



**SEN4LDN**

LAND DEGRADATION NEUTRALITY

## Algorithm Theoretical Baseline Document

Version 1.2  
16 January 2025

Authors: Carolien Toté, Daniele Zanaga, Giorgia Milli, Lars Eklundh,  
Zhanzhang Cai, Martin Herold, Katja Berger, Ruben Van De Kerchove

submitted by



CONSERVATION  
INTERNATIONAL



**Prepared under contract from the European Space Agency**

Contract No. 4000138770/22/I-DT

Project acronym:	SEN4LDN
Project full title:	Sentinel high resolution data to improve land degradation neutrality monitoring
Project duration:	01.10.2022 – 30.09.2024 (24 months)
Project coordinator:	Dr. Ruben Van De Kerchove, VITO Remote Sensing
Deliverable title:	Algorithm Theoretical Baseline Document
Deliverable n°:	D3.2
WP responsible:	Dr. Ruben Van De Kerchove, VITO Remote Sensing
Nature of the deliverable:	Document
Dissemination level:	Public
Lead partner:	VITO Remote Sensing, Belgium
Recommended citation:	Toté C., Zanaga D., Milli G., Eklundh L., Cai Z., Herold M., Berger K. & Van De Kerchove R. (2024). <b>Algorithm Theoretical Baseline Document</b> . SEN4LDN project deliverable D3.2.

**Deliverable status:**

Version	Status	Date	Author(s)
1.0	First version	2 September 2024	Carolien Toté (VITO), Daniele Zanaga (VITO), Giorgia Milli (VITO), Lars Eklundh (Lund University), Zhanzhang Cai (Lund University), Martin Herold (GFZ), Katja Berger (GFZ), Ruben Van De Kerchove (VITO)
1.1	Adaptations on algorithm improvements and output products definition	16 December 2024	
1.2	Minor corrections	16 January 2025	

## Table of contents

Key takeaway messages.....	9
List of abbreviations.....	10
1 Introduction .....	12
1.1 Scope and objectives.....	12
1.2 Document structure .....	12
1.3 Related documents .....	13
2 Review of user requirements.....	14
2.1 User requirements for trends in land cover.....	14
2.2 User requirements for trends in land productivity .....	15
2.3 User requirements for trends in carbon stocks .....	16
3 General concept.....	19
4 Trends in land cover.....	20
4.1 Outline.....	20
4.2 Background .....	20
4.3 Algorithm .....	21
4.3.1 Summary .....	21
4.3.2 Algorithm input variables .....	23
4.3.3 Algorithm description.....	25
4.3.3.1 Pre-processing .....	25
4.3.3.2 Deep Learning Model.....	28
4.3.3.3 CatBoost Model and Benchmarking.....	30
4.3.3.4 Post-processing.....	33
4.3.3.5 Land cover transitions.....	37
4.3.4 Algorithm output variables.....	44
4.3.4.1 Output products.....	44
4.3.4.2 Land Cover Maps (LCM) .....	45
4.3.4.3 Land Cover Transitions (LCT).....	46
4.3.4.4 Land Cover Degradation (LCD) products.....	46
4.4 Summary of algorithm validation.....	46
4.5 Conclusions, limitations and recommendations .....	47
5 Trends in land productivity .....	49
5.1 Outline.....	49
5.2 Background .....	49
5.3 Algorithm .....	51
5.3.1 Summary.....	51
5.3.2 Algorithm input variables .....	51
5.3.2.1 2-band Enhanced Vegetation Index (EVI2).....	51
5.3.2.2 Land cover.....	52

5.3.3	Algorithm description .....	52
5.3.3.1	Data processing using TIMESAT .....	52
5.3.3.2	Trend estimation .....	55
5.3.3.3	Performance estimation .....	56
5.3.3.4	Land Productivity Degradation .....	58
5.3.4	Algorithm output variables .....	59
5.4	Summary of algorithm validation .....	60
5.5	Conclusions, limitations and recommendations .....	61
5.5.1	Conclusions .....	61
5.5.2	Constraints/limitations .....	61
5.5.3	Recommendations .....	62
6	Trends in carbon stocks .....	63
6.1	Outline .....	63
6.2	Background .....	63
6.3	Algorithm .....	64
6.3.1	Summary .....	64
6.3.2	Algorithm input variables .....	64
6.3.3	Algorithm description .....	65
6.3.3.1	Stock change approach .....	65
6.3.3.2	Gain-loss approach .....	65
6.3.3.3	Postprocessing: hybrid method .....	66
6.3.4	Algorithm output variables .....	66
6.4	Summary of algorithm validation .....	67
6.5	Conclusions, limitations and recommendations .....	68
7	Integration .....	70
7.1	Outline .....	70
7.2	10AO integration process .....	70
7.3	Continuous sub-indicator integration method .....	72
7.4	Algorithm output variables .....	73
7.5	Conclusions, limitations and recommendations .....	73
	References .....	74
Annex A.	CatBoost model benchmarking .....	79
Annex B.	Summary of tested vegetation indices .....	84
	LAI and FAPAR .....	84
	NDVI .....	85
	PPI .....	85
Annex C.	Benchmarking of vegetation indices .....	87

## List of figures

Figure 1: General concept of the SEN4LDN algorithm to feed into SDG indicator 15.3.1 .....	19
Figure 2: Overview of the workflow for the trends in land cover sub-indicator .....	22
Figure 3: Summarized scheme of the Land Cover classification model.....	23
Figure 4: Training data for LC mapping: Locations density map showing the spatial stratification. Areas with heterogeneous land cover are sampled more to better capture the signal variability.....	24
Figure 5: Global training dataset: 30 000 locations stratified over space and land cover heterogeneity. For each point location, the colour is assigned in the visualization based on the dominant cover fraction for that patch.....	25
Figure 6: AgERA5 climatic regions embeddings for the year 2020 (displaying only the first 3 variables in an RGB composite for illustration purpose).....	26
Figure 7: Sample observation over tile 45QXE. a) RGB image, b) SCL cloud mask (custom colours), c) SEN4LDN cloud mask, showing better detection over the haze covering the scene. ....	27
Figure 8: Yearly RGB median for a patch (1km x 1km) over Colombia. a) RGB median computed using the SCL cloud masks for the yearly time-series: undetected clouds cause artefacts in the obtained image. b) The same data is masked using the merged SCL and SEN4LDN cloud masks, obtaining an improved yearly median image with reduced residual cloud artefacts. ....	27
Figure 9: Summarizing scheme of the L2A pre-processing steps yielding the temporal features.....	28
Figure 10: Confusion matrix of the deep learning model. It is to be noted that the algorithm is not able to learn to predict snow. This is due to the lower number of samples in the training data set. ....	29
Figure 11: Spatial accuracy map of the deep learning algorithm showing good performance in the northern hemisphere and homogeneous landscapes, while showing poorer performance in mixed landscapes, for example central Africa and in the Australian xeric shrublands deserts. This can be explained considering two aspects. Firstly, these are challenging areas, and secondly the training data is produced synthetically from ESA WorldCover which itself shows lower accuracy in these problematic areas. ....	29
Figure 12: Sampling probability based on H3 grid cell class fraction. The less pixels of a certain class are present in a patch, the higher the number of sampled pixels.....	31
Figure 13: Class distribution of the pixel training dataset obtained through the spatially balanced sampling described above. ....	31
Figure 14: Deep learning-based change detection model predictions for two areas in Portugal. Changes are detected partially. The algorithm shows low sensitivity.....	33
Figure 15: In the figure, an area of 2.5km * 2.5km is displayed over Portugal. The top row are the yearly RGB median composites for the 6 years 2018-2023. The second row are the predictions obtained from the Land Cover classifier and the third row are the predictions obtained by the post-processed predictions. The example highlights the stabilizing effect of the post-processing. In the absence of change, the raw predictions show higher shrubland/grassland fluctuations, which are reduced in the post-processed predictions, which appear more stable. ....	35
Figure 16: In this example from Portugal, the post-processing helps in stabilizing the predictions while preserving well the changes. For example, shrubland distribution between 2018 and 2019 is significantly more stable in the post-processed version. At the same time the deforestation event on the bottom left is captured and is not affected by the post-processing. ....	36
Figure 17: In this example from Uganda, we can observe a limitation of the classification algorithm. Residual clouds are present in 2019 and 2022. These residual thin clouds, which are not filtered out correctly by the pre-processing step, alter the features and lead to the spurious commission of cropland. The post-processing in this case slightly	

reduces the commission, however it is not able to override errors in the classification, as these depends on issues in the input data and are considered as ‘real changes’ from the perspective of the post-processing algorithm.....	36
Figure 18: Transitions probabilities workflow .....	39
Figure 19: Example of transition maps generated from the transition probabilities. The first row of images shows the RGB medians and Land Cover maps, and the transition maps obtained from the discrete land cover maps. In the second row it is presented a series of transition maps obtained with different thresholds on the transition probability, starting at 0.3 and finishing at 0.7. Although logically 0.5 would represent the default cutoff threshold, it was observed that 0.4 would be a more conservative choice, allowing to preserve more transitions. The last row shows a comparison of the prediction-based transition map and the transition map obtained from the transition probabilities. In red the pixels which are not considered transitions, showing how the method significantly reduces noisy transition pixels, and preserves the actual transitions. ....	40
Figure 20: Similarly to the previous figure, another example of transition maps obtained from the transition probabilities, highlighting the reduction of noise. The threshold chosen for the final maps is 0.4. ....	41
Figure 21: Example of a Degradation/Improvement probability (LCD-PROB). Providing a continuous probability enables users to eventually customize their results or use the probabilities as input for downstream applications.	42
Figure 22: Discretizing the Land Cover Degradation probabilities yields the Land Cover Degradation (LCD) (bottom right) for the area of interest.....	42
Figure 23: Land Cover Transition and Degradation probabilities for another nearby location.....	43
Figure 24: Transition and LCD maps for the same location, as well as the Land Cover maps. In this case the Land Cover maps are shown as well to highlight that although in the central patch there is a larger area predicted as cropland, this does not end up in the transition/degradation maps. The algorithm is not providing a high confidence prediction of cropland against grassland and the consequent transition probabilities are low enough to be able to discard this as a spurious change.....	43
Figure 25: Productivity for three years of data illustrated with the light blue areas. The grey points are vegetation index data from Sentinel-2 observations, and the blue line is the smooth seasonal vegetation curve estimated with a cubic spline function. ....	50
Figure 26: Overview of the workflow for the trends in land productivity sub-indicator.....	51
Figure 27: TIMESAT4 processing module in SEN4LDN.....	53
Figure 28: Time-series across six years of data. The blue line is the seasonal trajectory (ST), and the raw data values (grey points) represent EVI2 computed from Sentinel-2 TOC reflectance. Red arrows represent seasonal start and end dates. ....	54
Figure 29: Definition of productivity parameters SOS (start of season), EOS (end of season), and TPROD (total productivity). ....	55
Figure 30: Workflow of the trend estimation module.....	56
Figure 31: Workflow of the performance estimation module .....	57
Figure 32: Workflow for Land Productivity Degradation (LPD) class calculation. ....	58
Figure 33: Workflow for Land Productivity Degradation (LPD) value calculation .....	59
Figure 34 Overview of the workflow for the trends in carbon stocks sub-indicator .....	64
Figure 35: Example of AGB change mapping using ESA CCI dataset from 2010 as baseline and calculating change to 2018. Adapted from [52]. ....	65
Figure 36: Example of hybrid AGB change mapping method 2010-2018, Portugal case study. ....	66

Figure 37: $\Delta$ AGB comparisons between the 100 m map products (CCI and Flux), and the LiDAR and NFI reference data (left) and comparisons of 25 km $\Delta$ AGB aggregation between the map products and the three types of reference data [52].	68
Figure 38: Example on area of 2.5km x 2.5km in Portugal. (a, b) Sentinel-2 RGB Median composite for years 2018 and 2023. (c, d) Land Cover maps 2018, 2023. (e) Land Cover Transition map (LCT). (f) Land Cover Degradation map (LCD). (g) Land Productivity Degradation map (LPD). (h) Land Degradation indicator obtained through integration of LCD and LPD with the IOAO method, as in Table 15. The indicator highlights 'Improved' areas in green, 'Degraded areas in purple, and 'Stable' areas in light grey.	71
Figure 39: Example of continuous sub-indicator integration. Top left: LPDval. Top right: LCD-PROB. Bottom left: simple average. Bottom right: maximum absolute value.	72
Figure 40: Class distribution of the pixel training dataset obtained through the spatially balanced sampling described above.	79
Figure 41: Confusion matrices of the basic model using only L2A features and the candidate model with all the selected features. Accuracy of the s2 model (left) is 69%, while accuracy of the candidate model (right) is 78.9%.	80
Figure 42: Spatial accuracy map for the base 's2' model (top) and for the final model (bottom), showing improved global accuracy.	80
Figure 43: Models predictions for 2021 for test site 16TDM_093_48. The candidate model "s2_meteo_unet_cbprob_dem_latlon" shows the to match well with WorldCover 2021 (last row, WC2021). The LSC image in the last row is an RGB composite of the 3 "cprob" features, with red, green and blue being respectively 'not vegetated', 'vegetated' and 'water'. Adding the unet probabilities reduces salt and pepper noise and helps removing commission of wetland.	81
Figure 44: Models predictions for 2021 for test site 42QQM_002_21. The candidate model "s2_meteo_unet_cbprob_dem_latlon" shows a good match with WorldCover 2021, as well as a significant reduction of built-up commission thanks to the introduction of the deep learning features.	82
Figure 45: Comparison of the reference model with and without 'unet' features. The location does not present land cover changes. However, phenological fluctuations can be observed in the yearly RGB medians. The model is stable against these fluctuations. The presence of deep learning features improves the classification of built-up class. The road in the upper left corner of the patch is almost always omitted in the model without deep learning features, while correctly classified in the model with them.	82
Figure 46: Example of a patch with actual land cover changes. During the years water expansion is correctly classified. Presence of the deep learning features reduce wetland commission.	83
Figure 47: Example showing the improved stability against phenological fluctuations which can cause confusion between shrublands and grasslands in heterogeneous landscapes.	83

## List of tables

Table 1: Summary of user requirements by the SDG Early adopters (derived from SEN4LDN D1.2 Requirement Baseline). Where relevant, distinction is made between specific requirements for Uganda (U), Colombia (C) or Portugal (P). The requirements marked in green are addressed by SEN4LDN.	18
Table 2: Summary of feature groups that are used in different combinations for model training	32
Table 3: General land cover transition matrix using the 11 land cover classes (see §4.3.4.2). Land cover change processes are colour coded as positive <b>P</b> , stable <b>S</b> or negative <b>N</b> . Unlikely transitions are put within brackets. Note that this is the general starting point, but that specific transitions can be re-evaluated per use case, to adopt to local conditions.	37

Table 4: Definition, colour and map code of the Land cover transitions layer. Land cover transitions (degradation and improvement processes) are based on the land cover transition matrix (see Table 3). .....	38
Table 5: Key degradation and improvement processes for SEN4LDN SDG Early adopters, likely to be detectable based on trends in land cover .....	38
Table 6: Output products of trends in land cover algorithms .....	44
Table 7: Metadata fields of the output COG products .....	44
Table 8: Coding of the LCM layer and definition of the classes .....	45
Table 9: LCD classes .....	46
Table 10: Classification criteria for assigning pixels to slope classes .....	56
Table 11: Classification criteria for productivity performance (PP) .....	57
Table 12: Lookup table for codes defining land productivity degradation classes .....	58
Table 13: Datasets of land productivity .....	60
Table 14: Output products of trends in carbon stocks .....	67
Table 15: Combination of land cover degradation and land productivity degradation sub-indicators using the 10AO integration method .....	71
Table 16: SEN4LDN products on the LDN indicator .....	73
Table 17: Correlation coefficients ( $r$ ) between Gross Primary Productivity and smoothed vegetation index data per land cover class .....	87



## Key takeaway messages

- The document is the Algorithm Theoretical Baseline Document (ATBD) for the SEN4LDN project, prepared under contract from the European Space Agency. The project aims to improve land degradation neutrality (LDN) monitoring using high-resolution Sentinel data. The document outlines the development and validation of algorithms for monitoring trends in land cover, land productivity, and carbon stocks.
- The development of the land cover classification system within the framework of the SEN4LDN project, to create trends in land cover output products, marks a significant advancement over the previous ESA WorldCover system, particularly in ensuring consistency and accuracy across multiple years. The algorithm uses Sentinel-2 data, deep learning, and a CatBoost model, with post-processing to reduce noise and improve stability. Output products include annual land cover maps, transition maps, and land cover degradation maps.
- The productivity trend sub-indicators in SEN4LDN meet the user requirements with respect to spatial and temporal resolutions, spatial coverage, output reference system, delivery mode, data format, and free and open access. The algorithm uses Sentinel-2 data and TIMESAT software to derive productivity parameters like the 2-band Enhanced Vegetation Index (EVI2). It estimates trends and performance, combining them into a Land Productivity Degradation (LPD) indicator. Output products include trend values, performance values, and LPD maps.
- Due to the ongoing absence of high-resolution global trend data of SOC derived from EO data, within SEN4LDN the above-ground biomass (AGB) mapping using ESA CCI and WRI flux models was explored. The algorithm combines stock change and gain-loss approaches to create hybrid AGB change maps.
- Two methods are tested for integrating sub-indicators into a single land degradation indicator: the one-out-all-out (1OAO) method and a continuous sub-indicator integration method. The latter provides a more nuanced interpretation of degradation probabilities.
- The document also includes detailed validation results, highlighting the strengths and limitations of the algorithms, and recommendations for future improvements.

## List of abbreviations

ATBD	Algorithm Theoretical Baseline Document
BIOPAR	Biophysical Parameters
C	Colombia
CCDC	Continuous Change Detection and Classification
CCI	Climate Change Initiative
CLMS	Copernicus Land Monitoring Service
DEM	Digital Elevation Model
ESA	European Space Agency
EU	European Union
EO	Earth Observation
EVI2	2-band Enhanced Vegetation Index
FAO	Food and Agriculture Organization
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FRA	Forest Resource Assessment
GLC	Global Land Cover
GPG	Good Practice Guidance
H3	Hexagonal Hierarchical Geospatial Indexing System
HR	High Resolution
HR-VPP	High Resolution Vegetation Phenology and Productivity
LAI	Leaf Area Index
LC	Land Cover
LCCS	Land Cover Classification System
LCD	Land Cover Degradation
LCT	Land Cover Transition
LD	Land Degradation
LDN	Land Degradation Neutrality
LPD	Land Productivity Degradation
MLP	Multi-layer perceptron
MSWEP	Multi-Source Weighted-Ensemble Precipitation
NDVI	Normalized Difference Vegetation Index
NFI	National Forest Inventory
OLS	Ordinary Least-Squares
P	Portugal
PP	Productivity Performance
PDR	Preliminary Design Review
PPI	Plant Phenology Index
PROB	Probabilities
QFLAG	Quality Flag
RMSE	Root Mean Squared Error
RTM	Radiative Transfer Model
SAR	
SCL	Scene Classification Layer
SDG	Sustainable Development Goal
SEN4LDN	Sentinels for Land Degradation Neutrality
TPROD	Total Productivity
U	Uganda
UMAP	Uniform Manifold Approximation and Projection

UN	United Nations
UNCCD	United Nations Convention to Combat Desertification
VI	Vegetation Indicator
WP	Work Package
WRI	World Resources Institute

# 1 Introduction

## 1.1 Scope and objectives

The 2030 Agenda for Sustainable Development is fundamentally based on 17 Sustainable Development Goals (SDG) which are targets agreed upon by the UN members regarding various interlinked objectives that must be ensured to achieve sustainable development. These range from combating poverty, ensuring access to education, to economic development and the protection of life on water and land.

Diminished overall productivity and reduced resilience in the face of climate and environmental change, have made addressing land degradation a global priority formalized by the United Nations Convention to Combat Desertification (UNCCD) and the SDG. To this end, the 2030 Agenda for Sustainable Development defined target 15.3 of SDG 15, called '*Life on Land*', that strives to reach Land Degradation Neutrality (LDN) by 2030.

Efficient monitoring of LD requires constant monitoring of various biophysical and biochemical characteristics of the land. These disturbances can range from rapid land cover change (e.g., fire or logging) to continuous and slower degradation of soil and land quality [1]. While monitoring these at larger scale becomes a logistical impossibility if not using Earth Observation (EO) data, there are still a number of challenges and opportunities to address particularly related with increasing spatial and temporal resolution and diversity of sensor types [2]. Sentinels for Land Degradation Neutrality (SEN4LDN) aims to address these two limitations by developing and showcasing a novel approach for improving both the spatial and temporal resolution of the data required for LD monitoring. While LDN is agreed between the SDG signatories, each region/country will have its own specific challenges and drivers of LD and therefore the inclusion of local partners in the product development is extremely important. These stakeholders will provide insights on the user requirements and feedback on the final product and its actual usability for SGD 15.3.1 reporting.

The objective of SEN4LDN Work Package 2 (WP2) is to select several candidate algorithms for land cover change monitoring and land productivity dynamics and benchmark these algorithms on the test sites. A quantitative and qualitative cross-comparison of best candidate methods was performed on a subset of the Phase 1 test sites and presented at the Preliminary Design Review (PDR) (November/2023) and after further tests and improvements at the Critical Design Review (CDR) (May/2024).

The definitive versions of the SDG algorithms are specified in this Algorithm Theoretical Baseline Document (ATBD; D3.2) and will be presented at the Production Readiness Review (PRR) (September/2024). The ATBD as such provides a detailed specification of all the processing algorithms of the EO solution developed by the project.

## 1.2 Document structure

The document is structured as follows:

- Chapter 2 provides a summary and review of the user requirements.
- Chapter 3 describes the general concept of the SDG algorithms and how these relate to each other.
- The algorithm to monitor trends in land cover is described in Chapter 4 .

- The algorithm to monitor land productivity dynamics is described in Chapter 5.
- Chapter 6 describes the information access on carbon stocks.
- Chapter 7 describes how algorithms can be integrated.

### 1.3 Related documents

- D1.2 Requirements Baseline
- D2.1 Input Data Inventory
- D3.1 Proof of Concept results on test sites
- D5.2 Product Validation Report

Public deliverables are available on <https://esa-sen4ldn.org/en/deliverables>.

## 2 Review of user requirements

SEN4LDN D1.2 ‘Requirements baseline’ outlines the key user requirements for the SEN4LDN project based on a set of interviews that were carried out with our early adopters:

- Busitema University (Uganda),
- Ministry of Environment and Sustainable Development (Colombia), and
- Institute for Conservation of Nature and Forest (Portugal).

The interviews were conducted as part of the SEN4LDN Living Labs approach to product development which aims to provide continuous interaction with potential users of the data products that are to be developed within SEN4LDN. This first session focused on

- (i) identifying what are the drivers;
- (ii) what are the efforts being done in each country regarding Land Degradation; and
- (iii) an overview of specific user requirements regarding Remote Sensing data.

Table 1 summarizes the SDG early adopters’ user requirements and provides an overview of what would provide an ideal solution for the early adopter. The first living labs provided valuable insights into the unique challenges and data requirements of each country. Not all the requirements formulated by the early adopters are to be or can be addressed by SEN4LDN. Indeed, the fact that SEN4LDN works towards a global solution somehow puts constraints on local performance. Nevertheless, a major requirement expressed by all early adopters, is the improvement of spatial and temporal resolution, which would greatly enhance their ability to implement and monitor LDN policies.

The following paragraphs summarize the user requirements for trends in land cover, trends in land productivity and trends in carbon stocks.

### 2.1 User requirements for trends in land cover

The intention of the SDG 15.3.1 land cover sub-indicator is to identify where degradation has occurred specifically as a result of land cover change. Land cover (LC) as defined by the FAO is the “observed (bio)physical cover of the earth’s surface” and is a synthesis of the many processes taking place on the land” [3]. Land cover is typically described as a set of hierarchical classes, each denoting the dominant biotic and abiotic assemblages occupying the Earths’ surface [4].

Land degradation for the SDG 15.3 target is defined as ‘the reduction or loss of the biological or economic productivity and complexity of rain fed cropland, irrigated cropland, or range, pasture, forest and woodlands resulting from a combination of pressures, including land use and management practices’. The objective of the land cover sub-indicator is to identify where degradation has occurred by detecting trends in land cover. However, the discrete representation of land cover and monitoring of shifts from one class to another provides information on land cover conversions, thereby neglecting more subtle land cover or land use modifications [5].

The SDG early adopters have identified key degradation processes that should be included in the country’s assessment of land degradation/improvement. For all three early adopters, these key processes include deforestation or forest cover loss. Uganda has a special interest for wetland cover, conversion to tree crop

plantations (palm oil, coffee, sugar cane, tea, cacao, banana), overgrazing (incl. transboundary movement of livestock) and urban growth. In Colombia, monitoring of forest restoration over pasture (e.g. silvo-pasture systems) is essential, as well as conservation of dry forests. The use case in Portugal should focus on overall transitions between land cover types, the impact of drought, soil erosion, and invasive species. Both Uganda and Colombia mention degradation through (illegal) mineral exploration. Both Colombia and Portugal want to focus on intensive agriculture and (forest loss by and forest restoration or recovery after) forest fires. In all three cases, a distinction between natural tree cover and tree crops is desired.

In §4.3.2 we describe in detail to what extent the land cover legend classes allow monitoring of the degradation/improvement processes that were identified in the user requirements baseline. For some processes, the recognition is not straightforward, e.g. conversion to tree crop plantations (possibly labelled as vegetation establishment or afforestation). During the second round of living labs, in which a first version of the prototype products were presented to the early adopters, this (lack of) distinction between natural tree cover and tree crops was brought up several times. In the Portugal use case, the conversion from natural vegetation (e.g. shrubland) to non-native tree cover is non-desirable from a nature conservation point of view, as it also increases the susceptibility to forest fires. For Colombia, the distinction between natural grasslands and managed pastures is desired.

Concerning the overall user requirements on the land cover sub-indicator, SEN4LDN will be able to respond to most of the minimum requirements (see Table 1): spatial resolution up to 10x10 m<sup>2</sup>, annual temporal resolution, national spatial coverage, analysis-ready maps of standard land cover classes available in user specified CRS, openly available through FTP and/or a web service, with related documentation. However, since Sentinel-2 data is used as primary input, the temporal extent will start in 2018, which serves as the baseline period.

## 2.2 User requirements for trends in land productivity

Project requirements following SDG 15.3.1 with respect to land productivity as stated in RQ-C3 is “... to develop robust automated EO method to monitor land productivity dynamics at high spatial resolution (10-30m) for large-scale LDN applications (robust phenological analysis of vegetation parameters for LDN)”. This information is then, together with information from land cover change data, input to a “reliable and cost-effective land degradation monitoring system at high spatial resolution...” (RQ-C5). The methodology for addressing the requirements uses the UNCCD Good Practice Guidance (GPG) [6] as a springboard, while taking experiences from the first implementations in Trends.Earth and results from the discussions and views of the early adopters into account.

All early adopters mentioned the importance of the high spatial resolution data from Sentinel-2, to simplify the attribution of processes to specific land categories.

In Uganda, land productivity is assessed by using the Trends.Earth platform based on NDVI time series data. The early adopters from Uganda added the requirement to be able to link land productivity with specific agriculture lands, crops and grazing patterns. They also pointed to the fact that large parts of Mt. Elgon ended up in the “Degradation” class of the Performance sub-indicator. This brought up a discussion on the need to modify the terminology and the processing methodology of the Performance sub-indicator. Adding more information to clarify how the indicator is defined might also improve the interpretation.

In Portugal, land productivity for the last reporting period (2016-2019) is measured following the GPG and the PRAIS4 reporting manual. The three metrics of Trend, State and Performance are measured using annual Net Primary Productivity data from 2000 to 2019 and the guide's look-up tables. The early adopters from Portugal highlighted the need for higher temporal and spatial resolutions alongside no-cost data, and raised the question as to the possibility to use a harmonized Landsat/Sentinel data. The country suffers recurrent fire disturbance resulting in fire-scars may last for shorter or longer time periods. These affect the rather short time-series of Sentinel-2, affecting trend estimates. A discussion emerged concerning the challenge to harmonize national requirements with a global approach. Furthermore, the challenges in directly translating remotely sensed observations into degradation/improvement measures were discussed.

From the Colombian early adopter's team, it was clear that an improvement especially of the spatial resolution would provide improved possibilities for the Colombian government in terms of both understanding the dynamics of LD as well as to be able to improve their ability to implement policies and monitor their efficiency. Persistent cloud coverage of Colombian upland forest areas results in rather large areas of missing data complicating the analysis in these areas.

Overall, SEN4LDN will meet the desired user requirements with respect to spatial and temporal resolutions, spatial coverage, output reference system, delivery mode, data format, and free and open access. For temporal extent it will meet the minimum requirement (2018-). It is important to realize that the short time-series of Sentinel-2 available imposes rather severe limitations on the productivity trend and performance sub-indicators, further discussed in the respective sections below.

### 2.3 User requirements for trends in carbon stocks

In case of Uganda, carbon stocks is estimated using a combination of land cover and SOC data within the Trends.Earth application. The EA stated that if SEN4LDN can develop products based on AGB then it would be a significant improvement to current estimates, furthermore, plot data has been collected in permanent plots which could be used for the model development. Currently, SOC is also based on datasets with 250 m resolution and therefore there is a need for higher resolution data.

