the main data source for sample selection of foreground reference (= woody vegetation). For area where the density of SWF 2018 was too scarce, other data sources sur as HRL TCD 2018 have been used as complement. Reference samples have been automatically cleaned to ensure relevancy and remove from the reference dataset samples that do not match with woody vegetation. This step is mandatory as the input data derived from product with a coarser resolution than the input imagery, and therefore “woody pixels” from

HRL TCD or SWF2018 can correspond to mixed pixels on the VHR imagery. A NDVI filtering allow to exclude most of the non-vegetated mixed pixels from the reference dataset.

For increasing the accuracy of woody vegetation detection by the algorithm, it is worthy to include background samples (=non woody vegetation) to the reference database, to limit confusion between different types of vegetation and enhance small woody feature delineation within agricultural areas. CLMS CLCplus BB 2021 was used to identify agricultural areas (crop + grassland). In future operational productions, the annual versions of HRL VLCC layers GRA and CTY would be used in complement or replacement of CLCplus BB for this background reference sampling.

The methodology of classification and segmentation is based on Geographic Object-Based Image Analysis (GEOBIA) (Merciol & al, 2019). It relies on two main components which are feature extraction derived from attribute profiles, and a semi-supervised classification using a random forest algorithm. The first step consists in computing predefined indices such as the Normalised Difference Vegetation Index (NDVI), as well as texture characterisation such as Sobel gradient and Haar-like features based on integral image (Viola & Jones 2001) representations (Faucqueur & al. 2019). The texture information is extracted from each spectral band of the original VHR multispectral image, as well as from the NDVI. This textural information has a significant role in allowing the discrimination between woody vegetation and cropland or grassy vegetation, as their spectral response can be close, considering the limited spectral resolution of VHR imagery (Near Infra-Red, Red, Green and Blue). Multi-scale features called Attribute Profiles (AP) and Differential Attribute Profiles (DAP) were derived through a model of morphological tree, which can be seen as a hierarchical segmentation: these trees allow to identify, depending on the way they are pruned, morphological objects from the VHR scenes.

The second step was the use of the random forest classifier in a semi-supervised framework. The reference database previously mentioned ensured the quality of the training set by providing foreground (woody) and background (non woody) samples, needed to comply with the heterogeneous landscapes of the test sites. The output of this classification was a probability map where each pixel is given a probability to depict woody vegetation as show in Figure 53.

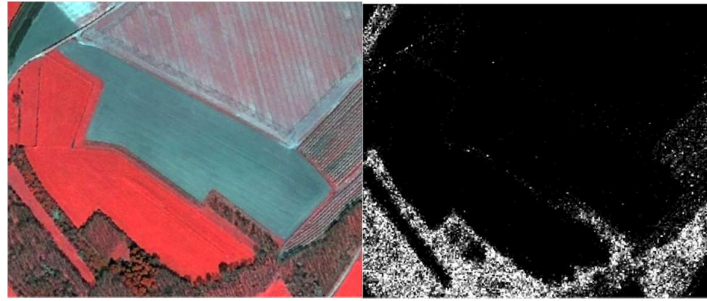


Figure 53. VHR imagery (Pleiades) and Woody vegetation probability map as result of the classification step.

To ensure consistency and allow comparison with current specification of the CLMS HRL SLF 2021-2024 products, and opposite to Phase 1 outputs, forested areas were not masked and woody vegetation mapped within the C6 prototype is now consistent with the Woody Vegetation Layer of the HRL SLF 2021-2024 product suite.

4.3.5.2 Index computing

Regarding additional features classes, the supervised approach described in previous section has not been used. Indeed, a supervised classification relies on a reference database, which in this case would contain examples of the new classes: ditches, grass strip, stone walls. Currently, it is not feasible to extract such dataset from European wide, harmonized datasets, as those classes are most of the times absent for existing geo-referenced database. One alternative would be to manually create this database, by digitizing examples of new classes based on visual interpretation of VHR imagery. This possibility has been rejected due to several points:

- Quantity of data needed to accurately feed a model is important, requiring a huge manual effort not foreseen in this project
- VHR imagery corresponds to single data acquisition, where accurate identification of new classes can be challenging. Ex: differentiate grass margin from neighbouring crop, correctly discriminate ditches from grass strip, etc... Inaccurate reference database would assuredly lead to inaccurate classifications, jeopardizing relevancy of the product.

The figure below illustrates a grass strip between two crop parcels. On this VHR image, spectral and textural response are very similar to the southern crop parcel, making it very challenging to be correctly delineated.

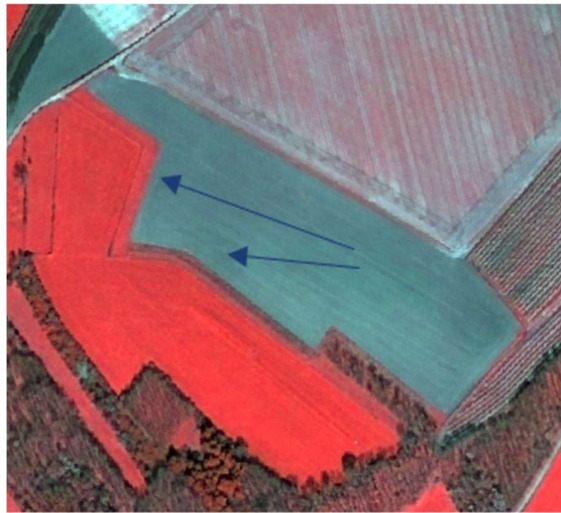


Figure 54. Grass strip between crop parcels (background: Pléiades).

Based on this conclusion, an unsupervised approach has been privileged. The main challenge relied here on determining the correct way to identify those features signatures and the correct approach to discriminate them from surrounding areas.

Grass margins, stone walls, ditches share several specifications: they exist in an agricultural context, they can be found at parcel boundaries, they are linear features, and they are stable elements across seasons, as opposed to the crop parcels themselves which show a continuously changing coverage across the year.

Based on this assumption, the use of Sentinel-2 time series could be valuable to identify their presence, especially if combined with super-resolution approach to reach a spatial resolution compatible with the specification of this prototype.

Figure 55 shows the same grass strip illustrated on previous figure, but on Sentinel-2 super resolution time series across the year, showing that it is always covered by vegetation.

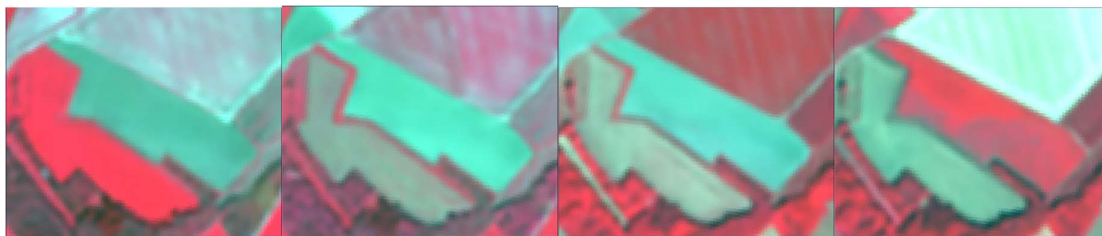


Figure 55. Grass strip on S-2 time series (from left to right: March, April, May, August).

This land cover stability over seasons can be highlighted by computing spectral indices such as NDVI and NDWI and evaluating its standard deviation over the time period. Crop will show a high variability, and stable landscape features such as woody vegetation, grass strips, tracks, ditches, etc. will show little to no variation, independently of their initial coverage (vegetation or not). This is illustrated in Figure 56, where cold colours show stable features, and warmer colours show features with high variability. The grass strip on the southern border of the central

parcel is easily identified (as always covered by the same amount of vegetation), as well as the access track (as always barely covered by vegetation) that separates the two upper parcels, track that was barely visible on the VHR imagery.

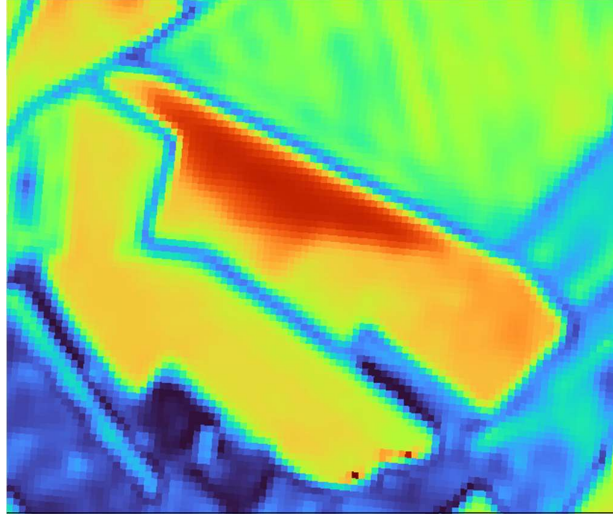


Figure 56. NDVI standard deviation between March and August 2020. Arrows indicate grass margin (down) and access track (up).

4.3.5.3 Features extraction and exclusion mask

Once new features' location has been identified, next step was to extract features from the surrounding landscapes. The assumption based on their linear characteristic was used here. Using algorithms based to edge detection, it was possible to extract linear features that corresponds stable elements.

4.3.5.3.1 Linear Filter

The Frangi filter was used in this context to detect stable linear landscape elements — such as hedgerows, ditches, or grass margins — that persist between crop rotations. These features often resemble vessel-like structures in their shape and continuity, which is precisely what the Frangi filter is designed to highlight. Originally developed for medical imaging to detect blood vessels, the Frangi filter enhances tubular structures by analysing the eigenvalues of the Hessian matrix of the image, making it well-suited for identifying elongated features in satellite imagery. In this implementation, the filter was applied to standardized spectral indices from Sentinel-2 imagery, and the output was binarized using a threshold value. This threshold determines the sensitivity of the detection: pixels with a Frangi response above the threshold are considered part of a linear feature. The binarized outputs from multiple indices are then combined using a logical AND operation, ensuring that only consistently detected features across indices are retained.

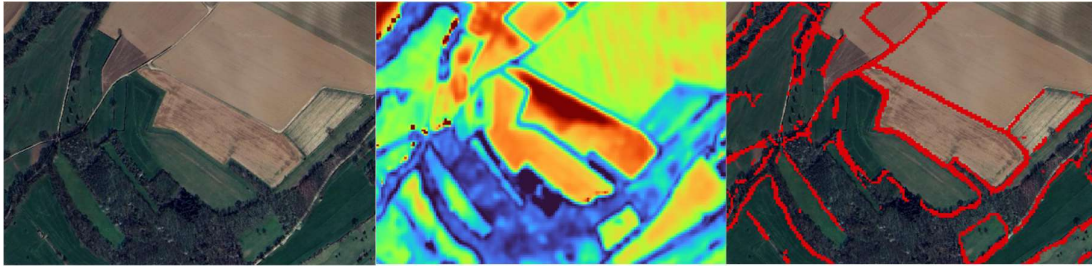


Figure 57. Frangi filter applied on the indices standard deviation in 2020 (from left to right: Google Satellite as background, NDVI standard deviation between March and August 2020, Frangi filter > 0.003 threshold confirmed on both standard deviation of NDVI and NDWI between March and August 2020).

As we also assumed that those new features are located in agricultural context, we used CLCplus BB 2021 to produce an agricultural mask where such elements would be identified, to exclude artefacts within urban areas for example. In future productions, HRL VLCC layers GRA and CTY could be used in replacement or complement of the CLCplus BB.

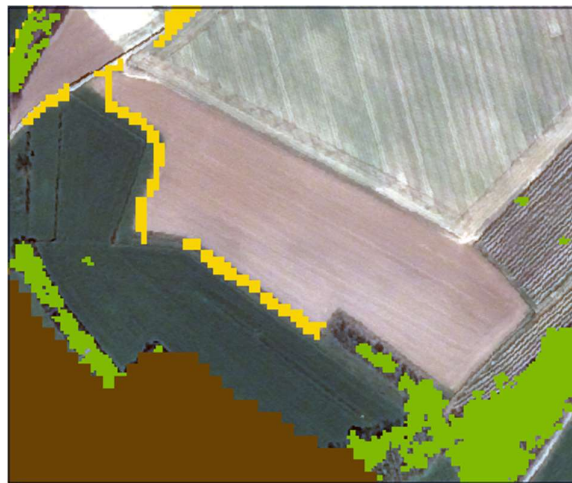


Figure 58. C6 SLF results illustrating mapping of a grass margin in Belgium (background: Pléiades).

At this stage of the workflow, the prototype contains 3 classes:

- Woody vegetation derived from VHR analysis
- Non-productive agricultural linear features which contain grass margins, ditches, low height hedges and stonewalls.
- Other landscapes

4.3.5.3.2 Inclusion of height data

Inclusion of height data was foreseen to help discriminate different classes within the new detected linear features. Indeed, grass margins, ditches and stone walls (most of the time covered by vegetation) present very similar spectral signatures on VHR data or S2SR imagery, but present different heights.

Digital Height model, if presenting a sub-metric spatial resolution, is an important source of information allowing to discriminate ground level structures (grass

margins and ditches) from stone walls and low height hedges. During phase 2 of EvoLand project, only the Belgian test site was covered by height data sufficiently recent and spatially resolved to be consistent with the C6 goals. Derived from LIDAR data, Digital Surface Model (DSM) and Digital Terrain Model (DTM) with 50cm spatial resolution were collected over the test site and a Digital Height Model (DHM) was derived. Using this DHM layer, non-productive agricultural linear features were separated into 3 classes:

- 0m high: grass margin or ditch
- 0-1m high: stone wall/fence/low hedgerow
- >1m high: high hedgerow (to be merged with woody vegetation)

For Spain and Eastern Europe test sites, the absence of recent and spatially relevant height dataset caused the grass margins, ditches, stone walls and low hedgerow to be merged in one unique class.

4.3.5.3.3 Ponds Detection

The detection of ponds in satellite imagery focuses on identifying small, isolated water bodies that are distinct from larger hydrological features like rivers or lakes. Ponds are typically shallow and limited in surface, and can play a crucial role in local biodiversity, agricultural water management, and landscape stability. In this workflow, ponds were detected using the NDWI (Normalized Difference Water Index), which highlights water surfaces based on their spectral reflectance. A threshold was applied to the NDWI image to isolate water pixels, and then river segments were excluded using OpenStreetMap (OSM) data to ensure only non-riverine water bodies were retained. A size filter removed large water bodies, keeping only features smaller than 5ha.

This class was then added to already mapped small landscape features in the status layer.



Figure 59. Pond detection (background: Google Satellite).

4.3.5.3.4 Change identification and Status Layers post-processing

Due to the nature of newly detected features (small, linear), their detection and precise location is strongly impacted by any mis-registration between pixels from different sources, which could lead to the same element being slightly shifted between two consecutive reference year. This would strongly impact change detection if only identified by a status layer difference process. To mitigate this

effect, a post-processing is applied to the two-status layer to confirm the detection of elements and decrease as much as possible false detection of change. This approach builds on comparison of time-series parameters and their Mean Absolute Error (MAE) to confirm the presence of an element in said reference year and within the change layer. The process is illustrated in Figure 60 below.



Figure 60. SLF Change identification workflow.

4.3.6 Results and Validation

4.3.6.1 Overview

At the end of the process, we obtain the C6 SLF layer containing 53 classes: small woody features grass margins/ditches (class 1), newly detected landscape features that we will call 'non-productive agricultural features' for stone walls / low hedges (class 2), woody vegetation (class 3), ponds (class 4) and remaining of the AOI (class 0). To ease manipulation and display of the results, the Forest mask re-used from SWF2018 has been included in the layers as an additional class (class 3). The Figure 61, Figure 62 and Figure 63 illustrate results over the 3 test sites. From left to right, Pleiades image, C6 prototype results and SWF 2018 Woody Vegetation Mask (WVM) are displayed. Forest mask is coloured in brown. Woody vegetation is coloured in shades of green.



Figure 61. C6 prototype results over Belgium site (background; Pléiades).



Figure 62. C6 prototype results over Spain site (background; Pléiades).

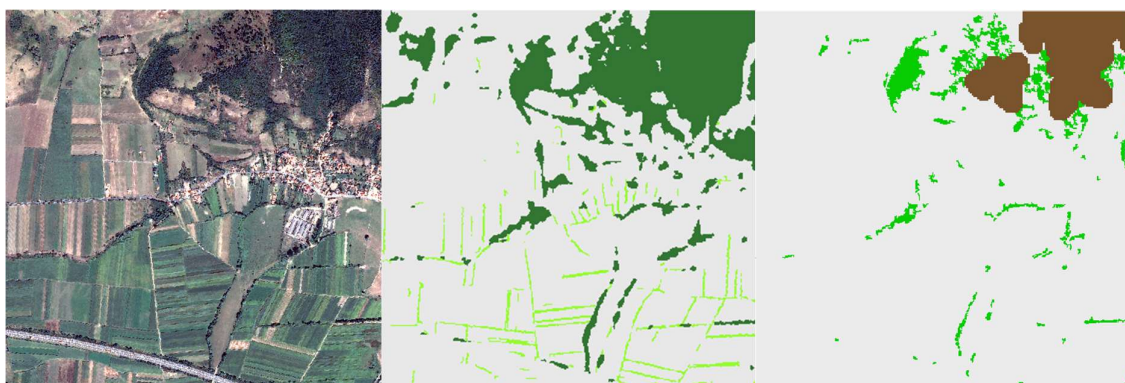


Figure 63. C6 prototype results over Romania site (background; Pléiades).

Over every test site, the use of Pleiades imagery at 1m spatial resolution significantly improved the detection and outlines of woody vegetation. It is especially visible in Spain where the low-density tree coverage was partly omitted in SWF2018 product. In the framework of EvoLand, no post-processing or geometrical rules were applied to the woody vegetation, to maximize the detection and avoid removing thematically accurate features because of their geometrical characteristics. Thus, the woody vegetation class of C6 prototype is similar to the Woody Vegetation Layer of SWF2021.

4.3.6.2 Thematic accuracy

Thematic assessment of the prototype was performed by comparing the product against a reference dataset derived from visual interpretation of VHR and Sentinel-2 data, over validation samples.

Validation samples were selected using a random stratified approach over each test site: the AOI was divided into several strata (see details below), within which samples (e.g. centre of 2.5m pixel) were randomly selected. Strata were defined and validation samples were distributed as illustrated in the table below, and this per Sentinel2 tile of each test site (Table 33):

Table 33. C6 prototype validation sample distribution per test site.

Strata	Number of samples
Omission strata	280
Grass margin / ditch commission	300
Low hedge/ stone walls commission	50
Woody vegetation commission	245
Ponds commission	175
Loss of SLF commission	175
Gain of SLF commission	175

To compensate for sampling effort bias (as the 3 strata are not equal in terms of area), a weighting factor is applied to each sample, according to the strata it belongs to.

Visual interpretation of samples was performed, and confusions matrices for each test site were derived. Results are presented in following tables.

Table 34. C6 Thematic accuracy for Spain test site for 2020.

ES 2020		Reference						User Acc.	CI95%	Commission
		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total			
MAP	other	236.21	0.04	2.74	26.55		265.5414	88.95%	3.07%	11.05%
	grass/ ditches	0.12	0.35	0.01	0.02		0.5025	72.33%	4.38%	27.67%
	Woody	2.50			131.42		133.9119	98.14%	1.33%	1.86%
	Ponds	0.00				0.04	0.0441	93.85%	2.36%	6.15%
	Total	238.83	0.40	2.75	157.99	0.04	400.0000			
Prod Acc.		98.90%	11.55%		83.18%	100.00%				
CI95%		1.02%	3.13%		3.67%	0.00%				
Omission		1.10%	88.45%		16.82%	0.00%				

Table 35. C6 Thematic accuracy for Spain test site for 2022.

ES 2022		Reference						User Acc.	CI95%	Commission
		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total			
MAP	other	236.21	0.04	2.74	26.55	0.02	265.5645	88.95%	3.07%	11.05%
	grass/ ditches	0.12	0.35	0.01	0.02		0.5001	72.44%	4.38%	27.56%
	Woody	2.50			131.42		133.9119	98.14%	1.33%	1.86%
	Ponds					0.02	0.0235	100.00%	0.00%	0.00%
	Total	238.83	0.40	2.75	157.99	0.04	400.0000			
Prod Acc.		98.91%	11.52%		83.18%	56.66%				
CI95%		1.02%	3.13%		3.67%	4.86%				
Omission		1.09%	88.48%		16.82%	43.34%				

Table 36. C6 Thematic Accuracy for Eastern Europe Test Site for 2020.

East. Europe 2020		Reference						User Acc.	CI95%	Commission
		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total			
MAP	other	220.28	0.23	0.03	8.48	1.77	230.7871	95.45%	2.04%	4.55%
	grass/ ditches	0.75	0.60	0.01	0.05		1.4062	43.40%	4.86%	56.60%
	Woody	3.66			364.07		367.7350	99.00%	0.97%	1.00%
	Ponds	0.01				0.07	0.0717	92.97%	2.51%	7.03%
	Total	224.70	0.83	0.04	372.60	1.84	600.0000			
Prod Acc.		98.04%	70.41%		97.71%	3.62%				
CI95%		1.36%	4.47%		1.47%	1.83%				
Omission		1.96%	29.59%		2.29%	96.38%				

Table 37. C6 Thematic Accuracy for Eastern Europe Test Site for 2022.

East. Europe 2022		Reference						User Acc.	CI95%	Commission
		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total			
MAP	other	220.28	0.13	0.03	8.47	1.80	230.7122	95.48%	2.04%	4.52%
	grass/ ditches	0.81	0.66	0.01	0.02		1.5075	44.70%	4.87%	55.30%
	Woody	3.66			364.07	0.00	367.7373	99.00%	0.97%	1.00%
	Ponds					0.04	0.0431	100.00%	0.00%	0.00%
	Total	224.76	0.80	0.04	372.57	1.84	600.0000			
Prod Acc.		98.01%	80.48%		97.72%	2.34%				
CI95%		1.37%	3.88%		1.46%	1.48%				
Omission		1.99%	19.52%		2.28%	97.66%				

Table 38: C6 Thematic Accuracy for Belgium Test site for 2020

BE 2020	Reference
---------	-----------

		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total	User Acc.	CI95%	Commission
MAP	other	284.28	3.50	0.07	7.34	0.05	295.24	96.29%	1.85%	3.71%
	grass/ ditches	0.32	0.97	0.04		0.04	1.37	70.71%	4.46%	29.29%
	hedges/ stone walls	0.63	0.09	1.78			2.51	71.06%	4.44%	28.94%
	Woody	2.87			97.70		100.58	97.14%	1.63%	2.86%
	Ponds	0.06				0.24	0.30	79.02%	3.99%	20.98%
	Total	288.17	4.57	1.89	105.05	0.33	400.00			
	Prod Acc.	98.65%	21.29%	94.10%	93.01%	72.18%				
	CI95%	1.13%	4.01%	2.31%	2.50%	4.39%				
	Omission	1.35%	78.71%	5.90%	6.99%	27.82%				

Table 39. C6 Thematic Accuracy for Belgium Test site for 2022.

BE 2022		Reference						User Acc.	CI95%	Commission
		other	grass/ ditches	hedges/ stone walls	Woody	Ponds	Total			
MAP	other	284.50	3.47	0.10	7.34	0.08	295.50	96.28%	1.86%	3.72%
	grass/ ditches	0.16	1.02	0.04		0.04	1.27	80.33%	3.90%	19.67%
	hedges/ stone walls	0.59	0.09	1.73			2.42	71.61%	4.42%	28.39%
	Woody	1.82			98.76		100.58	98.19%	1.31%	1.81%
	Ponds	0.03				0.20	0.24	85.16%	3.48%	14.84%
	Total	287.11	4.59	1.87	106.10	0.33	400.00			
	Prod Acc.	99.09%	22.26%	92.33%	93.08%	61.61%				
	CI95%	0.93%	4.08%	2.61%	2.49%	4.77%				
	Omission	0.91%	77.74%	7.67%	6.92%	38.39%				

Table 40. C6 Thematic accuracy for Change layers (all sites).

Change 2020-2022 All sites		Reference				User Acc.	CI95%	Commission
		Stable	Loss	Gain	Total			
MAP	Stable	1397.623	0.097	1.056	1398.776	99.92%	0.15%	0.08%
	Loss	0.640	0.020		0.660	3.02%	0.90%	96.98%
	Gain	0.505		0.059	0.564	10.49%	1.61%	89.51%
	Total	1398.768	0.117	1.115	1400.000			
	Prod Acc.	99.92%	17.03%	5.31%				
	CI 95%	1.13%	4.01%	2.31%				
	Omission	0.08%	82.97%	94.69%				

Thematic assessment of the C6 prototype confirms look and feel results. Woody vegetation is usually well captured with few commission errors. Some omission errors are found for smaller elements of open forested areas in Spain. Accuracy figures are similar to results from SWF2018 production, indicating that the proposed workflow can deal with new EO data with improved spatial resolution without impacting the accuracy of the class.

For new classified features, omissions errors are low except for the “ground level” class containing grass margins and ditches. Some of those elements can be confused with field boundaries depending on crops surrounding those elements, especially if the crops present different phenology (ex. Maize next to winter wheat). In such scenario, the field boundary is constituted of mixed pixels presenting a continuous vegetation coverage over the full period of analysis, leading to confusion with other stable elements. In Spain and Eastern Europe, where this class is merged with the low hedges/ stone walls class due to the absence of height data, this issue is less present as the results for the class stone walls/ low hedges are typically higher.

Commission error rates for woody vegetation and ponds is very low, reflecting the accurate mapping of those elements. For linear elements such as grass margin, stone walls, ditches or low hedges, commission errors vary between 20 to 50% depending on the site. This is consistent with the previous findings indicating that those class (especially for ground level elements) can present confusion with field boundaries, leading to the omission and commission errors displayed in the confusion matrices.

Thematic assessment of changes reflects the not solved challenge of accurately mapping changes for small landscape elements highly subject to technical specification of satellite imagery: absence of co-registration between subsequent reference years, impact of acquisition angle and spatial resolution. The rarity of the change class also amplifies in a negative way the results as it impacts greatly the weighting factor applied to sample to compensate over-representation of “change” samples in the stratification.

4.3.7 Conclusions

C6 Small Landscape Features prototype maps small woody features with an improved resolution compared to existing products and include a new class of agricultural non-productive landscape features.

Listing the positive aspects related to this prototype, it features an approach combining both supervised and unsupervised approaches. The new class of landscape features is now included in the products, and it does not rely on existing dataset or intensive collection of in-situ data, making it easily transferable to other areas. The method is also computational efficient and takes advantages of super-resolution Sentinel-2 data, making Sentinel-2 relevant for mapping smaller

landscape elements. The existing small woody features class has been successfully extracted at an improved spatial resolution compared to previous SWF products, through the use of Pleiades imagery.

However, this prototype has some known limitations. All new classes listed in the user requirements cannot be totally discriminated without the addition of very high precision height dataset, which availability and update frequency is very variable across Europe (only available for the Belgium site during this study). Some classes remain merged (grass margin and ditches; low hedges and stone walls) because they share very similar signatures: similar height, and most of the time covered by vegetation which limits further discrimination.

While a great achievement has been done by detecting those features from the Sentinel-2 super-resolution, it implies that they will be mapped with a spatial resolution of 5m at best, limiting the detection of smaller elements and subject to confusion at field boundaries where those elements can be confused with mixed crop pixels. Furthermore, Sentinel-2 super-resolution is not a processing level available from the usual data hub such as the Copernicus Data Space Ecosystem (CDSE). Therefore, a significant part of pre-processing is mandatory to compute those datasets before proceeding to extract landscape features from it.

Change detection of small landscape features remains a challenge due to the very nature of the mapped elements: small and mostly linear. These characteristics make them particularly sensitive to the EO data specification. Mis-registration of pixels between subsequent acquisitions, different acquisition angles or acquisition time can greatly impact the precise location of the element. Over several years, a stable element can thus be detected few meters away from the previously identified location. While not an issue for elements with a larger footprint such as ponds, large hedgerow or tree patches, it is problematic for small linear elements such as grass margins, smaller hedges or ditches. A small shift in the location of a small element would then implies a double wrong detection of change: disappearance of the element on one side, appearance of a new elements few meters away on the other side. While the workflow of this prototype tries to limit those effects, they are still significant and limit accurate change detection at finer spatial resolution.

4.3.8 Outlook beyond the EvoLand project

The C6 prototype has demonstrated promising capabilities in mapping small landscape features with improved resolution and minimal reliance on external datasets. To progress towards higher TRLs and operational phase, the focus should shift toward operational validation in diverse geographic and agricultural contexts, to complete observations from EvoLand project. Key priorities include integrating complementary datasets—such as high-precision elevation models—to enhance class discrimination and reduce ambiguity between visually similar features.

Future Earth Observation missions (e.g., CO3D) and expanding LIDAR coverage across Europe offer a pathway to overcome current limitations in height data

availability. Additionally, efforts to streamline the pre-processing of Sentinel-2 super-resolution imagery, potentially through automated workflows or cloud-based platforms, will be essential to support scalability and broader adoption.

To enable reliable change detection, further research is still needed to mitigate the effects of pixel misregistration and acquisition variability. This may involve developing more robust temporal alignment techniques or leveraging multi-sensor fusion. Ultimately, transitioning to higher TRLs will require close collaboration with end-users to refine product specifications and ensure the prototype meets operational needs in real-world settings.

5 Water prototypes

In the frame a world facing the numerous challenges linked with climate change, the monitoring of water resources is an unavoidable purpose. Indeed, the area of water bodies is considered an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS). It also plays a strong role on the answer given to the Sustainable Development Goal 6 on water and sanitation (SDG6) of the United Nations. To detect and provide accurate and regular information on the water area changes, space remote sensing imagery has a key role to play thanks to its continuous and recurring measurements at a global scale. The water prototype presented and validated in the frame of the EvoLand project is exposed in the following sections to enhance the currently existing products.

5.1 C7 – Improved Water Bodies

5.1.1 Introduction & Objectives

To better understand the changes in water surface area at a global scale, the production of accurate, regular and global water extent is needed. That is why in the frame of CLMS, water area delineation is a key component. Furthermore, near-real-time mapping of water bodies from satellite imagery plays a critical role in water management.

As part of the EvoLand project, the objective is to enhance current CLMS products (Copernicus High-Resolution Layer Water and Wetness; Copernicus Global Land Monitoring Water Bodies). The requirements must first be established with the end users. Then, starting from the existing current products, solutions will be developed following the requirements, before validating the approach on various sites in Spain, Belgium, Sweden and Ivory Coast.

5.1.2 State of Play

5.1.2.1 Existing CLMS Products

Water area mapping is done in two separate products within the CLMS frame:

- **High Resolution Layer Water and Wetness (HRL WaW):**

This product provides information about the presence and condition of water and wet surfaces across Europe. The method approach uses multi-temporal and multi-seasonal optical high-resolution satellite imagery and on radar information (Sentinel-1 data) with a geometric resolution of 10 m. Using a time series of seven years, products correspond to one image composite per season (3 months) and year during the observation period. It captures the intra-annual dynamics. Two products are available on two periods:

- ✓ Water and Wetness status 2015 (raster 20 m and 100 m), Europe, 3-yearly.
- ✓ Water and Wetness status 2018 (raster 10 m and 100 m), Europe, 3-yearly.

This product provides precious information about water occurrences on the time periods 2009-2015 and 2012-2018.

- **Copernicus Global Land Monitoring Water Bodies:**

As part of Copernicus Global Land Operations (CGLOPS), this product is a detection of areas covered by inland water along the year. With a monthly temporal resolution, it allows the analysis of seasonal dynamics in near-real time at a global scale.

Water bodies products have a spatial resolution from 100m to 300m, a temporal resolution from 10-daily to monthly and are provided since 2014.

The method to compute Water Bodies (WB) product is described in the figure below.

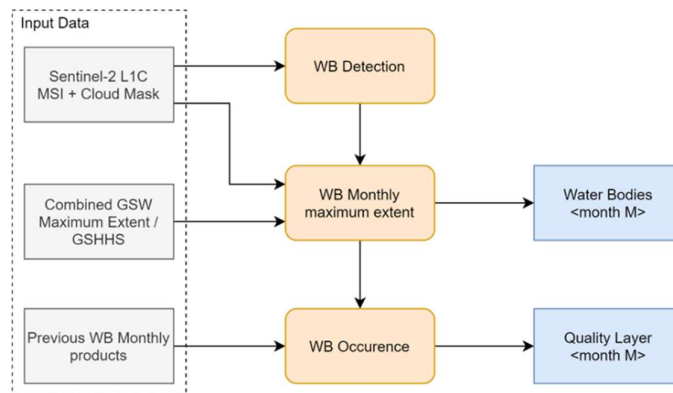


Figure 64. CGLOPS Water bodies production workflow.

This approach uses Sentinel-2 data, combined with Global Surface Water (GSW) maximum extent / GSHHG (Global Self-consistent, Hierarchical, High-resolution Geography), and previous WB products (when available) to create a global water mask within five days after a month is over. With a resolution of 100m, this monthly product monitors the changes in water surface areas at a global scale.

5.1.2.2 WP1 User requirements

During the realisation of the first Work Package, requirements about what the water product should become were defined as follows.

The product should be capable of capturing ponds smaller than 1 hectare (100x100m). Ideally, a resolution of 0.1 ha (10m) would be achieved. It should also be available globally and provide maps on at least a monthly basis. Respect these requirements is possible with the use of sentinel-2 data, its 10-meter resolution, its global coverage and its high frequency of revisit.

Additionally, the product should have included information on the volume of water contained in the bodies, however this is not feasible due to the unavailability of bathymetry information at global scale.

The product should distinguish between permanent and temporary water bodies and detect new water bodies. For this requirement, the idea in the first instance is to remove constraints on existing water bodies mapping datasets.

Respecting those requirements implies following a hybrid approach between existing CLMS products and bringing new ideas. Along with these requirements, as operational producers of the CGLOPS water bodies, we keep in mind and strongly advocate for the operability of the method to be developed. This means that we must propose a good trade-off between speed and accuracy, along with robustness and spatial coherence. Concerning the volume of water contained in a water body, a volume variation option is already proposed in the frame of CGLOPS products, so it was not explored in the frame of EvoLand. The specifications of the C7 prototype are described in Table 41.

Table 41. Technical thematic specifications for the C7 prototype.

Prototype C7	
Sub-prototype	Water Bodies extent
Prototype sites	1 – Spain 3 – Belgium 5 – Sweden 8 – Côte d'Ivoire
Description of the sub-prototype	Map water bodies extent and type of water bodies (permanent or temporary)
Expected temporal update	Monthly
Spatial resolution	10m
MMU	0.1ha
MMW	N/A
Envisaged accuracy	>95%

5.1.3 Selection of Prototypes sites

During Phase 1, the prototype was tested on 4 sites of 220*220 km² (representing four Sentinel-2 MGRS tiles): Ivory coast, Spain, Belgium, and Sweden. The study areas remained the same during Phase 2. This set allows us to evaluate the behaviour of this algorithm in different situations in terms of variety of climate, land use, land cover, etc.

5.1.4 Input and Ancillary Data & Pre-processing

Sentinel-2 L1C (S2L1C) data was used: it has the advantage of having high resolution (10 meters) optical earth observation from Copernicus Sentinel-2 mission with Top-of-atmosphere reflectance in cartographic geometry. We chose to work on L1C products to avoid the highly time-consuming step of levelling up to L2A products.

The prototype needs the use of a Digital Surface Model (DSM)/DEM. The choice of Copernicus GLO-30 DEM is relevant because this is a DSM with a nearly global cover at a resolution of 30 metres from data acquired with the TanDEM-X mission between 2011 and 2015.

A *priori* data of global water surfaces was also needed. Global Surface Water – maximum water extent (GSWE) maps maximum water extent at a global scale based on Landsat data from 1984 to 2021 (Pekel et al., 2016) and meets that need. A list of datasets is shown in Table 42 table:

Table 42. EO Ancillary data for water bodies mapping prototype.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Sentinel-2 LIC spectral bands.	10 - 20m	Daily
Copernicus GLO-30 DEM	Global DSM	30m	-
Global Surface Water, – maximum water extent (GSWE)	Maximum water extent at a global.	30m	-

5.1.5 Methodology and Workflow

The objective is to product a global water map which could capture ponds under 1 hectare, be produced at least at a monthly pace, and consider new and ephemeral water bodies, with an accuracy above 95%. That is why we chose to work on the native resolution of 10m, and on a tile-by-tile approach to have potentially daily actualized water areas. During phase 1, we choose to get a probabilistic approach to produce water likelihood map based on a multi-index' thresholding, filtered with an exclusion mask, and weighted with slope and GSWE data (see Figure 65).

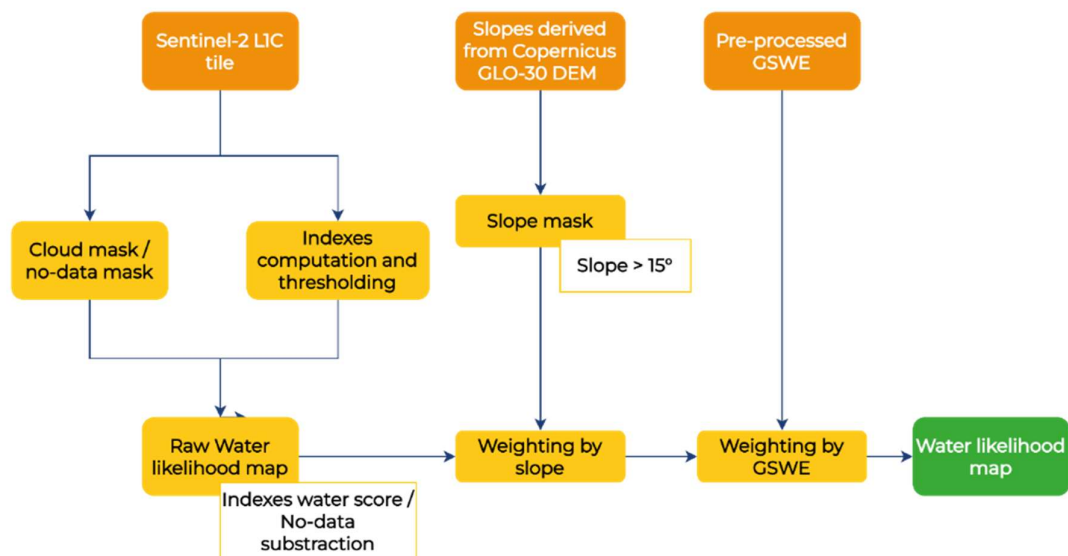


Figure 65. C7 prototype Workflow during phase 1.

Connected to the limitations of phase 1 version, we proposed to focus on the following points during phase 2 of the project:

- Tuning of the thresholds used on the three indexes.
- Explore new shadow exclusion approaches.
- Develop a synthesis approach to go from daily to weekly or monthly products.
- Improve as much as possible the calculation efficiency of the algorithm.

5.1.5.1 Multi-index thresholding

To choose the water indexes, we used 2 complementary techniques to finally select 3 indexes among 10 water indexes from literature: a Principal Component Analysis (PCA) and a Linear Discriminant Analysis (LDA). PCA is an unsupervised method that reduces dimensionality by capturing the directions of maximum variance in the data, helping to remove redundancy and noise. LDA, on the other hand, is a supervised technique that finds the axes that best separate predefined classes. When combined, PCA can be used first to simplify the dataset, and LDA can then maximize class separability in this reduced space. To do that we prepared a file of values samples on ten different indexes labelled as water class and non-water class on challenging test sites (shadows, topographic issues, clouded area, sun glint...) shown in Figure 66.

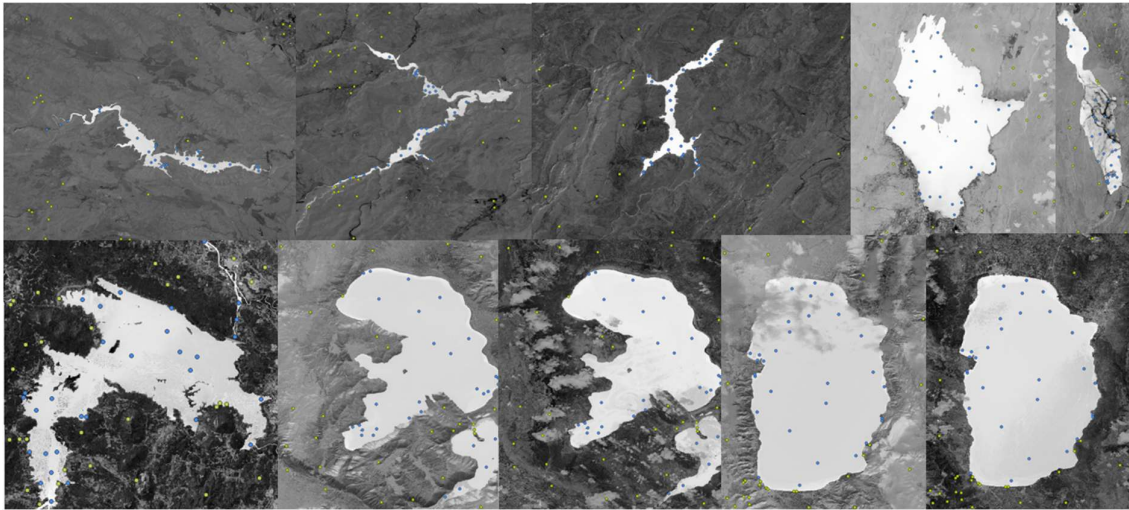


Figure 66. Value sampling on multiple test sites.

This study confirmed the choices made during phase 1. The indices NDWI, AWEIsh, and MBWI were selected based on their ability to discriminate water from non-water areas while minimizing redundancy between them. NDWI showed the highest correlation with the water class, indicating strong discriminatory power. AWEIsh complements NDWI by incorporating additional spectral features (shortwave infrared), offering a different perspective on water detection. MBWI, while slightly less correlated, remains informative and is less redundant with the other two indices. Together, these three indices provide a balanced combination of discriminative strength and spectral diversity, making them suitable for robust water/non-water classification.

During Phase 1, fixed thresholds were defined for each index and applied uniformly across all test sites. However, it was observed that these static thresholds could lead to inconsistencies in detection results, particularly due to seasonal and geographical variations in reflectance.

To mitigate this effect, the Solar Zenith Angle (SZA), available in Sentinel-2 metadata, was used to dynamically adjust threshold values. For each index, a linear regression model was established to derive threshold values as a function of the SZA, based on the reflectance extremes observed at the Belgium and Spain test sites (see Figure 67).

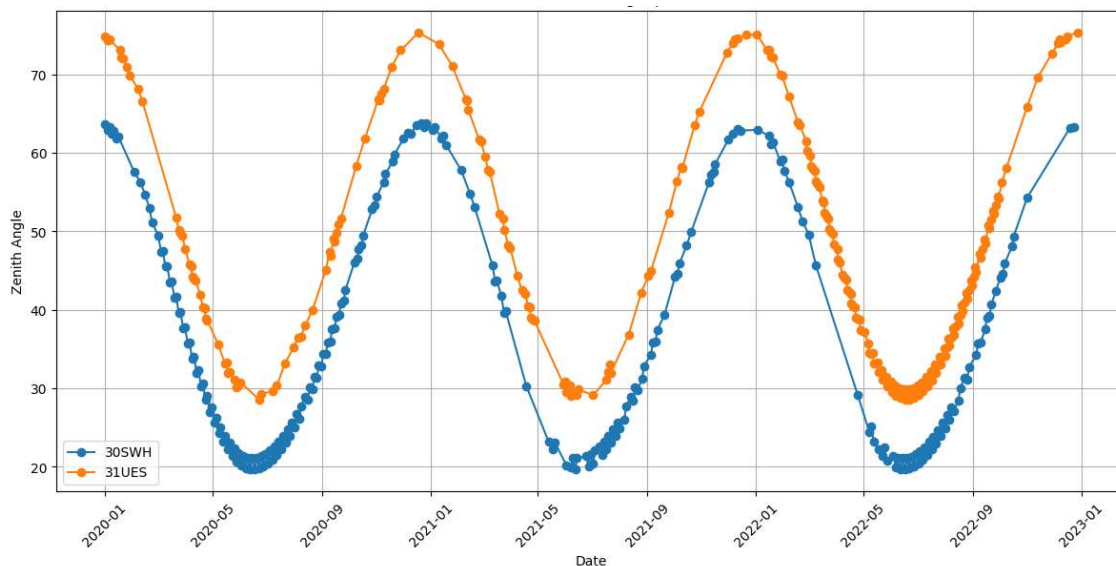


Figure 67. Solar Zenith Angle on Belgium and Spain test sites.

This approach enables scene-specific thresholding, enhancing the robustness of water detection by accounting for variations in illumination conditions. By normalizing reflectance values across different acquisition contexts, the method reduces false detections caused by changes in sunlight incidence and improves the overall consistency of the results.

After the thresholding step, the indexes water score was kept in the method: once the three indexes have been thresholded, it is possible to calculate for each pixel the proportion of indices that consider it to be water. This produces a water likelihood map with values of 0, 33, 66 and 100%.

5.1.5.2 Filtering and shadow weighting

During Phase 1, once the initial water likelihood map was generated, the data was refined using an exclusion mask and further weighted with slope information and GSWE data. In Phase 2, the methodology was optimized to streamline the process and remove dependencies on slope modelling and GSWE layers, allowing for greater flexibility and scalability across diverse geographic contexts. The main challenge was to tackle the absence of a priori information given by GSWE.

5.1.5.2.1 White exclusion mask

The exclusion mask of phase 1 was computed using filtering condition on green and near-infrared bands are inspired from Braaten-Cohen-Yang cloud detector (Braaten, Cohen and Yang, 2015), along with a filtering that consists in discarding values above 0.22 on the normalized blue band. During phase 2, testing revealed that a complex masking strategy was not necessary to effectively identify white areas and clouds. A simple reflectance threshold on the blue band —set at 0.23— was sufficient to detect and exclude all relevant bright zones, including cloud cover and saturated pixels. This streamlined approach simplifies the processing workflow while maintaining high detection accuracy.

5.1.5.2.2 Shadow mask

Shadows are an important source of commission for water mapping using optical sensors. Indeed, the low reflectance areas that are resulting from shadows due surrounding topography, cloud and urban buildings lead to misinterpreted water indexes, that may often be misclassified as water. If direct automatic detection of shadow may be tricky from a single Sentinel-2 tile, we can surround this limitation with high-resolution DEM for what is linked to topography induced shadows.

Sentinel-2 data are generally acquired about the same hour locally (+/- 30min) all year long for the same orbit. Topography does not change, so in the end the only thing that changes from one date to another is the sun's location in the sky, and thus its shadow from the local topography. Luckily for us, with only the date and the local hour, algorithms already exist to estimate the shadow from the local topography. The main issue will thus be the different parameters to use, and how to easily stock this information without taking too much space to make it the most suitable ancillary data.

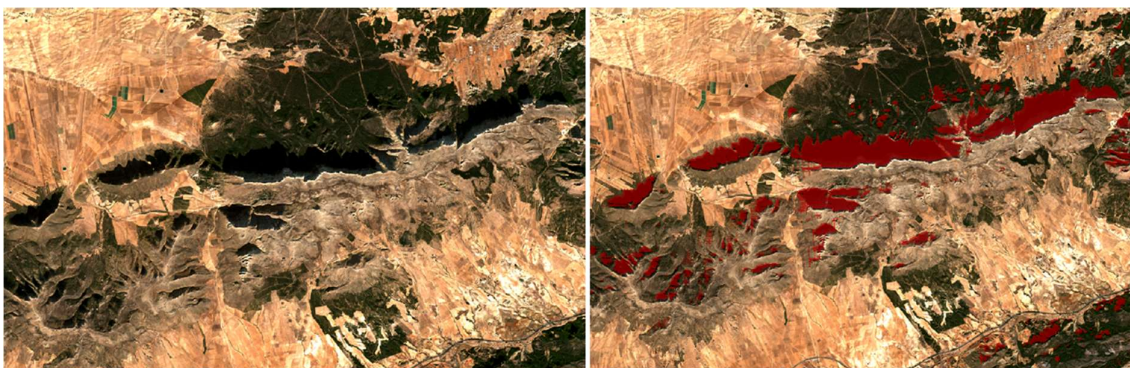


Figure 68. Example of shadow impact (left) and estimate (right) in Southern Spain.

If the shadow can be directly estimated over a whole Sentinel-2 tile in a matter of seconds with the proper DEM at hand, the production of ancillary shadow data can be done before the water body calculation. This calculation needs a few assumptions though to improve accuracy and precision, listed hereafter:

- On the same orbit, the sentinel-2 acquisition time over a tile is roughly done at the same local time (i.e around 10h30 AM).

- The topography data does not change in time; the use of Copernicus DEM GLO-10 or GLO-30 (depending on whether the tile is located in Europe) is suitable for such a use.
- There is no change in sun location from one year to another for the same date (from one year to the next, the difference in the Sun's position at the same time and date is usually a few arcminutes ~ 1 arcminute = $1/60$ of a degree).
- The shadow does not change significantly within every 15 days period

These assumptions lead to an optimized proposal to produce the shadow ancillary data:

- The shadow masks can be produced every 5th and 21st of each month; hence 24 shadow maps are estimated for a whole year over one tile. For a new Sentinel-2 tile, the closest "shadow date" must then be selected.
- All the 24 shadows information over a pixel can be stocked bitwise, which allows the production of one final file per tile to cover a whole year using int32 format.
- To avoid noises in the shadow maps and a robust approach to slight changes over every 15-day period, a morphologic opening of one pixel followed by a dilation of one pixel is performed to enhance the shadow maps.

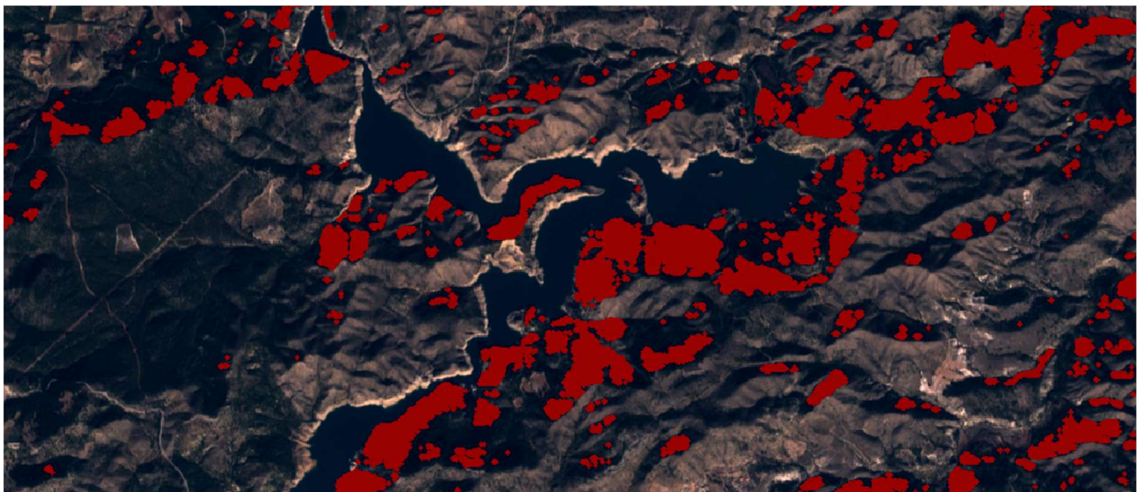


Figure 69. Example of raw shadow estimation with an opening and dilation of one pixel (in red) over tile 30SWH for January 21st.

The ancillary data produced may then be used for water body estimation.

5.1.5.3 “Permatemp” Products computation

5.1.5.3.1 Concept

The separation of water pixels into permanent and temporary water bodies is an enhancement we investigate in this activity. As a request for the new water body product within the EVOLAND project, this classification deals with the timewise nature of a water body. Nevertheless, the very definition of a temporal water body varies a lot from one definition to another. From our side, we chose to consider a

pixel as being a permanent water pixel if it is observed over a whole year as being water. If it is only partially observed as being water in time, it is then considered temporary. As our water classification approach relies on a pixel-by-pixel basis, the permanent/temporary classification is done at a pixel level. Thus, the permanent/temporary classification cannot be done over a whole water body, but only on a pixel-by-pixel basis. An example is shown in the following figure.

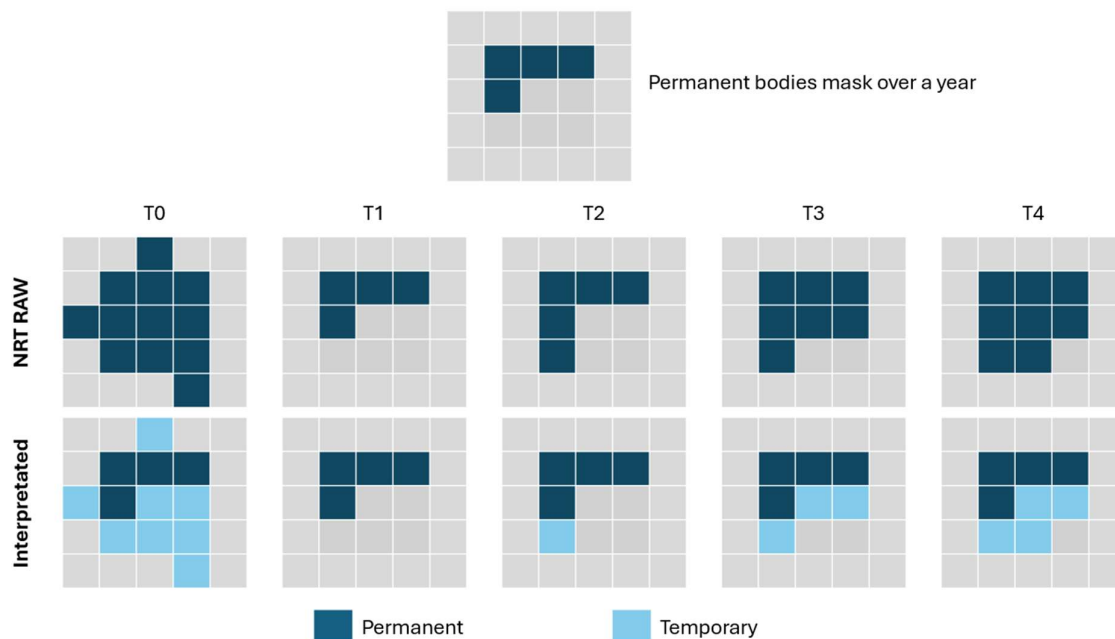


Figure 70 -Concept of permanent/temporal water pixel classification

To allow the classification of previously unobserved water body as being permanent within this frame, the probability of a pixel being water or not must be continuously updated to that endeavor.

5.1.5.3.2 Assumptions made

This classification needed some assumptions before being developed and applied:

- The updating of the permanent body mask over a year is enough to separate the permanent/temporal water pixels
- The clouds are properly detected. As the cloud detection algorithm we use has defects (the constraint being the calculation speed), a small margin must be considered to avoid being too protective with time.
- The permanent body mask is updated monthly from a composite of the previous month. For month M, the permanent body mask is updated from the NRT water masks from months M-1, and missing values are filled in depending on the number of observations and data acquired during months M-4 to M-1.

5.1.5.3.3 Initialization

The initial permanent mask has been tested from two different methodologies: using GSWO, or from one year of data using our own algorithm.

GSWO:

The initialization can be done using the Global Surface Water occurrence dataset and selecting the pixels with occurrences >80% as a first approach. The advantage is that it is straightforward and global. The inconvenience is that the occurrences from 40 years might not correspond to the ones observed over a particular tile for recent years.

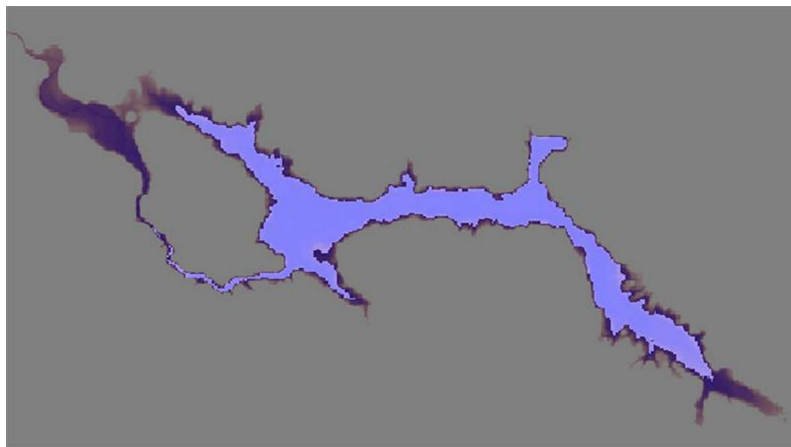


Figure 71. Example of permanent mask (purple) derived from GSW occurrences above 80%.

Own algorithm:

Using a whole year of raw water body classifications derived from the initial likelihoods (in general we used 2020), we calculated the occurrences excluding missing and clouded data. The values over 70% of occurrences were then kept for the initial permanent water mask. The main inconvenience encountered is that in some areas cloud cover can lead to imperfect classification even over a whole year.