In the case of Colombia, carbon stocks monitoring is based on the FAO methodology. Maps are obtained from 4329 soil profiles sampled across the entire country between 1980 and 2012 using a kriging regression algorithm with a scale of 1:100,000. The EA stated that the inclusion of higher resolution EO imagery could therefore improve the resolution of this data and empower the decision-making process of the Colombian government regarding LDN.

In Portugal, carbon stocks assessment is based on the National Forest Inventory (IFN615) from 2015 which details Above Ground Biomass from both alive and dead biomass. This dataset is based on an extensive sampling throughout the country which providing an excellent coverage of the country. For SOC estimates for PRAIS4, Trends.Earth default data was used.

Overall the project can address the need for higher resolution data with the AGB change mapping approach (100m), although not reaching the desired 10x10m<sup>2</sup>. In addition, there is currently no global EO based capacity for soil organic carbon monitoring available. As replacement, AGB change maps (hybrid approach) could be provided, which was fine for Uganda and Portugal, but Colombia partners raised some doubts and emphasized the need for SOC data to assess degradation. Also, the required time frame, i.e.



temporal extent from 2000 onwards could not be provided, since the AGB change mapping starts in 2010. Furthermore, there is no yearly update but 8 years interval (2010-2018).

Table 1: Summary of user requirements by the SDG Early adopters (derived from SEN4LDN D1.2 Requirement Baseline). Where relevant, distinction is made between specific requirements for Uganda (U), Colombia (C) or Portugal (P). The requirements marked in green are addressed by SEN4LDN.

Requirement	Land Cover		Productivity		Carbon stocks	
	Minimum	Desired	Minimum	Desired	Minimum	Desired
Spatial resolution	U, C: 30x30 m <sup>2</sup> P: 10x10 m <sup>2</sup>	U, C: 10x10 m <sup>2</sup> P: < 10x10 m <sup>2</sup>	U: 30x30 m <sup>2</sup> C: 1x1 km <sup>2</sup> P: 500x500 m <sup>2</sup>	10x10 m <sup>2</sup>	U, P: 30x30 m <sup>2</sup> C: 100x100 m <sup>2</sup>	≤ 10x10 m <sup>2</sup>
Temporal resolution	U, P: Annual C: Trimester	U: 10 – day C: Weekly P: 3-6 months	Annual	Annual	Annual	Annual
Temporal Extent	U: 2018 – C, P: 2015 –	U, C: 2000 – P: 2013 –	U, P: 2018 – C: 2015 –	2000 -	U: 2018 – C, P: 2015 –	2000 -
Spatial coverage	National	U, P: National C: South America	National	U, P: National C: South America	National	U, P: National C: South America
CRS	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06	U: WGS84/UTM C: EPSG:9377 P: ETRS-89/PT-TM06
Specific requirements	Standard LCLU classes for SDG reporting	Minimum + U: Plantations, Forest C: Tropical Dry Forest, Savanah P: Increased detail in all classes	Standard VI data required for SDG reporting	10-day time series of VI	Standard SOC data product for SDG reporting	Standard SOC + Above-carbon data product (10 - day time series)
Processing level	Analysis-ready	Analysis-ready	Analysis-ready	Analysis-ready	Analysis-ready	Analysis-ready
Delivery-mode	FTP, Web-service	FTP, Web-service	FTP, Web-service	FTP, Web-service	FTP, Web-service	FTP, Web-service
Data Format	GeoTiff	GeoTiff	GeoTiff	GeoTiff	GeoTiff	GeoTiff
Documentation	ATB Document	ATB Document	ATB Document	ATB Document	ATB Document	ATB Document
Licensing	GNU General Public License / Open source	GNU General Public License / Open source	GNU General Public License / Open source	GNU General Public License / Open source	GNU General Public License / Open source	GNU General Public License / Open source
Users	Technical staff	Wider SDG community	Technical staff	Wider SDG community	Technical staff	Wider SDG community

### 3 General concept

The extent of land degradation for reporting on SDG Indicator 15.3.1 is in principle calculated using three sub-indicators [7]. The general concept of the workflow used in SEN4LDN to provide the necessary inputs to the indicator is illustrated in Figure 1. The input data consist of Sentinel-2 Level 2A products, ESA CCI Biomass products, Ancillary layers, and training data. The following chapters provide a detailed specification of the SEN4LDN processing algorithms for

- (i) Trends in land cover, yielding annual land cover maps, information on land cover transition and a land cover degradation product (see Chapter 4);
- (ii) Trends in land productivity, yielding information on land productivity trends, performance and a land productivity degradation product (see Chapter 5); and
- (iii) Trends in carbon stocks, yielding information on above ground biomass change (see Chapter 6).

Each chapter mentioned above also lists the output products that are generated. Since the trends in carbon stocks output products cover a different time frame and are provided at a different spatial resolution, the integration is performed on the trends in land cover and trends in land productivity sub-indicator products only. This is described in Chapter 7.

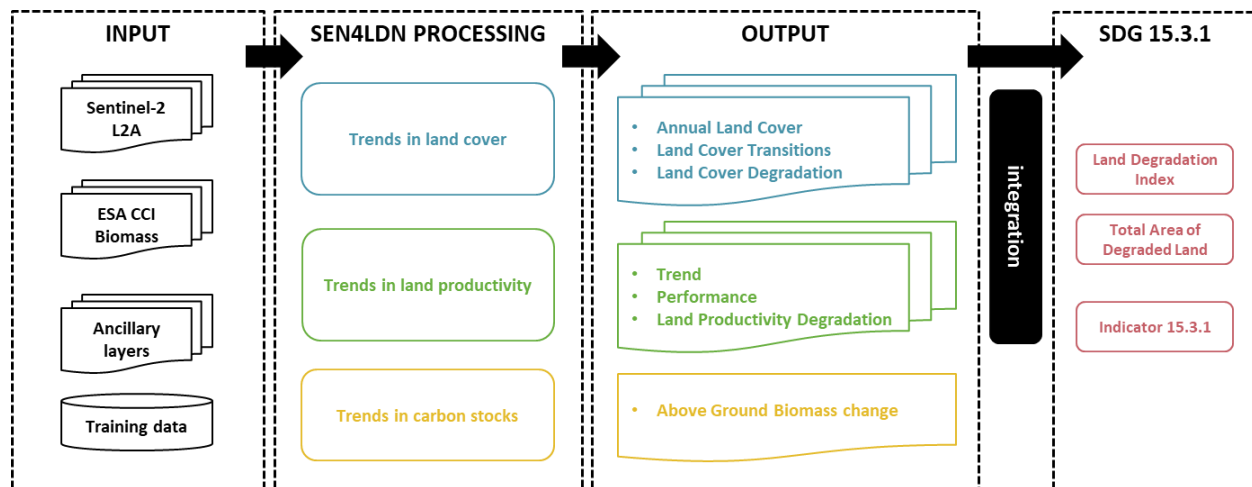


Figure 1: General concept of the SEN4LDN algorithm to feed into SDG indicator 15.3.1

## 4 Trends in land cover

### 4.1 Outline

The intention of the SDG 15.3.1 trends in land cover sub-indicator is to identify where degradation has occurred specifically because of land cover change. Within SEN4LDN the aim is to develop an automated method to monitor land cover dynamics at high spatial resolution, i.e. 10m, based on Sentinel-2 observations. This chapter describes the algorithm developed within SEN4LDN to monitor trends in land cover, including a detailed specification of the input data sources, processing steps, calibration data and resulting output products.

### 4.2 Background

The first satellite-based global land cover (GLC) map dates to 1994, was based on AVHRR data and had a spatial resolution of 1° [8]. Following these efforts, numerous global land cover maps were produced using medium resolution satellite data such as the GLC2000, MODIS, GlobCover, LC-CCI and GLCNMO [9], [10], [11], [12], [13]. In this respect, the European Space Agency played a key role establishing GlobCover 2005, GlobCover 2009, CCI-LC global LC maps per epoch (2000, 2005, 2010) and yearly (from 1992 to 2015). More recently, the yearly Copernicus Global Land Service Land Cover maps (CGLS-LC100) at 100 m resolution attained a high overall accuracy of 80% [14]. One of the main advantages of the CGLS-LC100 map, is the fact that it also presents cover fractions at the pixel level. This allows users to make their own interpretation of the discrete classes and to detect more subtle variability in land degradation.

Since the advent of the Copernicus programme and the Sentinels, the bar has raised, as it is now also possible to map land cover at 10m resolution, using 5-daily optical Sentinel-2 time series, possibly completed with Sentinel-1 Synthetic Aperture Radar (SAR) data. At first, two prototype maps at continental scale covering Africa (20m resolution) and Mesoamerica (10m resolution) were produced in the framework of the CCI Land Cover project (CCI Land Cover 2017). More recently, also first global 10m land cover products were released by ESRI<sup>1</sup> [15], Google<sup>2</sup> (i.e. Dynamic World, [16]) and in the framework of the ESA WorldCover project<sup>3</sup> [17]. The latter product has been produced by a consortium led by VITO and provides a new baseline global land cover product at 10m resolution for 2020 based on both Sentinel-1 and 2 data with 11 land cover classes. This map, which has been independently validated, attained an overall accuracy of 74.4%. Following positive feedback from users, ESA chose to expand on the WorldCover project by creating a new version of even higher quality for the year 2021 [18]. This version improved the product's global overall accuracy to 76.7%.

Despite the improved resolution of the WorldCover product and other 10 m global land cover products, so far global land cover mapping efforts based on HR Sentinel data are constrained to single maps, therefore lacking the change component crucial for LDN monitoring. In fact, to assess changes in land cover users need land cover maps covering the study area for the baseline (2000-2015) and target years.

---

<sup>1</sup> <https://www.esri.com/arcgis-blog/products/arcgis-living-atlas/imagery/global-land-cover-revealed/>

<sup>2</sup> <https://dynamicworld.app/>

<sup>3</sup> <https://esa-worldcover.org/en>

Such maps need to be of acceptable accuracy and created in such a way that allows for valid comparisons. Currently, the ESA CCI land cover maps (available at 300m resolution since 1992) are generally used as the default dataset (e.g., in Trends.Earth). One of the main challenges of moving towards HR datasets is therefore the integration with the SDG 15.3.1 baseline derived from moderate resolution satellite imagery acquired over the period 2000 to 2015.

Inconsistent spatial-temporal land cover change is a major problem when producing independent land cover maps for distinct time periods [19]. Small differences in image quality, the classification algorithm, and the definition of distinct classes may lead to inconsistencies [20]. In order to increase the quality of multi-temporal land cover maps, spatio-temporal cleaning is advised as a post-processing step. Several methodologies focus on discriminating real from false land cover changes. Examples are break-detection methods such as Breaks For Additive Seasonal and Trend (BFAST) [21] or the Continuous Change Detection and Classification (CCDC) algorithm [22].

Other change detection models are based on probabilistic models, such as the hidden Markov model or Markov random fields, which can take both spatial and temporal context into account [23]. Most of these algorithms focus on regions covering a limited area [19] are used to detect the dynamical character of one specific land cover class [24].

The original WorldCover classification algorithm was optimized to accurately map the land cover of a single year. These optimizations cannot be replicated dynamically for several years and therefore the algorithm is not suitable to produce consistent land cover time-series which are robust against errors between years. The consequent inability to discriminate between real and artificial (deriving from the algorithm) changes, or the omission of real changes, poses a limitation to direct application of the algorithm to land degradation detection.

Within SEN4LDN we aimed at extending the capabilities of the classification system to overcome this limitation by realizing an algorithm that achieves consistent and accurate land cover mapping across different years while retaining the same degree of spatial resolution and accuracy, enabling the detection of land degradation.

## 4.3 Algorithm

### 4.3.1 Summary

An overview of the workflow to monitor trends in land cover as a sub-indicator for land degradation is illustrated in Figure 2. The main input data source is Sentinel-2 L2A surface reflectances. In a pre-processing step, the Sentinel-2 time series are cleaned (including improved cloud detection), and a compositing algorithm is applied to extract features. Together with the training dataset and ancillary layers, this feeds into the land cover classification algorithm (LC model, see Figure 3). The land cover classification algorithm developed in the framework of the SEN4LDN project aims to extend the capabilities of the current state-of-the-art algorithms for global land cover classification recently released by ESRI, Google and VITO (see §4.2). The new system was designed as an evolution of the ESA WorldCover classification system, as this proved to be one of the best global land cover algorithms published in recent years [25], [26], [27].

The key limitation of ESA WorldCover algorithm is the poor generalization across different years. The development and calibration of the algorithm is performed on single years, and when the algorithm is applied to a different year, the prediction reliability is reduced, leading to errors. This limits the ability of using this algorithm for studying land cover changes and degradation events, and a solution needed to be designed.

The objective of the new design is to overcome this problem by adopting two key solutions. As a first step, the new models are trained on features computed from Sentinel-2 and AgERA5 data of multiple years (2018-2022). This ensures robustness against phenological fluctuations due to different meteorological conditions in different years. The second solution is to combine the advantages of single pixel classifiers with the generalization capabilities of deep learning semantic segmentation algorithms used for example by ESRI Land Cover and Google Dynamic World, thus preserving high spatial accuracy, but improving the generalization capabilities.

In total 11 land cover classes are predicted starting from yearly time-series of reflectance data for the Sentinel-2 L2A products bands. These 11 classes are considered adequate to allow monitoring of key degradation processes, although some of the land use changes might remain undetected (see also §2.1). A post-processing routine is applied to the LC probabilities obtained by the model predictions across different years, to stabilize them and reduce classification errors due to noise and phenological fluctuations of the input features. To determine changes for a given location, a combination of the LC probabilities is performed for each type of transition, obtaining continuous transition probabilities, which are then used to produce a discrete LC transition maps by means of a transition matrix. Land cover changes are specified as being either degradation, improvements, or neutral transitions, summarized in the Land Cover Degradation (LCD) output products.

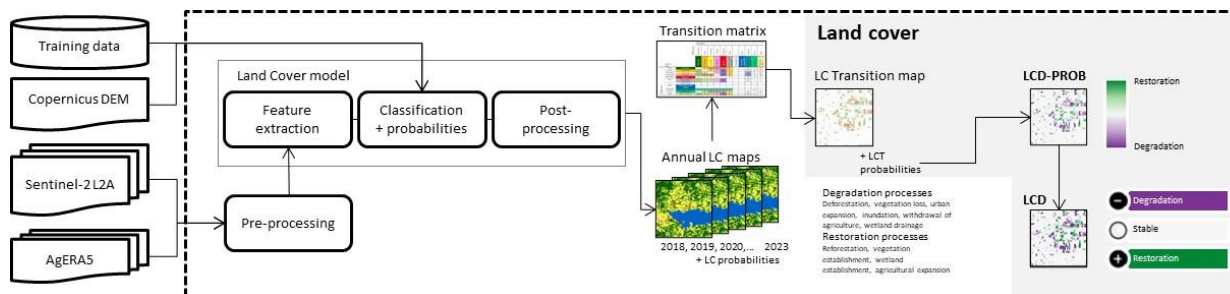


Figure 2: Overview of the workflow for the trends in land cover sub-indicator

The land cover classification workflow is summarized here and illustrated in Figure 3. Individual components are detailed in the sub-sections below. The algorithm starts from the input data which are used to produce features through a pre-processing step. The features are then used to train first a deep learning semantic segmentation algorithm. Due to technical constraints, only a subset of the computed features is used for the deep learning algorithm. The full set of features is then merged with the probabilities obtained from the deep learning model and fed into a CatBoost classifier. The output probabilities of the CatBoost classifier, computed for several years, are ingested in a post-processing

routine, which optimizes consistency of the predictions while preserving changes, yielding the final land cover time-series.

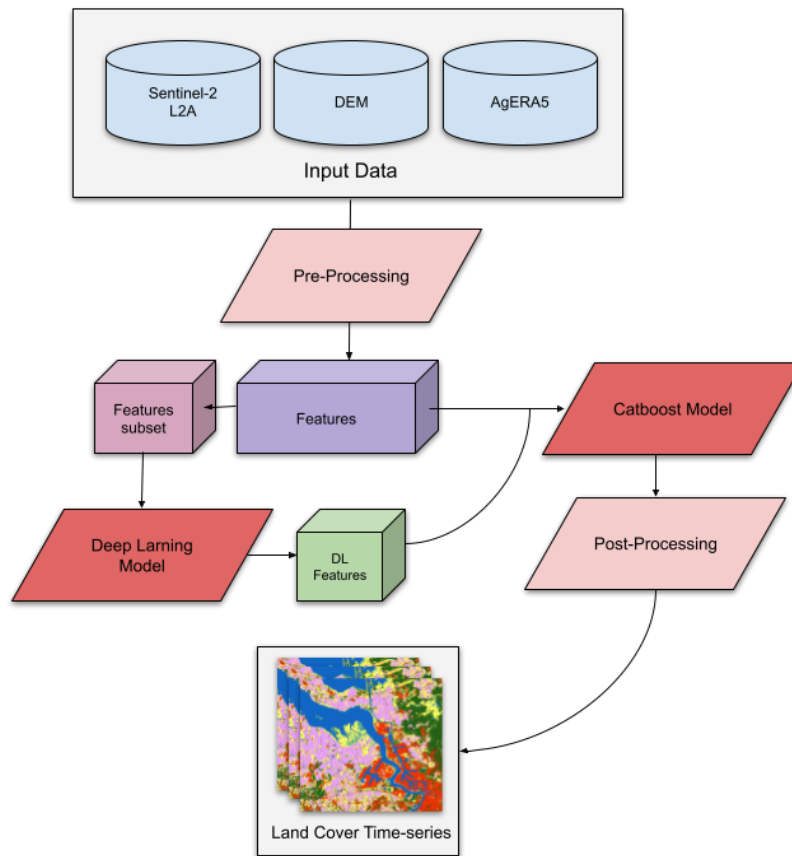


Figure 3: Summarized scheme of the Land Cover classification model

### 4.3.2 Algorithm input variables

The algorithm is based on the following input datasets:

- **Copernicus Sentinel-2 L2A:** The L2A data is downloaded/processed from the Copernicus Data Space Ecosystem. Quality checks are performed by ESA, resulting in quality flags in the product metadata. Upon ingestion of the products, products containing defective data are discarded.
- **Copernicus DEM** – Global and European Digital Elevation Model (COP-DEM<sup>4</sup>): The Copernicus GLO-30 DEM data, offering global coverage at a resolution of 30m, are obtained from the Copernicus Data Space Ecosystem, and resampled to 10m using bilinear interpolation. No quality checks are performed on the dataset.

<sup>4</sup> <https://doi.org/10.5270/ESA-c5d3d65>



- **Copernicus AgERA5 – Agrometeorological indicators:** daily aggregates of agronomic relevant variables [28]. The variables cover temperature, precipitation, snow depth, humidity, cloud cover and radiation.
- **Training data for LC mapping:** To train the land cover mapping neural network a global dataset of approximately 60,000 patches of 1280 x 1280 m<sup>2</sup> was produced. These patches were sampled using a sampling strategy that maximizes the density of sampling in heterogeneous areas and minimize it in homogeneous areas (Figure 4). For each Sentinel-2 tile over land, the tile area is split into 6400 patches. The patches are grouped together based on their land cover fractions (obtained from WorldCover 2021 (V200)). For each different group of cover fractions one sample is randomly selected. For example, for a homogeneous tile that is fully classified with 1 class, there will be only 1 sample patch extracted. For tiles with heterogeneous landscape, multiple patches will be selected based on the level of heterogeneity.

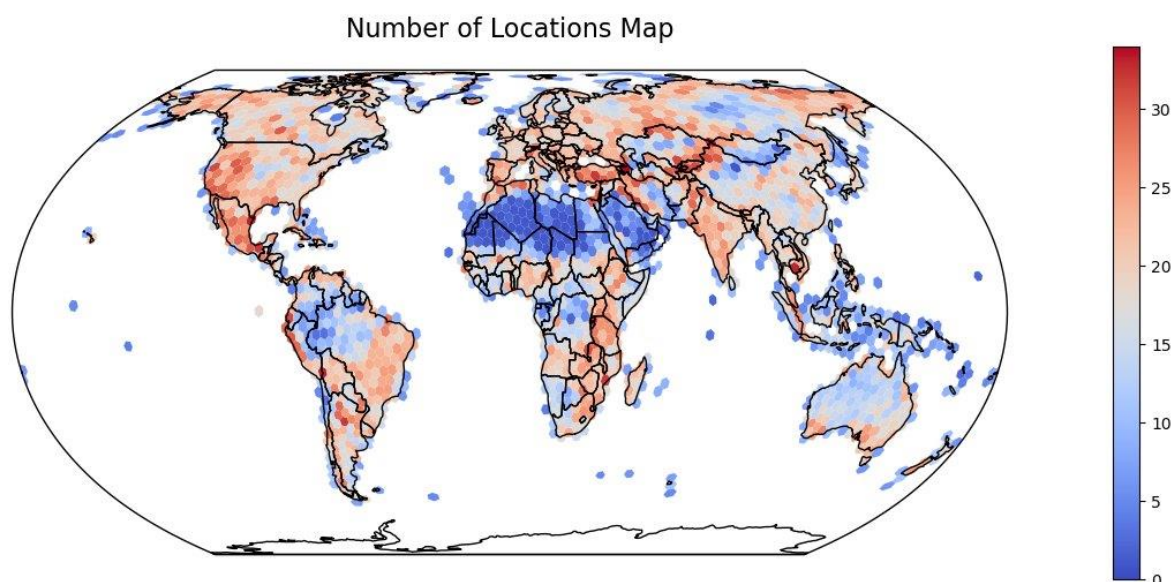


Figure 4: Training data for LC mapping: Locations density map showing the spatial stratification. Areas with heterogeneous land cover are sampled more to better capture the signal variability.

This dataset serves multiple purposes: (i) a class balanced subset of it (30,000 locations) is used for training AI models, using WorldCover 2021 as target annotations for the locations (Figure 5); (ii) a complementary independent subset (30,000 locations) is used for a global validation of the classification algorithm against WorldCover 2021 as well as to evaluate global performance of the land degradation metrics.

In practice, time-series for each patch are extracted from Microsoft Planetary Computer<sup>5</sup> for 5 years (2018-2022). Features are computed for each patch and year producing a multi-year training dataset. This leads to a training dataset of approximately 150,000 samples. However, for each location only the target annotation for 2021 is provided, as this is sampled from WorldCover 2021.

<sup>5</sup> <https://planetarycomputer.microsoft.com/>



Models training is performed using input features from all the available years and the single available target. Collecting a hand labelled dataset over multiple years was beyond the scope and resources of the current project, therefore it was needed to opt for this approximation, with the assumption that most locations will not be affected by changes. Inevitably, a certain amount of additional noise is introduced by this assumption, however, using input features from different years is essential to obtain a robust classification algorithm that can generalize well across different years, and the downsides of the approach are mitigated by the post-processing.

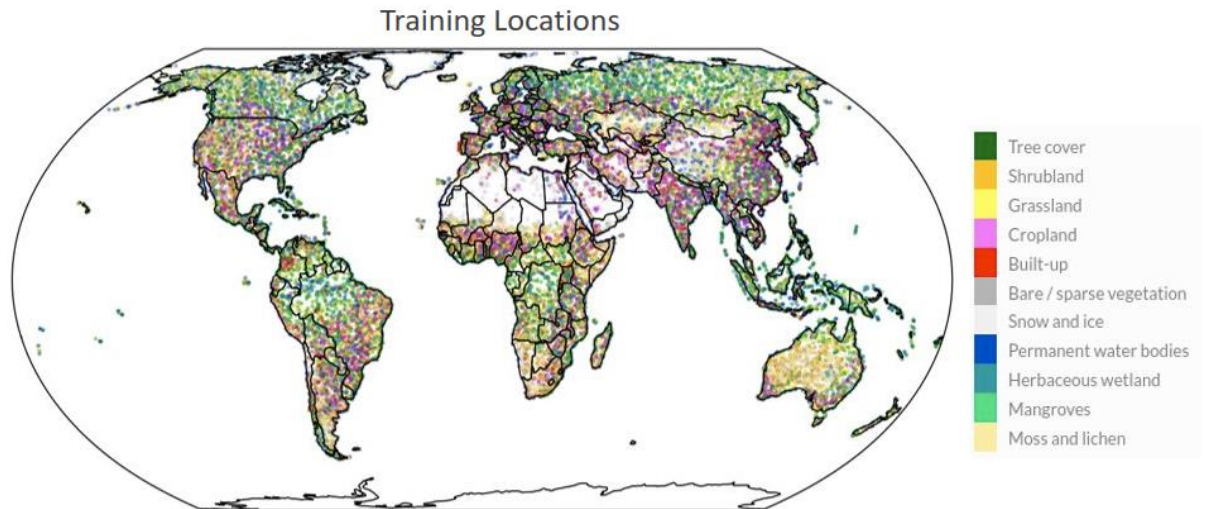


Figure 5: Global training dataset: 30 000 locations stratified over space and land cover heterogeneity. For each point location, the colour is assigned in the visualization based on the dominant cover fraction for that patch.

### 4.3.3 Algorithm description

#### 4.3.3.1 Pre-processing

##### **AgERA5 Climatic Regions Embeddings**

AgERA5 variables are pre-processed to produce a set of layers (or embeddings) describing the meteorological properties of different climatic regions. These layers help the algorithm to model interannual phenological variations, boosting robustness against fluctuations in the input features, due to different meteorological conditions, across different years.

Daily AgERA5 acquisitions are pre-processed to fill data gaps and outer borders on shorelines, by using the nearest valid values. The pre-processed layers are then reduced to seasonal aggregates for each variable (dewpoint temperatures, solar radiation flux, precipitation flux, minimum temperature, mean temperature, maximum temperature), computing the mean in the 90 days across the season, resulting in 24 seasonally aggregated variables.

For each of the 30,000 training points, the values of the seasonally aggregated variables are extracted and reduced to 6 variables using UMAP (Uniform Manifold Approximation and Projection) [27]. A multi-layer perceptron (MLP) model is then used to learn the parametric UMAP transformation to replicate it on the

global dataset, producing the 6 variables on global scale from the 24 seasonally aggregated input variables (Figure 6).

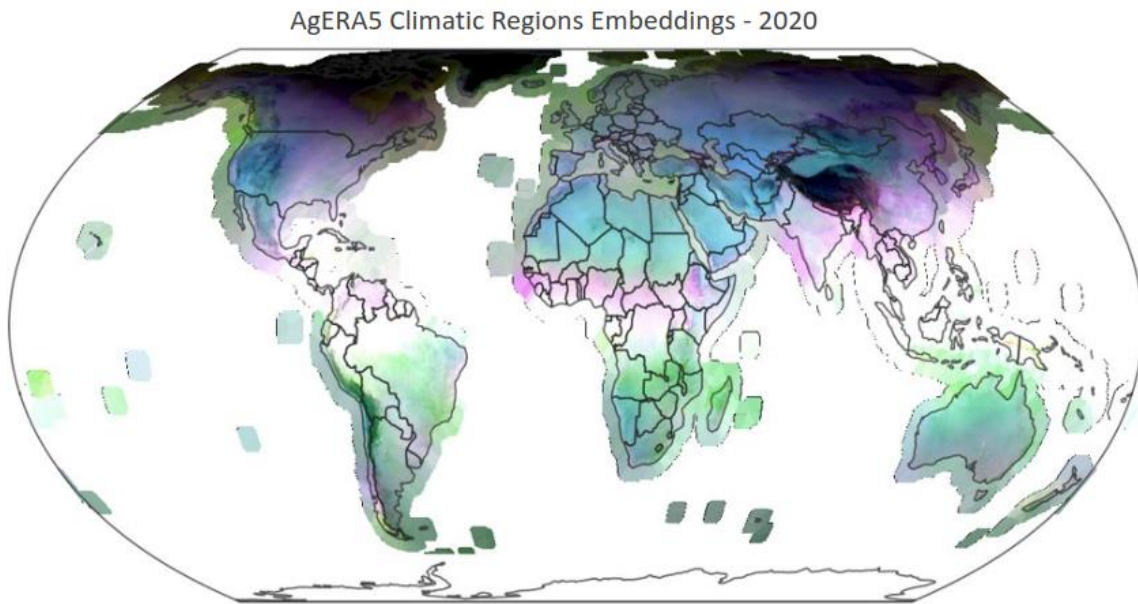


Figure 6: AgERA5 climatic regions embeddings for the year 2020 (displaying only the first 3 variables in an RGB composite for illustration purpose)

### SEN4LDN Cloud Masking

L2A products are distributed with a 20m scene classification (SCL) layer. This layer includes a classification of cloudy pixels. Cloud detection is sometimes affected by omissions (cloudy observations in reality, but no cloud masking set) which can cause artefacts in the derived yearly features. To improve the quality of input features, the CloudSEN12 [28] dataset was used to train a deep learning-based clouds detection. The model is trained at a resolution of 60m. A U-Net [29] architecture is used, with a ResNet50 backbone [32] and dice loss.

Comparisons with the SCL masks showed mixed results, with better performance for certain cases, but worse for others. To obtain the most conservative approach in reduction of clouds, the SCL and the new SEN4LDN cloud masks are merged, so that if a pixel is marked as cloud in either of the masks it will be masked from the data. The results show a noticeable improvement in the quality of the input features (Figure 7, Figure 8).