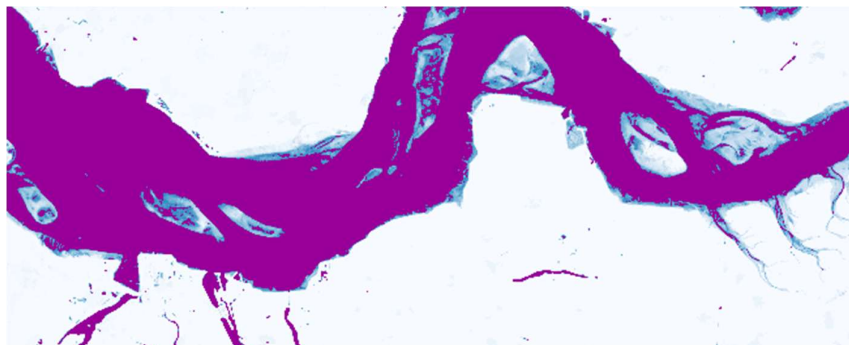


Figure 72. Example of permanent mask (purple) derived from a whole year water body occurrence.

5.1.5.3.4 Rolling a priori

The key element for operational use of this approach is the establishment of a rolling permanent water mask, updated at a monthly pace. This calls for two intricate approaches:

- The monthly synthesis, which will become the last entry of the rolling permanent water mask (updated monthly). The main issue here is the accuracy of the entry water maps, which can be deteriorated due to multiple effects (clouds, cloud shadows, sun glint and other effects that usually impact water surfaces).
- The update of the rolling permanent water mask, which is done bitwise, forgetting the oldest water map, and needs to be initialized. Using an INT16 coded raster, the size is optimized and permits working bitwise.

The figure below shows the rolling permanent map concept.

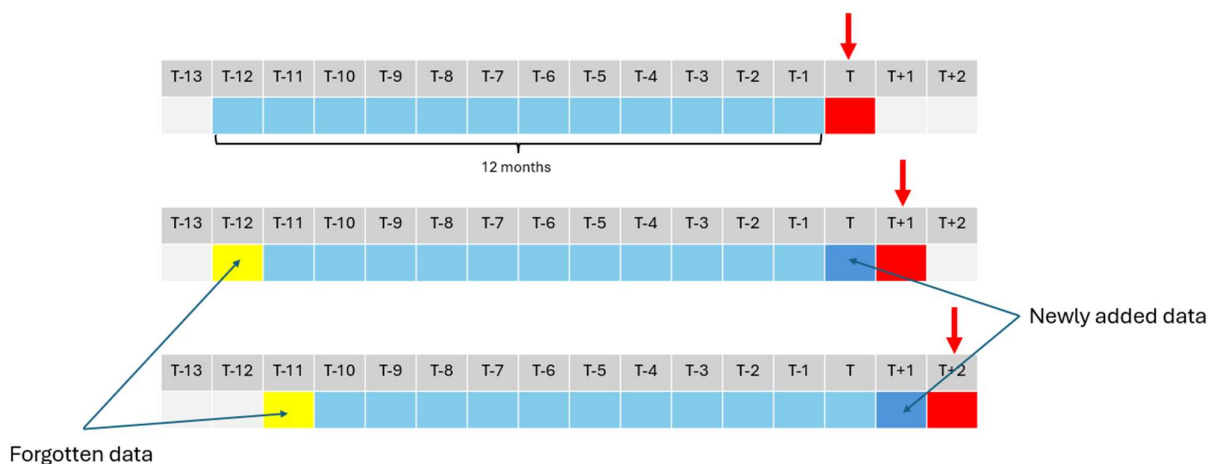


Figure 73. Rolling map updating for three consecutive months.

In practice, after a whole year, the permanent water mask might strongly differ from the initial mask. One way to limit this impact is to reprocess the first study year with its final permanent water mask, using the second methodology exposed in the previous section.

During the monthly synthesis calculation, the number of times a pixel is observed as being water is a key element for the success of the whole approach. Misclassifications are in this frame of major importance. To provide tolerance to the approach, a threshold can be used within a month to avoid applying too much weight to misclassifications – mainly clouds and their shadows for instance. If a pixel is not sufficiently observed at a monthly scale (cloud cover is too important or too few viable observations are available), the output of the previous month can be duplicated, or statistics on the last 4 months can be used to fill in the missing values.

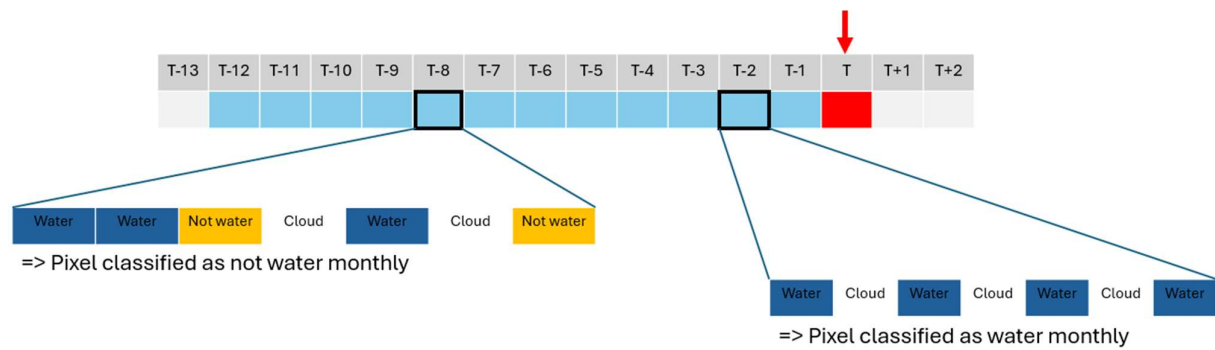


Figure 74. Example of monthly synthesis taking as an example one pixel in time.

From this rolling permanent water map, an NRT raw water body can then be split into permanent and temporary water pixels as shown in the next figure. The advantage of this approach is that it is suitable for operational use with minimum calculation time.

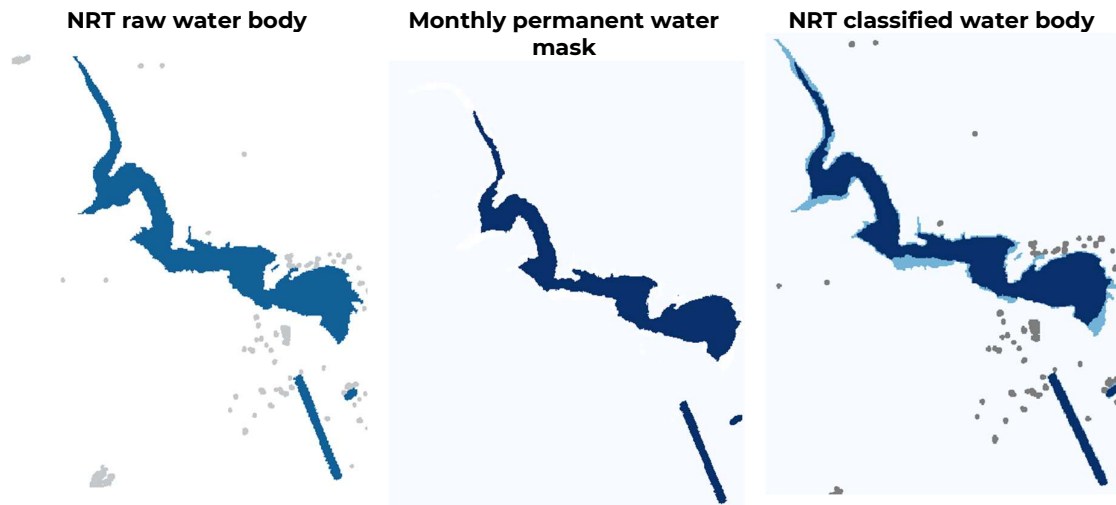


Figure 75. Example of permanent (dark blue) and temporary (light blue) water pixels classification (right) from the raw NRT water mask (left), using the rolling map (middle) on Cuevas de Almanzora Lake (Southern Spain) on September 10th, 2022.

5.1.5.4 Synthesis scheme

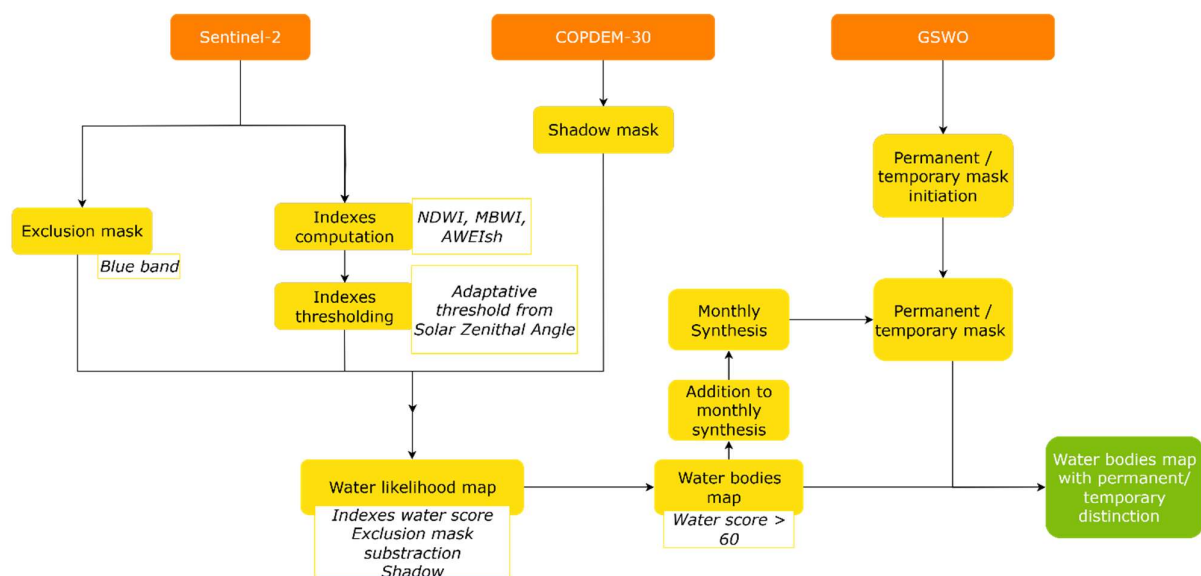


Figure 76. C7 prototype workflow during phase 2.

5.1.6 Results and Validation

5.1.6.1 Results overview on test sites

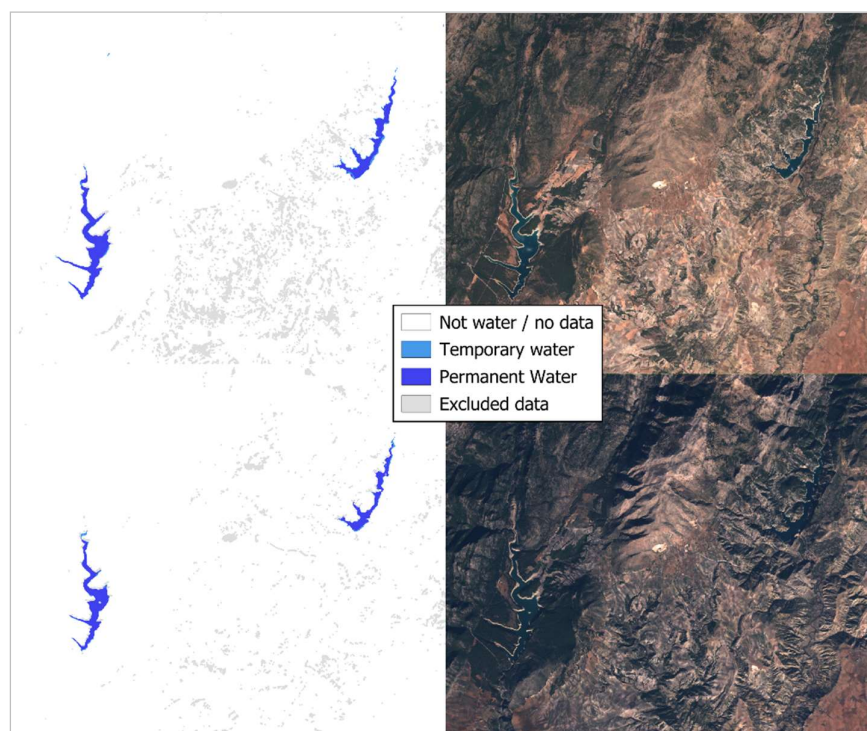


Figure 77. C7 prototype results (summer above, winter below) in Spain.

Figure 77 presents the classification results obtained for the Spanish site, with summer conditions illustrated in the upper panel and winter conditions in the lower panel. Overall, the performance is notably strong across both seasons. The shadow masking algorithm proves particularly effective in this region, with complex topography. Minimal cloud cover contributes to favourable detection conditions. Seasonal luminosity variations are evident: the site exhibits low brightness during winter and significantly higher levels in summer. This contrast is well managed by adapting the classification threshold based on the solar zenith angle (SZA), which enhances the robustness of the method under varying illumination conditions.

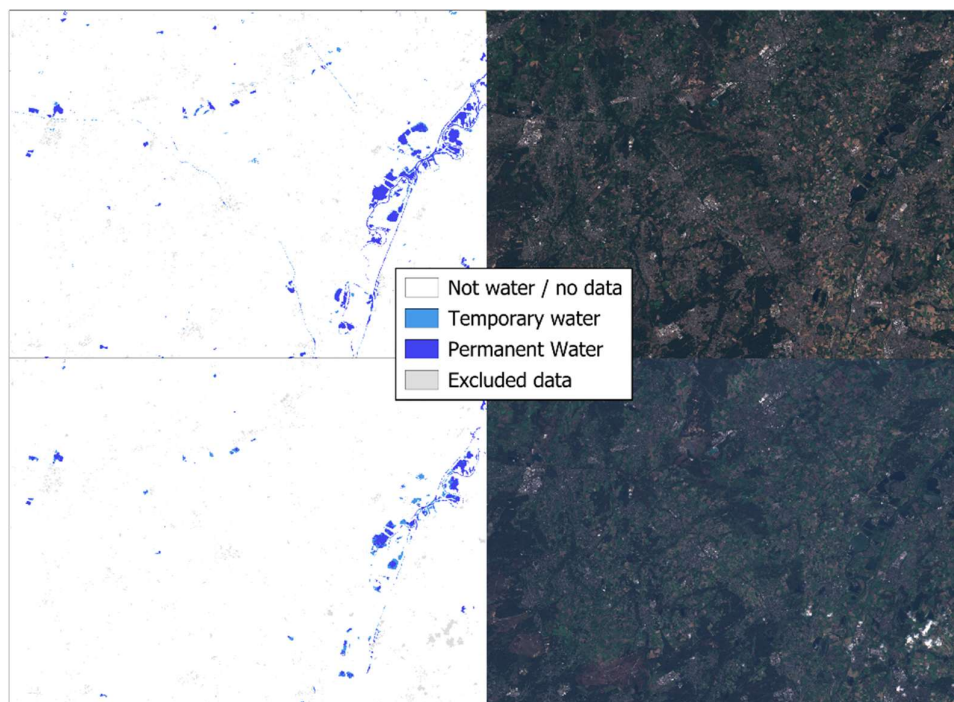


Figure 78. C7 prototype results (summer above, winter below) in Belgium.

Figure 78 presents the classification results obtained for the Belgium site, with summer conditions illustrated in the upper panel and winter conditions in the lower panel. This site is more complex than Spain, with a lot of built-up areas and, and small water bodies. We observe good results with some commissions on built-up areas and shaded vegetation.

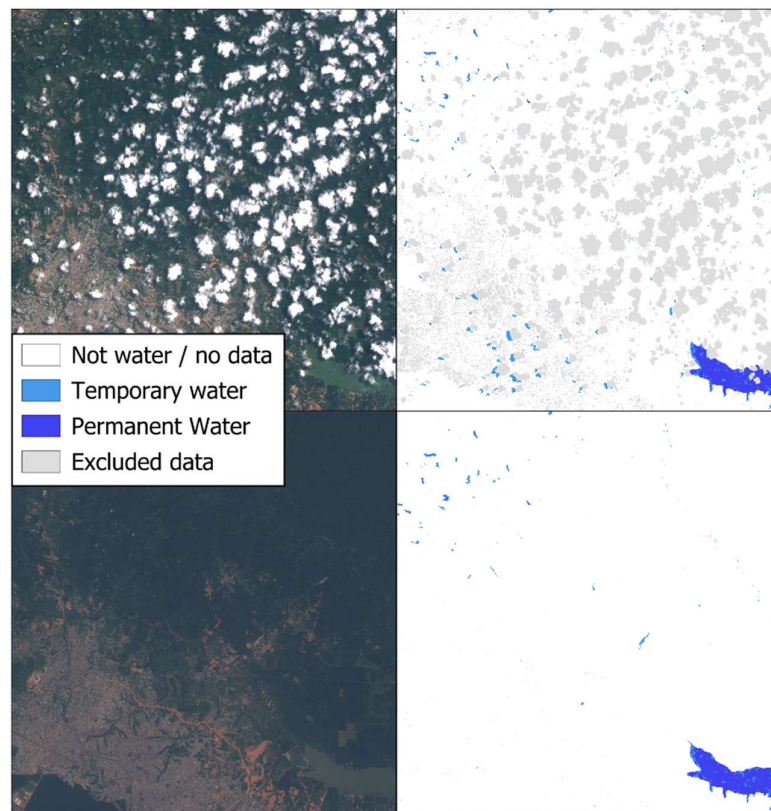


Figure 79. C7 prototype results (humid season above, dry below) in Ivory Coast.

Figure 79 presents the classification results obtained for the Ivory Coast site, with humid conditions illustrated in the upper panel and dry conditions in the lower panel. We observe good result when there are no clouds. Otherwise, the algorithm is still sensitive to cloud shadow, and this site is the more often clouded one.

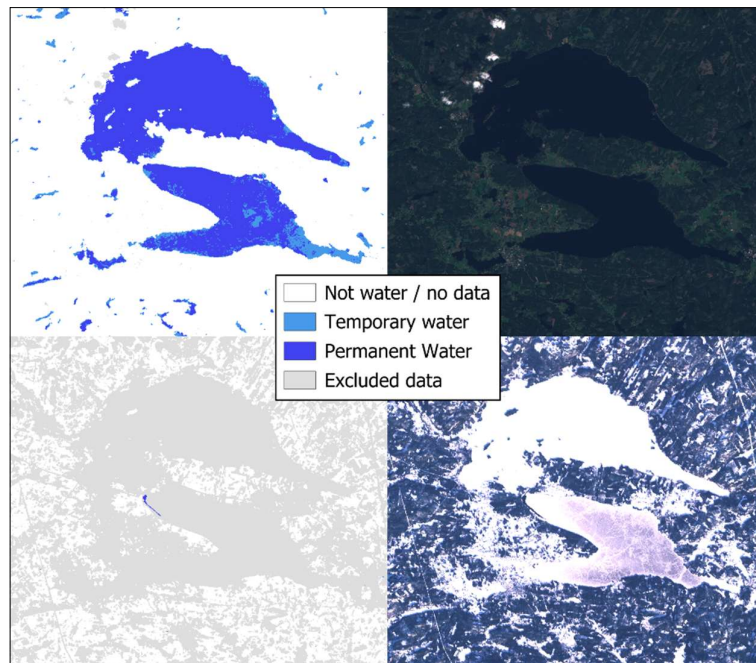


Figure 80. C7 prototype results (summer above, winter below) in Sweden.

Figure 80 presents the classification results obtained for the Sweden site, with summer conditions illustrated in the upper panel and winter conditions in the lower panel as well. We can observe that liquid water is well detected even in winter. Snow and ice are excluded from the product thanks to the exclusion mask. The few commissions we can observe in Sweden are generally in winter in shaded vegetation

5.1.6.2 Thematic accuracy

Using the same metrics as for the Imperviousness Prototype (6.1.6.2) we proceeded to the computation of thematic accuracy.

All products were assessed using the described methodology:

1. Thresholding of GSWO data (occurrence > 30% to select all pixels which could be water) to use it as reference data.
2. Transformation of NRT Water bodies map to binary map of water to use as a product to qualify.
3. Random Sampling of 110 points on each tile (80 on land, 30 on water according to the product). No sampling has been made on excluded pixels.
4. Comparison between reference data and the product to extract the number of True positives and negatives (TP, TN), False positives and negatives (FP, FN).
5. Human supervision (GIS) to correct eventual misclassification (among TP, TN, FP, FN) due to the source of the reference data.

6. Statistics (Ponderation and accuracy computation) on results recapitulated in following tables.

As outlined in deliverable D3.8, the estimation of water-to-soil proportions was refined to improve spatial relevance. These values were derived from the Global Surface Water Occurrence (GSWO) dataset on a per-tile basis, allowing for more accurate regional differentiation. Unlike Phase 1 of the project, where validation was based directly on GSWO data, the current phase adopts a more targeted approach: sampling is now performed using the Near Real-Time (NRT) water bodies map. This methodological shift ensures a more contextually relevant and dynamic validation process, better aligned with operational conditions.

The results on the four test sites are presented in the four following tables.

Table 43. Confusion matrix for the Belgium test site.

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	53,09	0,00	53,09	User Acc.	CI95%	Commission
	Water	11,95	254,97	266,92	95,52%	7,00%	4,48%
	Total	65,05	254,97	320,01			
		Prod. Acc.	100,00%		Overall Acc.	96,27%	
		CI95%	0,00%		CI95%	5,84%	
		Omission	0,00%				

Table 44. Confusion matrix for the Spain test site.

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	55,52	0,00	55,52	User Acc.	CI95%	Commission
	Water	9,67	546,51	556,18	98,26%	4,78%	1,74%
	Total	65,19	546,51	611,71			
		Prod. Acc.	100,00%		Overall Acc.	98,42%	
		CI95%	0,00%		CI95%	4,34%	
		Omission	0,00%				

Table 45. Confusion matrix for the Ivory Coast test site.

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	58,31	0,00	58,31	User Acc.	CI95%	Commission
	Water	254,37	606,00	860,37	70,43%	16,61%	29,57%
	Total	312,68	606,00	918,69			
		Prod. Acc.	100,00%		Overall Acc.	72,31%	
		CI95%	0,00%		CI95%	15,55%	
		Omission	0,00%				

Table 46. Confusion matrix for the Sweden test site.

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	49,90	0,35	50,25	User Acc.	CI95%	Commission
	Water	47,73	170,92	218,65	78,17%	13,69%	21,83%
	Total	97,63	171,27	268,91			
		Prod. Acc.	99,79%			Overall Acc.	82,12%
		CI95%	0,57%			CI95%	11,13%
		Omission	0,21%				

Across all sites, producer accuracy is high, exceeding 99%. This indicates a very low omission rate: pixels classified as water are indeed water, and almost no actual water pixels were misclassified as non-water. The only observed false negative occurred over a lake in Sweden, where a thin cloud layer interfered with the classification process, leading to a misdetection.

Otherwise, user accuracy is lower. Commission errors were observed, with several land pixels misclassified as water, particularly in regions such as Ivory Coast and Sweden. Most commission errors are associated with built-up areas, cloud shadows, shaded topography, vegetation shadows, and mixed pixels (e.g., partially water and partially land).

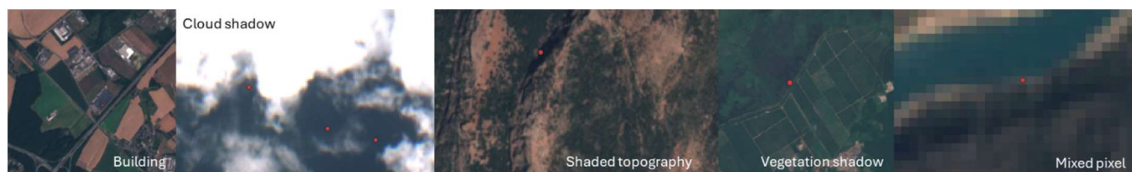


Figure 81. Commissions examples on all sites.

The increase in commission errors compared to Phase 1 is primarily due to the discontinuation of GSWE data as weighting parameters. Ivory coast is the site where there is the most cloud, which explains the high rate of commission (red samples in Figure 82).

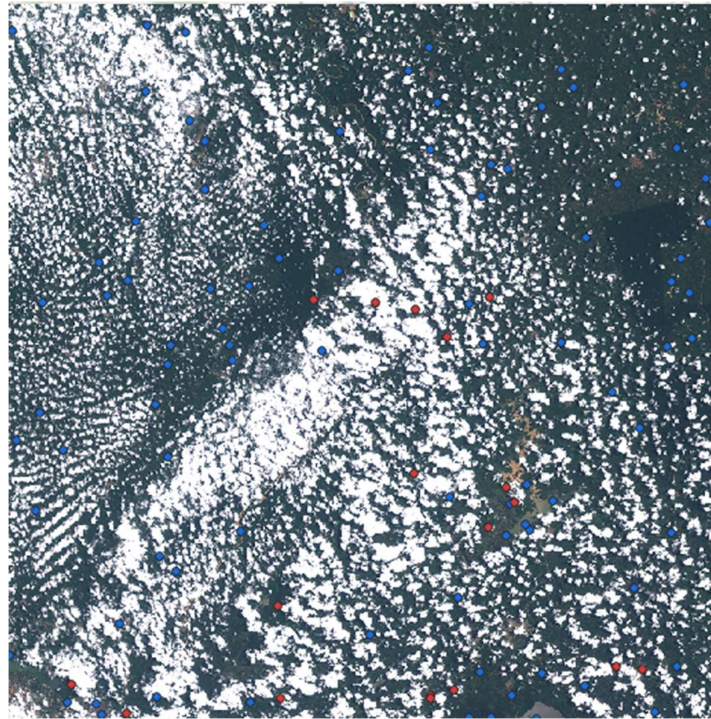


Figure 82. Commissions linked to cloud shadow in Ivory Coast (red points).

Across all sites, *Producer Accuracy* consistently exceeds the required specification threshold of 95%, with values surpassing 99%, demonstrating robust performance. Regarding *User Accuracy*, the results meet specification in Spain and Belgium, confirming reliable detection in these regions. While performance in Sweden and Ivory Coast is comparatively lower—78% and 70%, respectively—these figures remain within a reasonable range given the environmental challenges.

In Sweden, the reduced accuracy observed during winter is primarily due to the combined effects of low solar illumination and increased shadowing, both of which adversely affect detection reliability. While the summer results fully comply with the specified requirements (see Table 47), the winter results do not meet the expected user accuracy (see Table 48), thereby underscoring the seasonal variability in performance. To maintain the desired level of accuracy, it may be beneficial to filter input images based on their solar zenith angle. The widespread presence of snow cover which cannot be detected as water across most of Sweden during this period further supports the relevance of this approach.

Table 47. Confusion matrix for the Sweden test site (Summer).

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	50,80	0,00	50,80	User Acc.	CI95%	Commission
	Water	0,00	216,76	216,76	100,00%	0,00%	0,00%
	Total	50,80	216,76	267,56			
		Prod. Acc.	100,00%			Overall Acc.	100,00%
		CI95%	0,00%			CI95%	0,00%
		Omission	0,00%				

Table 48. Confusion matrix for the Sweden test site (Winter).

		REFERENCE					
		Not Water	Water	Total			
MAP	Not Water	49,00	0,71	49,70	User Acc.	CI95%	Commission
	Water	96,31	124,27	220,59	56,34%	16,43%	43,66%
	Total	145,31	124,98	270,29			
		Prod. Acc.	99,44%			Overall Acc.	64,11%
		CI95%	1,10%			CI95%	13,42%
		Omission	0,56%				

In Ivory Coast, the lower accuracy is primarily attributed to the presence of persistent cloud cover throughout the year, which contribute to false detections. During the dry season, when cloud cover is reduced, the prototype achieves the expected accuracy level (greater than 95%). Incorporating a cloud shadow model as an input variable could significantly improve the overall performance.

Despite these limitations, the overall performance remains commendable, particularly considering the diverse and

5.1.7 Conclusions

The main challenge between phase 1 and phase 2 was to remain as accurate without constraining to the GSWE mask. This was needed to be able to detect new water bodies, which was one of the requirements from the URD. The direct impact of not restraining the water detection area was a strong rise in the commissions, mostly due to shadows from either topography or low sun zenith angle. To tackle this issue, two main tasks were carried out:

- Normalization of reflectance with the sun zenith angle; this led to more stable reflectance, and thus to more homogeneous water indexes. This diminished commissions at higher latitudes mostly.
- Production of topographic shadow masks to identify complex areas. Depending on the season and on the latitude, the hills and mountains shadow could indeed be mistakenly classified as water.

These two additions to the algorithm allowed to overall accuracy on sites spreading between 72.31% (worst case in Ivory coast) and 98.42% (best case in Spain). The main

limitation of the proposed approach corresponds to snowy areas, low sun zenith angle sceneries, and shadows, along with proper cloud shadow detection. If the overall quality lowered slightly from Phase 1 to Phase 2, its coverage rose significantly. About results on Sweden in winter, it is to report that some water mask producers choose not to consider sceneries with low sun zenith angle to avoid such commissions.

Aside from the NRT water detection algorithm improvement, a novel approach is proposed to classify water pixels between permanent and temporary classes. This approach relies on the production of permanent water masks updated monthly, with occurrences from previous months used to tackle missing information linked to heavy cloud cover or low observation numbers. This approach proved to be close to an operational level for NRT classified water mask production. This classification is not only a novelty in Phase 2, but also something not proposed operationally in NRT water mask production currently done to our knowledge.

5.1.8 Outlook beyond the EvoLand project

The algorithm developed is close to an operational level and could be deployed in an operational environment for NRT or lower pace water masks. It answers all requirements listed in the URD and proposes a robust approach towards an NRT production. Nevertheless, some elements of the algorithm could still be improved to improve the accuracy without adding too much calculating time.

The first axis of improvement resides in the sun zenith angle normalization of the reflectance. If what is proposed in the current version of the algorithm showed very promising results, the thresholding of the indexes for the raw water masks production could be adapted to the sceneries observed to improve the water/land/urban areas discrimination.

From the four test sites, the lowest accuracy was observed in winter Sweden. This was mainly due to very low sun zenith angle, but also to the presence of snow. Our product would benefit from the outputs from snow cover producers at it is provided currently within CLMS for instance. The second improvement axis is thus to better tackle snow- and ice-covered areas, to better detect liquid water.

Finally, the improvement of cloud and cloud shadow detection algorithms is key for NRT water mask production. If it is an old and common issue with optical data, the major issue resides in the calculation time and the ancillary data needed depending on the algorithm used. During Phase 1, the cloud mask from the Braaten-Cohen-Yang algorithm proved to be efficient but rises the calculation time and might be insufficient for cloudy veil for instance. The cloud shadow detection is trickier and is one key axis for commission diminution during water mask production.

6 Urban prototypes

The following chapter presents the two urban CLMS prototypes of Task 3.4, which include: i) T 3.4.1 Continuous Imperviousness Monitoring, and ii) T3.4.2 Automated Land Use Mapping of Urban Dynamics.

6.1 C8 – Continuous Imperviousness Monitoring

6.1.1 Introduction & Objectives

The HRL imperviousness product monitors the percentage of soil sealing and captures the spatial distribution of artificially sealed areas including the level of sealing of the soil per area unit every three years.

Most CLMS products are updated on an annual or triannual basis. The HRL Imperviousness layer is currently produced every 3 years. The main objective of the candidate is to demonstrate the potential for continuous monitoring of imperviousness degree and demonstrate incremental updates of HRL Imperviousness. Continuous monitoring techniques and integration of new EO data will be assessed for this prototype development.

The initial objective was to develop a prototype service with a finer spatial resolution (5 meters) and a monthly update frequency. During Phase 1 of the project, the focus was placed on consolidating and testing the algorithms, which led to implementing the prototype at six-month intervals instead. Although monthly updates were initially planned for Phase 2, feedback gathered during a user webinar held between the two phases revealed that users preferred a six-month update cycle. During this webinar, stakeholders and end users responded to a series of questionnaires, including one on their preferred update frequency.

The algorithms have been developed with the objective of maintaining the level of quality (thematic accuracy) of the products over time (see section 6.1.2 for the existing portfolio), thereby ensuring temporal consistency in the times series. Table 49 summarizes the prototype specifications for the candidate C8.

Table 49. Technical thematic specifications for the candidate C8.

Prototype C8	
Sub-prototype	C8 Continuous Imperviousness Monitoring
Prototype sites	2 – Southern France 3 – Belgium 6 – Central Europe
Description of the sub-prototype	Monitors the percentage of soil sealing and captures the spatial distribution of artificially sealed areas including the level of sealing of the soil.
Expected temporal update	Continuous (every 6 months)

Spatial resolution	5m
MMU	1 pixel
Envisaged accuracy	HRL IMD 2018/HRL NVLCC 2021 as basis: min 90% UA/PA for urban/sealed classes

6.1.2 State of Play

Soil health is of increasing concern in Europe, as a combination of climate change and intensive human activities places increasing pressure on soils. Soil degradation manifests in various ways, including soil sealing, erosion, desertification, and reduction in soil organic carbon.

Given the importance of soil health, the EU has developed a soil strategy for 2030, which describes actions that the EC will take around soil health.

The most mature use of CLMS for soil health is in monitoring soil sealing by means of the HRL for imperviousness. The HRL imperviousness product monitors the percentage of soil sealing and captures the spatial distribution of artificially sealed areas including the level of sealing of the soil per area unit every three years.

To date the IMD layer has been produced for five reference years: 2006, 2009, 2012, 2015 and 2018. The reference years 2021 and 2024 are actually in production as part of the HRL Non-Vegetated Land Cover Characteristics (NVLCC) with the introduction of a new layer, the Permanent Bare Surfaces (PBS). For each of the iterations from 2009 onwards the results have contained two products: a status layer for any reference year (e.g. IMD2018), as well as an imperviousness density change layer between reference years (e.g. IMC2018 is 2015 to 2018) and based on the already existing imperviousness product for that previous reference year. The HRLs were originally produced at a 20 m spatial resolution, but for the 2018 versions this has been improved to 10 m spatial resolution as all imagery could be sourced through the Copernicus Sentinels (1 & 2).

Although not strictly part of CLMS, it is also worth highlighting the Global Human Settlement Layer¹⁰ produced by JRC in the framework of the Copernicus Emergency Management Service.

But under the new Copernicus 2.0 programme, CLMS aims to continue providing its core Copernicus information products and services for the 2021-2027 period, thus building time series that enable for long term change monitoring.

To address future needs, EvoLand conducted targeted interviews as part of the WP1 to identify the key geospatial needs required to support the implementation and monitoring of these policies. Following the results of WP1, user requirements were identified for each prototype, and particularly for the prototype C8. The CLMS HRL IMD and NVLCC products already partly meet the requirements for the imperviousness change detection

¹⁰ <https://ghsl.jrc.ec.europa.eu/about.php>

Further requirements include that the change products should support analysis of what land is being lost to soil sealing. Not being part of the technical specifications of the CLMS products, this requirement has not been addressed.

Additional requirements state that the products should be available in at least the EU27 and ideally the EEA38. The CLMS products already meet these requirements. There is additional interest in worldwide coverage (but not considered a priority).

6.1.3 Selection of Prototypes sites

For the Phase 2 of the project, the implementation was extended on the full extent of the in 3 test sites in Europe, namely Southern France, Belgium and Central Europe. Due to the absence of in-situ or reference data on imperviousness, the Côte d'Ivoire test site could not be processed. Given the low priority assigned to global product coverage in the user requirements, efforts have been concentrated on the three European test sites. The location of the test sites is presented in section 2.2.

6.1.4 Input and Ancillary Data & Pre-processing

6.1.4.1 Input Data

Inputs EO data are Sentinel-2 (Sentinel-2A & B) level-1C (top-of-atmosphere) for the production of both status and change layers.

For Phase 2, with the aim of continuously monitoring the status and dynamics of the degree of imperviousness, all the scenes covering the time frame from 01-January 2020 to 31-December 2022 were selected to produce the impervious products for all three test sites.

The pre-processing step of the Sentinel images includes all the pre-processing required to prepare the data which can be listed as: conversion from Level-1C (top-of-atmosphere) to Level-2A (bottom-of-atmosphere) including atmospheric correction, topographic normalisation pre-processing for S-2, cloud masking, the L3A Sentinel monthly synthesis – and finally the super-resolution (following the outcome of the WP2).

For the Phase 2 of the project, monthly syntheses were processed using the WASP algorithm (open-source solution developed by CNES). For each pixel and each band (10 and 20m bands), the WASP algorithm computes the monthly synthesis using a weighted average of the cloud free surface reflectance's gathered during a synthesis period of 91 days. Cloud-free pixels as close as possible to the "centre-date" are used to build a cloud-free image. For this synthesis, the algorithm considers all images +/- 45 days from the centre date and takes the cloud-free pixel closest to it.

Table 50 lists and summarizes the input data used for the C8 prototype development. The MAJA¹¹ processor is used for these pre-processing. Following the

¹¹ <https://www.cesbio.cnrs.fr/maja/>

outcomes of the WP2, 5m super-resolved images from Sentinel-2 L3A products have been generated using Single Image Super-Resolution model trained in the frame of the EvoLand project.

Table 50. list of input datasets used for the C8 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2	Sentinel-2 L3A monthly syntheses, bottom-of-atmosphere including atmospheric correction, topographic normalization preprocessing for S-2, cloud masking- and finally the super-resolution	5 m	Monthly synthesis with a revisit frequency of 5 day for the S-2 L2A

6.1.4.2 Ancillary data

The following ancillary datasets are used for the prototype C8 during the Phase 2:

- The HRL NVLCC and VLCC 2021 (see section 6.1.2).
 - The HRL IMD and IBU 2021 for the NVLCC layers
 - The HRL GRA, TCD and CTY 2021 for the VLCCC layers
- The HRL SWF 2018
- The OpenStreetMap (OSM)

The ancillary data are used as input for the Machine Learning algorithms and the absolute calibration of the Imperviousness degree. Indeed, a set of training data is required and for this purpose, the historical HRL layers have been selected as, along with the OSM data to derive reliable training samples.

Next table summarizes the list of ancillary data used for the C8 prototype development.

Table 51. list of ancillary datasets used for the C8 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
HRL IMD 2021	Provides at pan-European level in the spatial resolution of 10 m the sealing density in the range from 0% to 100% for the 2021 reference year.	10m	3-year
HRL IBU 2021	Provides at pan-European level in the spatial resolution of 10 m binary information (building / no building) within the sealing outline derived from the Imperviousness Density 2018	10m	3-year
HRL GRA 2021	Provides at pan-European level in the spatial resolution of 10 m a basic land cover classification with two thematic classes (grassland / non-grassland) for the 2021 reference year	10m	3-year
HRL TCD 2021	Provides at pan-European level in the spatial resolution of 10 m the level of tree cover density in a range from 0% to 100% for the 2021 reference year.	10m	3-year
HRL CTY 2021	Provides at pan-European level a classification of cropland into one of	10m	3-year

	the 19 crop type classes at a spatial resolution of 10 m and a MMU of 0.25 ha for the 2021 reference year.		
HRL SWF 2018	Provides at pan-European level information on information on linear structures such as hedgerows, as well as patches of woody features for the 2018 reference year.	5m	3-year
OpenStreetMap (OSM)	OSM is a free, open map database The that aims to collect data about stationary objects throughout the world, including infrastructure and other aspects of the built environment, points of interest, land use and cover classifications, and topography.	Vector	Continuous

6.1.5 Methodology and Workflow

6.1.5.1 Status layer production

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. The achieved level of automation ensures the effective application of the process to map impervious areas of almost the entirety of Europe.

Consequently, the production of the status layer has been completed for all the S-2 tiles of the three test sites over the entire six-month periods.

The workflow/methodological steps to produce the Imperviousness products are described in the next sections.

6.1.5.1.1 Biophysical variables and additional image parameters

For the purpose of the classification task, the following spectral indices are derived for each Sentinel-2 image:

- Normalized Difference Tillage Index (NDTI), commonly used to evaluate tillage intensity and crop residuals thanks to the difference of reflectance between the two Sentinel-2 shortwave infrared bands.
- Normalized Difference Built-up Index (NDBI), is the normalized difference between short wave infrared (SWIR1) and near infrared (NIR) reflectance, used to detect urban areas.
- Normalized Difference Vegetation Index (NDVI) highlights vegetation covered areas by exploiting red and near infrared reflectances.
- Normalized Difference Water Index (NDWI) uses the difference between green and near infrared reflectance to detect water bodies.

Therefore, the here above indices solely rely on multispectral information. This information is essential to discriminate landscape elements but is not sufficient. Textural indices represent advanced features for a more effective detection. Amongst numerous existing techniques, the PanTex index is extensively used for the extraction of the built-up areas (Pesaresi, et al., 2008) and is derived during the Phase 2.

Additionally, unused spectral bands not covered by these indices are included in the classification input to ensure full spectrum coverage.

Signal decomposition techniques are also used. They provide a multi-resolution representation of the original image in a series of components related to a specific direction. More precisely, the Haar wavelets which are used for face recognition and more generally for object recognition are generated for the EvoLand project.

6.1.5.1.2 Derivation of classification training samples from additional reference data

As input for the Machine Learning algorithms, a set of training data is required. The historical HRL NVLCC, VLCC and SWF have been used as training data, along with the OSM data to derive reliable training samples for the roads network. Samples in built up and non-built-up areas have been selected.

Based on the spectral information, biophysical indicators and texture parameters at the training sample points, the algorithm 'learns' how to classify the features (Tan et al. 2006, Camp-Valls, 2009) and identifies the most efficient combinations of input parameters to differentiate built-up areas from other land cover.

6.1.5.1.3 Production of initial built-up masks by automated supervised classification (Machine learning)

Then, the next step is the production of the initial built-up masks (binary sealed/non-sealed layer). The proposed methodology relies on two main components which are 1) feature extraction based on tree representation and attribute profiles, and 2) a semi-supervised classification using random forest algorithm.

Tree representation and Differential Attribute Profiles (DAP)

A hierarchical representation of the content of the images is made by using morphological trees, here inclusion trees. The main goal of the tree build is to create what we call the Attribute Profiles (AP).

DAP (Dalla Mura, Benediktsson, Waske, & Bruzzone, 2010) has been built upon Differential Morphological Profile (DMP) to achieve more discriminative power, a higher flexibility, for a lower computational cost and has been embedded in large-scale analysis performed by Joint Research Centre (JRC) such as the Global Human Settlement Layer (GHSL, Release 2019) (Pesaresi, et al., 2013) and European Settlement Map (ESM Release 2019) (Sabo, et al., 2019).

Supervised classification (Machine learning)

Among the different supervised classifiers available, we will use the Random Forest classifier that combines multiple decision trees to increase the robustness of the overall classification process. Furthermore, it comes with parallel, scalable, open-source implementations such as the Shark library for the Machine Learning classification. The preference for Machine Learning over Deep Learning is based on

the fact that it is less time-consuming and requires less training data. The functional architecture of the classification process is illustrated in Figure 83.

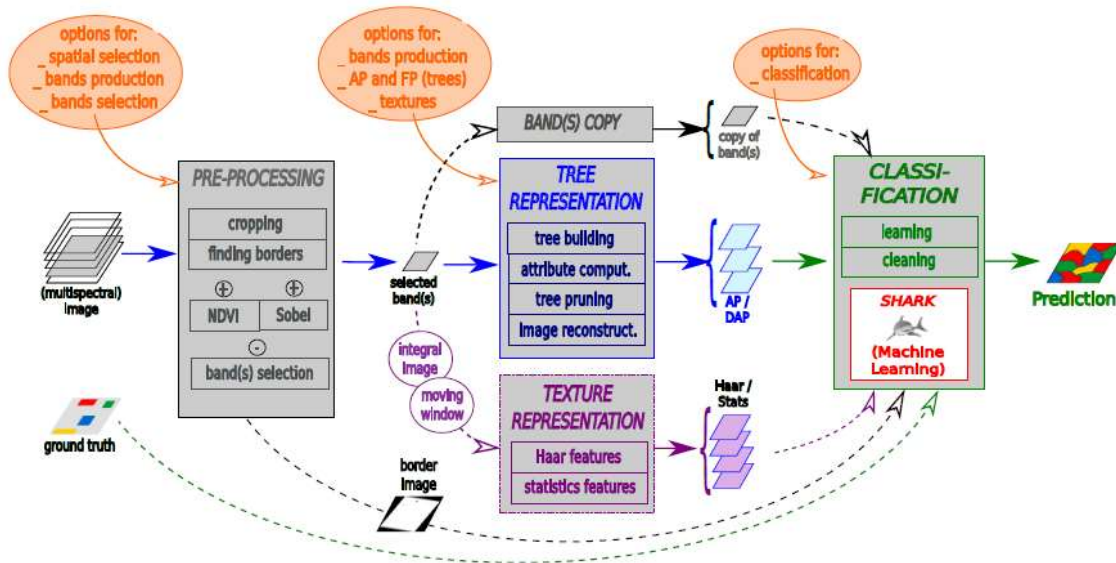


Figure 83. Functional architecture of the classification process (Source: Broceliande Manual - User Guide for Broceliande - François Merciol – Jérôme Moré).

The process involves single Image classification. This allows for continuous monitoring, for each L3A monthly synthesis within a 6-month period. To enhance the reliability of the classification maps, all 6 resulting binary built-up masks are aggregated to into a final built-up mask. A pixel is considered as built-up or non-built-up based on its status across the 6-month period (see Figure 84). To reduce the impact of potential outliers, a pixel is considered as built-up status if it appears at least five out of six-monthly.

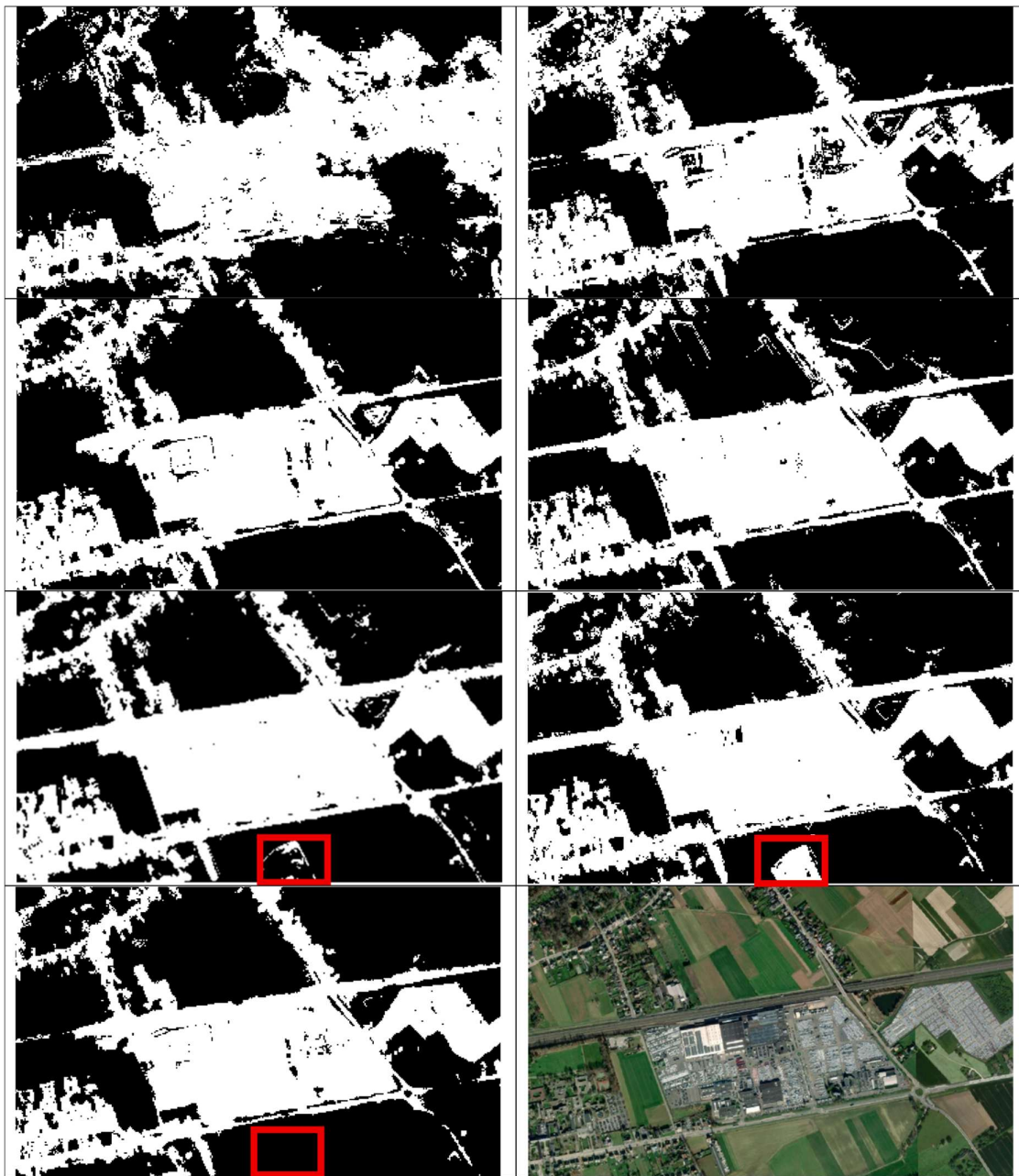


Figure 84. Temporal synthesis of built-up area classification. Panel shows monthly binary built-up/non-built-up masks for tile 31UFS from January to June 2021 (first three rows, left to right). The integrated mask (bottom left), derived from the logical aggregation of all six-monthly masks, is compared against ESRI basemap imagery (bottom right). The final composite mask is more accurate because the aggregation method reduces inconsistencies. For instance, it excludes the agricultural field (red box) that was falsely classified in May and June 2021 at the bottom.

6.1.5.1.4 Absolute calibration of the Imperviousness degree

Besides the production of initial sealed and non-sealed masks (previously described in section 6.1.5.1.3), one key step in the HRL Imperviousness production is the estimation of the degree of imperviousness. Each single pixel in the built-up

mask will be assigned an imperviousness density value of 1 to 100%. The linkage between the spectral variables and the IMD measurements is done through an absolute (linking the spectral variables to IMD) and relative calibration procedure. This combination improves the accuracy of imperviousness density estimates, correct any over-/underestimation of values and assure comparability and consistency over time.

The HRL IMD 2021 serves as calibration input for an absolute calibration of the IMD measurements. For the prediction of the imperviousness degree, a linear regression method is used to model the relationship between the collected reference samples and a meaningful metric from the biophysical variables: the NDVI max. The established linear equation is applied to transform the input data into imperviousness degree values between 1 – 100%. This results in absolutely calibrated IMD measurements.

6.1.5.1.5 Post-processing

A lot of commission errors were identified as coming from fields. Their spectral signature considerably changes whether it is bare or covered with vegetation which makes them very likely or very unlikely to be misclassified impervious areas, depending on the season. Unlike Phase 1 where a crop mask was applied, using such a mask to the classification results will exclude areas that were previously identified as cropland but later transitioned into sealed. This could lead to an underestimation of newly urbanized zones, as the crop mask filters out pixels regardless of their land cover change.

To minimize the scattered noise, isolated pixel groups smallest than 4 pixels are removed. Subsequently a morphological closing operation with a 3x3 square kernel is applied to fill gaps and smooth features. An example is shown in the Figure 85 below.



Figure 85. Example from tile 31UFS showing the combined binary mask (left) and the result after filtering (right).

6.1.5.2 Change layer detection

The change detection methodology employs a multi-temporal analysis of super-resolved Sentinel-2 L3A imagery to identify urban changes, focusing on

imperviousness degree modifications. The approach includes spectral index analysis, Principal Component Analysis (PCA), and spatial clustering to detect and validate changes with higher confidence.

Change detection was not addressed in Phase 1 and has only been explored in Phase 2. For this reason, processing has only been completed on one tile (31UFS) at the Belgian test site. The aim was to test and consolidate the methodology, rather than producing low-quality results for all the tiles in all the test sites.

6.1.5.2.1 Data pre-processing

As described above, super-resolved L3A time series imagery was used, covering 6-month period. Although L3A products generally contain fewer clouds, a number of acquisitions are still contaminated with clouds and shadows, which can lead to false detections. To mitigate these effects, cloud masks are applied whenever available. For central Europe test site, however, the masks were often missing, therefore, were passed.

The binary built-up/non-built-up mask derived from the status layer production as an intermediate output is used to restrict the analysis to urban area. This will reduce the computation time.

Super-resolved Sentinel-2 L3A products undergo a z-score normalization to mitigate the variations of illumination, atmospheric conditions, calibration difference etc. The normalization is employed per image instead of global time series to account for unique scene characteristics.

6.1.5.2.2 PCA-based detection

Since a new change map is produced every six months, six Sentinel-2 L3A acquisitions are included in total. To detect changes, the first L3A image serves as a reference and is compared individually with each of the five subsequent images, resulting in five reference-target pairs.

For each pair, a PCA model is trained on the spectral differences between the reference and target images. The first principal component, capturing the highest variance, is used as a change indicator.

To distinguish change signals from background noise, several adaptive thresholding techniques were evaluated, including Otsu's method and percentile-based approaches. Based on performance testing, the 99th percentile threshold was selected as the most effective criterion for separating significant changes from background variability.

6.1.5.2.3 Post-processing

While normalization reduces seasonal variability, several sources of false detections remain, especially in crop fields, bare soils, and solar panels areas. To mitigate these issues, additional masking steps were implemented as follows:

- Crop fields: Agricultural areas are often detected as changes due to seasonal cycles. To address this, NDVI statistics are thus computed across the time series. Pixels showing consistently high NDVI in multiple dates combined with high temporal variability (high standard deviation value) are excluded from the change map.
- Bare soil: Soil reflectance can fluctuate with moisture content, introducing another source of confusion. To solve this, the Normalized Soil Moisture Index (NSMI) is used to assess pixel stability. Changes are retained for a pixel with low NSMI variance and low NSMI, which indicate stable sealed surfaces, while unstable bare soil areas are discarded.
- Solar panels: This source of error remains unsolved in the current version of the method. However, spectral indices designed to detect photovoltaic surfaces could be included in the future to mask out and remove them from the change map.

Following the approach used in the status production layer, spatial filtering is applied to the final change map. Small objects smaller than 4 pixels are removed, and a 3x3 morphological closing operation is employed to improve the spatial coherence and reduce noise.

The HDBSCAN algorithm serves as an optional post-processing step to improve change detection results. It retains clusters larger than 5 pixels that contain high-confidence changes detected more than 3 times. Additionally, the algorithm also preserves adjacent lower-confidence pixels that are spatially adjacent to these high-confidences areas. This approach can reduce scattered low-confidence change detections. The algorithm focuses on preserving the complete shape of urban expansion changes areas where some pixels might have weaker individual detections but obviously belong to a larger changing area. However, due to its intensive computation, this clustering step remains optional and can be enabled or disabled based on user needs.

6.1.6 Results and Verification

This chapter depicts the results of the classification as well as their validation. The regression analysis performed is presented in section 6.1.6.1 and the thematic accuracies are summarized in section 6.1.6.2.

6.1.6.1 IMD scatterplots and regression analysis

A scatterplot is a way of displaying data against Cartesian coordinates to show and compare values for two variables within a dataset. The data is displayed as a series of points, where the x and y locations relate two variables assigned to a particular recording instance, in this case a PSU. The available measurements for each PSU are the original reference data (called REFERENCE in each figure shown below) and the mapped value from the product (called MAP on the figures shown below). In this way the relation of the reference and product information for a point can be compared to a 1:1 line which runs diagonally across the scatter plot. The closeness of a point to the point to the 1:1 line is an indication of the similarity between the

reference and mapped results. The points that lie exactly on the x and y axes are related to omission and commission rather than the calibration of the IMD values themselves.

Figure 86 shows the scatterplots for each of the 3 test sites selected for the Phase 2 and for the first date of the time period. There are some numbers of point on the x and y axes showing omission, where sealing is mapped that is not present in reality, and commission, where sealed areas are missed. These cases seemed to be limited in most cases to examples which represent sealing degrees of less than 20%. It can also be seen that the distribution is not centred on the 1:1 line but falls below the line indicating that the **mapped products are underestimating the actual imperviousness** as reported by the reference data. This conclusion is repeated consistently across the different test sites but is particularly true for the test site in Central Europe.

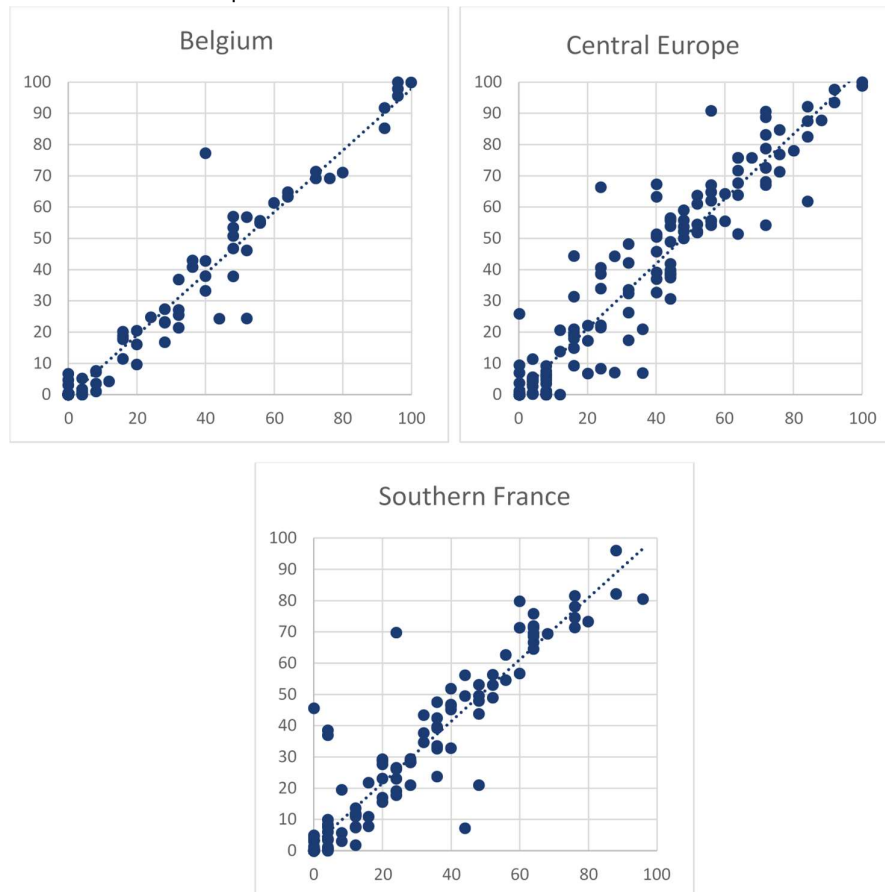


Figure 86. Scatterplots for the 3 test sites for the Imperviousness degree product.