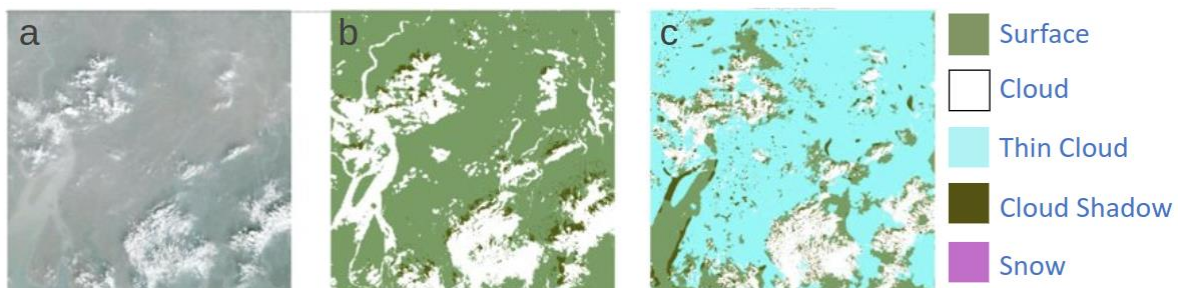


Figure 7: Sample observation over tile 45QXE. a) RGB image, b) SCL cloud mask (custom colours), c) SEN4LDN cloud mask, showing better detection over the haze covering the scene.

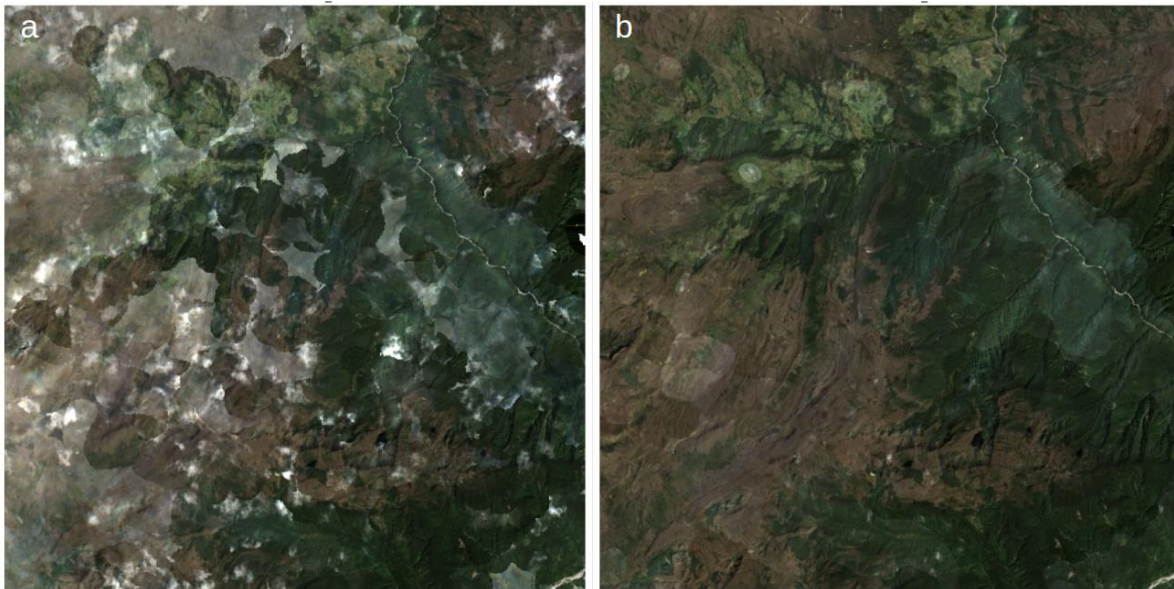


Figure 8: Yearly RGB median for a patch (1km x 1km) over Colombia. a) RGB median computed using the SCL cloud masks for the yearly time-series: undetected clouds cause artefacts in the obtained image. b) The same data is masked using the merged SCL and SEN4LDN cloud masks, obtaining an improved yearly median image with reduced residual cloud artefacts.

### L2A data and features computation

The Sentinel-2 L2A surface reflectance time-series are pre-processed in the following way (Figure 9):

1. Data is loaded for each band.
2. The Sen2Cor Scene Classification (SCL) is used to derive a cloud mask.
  - a. Clouds and shadows are first eroded by 30m to remove false positives over small bright buildings or small dark features, and then dilated by 120m to extend over the clouds borders, which are often not masked properly in the SCL.
  - b. Snow pixels are considered invalid if the pixel is classified as snow for less than 95% of the observations. If the pixel is classified as snow more than 95% of the observations, snow is not masked for that pixel as the pixel can be classified as permanent snow/ice.
  - c. The SCL derived cloud mask is merged with the SEN4LDN cloud mask.
3. Invalid values are masked out using the obtained cloud mask.
4. Median composites are then computed for each band with a moving window of size 20 days and a frequency of 10 days, resulting in 36 cloud-free composites for each year.
5. Missing data gaps are further filled in the time-series by linearly interpolating between the closest available values. In case of missing data at the beginning or end of the time-series, the values are filled with the median of the valid values.
6. From the pre-processed time-series, vegetation indices are computed and percentiles (10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>) are computed for bands B02, B03, B04, B08, B11 and B12 and for the NDVI index, yielding 35 Sentinel-2 features.



7. The percentiles are used as the features for the classification together with auxiliary data which serves as localizing features for the model: DEM altitude, 6 AgERA5 climatic regions embeddings, latitude and longitude.

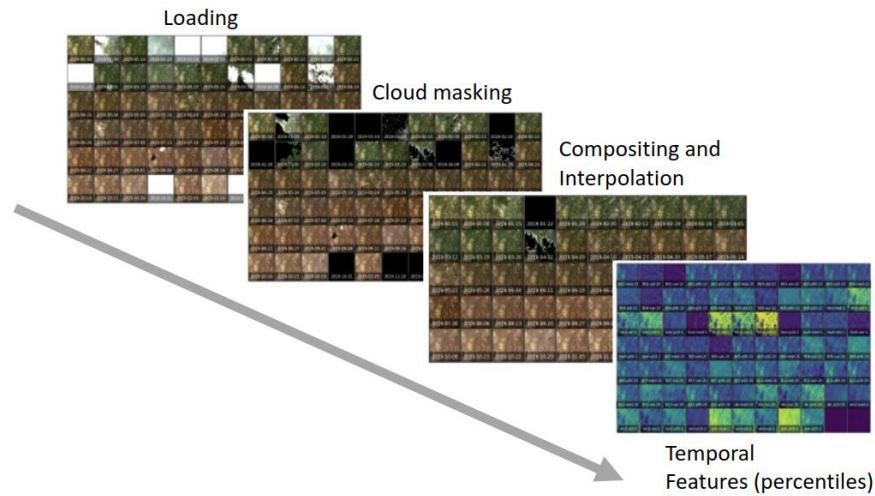


Figure 9: Summarizing scheme of the L2A pre-processing steps yielding the temporal features

#### 4.3.3.2 Deep Learning Model

For the deep learning algorithm, it was opted to use a UNet++ [30] architecture, based on a ResNet18 [32] backbone, which is considered state-of-the-art for semantic segmentation tasks.

The model was trained using a Python 3.10 ecosystem and Pytorch [34]. Training data for 30,000 locations and 5 years was used, for a total of 150,000 samples, with a hold-out validation set of 15,000 samples. Data augmentation is applied (90 degrees rotations and flipping), as well as scaling augmentation. Scaling of the data is performed using a sigmoid function for Sentinel-2 data (Eq. 1).

$$y(x) = \frac{2}{1+e^{-kx}} - 1$$

Eq. 1

The sigmoid scaling is performed with a default value of  $k=7$  and augmented during training by adding a random value. At loading time the scaling of each sample can have  $k = 7 \pm d$  where  $d$  ranges between 0 and 3.

The DEM data is scaled by dividing by 4000, while latitude and longitude are divided by 90 and 180 respectively. To avoid overfitting associated with the latitude and longitude values, values are also augmented by  $\pm 1^\circ$  before scaling.

Dice loss is used [35] with a starting learning rate of 0.008 and an exponential decay. Training is performed until convergence at approximately 40 epochs with a batch size of 50. Final learning rate after 40 epochs is 0.002.

Only a subset of bands is used to train the deep learning algorithm in order to reduce memory footprint and speed up the learning process<sup>6</sup>. Bands were determined through benchmarking models trained with different groups of bands, and selecting the best performing candidates.

The model intersection over union score is 0.64. Figure 10 and Figure 11 show the relative confusion matrix and the spatial accuracy map, respectively, where the performance of the model is evaluated globally to better understand potential regional issues or anomalies which cannot be evaluated from global metrics.

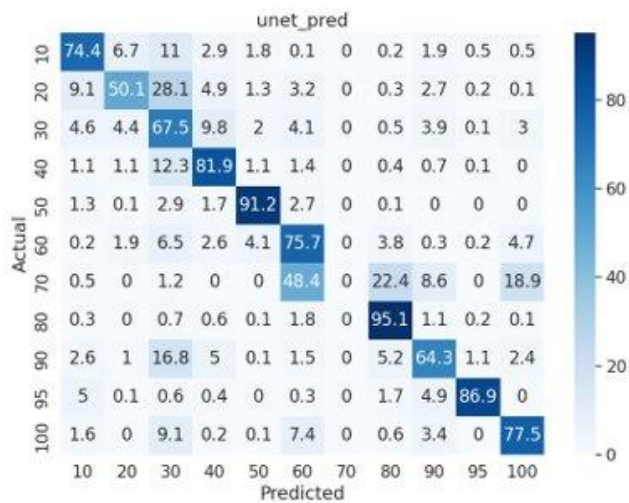


Figure 10: Confusion matrix of the deep learning model. It is to be noted that the algorithm is not able to learn to predict snow. This is due to the lower number of samples in the training data set.

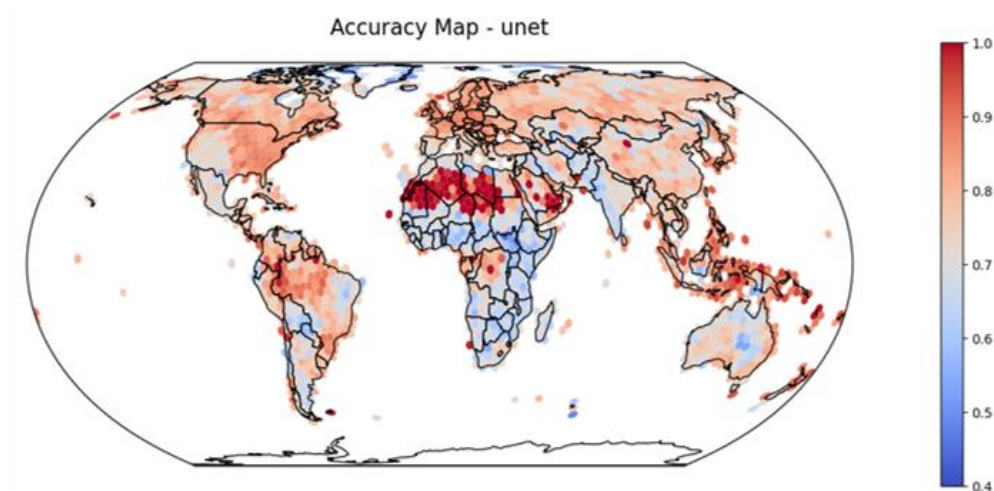


Figure 11: Spatial accuracy map of the deep learning algorithm showing good performance in the northern hemisphere and homogeneous landscapes, while showing poorer performance in mixed landscapes, for example

<sup>6</sup> The bands used are: s2-B04-p50, s2-B03-p50, s2-B02-p50, s2-B08-p50, s2-B11-p50, s2-B12-p50, s2-B04-p10, s2-B08-p10, s2-B12-p10, s2-ndvi-p10, s2-ndvi-p25, s2-ndvi-p50, s2-ndvi-p75, s2-ndvi-p90, cop-DEM-alt, latitude, longitude.

central Africa and in the Australian xeric shrublands deserts. This can be explained considering two aspects. Firstly, these are challenging areas, and secondly the training data is produced synthetically from ESA WorldCover which itself shows lower accuracy in these problematic areas.

#### 4.3.3.3 CatBoost Model and Benchmarking

CatBoost [32] is a state-of-the-art gradient boosting machine learning algorithm, trained and applied on tabular data. To train the CatBoost algorithm, a data frame of features is generated by sampling pixel features and annotations from the training patches. The sampling is designed to reduce the amount of total training samples (as there would be ~2,5 billion samples:  $128 \times 128$  pixels for each patch  $\times$  30,000 patches  $\times$  5 years) and at the same time to balance the dataset, such that minority classes are well represented in the training and test sets. This is done by applying a spatial sampling. The H3 grid<sup>7</sup> at resolution 2 is used to derive a set of global hexagonal cells. For each cell, all the training locations are considered, and the total cover fractions of the annotations is computed. The total cell fractions determine the sampling probability for each class. The number of pixels sampled for each class  $n_c$ , in every location, is given by:

$$n_c = \min(n_{ctot} \cdot p_c, n_{max})$$

Eq. 2

with  $p_c$  the probability of sampling for a given class  $c$  in the patch,  $n_{ctot}$  the total number of pixels for a given class  $c$  in the patch, and  $n_{max} = 200$  pixels.

The sampling probability (Figure 12) is derived for each class and each H3 cell as:

$$p_c = p_{min} + (p_{max} - p_{min}) \cdot e^{-a \cdot f_c}$$

Eq. 3

where  $f_c$  is the class cover fraction for the given class  $c$ , averaged over all the locations in the H3 cell and  $p_{min} = 0.01$ ,  $p_{max} = 0.05$  and  $a = 10$  are arbitrary constants that were empirically determined.

<sup>7</sup> <https://h3geo.org/>

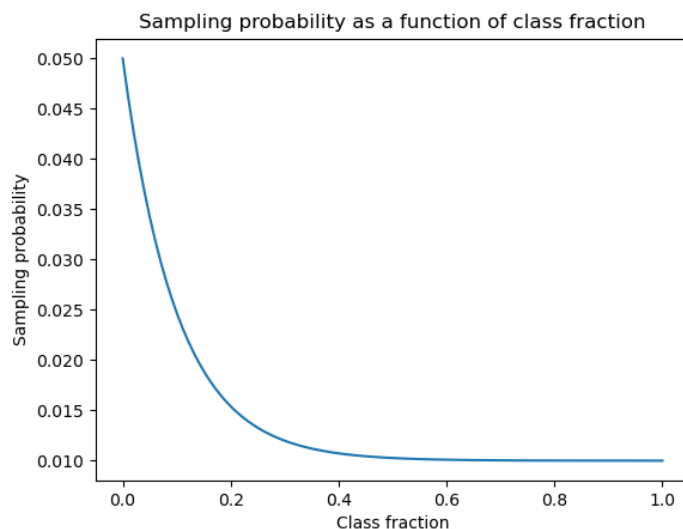


Figure 12: Sampling probability based on H3 grid cell class fraction. The less pixels of a certain class are present in a patch, the higher the number of sampled pixels.

From the obtained samples, a further reduction is applied by randomly sampling 60% of the pixels, yielding a final data frame with approximately 23 million samples, of which 80% are used for training and 20% for testing. In total, the sampling selects approximately 1% of the initial 2.5 billion samples available from the patch dataset. The class balance of the obtained dataset is shown in Figure 13. It is worth noting that this balance is obtained via the spatial stratification applied through the local balancing on the H3 cells, ensuring an optimized global distribution of samples.

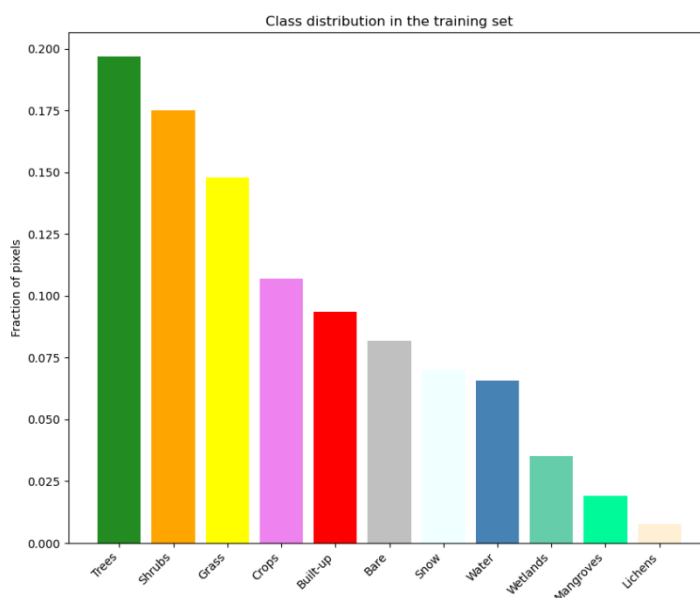


Figure 13: Class distribution of the pixel training dataset obtained through the spatially balanced sampling described above.

A series of models is benchmarked to evaluate the contribution of the different sets of features, summarized in Table 2. The results of this benchmarking process are presented in Annex A.

*Table 2: Summary of feature groups that are used in different combinations for model training*

Features Group	Features Group Id	Description
Sentinel-2 L2A	s2	35 L2A timeseries percentiles
Sentinel-2 L2A dimensionally reduced	umap	Reduction of the L2A timeseries percentiles to a set of 12 features using UMAP. These were tested as a replacement for the L2A features, to evaluate if dimensionality reduction would have improved the models performance.
Land Cover deep learning features	unet	11 class probabilities obtained from the deep learning model trained on the training dataset patches
AgERA5 Climatic Regions Embeddings	meteo	6 Climatic regions embeddings produced from AgERA5 data for each year
DEM Altitude	dem	Copernicus DEM altitude
Latitude and Longitude	latlon	Latitude and Longitude
Basic Land Cover Probabilities	cbprob	3 probabilities (vegetated, not vegetated, water) obtained from a 3 CatBoost regression models trained on an aggregated subset of the training data

The obtained model successfully achieves robustness on the classification across different years. This is also verified by applying the model over the 3 test areas, Portugal, Colombia and Uganda over 6 years (2018 – 2023), e.g. including year 2023, for which no training data was used from that year. The model is nevertheless able to generalize well over unseen data and produce consistent predictions. Examples of predictions from the year 2023 are shown in the next section.

Besides the robustness in the predictions across different years, the model also shows significantly less confusion between certain classes which share similar signal fingerprint. For example, wetland and cropland, built-up and bare, grassland and shrubland and grassland and cropland. The introduction of deep learning features expands the capabilities of a simple pixel classifier by adding contextual features, which help the algorithm in performing better over these challenging classifications.

Despite the good generalization achieved, the model is still suffering from a certain degree of noise in predictions from different years. This is due to the fluctuations of the percentiles values, which depends on the phenological evolution and fluctuations of vegetation, as well as the amount of cloudless data that is available for each pixel in different years. A different number of valid observations, between different years, can have the effect of shifting slightly the percentiles features making. The other main problem with the current algorithm is the absence of proper annotations for years before and after 2021. Training is performed under the assumption that no changes are present in the training locations, however a certain amount of change will always be present, and this introduces noise in the training data, which causes noise and instability in the predictions of challenging cases. A post-processing algorithm, presented in the next section, was implemented, and applied to mitigate this noise in the predictions, which would cause spurious change detections.

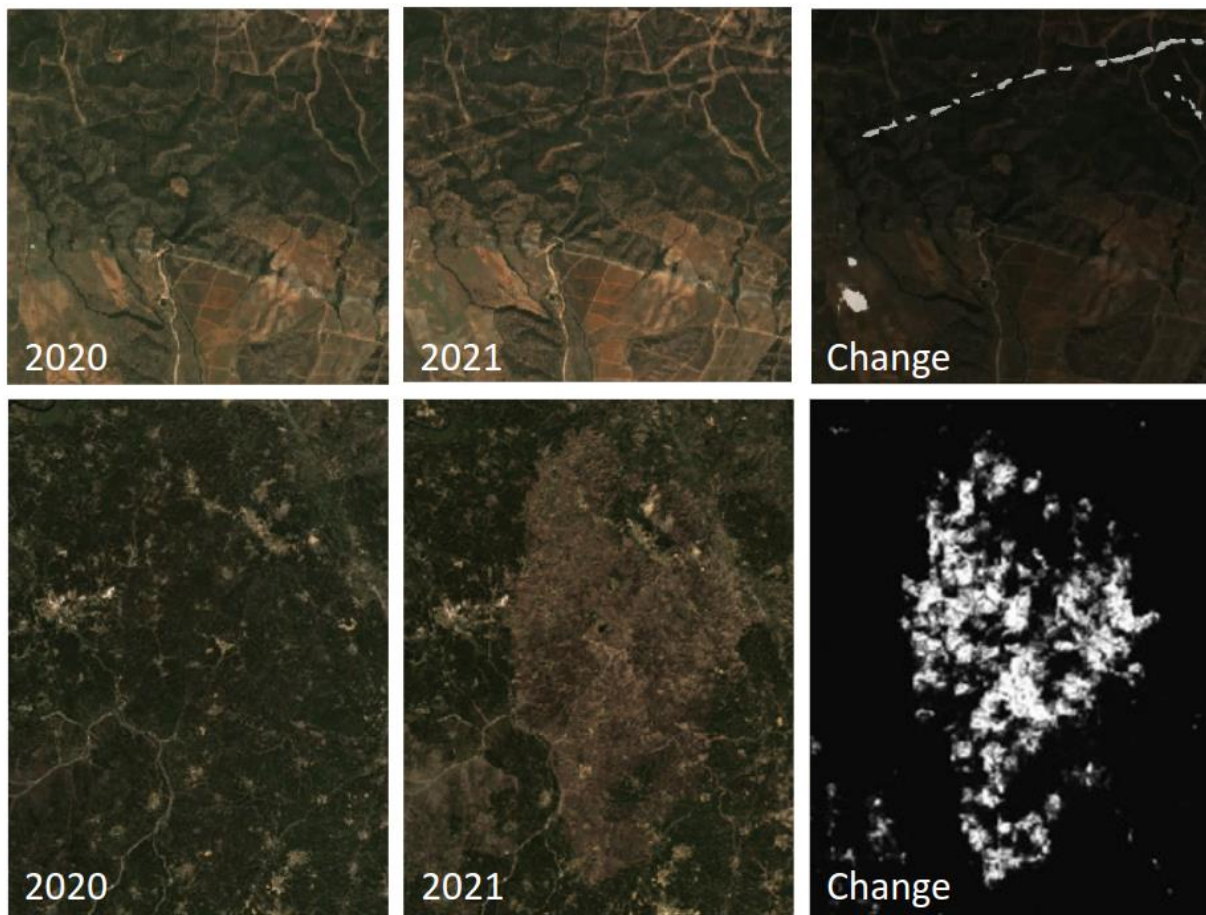


#### 4.3.3.4 Post-processing

The post-processing algorithm was designed to minimize differences between predictions of different years and further improve the robustness of the products against noise and fluctuations of the predictions which do not correspond to actual land cover changes.

##### **Change Detection**

One key challenge is to avoid altering and overwriting actual changes. The first step in the algorithm is therefore detecting real changes. For this purpose, a synthetic change dataset and a deep learning model were initially developed. The figure below shows an example of the model predictions.



*Figure 14: Deep learning-based change detection model predictions for two areas in Portugal. Changes are detected partially. The algorithm shows low sensitivity.*

Despite the promising results, the algorithm showed a low sensitivity for less dominant changes. This can be due to the deep learning architecture used and the training data, which was completely synthetic as a global change annotation dataset was not available at the time of development. As a result, it was opted to develop a simpler but more robust approach.

The approach is based on the following idea. The land cover classification algorithm produces probabilities which are then binarized into a discrete map. These probabilities are continuous and can fluctuate in difficult cases. For example, in certain conditions a pixel could be classified as grassland, with a grassland

probability of 0.51 and shrubland probability of 0.49. For the year after, due to noise and phenological fluctuations, the algorithm could predict a probability of 0.49 for grassland and 0.51 for shrubland which would lead to a classification as shrubland. If the comparison is made only based on the discrete classification, this would cause the detection of a change. In this example it is clear how the probabilities bear more information over the discrete classifications and can help distinguish real changes from those that might be due to noise and classification errors.

To exploit this, the cosine similarity of the probabilities distribution is used as a measure of the change between two pixels. The cosine similarity of two vectors  $A$  and  $B$  is defined as:

$$Cs(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

Eq. 4

If the cosine similarity is close to 1, it means that the probability distributions are similar and the probability of change is therefore low. For the previous example, the cosine similarity of two vectors  $A = [0.51, 0.49]$  and  $B = [0.49, 0.51]$  is 0.9992. On the other hand, the lower the cosine similarity, and the higher the probability of an actual change.

### **Multi-year normalization**

In the post-processing, the objective is to look at the pixel predictions trajectory over different years and homogenize them, but only for those predictions which show a low probability of change.

To achieve that, the probability distribution for a target year is generated as the weighted sum of all the predictions available in the timeseries. The weight of each different year is computed as a function of the cosine similarity with the target year. For example, the post-processing of the year 2018 prediction, will involve computing the cosine similarity between the predictions of the year 2018 against all other years available. The cosine similarity is used to obtain the weighting factors to average each year's probability with the target year. The weight is computed as 0 if the cosine similarity is between 0 and 0.5 (high probability of change) and grows linearly from 0 to 1 in the interval where the cosine similarity is between 0.5 and 1, such that the contribution of other probabilities to the weighted sum is higher for predictions which are similar, and lower as the probability of change increases.

The post-processing formula is given as:

$$P'_i = \frac{\sum_n^N (f(C_{ni}) \cdot P_n)}{\sum_n^N f(C_{ni})}$$

Eq. 5

Where  $P'_i$  is the post processed probability distribution for the target sample of year  $i$ ,  $C_{ni}$  is the cosine similarity between the target year  $i$  and the year  $n$ , and  $P_n$  is the probability distribution of year  $n$ . The weighting function  $f$  is defined as:

$$f(x) = 0 \text{ for } 0 < x \leq 0.5$$

Eq. 6

and

$$f(x) = 2x - 1 \text{ for } 0.5 < x \leq 1$$

Eq. 7

This post-processing step is applied recursively to probability distributions until convergence is achieved. Convergence is here defined as the difference between the results of the current iteration and those of the previous iteration being smaller than a specified tolerance threshold of  $1e-4$ . To prevent the recursive process from running indefinitely if convergence cannot be met, a maximum of 20 iterations is enforced.

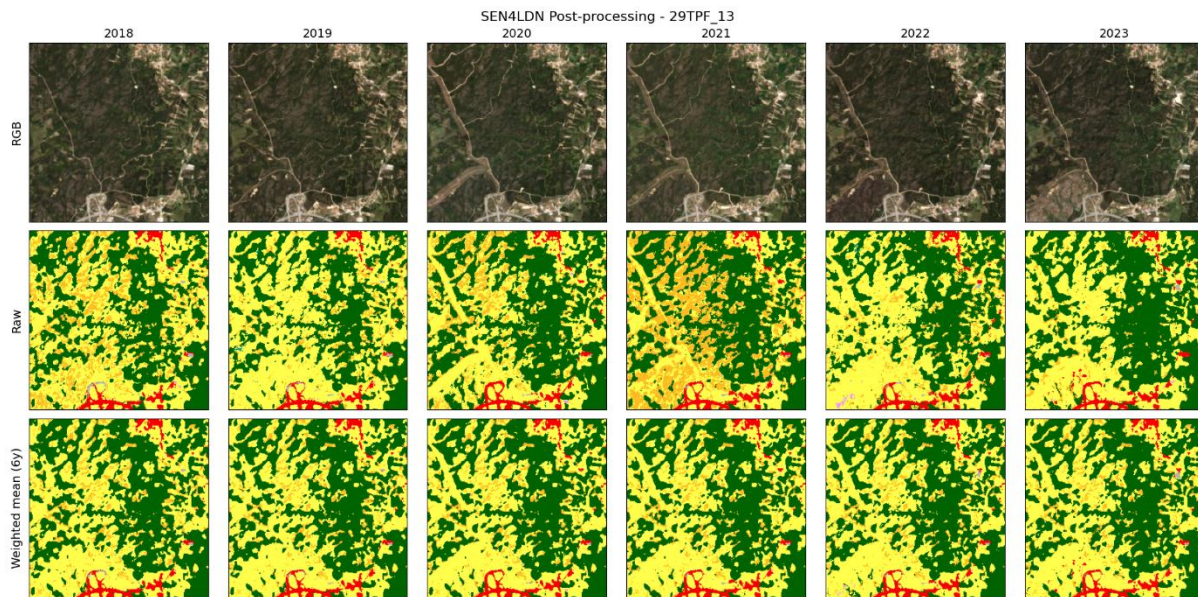


Figure 15: In the figure, an area of  $2.5\text{km} \times 2.5\text{km}$  is displayed over Portugal. The top row are the yearly RGB median composites for the 6 years 2018-2023. The second row are the predictions obtained from the Land Cover classifier and the third row are the predictions obtained by the post-processed predictions. The example highlights the stabilizing effect of the post-processing. In the absence of change, the raw predictions show higher shrubland/grassland fluctuations, which are reduced in the post-processed predictions, which appear more stable.