To quantitatively summarise the results displayed in the scatterplots above a linear regression analysis is performed to estimate the relationships between the reference and mapped product information. The analysis produces a coefficient of determination (R^2) which gives information about the goodness of fit of the estimated regression model. Coefficients of determination closer to 1 represent a better fit. In this case as the reference and map information are meant to represent

the same information then it is useful to also consider the slope and intercept of the estimated regression model. The slope should therefore approach 1 and the intercept should be close to 0 for the required relationships. Deviations from the expected values give an indication of the correspondence of the reference and mapped imperviousness data.

Table 52 shows a summary of regression line parameters for the scatterplots. Except one date in Central Europe, all regression coefficients are greater than 0.85 which indicate a very good relationship between the reference and map density values. The lowest value is related to the Southern France test site with a R^2 close to 0.81 and the better results are close to 0.95.

Regression slopes are most of the time smaller than 1,0 which is mainly a result of commission errors (IMD detected in non-sealed areas) and an under-estimation of the IMD value during the calibration. The intercept parameter tends to show values close to 0. Values under 0 are a result of the omission errors (negative intercept) and positive intercepts are a result of commission errors.

Phase 2 has yielded superior results in comparison to Phase 1, largely due to the enhancements made during the calibration step

Table 52. A summary of regression line parameters for the scatterplots derived from the plausibility analysis as reference data.

		3 test sites	Southern France	Belgium	Central Europe
15/01/2020	R^2	0,9153	0,8873	0,9503	0,9192
	Slp	1,0082	0,989	0,9845	1,0356
	Int	0,755	1,8287	-0,6409	0,4876
15/07/2020	R^2	0,8957	0,9065	0,9395	0,8688
	Slp	0,9897	0,9854	1,0083	0,983
	Int	0,764	0,7998	-0,8144	1,4977
15/01/2021	R^2	0,8932	0,877	0,9339	0,8879
	Slp	0,9687	0,9771	0,9723	0,9678
	Int	1,3952	2,4823	-1,0432	1,5349
15/07/2021	R^2	0,8732	0,9263	0,9263	0,8186
	Slp	0,9789	1,0261	0,9792	0,9555
	Int	1,1495	1,4349	-0,2757	1,4069
15/01/2022	R^2	0,8895	0,8922	0,8814	0,8932
	Slp	0,9675	0,98	0,9738	0,9614
	Int	2,102	2,8899	1,0655	1,8438
15/07/2022	R^2	0,8911	0,9011	0,9537	0,8537
	Slp	0,965	0,9665	1,0199	0,9383
	Int	0,4364	0,95	-0,3004	0,3576

6.1.6.2 Thematic accuracy

Thematic accuracies are presented in the form of an error matrix. Unequal sampling intensity resulting from the stratified systematic sampling approach is accounted for by applying a weight factor to each sample unit.

6.1.6.2.1 Status layers

For the status layer, a threshold has been applied to the density values for reference and map data to produce binary attributes of built-up for both the reference and map data layers. For IMD, the threshold was set to 1) 1% and 2) 30 % with density values respectively lower than 1% and 30 % classified as 0 (non-built-up) and density values greater than or equal to 30 % classified as 1 (built-up). The 1% threshold allows the assessment of the built-up mask whereas the 30% threshold is commonly used in HRL.

The below tables give a summary of the internal accuracy assessment extracted from the two confusion matrices of the Imperviousness Degree product for the test sites, namely the Southern France, Belgium and the Central Europe test sites.

Table 53. Summary of the internal accuracy assessment extracted from the confusion matrices of the Imperviousness Degree product with the threshold set to 1% (OA= Overall Accuracy; UA = User Accuracy; PA = Producer Accuracy). Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 10% and red cells fall more than 10% short.

	1%	3 test sites	Southern France	Belgium	Central Europe
20200115	OA	95,30%	95,09%	98,08%	94,14%
	CI95%	2,13%	4,14%	2,65%	3,39%
	UA	94,87%	92,46%	99,30%	94,38%
	CI95%	3,48%	7,47%	2,98%	5,18%
	PA	95,71%	98,08%	97,41%	92,65%
	CI95%	2,35%	2,62%	3,37%	5,05%
20200715	OA	95,61%	95,09%	96,48%	95,61%
	CI95%	2,08%	4,18%	3,44%	2,88%
	UA	93,65%	92,46%	99,32%	91,37%
	CI95%	3,82%	7,55%	2,85%	6,23%
	PA	97,00%	98,08%	94,29%	97,76%
	CI95%	2,16%	2,61%	5,28%	3,42%
20210115	OA	95,99%	94,90%	93,45%	98,06%
	CI95%	2,02%	4,04%	6,05%	2,01%
	UA	94,13%	92,02%	91,19%	97,68%
	CI95%	3,62%	7,37%	9,98%	3,32%
	PA	97,84%	97,95%	97,12%	98,14%
	CI95%	1,76%	2,83%	3,78%	2,71%
20210715	OA	95,93%	95,15%	98,06%	95,53%
	CI95%	1,94%	3,98%	2,69%	2,94%
	UA	95,93%	92,31%	99,31%	97,29%

	CI95%	3,10%	7,38%	2,87%	3,70%
	PA	95,55%	97,96%	97,07%	92,69%
	CI95%	2,50%	2,85%	4,01%	5,09%
20220115	OA	95,26%	93,16%	97,97%	95,62%
	CI95%	2,25%	4,98%	3,47%	3,01%
	UA	91,56%	87,56%	96,49%	92,36%
	CI95%	4,24%	9,06%	6,01%	5,89%
	PA	99,21%	100,00%	100,00%	98,04%
	CI95%	1,09%	0,00%	0,00%	2,87%
20220715	OA	95,33%	95,23%	98,04%	94,13%
	CI95%	2,11%	3,77%	3,48%	3,42%
	UA	94,64%	94,10%	96,44%	94,03%
	CI95%	3,58%	6,67%	6,32%	5,51%
	PA	95,61%	95,99%	100,00%	92,82%
	CI95%	2,47%	3,92%	0,00%	4,95%

Table 54. Summary of the internal accuracy assessment extracted from the confusion matrices of the Imperviousness Degree product with the threshold set to 30% (OA= Overall Accuracy; UA = User Accuracy; PA = Producer Accuracy). Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 10% and red cells fall more than 10% short.

	30%	3 test sites	Southern France	Belgium	Central Europe
20200115	OA	96,81%	97,61%	93,53%	97,75%
	CI95%	1,88%	2,65%	5,79%	2,35%
	UA	94,79%	94,96%	96,67%	93,42%
	CI95%	3,51%	6,19%	6,42%	5,57%
	PA	77,62%	83,08%	65,70%	82,93%
	CI95%	12,00%	19,10%	21,36%	19,45%
20200715	OA	97,50%	97,95%	96,48%	97,64%
	CI95%	1,68%	2,61%	4,36%	2,40%
	UA	95,23%	97,87%	96,88%	92,31%
	CI95%	3,33%	4,13%	6,03%	5,91%
	PA	80,06%	79,64%	77,88%	81,94%
	CI95%	12,70%	22,25%	23,47%	20,34%
20210115	OA	97,00%	97,56%	94,99%	97,51%
	CI95%	1,82%	2,76%	5,14%	2,43%
	UA	93,65%	94,23%	96,77%	91,14%
	CI95%	3,75%	6,34%	6,22%	6,27%
	PA	78,06%	79,09%	72,04%	82,16%
	CI95%	12,70%	22,70%	22,23%	20,15%
20210715	OA	96,79%	98,87%	96,48%	95,31%
	CI95%	1,93%	1,91%	4,35%	3,57%
	UA	95,86%	98,22%	96,88%	93,24%

	CI95%	3,13%	3,66%	6,03%	5,72%
	PA	75,71%	89,86%	78,15%	65,39%
	CI95%	12,44%	17,67%	23,27%	19,43%
20220115	OA	97,25%	97,41%	95,95%	97,74%
	CI95%	1,73%	2,74%	4,68%	2,39%
	UA	93,62%	93,00%	94,44%	93,59%
	CI95%	3,73%	7,00%	7,48%	5,44%
	PA	80,58%	81,19%	77,22%	82,49%
	CI95%	12,39%	20,86%	23,95%	19,84%
20220715	OA	97,19%	97,94%	98,04%	96,21%
	CI95%	1,83%	2,63%	3,17%	3,14%
	UA	95,77%	97,92%	97,33%	92,96%
	CI95%	3,07%	4,04%	5,50%	5,95%
	PA	78,00%	79,30%	89,09%	69,88%
	CI95%	12,47%	22,52%	18,71%	20,31%

The results show that the prototypes are mostly in line with the technical specifications. The minimum thematic accuracy requirement of 90 % (including the CI95%), based on the 1% threshold for IMD values, is met for both, producer and user accuracies for all test sites and all the dates. These results indicate that the built-up mask is of very good quality with few omission and commission errors.

The minimum thematic accuracy requirement of 90% (including the CI95%), based on the 30% threshold for IMD values, is met for both producer and user accuracies across all test sites and dates, with the exception of four cases concerning producer accuracy. However, when the confidence interval is not considered, the results tend to show lower producer accuracies compared to user accuracies. This suggests the presence of omission errors, particularly noticeable at the Belgium and Central Europe test sites. The results show an increase of the omission errors going from a threshold of 1% to 30% suggesting an underestimation of the imperviousness degree in the lower degrees. These results confirm the outcomes of the section 6.1.6.1. User accuracies consistently exceed 90%, indicating a very low rate of commission errors. Notably, these errors have been significantly reduced from Phase 1 to Phase 2, thanks to improvements made in the processing chain. However, this reduction in commission errors may partly explain the presence of omission errors, which are either comparable to or lower than those observed in Phase 1, depending on the test site.

The main source of omission errors are the roads. However, it should be noted that roads are not expected to be completely mapped in the HRL products and discontinuities are accepted. The main source of commission are the bare soils (quarries, crop fields without any crop development), the sands (riverbanks) the rocks and the snow. Based on a qualitative assessment, it appears that the results slightly underestimate the commission errors, especially in the mountainous areas in Central Europe and the bare soils in quarries, construction sites or mining sites.

On the other hand, the commissions with the crop fields visible in Phase 1 have been massively reduced in Phase 2. Figure 87 shows examples of commission errors.

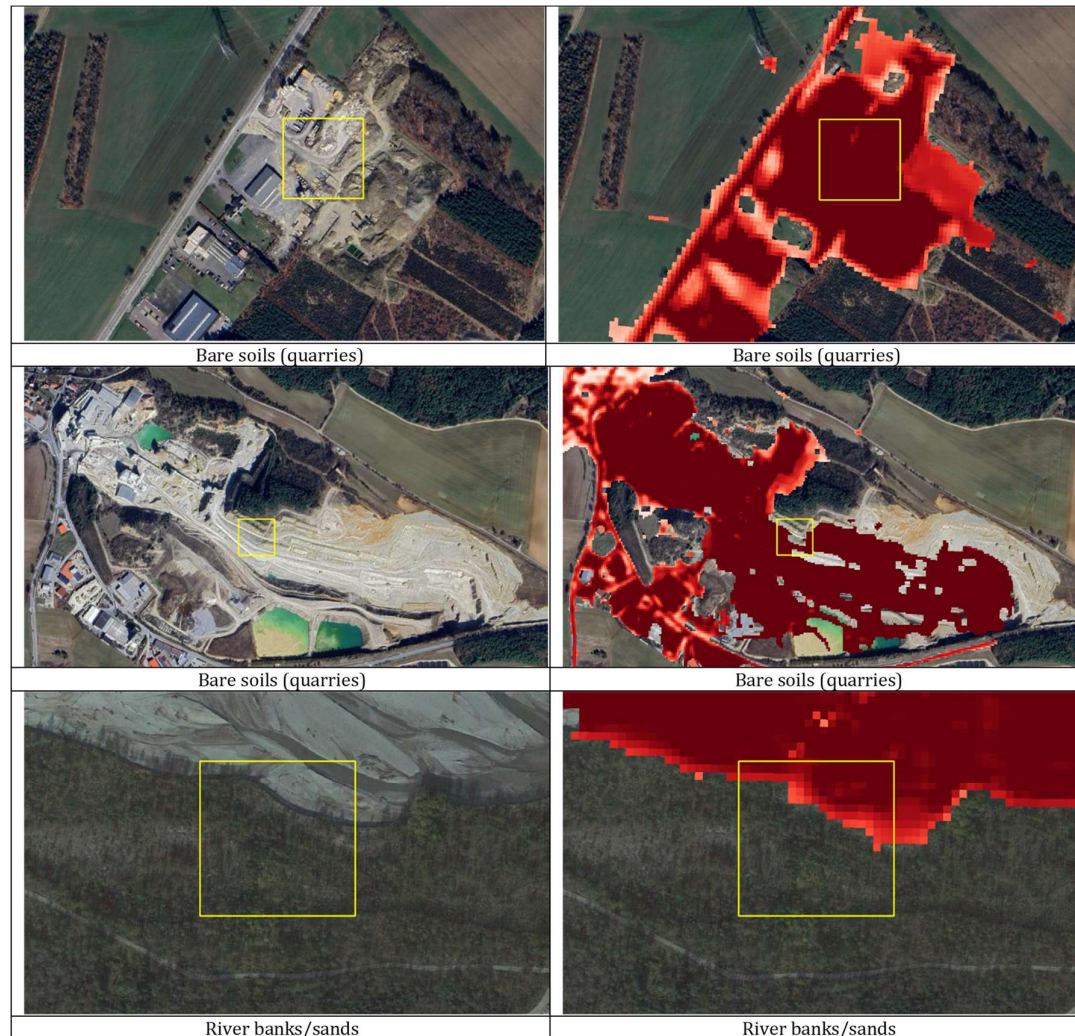


Figure 87. Example of omission errors.

6.1.6.2.2 Change layers

The below confusion matrices give a summary of the internal accuracy assessment of the Imperviousness Change product for the tile 31UFS located in the Belgium test site for the period 01/01/2021 – 30/06/2021.

Thematic assessment reflects that the change detection of urban features based on a 6-month period remains a challenge. But it should be noticed that the rarity of the change class also amplifies in a negative way the results as it impacts greatly the weighting factor applied to sample to compensate over-representation of “change” samples in the stratification. The unweighted confusion matrix highlights

this tendency, showing that producer accuracy has improved significantly without considering the weighting factors.

Table 55. C8 Thematic accuracy for the change layer (Tile 31UFS; Belgium test site) based on the weighted confusion matrix.

Weighted		Reference				
		No change	Change	Total		
Map	No change	1 207,71	3,23	1 210,94	User Acc.	Com.
	Change	0,04	0,02	0,06	40,13%	59,87%
	Total	1 207,75	3,25	1 211,00		
		Prod. Acc.	0,73%			Overall Acc.
		Omi.	99,27%			99,73%

Table 56. C8 Thematic accuracy for the change layer (Tile 31UFS; Belgium test site) based on the unweighted confusion matrix.

Unweighted		Reference				
		No change	Change	Total		
Map	No change	748,00	2,00	750,00	User Acc.	Com.
	Change	276,00	185,00	461,00	40,13%	59,87%
	Total	1 024,00	187,00	1 211,00		
		Prod. Acc.	98,93%			Overall Acc.
		Omi.	1,07%			77,04%

6.1.7 Conclusions

Phase 2 of the project has successfully demonstrated that the developed processing chain is capable of handling a substantial volume of input data efficiently and within a reasonable timeframe, while also delivering high-quality classification results. The integration of Sentinel-2 super-resolution technology, as outlined in the outcomes of WP3, has shown considerable promise. This advancement is particularly noteworthy for its potential to enhance the precision and overall quality of the processed data. The findings indicate that the incorporation of these innovative methodologies not only meets the initial project objectives but also sets a solid foundation for further developments. Figure 88 presents the improvement brought by the 5m super-resolution. Finer and better delineation of the urban areas are to be noted which has led to a reduction in the commission errors. Subsequently, it allows to derive meaningful and accurate statistics about the sealed surfaces in Europe, which is key for policy making.



Figure 88. Improvement brought by the 5m super resolution. Left is the HRL IMD 2018 with a spatial resolution of 10 meters and right the C8 prototype with a spatial resolution of 5 meters for the reference year 2020.

The positive results from WP3 underscore the robustness of the developed system and its applicability in various large-scale data processing scenarios. Overall, the results are very promising and satisfying for the C8 prototype and particularly the status layers. The results indicate that the prototypes generally align with the technical specifications applied to the current Imperviousness products within the CLMS portfolio. The minimum thematic accuracy requirement of 90 % (including the CI95%), based on a 1% and 30 % threshold for IMD values, is met for both, producer and user accuracies for most of the test sites and dates except for four producer accuracy based on the 30 %.

Improvements have been reached in Phase 2 compared to Phase 1 to enhance the quality of the prototypes. Specifically, the development in Phase 2 allows reducing the underestimation of the imperviousness degree as well as the commission errors observed in the current phase. These improvements have been achieved by leveraging various aspects of the processing chain:

- Training sample: Reliable training samples derived from historical HRL 2021 Imperviousness AND, Forest, Grassland, Water and Small Woody Features (and HRL VLCC/NVLCC 2021 if available) have been used, which helped reducing omission and commission errors.
- Absolute calibration: a linear regression method has been applied from the biophysical variables (e.g. NDVI_{max}) derived from the seasonal image composites. Improving the quality of the calibration helped reducing underestimation of the imperviousness degree.

Furthermore, the Phase 2 was an opportunity to expand the technical specifications in order to enhance and fulfil the user requirements to an even greater extent. If the objective was to improve the timeliness of the prototype/service and the monthly time interval, feedback gathered during a user webinar held between the two phases revealed that users preferred a six-month update cycle. But the Phase 2 was also an opportunity to work on the change layers, which include the sealing and de-sealing information. Nevertheless, as presented in section 6.1.6.2, change detection of urban features based on a 6-month period remains a challenge.

Phase 2 also provided an opportunity to explore how the product could support the analysis of land loss or gain due to soil sealing. Although this aspect was not part of the technical specifications of the imperviousness products, potential synergies with the C9 prototype were examined. However, no conclusive results were obtained, primarily due to differences in the temporal coverage of the datasets.

Overall, our work indicates that the status layer prototypes are largely in compliance with existing standards.

6.1.8 Outlook beyond the EvoLand project

The C8 prototype has demonstrated strong potential for mapping continuous imperviousness features with improved spatial resolution. The results indicate that the status layer prototypes largely meet the technical specifications (minimum user and producer accuracies) and the user requirements. However, commission errors appear to be underestimated with persistent confusion in differentiating rocks or bare soils. Moreover, the regression results have also shown an underestimation of the imperviousness degree. Nonetheless, this has been significantly reduced compared to Phase 1. These issues need to be addressed in case of future developments.

Thus, to advance toward higher TRLs and operational phase deployment, future efforts should prioritize the integration of advanced classification algorithms widely regarded as best-in-class by global scientific community that could bring significant improvements compared to existing techniques implemented in EvoLand. The WP2 has successfully demonstrated the potential of AI based time series change detection methodologies for forest. These techniques could be adapted and tested for urban domains in future developments.

Although the contribution of the Sentinel-2 super-resolution technology is promising in terms of enhancing data quality, the current computation time is prohibitively high for envisaging its full operationalization on a pan-European scale. WP4 will be instrumental in assessing the maturity and feasibility of this technology, as proposed in WP3.

Finally, the change detection component is not yet fully mature. Reliable change detection will require further research to address challenges such as pixel misregistration and acquisition variability. Developing more robust and resilient methodologies will be essential to ensure consistent and accurate results.

6.2 C9 – Automated Land Use Mapping of Urban Dynamics

6.2.1 Introduction & Objectives

One of the main points to emerge from the WP1 user requirements is that there was a strong desire for automated updating of LULC (Land Use Land Cover) products, such as the Urban Atlas, so that they could be updated annually instead of the current 3 years.

The prototype focuses on the detection of urban dynamics using a nomenclature derived from that of the Urban Atlas. This nomenclature contains a certain number of urban classes and rural classes.

This prototype must also work on areas where there is no input land-use data, and to be in line with this, the methodological approach was revised with the aim of obtaining a tool that can work with or without input land-use data for the detection of absence/presence of urban dynamics. The first objective is therefore to be able to detect these urban dynamics without any input product but using image data and potential in-situ data.

User requests are therefore focused on detecting urban dynamics in European territories, but also with a test outside European territory, which was carried out with Sentinel-2 tiles over Abidjan in Côte d'Ivoire. These detections and characterizations of these dynamics should lead, for example, to automated annual updates of change layers in the Urban Atlas, itself based on a Corine Land Cover-type nomenclature, i.e. classes nested at 3 levels.

Regarding the minimum mapping units, the rules of the Urban Atlas must be kept, 0.25 ha for urban classes and one hectare for rural classes, bearing in mind that for the type of product requested, rural classes are not considered and so we mainly used a minimum area of 0.25 ha, which is reasonable when using Sentinel-2.

The specifications of the C9 prototype are described in Table 57.

Table 57. Prototype specifications.

Prototype C9	
Sub-prototype	C9: Automated land use mapping of urban dynamics
Prototype sites	2 – Southern France 3 – Belgium 4 – Central Europe 8 – Côte d'Ivoire
Description of the sub-prototype	Mapping urban changes extents and type of change (renaturation, urban sprawl or urban renewal)
Expected temporal update	Yearly
Spatial resolution	5m
MMU	0.25ha
Envisaged accuracy	85%

6.2.2 State of Play

6.2.2.1 Existing CLMS products

Our reference product for this prototype is the Urban Atlas Land Cover/Land Use Change layer. This product provides land cover and land use change data between two reference years.

The covered areas by Urban Atlas are named Functional Urban Areas (FUA). The Minimum Mapping Unit (MMU) varies from 0.1 ha to 0.25 ha depending on the classes. The Minimum Mapping Width (MMW) is 10 m, and the product is available as vector data.

The existing CLMS products that have been produced to date are change layers, with 6 years of temporal resolution, 10m of spatial resolution and a thematic accuracy of over 80%. For the 2006-2012 product, the spatial coverage is EU27, and the sensors used are SPOT-5/6 and Formosat-2. For the 2012-2018 product, the spatial coverage is EEA38, and the sensors used are SPOT-5/6, Formosat-2, Pleiades, SuperView, Planet and KOMPSAT.

Here is an example of a change product using the Copernicus Land Monitoring Service viewer, available through the website [Data viewer — Copernicus Land Monitoring Service](#).

6.2.2.2 WP1 user requirements

During the development of the first Work Package, the requirements for what the Urban Dynamics product should become were defined as follows.

The product should retain consistency with existing Urban Atlas products. This is a general requirement but is emphasized in this case. Additionally, the product should be updated at least annually. It should also have a latency of no more than 1 year. i.e. products for a given reference year should be available before the end of the following year. Further, the product would also show what type of land is being lost to urbanization. The product should have spatial coverage including all cities, towns and suburbs. Regarding minimal mapping unit, product should have resolution of 0.25ha in urban areas and 1ha in rural areas. Regarding the quality, the product should have an accuracy of at least 85%. Note, however, that in this case precision is seen as a more important requirement than accuracy.

The product should have the highest possible precision. No numerical requirement is given here. Precision means the minimization of random errors, such that similar inputs always give similar output measurements, whereas accuracy refers to how close the measurement is to the underlying real value. Ideally, the product should align with the EUNIS classification of urban ecosystems. For EUNIS refer to the EEA

website. And lastly, the product should be available in at least the EU27 and ideally the EEA38. There is additional interest in worldwide coverage, but this is not considered a priority.

6.2.3 Selection of Prototypes sites

Four major geographic study areas were selected for this prototype, including 26 Sentinel-2 tiles:

- Southern France (31TCJ, 31TCK, 31TDJ, 31TDK, 31TEJ, 31TEK, 31TFJ, 31TFK)
- Belgium (31UER, 31UES, 31UFR, 31UFS)
- Central Europe (32TNT, 32TPT, 32UMU, 32UMV, 32UNA, 32UNU, 32UNV, 32UPA, 32UPU, 32UPV)
- Ivory Coast (30NUL, 30NUM, 30NVL, 30NVM)

This set allows us to evaluate the behaviour of the processing chain in different situations in terms of variety of climate and land cover.

6.2.4 Input and Ancillary Data & Pre-processing

6.2.4.1 EO data

The change detection is based on Sentinel-2 images at processing level 3A. Changes are derived from the comparison of two images from the same day of each reference year, which is important to match the periods of vegetation activity. L3A products provide a weighted average of the cloud-free observations, obtained over a 45-day synthesis period, which makes them an ideal input for the change detection method in place, as clouds could lead to the mapping of false changes.

The spatial resolution of the used images is 5 meters, obtained by applying the Super-Resolution Method from WP2 on the Sentinel-2 images which initially have a 10-meter resolution.

6.2.4.2 In-situ data

This prototype uses cadastral data in the aim of improving the change polygon geometries as explained in section 6.2.5.5. The limitation is that cadastral databases are not always available everywhere, depending on the country to work on. Sometimes it is open sourced and free to download and used, sometimes it needs payment, and sometimes it is not available at all. An automatization of the downloads is thus not possible by now.

Some areas are moving continuously because of their nature, such as car parks, reflection of sun on solar panels, or quarries, these classes were extracted from OpenStreetMap to be used as exclusion masks.

This prototype also uses CLC data, available for the 2018 year. This product provides pan-European wall-to-wall spatially detailed land cover inventory for the 2018 reference year, for each pixel showing the dominant land cover among the 11 basic

land cover classes. The dataset is available as 10 m raster. This data is used to train the random forest model for 2018, one is trained by site, which will then be applied to the other years.

As the CLC data are not available outside Europe, in order to characterize the classes for Ivory Coast, World Cover 2020 and 2021 were used, by taking the majority class inside the change polygons for each of these years.

Table 58. EO and ancillary datasets description.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-2 (super-resolved)	Sentinel-2 L3A spectral bands, synthesis over 45 days of L2A images around the 15 th of each month	5 m	Monthly
WorldCover	ESA WorldCover 2020 and 2021 Global land cover product	10 m	Yearly
Cadastral data	Vector layer delineating cadastral parcels. Availability depending on country	No minimum mapping unit	Depends on the country
CLC	Land cover for the reference year	10 m	3-years timestep

6.2.5 Methodology and Workflow

6.2.5.1 General workflow

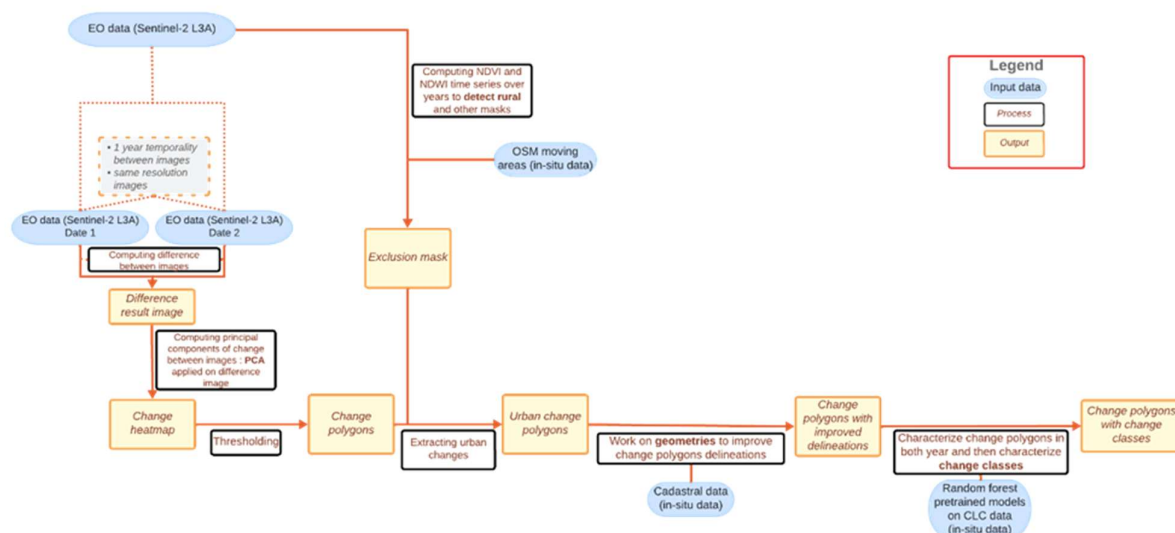


Figure 89. General workflow of C9 Prototype.