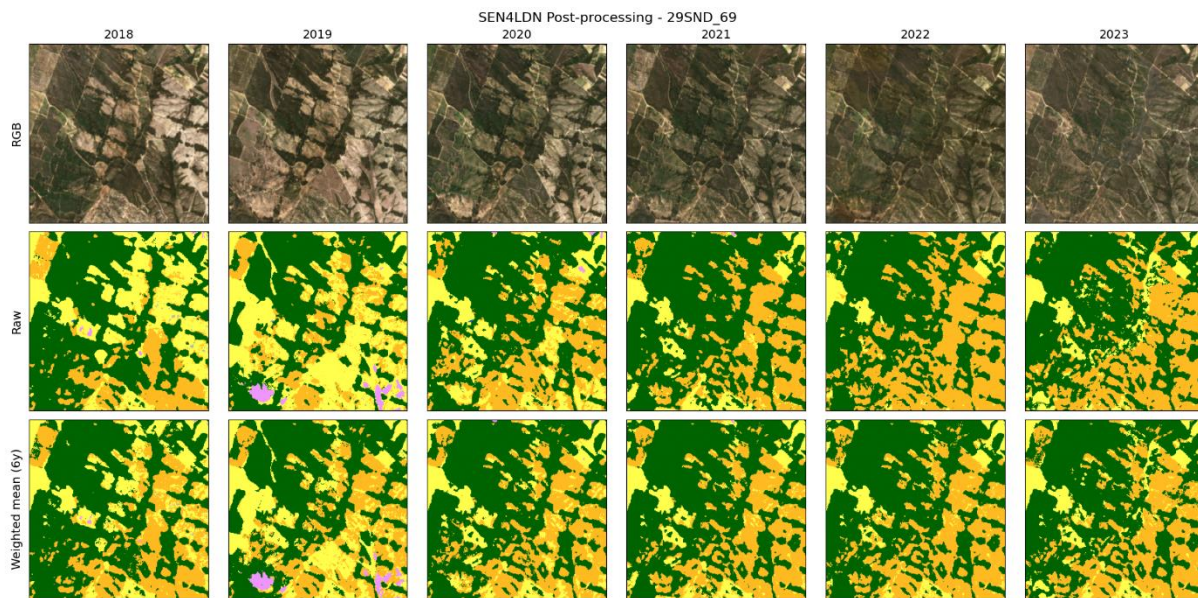


Figure 16: In this example from Portugal, the post-processing helps in stabilizing the predictions while preserving well the changes. For example, shrubland distribution between 2018 and 2019 is significantly more stable in the post-processed version. At the same time the deforestation event on the bottom left is captured and is not affected by the post-processing.

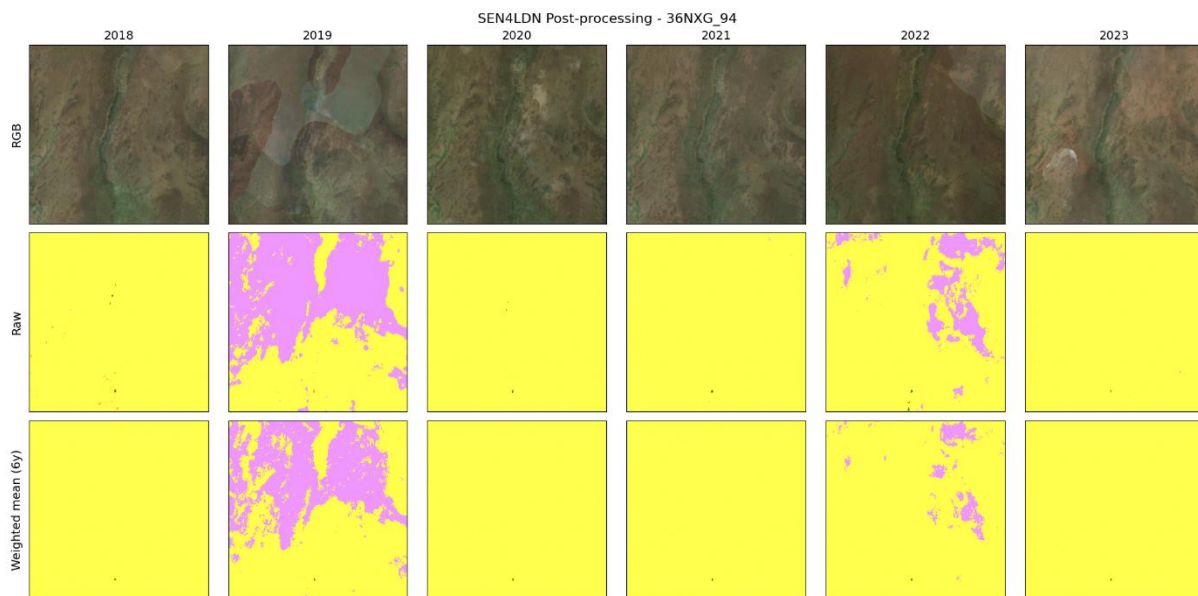


Figure 17: In this example from Uganda, we can observe a limitation of the classification algorithm. Residual clouds are present in 2019 and 2022. These residual thin clouds, which are not filtered out correctly by the pre-processing step, alter the features and lead to the spurious commission of cropland. The post-processing in this case slightly

reduces the commission, however it is not able to override errors in the classification, as these depends on issues in the input data and are considered as 'real changes' from the perspective of the post-processing algorithm.

#### 4.3.3.5 Land cover transitions

##### Transition Matrix

The Land cover sub-indicator is based on transitions of land cover from the initial year to the final year. A transition matrix is used to mark the transitions as degraded, stable or improved. Table 3 provides a proposed default matrix with predefined transition statuses based on the considered classes (see Table 8) and the default transition matrix as proposed in the GPG. However, our final algorithm aims to allow for adjustments to these transitions, taking into account local contexts and specific settings.

Table 3: General land cover transition matrix using the 11 land cover classes (see §4.3.4.2). Land cover change processes are colour coded as positive **P**, stable **S** or negative **N**. Unlikely transitions are put within brackets. Note that this is the general starting point, but that specific transitions can be re-evaluated per use case, to adopt to local conditions.

		Final class										
Original class	IPCC LC class	IPCC LC class										
		Forest land	Other	Grassland	Cropland	Settlement	Other	/	Other	Wetland	Forest	Other
Original class	Forest	Tree cover	Shrubland	Grassland	Cropland	Built-up	Bare / sparse vegetation	Snow and ice	Permanent water bodies	Herbaceous wetland	Mangroves	Moss and lichen
	Other	S	N	N	N	N	N	-	-	(N)	(S)	(N)
	Grassland	P	S	N	P	N	N	-	-	P	(P)	(N)
	Cropland	P	P	S	P	N	N	-	-	N	(P)	(N)
	Settlement	P	N	N	S	N	N	-	-	N	(P)	(N)
	Other	(P)	(P)	(P)	(P)	S	(P)	-	-	(P)	(P)	(P)
	Other	P	P	P	P	N	S	-	-	P	(P)	(P)
	/	-	-	-	-	-	-	-	-	-	-	-
	Other	-	-	-	-	-	-	-	-	-	-	-
	Wetland	-	-	-	-	-	-	-	-	-	-	-
	Forest	N	N	N	N	N	N	-	-	S	(P)	S
	Other	(S)	N	N	N	N	N	-	-	(N)	S	(N)
Original class	Forest	P	P	P	P	N	N	-	-	P	(P)	S
	Other	P	P	P	P	N	N	-	-	P	(P)	S
	Grassland	P	P	S	P	N	N	-	-	N	(P)	(N)
	Cropland	P	N	N	S	N	N	-	-	N	(P)	(N)
	Settlement	(P)	(P)	(P)	(P)	S	(P)	-	-	(P)	(P)	(P)
	Other	P	P	P	P	N	S	-	-	P	(P)	(P)
	/	-	-	-	-	-	-	-	-	-	-	-
	Other	-	-	-	-	-	-	-	-	-	-	-
	Wetland	-	-	-	-	-	-	-	-	-	-	-
	Forest	N	N	N	N	N	N	-	-	S	(P)	S
	Other	(S)	N	N	N	N	N	-	-	(N)	S	(N)
	Forest	P	P	P	P	N	N	-	-	P	(P)	S
	Other	P	P	P	P	N	N	-	-	P	(P)	S

**Degradation process**








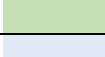


- Deforestation
- Vegetation loss
- Urban expansion
- Inundation
- Withdrawal of agriculture
- Wetland drainage

**Improvement process**

- Reforestation
- Vegetation establishment
- Wetland establishment
- Agricultural expansion

The land cover transitions map thus provides information on the degradation and improvement processes that have taken place between the initial year and the final year (i.e. the reporting period), based on the post-processed land cover class probabilities for the initial and final year at 10m resolution. Class definition and colour coding is specified in Table 4. Table 5 highlights some key degradation and improvement processes for the SEN4LDN SDG Early adopter countries, that are likely to be detectable based on trends in land cover.

*Table 4: Definition, colour and map code of the Land cover transitions layer. Land cover transitions (degradation and improvement processes) are based on the land cover transition matrix (see Table 3).*

Degradation / Improvement process	Definition	Colour	Map code
Deforestation	Deforestation refers to the intentional clearing of forested land, i.e. conversion of tree cover (or mangroves) to grassland, cropland or built-up areas.		1
Vegetation loss	Vegetation loss refers to the reduction or decline in vegetation cover, e.g. conversion from tree cover to shrubland, or conversion from other land to bare / sparse vegetation.		2
Urban expansion	Urban expansion refers to the physical extension settlements into the surrounding countryside, i.e. conversion from other land to built-up.		3
Inundation	Based on the annual LC maps, prolonged inundation can be detected. This involves the conversion of grassland or cropland to herbaceous wetland.		4
Withdrawal of agriculture	This degradation process refers to the conversion of cropland to shrubland or grassland.		5
Wetland drainage	This degradation process refers to the conversion of wetland to grassland, cropland, or other land.		6
Stable / Unlikely change			0
Reforestation	Reforestation is an improvement process that refers to the establishment or restoration of tree cover.		101
Vegetation establishment	Vegetation establishment is the improvement process of introducing vegetation cover, i.e. conversion of bare / sparse vegetation to other vegetation classes.		102
Wetland establishment	Wetland establishment refers to the conversion of shrubland, built-up areas or other land to wetlands.		103
Agricultural expansion	This improvement process refers to the conversion of grasslands or other land in cropland.		104

*Table 5: Key degradation and improvement processes for SEN4LDN SDG Early adopters, likely to be detectable based on trends in land cover*

Degradation process	Land cover transition			Highly relevant for SDG Early adopter		
				Uganda	Colombia	Portugal
Deforestation	Tree cover	⇒	Grassland	●	●	●
	Tree cover	⇒	Cropland	●	●	●
	Tree cover	⇒	Built-up	●	●	○
Vegetation loss	Tree cover	⇒	Shrubland	●	●	●
	Tree cover	⇒	Bare / sparse vegetation	●	●	●
	Shrubland	⇒	Grassland	●	●	●
	Shrubland	⇒	Bare / sparse vegetation	●	●	●
	Grassland	⇒	Bare / sparse vegetation	●	●	●
	Cropland	⇒	Bare / sparse vegetation	●	●	●
Urban expansion	Shrubland	⇒	Built-up	●	●	○
	Grassland	⇒	Built-up	●	●	○
	Cropland	⇒	Built-up	●	●	○
	Bare / sparse vegetation	⇒	Built-up	●	●	○
Inundation	Grassland	⇒	Herbaceous wetland	●	○	○
	Cropland	⇒	Herbaceous wetland	●	○	○
Withdrawal of agriculture	Cropland	⇒	Shrubland	●	●	●
	Cropland	⇒	Grassland	●	●	●

Wetland drainage	Herbaceous wetland	⇒	Tree cover	●	○	○
	Herbaceous wetland	⇒	Shrubland	●	○	○
	Herbaceous wetland	⇒	Grassland	●	○	○
	Herbaceous wetland	⇒	Cropland	●	○	○
	Herbaceous wetland	⇒	Built-up	●	○	○
	Herbaceous wetland	⇒	Bare / sparse vegetation	●	○	○
Improvement process	Land cover transition			Highly relevant for SDG Early adopter		
				Uganda	Colombia	Portugal
Reforestation	Shrubland	⇒	Tree cover	●	●	●
	Grassland	⇒	Tree cover	●	●	●
	Cropland	⇒	Tree cover	●	●	●
	Bare / sparse vegetation	⇒	Tree cover	●	●	●
Vegetation establishment	Grassland	⇒	Shrubland	●	●	●
	Bare / sparse vegetation	⇒	Shrubland	●	●	●
	Bare / sparse vegetation	⇒	Grassland	●	●	●
Wetland establishment	Shrubland	⇒	Herbaceous wetland	●	●	●
	Grassland	⇒	Herbaceous wetland	●	●	●
Agricultural expansion	Shrubland	⇒	Cropland	●	●	●
	Grassland	⇒	Cropland	●	●	●
	Bare / sparse vegetation	⇒	Cropland	●	●	●

### Transition Maps & Transition Probabilities

Transition maps are conventionally computed starting from the discrete maps. As shown earlier for the change detection, this can lead to an overestimation of transitions and noise in the maps due to noise in the discrete predictions. A new algorithm to derive transition probabilities starting from the classification probabilities was developed to mitigate this issue and reduce overestimation of transitions. Figure 18 shows a graphical workflow depicting the algorithm.

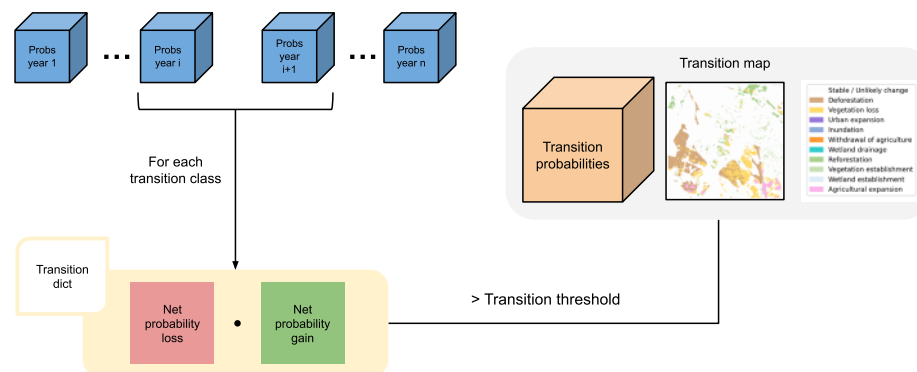


Figure 18: Transitions probabilities workflow

For each type of transition that is reported in the transition matrix, there is a group of starting classes and a group of target classes. If the land cover class probability flows from one of the starting classes to one of the target classes, this is considered as a probability for that transition. Since it's not possible to control the probabilities flow on single classes, the algorithm aggregates the probabilities of the starting and target classes. For each class the probability difference between the start year and end year is computed. The differences are then summed together. The same is done for the probabilities of the target transition.



For example, for a deforestation transition, the starting class is *Tree cover* and the target classes are *Grassland*, *Cropland*, *Built-up*. Therefore, the first term is obtained by the difference between the tree cover in the target year and the starting year, while the second term is obtained by the sum of the differences of the three target classes across the two years. If the first term is negative, it indicates a loss in the initial class probabilities, which translates into a gain for some other landcover class. If the gain occurs in one of the transition destination classes, the second term will be positive. The crucial information indicating a transition is represented by negative differences in the first term and positive differences in the second term. Positive differences in the first term and negative differences in the second term are discarded. Finally, a transition threshold is applied to the resulting probabilities to generate the transition map.

Figure 19 and Figure 20 show two examples of transition maps generated from the transition probabilities.



Figure 19: Example of transition maps generated from the transition probabilities. The first row of images shows the RGB medians and Land Cover maps, and the transition maps obtained from the discrete land cover maps. In the second row it is presented a series of transition maps obtained with different thresholds on the transition probability, starting at 0.3 and finishing at 0.7. Although logically 0.5 would represent the default cutoff threshold, it was observed that 0.4 would be a more conservative choice, allowing to preserve more transitions. The last row shows a comparison of the prediction-based transition map and the transition map obtained from the transition probabilities. In red the pixels which are not considered transitions, showing how the method significantly reduces noisy transition pixels, and preserves the actual transitions.



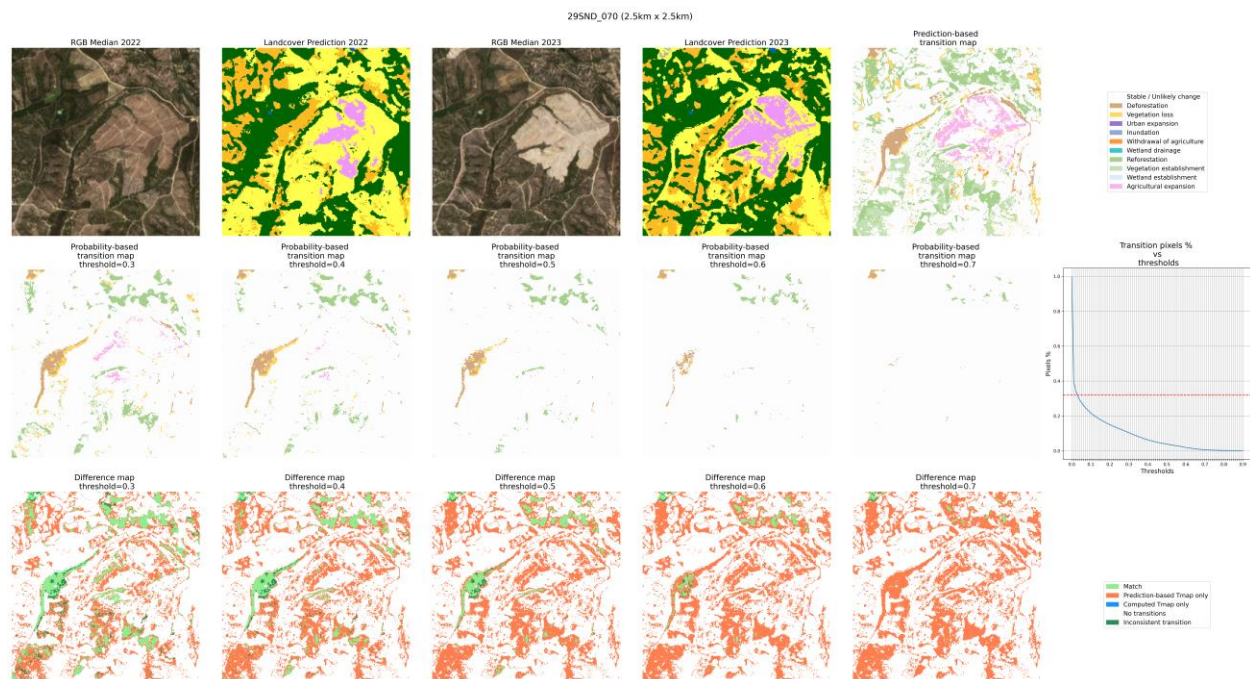


Figure 20: Similarly to the previous figure, another example of transition maps obtained from the transition probabilities, highlighting the reduction of noise. The threshold chosen for the final maps is 0.4.

### Land Cover Degradation Maps

Finally, the obtained transition probabilities are used to derive a degradation or improvement probability LCD-PROB (e.g. Figure 21), with values between 0 and -1 for degradation processes and between 0 and 1 for improvement processes, by taking the absolute maximum of degradation resp. improvement transition probabilities, and assigning negative values to degradation processes. From the LCD-PROB, using the same threshold applied for the transition maps, a Land Cover Degradation (LCD) map can be derived (e.g. Figure 22). The development of continuous transition and degradation probabilities expands and refines the capabilities of the system, on one side mitigating over detection of transitions, occurring when using discrete land cover maps would be used as starting point (e.g. Figure 23), and on the other side providing more information to the user which can either use the improved discrete transition and degradation maps or use their probabilities for more advanced applications.

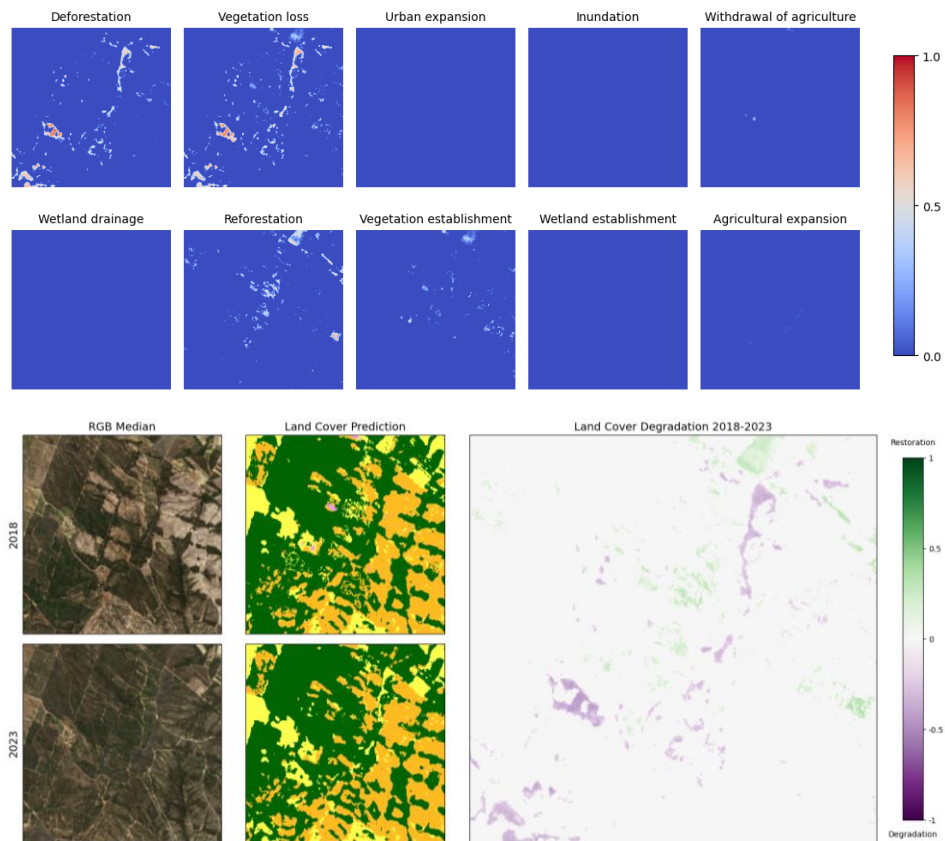


Figure 21: Example of a Degradation/Improvement probability (LCD-PROB). Providing a continuous probability enables users to eventually customize their results or use the probabilities as input for downstream applications.

29SND (2.5km x 2.5km)  
lat: 39.15, lon: -8.59

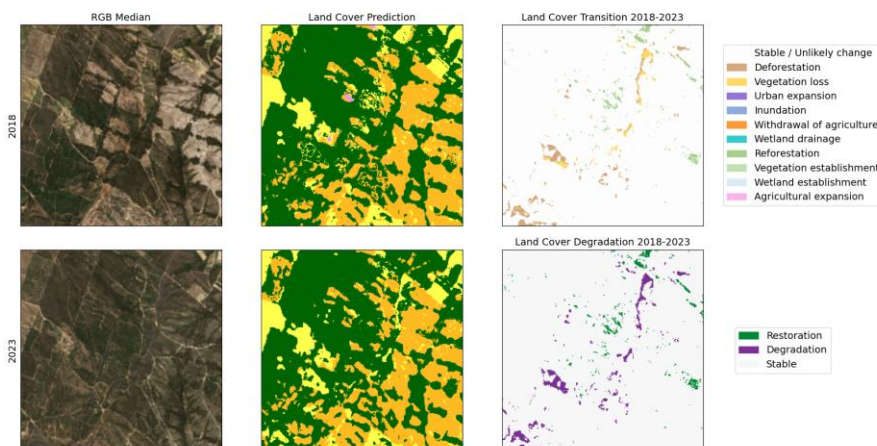


Figure 22: Discretizing the Land Cover Degradation probabilities yields the Land Cover Degradation (LCD) (bottom right) for the area of interest.

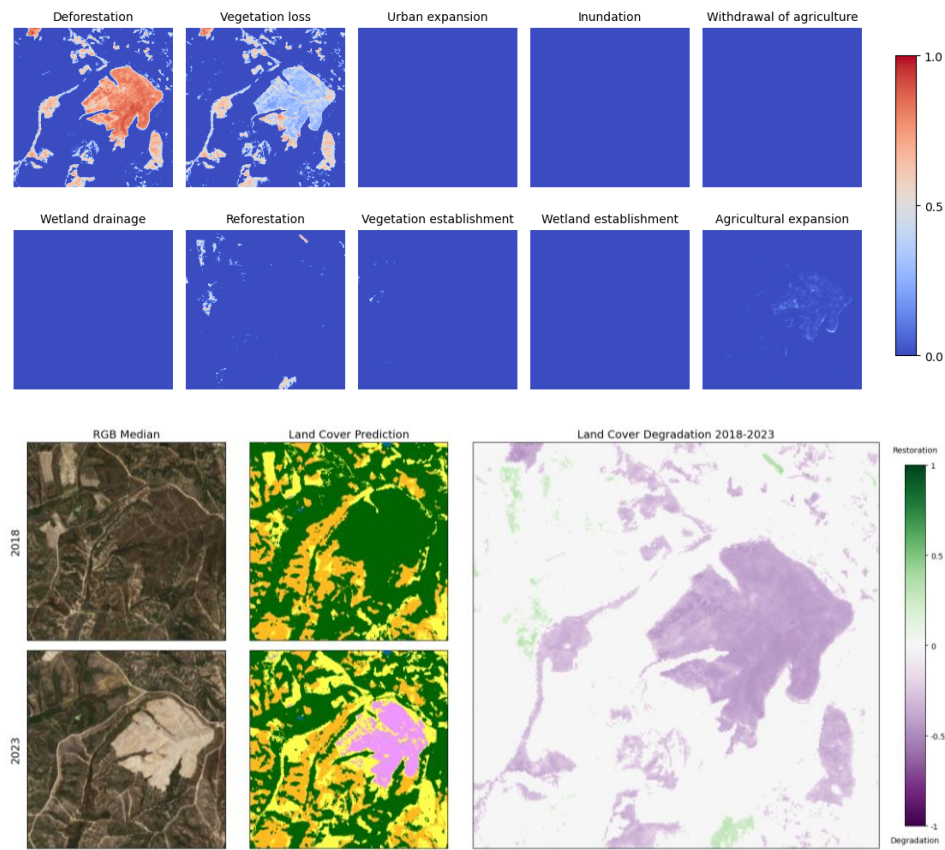


Figure 23: Land Cover Transition and Degradation probabilities for another nearby location.

29SND (2.5km x 2.5km)  
lat: 39.15, lon: -8.47

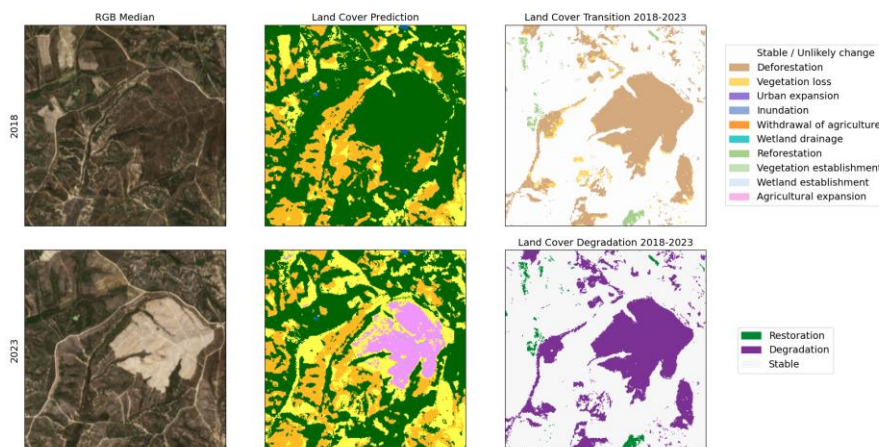


Figure 24: Transition and LCD maps for the same location, as well as the Land Cover maps. In this case the Land Cover maps are shown as well to highlight that although in the central patch there is a larger area predicted as cropland, this does not end up in the transition/degradation maps. The algorithm is not providing a high confidence prediction of cropland against grassland and the consequent transition probabilities are low enough to be able to discard this as a spurious change.

### 4.3.4 Algorithm output variables

#### 4.3.4.1 Output products

The final output products of the trends in land cover algorithms are:

- Annual Land Cover Maps (LCM) (timeseries 2018 – 2023)
- Land Cover Transition map (LCT) (between year 2018 and year 2023)
- Land Cover Degradation Map (LCD) (between year 2018 and year 2023)
- Land Cover Degradation Probabilities (LCD-PROB) (between year 2018 and year 2023)

The products (see details in Table 6) are delivered as single band Cloud Optimized Geotiffs (COGs) in a 3x3 degree lat/lon (WGS84) tiling grid. The aim of the operational workflow would however be to make the spatial extent and projection grid to be user customizable.

The products naming will follow the scheme:

SEN4LDN\_<PRODUCT>\_V100\_<PERIOD>\_<LATLON\_TILE\_3>\_MAP.tif

For example, for a tile of the annual land cover maps:

SEN4LDN\_LCM\_V100\_2018\_N09W075\_MAP.tif

The date identifier for transition and degradation maps include the starting and ending years (for example 2018-2023).

The metadata included with the products follows the scheme delineated in Table 7.

Table 6: Output products of trends in land cover algorithms

Product	Content	Data type and range	No data value	Scale Offset	Legend
LCM	Time series of Annual Land Cover Maps	UINT8 [10-100]	0	-	See Table 8
LCT	Land Cover Transition Map	UINT8 [1-104]	0	-	See Table 4
LCD	Land Cover Degradation Map with classes indicating Stable (0), Improvement (1), Degradation (2)	UINT8 [0,1,2]	255	-	See Table 9
LCD-PROB*	Land Cover Degradation Probabilities scaled between -1 (degradation) and 1 (improvement)	UINT8 [0-250]	255	0.008 -1	-

\* The LCD-PROB is currently not distributed to users.

Table 7: Metadata fields of the output COG products

Field	Description
copyright	SEN4LDN project 2024 / Contains modified Copernicus Sentinel-2 data for the year/years {year/years} processed by the SEN4LDN consortium.



product_crs	EPSG:4326
product_grid	3x3 degrees lat/lon (WGS84) tiling grid
product_tile	{tile_id}. For example W003N21 or E120S12
version	{product version}
reference	Products DOI
time_start	Interval starting time for the products used in the generation.
time_end	Interval ending time for the products used in the generation.
creation_time	File creation time
title	Product title
product_type	Product type
license	CC-BY 4.0 - <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
legend	Product legend (if applicable, e.g. for discrete maps)
bands	Product bands (if applicable, e.g. for probabilities)

#### 4.3.4.2 Land Cover Maps (LCM)

The discrete annual LCM provide annual maps classified in 11 classes (see Table 8), defined using the Land Cover Classification System (LCCS) developed by the UN Food and Agriculture Organization (FAO). The UN-LCCS system was designed as a hierarchical classification, which allows adjusting the thematic detail of the legend to the amount of information available.

Table 8: Coding of the LCM layer and definition of the classes

Map code	Land Cover Class	LCCS code	IPCC LC class	Definition	Color Code (RGB)
10	Tree cover	A12A3 // A11A1 A24A3C1(C2)- R1(R2)	Forest land	This class includes any geographic area dominated by trees with a cover of 10% or more. Other land cover classes (shrubs and/or herbs in the understory, built-up, permanent water bodies, ...) can be present below the canopy, even with a density higher than trees. Areas planted with trees for afforestation purposes and plantations (e.g. oil palm, olive trees) are included in this class. This class also includes tree covered areas seasonally or permanently flooded with fresh water except for mangroves.	0, 100, 0
20	Shrubland	A12A4 // A11A2	Other (Shrubland)	This class includes any geographic area dominated by natural shrubs having a cover of 10% or more. Shrubs are defined as woody perennial plants with persistent and woody stems and without any defined main stem being less than 5 m tall. Trees can be present in scattered form if their cover is less than 10%. Herbaceous plants can also be present at any density. The shrub foliage can be either evergreen or deciduous.	255, 187, 34
30	Grassland	A12A2	Grassland	This class includes any geographic area dominated by natural herbaceous plants (Plants without persistent stem or shoots above ground and lacking definite firm structure): (grasslands, prairies, steppes, savannahs, pastures) with a cover of 10% or more, irrespective of different human and/or animal activities, such as: grazing, selective fire management etc. Woody plants (trees and/or shrubs) can be present assuming their cover is less than 10%. It may also contain uncultivated cropland areas (without harvest/ bare soil period) in the reference year.	255, 255, 76
40	Cropland	A11A3(A4)(A5) // A23	Cropland	Land covered with annual cropland that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation. Note that perennial woody crops will be classified as the	240, 150, 255

				appropriate tree cover or shrub land cover type. Greenhouses are considered as built-up.	
50	Built-up	B15A1	Settlement	Land covered by buildings, roads and other man-made structures such as railroads. Buildings include both residential and industrial building. Urban green (parks, sport facilities) is not included in this class. Waste dump deposits and extraction sites are considered as bare.	250, 0, 0
60	Bare / sparse vegetation	B16A1(A2) // B15A2	Other (Sparse vegetation)	Lands with exposed soil, sand, or rocks and never has more than 10 % vegetated cover during any time of the year	180, 180, 180
70	Snow and ice	B28A2(A3)	/	This class includes any geographic area covered by snow or glaciers persistently.	240, 240, 240
80	Permanent water bodies	B28A1(B1) // B27A1(B1)	Other (Water)	This class includes any geographic area covered for most of the year (more than 9 months) by water bodies: lakes, reservoirs, and rivers. Can be either fresh or salt-water bodies. In some cases the water can be frozen for part of the year (less than 9 months).	0, 100, 200
90	Herbaceous wetland	A24A2	Wetland	Land dominated by natural herbaceous vegetation (cover of 10% or more) that is permanently or regularly flooded by fresh, brackish or salt water. It excludes unvegetated sediment (see 60), swamp forests (classified as tree cover) and mangroves (see 95).	0, 150, 160
95	Mangroves	A24A3C5-R3	Forest	Taxonomically diverse, salt-tolerant tree and other plant species which thrive in intertidal zones of sheltered tropical shores, "overwash" islands, and estuaries.	0, 207, 117
100	Moss and lichen	A12A7	Other (Sparse vegetation)	Land covered with lichens and/or mosses. Lichens are composite organisms formed from the symbiotic association of fungi and algae. Mosses contain photo-autotrophic land plants without true leaves, stems, roots but with leaf-and stemlike organs.	250, 230, 160

#### 4.3.4.3 Land Cover Transitions (LCT)

Land Cover Transitions are derived from the transition matrix (Table 3) and provided using the coding as specified in Table 4 (see e.g. Figure 19).

#### 4.3.4.4 Land Cover Degradation (LCD) products

The LCD product consist of a discrete LCD map, using coding as specified in Table 9 (see e.g. Figure 24). LCD-PROB provides continuous land cover degradation probabilities, scaled between -1 (degradation) and 1 (improvement) (see e.g. Figure 21).

Table 9: LCD classes

Degradation / Improvement process	Definition	Colour	Map code
Stable	Stable Land Cover		0
Improvement	Land Cover Improvement as defined by Table 4		1
Degradation	Land Cover Degradation as defined by Table 4		2

## 4.4 Summary of algorithm validation

With independent validation, the SEN4LDN land cover maps showed varied performances in the three demonstration countries, namely Uganda, Portugal, and Colombia. The LC map achieved the highest overall accuracy in Colombia ( $90.1\% \pm 3.4\%$ ), while the lowest accuracy in Uganda ( $69.6\% \pm 5.5\%$ ). In terms of the land cover classes, wetlands (i.e., open water and wetland herbaceous vegetation), trees, and low

vegetation were mapped with higher accuracies among the IPCC classes. However, the SEN4LDN LC map had considerable confusion between low vegetation (i.e., grass and shrubs), trees and crops in Uganda, which is also a common issue in global land cover products [37].

With an independent and direct validation of the land cover change map in Uganda, the SEN4LDN LCC product achieved an overall accuracy of 73.7% at the change vs. no change level, and 72.9% when considering the specific transition classes. The LCC map in Uganda had a good performance in detecting change related to forest, i.e., deforestation and reforestation, as well as stable/unlikely changes. However, the map underestimated considerable changes as well. By comparing SEN4LDN LCC maps with the national LC map-derived changes in Portugal and Uganda, it is observed that SEN4LDN tends to be more conservative in predicting change areas in Portugal compared to the COSc product, while its estimation of change areas in Colombia is more logical compared to the MapBiomass product.

## 4.5 Conclusions, limitations and recommendations

The development of the land cover classification system within the framework of the SEN4LDN project marks a significant advancement over the previous ESA WorldCover system. The primary objective was to enhance the existing classification by addressing its limitations, particularly in ensuring consistency and accuracy across multiple years.

The newly developed algorithm addresses the problem of yearly instability in predictions, which was a major limitation of the previous system. This instability derived from noise in the training data, landscape complexity, heterogeneity, and varied meteorological conditions that influenced the behaviour of vegetated classes across different years. Such factors led to unstable input features, making it challenging for the model to be consistent across different years. The creation of a multi-year training dataset, with an optimized spatial stratification based on the land cover, and the AgERA5 Climatic Regions Embeddings, have successfully mitigated these issues, enabling the algorithm to be robust on input features from different years.

Moreover, the deep learning component of the algorithm offers enhanced generalization capabilities, and with the integration of contextual information, allows a more effective differentiation between classes that typically show high confusion when using pixel classifiers. The inclusion of change detection and post-processing modules further refines the predictions, ensuring greater consistency in land cover maps across years in the absence of true change.

Despite these advancements, the quality of the input data remains a critical factor in the accuracy of land cover products. Residual cloud contamination and artifacts in the Sentinel-2 L2A data have historically posed significant challenges, leading to noise and errors in the final outputs, particularly in regions with persistent cloud cover. To address these issues, the new system incorporates an optimized cloud detection algorithm designed to mitigate the effects of undetected or residual clouds. This enhancement reduces noise and artifacts in the input data, resulting in cleaner and more reliable input datasets.

The previous version of the algorithm, as used in WorldCover, was meticulously calibrated for single-year applications, both in its training approach and post-processing steps. This made it unsuitable for degradation detection and operational tasks. The current iteration, however, delivers robust and consistent predictions across multiple years while utilizing a streamlined and generalized post-processing routine. This enhancement makes it suitable for generating degradation products and enables its

application in an operational context, supported by its implementation in openEO, which allows users to run the algorithm on their areas of interest.

A significant advancement was also made with the development of continuous transitions and degradation probabilities. Traditional approaches, which relied on directly comparing discrete land cover maps, were prone to producing a large number of false positives due to noise in the discrete maps across different years. By leveraging land cover probabilities, all the information generated by the classifier contributes to the creation of transition maps, resulting in more accurate and refined transition and degradation products. This additional information is also made available to users in the form of Land Cover Degradation Probabilities, enabling them to perform custom and advanced downstream analyses.

Finally, it is worth noting as certain land cover transition processes remain challenging to classify accurately. For instance, the algorithm will classify plantations—such as palm oil, coffee, sugar cane, tea, cacao, and bananas—as ‘tree cover’. As a result, the conversion from any land cover type to plantations could potentially be labelled as vegetation establishment or reforestation. This raises questions about the ecological implications of such classifications and highlights the need for ongoing refinement in this area of the land cover classification system.



## 5 Trends in land productivity

### 5.1 Outline

The aim of this component is to develop an automated method to monitor trends in land productivity dynamics at high spatial resolution. The biological production provides a fundamental resource in the form of food, fodder, fuel and fibre for sustaining human economy and wellbeing. Negative trends in this production can be an indicator of land degradation. Land productivity can be estimated from satellite-observed vegetation index data over growing seasons, summarized into annual values. Frequently, the Normalized Difference Vegetation Index (NDVI) is used as a proxy for this plant growth. However, since this index saturates over high biomass areas, we are here assessing also alternative indices with better dynamic range.

Optical observations from satellite are constrained in cloudy areas. Since this limits data coverage and frequency, a data processing method that is robust to noise and short-term variability is necessary. Therefore, while it is possible to develop trend estimates directly from raw satellite sensor data, we here first implement robust data processing methodology to handle gaps in time series, then aggregate data seasonally, and finally derive trends from the aggregated data.

For data aggregation and smoothing, we are using the TIMESAT software system. TIMESAT has been used since ca. 2000 [33] for pre-processing of satellite time-series, and was recently upgraded to work with high-spatial resolution data [34]. TIMESAT is the core algorithm in the High-Resolution Vegetation Phenology and Productivity (HR-VPP) portfolio of the Copernicus Land Monitoring System (CLMS)<sup>8</sup>. The software system fits mathematical functions over time to each pixel location of a stack of vegetation index data to generate smooth seasonal trajectories. From these trajectories, seasonal parameters are derived that describe the vegetation phenology and productivity. Vegetation productivity is the aggregated smoothed seasonal vegetation index across the growing season. We focus on changes in the sum of vegetation productivity between years.

Apart from deriving the trend in the annual productivity parameters, we generate data on the productivity performance, which is a measure of the productivity of a particular land area in comparison to other land areas with similar characteristics. Finally, we generate an aggregate measure termed Land Productivity Degradation that combines aspect of trend and performance. The methodology in this chapter broadly follows the suggestions in the UNCCD Best Practice Guidance [6], but with necessary modifications due to the use of high-resolution data.

### 5.2 Background

Productivity is one of several phenological parameters, derived from the smooth seasonal curve of the vegetation indicators (VI) that represents the seasonal development of photosynthesizing canopy foliage. The seasonal curve starts at the early-season onset, passes the mid-season maximum and ends at the late-season growth cessation. The productivity is thus the sum of VI data from the start of the season to the

---

<sup>8</sup> <https://www.eea.europa.eu/en/datahub/datahubitem-view/b6cc3b37-0686-4bb1-b8c3-3c08520743c3>

end of the season (Figure 25). By employing a consistent methodology for deriving these parameters, data can be used for identifying trends and variations in the productivity, thus, enabling conclusions about possible ongoing degradation or recovery that may signal land degradation.

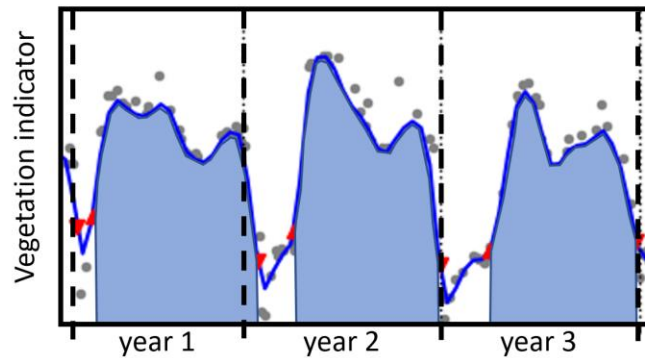


Figure 25: Productivity for three years of data illustrated with the light blue areas. The grey points are vegetation index data from Sentinel-2 observations, and the blue line is the smooth seasonal vegetation curve estimated with a cubic spline function.

The algorithm for phenological parameter extraction is based on the methods developed for the CLMS HR-VPP service, with modifications developed during the generation of coarse-resolution global land surface phenology in CLMS. The methodology is based on principles in [34], in which TIMESAT was modified from its original version to handle irregular data from high-spatial resolution sensors.

Land productivity, and losses of productivity in connection with land degradation, can be estimated based on the **trend**, **state** and **performance** of vegetation productivity [6,35,36]. The trend measures the rate and direction of change of land productivity over a time period. The state compares the productivity to historical productivity, and the performance compares the local productivity to similar land units over a large area.

In SEN4LDN, the **trend** is estimated for the period 2018-2023 at 10 x 10 m spatial resolution. The resulting map product displays the value of the slope coefficient of this trend, thus the rate of change. Apart from the trend values, a map of classes of trend values is generated that shows areas that are degrading (significant negative trend), stable (no significant trend) or improving (significant positive trend). We use the non-parametric Theil-Sen estimator to derive the trend and the Kendall-Tau statistic to derive the significance of the trend, acknowledging the need for robustness when fitting to short time-series.

A **state indicator** would depend on comparing current productivity with historical values. However, no historical satellite data of the same spatial resolution exist. While theoretically, historical estimates could be derived from low-resolution data (kilometer scale), comparing trends at such different resolutions would be misleading, and therefore the state indicator has not been computed in SEN4LDN.

The **performance** indicator is computed by comparing productivity at local pixel level with the average over similar land cover areas across the whole country.

A joint sub-indicator, **land productivity degradation**, that combines trend and performance is also developed.

All the indicators are described in further detail below.

## 5.3 Algorithm

### 5.3.1 Summary

Land productivity parameters are generated from the same workflow on Sentinel-2 TOC data as used for the trends in land cover. Figure 26 outlines the general methodology. Following pre-processing of the data, vegetation indices (VIs) are calculated. Then, time-series for each pixel are fed into TIMESAT for fitting smooth functions and outputting phenological parameters. Out of these, productivity is used as the main parameter in the ensuing process. Trend estimation and performance estimation are done for each pixel, generating sub-indicators Trend and Performance. Finally, an aggregated sub-indicator, Land Productivity Degradation (LPD), is developed.

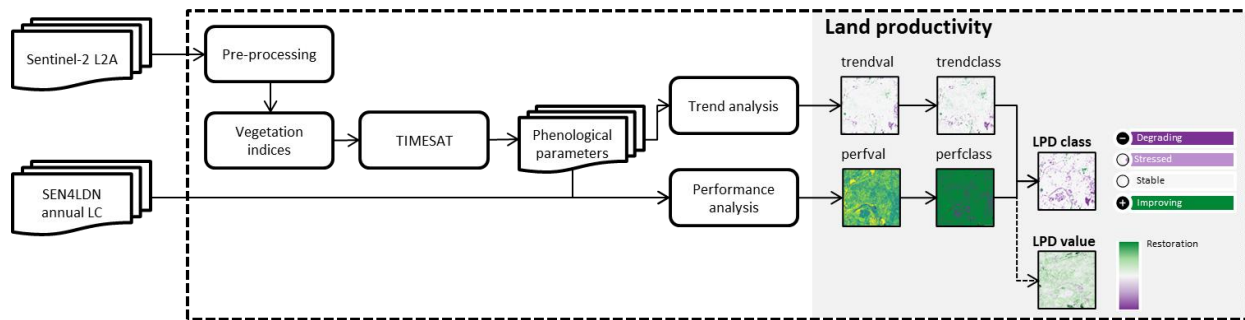


Figure 26: Overview of the workflow for the trends in land productivity sub-indicator

### 5.3.2 Algorithm input variables

During the algorithm development, several VIs (see Annex B) from the input Sentinel-2 TOC data were analysed by cross-correlating the VIs against gross primary productivity (GPP) data from eddy covariance towers. The tested vegetation indicators are Leaf Area Index (LAI), fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Normalized Difference Vegetation Index (NDVI), the 2-band Enhanced Vegetation Index (EVI2), and the Plant Phenology Index (PPI). Details about the results of the benchmarking through correlation analysis are shown in Annex C. The outcome of the analysis was to proceed with the 2-band Enhanced Vegetation Index (EVI2), described below. Quality flags propagated from Sen2COR Scene Classification Layer (SCL) and cloud mask (the same as in the generation of HR-VPP) are included to weight down data points of poor quality.

#### 5.3.2.1 2-band Enhanced Vegetation Index (EVI2)

EVI2 [37] is a vegetation index related to the Enhanced Vegetation Index [38] which was derived to be less sensitive than NDVI to the influence of atmosphere and background soil. EVI and EVI2 are also more dynamic than the NDVI at high biomass levels.

The formula of EVI2 is

$$EVI2 = 2.5 \frac{DVI}{NIR + 2.4 * red + 1}$$

Eq. 8

where DVI is the difference vegetation index (NIR-red). The index is used in several phenology products, e.g., in the Landsat MuSLI Land Surface Phenology product over the USA [39] and in the 500-m Global Land Surface Phenology (GLSP) product based on VIIRS data [40].

#### 5.3.2.2 Land cover

The annual Land Cover Map (see Chapter 4) from 2021 was used as one of the inputs for generating the performance sub-indicator.

### 5.3.3 Algorithm description

#### 5.3.3.1 Data processing using TIMESAT

The methodology for extraction smooth seasonal time-series curves follows previous developments for the European CLMS HR-VPP phenology estimation, with modifications introduced for generating global land surface phenology in CLMS.

TIMESAT ingests EVI2 time-series data for each pixel location, which may include some noise and temporal gaps. The first step is the pre-processing, consisting of two parts: removing outliers and extending the time-series. Based on the number of clear-sky data points, TIMESAT will detect whether there are enough good-quality data to proceed with the detailed fitting operation. Growing seasons are determined using sinusoidal functions making a preliminary analysis of seasonality (number of seasons) and to establish the rough timing of each growing season. For the detailed seasonal data fitting, we chose one of the methods available in TIMESAT: cubic spline functions. This method uses piecewise cubic polynomials to fit time series segments, and this was also the method applied for generating global land surface phenology in CLMS. Cubic splines model time-series data in a flexible manner and have also been used with e.g. Landsat data for phenological data extraction [41]. Finally, a search algorithm detects the points of the start and end of seasons based on a chosen amplitude threshold.

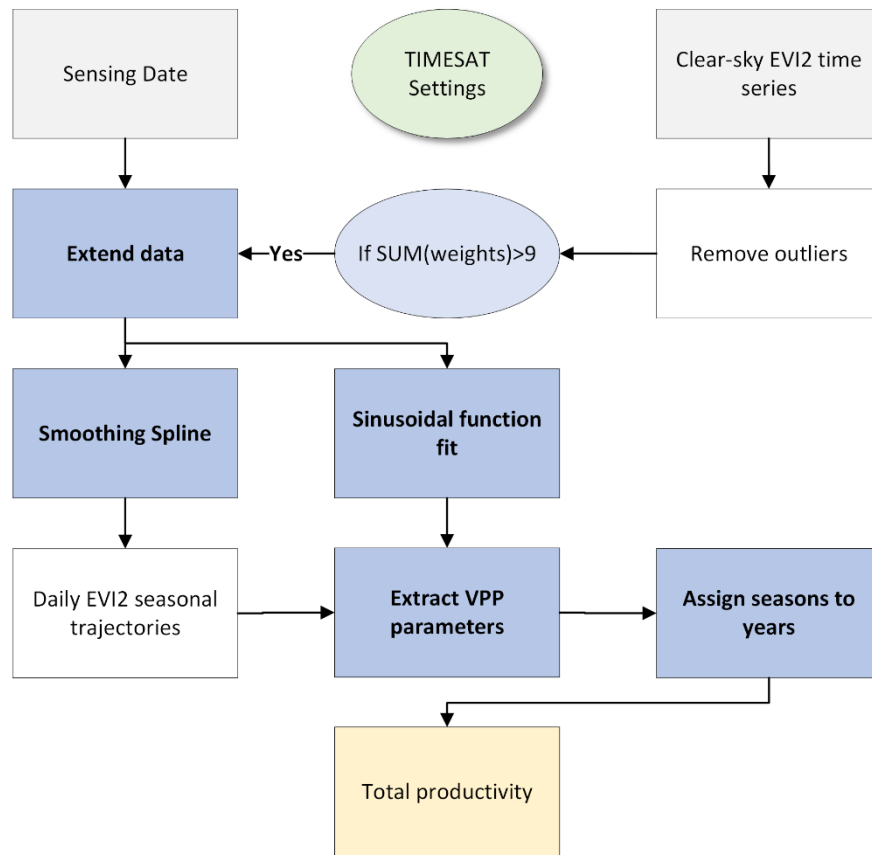


Figure 27: TIMESAT4 processing module in SEN4LDN.

To identify outliers in the dataset, a smoothing spline with a high smoothing parameter of 10,000 was applied initially to smooth the data. Then, both the global median difference (GD) and local median difference (LD) were calculated by comparing the original data to the smoothed curve, where the LD was determined using a 91-day moving window. A data point was flagged as an outlier if the deviation from the smoothed data exceeded six times the values of both GD and LD.

TIMESAT 4 checks for the presence of sufficient high-quality data in the time series by ensuring that the sum of the weights is greater than 9 over the whole time series. This check is crucial to guarantee that there is enough reliable data available to perform the fitting of double logistic functions.

To extract phenological details for the initial and final seasons from the time series, TIMESAT 4 extended the series by duplicating the data from the first and last years. The procedure is the same as for the HR-VPP product [33].

### Constructing the seasonal trajectory

Following the preprocessing, the EVI2 time series is fitted with a cubic smoothing spline  $S_p(t)$  [42], which minimizes the value of a criterion function  $C_p$

$$C_p = \sum_{i=1}^n \left\{ w_i [y_i - S_p(t_i)]^2 + p \int_{-\infty}^{+\infty} |S_p''(t)|^2 dt \right\},$$

Eq. 9

where  $t_i, i = 1, \dots, n$  is time vector and  $y_i$  are corresponding EVI2 values. Each point in the time series is associated with an input weight  $w_i$ . The smoothing parameter  $p$  controls the shape of the spline, varying from an exactly interpolating spline ( $p = 0$ ) to a straight line ( $p \rightarrow \infty$ ). The  $p$  value is set to 1,000, which can reduce the impact of noise while preserving sufficient local variation. Figure 28 shows an example of a pixel location with spline-fitted seasonal trajectory. To extract the phenological information of the first and last seasons from the time series accurately, TIMESAT extends the length of the time series by duplicating the data of the first and last years [33].

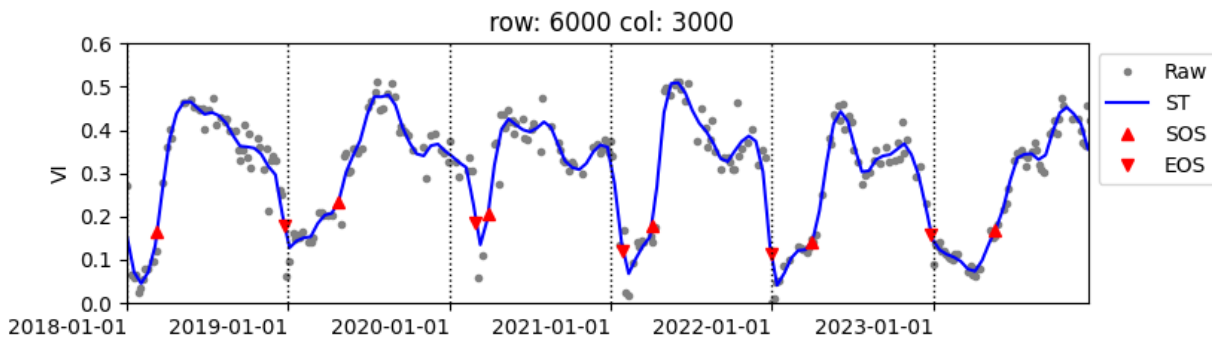


Figure 28: Time-series across six years of data. The blue line is the seasonal trajectory (ST), and the raw data values (grey points) represent EVI2 computed from Sentinel-2 TOC reflectance. Red arrows represent seasonal start and end dates.

### Defining the number of growing seasons

A first estimate of the growing seasons' start and stop dates ("coarse seasons") is determined by applying a sinusoidal function fitting to each pixel's time series:

$$f(t) = c_1 + c_2 \sin(\omega t) + c_3 \cos(\omega t) + c_4 \sin(2\omega t) + c_5 \cos(2\omega t),$$

Eq. 10

where  $\omega = 2\pi/n$ , and  $n$  is the number of points,  $f(t)$  is the corresponding PPI value at time  $t$ . Coarse seasons are used to identify one or two seasons and to determine the search period for each growing season.

### Extraction of phenology and productivity parameters

In TIMESAT, thirteen VPP parameters and associated quality values can be generated. In Sen4LDN only the parameter *total productivity* (TPROD) is used, as it captures the total seasonal productivity of the EVI2 time series.

The productivity, TPROD, is defined for the period between the start and end of each season, which calculates the integral of the vegetation index values counting from the zero level (Figure 29). TIMESAT determines the timing of start of season (SOSD) and end of season (EOSD) by applying a threshold value in relation to the seasonal amplitude. Following results from an extensive analysis during the HR-VPP project [43], we selected these thresholds to be 25% of the amplitude for SOSD and 15% for EOSD.

In certain cases, TIMESAT outputs missing data for a pixel. This occurs when the number of missing data points is less than 1% of the total number of points. It also occurs when the seasonality is extremely weak, which could occur in areas of evergreen vegetation. Furthermore, if EVI2 values are extremely low (e.g. for water surfaces), values are filtered out.

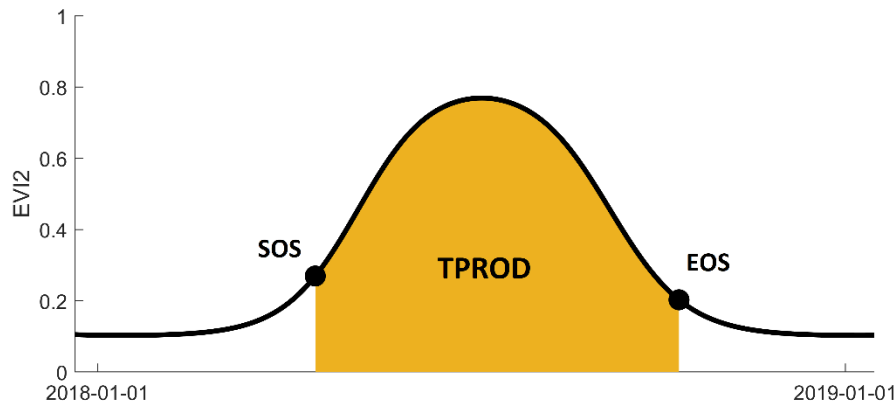


Figure 29: Definition of productivity parameters SOS (start of season), EOS (end of season), and TPROD (total productivity).

While several productivity parameters are generated in TIMESAT<sup>9</sup>, in Sen4LDN the choice to use TPROD was based on the following reasons: SPROD (seasonal productivity) may only capture a fraction of the total yearly productivity in the case of weak seasonality, e.g. over evergreen tropical forests. AMP (amplitude) is more sensitive to random variations since it is controlled by extreme values that may fluctuate due to noise. LENGTH (length of the season) only carries useful information on the productivity in areas where the amplitudes are similar. Consequently, TPROD is selected as it combines the qualities of all the other parameters and is robust against noise.

### Assigning seasons to years

Seasons may traverse years, thus a method for assigning them to the correct year is necessary. Seasons are assigned to years based on the peak of coarse seasons. For example, the peak of a coarse season (from sinusoidal functions) is located on 2021-12-31, then the corresponding season (from smoothing spline) will be in 2021 even the peak of the corresponding season is located on 2022-01-15. Maximum two seasons are stored in a year. The two seasons in a year are present chronologically, irrespective of their relative size. Thus, the main season may occur in season 2 and a minor season in season 1.

### 5.3.3.2 Trend estimation

Productivity data (TPROD) have been computed from EVI2 for each 10 x 10 m<sup>2</sup> Sentinel-2 pixel for all growing seasons between the start year (2018) and the end year (2023) (Figure 30). We only compute the trend value for pixels that have an EVI2 value for all the years.