6.2.5.2 Change detection process

As input, two Sentinel-2 images at level 3A one year apart are selected. July data has been chosen, being a time of year when vegetation productivity is active, to limit confusion between bare or almost bare ground and sealed or built-up areas.

These images are synthesized monthly (45 days) to obtain cloud-free data for each year. Using two images of the same area on two different dates, as shown in the two figures below.

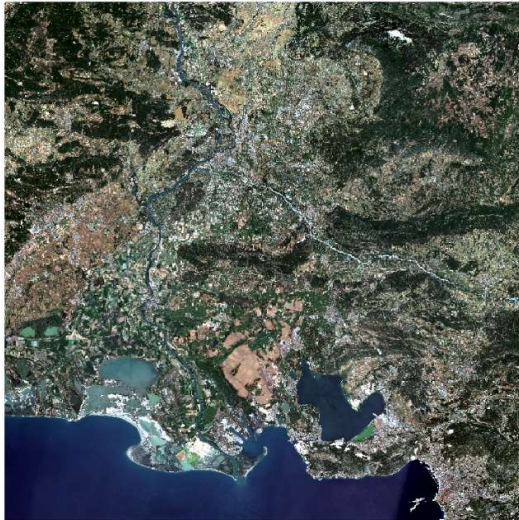


Figure 90. 31TFJ Tile (Sentinel-2 at level 3A) - July 2020.



Figure 91. 31TFJ Tile (Sentinel-2 at level 3A) - July 2021.

The most obvious approach when trying to detect changes between several images of the same region is to calculate the difference between these different acquisitions and then select the changes that are deemed significant.

Principal Component Analysis (PCA) (Deng, J. S. et al., 2008) can be used to generate change maps by retaining only the principal component of the difference image, which is a method for eliminating parasitic noise induced by the sensor.

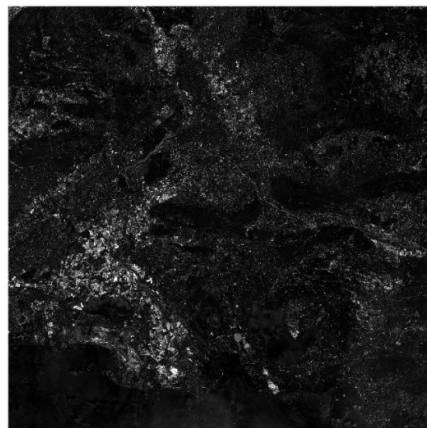


Figure 92. Resulting image of the PCA on the difference between the two images above.

After extracting the principal component, we define a threshold above which a pixel is considered to be a change. To do this, we tested several threshold definition methods (Section 6.2.5.4).

6.2.5.3 Exclusion mask

As the difference is used between two images, depending on variations in crops, snow, vegetation or water, these elements can stand out in the result of the detection of change by PCA.

To exclude these cases, and thus detect only urban-related changes, an exclusion mask has been set up, comprising:

- A water mask, which extracts pixels where the NDWI on both dates is > 0 .
- A forest mask, which extracts pixels where the NDVI on both dates is > 0.85 .
- A cloud and snow mask, provided by downloaded Sentinel-2 L3A data.
- An agricultural mask, provided by stacking the NDVI across the year and thresholding the mean and standard deviation.

As part of our analysis, several types of spatial features were also extracted from OpenStreetMap (OSM) that are considered as moving areas but not as real changes and were used as an exclusion mask in our processing. These include parking areas identified by the tag `amenity=parking`, excluding those with values indicating enclosed or underground structures (parking in underground, `garage_boxes`, `garages`). Quarries and landfills were also included, using the tags `landuse=quarry` and `landuse=landfill`, as well as solar panel installations (`generator:source=solar`). Additional features considered as moving areas are beaches (`natural=beach`), salt ponds (`landuse=salt_pond`), natural sand areas (`natural=sand`), and wetlands (`natural=wetland`). These areas are treated as dynamic due to their potential for environmental change or human activity.

6.2.5.4 Change heatmap thresholding

Finding the threshold is an important step, as it determines the binary boundary between what is a change and what is not. As PCA extracts the principal components of the image, the threshold is determined based on the distribution of values in the image. The PCA histogram for each zone follows a right-skewed distribution.

The first thought was that pixels with no major changes would have distances close to the average, while pixels with significant changes would deviate further from the average. The first threshold tested was $\text{mean} + 2 \times \text{std}$. However, this threshold would be better suited to a normal distribution, so another threshold search method was studied (Vázquez-Jiménez, R. et al, 2018).

This second method consists of defining the threshold automatically by selecting the pixel value corresponding to the point in the distribution of the histogram, where the maximum perpendicular line intersects the secant line between the highest and lowest points of the histogram.

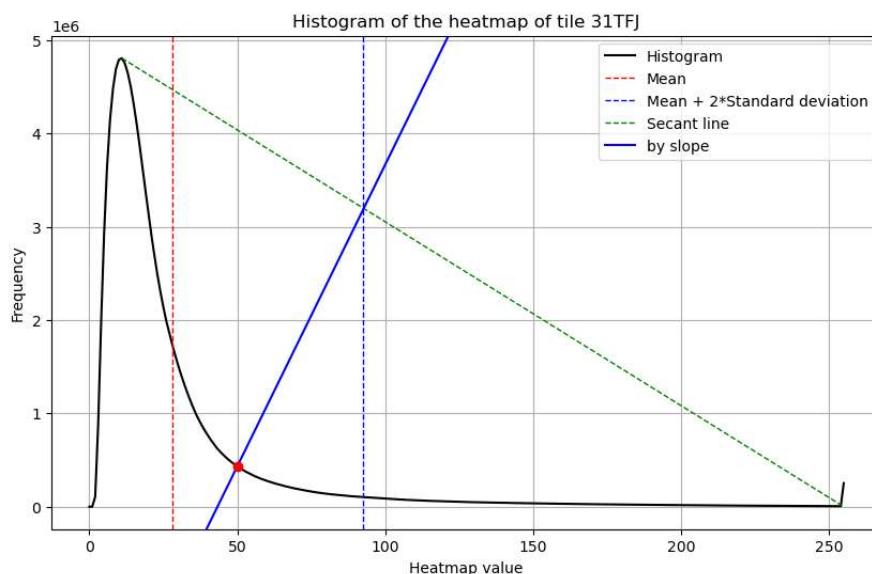


Figure 93. Histogram of the heatmap (Figure 92. Resulting image of the PCA on the difference between the two images above) of tile 31TFJ (method 1 thresholding (dotted blue and dotted red) - method 2 thresholding (solid blue & dotted green)).

To determine which method to choose for the 'truth' of this intermediate test, small areas were selected from the tiles studied, and within these areas, all changes were marked manually.

The ground truth and the changes detected were compared to extract omissions within these zones after the application of these 2 thresholds. It should be noted that this test is an intermediate step in defining the heatmap thresholding method, and does not constitute the final validation results, which are presented later in the 'Results and Validation' section 6.2.6.

Table 59. Intermediate omission rates.

Area	Thresholding method 1 (Mean + 2*std)	Thresholding method 2 (Perpendicular)
South of France - area 1	93,8%	15,0%
South of France - area 2	12,0%	0,0%
Central Europe - area 1	70,6%	5,9%
Central Europe - area 2	61,4%	10,5%
Belgium - area 1	84,6%	40,4%
Belgium - area 2	74,2%	32,3%

There are fewer omissions in most areas with the second method, so method 2 is chosen to be the follower method in the chain. However, as the rates are still high in some areas, more exploration of this issue needs to be done.

6.2.5.5 Improving geometries

Once these polygons have been detected, the next step is to improve the geometry of the detected changes by cross-referencing them with in-situ data, which is mainly cadastral parcel data where available. Cut and overlap system are used to determine whether to retain parts of the changes detected, and other parts can be slightly enlarged to best fit the geometry of the cadastral parcels, where available. This work makes it possible to obtain geometric renderings that are much more accurate and realistic than those based solely on image analysis. This greatly improves the geometric accuracy of the product.



Figure 94. Example of improved geometry detected by the chain - red: initial geometry, green: improved geometry – OSM background map.

6.2.5.6 Characterization of the detected changes

6.2.5.6.1 Single-date image classification

To characterize the changes detected, Random Forest classification models were trained for each site on an L3A reference image in 2018 when available, with a CLC 2018 vector dataset inside Europe sites. These models were then inferred on the images used for change detection (one image in July in year N and one image in July in year N+1). These classifications were then cross tabulated with the change polygons in order to extract the majority class for each year for each polygon. The chosen classes are as follows:

- Construction/landfill
- Sealed
- Natural/Agricultural
- Forest
- Water

It should be noted that these classifications are based on level 2 of the Urban Atlas nomenclature.

For the Côte d'Ivoire site, in absence of Corine Land Cover data on this area, the WorldCover 2020 and 2021 datasets were applied to map the majority class inside change polygons, but this source of data is less precise and less detailed than the Corine Land Cover ones, which could have an impact on the results.

6.2.5.6.2 Multi-date image classification

The characterization of changes is based on the classification of each year's polygons. To focus mainly on urban changes, reduced classes of change were chosen:

Table 60. Change classes.

Year 1 class	Year 2 class	Change type
Sealed or construction	Sealed or construction	Urban renewal
Neither sealed nor construction	Sealed or construction	Urban sprawl
Sealed or construction	Neither sealed nor construction	Renaturation

There are some cases in which the PCA detected a change that was subsequently coded as non-urban in both dates. This situation corresponds to non-urban changes that are therefore not be retained in the output layer.

6.2.6 Results and Verification

This chapter shortly describes the approach adopted for the validation and depicts the results of the classification as well as their validation. Firstly, the validation plan is presented, then the thematic accuracies are summarized.

The Land Use / Cover Area Frame Survey (LUCAS) database was not used for the validation process. (see [AD14]). A visual analysis was preferred to better evaluate the accuracies, using mainly the Sentinel-2 images as reference.

6.2.6.1 Overview

Product validation is an essential stage in a tool development process such as the one we are working on here. This stage provides a quantitative and qualitative assessment of the generated products. The expected overall degree of reliability of the products is 85%, which corresponds to the degree of reliability expected of Urban Atlas products.

Methodology for extracting reference datasets

The sampling design is based on the validation of the accuracy of the detected changes (change / no change). The samples were extracted thanks to 3 strata:

- A change stratum corresponding to the detected change polygons, which will assess the thematic accuracy and the potential commissions.

- A buffer stratum which is a buffer of 100 meters around the detected change polygons assuming that some changes could be partially detected in dynamic areas.
- A non-change stratum everywhere else within the area of interest (S-2 Tiles). It allows to point out some potential omissions.

The redistribution of these points by stratum is established according to the proportion of each stratum in terms of surface area in relation to the total surface area of the study zone. Thanks to this approach, the results can then be displayed according to a weighted confusion matrix and associated indicators, enabling the results to be analysed according to the surface representativeness of each stratum.

Validation of the product released for each test site

The interpretation of the reference data has been done by means of visual interpretation of VHR (Google Earth or Bing Aerial) and HR data (Sentinel-2) set by blind analysis. Indeed, an experienced photo interpreter can easily detect the change areas based on VHR/HR images comparison. Based on the satellite imagery, the photo interpreter had the choice between two categories (change / no_change) and for each change he qualified the classes for both years, using the same nomenclature as the one used during the random forest classification and characterization.

A comparison is then made between this visual analysis, which constitute the reference, and the extraction of the product classes on these same points, this leads to a confusion matrix to get the raw accuracies.

A second interpretation, carried out by a second interpreter as part of the plausibility analysis, is then done only on the sample units in disagreement with the product classification. The Plausibility analysis resulted in revised accuracy metrics. Indeed, the blind approach described above guarantees independence from the map products but may underestimate the accuracies for complicated points when sometimes the point is located slightly out of a detected change perimeter or when the notion of change is visible on VHRs (Google Earth for example) but almost indistinguishable by comparison of Sentinel images. This is resolved by the plausibility approach. The results presented in the section above are based on the plausibility analysis.

A comparison will then be made between this visual analysis, which will constitute the reference, and the extraction of map classes on these same points. A confusion matrix is then extracted, and the degrees of reliability obtained.

6.2.6.2 Accuracy evaluation

The results on the four test sites are presented in the following tables.

Table 61. Change / no change confusion matrix for the Southern France site for 2020-2021.

	FR_change 2020-2021	REFERENCE		Total	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	651	21	672	96,9%	3,1%	0,0%	1,3%	0,97
	change	24	104	128	81,3%	18,8%	0,1%	6,8%	0,82
	Total	675	125	800					
	Prod. Acc.	96,4%	83,2%		Overall Acc.	94,38%			
	Omission	3,6%	16,8%		CI95%	1,55%			
	Variance	0,0%	0,1%						
	CI 95%	1,2%	6,0%						

Table 62. Change / no change confusion matrix for the Southern France site for 2021-2022.

	FR_change 2021-2022	REFERENCE		Total	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	655	17	672	97,5%	2,5%	0,0%	1,2%	0,97
	change	29	99	128	77,3%	22,7%	0,1%	7,3%	0,81
	Total	684	116	800					
	Prod. Acc.	95,8%	85,3%		Overall Acc.	94,25%			
	Omission	4,2%	14,7%		CI95%	1,53%			
	Variance	0,0%	0,1%						
	CI 95%	1,3%	6,0%						

Table 63. Change / no change confusion matrix for the Belgium site for 2020-2021.

	BE_change 2020-2021	REFERENCE		Total	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	527	16	543	97,1%	2,9%	0,0%	1,4%	0,97
	change	16	41	57	71,9%	28,1%	0,4%	11,7%	0,72
	Total	543	57	600					
	Prod. Acc.	97,1%	71,9%		Overall Acc.	94,67%			
	Omission	2,9%	28,1%		CI95%	1,70%			
	Variance	0,0%	0,3%						
	CI 95%	1,2%	10,3%						

Table 64. Change / no change confusion matrix for the Belgium site for 2021-2022.

	BE_change 2021-2022	REFERENCE		Total	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	529	16	545	97,1%	2,9%	0,0%	1,4%	0,98
	change	2	53	55	96,4%	3,6%	0,1%	4,9%	0,85

<i>Total</i>	531	69	600		
Prod. Acc.	99,6%	76,8%		Overall Acc.	97,00%
Omission	0,4%	23,2%		CI95%	1,36%
Variance	0,0%	0,2%			
CI 95%	0,5%	8,7%			

Table 65. Change / no change confusion matrix for the central Europe site for 2020-2021.

	Central_change 2020-2021	REFERENCE		<i>Total</i>	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	743	12	755	98,4%	1,6%	0,0%	0,9%	0,96
	change	50	90	140	64,3%	35,7%	0,2%	7,9%	0,74
	<i>Total</i>	793	102	895					
	Prod. Acc.	93,7%	88,2%		Overall Acc.	93,07%			
	Omission	6,3%	11,8%		CI95%	1,45%			
	Variance	0,0%	0,1%						
	CI 95%	1,3%	6,0%						

Table 66. Change / no change confusion matrix for the central Europe site for 2021-2022.

	Central_change 2021-2022	REFERENCE		<i>Total</i>	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	745	10	755	98,7%	1,3%	0,0%	0,8%	0,97
	change	42	98	140	70,0%	30,0%	0,2%	7,6%	0,79
	<i>Total</i>	787	108	895					
	Prod. Acc.	94,7%	90,7%		Overall Acc.	94,19%			
	Omission	5,3%	9,3%		CI95%	1,37%			
	Variance	0,0%	0,1%						
	CI 95%	1,3%	5,3%						

Table 67. Change / no change confusion matrix for the Côte d'Ivoire site for 2020-2021.

	IC_change 2020-2021	REFERENCE		<i>Total</i>	User Acc.	Commission	Variance	CI 95%	F-Score
		no change	change						
MAP	no change	539	5	544	99,1%	0,9%	0,0%	0,8%	0,96
	change	33	13	46	28,3%	71,7%	0,4%	13,0%	0,41
	<i>Total</i>	572	18	590					
	Prod. Acc.	94,2%	72,2%		Overall Acc.	93,56%			
	Omission	5,8%	27,8%		CI95%	1,26%			
	Variance	0,0%	1,0%						
	CI 95%	1,0%	19,8%						

Table 68. Global change / no change confusion matrices for all sites.

	change 2020- 2021	REFERENCE		Total	User Acc.	Commission	Variance	CI 95%	F- Score
		no change	change						
MAP	no change	4389	97	4486	97,8%	2,2%	0,0%	0,4%	0,97
	change	196	498	694	71,8%	28,2%	0,0%	3,4%	0,77
	Total	4585	595	5180					
	Prod. Acc.	95,7%	83,7%		Overall Acc.	94,34%			
	Omission	4,3%	16,3%		CI95%	0,58%			
	Variance	0,0%	0,0%						
	CI 95%	0,5%	2,8%						

The results show that the prototype does not reach the expected accuracy of 85%. The average accuracy is around 72%. Some omissions are highlighted by these results. The weighted approach overestimates a lot the omissions because the no change strata represent a very high weight compared to the change strata and this imbalance has a disproportionate consequence on the accuracy results. For example, regarding the Belgium site and considering the 2020-2021 production accuracy we are moving from 71,9% without weighting to 9% with weighting. This is why we choose to present only the non-weighted matrices. We chose to present only the non-weighted matrices, prioritizing the analysis of detected changes. While a more representative sampling across strata — fewer points in change and buffer zones, and more outside — would have been ideal, this was not feasible due to time constraints and the limitations of visual analysis.

The Côte d'Ivoire sites were only produced for the 2020-2021 period because the necessary data required as input for the tool was not available on the 2021-2022 period. For this site, the results are significantly weaker than those from other sites, especially regarding the change commissions. This could be explained by the fact that we did not have the CORINE LAND COVER data and had to rely on World Cover instead, which is less precise and less adapted.

We made special efforts to try to reduce false positives by using several additional masks (clouds, parking lots, storage areas, photovoltaic panels), but these masks — often derived from OpenStreetMap database — were not sufficient to eliminate all false positives.

The following images show some remaining artefacts that cannot be masked if the in-situ data is not sufficient.

- Clouds and cloud shadows, even using the associated cloud masks there is still some cases of unmasked clouds.

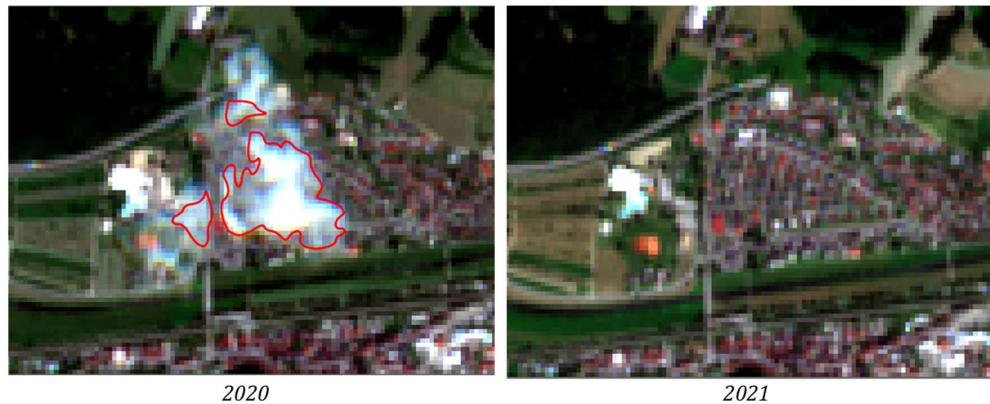


Figure 95. Unmasked clouds and changes detected (in red) – source: Sentinel-2 at level 3A – tile 32UMU – 15th July 2020 (1) and 15th July of 2021 (2).

- Some moving disposals within quarries, landfills or parking areas.

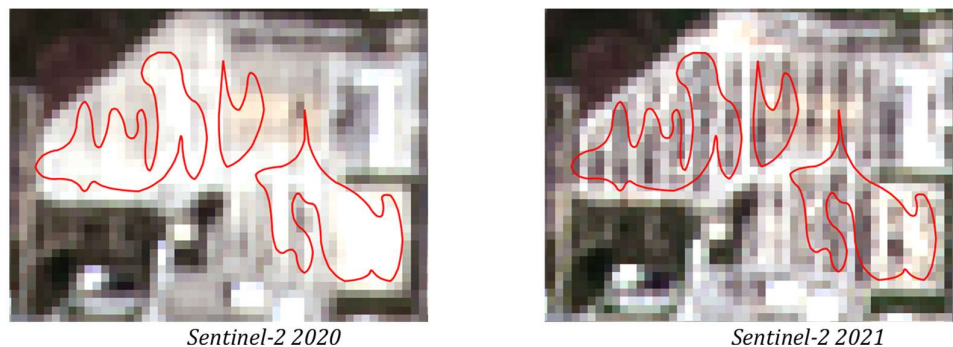


Figure 96. Parking area and changes detected (in red) – source: Sentinel-2 at level 3A – tile 31TFJ – 15th July 2020 (1) and 15th July of 2021 (2).



Figure 97. Parking area on the same location and year that Figure 96 – source: Google Earth.

- Greenhouses or big flat roofs, often corresponding to industrial or commercial buildings, especially those with photovoltaic panels or elements with potentially high reflectance and which could give very different pixel values depending on the sun lightning incidence.

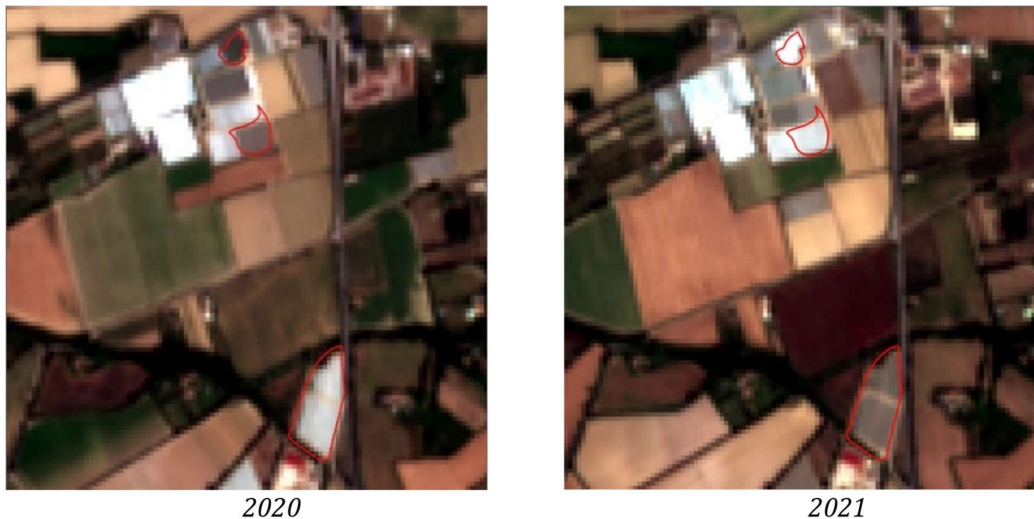


Figure 98. Greenhouse's reflectance effects and changes detected (in red) – source: Sentinel-2 at level 3A – tile 31TFJ – 15th July 2020 (1) and 15th July of 2021 (2).

- Some rural false positives could also still be found, even if the proportion of these cases in the commission errors is lower than the previous cases.

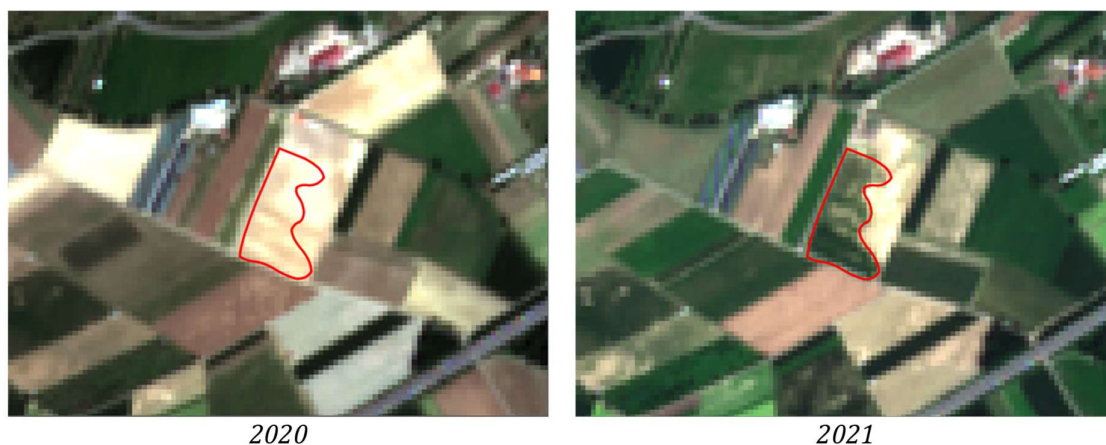


Figure 99. Rural commission and changes detected (in red) – source: Sentinel-2 at level 3A – tile 32UMU – 15th July 2020 (1) and 15th July of 2021 (2).

To conduct a more detailed analysis of the resulting products, validation was also performed on the type of change derived from the comparison of the assigned classes for each year. The results are presented in Table 69.

Table 69. Global change category confusion matrices for all sites.

	change_cat	REFERENCE									
		no change	renaturation	urban renewal	urban sprawl	Total	User Acc.	Commission	Variance	CI 95%	F-Score
MAP	no change	4389	2	37	53	4481	97,80%	2,20%	0,00%	0,40%	0,97
	renaturation	18	10	10	0	38	23,80%	76,20%	0,40%	12,90%	0,34
	urban renewal	147	3	316	6	472	65,70%	34,30%	0,00%	4,20%	0,73
	urban sprawl	14	0	15	105	134	76,60%	23,40%	0,10%	7,10%	0,71
	Total	4568	15	378	164	5125					
Prod. Acc.						Overall Acc.		93,38%			
Omission						CI95%		0,60%			
Variance											
CI 95%											

The table shows various results depending on the class. The "renaturation" class is the weakest, due to confusion during processing between bare agricultural soils and soils under construction or landfills.

The commissions on the "urban renewal" category can be explained by the difficulties already exposed regarding the reflectance of rooftops that can vary greatly from one image to another and leads to false positives.

To further refine the analysis, Table 70 presents the results of the class comparison assigned across all test sites for the year 2020. This table includes only polygons with assigned classes, those detected as belonging to the "change" stratum.

During production and validation, a numeric code was assigned to each class for more simplicity. They are displayed according to this coding system to avoid overloading the table. Below is a detailed list of these classes:

0: Construction site or landfill

1: Sealed / artificialized area

2: Agricultural / Natural area

3: Forest

4: Water and wetlands

Table 70. Global 2020-2021 classes confusion matrices for all sites.

	class	REFERENCE									
	2020	0	1	2	3	4	Total	User Acc.	Commission	Variance	CI 95%
	0	193	70	27	0	11	301	64,10%	35,90%	0,10%	5,40%
MAP	1	34	150	35	1	2	222	67,60%	32,40%	0,10%	6,20%
	2	9	4	121	1	2	137	88,30%	11,70%	0,10%	5,40%
	3	0	0	2	5	1	8	62,50%	37,50%	2,90%	33,50%
	4	1	5	0	0	21	27	77,80%	22,20%	0,60%	15,70%
	Total	237	229	185	7	37	695				
	Prod. Acc.	81,40%	65,50%	65,40%	71,40%	56,80%	Overall Acc.		70,50%		
	Omission	18,60%	34,50%	34,60%	28,60%	43,20%	CI95%		3,32%		
	Variance	0,00%	0,10%	0,10%	2,40%	0,40%					

The resulting accuracy varies depending on the classes analysed. There is notable confusion between class 0 (construction sites, deposits, landfills) and class 1 (impervious surfaces). This can be explained by the fact that these two land cover types may have quite similar spectral indices, making them difficult to distinguish accurately.

6.2.7 Conclusions

The C9 prototype introduced a novel processing chain for the automated mapping of urban dynamics using Sentinel-2 imagery. It enables the detection of both the extent and type of urban changes with improved temporal resolution compared to existing products, thanks to its fully automated workflow. The chain includes the use of spatial super-resolution (T2.5) applied to Sentinel-2 L3A data, enhancing the spatial detail of the outputs.

The results demonstrate the feasibility of identifying urban transformations, such as the emergence of new infrastructure or modifications in built-up areas. In terms of user requirements, the use of regularly available Sentinel-2 data allows for annual updates, and the automated nature of the method supports a latency of no more than one year. The prototype can be deployed across cities, towns, and suburbs where Sentinel-2 L3A data is available. However, due to weaker performance observed in the Côte d'Ivoire test sites, it is recommended to restrict the application of this method to European regions, using the same input data and conditions as those employed during development.

To mitigate commission errors — particularly in areas such as parking lots, greenhouses with flat roofs, and disposal sites — OpenStreetMap (OSM) data were used to extract landfill, quarry, and parking area polygons as exclusion masks. For greenhouses and large flat-roofed buildings, attempts were made to incorporate in-situ data to reduce misclassifications, but no sufficiently accurate and globally consistent datasets could be identified.

Despite its strengths, the prototype has known limitations. Not all land use classes defined in the Urban Atlas change layer can be reliably discriminated. Many urban use classes require detailed in-situ data for accurate detection, which cannot be achieved using Sentinel-2 imagery alone. Furthermore, the spatial super-resolution applied to Sentinel-2 is not a standard processing level available through commonly used data hubs such as the Copernicus Data Space Ecosystem (CDSE). As a result, significant pre-processing is required to generate suitable input datasets before change detection can be performed.

6.2.8 Outlook beyond the EvoLand project

Looking ahead, several avenues can be explored to enhance the C9 prototype and prepare it for potential operational phase. One promising direction is the integration of a T0 layer to support neighbourhood-based analyses. This would allow for improved class discrimination and more accurate geometry refinement, particularly in complex urban environments.

To support operationalisation, automating the acquisition of input imagery — either through direct download or cloud-based ingestion — should be considered.

This would streamline the workflow and reduce manual intervention, making the chain more scalable and efficient.

During development, the prototype was tested using both 10-meter and 5-meter Sentinel-2 imagery. While the 5-meter resolution offers improved spatial detail, it significantly increases computational demands. Therefore, a scalability assessment of the chain is essential before transitioning to operational use, to ensure that processing resources and infrastructure can support large-scale implementation.

The current use of Principal Component Analysis (PCA) for change detection has shown its limitations. To improve classification performance and robustness, further exploration of deep learning approaches is recommended.

It is also important to note that the resulting product is best suited as an intermediate change detection layer — potentially to support or guide manual production — rather than a fully updated change detection layer. Depending on the input data used, it may be challenging to meet all specifications required for a comprehensive update product.

In summary, while the C9 prototype shows strong potential for automated urban change detection, its operationalisation will require targeted improvements in scalability, automation, and quality.

7 General Land Cover prototypes

The results of the T3.5 general land cover prototypes of the first development cycle include T3.5.1 Continuous mapping of land surface categories and T3.5.2 on-demand land cover.

7.1 C10 – Continuous Mapping of Land Surface Categories

7.1.1 Introduction & Objectives

Most operational global land cover datasets, such as the Copernicus Global Land Service (CGLS) Land Cover product at 100m resolution, National Aeronautics and Space Administration (NASA) MCD12Q1 dataset at 500m resolution, or the European Space Agency (ESA) Climate Change Initiative (CCI) land cover dataset at 300m resolution, focus on classifying the Earth surface at a yearly interval. While they are an invaluable source of information for a suite of applications, including monitoring and modelling the Earth's surface, their fixed legend system does not always serve the needs of all users. In addition, the yearly update interval does not support the near-real time monitoring of fast changes (e.g. forest loss) or seasonal change (e.g. cropland rotations or changing water levels in ephemeral water bodies). Timely information on land cover and land cover change is however crucial to understand anthropogenic pressures on our environment, the state of ecosystems, and their evolution.

The near-real time mapping of a set of basic, independent diagnostic categories describing the properties of the Earth's surface, called Land Surface Categories (LSC), may complement the information provided in the yearly land cover maps. For instance, continuous, near-real time information on the presence of land surface categories has the potential to capture both land cover dynamics (e.g. forest loss) and to get information on seasonal dynamics (e.g. seasonal presence of water, snow, and ice). The dynamic nature of the LSC may support downstream applications, such as the production of land cover and land cover change products. For instance, temporal information on the presence of bare and herbaceous vegetation may support the classification of cropland and grassland on an annual basis. Other applications may extend to mapping of dynamics (e.g. the mapping of snow, water, or vegetation cycles) or the generation of user-defined classification systems (e.g., the mapping of ephemeral versus persistent water bodies based on user-defined thresholds).

Table 71. Technical thematic specifications for the C10 prototype.

Prototype C10	
Prototype	Continuous Mapping of Land Surface Categories
Thematic specifications	Land Surface Categories (LSC) are basic, independent diagnostic categories and characteristics describing the

	properties of the Earth's surface. This deliverable describes the LSC prototype maps, produced at 10m resolution for single scenes over the test tiles
Prototype sites	Spain (1), Southern France (2), Belgium (3), Central Europe (4), Sweden (5), Eastern Europe (6), Côte d'Ivoire (8)
Temporal frequency	Single scenes (one scene per month for the year 2020, 2021, and 2022)
Spatial resolution	10 meters
MMU	1 pixel
Envisaged accuracy	85%

7.1.2 State of Play

Land cover information that is currently available within the CLMS portfolio has been produced on an annual or multi-annual basis. The Land Surface Categories is thus a novel product within CLMS. Yet, within the Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring Service (LCFM) service, which is part of the global component of the Copernicus Land Monitoring Service, global LSC maps are also foreseen. Towards the end of the EvoLand project, LSC will be mapped at a monthly interval within the LCFM service. Within the EvoLand project, we aim to map LSC at a higher cadence, i.e. using NRT updates.

A set of nine user requirements with respect to the LSC have been collected within the first EvoLand Work Package. The first requirement mentions that the service should be possible to use worldwide. Yet, a Europe only service would still bring considerable value. Second, the service should provide outputs in a format that can be used as an input to other services. This could be seen as providing analysis ready data to other services. Third, the service should support the mapping of wetlands and water extent. Fourth, the service should support the detection of harvesting and emergence events. Fifth, the service should provide data at a 10m resolution (UR-C10-05). Sixth, the service should provide data at a cadence of 3-10 days. Seventh, the service should provide data with minimum latency. No numerical requirement was given for this requirement. It was rather the aim to give the best possible performance. Eighth, the product should have an accuracy of at least 85%. Ninth, the product should have the highest possible precision. Precision means the minimisation of random errors, such that similar inputs always give similar output measurements, whereas accuracy refers to how close the measurement is to the underlying real value.

It should be possible to meet most of the user requirements. The prototype that is being developed uses a spatial resolution of 10m and a NRT cadence. It is based on an automated workflow based on machine and Deep Learning algorithms using auxiliary data and Sentinel-2 imagery, which facilitates the consistency of the outputs and minimizes latency. The selected LSC classes are generic and should support downstream applications, such as mapping water dynamics. Whereas woody and herbaceous vegetation have been included to support the detection of harvesting and emergence events, the probability of the woody vegetation class tends to fluctuate over the seasons, mostly in dry areas. The current prototype has

a slightly lower accuracy (82%) than the targeted overall accuracy of 85%. All classes, except for the shrub class, reach a producer accuracy of 85%, whereas the three vegetation classes (tree, shrub, herbaceous vegetation) show lower user accuracies than 85%. However, note that the accuracy may vary spatially and may depend on the number of classes that are included.

7.1.3 Selection of Prototypes sites

Seven prototype sites were selected to cover a variety in biogeographical regions and LSC dynamics. The prototype area in Sweden for instance shows seasonal change in snow cover and has a significant woody cover, while the prototype area in Belgium contains more built-up and cropland areas, with an alternation between bare and herbaceous vegetation.

7.1.4 Input and Ancillary Data & Pre-processing

We trained a deep learning model (EVONet; (Copernicus, 2025)) across a global set of 1280 by 1280m patches and predicted the LSC over the pre-selected prototype areas within Europe.

Sentinel-1, Sentinel-2 data, the Copernicus DEM, and climatic regions embeddings data were used to train the LSC models and to predict maps over the prototype areas (Table 72). A land cover map of the year 2020 was additionally used to generate annotations.

Table 72. List of datasets used for the C10 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-1 data	Sentinel-1 GAMMA0 monthly composites	20m	monthly
Sentinel-2 data	Sentinel-2 L2A spectral bands	10 - 20m	Single scenes
Copernicus DEM (GLO-30)	Digital Elevation Model	30m	NA
Climatic regions embeddings	AgERA5 variables aggregated across the year and subsequently reduced and normalized into a set of six variables	0.1 degree	annual

7.1.5 Methodology and Workflow

The aim of the subtask is to develop a candidate method that focuses on mapping LSC on single scenes using a legend covering 10 classes (Table 73).

Table 73. Class legend for the C10 prototype development.

Label	Description	Colour
1	Tree	Dark Green
2	Shrub	Orange
3	Herbaceous vegetation	Yellow
4	Bare	Grey
5	Water	Blue

6	Snow/ice	
7	Built-up	
254	Shadow	
255	Clouds	
0	No data	

Whereas we focussed on a CatBoost model within the first phase of the project, a deep learning architecture (EVONet) was tested and selected within the second phase. The EVONet architecture aims to include contextual information while preserving fine spatial details and consists of two components. The first component consists of a convolutional neural network and aims to extract spatial context, resulting in a set of spatial features. The second component of the model is a pixel-based multi-layer perceptron (MLP) and uses the spatial features, spectral information, DEM, and set of localizing information (latitude, longitude, time, and climate) to predict the LSC classes. More detailed information about the EVONet model can be found in Copernicus (2025).

During training the LSC classes were divided into three groups. The first group are the physical surface categories, which consist of the tree, shrub, herbaceous vegetation, not-vegetated land, and water classes. The second group consists of occlusion classes, which include the snow, clouds, shadow, and surface classes. The last group consists of attribute classes, including cropland, mangrove, built-up, herbaceous wetlands, and lichens classes. Each of the groups of LSC classes were trained with their own loss and weights. During inference, each of the classes are predicted and subsequently merged into the final set of LSC classes.

Annotations

In order to train the EVONet model, an extensive training dataset is required. Yet, since manually collecting annotations from single scenes is very costly and time consuming, we applied an automated workflow to produce the annotations.

We used the land cover labels to derive the most likely LSC label and additionally applied a set of rules on the spectral data to ensure that we capture intra-annual changes. We for instance employed a rule on the Normalized Difference Water Index (NDWI) to ensure that water was present in pixels that were labelled as water according to the annual land cover labels and checked for the presence of snow using snow indices and temperature data. To distinguish between herbaceous vegetation and bare surfaces, we employed a threshold of 0.15 on FCOVER. Information on the presence of clouds and cloud shadows was derived from a U-Net model trained on the CloudSen12 dataset (Aybar et al., 2022b) using the Sentinel-2 bands B02, B03, B04, B8A as input.

These rules resulted in a set of probabilities for each of the physical surface, occlusion, and attribute classes.

Besides the probabilities, weights were defined for each of the three groups of LSC classes. These weights allow to give lower importance to pixels for which the

annotation is unsure or unknown. For example, pixels covered by thick clouds were given a weight equal to one for the occlusion classes and zero for the physical cover classes. As such, the model is not penalized when it cannot distinguish the surface classes over cloudy pixels. In addition, low weights were given for land cover classes that may contain a mixture of LSC classes (e.g. wetland), are prone to confusion, or for pixels for which the inferred LSC labels conflicts with the spectral information (e.g. pixels classified as water showing a low NDWI).

The annotations were prepared using single Sentinel-2 scenes of the year 2020 over a global set of about 45000 patches. For each location, we tried to collect one observation per month that had a minimal cloud cover and no missing data. The patches covered an area of 1280x1280m and had a spatial resolution of 10m. The annotations were finally split into a training and validation set based on location using a division of 80% and 20%, respectively.

An example of annotations generated over a patch located in a cropland area is shown in next figure.

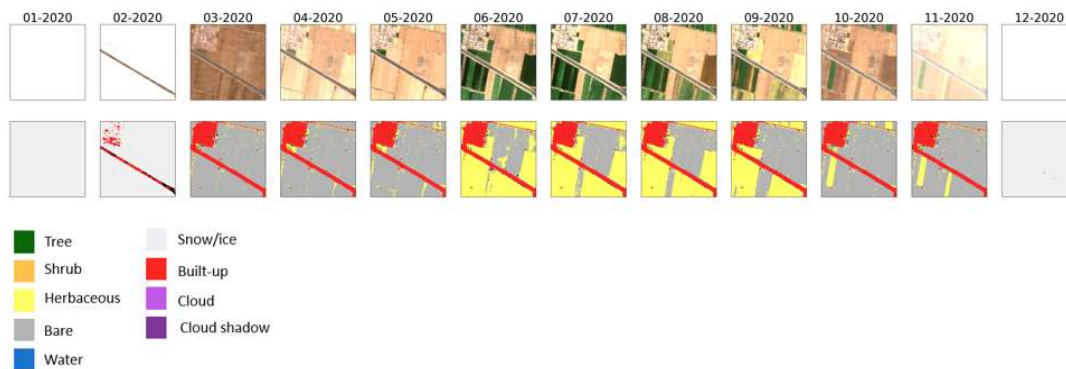


Figure 100. Example of annotations generated over a patch located in a cropland area. The upper row shows the Copernicus Sentinel-2 data for each month and the bottom row shows the associated annotations.

EVONet model

Two EVONet models were subsequently trained on the patch annotations using different inputs. First, a model was trained using the DEM and nine Sentinel-2 bands (B02, B03, B04, B08, B05, B06, B07, B11, B12) as input. On top of these data, the MLP head of the model was fed with additional localizing features, including the location (latitude and longitude), circularly encoded day of the year, and climatic regions embeddings. Second, a model was trained with monthly Sentinel-1 mosaics (VV, VH, and VH to VV ratio) on top of the features of the first model. 10% of the Sentinel-1 data were randomly filled with a fill value to learn the model to handle missing Sentinel-1 observations. As such, inference of this model can be performed using both Sentinel-1 and -2 data as input, but also without Sentinel-1 observations. For both models, an MSE loss was used, where classes were weighted based on their abundance in the training dataset.

7.1.6 Results and Verification

Validation dataset

To evaluate the models, we prepared an additional validation dataset derived from manually labelled patches over each of the prototype areas. The labels were collected based on a set of land cover classes and using the annotation framework developed within EvoLand, resulting in a set of annotated segments over the annotation patches. Road annotations from OpenStreetMap were additionally merged into the validation dataset. Similar to the training data, we selected one Sentinel-2 observation per month for the year 2020, without missing data and with the least cloud cover, and applied a set of rules to convert the annual land cover labels to LSC labels matching the selected Sentinel-2 observations.

In total, annotations were collected over 167 locations (32 in Spain, 28 in Southern France, 12 in Belgium, 40 in central Europe, 8 in Sweden, 26 in Eastern Europe, 12 in French Guyana, and 9 in Côte d'Ivoire), resulting in 1996 samples in total.

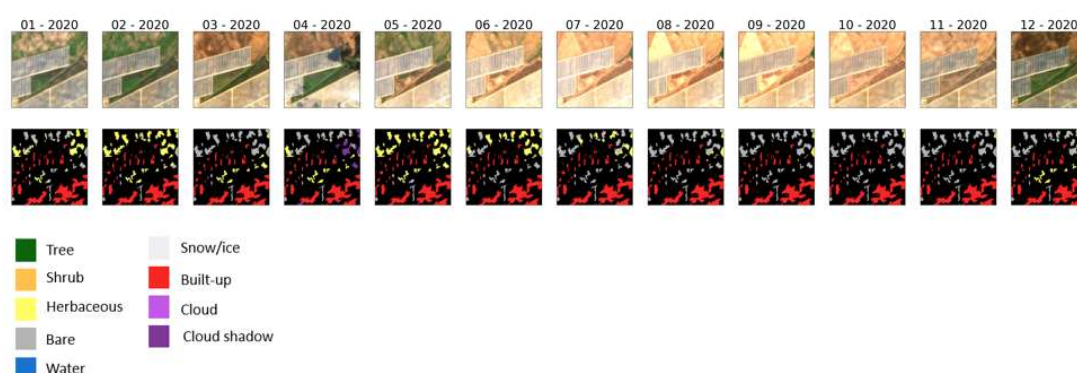


Figure 101. Example of LSC annotations over the validation dataset.

Benchmark

A benchmark has been performed on the EVONet models using different sets of input data. Specifically, the effect of Sentinel-1 data was compared (Table 74).

Table 74. Overview model benchmark.

Name	Model trained with Sentinel-1 data	Sentinel-1 data used during inference
Tr-_Inf-	No	No
Tr+_Inf-	Yes	No
Tr+_Inf+	Yes	Yes

This benchmark showed that the models had a similar accuracy (Tr+_Inf+: 82.27%, Tr-_Inf-: 82.45%), with slightly lower values when the Sentinel-1 data was masked during inference (81.03%). Since the availability of the monthly Sentinel-1 Gamma0 mosaics is not guaranteed overall years and locations, the model trained without Sentinel-1 data was selected. The accuracy metrics of this model are shown in Table 75.

Table 75. Confusion matrix of the LSC model over the test data.

		REFERENCE									Total	User Acc.	F1 score
		Tree	Shrub	Herb	Bare	Water	Snow	Built	Shadow	Cloud			
MAP	Tree	907728	113350	81504	4343	4238	597	193043	3506	8919	1317228	69%	77%
	Shrub	14937	99112	27777	2343	0	0	2985	0	416	147570	67%	40%
	Herb.	60814	105950	936636	13037	897	18	132381	1109	10975	1261817	74%	80%
	Bare	7179	30405	5065	394716	3127	18	36414	690	2947	480561	82%	87%
	Water	708	1	331	965	459354	1	837	1112	1520	464829	99%	98%
	Snow	369	267	123	787	669	131774	182	133	3434	137738	96%	99%
	Built	55450	2044	20219	10373	1366	34	1335994	2736	9810	1438026	93%	85%
	Shadow	2346	269	2178	175	118	0	3089	82344	4607	95126	96%	93%
	Cloud	3186	335	2259	219	299	1	4530	3762	397018	411609	87%	86%
	Total	1052717	351733	1076092	426958	470068	132443	1709455	439646	95392	5754504		
		86%	28%	87%	92%	98%	99%	78%	90%	86%			
Prod. Acc.		83%											
User Acc.		82%											
Overall Acc.		82%											

This model has been applied to test tiles of the prototype areas within Europe. To minimize the data volume and facilitate the inspection of the dataset, we selected and produced LSC maps for single scenes for each month of the years 2020, 2021, and 2022 having a minimal number of missing data and cloud cover.

Qualitative evaluation

LSC time series over three tiles in Romania, Sweden, Spain are shown in Figure 102, Figure 103, and Figure 104, respectively. The examples illustrate the capacity of the LSC maps to capture temporal dynamics, such as fluctuations between sparsely vegetated areas and herbaceous cover over cropland areas (Figure 102) and snow/ice dynamics in Scandinavia (Figure 103). Yet, over dry areas with sparse woody vegetation, such as in Spain (Figure 104), the model tends to confuse tree and shrub vegetation, resulting in seasonal fluctuations between shrubs and trees.

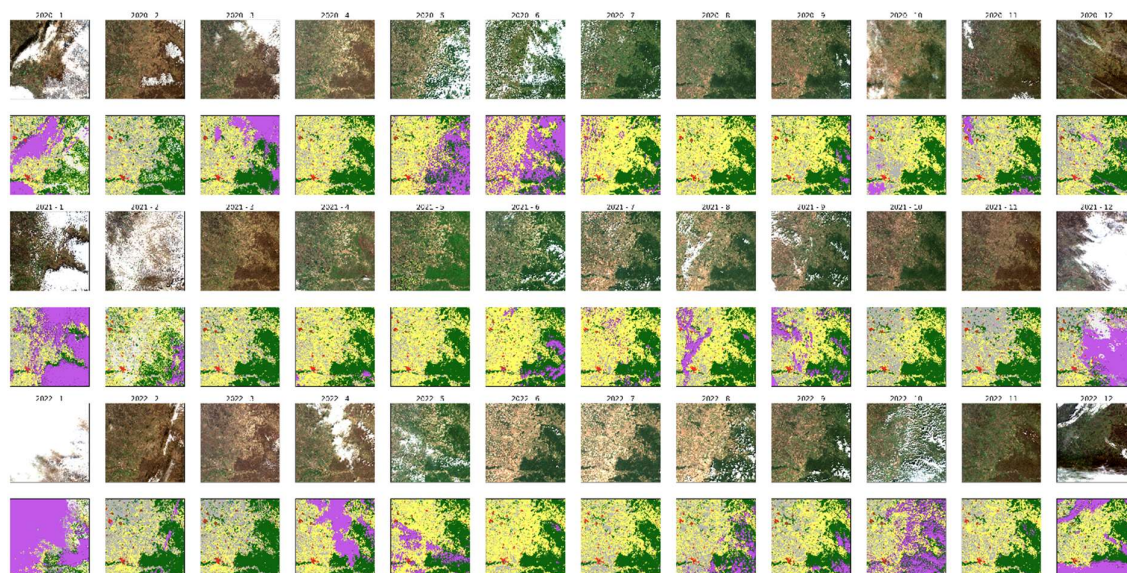


Figure 102. Example of LSC maps over tile 34TES (Romania, Hungary). Each column represents a month (Jan – Dec) and each row represents a year (2020-2022). The corresponding Sentinel-2 imagery is shown together with the maps. The time series of LSC maps show fluctuations between herbaceous and bare cover over the cropland area, while the tree class remains stable over time.

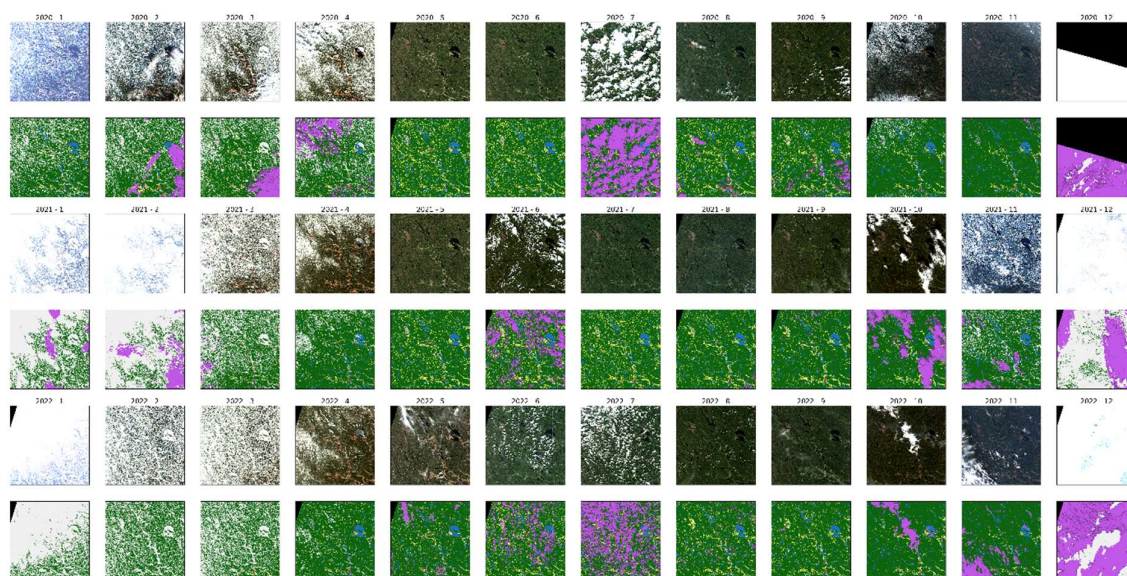


Figure 103. Example of seasonal changes in snow and ice cover over tile 33VWJ (Sweden).

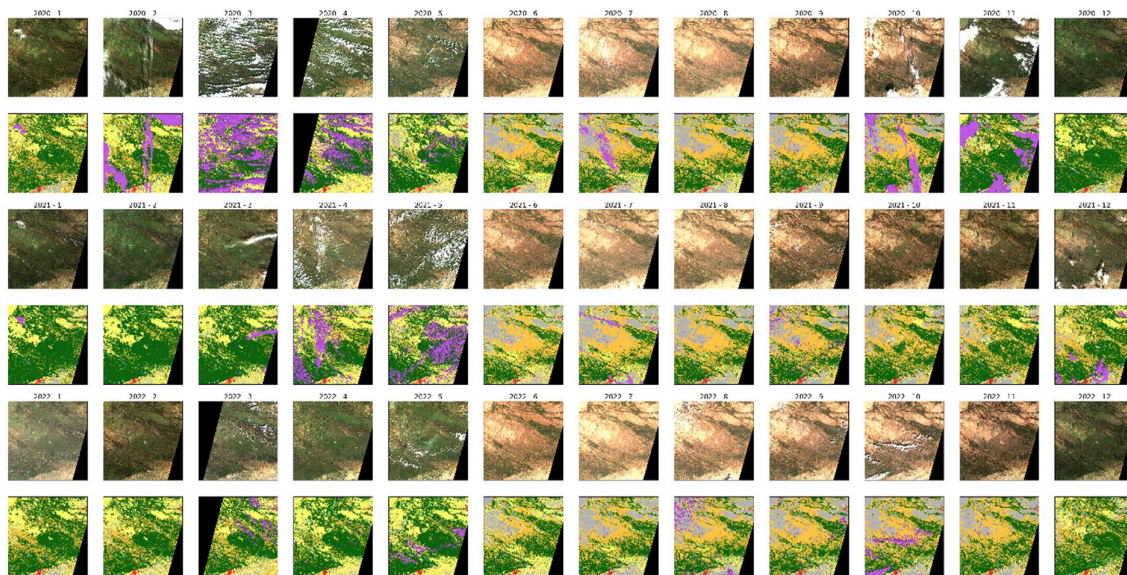


Figure 104. Example of seasonal confusion between the tree and shrub class within tile 30SUH (Spain).

7.1.7 Conclusions

A deep learning-based model for NRT mapping of the land surface categories was developed during the EvoLand project. Visual inspection and evaluation on the validation set showed that the LSC captures the classes with reasonable accuracy. The maps capture seasonal fluctuations, such as emergence and harvest of crops, water dynamics, or snow dynamics. Yet, the probability of the woody vegetation classes tends to show seasonal fluctuations, mostly over dry or sparsely vegetated areas.

7.1.8 Outlook beyond the EvoLand project

The prototype demonstration of continuous mapping of Land Surface Categories shows the added value and potential of the maps to capture dynamics and provide information to downstream applications.

Further development of the LSC model is planned within the Tropical Forest Mapping and Monitoring Service (LCFM) service, which plans to produce NRT LSC maps. Here, special attention will go to the seasonal confusion of the vegetation classes.

7.2 C11 – On-demand Land Cover Mapping

7.2.1 Introduction & Objectives

A wide variety of ground-truth datasets for land cover classification exist in private and public data storage spaces and are used to train land cover classification models. Oftentimes, many of these datasets are subjected to privacy regulations,

or users may prefer to use the datasets within their own applications without sharing the data. Additionally, building and training a land cover classification model is an intensive process requiring specific skills and expertise, which not all potential users might have. Furthermore, users might face resource and infrastructure challenges when one needs to train a model and conduct inference on satellite data. They must be capable of extracting, downloading, and processing large quantities of data which is difficult without proper resources.

The C11 prototype aims to address part the issues above by demonstrating a cloud-based model inference pipeline on openEO. The openEO platform can access public data from the CDSE backend and private data through user-created STACs. In this prototype, we demonstrate the usage of private STACs. Furthermore, we illustrate how users can utilize openEO to generate land cover maps based on their region of interest, time period, machine-learning model and its parametrizations. The use case focusses on forest management type classification and uses the technical specifications defined in Table 76:

Table 76. Technical thematic specifications for the C11 prototype.

Prototype C11	
Prototype	On-demand Land Cover Mapping
Thematic specifications	On-demand forest management type classification
Prototype sites	Belgium, Central Europe, French Guyana, and Côte d'Ivoire
Temporal frequency	Single map for the year 2020
Spatial resolution	10 meters
MMU	1 pixel
Envisaged accuracy	TBD

7.2.2 State of Play

The on-demand land cover prototype is novel and there is no precursor CLMS product. The prototype is highly innovative though its dynamic nature enabling customizable classes at different spatial resolutions and observation periods. This candidate relies on developments in in-situ data, generic embeddings, improved resolution and novel infrastructure.

7.2.3 Selection of Prototypes sites

Four prototype sites were selected to demonstrate the use case on on-demand forest management type mapping, i.e., those in Belgium, Central Europe, Côte d'Ivoire, and French Guyana.

7.2.4 Input and Ancillary Data & Pre-processing

The following datasets were collected for the use case (Table 77):

Table 77. List of datasets used for the C11 prototype development.

Dataset	Description	Spatial Resolution	Temporal resolution
Sentinel-1 data	Monthly Sentinel-1 GAMMA0 mosaics	20m	Entire year 2020
Sentinel-2 data	Sentinel-2 L2A spectral bands	10m - 20m	Entire year 2020
Climatic region embeddings	AgERA5 variables aggregated across the year and subsequently reduced and normalized into a set of six variables	0.1° x 0.1°	Entire year 2020
Copernicus DEM (GLO-30)	Digital Elevation Model	30m	NA
Spatial embeddings	Spatial features extracted from an EVONet model trained on land cover data	10m	2020

7.2.5 Methodology and Workflow

This prototype aims to develop a generally applicable cloud-based pipeline for inference of a classification model in openEO using input data (both public and private) for given spatial and temporal extents. We have used forest management type classification as a use case to develop and test the modelling pipeline. The following forest management type classes were distinguished (Table 78):

Table 78. Class legend for the C11 prototype forest management type classification.