To compute the trend, we have used the Theil-Sen robust statistical regression method that computes the median change over time, and the Kendall-Tau estimator of the Mann-Kendall test to investigate the

<sup>9</sup> <https://land.copernicus.eu/en/technical-library/product-user-manual-land-surface-phenology-300-m-version-1.0/@download/file>



statistical probability of finding a trend [44,45]. This non-parametric method has the advantage over ordinary least-squares regression in being robust to outliers and not requiring normally distributed data [46].

The p-value of the Kendall test [47,48] indicates the probability of no trend. Thus, a high p-value indicates that the change is not significant and a low p-value that the trend is significant. Since the number of data points ( $n$ ) is always constant between pixels (all years present), the p-value also reflects the slope value of the trend. We use the p-value to assign pixels a slope class according to Table 10.

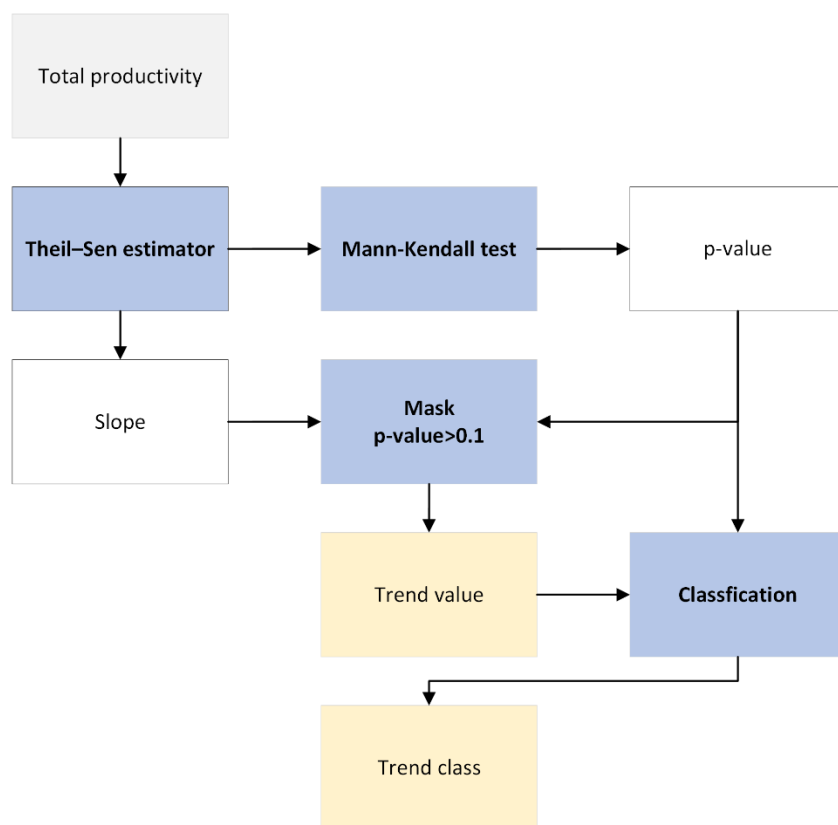


Figure 30: Workflow of the trend estimation module

Table 10: Classification criteria for assigning pixels to slope classes.

Class name	Criteria	Class id
Degrading	$\text{Slope} < 0$ and $p \leq 0.1$	1
Stable	$\text{Slope} = 0$ or $p > 0.1$	2
Improving	$\text{Slope} > 0$ and $p \leq 0.1$	3

### 5.3.3.3 Performance estimation

The performance indicator ( $PP$ ) is based on computing the TPROD value for each pixel in relation to the productivity across a country for the same land cover class as the pixel (Figure 31). The pixel TPROD is derived for the period 2021-2023, and the TPROD across area is defined per land cover class and for the

full period 2018-2021. Thus, the performance indicator provides a measure of how an area compares in productivity to other areas across the analyzed zone. However, there is also an element of change since the indicator focuses the recent years compared to a longer period:

$$PP = \frac{TPROD_{max,p}}{TPROD_{ref,c}},$$

Eq. 11

where  $PP$  is productivity performance,  $TPROD_{max,p}$  is the maximum 2021-2023 TPROD value for a pixel belonging to a certain land cover class, and  $TPROD_{ref,c}$  is the reference TPROD value for all pixels of the same stable land cover class in the country. To reduce uncertainties from land cover map, we use the 90'th upper percentile to determine the national maximum TPROD of each land cover class.

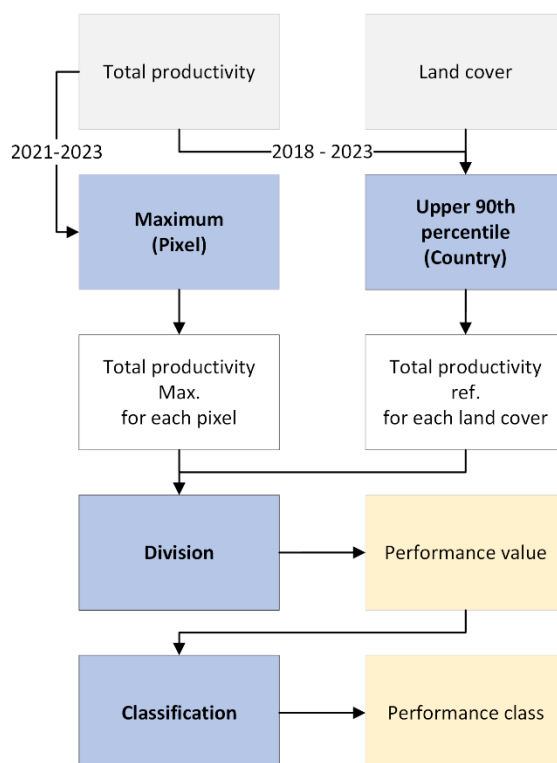


Figure 31: Workflow of the performance estimation module

PP is categorized into the classes *Degrading* and *Stable* following recommendations from the UNCCD Best Practice Guidance [35], Table 11.

Table 11: Classification criteria for productivity performance (PP)

Class name	Criteria	Class id
Degrading	0.0 – 0.5	1
Stable	> 0.5	2

### 5.3.3.4 Land Productivity Degradation

Land Productivity Degradation (LPD) is a joint sub-indicator combining Trend and Performance according to similar logic and nomenclature as in the UNCCD Good Practice Guidance [6], with the exception that no State indicator is available as input. The methodology for calculating LPD integrates trend analysis, performance evaluation, and classification into a cohesive process. The following steps outline the core components of the LPD calculation algorithm (Figure 32 and Figure 33).

#### *LPD class calculation*

The data preparation process involves inputting data that includes both Performance and Trend classifications (Figure 32). Once the data is prepared, the Land Productivity Dynamics (LPD) values are calculated and subsequently categorized into specific degradation classes based on a predefined lookup table, as detailed in Table 12. This table shows the combinations of Trend and Performance classifications to determine the final LPD classes, categorizing them as Degrading, Stable, or Improving. Where the Trend indicates Stable or Improving but the Performance shows Degrading, the resulting classes are labelled as Stressed.

Table 12: Lookup table for codes defining land productivity degradation classes.

<b>Performance class \ Trend class</b>	<b>Degrading</b>	<b>Stable</b>	<b>Improving</b>
<b>Degrading</b>	Degrading	Stressed	Stressed
<b>Stable</b>	Degrading	Stable	Improving

A 5x5 weighted spatial filtering window is applied to the LPD product. Since the third class (Stable) occupies the majority of the product's area, its weight is reduced to 0.5. This approach not only reduces spatial noise, ensuring that the final product appears continuous, but also highlights areas of interest requiring attention.

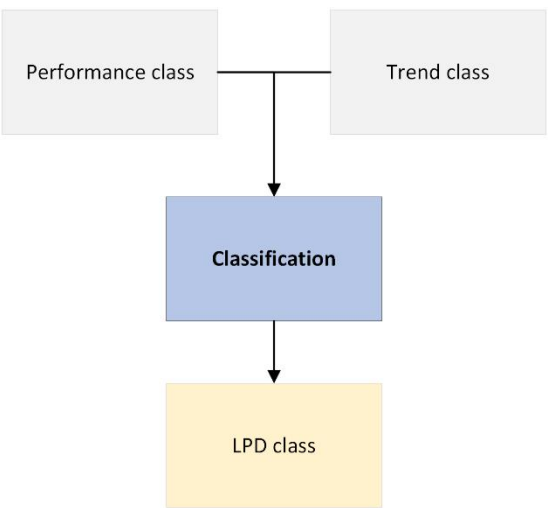


Figure 32: Workflow for Land Productivity Degradation (LPD) class calculation.

### LPD value calculation

This sub-indicator is generated as an intermediate dataset as input to the integration into a single land degradation indicator, as described in chapter 7.

1) Data preparation and Input:

Input data contains *total productivity* and *performance value*. Instead of using *trend value* directly, the slope of total productivity from trend estimation (5.3.3.2) is utilized. This approach allows for the inclusion of all observed trends, even those typically masked due to a  $p\text{-value} > 0.1$ , to ensure a more comprehensive evaluation of land productivity changes.

2) Normalization through linear stretching:

Slope normalization: The slope values (range: -10 to 10) and performance values (range: 0 to 1) are normalized through a linear stretching process, which scales the data to the range -1 to 1.

3) LPD value calculation:

In order to generate a continuous value that combines trend and performance we compute LPD values. It can be regarded as an intermediate value, derived to provide a basis for a continuous integrated indicator (see Chapter 7). The normalized slope and performance values are combined to calculate the LPD index. This combination is achieved by averaging the two normalized values to produce a single index value for each land cover type.

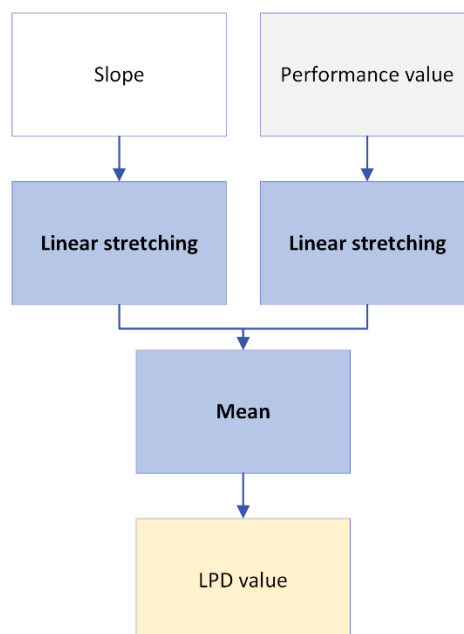


Figure 33: Workflow for Land Productivity Degradation (LPD) value calculation

### 5.3.4 Algorithm output variables

The datasets in Table 13 are generated as part of the Productivity Trends sub-indicator.

Table 13: Datasets of land productivity

Name	Sub-Indicator of land productivity	Content	Data type and range	No data value	Scale Offset
trendval	Trend	Values of trend coefficient of productivity over the period 2018-2023 ( $\text{day} \cdot \text{year}^{-1}$ ) [-10, 10]	UINT8	255	1/10 -10
trendclass	Trend	Classes indicating trend / no-trend: Degrading (1), Stable (2), Improving (3)	UINT8 [1,2,3]	0	-
perfval	Performance	Maximum performance 2021-23 over the land cover class reference [0.0 – 2.0]	UINT8	255	1/100 0
perfclass	Performance	Classes of performance indicating Degradation or No degradation	UINT8 [1,2]	0	-
LPD	Land productivity degradation	Classes of degradation or no degradation by combining slope and performance: No data (0), Degrading (1), Stressed (2), Stable (3), Improving (4)	UINT8 [1,2,3,4]	0	-
LPD Index*	Land productivity degradation	Continuous values of land productivity degradation [-1 - 1]	UINT8	255	1/10 -1

\* The LPD Index sub-indicator is currently not available to users.

## 5.4 Summary of algorithm validation

Validation focuses the phenological parameters and the productivity trend parameter. Due to the lack of objective validation data on performance and land productivity degradation no direct validation of these parameters was carried out. The annual productivity data were indirectly verified by comparison against outputs from the HR-VPP product.

To validate trend, statistical testing of differences in slope values for trends in GPP (estimated from flux tower data) and satellite-derived productivity is done. If the confidence intervals of these two trends overlap, the hypothesis that the slopes are different is rejected. Both data are standardized before trend computation to make them comparable. Furthermore, checks for completeness and cross-tile consistency are performed.

Independent validation of the SEN4LDN output products on trends in land productivity is based on visual inspection, internal consistency analysis and qualitative indirect validation with external data. Visual inspection indicated no important issues related to spatial consistency, except the effect of persistent cloud coverage over some areas in Colombia and – to a lesser extent – Uganda. This however does not result in erroneous identification of degraded areas in the LPD product. Internal consistency of the products was evaluated based on intercomparison and error evaluation of the products of adjacent tiles on a sample in Sentinel-2 tile overlap area. Very high consistency is found for the products over Portugal and Uganda. The results for Colombia are slightly less good, probably related to larger uncertainties in the atmospheric correction, more pronounced topography, and higher cloud coverage. Indirect validation was done through qualitative cross-comparison with an external product from the Copernicus Land Monitoring Service (CLMS) Gross Dry Matter Productivity (GDMP) 300m to evaluate the temporal consistency of interannual temporal variation. Although there is a large discrepancy in spatial resolution between the SEN4LDN products (10 m) and the CLMS GDMP 300m product, in most cases a good

agreement can be found between the temporal evolution of GDMP, the temporal evolution of TPROD and the derived *Trendval* and *Trendclass*.

## 5.5 Conclusions, limitations and recommendations

### 5.5.1 Conclusions

The productivity trend sub-indicators in SEN4LDN meet the user requirements with respect to spatial and temporal resolutions, spatial coverage, output reference system, delivery mode, data format, and free and open access. In Colombia, rather large areas of highland rain forest are affected by persistent cloud cover, resulting in missing data for these areas. For temporal extent, the products meet the minimum requirement; however, the rather short time span imposes limitations. The early adopters of the project particularly highlighted the benefit of high spatial resolution of the data in comparison to previous products. Limitations to the accuracy of the products still exist, as outlined below, and the degree to which the data fulfils national requirements is more uncertain. The products will need to undergo a more thorough evaluation by national expertise to evaluate their usefulness in operational land degradation planning and monitoring.

### 5.5.2 Constraints/limitations

Satellite observations are affected by numerous factors that reduce reliability (effects of residual cloud contamination, different viewing and solar angles, atmospheric aerosols, moisture variations, geometric errors, mixed target areas etc.). This adds variability to the data, which means that weak trends may be difficult to distinguish. Similarly, random variations can lead to false trends in the data. It is important to realize that a state or trend in a single isolated pixel may not be trustworthy, thus, emphasis of the interpretation should be placed on areas where observations are spatially consistent across several pixels.

**Productivity Trend.** The period for estimating trends is short, and extreme values may have large impacts on the estimates. While we use a robust statistical method to reduce the impact of single outliers, the trend maps and the combined products should be interpreted with caution. Since a rather strong trend is needed to achieve statistical significance for this short time series, we present trends at p-values of 0.10 rather than the more commonly used 0.05 level. In agricultural systems with crop rotation, the observed trend may not indicate degradation but could be due to variations in crop types. Management factors (e.g. irrigation) may also affect vegetation productivity in ways not related to land degradation.

**Productivity State.** This sub-indicator is defined in the SDG 15.3.1 Good Practice Guidance as the mean of the most recent three years of NPP in comparison to the preceding 13 years. This is not possible to derive with the short time-series of Sentinel-2, and productivity state is thus not included in Sen4LDN.

**Productivity Performance.** This is defined in the SDG 15.3.1 Good Practice Guidance as the local plant productivity relative to other regions with similar productivity potential. The computation of the maximum productivity across areas ( $TPROD_{max,a}$ ) is strongly dependent on the definition of LCEUs (land cover and ecosystem functional units), which can be defined based on e.g. land cover, soil type, climate conditions, elevation and aspect. Most of these ancillary data are only globally available at coarse spatial resolution. For some countries they may exist at high spatial resolution, but due to practical reasons and to generate harmonized output products for the pilot countries we have not utilized national datasets.

Thus, performance is only based on land cover, as this input variable is available at high spatial resolution globally. Since similar land cover can exist for areas with quite dissimilar environmental conditions across a country, the national  $TPROD_{max,a}$  may not represent typical regional growing conditions, particularly for large countries. A further weakness of the performance indicator is the relatively short time-span for the indicator. In the Good Practice Guidance [35] it is recommended to use data across 15 years, but we only have access to six years of data. In areas with strong variability of conditions, e.g. croplands, the baseline may not represent a good average. It should also be noted that errors in land cover classification will propagate to the performance indicator, however, by using the 90'th percentile rather than the actual maximum value, we have reduced the effect of these potential misclassification errors.

**Land productivity degradation (LPD).** This sub-indicator combines classifications of trend and performance and thus depends on the accuracy of these classifications. The classification of trend is based on p-values, which can be considered as an objective approach. However, the classification of performance is based on a rather arbitrary threshold value of 0.5, and it is also affected by land cover classification errors. As errors in both these input parameters will propagate to the LPD parameter, LPD is an output that should be used with caution.

While we have used state-of-the-art methodology for selection of satellite-derived parameters and processing of data, the analysis is based on a series of assumptions and methodological choices. Consequently, the different choices may affect the final output products. In conclusion, all the land productivity parameters are to some degree affected by data errors and processing uncertainties and should be used with caution. National and local knowledge is vital for assessing the final outputs.

### 5.5.3 Recommendations

The fidelity of the productivity trend estimates will increase as time-series of Sentinel-2 data become longer. The analysis will also benefit from an increased accuracy of the land cover classification. Improved base data to generate the LCEUs that can express local geographical productivity at high spatial resolution will be beneficial for future estimates of performance. The principle for analysing and combining the trend and performance to a joint sub-indicator will benefit from thorough review and analysis involving national land degradation expertise.



## 6 Trends in carbon stocks

### 6.1 Outline

The aim of this component is to develop an automated method to monitor carbon stocks dynamics at high spatial resolution. Currently, carbon stocks are estimated using a combination of land cover and SOC data within the Trends.Earth application. This means that only changes in SOC are considered for LDN monitoring, which is, however, difficult to observe with multi-spectral data alone (such as Sentinel-2) due to coverage of the soil with vegetation and lack of spectral information (imaging spectroscopy is better suited for SOC prediction). In addition, SOC storage is highly variable across sites depending on soil-forming factors, including time, parent material, topography, climate, organisms and impact by humans. Accurate predictions of SOC stocks require therefore detailed data on site properties and management, which are often not available. Those factors inhibit the establishment of SOC time series preventing the estimation of trends that would be needed. Hence, we propose the exploration of above-ground biomass (AGB) maps and derived trends as a first feasibility study in this context.

There are two research demonstration options available:

- 1) European Space Agency's Climate Change Initiative (CCI) offers a dataset of global AGB maps. This collection offers forest AGB estimates for the years 2010, 2017, 2018, 2019, and 2020. Maps are provided at a spatial resolution of 100 m.
- 2) The Word Resources Institute (WRI) carbon flux model incorporates spatial datasets to characterize carbon emissions and removals based on activity data (forest gains and losses) combined with biomass maps information.

### 6.2 Background

Relying on multi-temporal AGB maps allows for the direct estimation of biomass change and carbon emissions. Therefore,  $\Delta\text{AGB}$  provides an assessment of trends in carbon stocks, which can be affected by land-use conversions but also dynamics within intact ecosystems (forests), such as degradation, regrowth, management activities, and natural disturbances [41].

AGB has been defined as the Essential Climate Variable (ECV) T12. In the context of ECVs, AGB is generally defined as the mass of live organic matter above the soil in terrestrial vegetation, thus excluding roots, litter, and dead wood [50]. ECV AGB is quantified in tons of dry matter, while biomass itself is measured in  $\text{g/m}^2$  (dry matter) or  $\text{mg/ha}$ . Carbon can then be conventionally converted from dry weight with the factor of 0.47-0.5. Information of uncertainty associated with AGB estimates is also crucial, and can involve various metrics, including relative and absolute systematic deviation, confidence intervals, and root-mean-square error (RMSE). These uncertainties are calculated for the overall biomass estimate and for individual biomass classes/ranges, leveraging higher-quality reference data where available [41].

The availability of satellite-based AGB and AGB change ( $\Delta\text{AGB}$ ) products has increased in recent years [52]: Estimation relies on remote sensing data acquired across the electromagnetic spectrum, encompassing visible light to microwaves. For instance, optical imagery reveals information about vegetation cover and greenness, enabling regional-scale AGB mapping with moderate accuracy. Microwave remote sensing

penetrates through clouds and vegetation layers, providing valuable information about biomass in dense forests and areas with limited accessibility. LiDAR data excels at capturing tree height and canopy structure, offering precise estimates of AGB in individual trees or small forest patches.

## 6.3 Algorithm

### 6.3.1 Summary

Two research demonstration approaches are used for mapping changes in AGB based on Earth observation data and combined [52]:

- **Stock change approach:** This method directly calculates  $\Delta\text{AGB}$  by differencing biomass maps created at different points in time.
- **Gain-loss approach:** This approach focuses on land-use specific carbon fluxes, utilizing pre-determined emission and removal factors to estimate  $\Delta\text{AGB}$  from an initial biomass estimate for a specific timeframe.
- **Hybrid approach:** The final optimal solution is to average the resulting maps from stock change and gain-loss methods, and the resulting standard deviation is considered as absolute uncertainty measure. See also Figure 34.

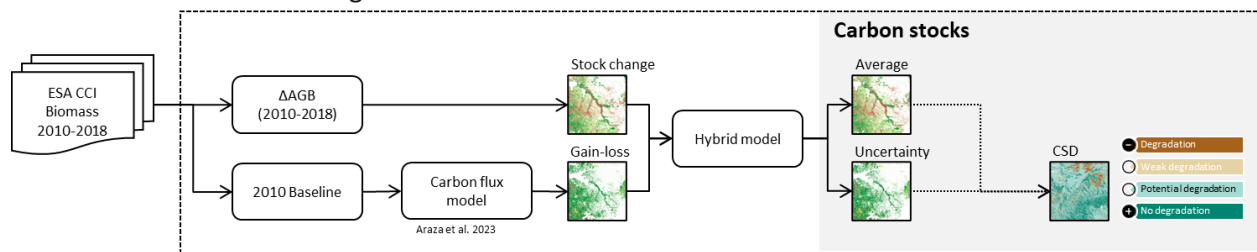


Figure 34 Overview of the workflow for the trends in carbon stocks sub-indicator

In the following section, the input datasets and algorithms that are applied are described in more detail.

### 6.3.2 Algorithm input variables

The ESA CCI Biomass data sets [53] are freely available from the Centre for Environmental Data Analysis (CEDA)<sup>10</sup>. The current release of the CCI BIOMASS data products (v4) consist of global maps of:

- **AGB** (dry mass per unit area, units: Mg ha<sup>-1</sup>) for the years 2010, 2017, 2018, 2019 and 2020, together with per-pixel standard deviation with a pixel size of 1 ha, i.e., for a 100 m x 100 m large area.
- **AGB change** expressed as the difference of AGB maps for two consecutive years (2018-2017, 2019-2018 and 2020-2019) and for a decade (2018-2010), together with an estimate of their standard deviation and a quality flag map, detailing the level of reliability of the AGB change estimate.

<sup>10</sup> <https://catalogue.ceda.ac.uk/uuid/af60720c1e404a9e9d2c145d2b2ead4e>

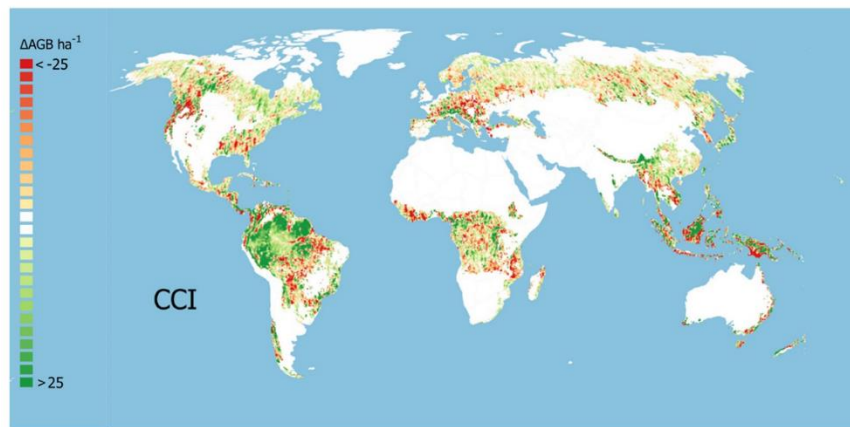


Figure 35: Example of AGB change mapping using ESA CCI dataset from 2010 as baseline and calculating change to 2018. Adapted from [52].

In the WRI carbon flux model, we incorporated three key datasets specific to 2010 to initialize the model. First, the CCI AGB map from 2010 providing the initial values for the amount of carbon stored in the above-ground parts of vegetation within the modelled area. A primary forest map helped to define areas considered to be primary forests, which play a crucial role in global carbon cycling. Finally, a tree cover map was included serving as a reference for the presence or absence of trees within the modelled region. Since the baseline AGB values used to initiate the model simulations were obtained from the CCI 2010 map, resulting into carbon fluxes were as well obtained at 100 m spatial resolution and the simulation period was adjusted to 10 years.

### 6.3.3 Algorithm description

#### 6.3.3.1 Stock change approach

Regarding the 'stock change' approach, ESA Biomass CCI forest maps have been generated using backscatter intensity of SAR data, which captures reflections from both the forest canopy and underlying ground. Then a physically-based model was used that accounts for the individual contributions from the forest canopy and the ground below the canopy to the backscattered signal. Both canopy density and height are linked to AGB through allometric equations to create the maps [54]. The canopy density and height data originate from spaceborne LiDAR observations, while the AGB equations are derived from the GlobBiomass dataset. Along with the CCI AGB maps, uncertainties inherent to the model inputs and AGB retrieval process are provided using a first-order Taylor series approximation. This is included in the final maps as a standard deviation layers. Biomass CCI data sources vary depending on the year, with for instance, 2010's data coming from the Copernicus Sentinel-1 mission, while 2017 relies on the ASAR instrument onboard ENVISAT. JAXA's Advanced Land Observing Satellite (ALOS-1 and ALOS-2) takes over in 2018.

#### 6.3.3.2 Gain-loss approach

Regarding the 'gain-loss' approach, land use-specific carbon emission and removal factors are used to derive an estimate of  $\Delta\text{AGB}$  starting from an initial estimate of AGB for a given baseline year [52,55]. In detail, this approach is based on a modified version of the carbon flux model developed by [56]. This

adaptation was crucial to capture the net exchange of carbon, which refers to the difference between the amount of carbon emitted and absorbed by woody vegetation in a given timeframe. The original model started simulations a decade earlier, in 2000. To address the specific timeframe of interest, we adjusted the starting year to 2010.

### 6.3.3.3 Postprocessing: hybrid method

Finally, we recommend calculating an average map (example Figure 36), calling it hybrid approach, from the two methods. This averages out errors from the two individual, relatively independent approaches, and gives the advantage of reducing the variance in the final map, leading eventually to a more robust estimate. In addition, retrieval uncertainties are given through the standard deviation from both mapping demonstrators, indicating the fidelity and confidence of certain areas within the generated map. Overall, the hybrid map gives a valid carbon stock degradation indicator.

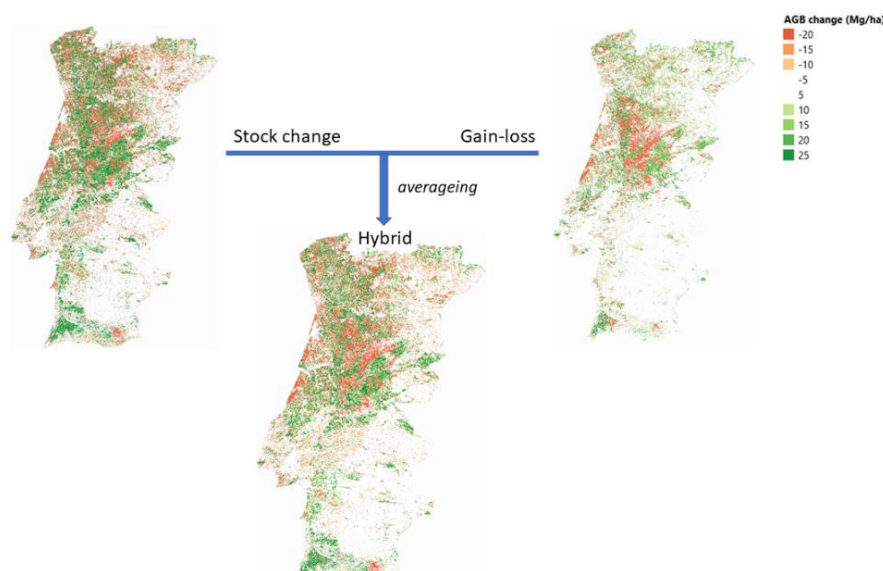


Figure 36: Example of hybrid AGB change mapping method 2010-2018, Portugal case study.

### 6.3.4 Algorithm output variables

The final output products of the carbon stocks algorithms are:

- Stock change AGB map (time span 2010 – 2018)
- Gain-loss Land AGB map (time span 2010 – 2018)
- Hybrid AGB change map (time span 2010 – 2018)
- Uncertainty AGB change map (standard deviation of hybrid approach, time span 2010-2018)

Note that the carbon stock change product serves rather as a prototype demonstration case. The data availability does not correspond to the same time span and resolution as the products generated to monitor trends in land cover and productivity.

The products naming will follow the scheme:

SEN4LDN\_<PRODUCT\_ID>\_<VERSION>\_<DATE\_ID>\_<ROI>.tif

with ROI the region of interest (e.g. country name).