Label	Description	Colour
0	No tree	
11	Naturally regenerating forests without any signs of management	
20	Naturally regenerating forests with signs of forest management	
31	Planted	
32	Plantation	
33	Rubber	
40	Oil palm	
50	Fruit tree	
53	Agroforestry	
100	Other trees	

First, the input data used for training the machine learning model and its inference is described, followed by a brief description of the model, its training, and post-processing steps. Finally, we discuss the main objective of this prototype, which is porting the inference and post-processing pipeline to openEO platform.

Input layers

From the annual time series of Sentinel-1 and -2 data, temporal statistics were computed. These temporal statistics included the 10th, 25th, 50th, 75th, 90th percentile of observations over the year 2020 for the Sentinel-2 bands B02, B03, B04, B08, B11, B12, the NDVI and the Sentinel-1 polarizations VH, VV, and the ratio between VH and VV. In addition, the number of available monthly Sentinel-1 mosaics and two quality indicators related to cloud coverage of the Sentinel-2 data were generated.

Next to the Sentinel data, a set of localizing features, including AgERA5 climate embeddings, the location, and DEM were used. These features provide biogeographic context to the model.

Finally, a set of spatial embeddings, i.e., spatial embeddings extracted from an EVONet model trained on land cover data, were utilized in the model. These spatial embeddings add spatial context to the spectral and localizing information.

An overview of the model inputs is shown in the table below (Table 79):

Table 79. List of input data for the ML model of the C11 prototype.

Dataset	Input layers
Sentinel-1 data	Temporal statistics of VH, VH, VH/VV ratio, and the number of available observations for the year 2020
Sentinel-2 data	Temporal statistics of B02, B03, B04, B08, B11, B12, NDVI, and two input quality bands for the year 2020
AgERA5 climate embeddings	AgERA5 variables aggregated across the year 2020 and subsequently reduced and normalized into a set of six variables
Copernicus DEM (GLO-30)	Elevation
Location	Latitude/longitude
Spatial embeddings	40 bands of spatial features extracted from an EvoNET model trained on land cover data

Model training

A training dataset with manually collected training labels over about 230.000 100 m by 100 m patches was used to train the forest management type model. This dataset contained all forest management classes, except for the 'no tree' and 'other tree' classes.

Since contextual information might be relevant to distinguish forest management types, an EVONet model (Copernicus, 2025) was selected for the development of the use case. As the limited patch size of the training data might hamper the model to learn relevant spatial features, only the MLP component of the EVONet model was trained and the spatial features were extracted from a model that was trained on land cover data. The forest management type model was trained using MSE loss where weights were applied inversely proportional to the abundance of each forest management type class. The model was subsequently evaluated using a qualitative and quantitative assessment.

Postprocessing

Several postprocessing steps were developed to introduce the missing 'no tree' and 'other tree' classes.

First, all pixels that were not covered by trees in the LCFM land cover map (LCM-10) of the year 2020 were assigned to the 'no tree' class. Second, the 'other tree' class consists both of urban trees and sparse trees that do not belong to any other class (e.g. rows of trees alongside a road). Urban trees were detected as treed pixels within urban areas (areas with high built-up density) that were classified as agroforestry. Sparse 'other trees' were identified by treed pixels of the 'agroforestry',

‘planted’, and ‘naturally regenerating forests with signs of forest management’ classes that showed high confusion between those classes and a low tree density. The built-up and tree density were extracted by extracting binary presence layers of those classes from the LCFM land cover map of the year 2020 and running a gaussian kernel over those presence layers.

OpenEO pipeline

The model inference was implemented in openEO. The first part of the openEO pipeline was data curation. Input data was extracted from two sources: the DEM was extracted from CDSE while STACs were created for the other input data. The creation and usage of STACs illustrates the storage of private data and reduces processing time. The latter is for instance the case for the annual Sentinel-1 and Sentinel-2 features, which were not computed on the fly in openEO to reduce processing time and costs. The tree datasets used in the post-processing step, i.e., the LCFM land cover map (LCM-10) of the year 2020, tree density, and built-up density were additionally loaded from a STAC collection.

After the data was curated, the machine learning model was ported to ONNX format to facilitate the deployment in openEO. Lastly, the code was ported to openEO as shown in Figure 105.

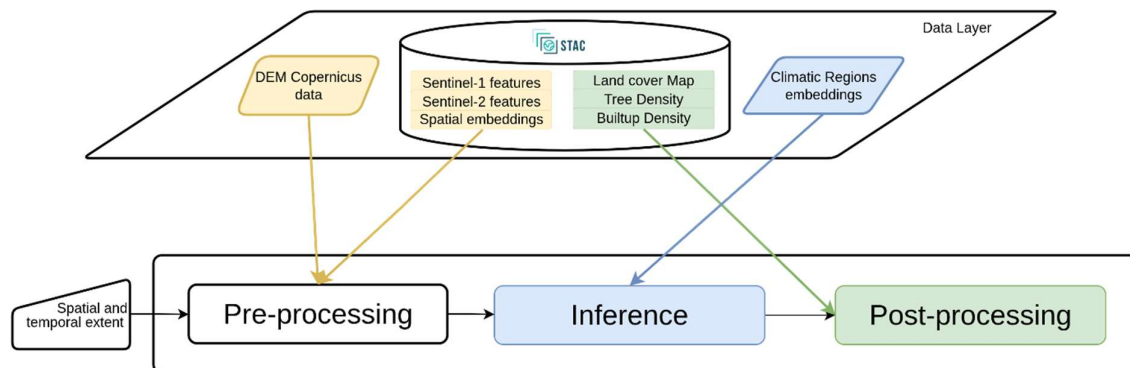


Figure 105. Overview of the openEO pipeline.

7.2.6 Results and Verification

The classification pipeline was applied to the prototype areas in Belgium, Central Europe, Côte d'Ivoire, and French Guyana. An example of the inference output within Côte d'Ivoire can be seen in Figure 106 showing naturally regenerating forest without management surrounded by a mosaic of agroforestry and naturally regenerating forest with management, and several patches of rubber, oil palm, and fruit trees.

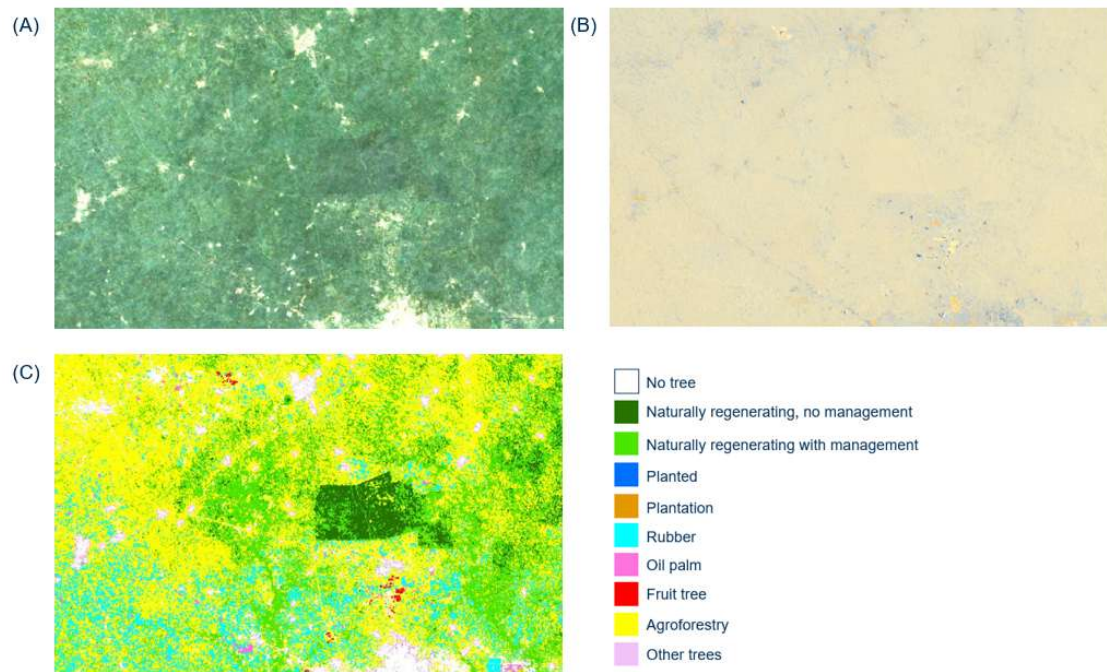


Figure 106. Example of forest management type classification in Côte d'Ivoire. (A) annual median Sentinel-2 composite, (B) annual median Sentinel-1 composite, and (C) forest management classification.

To validate the openEO pipeline, the maps produced via openEO were compared against a local run of the pipeline as shown in Figure 107. The two runs showed differences in 0.068% of the pixels in a 1024x1024 pixels area. This difference is due to the ONNX conversion and slight difference between the local and CDSE DEM collection.



Figure 107. Forest management layer processed in openEO (left) and locally processed forest management layer (right) over an area of 1024x1024 pixels (10240 by 10240 m). About 0.068% of the pixels are different.

7.2.7 Conclusions

Within the on-demand land cover mapping prototype, we built and tested various aspects of the prototype pipeline, including the usage of user generated STACs, model upload, and inference in openEO. We were able to demonstrate the feasibility of this prototype and deliver classified forest management type maps covering the prototype areas.

7.2.8 Outlook beyond the EvoLand project

Although the on-demand land cover mapping pipeline enables users to perform inference of custom models on their areas of interest, the pipeline could be further expanded, allowing for more flexibility to users.

For instance, expansion of the pipeline could allow for optimization of a global model over their area of interest. Users would be provided with a pre-trained global model. They could perform inference of this model at custom locations or years but also fine-tune and optimize the global model to enhance it over their area of interest. This could be achieved by first uploading external training labels or collecting training labels within a user interface. These training labels could subsequently be used to perform a fine-tuning of the pre-trained global model. As the model has been pre-trained with global data, this fine-tuning step should be rather light-weighted and fast. The fine-tuned model would then be available again for inference over their area of interest.

8 Summary and Conclusions

The prototyping part of the EvoLand project has been successfully concluded, with all the eleven foreseen prototypes in the fields of monitoring forest, agriculture, water, urban areas and general land cover having been produced. Beyond that, even a 12th prototype has been implemented during Phase 2, as a direct reaction to evolving stakeholder requirements and answering user needs. The prototypes, available towards the end of Phase 1, have already been presented to the main stakeholders (i.e. EEA and JRC) in a specific meeting in mid-June 2024, at their almost final version. Whereas some prototypes are still in an early development status at the end of the project, there are others which are already more advanced and reached a close-to-operational level, i.e., reaching a TRL of ~7.

Overarching over all twelve prototypes it can be reported that the targets of demonstration in minimum three prototype sites and over three reference years (2020 – 2022) were reached for all but two prototypes, as explained in section 1.2 (see also Table 3 in section 2.2) and further below. The method transfer from WP2 method development to WP3 prototyping was established in Phase 1 and further method uptake was considered in Phase 2 for potentials already identified but not implemented in Phase 1 (see Table 4 in section 2.3). For further ideas and methods that have been partially already discussed with the stakeholders, it is referred to the paragraphs further down in this section, where these are described specifically for every prototype.

The target of demonstration in **minimum three prototype sites** was achieved by eight prototypes, namely C1 - Continuous Forest Monitoring, C2 - Forest Disturbances Mapping, , C6 - Small Landscape Features, C7 - Improved Water Bodies Mapping, C8 - Continuous Imperviousness Monitoring, and C9 - Automated Land Use Mapping of Urban Dynamics, C10 - Continuous mapping of Land Surface Characteristics and C11 - On-demand Land Cover Mapping. C3 - Biomass Mapping was rolled out in 4 demonstration sites, while the results for two of them were not considered (see 3.3). C4 - Cover Crop Type Mapping and C5 - Cropland/Grassland GPP Monitoring were rolled out for a subset of only one prototype site, due to data limitations. C12 was added later in Phase 2 and produced for one enlarged prototype site.

The target of covering **three full reference years (i.e., 2020–2022)** was achieved by all but three prototypes. C5 was implemented for 2018 and 2022 only and C12 was implemented solely for 2021, as it was introduced in Phase 2 only and the inclusion of further reference years was therefore not possible. A further exception is C11 - On-demand Land Cover Mapping which is a single scene implementation. For further information see Table 3 in section 2.2).

Considering **method transfer**, all prototypes were suited to pick up one or more methodologies developed in WP2, transfer it to a prototype and deploy it over larger demonstration sites. For further details see Table 4 and method descriptions in section 2.3).

In terms of **topic-wise conclusions**, the main messages from the twelve prototypes are as follows:

The **C1** prototype can successfully capture subtle forest disturbances, but there is a trade-off between sensitivity of the detection of losses of tree cover on a small spectral magnitude and false-detections, leading to commission errors. As an unsupervised approach having the advantage of not needing any training data, it still requires a large amount of EO input data in terms of dense time series of Sentinel-2 data, adding a 3-year training period before the start of the analysis period. The cloud masking which turned out to be a major source of misclassifications in Phase 1 was improved in Phase 2, enabling a substantial reduction of false detections. Even if the approach requires massive amounts of data, in principal data and processing requirements appear feasible for a pan-European approach, with sufficiently scaled cloud infrastructure.

The Forest Disturbance classification prototype (**C2**) is still in a development stage after Phase 2 of the project. Whereas it has shown to be able to detect patches of the main disturbance agents correctly, it is still not reliable enough on a pixel-basis. Other factors like the dependence on the quality of C1, especially with regards to the detection delay, and most importantly missing reference data across Europe, hamper its implementation as an operational product. The prototype has high demands with regards to EO data download and time series pre-processing which are time-consuming. But in principle, the data needs would be feasible for a pan-European approach. As mentioned, the greater limitation is the high demand of reference data which is not available for a Europe-wide coverage (section 3.2.7).

The derived models of **C3** for both AGB and FCH, based on a combination of optical and SAR time series, DEM-based information and GEDI LiDAR data, are able to depict land cover structures that are visible in VHR reference data. However, there is still room for improvement in terms of accuracy, and further investigations on the variability of AGB and FCH results in different scenarios are recommended. The models seem to saturate at higher biomass values as well as larger tree heights, thus underestimating AGB and FCH. For AGB the model further seems to also overestimate the result at lower biomass values. If available, NFI data should be preferably used, but also GEDI seems to be a reasonable data source for training and as a reference. The testing showed that aggregation for final products can reduce noise, and acceptable accuracies could be achieved with an aggregation to 0.25 ha MMU (in comparison: current global products apply 1ha MMU). These are only initial results and the potential for a Europe-wide application without data from the NISAR mission (or comparable data) is limited, as concluded after Phase 2. In any case, it is recommended to add FCH to an AGB product, since the workflow is similar as for AGB estimation, and it can be produced without a lot of additional effort while serving additional use cases.

The **C4** results show a relatively good separability of grass-like cover crops from other winter covers, but at this stage the accuracy does not yet meet the requirements. One of the main limitations for cover crop mapping is the lack of

reference data. LPIS data rarely contain information on cover crops and secondary crop declarations carry some uncertainty. Another challenge are mixed classes of cover crop families, up to now only two broad groups were differentiated (section 4.1.7).

The **C5** prototype's high-resolution cropland and grassland specific GPP product was increased in resolution from 300m to 10m which allows to observe within field GPP variations, which was one of the user requirements of this prototype. Furthermore, the existing LUE method used for GPP estimation was adjusted by calibrating the LUE factor specifically for crop and grassland locations. A validation through ICOS sites proved the precision to be rather high, with especially for cropland more reliable results. Furthermore, this consistency is also relevant for continuity of the products in the CLMS portfolio. Currently, the GPP is still somewhat overestimated during drought periods and underestimated during the off-season, which needs to be further investigated (section 4.2.7).

The **C6** Small Landscape Features prototype maps small woody features (SWF) with an improved resolution compared to existing SWF products, through the use of Pleiades imagery. Moreover, it includes new classes of agricultural landscape features, which are, however, still generic and does not yet discriminate requested landscape features (grass margins, ditches, stone walls) when detailed Digital Height Model are not available. While it does not need to rely on in situ data for training or calibration, in situ data (e.g. height dataset) bring significant added value to this prototype. The generalisation of height data availability (LIDAR, future EO mission) will be a major step forward for the operationalisation of this prototype. It is computationally efficient and takes advantages of super-resolution Sentinel-2 data, implying that newly included features will be mapped with a spatial resolution of 5m at best. A significant amount of pre-processing is mandatory to compute those datasets before proceeding to extract landscape features from it (section 4.3.7).

The **C7** water body detection algorithm is close to operational, fulfilling key requirements with short processing time and preserving the native 10 m resolution. Temporal flexibility allows for daily or monthly outputs, depending on user needs. Detection is no longer constrained by the GSWE mask, enabling the identification of new water bodies, unlike CGLOPS. To address increased commission errors—mainly due to topographic shadows and low solar angles—two improvements were introduced: reflectance normalization based on solar zenith angle and topographic shadow masking. A new method was also developed to classify water bodies as permanent or ephemeral, using monthly updated masks and past occurrences to compensate for missing data. This approach is close to operational for NRT water classification and represents a novelty in recurrent products. Currently, only Sentinel-2 data are used, but integrating Sentinel-1 SAR data could help mitigate cloud-related issues. The exclusion mask may still be too strict in some cases, potentially omitting valid water bodies (section 5.1.7).

The developed processing chain for **C8**, “Continuous Imperviousness monitoring” is capable of handling a substantial volume of input data efficiently and within a reasonable timeframe, while also delivering high-quality classification results. The integration of Sentinel-2 super-resolution technology on Sentinel-2 L3A has shown considerable promise. The positive results underscore the robustness of the developed system and its applicability in various large-scale data processing scenarios. While errors of omission and commission are present, especially the confusion with rocks or bare soils and omissions in roads or scattered built-up areas, the underestimation of imperviousness and commission errors has significantly reduced by the use of more comprehensive training data, aggregation of time series, and improved calibration methods, as reflected in the regression and thematic accuracy results. The processing chain has also been extended to include change layers describing sealing and de-sealing processes, produced consistently with the 6-month update interval of the imperviousness status product.

The **C9** developments for “Automated Land Use Mapping of Urban Dynamics” contained the automation of a process that until then had been carried out only by visual analysis. Sentinel-2 L3A proved to be more suitable than the originally intended L2 data, particularly for eliminating clouds that could otherwise be detected as false changes. Annual updates of the products should be possible with latency of no more than one year. The assumption is taken that the approach is transferable to other regions, preferably where L3A data exist. The MMU of 0.25ha has been found suitable to detect changes. The prototype is based on a robust method that always provides similar output measurements with similar input, thus minimizes errors.

C10 developments contain an initial methodology for NRT mapping of LSC. Visual inspection and evaluation on the test set showed that the LSC capture the classes with reasonable accuracy. Changes, such as alternations between bare and herbaceous vegetation or forest loss were detected. Yet, the probability of the woody vegetation class over sparsely vegetated areas tends to show seasonal fluctuations and there seems to be confusion between the built-up and the bare class. In addition, some artefacts such as the prediction of water over topographic shadows are still present (section 7.1.7).

Various aspects of the **C11** prototype pipeline were built and tested, including training data preparation, feature extraction, model training and upload, and inference in openEO. The feasibility of this prototype to deliver classified tree species maps covering areas up to entire Sentinel-2 tile was demonstrated. Further refinements of the pipeline should focus on improving the usability and customization available to the end user (section 7.2.7).

The newly introduced **C12** showed strong potential for tree type classification with Sentinel-2 time series data. The integration of diverse data sources, namely GBIF and TreeSat, combined with a rigorous filtering process, was a key factor in achieving an overall accuracy of 76.3% on the test set. While the model excels at classifying common species such as *Quercus* and *Pinus*, performance for less-

represented classes is limited by the number of available training samples. Classification performance was particularly strong for dominant species with distinct phenological signatures, including *Pinus* and deciduous *Quercus*, whereas smaller classes which are often also mixed with other species suffered from confusion. The overall accuracy values of around 76% are therefore promising for operational applications in Central Europe, but class imbalance remains a significant challenge, especially for underrepresented species.

In conclusion, the prototyping part of the EvoLand project can be considered successful, with all the eleven foreseen prototypes developed and an additional one introduced on the way to account for evolving requirements from stakeholder side. Phase 2 prototypes have mostly shown very good advances and results are pointing towards a way to go for future evolutions and prototypes for new or improved products. The development level of the different prototypes, however, varies from still early development states to close to operational. In WP4 the level of operability is analysed and in-depth information will be provided in a dedicated report.

An outlook beyond the EvoLand project is provided in the following Chapter 9.

9 Outlook and Way Forward

As concluded in chapter 8, some of the 12 developed prototypes are more mature than others, which is also reflected in the outlook and the way forward with regards to requirements that have to be fulfilled to transform the different prototypes into operational products.

For some prototypes the availability of adequate EO and/or reference data, turned out to be an issue. The absence of high-quality training data, covering the whole of Europe remains a limiting factor for parts of the prototype portfolio which might be true also in the near future. Regarding the availability of new EO data sources, a next step has been made with the launch of the NISAR mission (unfortunately too late for this project), which is expected to provide an added value, especially for biomass mapping. Future work will have to demonstrate the potential of such new data sources to contribute to the feasibility of pan-European implementations, where in EvoLand it turned out to still be limited.

The output of this prototype report and adjoint data deliverable [AD13] and source codes deliverable [AD12] are direct input to the benchmarking in WP4. The aim is to identify, on the one hand, the most urgently and topically needed candidate services, and on the other hand, to assess which ones are the most technologically and scientifically advanced candidates. This concludes to the potential for a pan-European roll-out while providing significant added-value to the CLMS portfolio with potential for cost-efficient production. It will be determined which products qualify for being suggested as most promising candidates for future implementation into the CLMS operational product portfolio from 2027 onwards.

References

- Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B., & Kleinschmit, B. (2023). *TreeSatAI Benchmark Archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing*. *Earth System Science Data*, 15(2), 681–695. <https://doi.org/10.5194/essd-15-681-2023>
- Aybar, C. (2024). cloudSEN12. <https://huggingface.co/csaybar/cloudSEN12>. <https://doi.org/10.57967/hf/2291>. Publisher: Hugging Face.
- Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., Espinoza, W., Prudencio, F., Llactayo, V., Montero, D., Sudmanns, M., Tiede, D., Mateo-García, G., & Gómez-Chova, L. (2022). CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Scientific Data*, 9(1), 782. <https://doi.org/10.1038/s41597-022-01878-2>
- Bartaloš, T., Dufek, J., Ruf, K., Mancosu, E., Plouvier, W., Maucha, G., Turos, V., & Kerékgyártó, É. (2025). *Task 2: Product development support—HRL vegetated, Copernicus 2024/2025* (Negotiated Procedure EEA/DIS/R0/24006 No. Version 0.9; p. 67). EEA, ETC DI. <https://land.copernicus.eu/en/technical-library/hrl-vegetation-layers-comparative-analysis-forest-classification-and-temporal-consistency>
- Blickensdörfer, L., Oehmichen, K., Pflugmacher, D., Kleinschmit, B., & Hostert, P. (2024). National tree species mapping using Sentinel-1/2 time series and German National Forest Inventory data. *Remote Sensing of Environment*, 304, 114069. <https://doi.org/10.1016/j.rse.2024.114069>
- Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., & Asner, G. P. (2018). An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. *Remote Sensing of Environment*, 206, 156–173. <https://doi.org/10.1016/j.rse.2017.12.030>
- Braaten, J.D., Cohen, W.B. and Yang, Z., 2015. Automated cloud and cloud shadow identification in Landsat MSS imagery for temperate ecosystems. *Remote Sensing of Environment*, 169, pp.128-138.
- Brooks, E. B., Wynne, R. H., Thomas, V. A., Blinn, C. E., & Coulston, J. W. (2014). On the-Fly Massively Multitemporal Change Detection Using Statistical Quality Control Charts and Landsat Data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6), 3316–3332. <https://doi.org/10.1109/TGRS.2013.2272545>
- Camp-Valls, G., & Bruzzone, L. (2009). *Kernek methods for remote sensing data analysis*. John Wiley & Sons.
- Copernicus (2025). *Land Cover 2020 (raster 10 m), global, annual - version 1*. Available at: <https://land.copernicus.eu/en/products/global-dynamic-land-cover/land-cover-2020-raster-10-m-global-annual>
- d'Andrimont, R., Verhegghen, A., Lemoine, G., Kempeneers, P., Meroni, M., & Van Der Velde, M. (2021). From parcel to continental scale—A first European crop type map

- based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sensing of Environment*, 266, 112708.
- Dalla Mura, M., Benediktsson, J., Waske, B., & Bruzzone, L. (2010, October). Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10), 3747-3761.
- Deng, J. S., Wang, K., Deng, Y. H., & Qi, G. J. (2008). PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16), 4823–4838. <https://doi.org/10.1080/01431160801950162>
- EEA. (2025). *Feasibility study on the extension of the HRL Tree Cover & Forest portfolio towards a European Forest Type map* (Project Report No. D1.4; Copernicus Land Monitoring Service – High Resolution Layer - Vegetated Land Cover Characteristics, p. 46). European Environment Agency. <https://land.copernicus.eu/en/technical-library/feasibility-study-on-the-extension-of-the-hrl-tree-cover-and-forest-portfolio-towards-a-european-forestry-map>
- Feyisa, G.L., Meilby, H., Fensholt, R. and Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, 140, pp.23-35.
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342 (15 November): 850-53. Data available on-line from: <https://glad.earthengine.app/view/global-forest-change>
- Hermosilla, T., Bastyr, A., Coops, N. C., White, J. C., & Wulder, M. A. (2022). Mapping the presence and distribution of tree species in Canada's forested ecosystems. *Remote Sensing of Environment*, 282, 113276. <https://doi.org/10.1016/j.rse.2022.113276>
- Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, 9, 95.
- Kemper, T., Mudau, N., Mangara, P., & Pesaresi, M. (2015). *Towards a country-wide mapping & monitoring of formal and informal settlements in South Africa*. Ispra: Publications Office of the European Union.
- Langner, A. and Carboni, S., 2021, July. Forest Degradation Derived by a Newly Developed Sentinel-1 Change Detection Approach. In 2021 *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, (pp. 1019-1022), Brussels, Belgium, DOI: 10.1109/IGARSS47720.2021.9554574
- McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), pp.1425-1432.

- Pekel, J.F., Cottam, A., Gorelick, N. and Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), pp.418-422.
- Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* (2019), 11, 523. <https://doi.org/10.3390/rs11050523>
- Persson, M., Lindberg, E., & Reese, H. (2018). Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 10(11), 1794. <https://doi.org/10.3390/rs10111794>
- Pesaresi, M., G. Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, M. A. Marin-Herrera, G. K. Ouzounis, M. Scavazzon, P. Soille, V. Syrris & L. Zanchetta (2013) A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6, 2102-2131.
- Pesaresi, M., Gerhardinger, A., & Kayitakire, F. (2008, September). A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3), 180-192.
- Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C. and Defourny, P. (2014). Automated training sample extraction for global land cover mapping. *Remote Sensing*, 6, 3965-3987.
- Sabo, F., Corban, C., Politis, P. and Kemper, T., The European Settlement Map 2019 release, EUR 29886 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-12184-8 (online), 978-92-76-12183-1 (print), doi:10.2760/25824 (online), 10.2760/979189 (print), JRC118076.
- Schulz, C., Holtgrave, A. K., & Kleinschmit, B. (2021). Large-scale winter catch crop monitoring with Sentinel-2 time series and Machine Learning—An alternative to on-site controls?. *Computers and Electronics in Agriculture*, 186, 106173.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson.
- Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2023.113888>
- Vázquez-Jiménez, R., Ramos-Bernal, R. N., Romero-Calcerrada, R., Arrogante-Funes, P., Tizapa, S. S., & Novillo, C. J. (2018). Thresholding algorithm optimization for change detection to satellite imagery. In *InTech eBooks*. <https://doi.org/10.5772/intechopen.71002>
- Viana-Soto, A., & Senf, C. (2024). European Forest Disturbance Atlas (2.1.0.) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.10885507>

- Wang, X., Xie, S., Zhang, X., Chen, C., Guo, H., Du, J. and Duan, Z., 2018. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 68, pp.73-91.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S. and Bates, P.D., 2017. A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, 44(11), pp.5844-5853.
- Zanaga, D. , Van De Kerchove, R. , Daems, D. , De Keersmaecker, W. , Brockmann, C. , Kirches, G. , Wevers, J. , Cartus, O. S. , Lesiv, M. , Herold, M., Tsendbazar, N.-E. , Xu, P. , Ramoino, F. , & Arino, O. (2022). *ESA WorldCover 10 m 2021 v200 (Version v200)*.



EvoLand
LAND MONITORING EVOLUTION

EvoLand Consortium