Table 14: Output products of trends in carbon stocks

Product name	Content	Data type and range	No data value	Scale Offset	Unit
Stock change map*	ESA CCI Biomass mapping approach	FLOAT32	-9999	-	mg/ha
Gain-loss map*	WRI Flux model approach	FLOAT32	-9999	-	mg/ha
Hybrid map	Average of stock change and gain-loss maps	FLOAT32	-9999	-	mg/ha
Uncertainty layer	Standard deviation between stock change and gain-loss maps	FLOAT32	-9999	-	mg/ha

\* The Stock change map and Gain-Loss map are currently not distributed to users.

## 6.4 Summary of algorithm validation

Both the stock change and gain-loss methods were validated by [52] using three kinds of reference data sets: (a) re-measured National Forest Inventories (NFI) plot data from diverse European countries, (b) local AGB maps derived in forests with re-measured plot inventories and two airborne LiDAR campaigns between 2010 and 2019, mainly from Brazil, the US and some eastern European countries; and (c) country-level estimates of  $\Delta$ AGB from UN Forest Resource Assessment (FRA) reports, derived by differencing the reported AGB 2018 and 2010 (2018 was computed as the average of AGB 2015 and 2020). However, caution is needed with some of these reference data. For instance, NFIs are not primarily designed for comparisons or validation with spatial products, and the plots thus may not have been properly sampled for comparison with  $\Delta$ AGB at coarser aggregation levels. Figure 37 shows the results when comparing both mapping approaches datasets with these reference datasets at different spatial aggregations as demonstrated in [52].

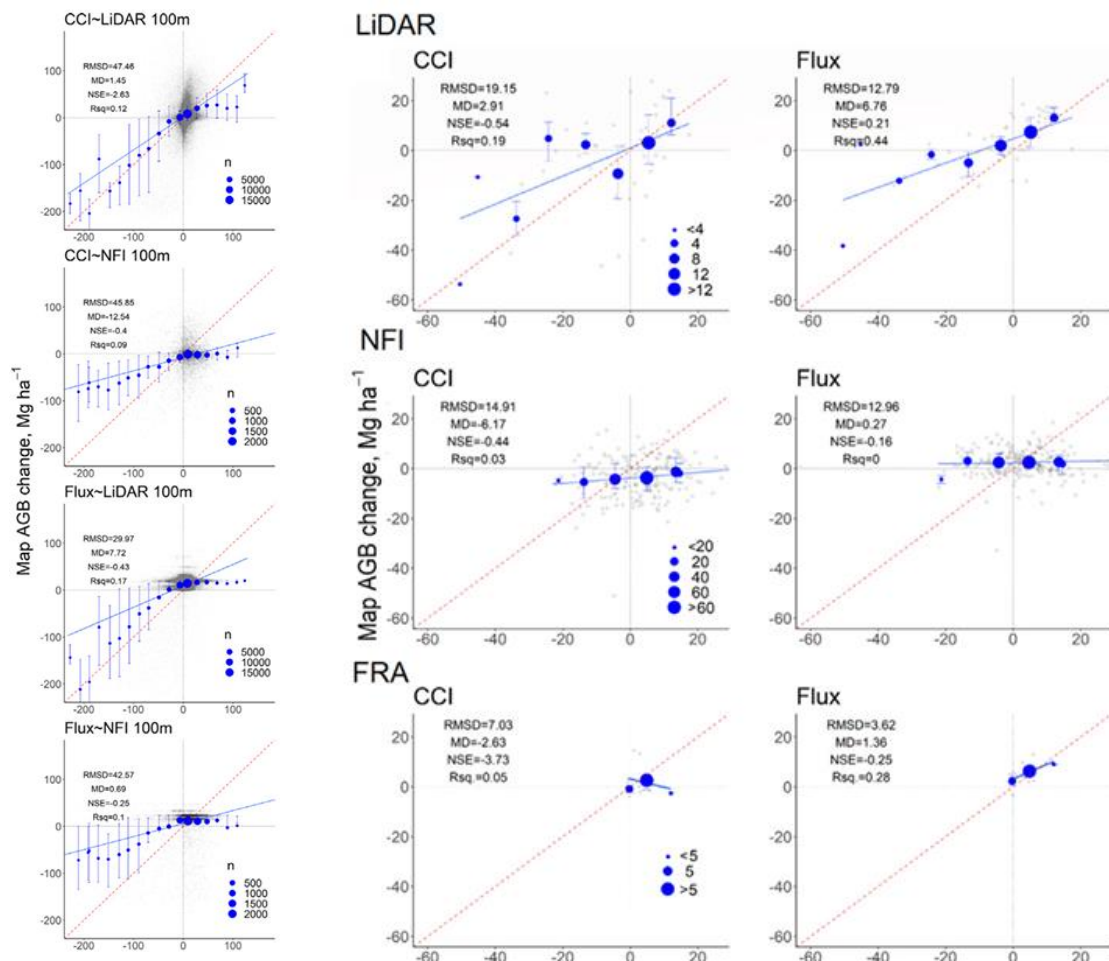


Figure 37:  $\Delta$ AGB comparisons between the 100 m map products (CCI and Flux), and the LiDAR and NFI reference data (left) and comparisons of 25 km  $\Delta$ AGB aggregation between the map products and the three types of reference data [52].

## 6.5 Conclusions, limitations and recommendations

Due to the ongoing absence of high-resolution global trend data of SOC derived from EO data, we explored two valuable alternatives for AGB mapping – the ESA CCI and WRI flux models. While these models provide crucial insights into carbon storage dynamics, it is important to acknowledge potential biases within both the i.e. stock change and gain-loss approaches.

When applying at national levels (Uganda, Portugal and Colombia) both methods showed somewhat different spatial patterns of sinks and sources, yet the overall net flux estimates were rather comparable. Nonetheless, the averaging of both methods has been currently recognized as the best solution, providing national maps with potentially reduced variance and robust estimates. As a next step, for performing validation exercises and improving the validity and transferability of the proposed method, it is highly recommended to incorporate in situ / local reference data.



In summary, the following limitations of the underlying AGB algorithms and stock change models must be considered:

- **Algorithm accuracy:** The carbon flux model might not capture the full complexity of land cover types within the specific country, since the reference data sources under-represent most eco-zones especially in the tropics. This can lead to under or overestimation of flux changes, particularly in areas with diverse ecosystems or land use practices.
- **Spatial resolution:** The resolution of the satellite imagery (ESA CCI maps, 100 m) used to train the models may not be fine-grained enough to detect subtle changes in AGB, especially for smaller land-use features.
- **Temporal resolution:** The relatively long time span between the two map products used (8 years) may not be sufficient to capture rapid or even medium-term changes in AGB, such as those caused by deforestation and fast regrowth events.

In addition, there are some limits regarding the reference and remote sensing data:

- **Limited coverage of reference data:** Reference data used for validation of global maps, i.e. NFI plot data, local  $\Delta$ AGB maps from airborne LiDAR, and selected FRA country data might not be geographically representative of the specific target country. Therefore, local inaccurate estimates of AGB changes may not be discovered.
- **Remote sensing data quality:** Accuracy and consistency of the baseline map (ESA CCI Biomass 2010) may vary depending on the measurements, which further impact the reliability of model calculation and outputs (error propagation).

Future developments could incorporate country-specific plot-based biomass and biomass change measurements to assess data quality and refine ("recalibrate") biomass changes considering specific national circumstances. While such data do exist, the project team did not yet have access to it.

Within a production cycle beyond the project's lifetime, the carbon flux model could utilize the output from the land cover change model as input to ensure alignment with the temporal and spatial resolutions of the other two sub-indicators.

For the near future, the ESA WORLDSOILS project<sup>11</sup> may offer prospects for acquiring the missing (global/European) SOC trend data. However, the ideal solution for a comprehensive understanding of trends in global carbon stocks lies in a combined approach. This would involve utilizing both AGB and SOC maps alongside their respective trends. Such a combined approach would provide a holistic view on global carbon dynamics.

---

<sup>11</sup> <https://world-soils.com/>

## 7 Integration

### 7.1 Outline

The goal of the integration step is to generate a product that allows to calculate the extent of land degradation for reporting on UN SDG indicator 15.3.1, expressed as the proportion (percentage) of land that is degraded over total land area:

$$P_n = \frac{A(Degraded)_n}{A(Total)}$$

Eq. 12

with

$$A(Degraded)_n = A(persistent)_n + A(recent)_n - A(improved)_n$$

Eq. 13

The three sub-indicators – trends in land cover, trends in land productivity, and trends in carbon stocks – are proxies to monitor the essential variables that reflect the capacity to deliver ecosystem services, and are used to delineate  $A(recent)_n$ , area that is degraded during the reporting period, and  $A(improved)_n$ , area that has recovered during the reporting period.

In SEN4LDN, we aim to test two different algorithms:

- The default one-out-all-out (1OAO) integration method, as described in the Good Practice Guidance document [6].
- A continuous sub-indicator integration method, that combines the continuous Land Cover Degradation (see §4.3.4) and Land Productivity Degradation (see §5.3.4) products into a continuous Land Degradation Probability Index. This allows for a more in depth interpretation of the combined product, including an assessment of the magnitude or probability of degradation and an interpretation of possible contrasting sub-indicators.

As this continuous sub-indicator integration method is a novel approach that still needs to be thoroughly tested and discussed, we present the first version of the concept, with some examples over the test areas.

### 7.2 1OAO integration process

In the default set-up, as described in [6], the UN SDG indicator 15.3.1 is calculated by integrating the three sub-indicators (trends in land cover, trends in land productivity and trends in carbon stocks) using a one-out-all-out (1OAO) method, in which a significant reduction or negative change in any one of the three sub-indicators is considered to comprise land degradation. In this methodology, the indicator is reported as a binary quantification (i.e. degraded/not degraded). If one of the sub-indicators is declining or negative (or stable when degraded in the baseline or previous reporting period) for a particular land unit, then it may be considered potentially degraded. In the current setup, SEN4LDN uses only two sub-indicators in the 1OAO aggregation method: land cover degradation and land productivity degradation (Table 15).

Table 15: Combination of land cover degradation and land productivity degradation sub-indicators using the 10AO integration method

Land Cover Degradation	Land Productivity Degradation	⇒	10AO
Improving	Improving		Improving
Improving	Stable		Improving
Improving	Declining		Declining
Stable	Improving		Improving
Stable	Stable		Stable
Stable	Declining		Declining
Declining	Improving		Declining
Declining	Stable		Declining
Declining	Declining		Declining

The input variables for the 10AO algorithm are discrete maps of LCD (see §4.3.4) and LPDclass (see §5.3.4). The output variable is a discrete map delineating pixels that have a positive, stable or negative evolution in the reporting period. Figure 38 shows an example of the 10AO integration principle applied over a 2.5 x 2.5 km<sup>2</sup> area in Portugal where deforestation processes and vegetation establishment have occurred in the time frame 2018-2023.

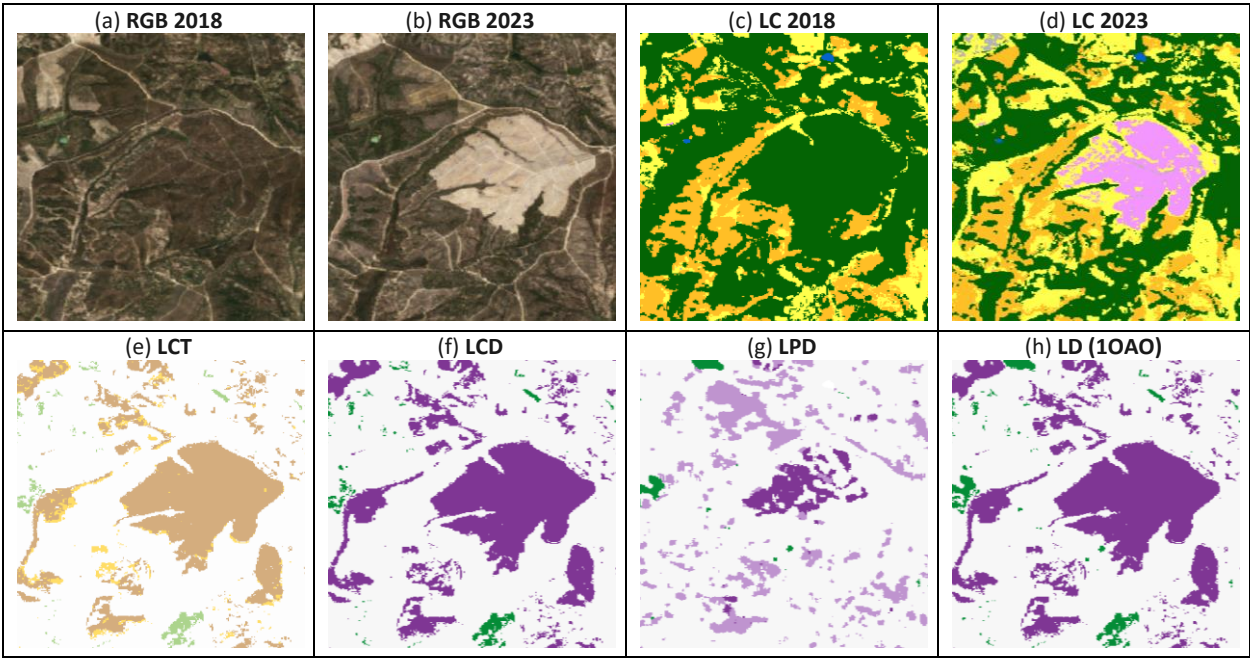


Figure 38: Example on area of 2.5km x 2.5km in Portugal. (a, b) Sentinel-2 RGB Median composite for years 2018 and 2023. (c, d) Land Cover maps 2018, 2023. (e) Land Cover Transition map (LCT). (f) Land Cover Degradation map (LCD). (g) Land Productivity Degradation map (LPD). (h) Land Degradation indicator obtained through integration of LCD

and LPD with the 10AO method, as in Table 15. The indicator highlights 'Improved' areas in green, 'Degraded areas in purple, and 'Stable' areas in light grey.

### 7.3 Continuous sub-indicator integration method

In contrast to the 10AO integration method, where discrete degradation sub-indicator products are used as input (see above), for the continuous sub-indicator integration method continuous maps of LCD-PROB (see §4.3.4) and LPDval (see §5.3.4) are considered as input.

Two different methods were tested for the continuous sub-indicator integration method: (i) the average of LCD-PROB and LPDvalue, and (ii) the absolute maximum of LCD-PROB and LPDvalue (either positive for improving pixels, or negative for degrading pixels). In both cases, the output variable is a continuous map delineating the probability of the pixel to have experienced a positive or negative evolution in the reporting period. A user customizable threshold can be used to convert this map into a discrete map and to delineate  $A(\text{recent})_n$  and  $A(\text{improved})_n$ , similar to the output of the 10AO (see above).

Figure 39 shows an example of both methods applied over the same area in Portugal. The average method (bottom left) results in an attenuated signal. The product can provide a measure of uncertainty of the integrated map: where both input indicators disagree (degradation vs. improvement), the average value will be close to zero, i.e. either degradation or improvement is less likely to have occurred; high absolute values in contrast indicate low uncertainty in the degradation/improvement signal. The absolute maximum (bottom right) retains the magnitude of the signal and highlights areas with high probability of degradation resp. improvement.

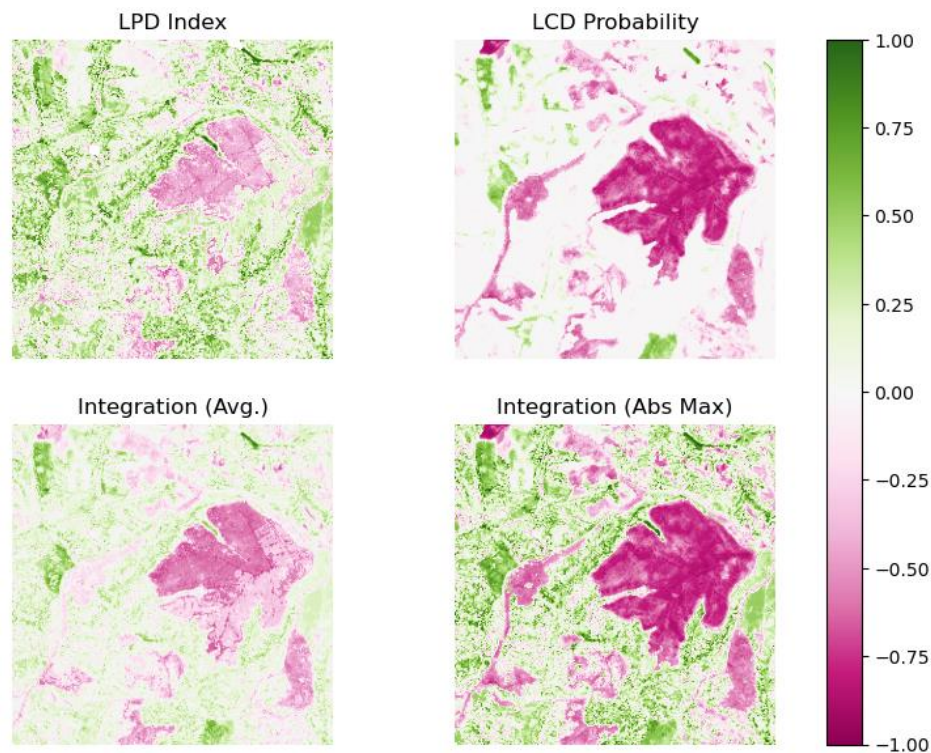


Figure 39: Example of continuous sub-indicator integration. Top left: LPDval. Top right: LCD-PROB. Bottom left: simple average. Bottom right: maximum absolute value.

## 7.4 Algorithm output variables

Table 16: SEN4LDN products on the LDN indicator

Product	Content	Data type and range	No data value	Scale Offset
LD	Land Degradation Neutrality Map with classes indicating Stable (0), Improvement (1), Degradation (2)	UINT8 [0,1,2]	255	-
LD-INDEX*	Land Degradation Neutrality Index map scaled between -1 (degradation) and 1 (improvement)	UINT8 [0-250]	255	0.008 -1

\* The LD-Index is currently not available to users.

## 7.5 Conclusions, limitations and recommendations

Both the 10AO principle and a customizable continuous sub-indicator aggregation method were implemented and tested. Further discussion on the advantages and disadvantages of these methods, and identification of possible improvements, is needed.



## References

- [1] S. Feng, W. Zhao, T. Zhan, Y. Yan, and P. Pereira, "Land degradation neutrality: A review of progress and perspectives," *Ecol Indic*, vol. 144, no. August, p. 109530, Nov. 2022, doi: 10.1016/j.ecolind.2022.109530.
- [2] K. Anderson, B. Ryan, W. Sonntag, A. Kavvada, and L. Friedl, "Earth observation in service of the 2030 Agenda for Sustainable Development," *Geo-spatial Information Science*, vol. 20, no. 2, pp. 77–96, Apr. 2017, doi: 10.1080/10095020.2017.1333230.
- [3] L. J. M. Di Gregorio, A., and Jansen, "Land Cover Classification System (LCCS): Classification Concepts and User Manual," *Fao*, vol. 53, p. 179, 2000, doi: 10.1017/CBO9781107415324.004.
- [4] G. Grekousis, G. Mountrakis, and M. Kavouras, "An overview of 21 global and 43 regional land-cover mapping products," *Int J Remote Sens*, pp. 1–27, 2015, doi: 10.1080/01431161.2015.1093195.
- [5] E. F. Lambin, H. J. Geist, and E. Lepers, "Dynamics of Land-Use and Land-Cover Change in Tropical Regions," *Annu Rev Environ Resour*, vol. 28, no. 1, pp. 205–241, Nov. 2003, doi: 10.1146/annurev.energy.28.050302.105459.
- [6] N. C. Sims *et al.*, *Good Practice Guidance for Sustainable Development Goal (SDG) indicator 15.3.1, Proportion of land that is degraded over total land area - Version 2.0*. United Nations Convention to Combat Desertification, Bonn, Germany, 2021. [Online]. Available: [https://catalogue.unccd.int/1768\\_UNCCD\\_GPG\\_SDG-Indicator-15.3.1\\_version2\\_2021.pdf](https://catalogue.unccd.int/1768_UNCCD_GPG_SDG-Indicator-15.3.1_version2_2021.pdf)
- [7] N. C. Sims *et al.*, "Developing good practice guidance for estimating land degradation in the context of the United Nations Sustainable Development Goals," *Environ Sci Policy*, vol. 92, pp. 349–355, Feb. 2019, doi: 10.1016/j.envsci.2018.10.014.
- [8] R. S. Defries and J. R. Townshend, "ndvi-derived land cover classifications at a global scale," *Int J Remote Sens*, vol. 15, no. 17, 1994, doi: 10.1080/01431169408954345.
- [9] E. Bartholomé and A. S. Belward, "GLC2000: A new approach to global land cover mapping from earth observation data," *Int J Remote Sens*, vol. 26, no. 9, pp. 1959–1977, May 2005, doi: 10.1080/01431160412331291297.
- [10] O. Arino *et al.*, "The globcover initiative," *European Space Agency, (Special Publication) ESA SP*, no. 597, pp. 171–175, 2005.
- [11] M. A. Friedl *et al.*, "Global land cover mapping from MODIS: algorithms and early results." [Online]. Available: [www.elsevier.com/locate/rse](http://www.elsevier.com/locate/rse)
- [12] ESA, "ESA Climate Change Initiative - CCI Land Cover products." [Online]. Available: <https://www.esa-landcover-cci.org/?q=documents>
- [13] R. Tateishi *et al.*, "Production of global land cover data – GLCNMO," *Int J Digit Earth*, vol. 4, no. 1, pp. 22–49, Jan. 2011, doi: 10.1080/17538941003777521.
- [14] M. Buchhorn, M. Lesiv, N. E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus global land cover layers-collection 2," *Remote Sens (Basel)*, vol. 12, no. 6, 2020, doi: 10.3390/rs12061044.



- [15] P. Gong *et al.*, “Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017,” *Sci Bull (Beijing)*, vol. 64, no. 6, 2019, doi: 10.1016/j.scib.2019.03.002.
- [16] C. F. Brown *et al.*, “Dynamic World, Near real-time global 10 m land use land cover mapping,” *Sci Data*, vol. 9, no. 1, 2022, doi: 10.1038/s41597-022-01307-4.
- [17] D. Zanaga *et al.*, “ESA WorldCover 10 m 2020 v100,” 2021. doi: 10.5281/zenodo.5571936.
- [18] D. Zanaga *et al.*, “ESA WorldCover 10m 2021 v200,” 2022. doi: 10.5281/zenodo.7254221.
- [19] W. Gong, S. Fang, G. Yang, and M. Ge, “Using a hidden markov model for improving the Spatial-temporal consistency of time series land cover classification,” *ISPRS Int J Geoinf*, vol. 6, no. 10, 2017, doi: 10.3390/ijgi6100292.
- [20] P. A. J. Van Oort, “Improving land cover change estimates by accounting for classification errors,” *Int J Remote Sens*, vol. 26, no. 14, 2005, doi: 10.1080/01431160500057848.
- [21] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, “Detecting trend and seasonal changes in satellite image time series,” *Remote Sens Environ*, vol. 114, no. 1, 2010, doi: 10.1016/j.rse.2009.08.014.
- [22] Z. Zhu and C. E. Woodcock, “Continuous change detection and classification of land cover using all available Landsat data,” *Remote Sens Environ*, vol. 144, no. January 2013, pp. 152–171, 2014, doi: 10.1016/j.rse.2014.01.011.
- [23] S. P. Abercrombie and M. A. Friedl, “Improving the Consistency of Multitemporal Land Cover Maps Using a Hidden Markov Model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, 2016, doi: 10.1109/TGRS.2015.2463689.
- [24] X. Li, P. Gong, and L. Liang, “A 30-year (1984-2013) record of annual urban dynamics of Beijing City derived from Landsat data,” *Remote Sens Environ*, vol. 166, 2015, doi: 10.1016/j.rse.2015.06.007.
- [25] H. Kerner *et al.*, “How accurate are existing land cover maps for agriculture in Sub-Saharan Africa?,” *Sci Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1038/s41597-024-03306-z.
- [26] T. Zhao *et al.*, “Assessing the Accuracy and Consistency of Six Fine-Resolution Global Land Cover Products Using a Novel Stratified Random Sampling Validation Dataset,” *Remote Sens (Basel)*, vol. 15, no. 9, May 2023, doi: 10.3390/rs15092285.
- [27] X. Ji *et al.*, “Comparison and Validation of Multiple Medium- and High-Resolution Land Cover Products in Southwest China,” *Remote Sens (Basel)*, vol. 16, no. 6, Mar. 2024, doi: 10.3390/rs16061111.
- [28] H. Boogaard, J. Schubert, A. De Wit, J. Lazebnik, R. Hutjes, and G. Van der Grijn, “Agrometeorological indicators from 1979 to present derived from reanalysis,” Copernicus Climate Change Service (C3S) Climate Data Store (CDS).
- [29] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” Feb. 2018, [Online]. Available: <http://arxiv.org/abs/1802.03426>

- [30] C. Aybar *et al.*, “CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2,” *Sci Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1038/s41597-022-01878-2.
- [31] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2006, pp. 770–778.
- [33] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” 2018, pp. 3–11. doi: 10.1007/978-3-030-00889-5\_1.
- [34] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada., Dec. 2019. doi: 10.48550/arXiv.1912.01703.
- [35] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, Oct. 2016, pp. 565–571. doi: 10.1109/3DV.2016.79.
- [36] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” Jun. 2017.
- [37] P. Xu *et al.*, “Comparative validation of recent 10 m-resolution global land cover maps,” *Remote Sens Environ*, vol. 311, p. 114316, Sep. 2024, doi: 10.1016/J.RSE.2024.114316.
- [38] P. Jönsson and L. Eklundh, “TIMESAT—a program for analyzing time-series of satellite sensor data,” *Comput Geosci*, vol. 30, no. 8, pp. 833–845, Oct. 2004, doi: 10.1016/j.cageo.2004.05.006.
- [39] P. Jönsson, Z. Cai, E. Melaas, M. A. Friedl, and L. Eklundh, “A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data,” *Remote Sens (Basel)*, vol. 10, no. 4, 2018, doi: 10.3390/rs10040635.
- [40] F. O. Akinyemi, G. Ghazaryan, and O. Dubovyk, “Assessing UN indicators of land degradation neutrality and proportion of degraded land for Botswana using remote sensing based national level metrics,” *Land Degrad Dev*, vol. 32, no. 1, pp. 158–172, Jan. 2021, doi: 10.1002/ldr.3695.
- [41] Z. Jiang, A. R. Huete, K. Didan, and T. Miura, “Development of a two-band enhanced vegetation index without a blue band,” *Remote Sens Environ*, vol. 112, no. 10, pp. 3833–3845, Oct. 2008, doi: 10.1016/j.rse.2008.06.006.
- [42] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, “Overview of the radiometric and biophysical performance of the MODIS vegetation indices,” *Remote Sens Environ*, vol. 83, no. 1–2, pp. 195–213, 2002.
- [43] D. K. Bolton, J. M. Gray, E. K. Melaas, M. Moon, L. Eklundh, and M. A. Friedl, “Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery,” *Remote Sens Environ*, vol. 240, p. 111685, Apr. 2020, doi: 10.1016/J.RSE.2020.111685.
- [44] X. Zhang *et al.*, “Generation and evaluation of the VIIRS land surface phenology product,” *Remote Sens Environ*, vol. 216, pp. 212–229, Oct. 2018, doi: 10.1016/J.RSE.2018.06.047.

- [45] E. K. Melaas *et al.*, “Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat,” *Remote Sens Environ*, vol. 186, pp. 452–464, Dec. 2016, doi: 10.1016/j.rse.2016.09.014.
- [46] P. Craven and G. Wahba, “Smoothing noisy data with spline functions,” *Numer Math (Heidelb)*, vol. 31, no. 4, pp. 377–403, 1978.
- [47] F. Tian *et al.*, “Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe,” *Remote Sens Environ*, vol. 260, p. 112456, Jul. 2021, doi: 10.1016/j.rse.2021.112456.
- [48] M. G. Kendall, “A New Measure of Rank Correlation,” *Biometrika*, vol. 30, no. 1/2, p. 81, Jun. 1938, doi: 10.2307/2332226.
- [49] H. Theil, “A Rank-Invariant Method of Linear and Polynomial Regression Analysis,” 1992, pp. 345–381. doi: 10.1007/978-94-011-2546-8\_20.
- [50] R. R. Wilcox, *A guide to robust statistical methods*. Springer Nature, 2023.
- [51] H. B. Mann, “Nonparametric Tests Against Trend,” *Econometrica*, vol. 13, no. 3, p. 245, Jul. 1945, doi: 10.2307/1907187.
- [52] M. G. Kendall, *Rank correlation methods*. Oxford, England: Griffin, 1948.
- [53] M. Herold *et al.*, “The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy,” Jul. 15, 2019, *Springer Netherlands*. doi: 10.1007/s10712-019-09510-6.
- [54] “The Global Observing System for Climate: Implementation Needs GCOS-200.”
- [55] M. Herold *et al.*, “The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy,” Jul. 15, 2019, *Springer Netherlands*. doi: 10.1007/s10712-019-09510-6.
- [56] A. Araza *et al.*, “Past decade above-ground biomass change comparisons from four multi-temporal global maps,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, no. March, p. 103274, 2023, doi: 10.1016/j.jag.2023.103274.
- [57] M. Santoro and O. Cartus, “ESA Biomass Climate Change Initiative (Biomass\_cci): Global datasets of forest above-ground biomass for the years 2010, 2017, 2018, 2019 and 2020, v4,” Apr. 21, 2023, *NERC EDS Centre for Environmental Data Analysis*. doi: 10.5285/af60720c1e404a9e9d2c145d2b2ead4e.
- [58] M. Santoro *et al.*, “Global estimation of above-ground biomass from spaceborne C-band scatterometer observations aided by LiDAR metrics of vegetation structure,” *Remote Sens Environ*, vol. 279, p. 113114, Sep. 2022, doi: 10.1016/J.RSE.2022.113114.
- [59] R. McRoberts, E. Næsset, C. Sannier, S. Stehman, and E. Tomppo, “Remote Sensing Support for the Gain-Loss Approach for Greenhouse Gas Inventories,” *Remote Sens (Basel)*, vol. 12, no. 11, p. 1891, Jun. 2020, doi: 10.3390/rs12111891.

- [60] N. L. Harris *et al.*, “Global maps of twenty-first century forest carbon fluxes,” *Nat Clim Chang*, vol. 11, no. 3, pp. 234–240, Mar. 2021, doi: 10.1038/s41558-020-00976-6.
- [61] M. Weiss and F. Baret, “S2ToolBox Level 2 products: LAI, FAPAR, FCOVER - Version 1.1,” 2016.
- [62] S. Jacquemoud *et al.*, “Remote Sensing of Environment PROSPECT + SAIL models : A review of use for vegetation characterization,” *Remote Sens Environ*, vol. 113, pp. S56–S66, 2009, doi: 10.1016/j.rse.2008.01.026.
- [63] J. B. Feret *et al.*, “PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments,” *Remote Sens Environ*, vol. 112, no. 6, pp. 3030–3043, Jun. 2008, doi: 10.1016/J.RSE.2008.02.012.
- [64] W. Verhoef and H. Bach, “Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models,” *Remote Sens Environ*, vol. 87, no. 1, pp. 23–41, Sep. 2003, doi: 10.1016/S0034-4257(03)00143-3.

## Annex A. CatBoost model benchmarking

This annex summarizes the results of the CatBoost model benchmarking process. The models are trained using a CatBoost classifier with a 'MultiClass' loss and eval functions, a tree depth of 8 and a learning rate of 0.05. Training is stopped at convergence using CatBoost early stopping which is set at 10 rounds, i.e., training is stopped if the test loss function is not improving after 10 iterations.

In the training of the models different sets of features (Table 2) are combined, the model id is obtained by separating the different features groups ids with and "\_". The models benchmarking (Figure 40) enables us to gain the following insights:

- Dimensionality reductions of the S2 features (umap features instead of s2) improves the computationally efficiency of the models at the cost of its accuracy, therefore it was discarded.
- AgERA5 climatic embeddings improve the models accuracy by 5.3%.
- Land Cover Deep Learning features (unet) improve accuracy by 3.4%
- DEM/Lat/Lon features improve by 1.2%
- Highest accuracy is obtained by the 's2\_meteo\_unet\_dem\_latlon' model which is selected as candidate land cover classifier model for the system.

Models were also evaluated on confusion matrices (Figure 41) and spatial accuracy maps (Figure 42), highlighting the improvements of the candidate model.

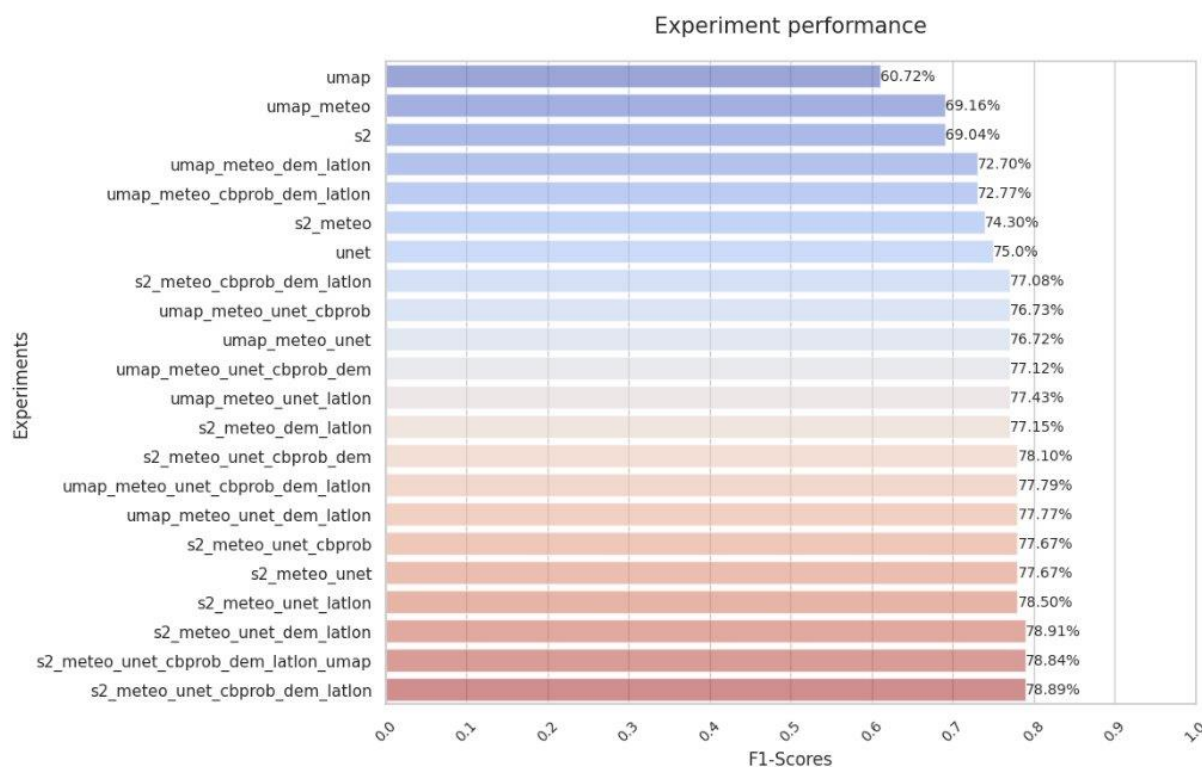


Figure 40: Class distribution of the pixel training dataset obtained through the spatially balanced sampling described above.



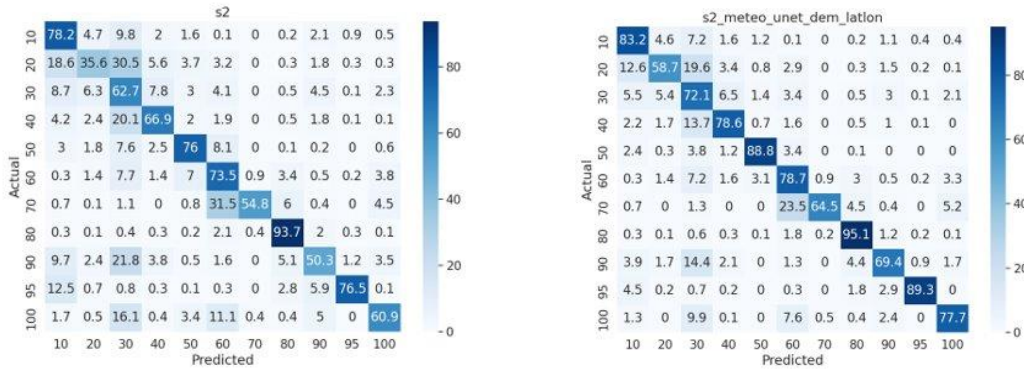


Figure 41: Confusion matrices of the basic model using only L2A features and the candidate model with all the selected features. Accuracy of the s2 model (left) is 69%, while accuracy of the candidate model (right) is 78.9%

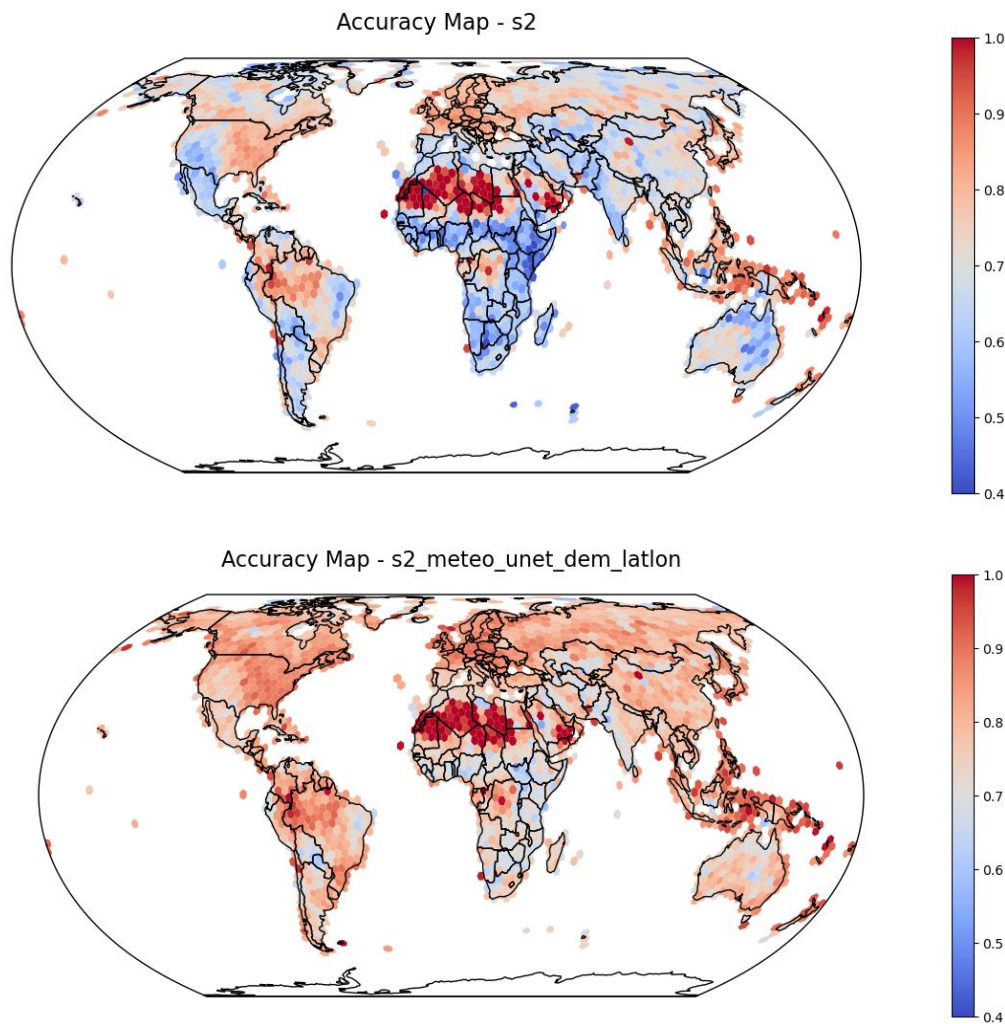


Figure 42: Spatial accuracy map for the base 's2' model (top) and for the final model (bottom), showing improved global accuracy



Figure 43 and Figure 44 show a few examples of the maps computed with different models on selected test sites. Comparing the candidate model, with and without deep learning features, over multiple years highlights their impact on the classification quality (see Figure 45, Figure 46 and Figure 47). It is also possible to qualitatively evaluate the models' robustness across different years and against phenological fluctuations and land cover changes.

Location: 16TDM\_093\_48

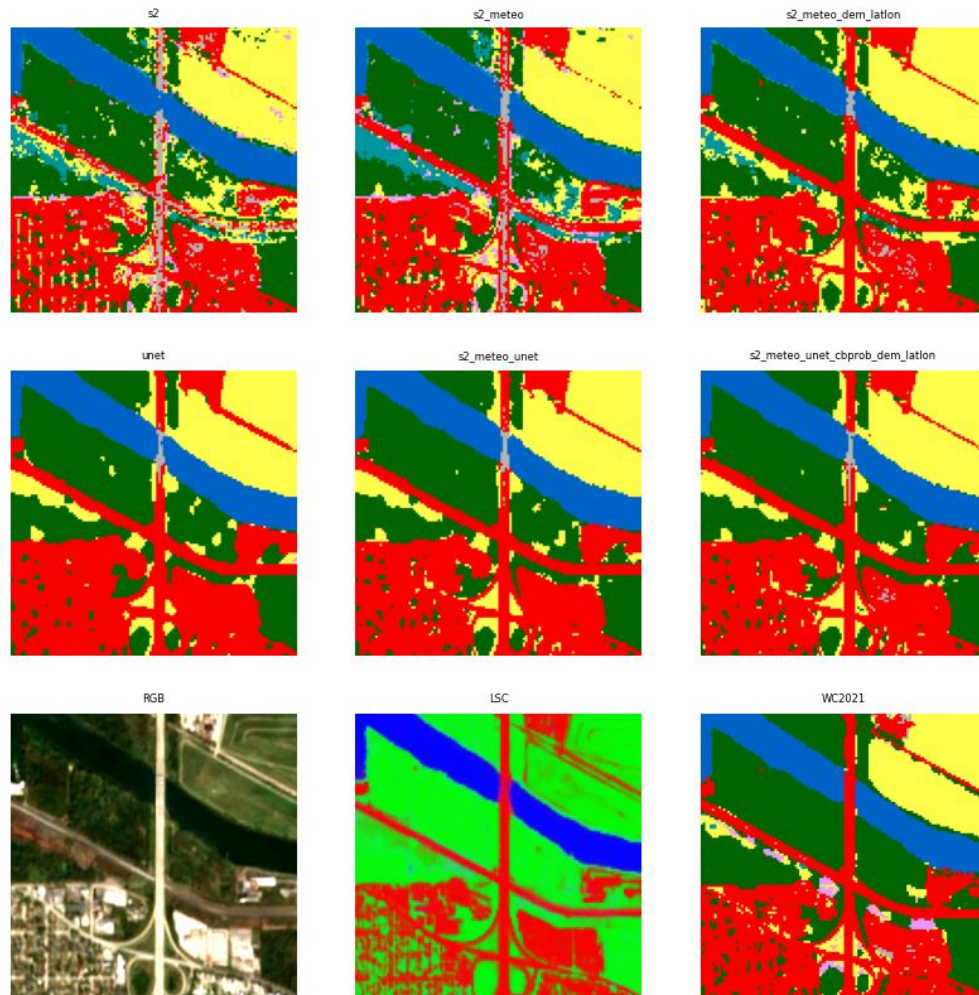


Figure 43: Models predictions for 2021 for test site 16TDM\_093\_48. The candidate model “s2\_meteo\_unet\_cbprob\_dem\_latlon” shows the to match well with WorldCover 2021 (last row, WC2021). The LSC image in the last row is an RGB composite of the 3 “cprob” features, with red, green and blue being respectively ‘not vegetated’, ‘vegetated’ and ‘water’. Adding the unet probabilities reduces salt and pepper noise and helps removing commission of wetland.

Location: 42QWM\_002\_21

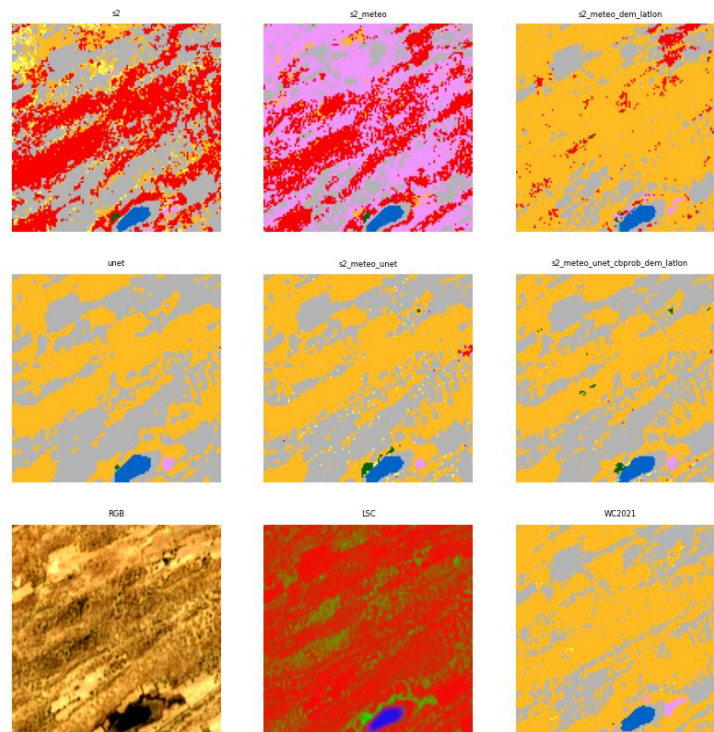


Figure 44: Models predictions for 2021 for test site 42QWM\_002\_21. The candidate model “s2\_meteo\_unet\_cbprob\_dem\_latlon” shows a good match with WorldCover 2021, as well as a significant reduction of built-up commission thanks to the introduction of the deep learning features.

Location: 39UUT\_037\_23

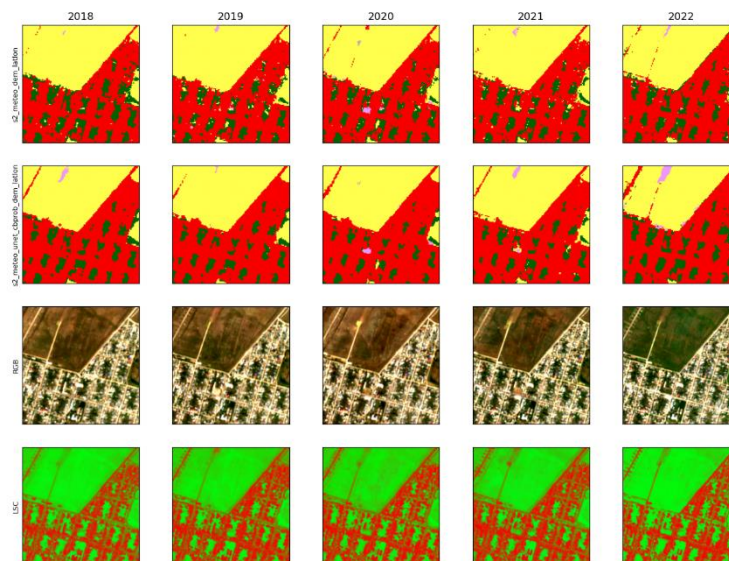


Figure 45: Comparison of the reference model with and without ‘unet’ features. The location does not present land cover changes. However, phenological fluctuations can be observed in the yearly RGB medians. The model is stable

against these fluctuations. The presence of deep learning features improves the classification of built-up class. The road in the upper left corner of the patch is almost always omitted in the model without deep learning features, while correctly classified in the model with them.

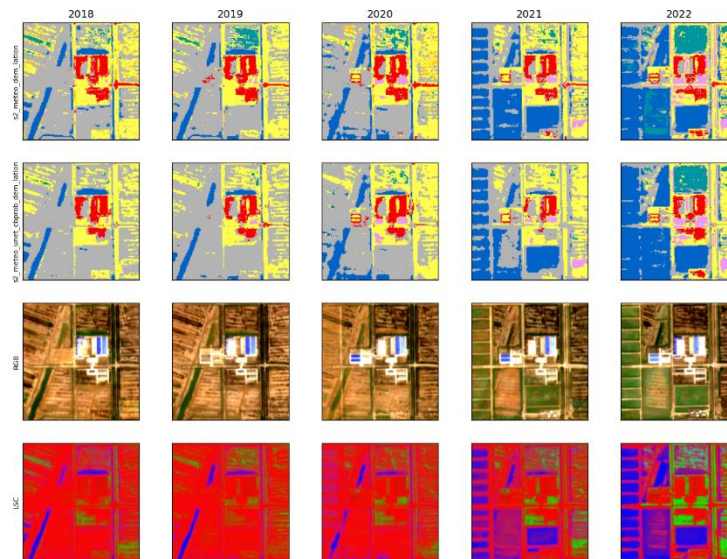


Figure 46: Example of a patch with actual land cover changes. During the years water expansion is correctly classified. Presence of the deep learning features reduce wetland commission.

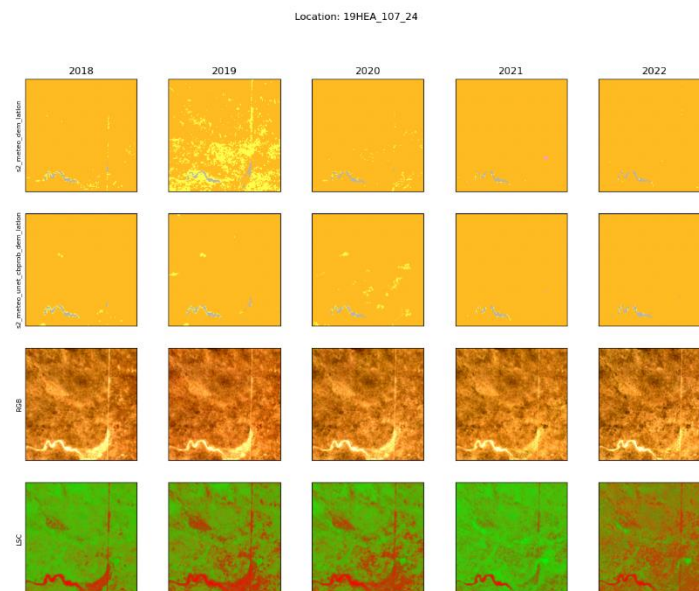


Figure 47: Example showing the improved stability against phenological fluctuations which can cause confusion between shrublands and grasslands in heterogeneous landscapes



## Annex B. Summary of tested vegetation indices

### LAI and FAPAR

Sentinel-2 LAI and FAPAR is based on Weiss et al. [57], implemented in the SNAP Toolbox (SNAP - ESA Sentinel Application Platform v2.0.2<sup>12</sup>). A short summary is provided here.

A hybrid approach combining artificial neural networks and radiative transfer is used to derive LAI and FAPAR. To test and train a neural network, a large database of input and output data is used. The input data are the reflectances of the 3 spectral bands (8 bands used for 20m), and the angular configuration. A large database of input and corresponding output data is simulated using Radiative Transfer Models (RTMs). The RTMs are used to surface reflectance given a set of input variables describing the leaf optical properties (e.g. chlorophyll), the canopy properties (e.g. LAI), the background properties (e.g. background soil reflectance), and the sun and observational geometry. The simulation database was created using the RTM PROSAIL [58] that consists of PROSPECT v5 [59] and 4SAIL [60]. In order to generate a globally representative database of all possible vegetation conditions, the parameterisation of the PROSAIL is based on literature review of all in the input variables used. The neural network is defined by the type of neurons used (the transfer function), the way they are organised and connected (the architecture) and the learning rule. The optimal architecture has been defined by testing several neural networks and compare their outputs with the simulated 'true' values. In addition, networks with less coefficients were preferred. After the training of neural network, its theoretical performance was tested on a test dataset, which is one third of the total simulation database that is not used for training the network. [57] reports for the 8-band network RMSE values of 0.89 for LAI and 0.05 for FAPAR, which demonstrate a good performance of the network. FAPAR shows the best performance, with higher RMSE values for mid-range values of the product. LAI is well estimated up to values of LAI=6, and increasing uncertainties with LAI. Furthermore, the networks are unbiased between the BIOPAR variables as expected. INRAEMMAH provided us, through the Belgian Terrascope platform<sup>13</sup>, the coefficients of the neural networks based on the three 10 m spectral bands. In order to establish these neural networks, the same method as described above was used, however no separate documentation was delivered for this network and hence the theoretical performance is unknown.

The Neural Network is implemented using the Tensorflow library and the processing is done per pixel according the following procedure:

1. The status map is used to check if the pixel is of good quality. If not, the pixel is labelled in the output product as 255. If yes, the pixel is further processed.
2. The input reflectance values are read and normalized according to specifications. The angular information is read and the cosine of the view and sun zenith angle and the relative azimuth is calculated as one set of angles.
3. The neural network is run.
4. Set output values to NoData if outside the definition domain.
5. The output is denormalized according to the specifications.

---

<sup>12</sup> <http://step.esa.int>

<sup>13</sup> <https://terrascope.be/en>

## NDVI

The formula of NDVI is

$$NDVI = \frac{DVI}{NIR + red}$$

Eq. 14

where  $DVI$  is the difference of NIR and  $red$ , and  $NIR$  and  $red$  are reflectance values in bands 4 and 8 of the Sentinel-2 MSI sensor.

## PPI

The formula of PPI (unit:  $m^2 \cdot m^{-2}$ ) is

$$PPI = -K \times \ln\left(\frac{MDVI - DVI}{MDVI - DVI_s}\right),$$

Eq. 15

where  $DVI$  is the difference of NIR and  $red$ .

$MDVI$  is the maximum DVI, which represents infinite leaf layers for a pixel. The steps for estimating the maximum of DVI ( $MDVI$ ) are

1. The temporal maximum of DVI ( $DVI_{max}$ , 10980 pixels  $\times$  10980 pixels) of each pixel is estimated.
2. The temporal 95% quantile of pixel DVI ( $DVI_{Q95}$ , 10980 pixels  $\times$  10980 pixels) of each pixel is estimated.
3. The difference ( $\delta$ ) between  $DVI_{max}$  and  $DVI_{Q95}$  is calculated for each pixel.
4.  $\delta$  is divided into 100 sub-areas, each sub-area with 1098 pixels  $\times$  1098 pixels.
5. The spatial mean  $\delta$  is estimated for each sub-area, then interpolated to the whole Sentinel-2 tile size (10980  $\times$  10980 pixels) by using a bicubic interpolation from the center point of each sub-area, to obtain  $\delta_{pixel}$  (10980  $\times$  10980 pixels).
6. The maximum DVI is computed as  $MDVI = DVI_{Q95} + \delta_{pixel}$ .
7. The upper limit of  $MDVI$  is set to 0.8 to avoid effects of reflectance outliers.

$DVI_s$  is the soil DVI, which is estimated based on the following steps:

1. A Sentinel-2 tile stack (10980 pixels  $\times$  10980 pixels  $\times$  n images) is divided into 100 10.98 km  $\times$  10.98 km image chunks.
2.  $DVI$  is calculated for each chunk from red and NIR reflectance. The ratio ( $p$ ) of the number of vegetated pixels (SCL = 4) to the number of the total valid pixels (SCL = 4 or SCL = 5) in each chunk is calculated.
3. The Soil DVI of each temporal layer in each chunk ( $SDVI_{layer}$ ) is estimated using 5% to 25% quantiles of all valid pixels (SCL = 4 or 5,  $DVI \leq 0.5$ ) in the chunk, and the quantile DVI value ( $Q_{soil}$ ) is chosen based on the ratio  $p$ :  $Q_{soil} = 25\%$  if  $p \leq 0.2$ ;  $Q_{soil} = 5\%$  if  $p \geq 0.8$ ; and  $Q_{soil}$  varies linearly from 25% to 5% if  $p > 0.2$  and  $p < 0.8$ . In this way, a large  $Q_{soil}$  (25%) is used when the

proportion of vegetation pixels is small, and a small  $Q_{soil}$  (5%) is used when the proportion of vegetation pixels is large. The number of  $SDVI_{layer}$  is  $10 \times 10 \times n$ .

4. The soil DVI of each chunk ( $SDVI_{chunk}$ ) is estimated by averaging  $SDVI_{layer}$  in the temporal dimension. The number of  $SDVI_{chunk}$  is  $10 \times 10$ .
5. The soil DVI of each pixel ( $DVI_s$ ) is estimated from upscaling  $SDVI_{chunk}$  to the whole Sentinel-2 tile size (10980×10980 pixels) by using a bi-cubic interpolation from the center point of each chunk.

The  $K$  factor in PPI is a gain factor given by

$$K = \frac{1}{4Q_E} \cdot \frac{1 + MDVI}{1 - MDVI},$$

Eq. 16

where  $Q_E$  is the canopy light extinction efficiency, which depends on leaf inclination angle, solar angle and the diffuse fraction of solar radiation:

$$Q_E = d_c + (1 - d_c) \cdot \frac{G}{\cos(\theta_i)},$$

Eq. 17

where  $G$  is a geometric function of leaf angular distribution, set to 0.5, and  $d_c$  is an instantaneous diffuse fraction of solar radiation, which can be estimated from solar zenith angle,  $\theta_i$ , by

$$d_c = 0.0336 + \frac{0.0477}{\cos(\theta_i)}.$$

Eq. 18

The range of PPI is restricted between 0 and 5 to reduce the noise sensitivity in high PPI values.



## Annex C. Benchmarking of vegetation indices

The analysis to select the best VI for describing vegetation productivity was carried out by cross-correlating values of gross primary productivity (GPP) estimated from eddy-covariance data for 60 ICOS flux tower sites across Europe.

The result of the analysis is shown in Table 17. It shows that EVI has slightly higher correlation than the other VIs.

*Table 17: Correlation coefficients ( $r$ ) between Gross Primary Productivity and smoothed vegetation index data per land cover class.*

	fAPAR	LAI	NDVI	EVI2	PPI
Tree cover (117 site seasons)	0.42	0.47	0.21	0.39	0.33
Grassland (38 site seasons)	0.81	0.86	0.71	0.84	0.81
Cropland (27 site seasons)	0.71	0.51	0.65	0.66	0.37
Wetland (10 site seasons)	0.72	0.63	0.55	0.94	0.76
<b>Average</b>	<b>0.64</b>	<b>0.61</b>	<b>0.50</b>	<b>0.67</b>	<b>0.54</b>